PhD in Computer Science

Research line: Artificial Intelligence

Integrating AI in Assisted Reproduction: Enhancing Fertility Treatments through Deep Learning

PhD Student: Georgios Athanasiou
PhD Advisor: Jesus Cerquides Bueno
Josep Lluís Arcos Rosell
Contact mail: gathanasiou@csic.es

Bellaterra (Cerdanyola del Vallès), May 7, 2024

Abstract

The advent of assisted reproductive technologies marks a revolutionary stride in the annals of biological science, signaling a new era not only in human healthcare but also in wildlife conservation. Infertility, affecting approximately 15% of couples globally, poses a significant challenge to the dream of parenthood. Assisted reproduction emerges as a ray of hope, offering innovative solutions such as In Vitro Fertilization and surrogacy, which have already brought joy to millions of families worldwide. In 2018 alone, IVF procedures resulted in the birth of over 2.5 million babies, reshaping the landscape of modern family dynamics. Expanding the horizon beyond human concerns, the narrative then delves into the world of mammals, particularly endangered species. In this realm, assisted reproduction transforms from medical interventions to instruments of conservation. With over 1,000 mammalian species currently facing the threat of extinction, the application of techniques like artificial insemination and embryo transfer becomes pivotal, underscoring their role in maintaining the delicate balance of our ecosystems.

At the heart of these advancements lies the flourishing field of artificial intelligence. Artificial intelligence's integration into assisted reproduction is not just innovative; it's revolutionary. Algorithms are enhancing the precision and efficiency of assisted reproductive technologies, from analyzing genetic data to improving success rates of treatments, to predicting the viability of embryos. This convergence of artificial intelligence and assisted reproduction is a prime example of how interdisciplinary approaches are pushing the boundaries of what's possible, offering hope where there was once despair.

This thesis, firstly, embarks on an insightful journey into the role of cumulus oocyte complexes in ensuring healthy oocyte and embryo development in mammals. It inspects closely and thoroughly several aspects such as the expansion during the immature and mature phases, and the cell density. To dissect these biological phenomena, advanced deep learning techniques and neural networks have been harnessed. We developed pipelines capable of evaluating the most effective scoring methods, subsequently enabling the automatic segmentation and estimation of crucial study elements. The findings underscore a compelling link between expansion and density, and the development of a healthy embryo. This revelation holds profound implications for the future trajectory of assisted reproduction, potentially revolutionizing our approach toward enhancing fertility treatments.

Simultaneously, this thesis confronts the challenge of acquiring accurate labels for these complex biological structures. Label acquisition typically involves multiple experts, each bringing their own subjective perspectives to the determination of "true" segmentation labels. This diversity of opinion, while enriching, often introduces a level of noise that can set back the effectiveness of segmentation algorithms. To navigate this issue, we implemented coupled convolutional neural networks, tailored for a small-sized, real-world dataset, necessary for healthy embryo develop-

ment. Our methodology pioneers a novel approach that focuses on areas of high uncertainty. By comprehensively analyzing these zones, we reach deeper insights into the individual characteristics of the structures in question. Having this, we were able to propose a more sophisticated ground truth, one that is refined through the use of maximum likelihood estimation. This approach contributes to the discussion of leveraging machine learning algorithms for segmenting medical images, especially in situations where multiple human annotators are involved.

Through this dual exploration of cumulus oocyte complexes role and the refinement of best methods to disentangle from biases in experts' annotations, this thesis lays the groundwork for significant advancements in the realm of assisted reproduction. By introducing cutting-edge artificial intelligence methodologies to intricate biological processes, we uncover new avenues for research and application that promise to elevate the efficiency and performance of reproductive technologies.

Contents

List of Figures

List of Tables

Acknowledgements

Undertaking a PhD journey is undoubtedly challenging, yet it is an enriching and enlightening experience. Completing it has involved many people from various facets of my life, an array of circumstances, and a multitude of emotions.

I would like to begin by extending my deepest thanks to Josep-Lluis Arcos, who provided me with the chance to pursue this journey. His consistent and productive collaboration throughout his supervision period, his openness to my ideas and concerns, and his ability to inspire professionalism while maintaining kindness have been cornerstones of my PhD experience. I must express my profound gratitude to Jesus Cerquides for stepping into the supervisory role when needed, pushing me beyond my perceived limits, and offering guidance on navigating through challenging situations. My sincere thanks go to Rita Vassena, who instilled in me the drive to demand excellence from myself, to aim for perfection, and to uphold professionalism. Despite our limited interactions, her influence has been profound, teaching me essential skills for both my scientific career and personal growth. I am grateful to Ann Van Soom for her faith in my project and the opportunity to work with her esteemed team, especially with Annelies Raes, leading to significant accomplishments. Lastly, my appreciation extends to Trudee Fair for her unwavering support and for being a sounding board for the various elements of the project, offering guidance at every juncture over these years.

I am also deeply thankful to my colleagues in the EUROVA project, whose passion and willingness to explain to me things I did not have previous knowledge have deepened my love for the fascinating field of assisted reproduction. The enjoyable moments and thought-provoking discussions we shared have greatly enriched my journey. I owe special thanks to those with whom I had the most interaction. To Stella, for her enduring friendship and support over the years, in whom I found a lifetime friend. To Raffaella, for our collaborative spirit, the laughter-filled meetings, her support during my time in Berlin, and the memorable time we shared in Copenhagen. To Cristina, for ensuring that my stay and visits to her lab in Copenhagen were not only smooth but also delightful. To Aimilia, for the valuable discussions and support during our few in-person meetings. To Pritha and Male for a truly enjoyable, though unexpected, day in their lab and the time spent in Loddi. To Julieta, for a memorable walk and philosophical discussion in Brussels. And to each and every one of the team members, though our interactions were brief, the moments we shared are fondly etched in my memory.

I extend my thanks to everyone at IIIA-CSIC, EUROVA, Universitat Autonoma de Barcelona, Department of Internal Medicine, Reproduction and Population Medicine at Ghent University, Laboratory of Reproductive Biology at Rigshospitalet in Copenhagen, Leibniz Institute for Zoo and Wildlife Research in Berlin, and the EUGIN Group. Your collective support has been greatly appreciated.

Last, but most importantly, I want to express my special thanks to my family and friends. Completing a PhD takes years, but it takes many more to build a personality, discover your passions and meanings in life, overcome difficulties, question yourself, reconsider who you are and where you are heading, fall but then stand up again, learn and relearn what is necessary about yourself, others, and life. To my family, for their unconditional love and tolerance of me through my life, the sacrifices they have been through to make my path easier, their mistakes that led me to discover new approaches, for knowing that they are present, even when we live far away from each other.

To my friends, the ones who have always been there, the ones who joined my path in recent years, and those who departed at some point. They comforted me in difficult times, amplified my joy during happy times, believed in me, showed patience through my darkest periods, and through their challenges, made me stronger and wiser. I am thankful for each one of them.

Last, to my partner for most of these years, for her presence, support, patience, inspiration, motivation. She made this journey immeasurably more meaningful.

Life is not about degrees. Life transcends academic achievements. It is more about how we evolve through our choices, the paths we follow, the struggles we face. Life is about the people we encounter, and the people that become part of the bigger journey. I had the luck to have all these individuals, and for this, I am deeply thankful.

Chapter 1

Introduction

In recent years, assisted reproductive technologies (ARTs) have experienced a surge of innovations. Techniques such as intracytoplasmic sperm injection (ICSI) [112] and time-lapse tracking of morphokinetics [96] have significantly boosted the success rate of ART. Similarly, the potential integration of artificial intelligence (AI) in oocyte, embryo, and sperm selection is under rigorous examination. While many studies have explored this, the actual implementation of AI in assisted reproduction clinics remains in its infancy. This hesitancy is due to the myriad of proposed AI techniques for clinic operations, leading to uncertainties while the topic is very sensitive.

1.1 Assisted reproductive technology

ART refers to a group of medical procedures designed to treat fertility or genetic problems, assisting individuals or couples in achieving pregnancy. ART covers a broad spectrum of treatments, including procedures like in-vitro fertilization (IVF), ICSI, cryopreservation of gametes or embryos, and embryo transfer, among others.

The existence and development of ART are deeply rooted in the need to address various infertility challenges faced by individuals and couples. According to the World Health Organization, infertility is defined as a disease of the reproductive system, described by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse[1]. The reasons for infertility can be numerous and complex, encompassing factors like age, lifestyle, physical abnormalities, genetic disorders, or environmental stressors.

In this context, ART emerges as a significant solution, helping those who struggle with natural conception. The purpose of these technologies is binary: to enable the possibility of parenthood for individuals or couples who could not otherwise conceive and to prevent the transmission of certain genetic conditions or diseases.

Firstly, ART makes parenthood attainable for a broad range of people who might struggle with natural conception due to various reasons like low sperm count, ovulation disorders, fallopian tube blockages, or unexplained infertility. ART also offers hope to single individuals or same-sex couples who wish to have biological children.

[1] See Infertility by World Health Organization

Secondly, through techniques like preimplantation genetic diagnosis, ART can help prospective parents avoid passing on certain inherited diseases to their children. By analyzing the genetic profile of embryos before they are implanted in the uterus, doctors can select embryos free of specific genetic disorders, thereby ensuring healthier offspring.

ART also plays a crucial role in the preservation of endangered mammalian species. Many of these species face a high risk of extinction due to habitat loss, climate change, poaching, and diseases. Traditional conservation efforts like protected areas and anti-poaching laws are crucial but can be insufficient. Here, ART comes into play as an innovative tool for biodiversity conservation. Using techniques such as artificial insemination, in-vitro fertilization, and embryo transfer, scientists can assist in reproducing endangered species under controlled conditions. Moreover, cryopreservation allows the storage of sperm and egg cells from these species, creating a valuable genetic resource for future use. These efforts can help bolster population numbers, enhance genetic diversity, and ultimately contribute to endangered species' survival and long-term viability.

Moreover, ART is significantly contributing to the lives of cancer patients, particularly those diagnosed at a young age. Specific cancer treatments, like chemotherapy and radiation therapy, can potentially impact fertility by damaging the ovaries or testes. As a result, cancer survivors may face challenges in conceiving a child naturally after their recovery. However, through techniques such as gamete (sperm or egg) freezing, also known as cryopreservation, individuals have the option to preserve their fertility before undergoing treatment. These preserved gametes can later be used in ART procedures to achieve pregnancy. This offers an invaluable lifeline for cancer patients who wish to have biological children in the future, enabling them to look forward to life beyond cancer with optimism.

Thus, whether it provides an invaluable avenue for individuals and couples to realize their dreams of parenthood or helps preserve our planet's biodiversity while offering hope to cancer patients for a family, the contribution of ART extends beyond addressing human infertility. Its growing role in modern medicine and conservation efforts underscores its importance and potential for the future.

1.2 Artificial Intelligence

AI is a broad field of computer science that simulates human intelligence in machines, programming them to think like humans and mimic their actions. AI can perform learning, reasoning, problem-solving, perception, and language understanding tasks, driving innovation and improvement in many industries. AI is reshaping how we live and work, from virtual assistants like Siri or Alexa to more complex technologies like self-driving cars.

AI can be classified into two types: Narrow AI, designed to perform a narrow task such as voice recognition, and General AI, which can theoretically perform any intellectual task that a human being can. While we have made significant strides in Narrow AI, General AI remains a far-off goal, currently confined to the realm of science fiction. AI contains subcategories, structured as found in Figute 1.1.

1.2.1 Machine learning

Machine Learning (ML) is a subset of AI that allows machines to learn autonomously based on experiences, observations, and ingesting data. Instead of programming specific tasks, ML systems

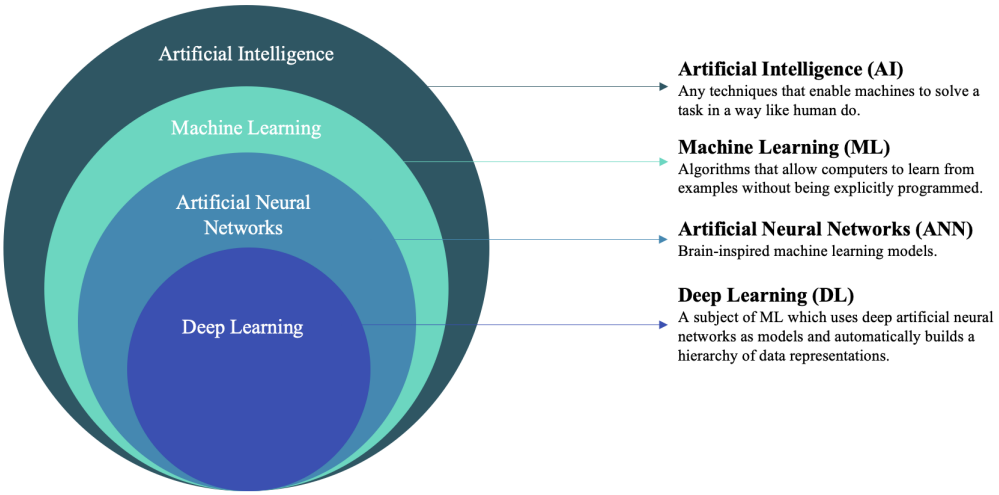


Figure 1.1: The family of Artificial Intelligence, Machine Learning, Artificial Neural Networks, and Deep Learning.

are trained using large amounts of data and algorithms that enable them to learn how to perform the task.

Machine learning models are typically classified into three types: supervised learning, where the model makes predictions based on a set of labeled examples; unsupervised learning, where the model identifies patterns in unlabeled data; and reinforcement learning, where an agent learns how to behave in an environment by performing actions and seeing the results.

1.2.2 Deep learning

Deep Learning (DL) is a further subset of ML based on artificial neural networks with multiple layers - or 'deep' networks. The 'deep' in deep learning refers to the number of hidden layers in the neural network. While a single-layer neural network can still make approximate predictions, additional hidden layers can help optimize the results.

Deep learning models are excellent at recognizing patterns, which they can do after being trained with a massive amount of labeled data. These models power high-profile AI applications, including voice assistants, image recognition, and recommendation systems.

1.2.3 Bioinformatics

Bioinformatics is an interdisciplinary field that combines biology, computer science, information engineering, mathematics, and statistics to analyze and interpret biological data. Bioinformatics involves the computational analysis of genetic and genomic information and other biological data.

AI, ML, and DL are increasingly becoming indispensable tools in bioinformatics, transforming the field by enhancing our ability to analyze and interpret complex and large-scale biological data. The utility of AI, ML, and DL in bioinformatics extends across various sub-domains. In genomics,

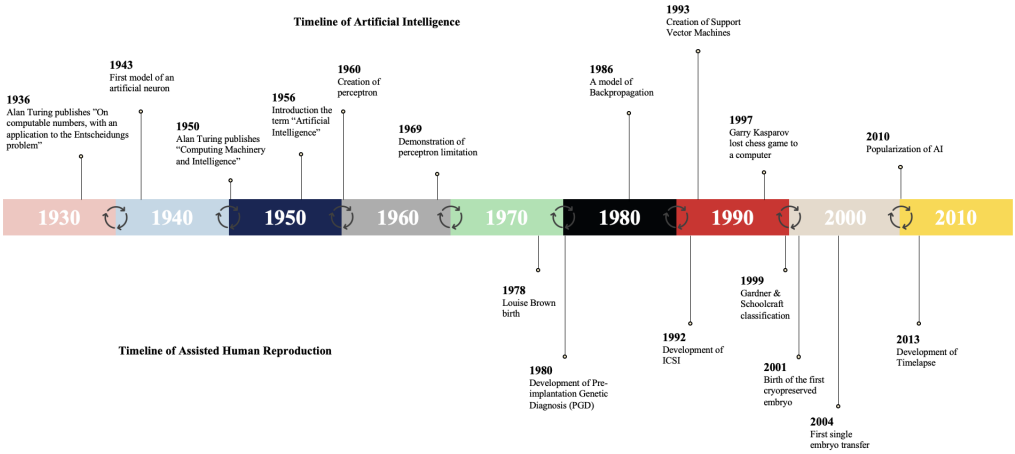


Figure 1.2: Timeline illustrating key events in Artificial Intelligence and Assisted Reproductive Technology.

for example, machine learning algorithms are used to predict gene function and analyze gene expression data, offering insights into the genetic basis of disease.

Similarly, deep learning is revolutionizing protein structure prediction, one of the most critical tasks in molecular biology, enabling us to understand disease mechanisms and develop new drugs. Furthermore, these techniques are used in sequence alignment, phylogenetic analysis, predicting protein-coding regions, and various other tasks that were traditionally labor-intensive and time-consuming.

The future of bioinformatics will continue to be shaped by advancements in AI, ML, and DL. As the volume and complexity of biological data continue to grow, applying these advanced computational methods will be crucial. Integrating these technologies further into the bioinformatics domain will not only enhance our understanding of complex biological systems but also accelerate the discovery and development of new therapies, diagnostics, and potentially even cures for various diseases.

Integrating AI into assisted reproduction clinics' daily operations seems inevitable. In Figure 1.2 by Fernandez et al. [63] there is a timeline of the development of ART along with the development of AI.

1.3 Scope of the research

The intersection of ART and AI presents a groundbreaking frontier in medical science, particularly in addressing challenges in fertility treatments and embryology. The application of AI, especially in the realms of machine learning and deep learning, offers innovative solutions to refine and enhance the effectiveness of ART procedures. This integration is evident in various aspects of reproductive technology, from optimizing the measurement and analysis of cumulus-oocyte complexes (COCs) to developing sophisticated models for predicting embryo viability. Our research delves into this synergy, exploring how advanced AI techniques can revolutionize ART.

The research questions tackled in this dissertation are the following:

1. **Q1:** What methodologies prove most effective and accurate in quantifying the expansion of COCs during mammalian embryo development?
2. **Q2:** Can the measurement of COC expansion be automated using advanced deep learning techniques, and what are the benchmarks for evaluating the performance of such automation?
3. **Q3:** How can a computational model be developed to accurately compute the expansion of COCs across different mammalian species?
4. **Q4:** What is the precise biological role of COC expansion in advancing to a healthy blastocyst stage in mammalian embryos?
5. **Q5:** Beyond COC expansion, what other morphological or cellular characteristics within the cumulus-oocyte complex are critical for achieving a healthy blastocyst stage, and how can these characteristics be quantitatively assessed?
6. **Q6:** How can a novel consensus segmentation method, which emphasizes areas of disagreement among experts, improve the accuracy and objectivity of medical image segmentation compared to existing dual and simultaneous consensus approaches?
7. **Q7:** How does the ratio between the oocyte's area and the follicle's area influence the healthy development of the oocyte, and how can this relationship be effectively modeled within a deep learning framework to enhance our understanding and predictive capabilities?
8. **Q8:** How can transfer learning be effectively applied to the study of early stage embryos across different mammalian species, particularly in cases where data on endangered species is limited, to enhance our understanding and conservation efforts for these species?

1.4 Contributions

This thesis provides several important contributions in the realm of deep learning and reproductive biology:

In collaboration with Ghent University's Department of Obstetrics, Reproduction, and Herd Health[2], a pioneering step was taken in this thesis: the application of deep learning for the detailed study and automation of cumulus-oocyte complexes (Figure 1.3). A cutting-edge deep learning pipeline was introduced, tailored to segment cumulus areas with a precision mirroring human expertise, even within limited resources. This innovative framework not only provides tools for in-depth analysis of COC expansion but also explores sensitive characteristics of the segmented areas, such as cellular density. The insights gathered from this study hold high importance in the realm of predicting the health and development of oocytes and embryos. Furthermore, from an applied standpoint, this pipeline dramatically reduces the manual labor burden on experts, equipping them with indicators rich in biological significance. These contributions correspond to research questions 1,2,3,4,5, and are presented in Chapter 4.

[2]https://www.ugent.be/di/irp/en

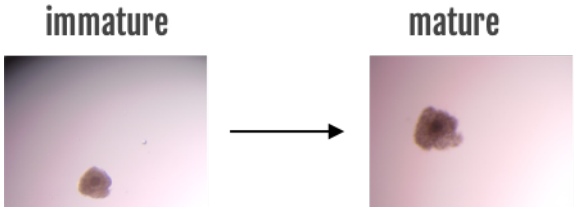


Figure 1.3: An example of an oocyte from the immature stage to the mature stage.

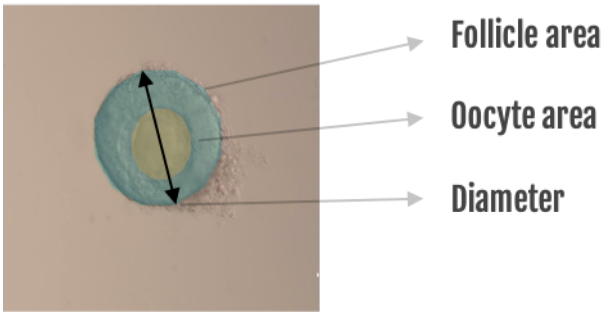


Figure 1.4: An example of an oocyte within a follicle, containing annotations for the features needed for this research.

A recurring and rather frustrating challenge in medical image segmentation is the biases experts carry. This thesis undertakes a comprehensive journey into this issue, concluding in proposing a novel strategy. By deploying a coupled CNN pipeline – a segmentation network and a complementary regularization network – a focused approach was adopted. The highlight of this method is its focus on areas of disagreement amongst experts. By meticulously extracting these diverse expert profiles, our system ingeniously circumvents individual biases during the annotation phase by proposing a new way of estimating the ground truth, disentangled from the human biases in uncertain areas. These contributions correspond to research question 6 and are presented in Chapter 5.

In cooperation with the Laboratory of Reproductive Biology at Rigshospitalet Copenhagen[3], a study was conducted to understand the impact of the ratio between the oocyte's area and the follicle's area on oocyte healthy development (Figure 1.4). Despite the constrained dataset, preliminary results unveiled a weak correlation between the areas and healthy development. However, any other factor available was of less significance than the areas. This finding underscores a need for broader data acquisition to arrive at more definitive conclusions. The attempts in this domain correspond to research question 7 and are presented in Chapter 6.

Lastly, a collaboration was established with the Leibniz Institute for Zoo and Wildlife Research[4]. The mission was to leverage transfer learning techniques for the in-depth study of mammalian embryos, particularly those hailing from endangered species (Figure 1.5). Our exploration confirmed a positive use of transferring knowledge between disparate mammalian embryos, such as naked

[3]https://www.rigshospitalet.dk/english/departments/juliane-marie-centre/fertility-department/laboratory-of-reproductive-biology/Sider/default.aspx

[4]https://www.izw-berlin.de/en/home.html

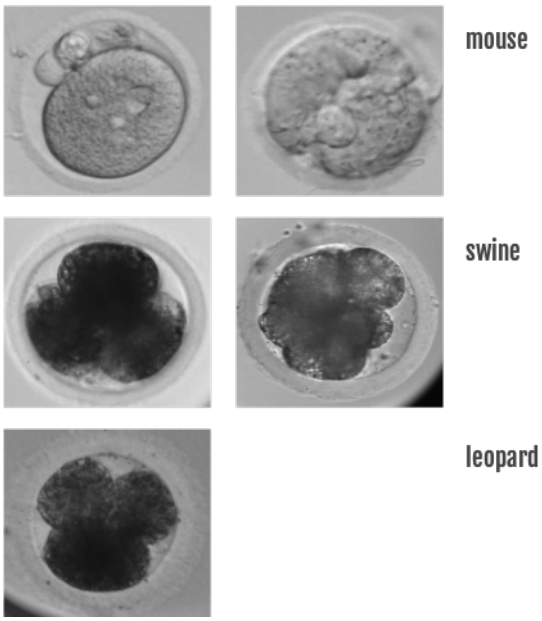


Figure 1.5: A sample representation of different mammalian embryos retrieved by the partners in Berlin.

mole rats and larger mammals or humans. This revelation not only underscores the method's adaptability to many species but also augments research prospects for species with barely sufficient data, like the white rhino, without jeopardizing the integrity of their invaluable data. The attempts in this domain correspond to research question 8 and are discussed in Chapter 6.

1.5 Overview

The structure of this dissertation unfolds as follows: Chapter 2 delves into the foundational aspects of oocyte biology. In Chapter 3, we embark on a comprehensive review of the convolutional neural networks, which play a core role in this research. Chapter 4, serving as the heart of this thesis, is dedicated to the research on Cumulus Oocyte Complexes (COCs). Our exploration into consensus segmentation is detailed in Chapter 5, while Chapter 6 focuses on side projects, including the study of follicle maturation and embryo characteristics across mammalian species. Finally, Chapter 7 wraps up the thesis by summarizing and discussing our major contributions to the field.

Chapter 2

The Language of Oocyte Biology

This chapter provides a comprehensive overview of the fundamental aspects of oocyte biology, which serves as the cornerstone of our current research project. Given the pivotal role of oocyte biology in our study, a thorough exploration of its theoretical foundations was deemed essential. The foundational information presented in this chapter draws upon the extensive research conducted by Gosden et al. [80].

2.1 Introduction

Oocytes hold a crucial role in the mammalian life cycle. They arise from germ cells in the unborn baby and mature within a follicular environment. This environment ensures their development properly aligns with puberty onset, ovulation, and hormone release. It's worth noting that many mammalian oocytes lack efficient developmental capability and are unusually susceptible to chromosomal abnormalities, particularly as females advance beyond their prime fertile years. Consequently, many women may seek ARTs or consider egg donation. Therefore, it is essential that research concentrates on understanding the molecular factors and the particular environmental conditions necessary for the optimal development of oocytes. This potentially enhances the quantity and quality of these techniques, as these factors frequently constrain human fertility.

The mysteries of egg development have intrigued philosophers and scientists since the time of Aristotle, particularly in relation to the development of chicks. However, the small mammalian egg remained a puzzle until its discovery by von Baer in 1826 [80]. Following this discovery, several generations of biologists have come to understand the process of creating haploid gametes through meiosis and their union during fertilization [136]. While it's true that the sperm makes an equal genetic contribution to the resulting zygote, the oocyte plays a far more significant role as the primary donor of cytoplasm. It provides almost all of the organelles and nonchromosomal molecules necessary for development. Even though fertilization signals the beginning of a new genetic being, it's obvious that the process of embryogenesis is heavily dependent on oogenesis, the production of an egg cell.

Oocytes are remarkably specialized to carry out the unique processes of meiosis and fertilization and initiate a molecular blueprint for development. Upon fertilization, they fulfill their reproductive role by generating blastomeres, cells formed by the initial divisions of the fertilized oocyte.

These blastomeres later become precursors to all fetus cell lineages and surrounding membranes. Therefore, oocytes represent a fascinating paradox, being highly differentiated and simultaneously the origin of totipotent cells. However, the scientific interest in oocytes extends beyond their intriguing biology. They determine the boundaries of a woman's reproductive lifespan, are frequently implicated in infertility cases, and play a role in significant birth defects, including Down syndrome. Regrettably, their extreme scarcity and the substantial bioethical considerations they entail, even more so than other cell types except for their fertilization products, hamper research efforts.

It was not until the advent of IVF techniques in the early 1980s that gynecologists regularly started to see living female gametes from our own species. Although IVF procedures have become commonplace and ovarian stimulation is the norm, oocytes are still incredibly rare and valuable. Most oocytes used for research are often donated from IVF programs where they were either not fertilized or immature. Freshly harvested, presumed mature, and fertile cells are seldom available for study. These challenges are further amplified by the oocyte's post-ovulation lifespan, which lasts only a day at most, and the inability to breed oocytes like cell lines. As such, advancements have largely relied on animal models, including non-mammalian species, despite the egg size and polarity variances. The cytoplasm in human oocytes appears comparatively uniform, which has profound implications for invasive ARTs like ICSI. ICSI is a procedure where a single sperm is injected into the oocyte's cytoplasm to address most forms of male infertility. It also affects embryo biopsy for preimplantation genetic diagnosis, where one cell is removed for genetic diagnosis and/or screening for abnormal numbers of chromosomes. Unlike amphibians, flies, and many other animals, where materials deposited during oogenesis visibly polarize the eggs, determine the first cleavage division's plane, and become sequestered in specific early embryo lineages, human eggs lack these visible traits. Another distinct characteristic between species is the absence of active germline stem cells in mammals after birth. Mammalian ovaries are typically thought to slowly use up a finite supply of non-renewable oocytes during adulthood. Human ovarian fecundity is more restricted than in most other species due to the mid-life exhaustion of the oocyte store (i.e., menopause). Additionally, the fertility of oocytes starts to decline dramatically early in mid-life (beginning around the age of 30), when chromosomal abnormalities become exceptionally prevalent [91]. An evolutionary explanation is necessary, given humans' universal occurrence of rapid ovarian aging.

Meiosis in oocytes, a process known as oogenesis, is a critical aspect of female reproduction and begins uniquely during fetal development in females. Unlike males, where sperm production is a continuous process, females are born with a finite number of potential oocytes. Each of these oocytes enters the first stage of meiosis, called prophase I, but then undergoes a pause in this stage. These cells, now termed primary oocytes, remain in a suspended state of development until the female reaches puberty. Once puberty is reached, the process of meiosis in oocytes resumes. Every menstrual cycle, a few primary oocytes reinitiate meiosis. The first stage of meiosis I is completed just prior to ovulation, resulting in two cells of unequal size: a larger secondary oocyte and a smaller first polar body. This stage is crucial as it reduces the chromosome number from diploid to haploid, essential for ensuring the correct genetic composition in the offspring. The journey of the oocyte through meiosis continues with the onset of meiosis II, which begins in the secondary oocyte. However, in a unique twist to the process, the oocyte pauses again, this time in metaphase II, and is ovulated in this state. It's only if fertilization occurs that meiosis II resumes, concluding with the production of a mature ovum and a second polar body. This second division mirrors mitotic cell division, focusing on the separation of sister chromatids. Several unique aspects mark oocyte meiosis. The process involves asymmetrical cytokinesis, where the division results in one

large oocyte and smaller polar bodies, which typically degenerate. Notably, oocyte meiosis is more prone to errors, particularly in chromosome segregation. This susceptibility increases with maternal age and can lead to chromosomal anomalies such as Down syndrome. Furthermore, the extensive duration between the start of meiosis in the fetus and its conclusion potentially decades later is a distinct characteristic of oocyte meiosis. The regulation of this entire process is intricately linked to hormonal changes in the female body. Hormones such as Follicle Stimulating Hormone and Luteinizing Hormone are instrumental in controlling the maturation and release of the oocyte. This precise hormonal regulation is essential for the proper timing and development of the oocytes, ensuring their readiness for potential fertilization and the subsequent restoration of the diploid chromosome number in the zygote. This complex and meticulously timed process underscores the sophistication of the reproductive system in females.

Oocytes collected for IVF or during ovulation are halted at metaphase II, surrounded by a cluster of "cumulus" cells. These cells are a subgroup of the granulosa cells within the follicle, and their main functions include supporting oocyte development and contributing to hormone and growth factor production. This phase marks the final step in a lengthy and intricate journey that starts with primordial germ cells in the epiblast of an implanting embryo [186]. After these cells multiply and migrate to the gonadal anlagen, they divide via mitosis until they enter the meiotic prophase. Consequently, by birth, virtually the entire 1–2 million germ cells, now referred to as oocytes, have arrived at the diplotene stage of prophase I [12]. At this point, they have already undergone meiotic recombination, which gives each oocyte a unique genetic constitution. They then become enveloped in a layer of pregranulosa cells, forming primordial follicles. However, these oocytes cannot ovulate, resume meiosis, or undergo fertilization until they've experienced several weeks of growth within the follicle. During this period, the follicle grows from the tiny primordial stage (around 35 microns in diameter) to the grape-sized Graafian stage, which is ready for ovulation. Later, the oocyte development across the complete range of follicle growth is discussed.

2.2 Development

To fully comprehend oocytes, it is crucial to consider them in conjunction with their follicles, which offer a vital environment for their maturation, making them capable of undergoing fertilization and producing a viable embryo. Follicles, particularly granulosa cells, provide a fundamental niche for oocyte survival, nourishment, and regulation (Figure 2.1). After lying dormant in the ovary for a period ranging from one to over fifty years, primordial follicles commence the growth of their oocyte and pregranulosa cells [178]. This process involves the PI3K signaling pathway but operates independently of the follicle-stimulating hormone (FSH) from the pituitary gland [159]. Only a minor portion of the total oocyte population (seemingly chosen randomly) grows at once; otherwise, the ovary would be quickly exhausted. The growth of the oocyte is synchronized with that of the granulosa cells, which progressively increase their activity in steroid and inhibin production under the impact of gonadotropin stimulation, providing feedback on the hypothalamic-pituitary unit. The granulosa and theca cells play complementary roles in producing estrogens, androgens (and progesterone, which begins shortly before ovulation). Despite the presence of steroid receptors in oocytes, they are not known to be affected by these hormones, except in frogs, where progesterone induces meiosis. Anti-Mullerian hormone is also secreted by granulosa cells, but at an earlier stage and as a paracrine factor impacting small follicle commitment to growth [32]. Previously, oocytes were perceived as passengers in follicles, passive and reliant [78]. However, this metaphor has evolved to view oocytes as captains of their vessels. This

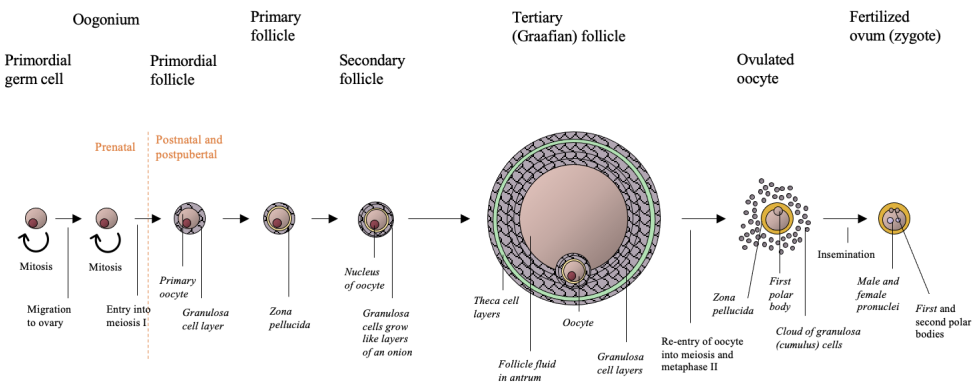


Figure 2.1: Illustrative representation of follicle development in the ovary: Before birth, primordial germ cells travel to the forming ovary and, following a growth phase, initiate meiosis I. The ovary is filled with numerous primordial-stage follicles when a child is born. These have tiny, dormant oocytes encircled by a layer of granulosa cells. As these follicles evolve, they transition through stages until they become preovulatory, or Graafian-stage follicles, housing mature oocytes. These oocytes progress to metaphase II of meiosis due to the mid-menstrual cycle surge of gonadotropins. However, the meiosis process is only finalized when the oocyte is fertilized.

change came with compelling evidence that mid-growth oocytes could accelerate follicle development to the Graafian stage after being combined with granulosa cells from an earlier stage of follicle development [145].

Members of the TGF-β family are some of the most significant secretory products of growing oocytes, notably the growth differentiation factor 9 (GDF9) and the closely associated bone morphogenetic protein 15 (BMP15). These proteins have synergistic roles in follicle growth [51, 230]. BMP15 stimulates the expression of KIT ligand in granulosa cells, which acts via the KIT receptor on oocytes to inhibit BMP15 expression. However, the physiological process is likely more intricate, also involving the theca cell layer [160, 209, 107]. This local feedback loop helps explain why differentiation in follicles where the oocyte has been removed goes awry. When the oocyte is present, granulosa cells are stimulated to multiply and synthesize hyaluronic acid. At the same time, the secretion of plasminogen activator and progesterone is inhibited, partly due to the actions of GDF9 and BMP15. These interactions effectively coordinate the compartments of the follicle to prevent premature ovulation and luteinization. The latter is the process where a post-ovulatory ovarian follicle transforms into a corpus luteum [25, 212, 213, 30].

Although oocytes play a crucial role, communication within follicles is two-way, and the cells within are interdependent. The survival of oocytes depends on metabolic cooperation, as oocytes lack certain metabolic pathways. For instance, they can't use glycolysis to generate energy and instead rely on pyruvate as a primary energy source [90]. Oocytes nearing maturity secrete GDF9, BMP15, and FGF8, which promote the expression of genes in granulosa cells. These genes encode enzymes that stimulate the production of donor metabolites for oxidative metabolism [200]. Similarly, they promote the expression of genes encoding amino acid transporters and enzymes and outsource cholesterol biosynthesis to granulosa cells for membrane growth and other functions [84, 199]. Nutrients and informational molecules can diffuse through the porous zona pellucida, the thick glycoprotein shell secreted by growing oocytes. However, cytoplasmic continuity cre-

ated by heterologous gap junctions is also vital for allowing the bidirectional transport of small molecules between the granulosa cells and oocytes. These gap junctions form at contact points between transzonal projections from granulosa cells and the oocyte membrane or microvilli [2, 180]. If connexin 37, a component of these junctions, is genetically deleted, both follicle and oocyte development stall in mid-growth [193]. FSH, a critical survival factor for granulosa cells, influences the density of transzonal projections, although the mechanism is unknown. This implies that protocols stimulating the ovaries could impact oocyte quality [39]. During conventional stimulation for harvesting multiple oocytes for IVF treatment, endogenous FSH is pharmacologically suppressed to gain greater control of the ovarian response to FSH administration. If oocyte quality is somewhat lower in stimulated than natural cycles, it's not necessarily due to any adverse effects of FSH per se. It's more likely a result of recruiting a wider range of follicle stages than the narrow selection window for the single, dominant follicle in spontaneous cycles.

The crucial harmony of growth and differentiation between oocytes and granulosa cells is vital for the proper timing of maturation, ovulation, and steroidogenesis. These processes are safeguarded by the cellular interactions and inter-dependencies within the follicles, which are the fundamental developmental units of the ovary [58]. Research seeking to culture small oocytes as an alternative to in vivo ovarian stimulation or following frozen banking for fertility preservation must respect the physiological needs of the granulosa cells in which they are nurtured. The structure of the follicle can also help explain why the quality of oocytes (and, consequently, embryos) is not uniform within a cohort, as each oocyte has a unique follicular microenvironment and developmental trajectory. As our understanding deepens, we might see a revival of the ancient debate about whether the intrinsic quality of its oocyte influences the likelihood of a particular follicle ovulating. This indicates that each oocyte and its surrounding environment can greatly influence fertility. The complexity and delicacy of these processes underline the importance of further research in this area.

2.3 Growth and differentiation

Although relatively small, mammalian eggs are much larger than any other somatic (body) cell. For instance, they expand to 120 micrometers in diameter in humans and grow around 100-fold in volume and the number of organelles and structural and soluble components [225, 79]. The enlarged nucleus in these oocytes is called a "germinal vesicle." In this state, the chromosomes are diffused and stained only weakly with basic dyes, making them less visible. While the appearance is different from the "lampbrush" chromosomes of the much larger frog egg – where the chromatin forms brush-like loops acting as factories for RNA synthesis – mammalian oocytes are transcriptionally highly active [43]. This heightened activity level is evident in the large nucleolus, indicating heavy ribosomal RNA production. However, when transcription stops, a cap or "hood" of heterochromatin (tightly packed DNA) forms around the nucleolus in fully grown oocytes that are competent to resume meiosis, which is the process of cell division that results in eggs [144].

The small oocytes found in the earliest stage of development, called primordial follicles, contain a dense aggregation of organelles, including mitochondria, Golgi elements, and the endoplasmic reticulum (ER). This cluster of organelles, reminiscent of the "Balbiani body" observed in non-mammalian eggs, disperses once oocyte growth begins [167]. The Golgi apparatus, an organelle that modifies and transports proteins, breaks into flattened sacs in the cell's periphery. Here, proteins for the zona pellucida (a protective layer around the oocyte) are processed, and cortical

granules (small sacs containing enzymes) are prepared for their role in hardening the zona pellucida after fertilization [201]. This hardening helps prevent the entry of additional sperm, a crucial step in ensuring the normal development of the fertilized egg. The cytoplasm of growing oocytes includes a network of ER and a significant increase in polyribosomes, which are clusters of ribosomes that synthesize proteins. In human oocytes, most centrosomal material needed to form the mitotic spindle in the first embryonic division is not inherited maternally as in mice. Still, it comes from the fertilizing sperm [188]. As a consequence, male infertility resulting from defects in this process is unlikely to be resolved by intracytoplasmic sperm injection (ICSI). During this procedure, a single sperm cell is injected directly into an egg. Oocytes have only small reserves of glycogen (a form of energy storage) and lipids. They lack specific yolk proteins, comparable to vitellogenins in non-mammals, which have disappeared during the evolution of internal fertilization, placentation, and lactation. The only exception is the platypus, an egg-laying mammal that still possesses these yolk proteins [23].

In rodent oocytes, lattice-like structures were once thought to be yolk proteins as they occupy more than 5% of the cytoplasm at maturity, and they gradually disappear after fertilization until they are completely lost in blastocysts (early-stage embryos) [203]. However, recent studies on peptidyl arginine deiminase 6 (PADI6), an enzyme that modifies proteins, have supported an alternative theory suggested years ago. This theory proposes that these cytoplasmic lattices are actually storage sites for the components of the protein synthesis machinery, including mRNA molecules and ribosomes [10, 216]. In line with this, it has been found that these lattices contain PADI6, and when the $Padi6$ gene is disrupted (knocked out), the lattices are completely absent from eggs. These $Padi6$ knockout eggs show defective ribosome function and abnormal activation of the embryonic genome, and they never develop beyond the 2-cell stage, the earliest stage of embryonic development [61, 235]. These findings imply that the lattices are involved in protein synthesis, and the presence of a human version (ortholog) of the $Padi6$ gene suggests that similar mechanisms might exist in human eggs. This new understanding of the role of these lattices could provide valuable insights into the processes involved in early embryonic development.

Oocytes in mammals, including humans, demonstrate radial symmetry, meaning they are symmetrical in all directions around a central point. This is true except for the slight eccentric positioning of the nucleus, which is the central control center for the cell. As the oocyte prepares for ovulation, the meiotic spindle, a structure involved in cell division, moves to the cortex of the oocyte. This repositioning of the spindle defines the plane for forming the first polar body. The first polar body is the small cell that results from the unequal division of the oocyte during the first meiotic division. The side where the polar body is formed is sometimes called the "animal pole." However, this term is technically a misnomer in humans because, unlike in certain animals like frogs, human oocytes do not have predefined cytoplasmic domains destined to develop into specific cell lineages after fertilization. In other words, mammalian oocytes are flexible in their developmental potential, not constrained to a predetermined embryonic axis, a concept referred to as regulative character [109]. This flexibility in mammalian oocyte development is beneficial. For example, in the case of fragmentation, cryoinjury, or biopsy (such as in preimplantation genetic diagnosis where one or more cells are removed from an embryo to test for genetic conditions), the loss of a blastomere (an early embryonic cell) is unlikely to cause maldevelopment. However, significantly reducing the volume of maternal cytoplasm can indeed reduce the viability or ability of the embryo to develop and survive [217].

The zona pellucida is a glycoprotein membrane surrounding the plasma membrane of an oocyte. It is involved in several important processes in the life of an oocyte and early embryo, includ-

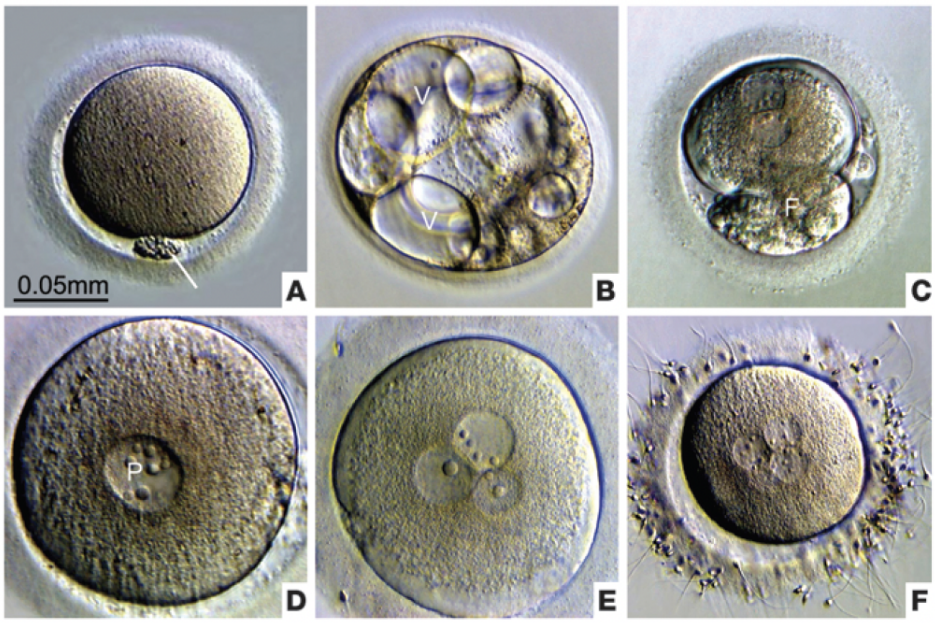


Figure 2.2: A visual display presents 6 human oocytes retrieved post ovarian stimulation for assisted reproductive techniques, spanning both before and after fertilization, with varying health and developmental prospects. (A) A standard oocyte in metaphase II phase, accompanied by its first polar body (highlighted by a white demarcation). (B) An unfertilized oocyte showcasing multiple prominent vacuoles (labeled as V). (C) A pronuclear stage oocyte exhibiting considerable cytoplasmic fragmentation (denoted as F). (D) An oocyte displaying a singular pronucleus (marked as P) following fertilization via ICSI, making it haploid. (E) A zygote with a triploid nature, evident from two dominant and one diminutive pronucleus, hinting at dispermic fertilization. (F) A zygote with tetraploid nature, showing four pronuclei, suggestive of a potential trispermic fertilization occurrence. (Reproduced from Gosden et al. [80].)

ing protection from the external environment, communication with the surrounding cells, sperm binding and induction of the acrosome reaction, prevention of polyspermy, and hatching of the blastocyst during embryo implantation. Regarding protein synthesis, the zona pellucida accounts for more than 5% of the peak output. This membrane contains the primary sperm receptor, zona pellucida 3 (Zp3) [224]. Due to its specificity and function, the zona pellucida has been a key focus of immunocontraception research. In mice, the zona pellucida comprises three highly glycosylated proteins: Zp1, Zp2, and Zp3. These proteins are organized as Zp2-Zp3 fibrils, which are non-covalently crosslinked by Zp1. The expression of these proteins is coordinated by the transcription factor known as FIGLA (factor in the germline α), which is also required for the establishment of germ cells [128] (also present in human [104]). The zona pellucida composition in humans is slightly different, containing an additional protein, ZP4 (found in mice as truncated pseudogene [42]). This makes four proteins: ZP1, ZP2, ZP3, and ZP4, contributing to the structure and function of the human zona pellucida.

In mice, each of the three zona pellucida proteins (Zp1, Zp2, and Zp3) plays a unique role, as evidenced by the different outcomes observed when these proteins are individually knocked out [45]. When Zp1 is absent, the zona pellucida becomes weaker than normal, resulting in reduced fertility in female mice. In the absence of Zp2, the zona pellucida is reduced to a thin membrane

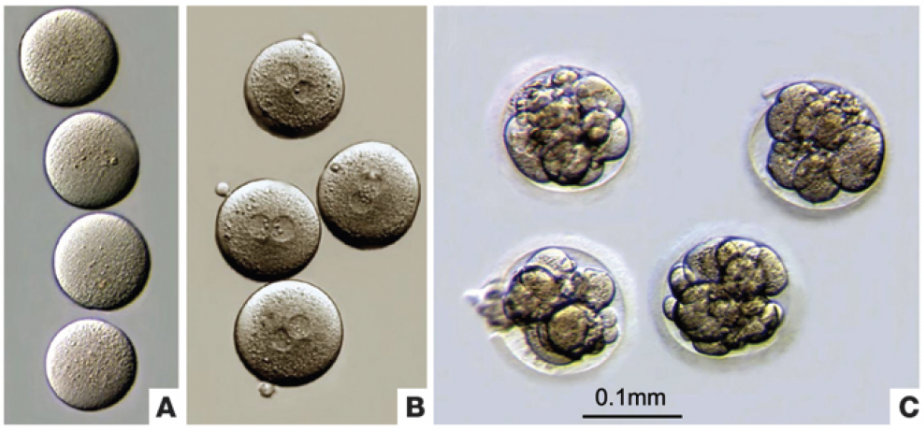


Figure 2.3: A unique instance of mature human oocytes devoid of zona. (A) All the oocytes retrieved from an infertile patient conspicuously lacked the zona pellucida. (B and C) After undergoing ICSI, these eggs exhibited pronuclei and progressed to the 2-4 cell stage when cultured in vitro (B). However, they failed to lead to a sustained pregnancy post-transplantation into the uterus (C), even after being inserted into the evacuated zone. (Reproduced from Gosden et al. [80].)

of Zp1 and Zp3 fibrils. This membrane disappears before ovulation, and the eggs fail to develop afterward. The most severe impact is observed when Zp3 is missing, which entirely inhibits the zona pellucida formation and results in infertility. However, this condition can be remedied by inserting the coding sequence for human Zp3 into the cells. However, the resulting oocytes cannot bind human sperm because they require either Zp4 or species-specific glycosylation [177]. In humans, mutations in these zona pellucida genes are likely rare and haven't been widely reported. At your center, among thousands of patients, you've encountered only one case where every oocyte lacked a zona pellucida in recurring stimulation cycles and a natural cycle [214], seen in Figure 2.3. These "naked" oocytes were fragile, requiring careful handling during follicular retrieval and intracytoplasmic sperm injection (ICSI). Despite reaching the 8-cell stage after fertilization and being transferred into empty donor zonae before being placed in the uterus, successful implantation failed in all cycles except for one, which resulted in a brief pregnancy. Nonetheless, this woman was able to achieve a successful twin pregnancy through egg donation. This case underscores the crucial role of a normal zona pellucida in human fertility, highlighting the significance of these proteins for successful conception.

Judging the competence of living oocytes based on their morphology can be quite challenging, as there are often no visible indicators of issues such as aneuploidy, which refers to an abnormal number of chromosomes [13]. While there are some exceptions, such as noticeable large cytoplasmic inclusions, central granularity (giving a bull's eye appearance), or endoplasmic reticulum (ER) clusters, these features are not always reliable predictors of an oocyte's competence [191, 113, 161]. Sometimes, the oocyte's cytoplasm may appear highly vacuolated (possibly due to blocked ER tubes) or even break up into fragments during meiotic maturation. These phenomena generally predict poor outcomes, either before or after implantation. Aneuploidy is a common occurrence with numerous possible causes, but its presence is difficult to determine based purely on oocyte morphology. The morphology of the oocytes often remains unclear until the surrounding cumulus cells have been dispersed for intracytoplasmic sperm injection (ICSI) or after fertilization. In general, embryo appearance and growth rate provide better indicators of implantation potential than oocyte morphology. This means that embryo development, rather than oocyte appearance, tends

to be a more reliable measure of potential fertility success [57].

2.4 Molecular side

Growing oocytes show a significant increase in mRNA synthesis, peaking when they reach full size. This synthesis stops when the germinal vesicle breaks down, indicating the nuclear membrane's dissolution and the continuation of meiosis after a lengthy pause at prophase I [169, 11]. While newly-formed mRNAs are typically processed and translated swiftly, many in oocytes transform into stored versions. Therefore, gene expression regulation in oocytes shifts predominantly to translation rather than transcription. There's a pivotal time frame starting from when the oocyte resumes meiosis to the embryo's initial cleavage, marking the activation of the embryonic genome. This suggests that the earliest stages of development rely on protein synthesis derived from the stored maternal transcripts. The well-being of the initial embryos hinges on the components and synthetic activity present in the egg. As hinted previously, the differences in mRNA compositions and amounts within a group of oocytes, or due to aging, could provide insights into their viability [164]. Some research indicates that lacking the spindle checkpoint assembly regulator, MAD2, and similar molecules may lead to aneuploidy [197].

From an economic perspective, one might anticipate that gene products would emerge promptly and in amounts directly corresponding to physiological demand. However, reality often defies this expectation. For instance, components of the maturation-promoting factor (MPF), integral to the molecular mechanism propelling the cell cycle, manifest way before oocytes are ready for meiosis. Furthermore, the protein lactate dehydrogenase appears in extremely high quantities, much more than required for carbohydrate metabolism. These apparent discrepancies could, in time, be clarified by understanding that these components might have multiple, diverse functions.

The collection of oocytes formed before birth is designed to last throughout the reproductive life span. Given this, certain germ cells must endure for many decades, and it's likely that over time, they undergo degradation due to the cumulative effects of aging. Once these oocytes are liberated from their follicular surroundings during ovulation, they begin to degrade at a swift pace. From the initial primordial stage to ovulation, these oocytes don't adhere to a strict developmental timeline. Their growth rates can vary, influenced by the particular trajectory of the follicle they're in. They also possess a degree of adaptability in their developmental schedule leading up to ovulation, which is eventually triggered by a surge in luteinizing hormone. For instance, in patients with polycystic ovaries susceptible to ovarian hyperstimulation syndrome (OHSS) due to FSH stimulation, there's an option to "coast" or delay the maturation of the follicles for as long as three days by halting the stimulation. This doesn't seem to affect the quality of the oocytes significantly [126]. The stability of mRNA and proteins likely remains intact during this period.

In oocytes, the lifespan of mRNAs extends to days or even weeks, which is in stark contrast to the typical few hours observed in most body cells. Once mobilized from storage for oocyte maturation, fertilization, and cell division, the vast majority (over 90%) of these transcripts vanish by the time the first cell division occurs. By the blastocyst phase, they're all gone [86]. So, how do these molecules maintain such impressive stability, and what signals their eventual breakdown? Upon being freshly transcribed, mRNA undergoes splicing, resulting in a mature molecule that's then exported to the cytoplasm. Here, it becomes part of a ribonucleoprotein (RNP) complex. Along with some initiation factors, this complex begins forming new polypeptide chains. A polyadenylation

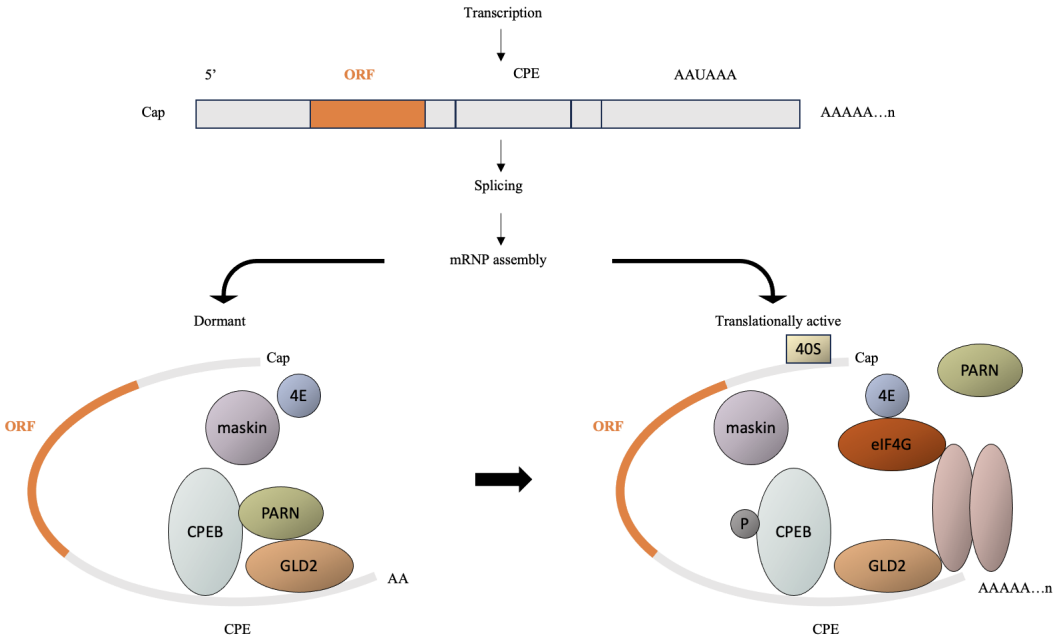


Figure 2.4: Illustrative depiction of the regulation of mRNA translation within RNP structures in oocytes. When a freshly synthesized and spliced transcript gets integrated into an RNP structure, it has two routes: directly translated or remaining translationally inactive for an extended time. Transcripts primed for storage have a CPE located at their 3'-untranslated tail, which becomes the binding site for the CPEB protein. When CPEB gets phosphorylated, it leads to the disengagement of PARN from the complex. This permits GLD2 to extend the poly(A) tail, facilitating translation through interactions with other proteins. For mRNAs in the dormant state, another protein associated with CPEB, termed maskin, obstructs an initiation element (eIF4G). This factor is essential to bring the 40S ribosomal component to the mRNA's 5' end by engaging with the cap-binding initiation factor, known as 4E. The term 40S stands for the 40S ribosomal component.

at the molecule's 3' end is crucial for this translation to occur. However, stored mRNAs experience a shortening of their poly(A) tails to a mere 20–40 nucleotides, rendering them "masked" [150]. The intricate regulation of translation encompasses the following primary components: (i) For storage in oocytes, transcripts have U-rich, cis-acting cytoplasmic polyadenylation elements (CPEs) located ahead of the processing signal, which is denoted as the sequence AAUAAA, with A, U being nucleobases, (A for Adenine, U for Uracil), in the 3' untranslated region [103, 215]. (ii) The CPE-binding protein (CPEB) is vital for the repression or activation of translation; it identifies the CPE in coordination with other factors, which leads us to the third component [179]. Figure 2.4 offers a basic representation of how these molecular participants interact, shifting from a state of translational dormancy to one of expression. This shift is triggered by the Aurora A kinase in frog eggs. When the poly(A) ribonuclease (PARN) is ejected from the complex, the germline development factor 2 (GLD2) extends the tails by 200–250 nucleotides, a necessary step for translation to occur. Furthermore, a protein associated with CPEB, called maskin, hinders translation. It does so by contending with other factors for recruiting the 40S ribosomal subunit to the AUG start codon. While these details are more accessible in species with sizable eggs (like frogs), they likely hold true across a broad range of mammals.

Polyadenylation's critical role becomes evident in the case of Mos RNA, which instructs a key regulator of meiosis. If this protein is missing, oocytes get stuck in meiosis I and don't proceed to meiosis II. However, a workaround exists: when the Mos RNA (with the correct CPE to facilitate polyadenylation) is directly injected into the cells, they can successfully transition to the next phase [71]. Another vital translational regulator is the Stem-loop binding protein (SLBP). It's abundant in oocytes and attaches itself to the 3' region of histone mRNAs, playing a pivotal role in regulating their translation [1]. Histones are essential because they step in to swap out the acidic protamines in the male pronucleus, embedding themselves in the DNA of expanding embryos. A deficiency of SLBP means that histones H3 and H4 can't accumulate. As a result, embryonic development grinds to a halt at the 2-cell stage. Although, this obstacle can be overcome by reintroducing the missing protein [4]. Nevertheless, despite the successes in reversing certain gene product deficiencies, the direct microinjection of mRNAs or proteins into oocytes doesn't seem poised to become a widespread technique to enhance oocyte viability soon. This is due to a few reasons: many vital genes spring into action at earlier, more challenging-to-access phases of oogenesis, and the degree to which transcriptional deficiencies during oocyte growth (or translation deficiencies afterward) impact the overall quality of the oocyte remains ambiguous.

Recently, there's growing evidence that small RNAs play a pivotal role as translation controllers, with many following the same timing patterns as mRNAs that code for proteins. An experiment involving the removal of Dicer, an enzyme responsible for converting precursor molecules into microRNAs (miRNAs), led to significant findings. Without Dicer in oocytes, almost the entire diverse range of miRNAs disappeared. This absence resulted in spindle defects and stopped embryonic development by the 2-cell stage [206]. Moreover, as a defense strategy for the germ line, siRNAs originating from pseudogenes (also processed by Dicer) collaborate with Piwi-interacting RNAs. Through RNA interference pathways, their combined action inhibits mobile genetic elements, specifically transposons [205, 226]. Such revelations are likely just the tip of the iceberg, ushering in a fresh perspective on the intricate mechanisms governing translation in oocytes.

Revisiting the earlier puzzle of how oocytes can simultaneously exhibit differentiation and totipotency, a credible theory suggests their central reliance on RNA, transitioning developmental control from transcription to translation. This temporary separation between the genome, which halts its synthesis, and the cytoplasm, where the pre-stored program undergoes translation, grants the cell a unique developmental flexibility. This allows for specific transcript deployment for dedicated functions, while the chromatin undergoes reshaping for embryogenesis [192]. This concept spans across evolutionary lines, even though the exact timing for the embryonic genome activation varies among species. In humans, this activation is detected during the 4- to 8-cell stage by inhibiting the activity of RNA polymerase [22]. In contrast, it happens 1–2 divisions sooner in mice. Yet, both are notably early compared to frogs, where transcription doesn't restart until the tadpole phase. However, it's worth noting that frogs reach this stage within hours post-fertilization due to their swift cell division.

Imprinted genes, crucial for development, are expressed from only one parental allele. In contrast, the other allele remains inactive due to epigenetic changes, specifically DNA methylation at CpG islands in the 5' region. For instance, the $H19$ gene is unmethylated and thus active in mouse oocytes. In contrast, maternally imprinted genes like small nuclear ribonucleoprotein N (Snrpn) and insulin-like growth factor 2 receptor undergo increasing methylation (thus becoming silent) as the oocyte grows [137]. Interestingly, even if $H19$ and Snrpn expression remains normal in mouse embryos formed post-superovulation, unexpected transgenerational methylation discrepancies appear in the sperm of the resulting male progeny [198]. Further, suboptimal culture conditions can

activate the typically silent paternal $H19$ allele in mouse embryos [50]. This raises concerns regarding in vitro maturation (IVM) of oocytes taken from preovulatory follicles. IVM is an assisted reproductive technique (ART) that largely circumvents the need for gonadotropin-triggered ovarian stimulation, a step in IVF linked with potential health hazards like ovarian hyperstimulation syndrome (OHSS). The concern is especially poignant if the culture initiates during early oocyte maturation phases when epigenetic reshaping occurs. However, despite these concerns, the health of the millions of babies born via ARTs has been largely positive, with imprinting anomalies being extremely uncommon [26].

Evaluating the safety of Assisted Reproductive techniques is complicated, given the naturally high rates of aneuploidy (abnormal number of chromosomes) in oocytes. This aneuploidy rate accelerates with age much faster in oocytes than in sperm or lymphocytes [166]. The key causes of this appear to be weak checkpoint controls and premature segregation of chromatids during metaphase, attributable to deficiencies in meiosis-specific cohesin. These factors shed light on the puzzling susceptibility of eggs [100, 131]. Moreover, human oocytes show moderate sensitivity to ionizing radiation, particularly at the primordial follicle stage. Such radiation can wipe out these follicles through apoptosis after exposure to therapeutic levels of radiotherapy or intensive chemotherapy [218]. An intriguing study on mice treated with ethyl nitrosourea, a strong mutagen, revealed that the predominant infertility effects were observed in males. These males either showcased reduced sperm production or were rendered completely infertile. On a brighter note, these effects seem to be binary in nature. Young cancer survivors of either gender who retain their fertility post gonadotoxic treatments don't demonstrate an increased risk of birthing children with major congenital defects [67].

Few mutations that directly impact the quality of oocytes and embryos have been identified. However, genetic variations might be the cause behind occasional infertility cases where processes like meiotic maturation, fertilization, or cellular division are halted. Maternal effect genes are often viewed as the primary suspects. This is because females carrying mutations in these genes tend to be infertile (despite being otherwise healthy), while males are entirely unaffected. This is largely due to the lack of significant gene product carryover from sperm to the embryo. The proteins these maternal effect genes produce are among the most prevalent in oocytes, and many play a role in processes like transcription and translation. However, some are also present in cortical complexes [242]. In mammals, roughly a dozen such maternal effect genes have been pinpointed, with $Padi6$ being a typical example. Mutations in these genes generally halt the development of mouse oocytes or embryos from the later stages of oogenesis to the blastocyst phase. Some genes with a broader expression can also halt this, but only maternal-effect genes have fertility-specific impacts. To clarify further, although genes like the CCCTC-binding factor and postmeiotic segregation increased 2 ($Pms2$) produce proteins from mouse oocyte mRNAs crucial for cellular division, they aren't labeled as maternal effect genes. These genes also have roles outside of reproduction, such as regulating the epigenome [219] and DNA mismatch repair [83], respectively. Humans probably have similar defects in the counterparts of these maternal effect genes. Such defects could be the reason why a significant number of human embryos have suboptimal quality. There's a possibility that proteins resulting from slightly penetrative genetic variations have only subtle impacts on fertility. However, if multiple such variations are present, their combined effect could be substantial, especially since many maternal effect proteins collaborate in the subcortical complex [127].

2.5 Oocyte maturation

Oocyte maturation is the important transition where a cell evolves into a mature gamete, standing at the crossroads between its own demise and the potential of giving rise to a vast lineage of descendants. Initiated by the surge of gonadotropin around the midpoint of a woman's menstrual cycle, the cell progresses from this trigger, emits a polar body, and continuously advances through the cell cycle, eschewing any hiatus for DNA replication until it reaches its next meiotic division. This division comes to a standstill at metaphase II, roughly a day and a half post the gonadotropin surge, as depicted in Figure 2.5. The first division of meiosis is distinct in that it incorporates genetic recombination and spans several years. In contrast, the subsequent division is brief, mirroring mitosis, and its culmination hinges on successful fertilization. Interestingly, it's been observed that oocytes at the germinal vesicle stage can spontaneously embark on the maturation process in a cultured environment even without the influence of hormones. This autonomous maturation seems to hinge on protein phosphorylation and removing suppressive effects inherent within the follicle. Given the scarcity of mature oocytes in the earlier days of IVF research, early trailblazers in the field often had access only to immature oocytes, prompting them to rely on in-vitro maturation (IVM) [59]. Currently, IVM is witnessing a resurgence, primarily as a method to sidestep the complications associated with ovarian hyperstimulation syndrome (OHSS), especially in individuals diagnosed with polycystic ovaries. While the outcomes in terms of pregnancy rates are commendable, the competency of oocytes derived from IVM trails behind that of conventional IVF cycles. This is likely because procuring oocytes prematurely results in cells that, while they may be prepared for nuclear maturation, haven't fully achieved cytoplasmic maturity [36].

During oocyte maturation, the cell's internal structures are significantly reshuffled and reorganised. Notably: (i) Mitochondria change to ensure localized energy provision. (ii) Cortical granules move in anticipation of preventing the entry of multiple sperm, a phenomenon known as polyspermy. (iii) The endoplasmic reticulum (ER) undergoes shifts, preparing for the sequenced release of calcium ions (Ca^{2+}) from its stores post-fertilization [134, 66, 132]. This maturation journey is primarily powered by cell cycle kinases, most of which also play a role in managing mitotic cycles [110, 24]. MPF (M-phase promoting factor) and MAPK (mitogen-activated protein kinase) are two pivotal molecules in this process. These molecules are central in orchestrating downstream targets and are instrumental in the assembly of the spindle [171, 62, 234]. Their activation is synchronous with the breakdown of the germinal vesicle. As the maturation process unfolds, their activity escalates, though MPF does experience a brief dip post-metaphase I. Both these kinases reach their zenith in activity levels in fully matured oocytes [171, 119, 87] as shown in Figure 2.5. They embody the cytostatic factor that has long been understood to sustain the arrest at metaphase II [143].

Maturation is critical for ensuring proper development and the correct number of chromosomes in the cell. Fertilizing an oocyte prematurely at metaphase I can lead to the formation of cells with three sets of chromosomes, known as triploidy. However, many immature oocytes lack the ability to generate the necessary repeated calcium (Ca^{2+}) signals required for successful activation by a sperm [147, 111, 125]. This signalling capability is only fully developed by metaphase II, but starts declining with age post-ovulation. As the oocyte ages, the signal transitions from one that initiates activation to one that causes cell death or apoptosis [147, 68, 77]. Following fertilization, an initial surge of intracellular Ca^{2+} oscillations repeat approximately every 20-30 minutes. These oscillations set off a sequence of events that includes the release of contents from cortical granules, continuation of meiosis, mRNA utilization, formation of the male and female pronuclei, and

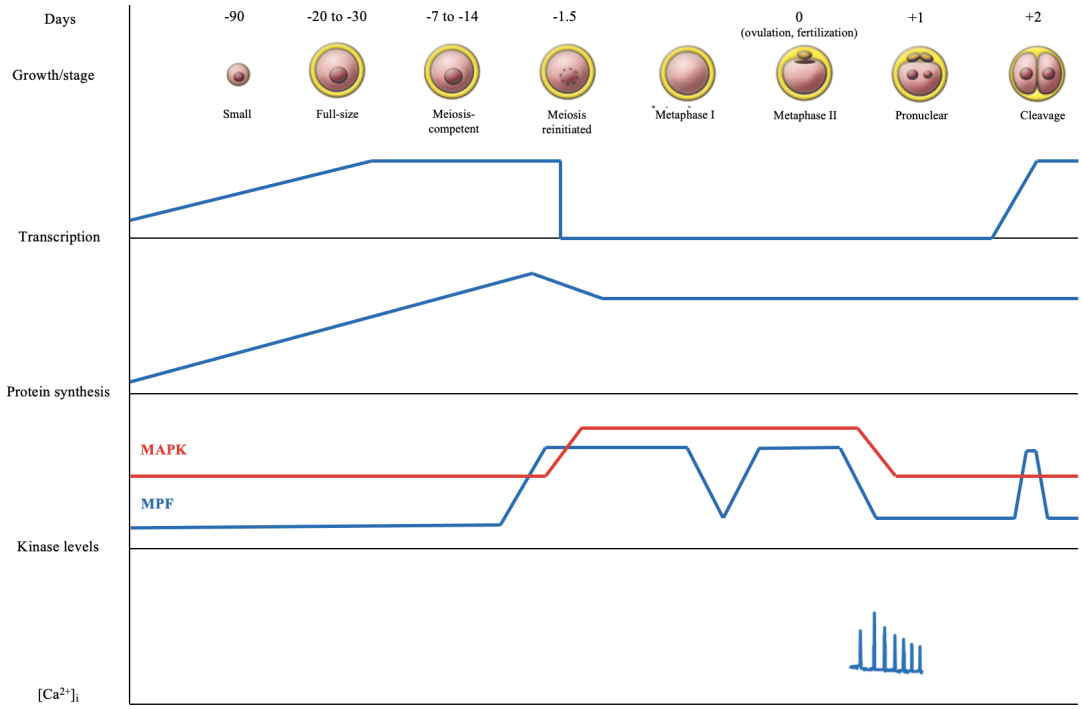


Figure 2.5: Graphical representation of human oocytes' molecular and cellular evolution. The development trajectory begins with the onset of oocyte growth in smaller follicles and spans meiotic maturation, fertilization, and eventual cellular division. Fully grown oocytes can recommence meiosis a few days before ovulation. Post the resumption of meiosis, transcription remains inactive until the embryo reaches the 4- to 8-cell stage. Subsequent gene expression hinges on translating the pre-stored mRNAs until the activation of the embryonic genome approximately 3–4 days later. As the oocyte reaches its full growth, protein synthesis escalates and plateaus. Progressing through meiotic stages is steered by the phosphorylation alterations in both MPF and MAPK. Once the oocyte reaches cytoplasmic maturity, oscillatory patterns of intracellular calcium concentration ($[Ca^{2+}]_i$), sourced from the endoplasmic reticulum (ER) stores, are initiated by the sperm that fertilizes the oocyte.

progression to the first cell division or mitosis [190]. The significance of these Ca^{2+} oscillations was demonstrated in studies where Ca^{2+} injections induced the development of oocytes without fertilization, a phenomenon called parthenogenesis [70]. Conversely, a substance that binds and removes Ca^{2+} prevented fertilization [117, 229]. When electric pulses were used to induce Ca^{2+} release from internal storage areas of the cell, the degree of parthenogenetic development was linked to the number of Ca^{2+} oscillations produced. For instance, the release of cortical granules, an early event, required fewer Ca^{2+} signals than the formation of pronuclei and the onset of mitosis, which are later events [163, 54]. To reach stages beyond implantation, a high frequency of these oscillations was necessary [162]. In some cases, patients undergoing IVF treatments have oocytes that consistently fail to activate post-intracytoplasmic sperm injection (ICSI). This could be due to a failure in Ca^{2+} signaling. To overcome this, a Ca^{2+} ion-inducing substance has been used to encourage fertilization [95]. However, a more natural or physiological method would be preferable, as will be discussed further.

The oscillations in calcium ions ($[Ca^{2+}]_i$) that are triggered by fertilization stem from the phosphoinositide signaling pathway. Here, phosphatidylinositol 4,5-bisphosphate, found on the plasma membrane, gets cleaved by phospholipase C (PLC) into two secondary messengers: inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG) [17]. IP_3 attaches to its receptors (IP_3Rs) on the endoplasmic reticulum (ER) membrane, prompting the release of Ca^{2+} from ER storage into the cell's cytoplasm [243]. On the other hand, DAG aids in the influx of Ca^{2+} by activating protein kinase C (PKC) [158]. In mammalian eggs, the oscillations in calcium following fertilization are primarily mediated by the type 1 IP_3R (known as IP_3R-1) [165, 65]. Evidence for this comes from using IP3R-1 specific antibodies, which have been found to inhibit the calcium response and the activation of the oocyte [229, 153]. If there are irregularities in the abundance, location, or any post-translational changes of IP_3R-1 during oocyte maturation, they could potentially hinder successful fertilization [125, 111, 148]. Addressing irregularities in this signaling pathway presents a challenge. However, in cases where there's a deficiency in the "sperm factor" that initiates these calcium oscillations, one solution could be to inject the sperm-specific form of PLC, known as PLCζ. This could potentially activate eggs in situations where intracytoplasmic sperm injection (ICSI) doesn't work, especially in scenarios related to male infertility [232, 97].

Following the activation of the oocyte, the sperm's nucleus goes through decondensation and replaces its protamines with histones, forming the male pronucleus. This pronucleus gains the potential for future transcriptional activity. The process of decondensation is facilitated by the reduction of disulfide bonds in protamines [168]. This reduction likely involves glutathione, which accumulates during oogenesis, and a product of a maternal effect gene known as nucleoplasmin 2 (NPM2) [29]. In situations where fertilization is halted due to a failure in decondensation, it could result from either an ineffective oocyte or a sperm nucleus resistant to decondensation.

2.6 Future

Human oocytes have transitioned from being relatively unknown to playing a pivotal role in contemporary biology. Even with the rising use of Assisted Reproductive Technologies, the efficiency remains relatively low. For younger women, about 25 oocytes are harvested for every live birth. However, this ratio can vary based on factors such as the treatment facility, maternal age, and whether the oocytes are frozen. A significant reason for this inefficiency is the subpar quality of many oocytes, especially as a woman gets older. Hence, there's a pressing need for advanced methods to discern high-quality oocytes. Screening single cells presents inherent challenges. While non-invasive techniques, like analyzing polar bodies, surrounding cumulus cells, or the culture medium left after fertilization, are promising, their practical applications are limited or still in the experimental phase. It might be more feasible to assess the potential of embryos for implantation than to screen oocytes, given that issues typically become more apparent post-fertilization. Enhancing the quality of oocytes is a compelling approach, but realizing it is complex. Regarding increasing the number of oocytes for medical procedures and research, the current ovarian stimulation methods might have maxed out their potential. However, there's a pressing demand for higher-quality and more abundant oocytes. Meeting this demand could alleviate the strain on egg donation services and encourage the transfer of single embryos, reducing the risk of multiple pregnancies. Furthermore, there's a need for research-quality oocytes to investigate the reasons behind chromosomal abnormalities and to pave the way for regenerative medicine through somatic cell cloning.

In the coming years, due to biological and ethical constraints, the availability and quality of oocytes will be inherently limited. This reality brings forth a mix of great anticipation, tempered by caution, towards advancements via innovative cultivation technologies. These advancements include the potential to cultivate gametes from the relatively more numerous small-follicle stages [81], or from the germline stem cells that continue to exist after birth [28, 246]. To achieve success, the culture setup must be intricate enough to replicate the environment of the follicle accurately. It should also have the capability to provide stage-specific cues for regulating growth, differentiation, epigenetic alterations, and meiotic processes. Standard cultivation methods might fall short in achieving these requirements. However, the evolution of microfluidics technology could pave the way to create a more organic microenvironment. Recent research has managed to cultivate cells resembling oocytes using embryonic stem (ES) cells, and various bodily cells like those from the pancreas, fetal skin, and the outer layer of postmenopausal ovaries [142]. And while it's undeniable that germ cells can develop in embryoid bodies derived from ES cells, present-day culture conditions haven't yet been able to fully simulate the unique and dynamic conditions vital for producing a fertile oocyte – whose true definition lies in its ability to result in a healthy offspring.

The pursuit of enhancing the quality of oocytes by cytoplasmic transfer from younger donor oocytes has been met with skepticism. The reservations stem from the minuscule amount of cytoplasm that can be introduced and the inadequate knowledge regarding the determinants of poor quality and aging processes in oocytes. Indeed, while several babies have been successfully conceived following such procedures [15], worries regarding genetic safety led to the suspension of these practices in the USA. Another method, transferring the germinal vesicle to an enucleated donor oocyte, provides a more extensive cytoplasmic replacement. However, this technique is limited by the efficiency of in-vitro maturation (IVM) protocols and concerns about mitochondrial heteroplasmy. The latter refers to the presence of mitochondria from multiple sources in a single cell, which raises potential concerns for offspring health. Interestingly, the transfer of the spindle-metaphase II chromosome complex to enucleated donor oocytes has shown promise. This method has resulted in the birth of three healthy monkeys [204]. One significant advantage of this method is that since mitochondria are maternally inherited, the risk of passing on mitochondrial DNA mutations to the offspring is virtually nil. This is especially crucial for families with a history of mitochondrial disorders, as it provides a method for ensuring that the child does not inherit these disorders. Nevertheless, a recurring challenge for all these approaches is the restricted availability of donor oocytes. This scarcity underscores the need for continued research and development of alternative, more efficient methods to address oocyte quality and quantity issues.

Despite the challenges, there is a bright horizon for advancements in this domain. With the aid of advanced molecular tools, it's now possible to study individual oocytes – a critical development when examining rare and diverse cells. Establishing an international oocyte bank could propel this field forward, especially given the improved efficacy of low-temperature storage. This means that researchers wouldn't rely on IVF clinics for samples and invaluable cells wouldn't go to waste. Over time, we'll gain a more profound insight into oocyte health and aging, paving the way for innovative oocyte cultivation techniques and even molecular interventions to rectify defects. Furthermore, if induced pluripotent stem cells (iPS cells) continue to show promise, they might become a fresh source of oocytes, offering chances to rectify genetic defects at a nascent stage. iPS cells strongly resemble ES cells but originate from modified somatic cells. Healthy mice, birthed from these cells, have been found to possess fertile germ cells [241]. This suggests the potential for human iPS cells to eventually serve as a germ cell reservoir for producing oocytes. While it's hard to foresee the speed of these advancements or the safety of these gametes for fertility treatments, these explorations enrich our understanding of oocyte biology and any associated disorders.

Chapter 3

The Language of Convolutional Neural Networks

As we transition from the intricate biology of oocytes explored in the previous chapter to the innovative realm of AI in this chapter, we introduce Convolutional Neural Networks (CNNs), a transformative force in computer vision. These networks, mirroring the pattern recognition capabilities of the human brain, provide a novel lens through which we can analyze and interpret the complex imagery of oocytes. Bridging the gap between the detailed biological understanding previously established and the technological prowess of CNNs, this chapter aims to explore how the synergy of biology and advanced computing can unlock new dimensions in the study of reproductive biology, propelling us beyond traditional methodologies to discover groundbreaking insights in the field.

3.1 Introduction

Computational neural models have a rich history, dating back to the mid-20^{th} century, with early foundational works like the Hebb and Perceptron models [94, 182]. A critical moment came with Rumelhart et al.'s introduction of backpropagation for learning family tree structures [184]. Today, backpropagation is the primary method for training diverse neural network architectures. LeCun et al. later utilized backpropagation to train highly complex neural networks, introducing convolution layers for practical applications such as recognizing handwritten digits [123, 124]. It's crucial to note that while these introductions of backpropagation and convolution layers were impactful, they weren't the first of their kind since others have been mentioned thoroughly [189].

Deep Convolutional Neural Networks (DCNNs) are a class of deep learning models primarily used in processing structured array data such as images, leveraging convolutional layers for feature extraction. Their development was significantly influenced by advancements in computational resources and training techniques, allowing them to achieve outstanding performance in tasks like image and video analysis. In a later period, interest in these networks was reduced, mainly because of computational constraints and the absence of large labeled datasets. However, a revived interest emerged when Hinton et al. [98] successfully trained deep layers of an auto-encoder for dimensionality reduction. Advances in computational power, primarily through graphics processing units (GPUs), and a deeper understanding of how to efficiently initiate and train these

networks soon led to the remarkable success of a Deep Convolutional Neural Network (DCNN) named 'AlexNet' [118]. This model excelled in the 'ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)', bringing deep learning to the forefront. ILSVRC is a yearly challenge dedicated to detection and classification tasks [185].

It's essential to recognize that while deep learning often centers around deep neural networks, other deep architectures exist, such as hierarchies based on k-means clustering [38]. Nevertheless, deep neural networks remain the most representative model in the deep learning domain. This thesis uses 'deep learning' to denote 'deep neural networks'.

3.1.1 Definition and Overview

A standard neural network consists of input and output layers, with at least one intermediary layer known as the hidden layer. These layers are made up of units named neurons. Neurons within a layer connect to those of the layers before and after them using connection weights, or simply, network weights. A particular neuron, y_j, calculates a weighted sum of its input values x_i, using corresponding weights w_{ij} and an added bias b_j. Based on the network's design, a linear or non-linear transformation, known as an activation function f, may be applied to this sum, expressed as:

$$y_j = f(\sum_i x_i w_{ij} + b_j). \tag{3.1}$$

The network's performance is measured using a 'loss' function. During the training phase, the objective is to minimize this loss function relative to the set of connection weights W, which consist of the weights w_{ij}. To simplify it, we'll focus on the weights and not include bias in our explanations. The training process typically employs backpropagation, which comprises two iterative procedures: the forward pass and the backward pass. The network's operations boil down to straightforward matrix multiplications if the non-linear activations are removed.

3.1.1.1 Network depth

The depth of a network is usually determined by counting its layers, excluding the input layer. So, a conventional shallow neural network, consisting of an input, an output, and a single hidden layer, is said to be two layers deep. A layer's width is related to the number of neurons it contains. Generally, a network that is both deeper and wider might have a higher capacity to model complex patterns. However, this notion varies based on the neural network's design and the specific problem it addresses. It's worth noting that deeper networks can potentially overfit data. While the distinction between 'shallow' and 'deep' networks can be subjective, networks with several hidden layers are typically labeled 'deep'.

Figure 3.1 provides a generalized depiction of a deep network, specifically a DCNN. This includes two successive convolution layers (alternating with pooling layers), culminating in the fully-connected layers (all layers defined later on in Section 3.2). The convolution layers produce multiple feature maps derived from the preceding layer or the input image. The final convolution or pooling layer's features then lead to a fully connected layer, which calculates a weighted sum of these inputs and typically applies a non-linear transformation, or activation, to them.

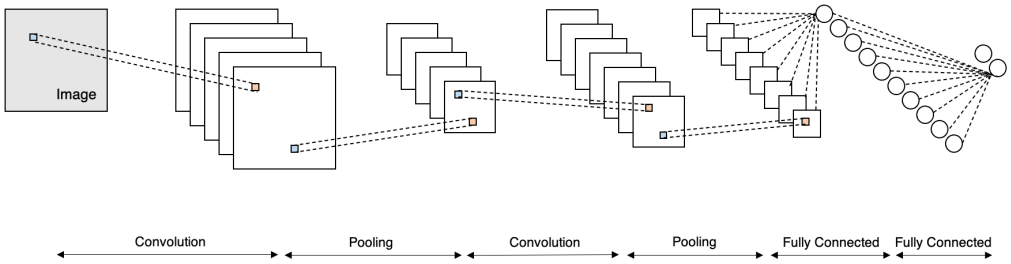


Figure 3.1: A Deep Convolutional Neural Network (DCNN) that consists of two convolutional layers responsible for calculating multiple feature maps. These convolutional layers are alternated with two pooling layers, which achieve a dual purpose: they reduce dimensionality through response pooling and introduce a degree of translation invariance. To conclude, the network incorporates two fully connected layers at the end, which carry out a weighted sum of the feature vectors derived from the convolutional layers.

3.1.1.2 Network orientation

The terms 'up' and 'down' are conventionally used to indicate directions pointing toward the output and input layers, respectively. This means the input layer is thought to be at the base, while the output layer sits at the network's peak. As a result of using back propagation during the training phase, data ascends in a forward pass, while gradient updates descend in a backward pass.

3.1.1.3 Layers, activations and variations in taxonomy

Theoretically, activation functions are deemed network layer components rather than independent layers themselves. But with the surge in the development of deep learning tools and methodologies over the past couple of years, it's not unusual to see instances where activation functions are designated separate layers for implementation purposes [108], even if their operations remain unchanged. This shift has led to a growing trend in literature where activation functions are discussed as 'layers'. Details on frequently utilized network components, like layers, activations, and loss functions, essential for crafting contemporary deep learning workflows, can be found in the following sections.

3.2 Blocks of CNN architecture

The architecture of a CNN is comprised of various components, including convolution layers, pooling layers, and fully connected layers. Generally, a CNN architecture is characterized by repeated sequences of multiple convolution layers paired with a pooling layer, followed by one or several fully connected layers. The process during which the input data undergoes transformation to produce output via these layers is known as forward propagation.

3.2.1 Convolutional layer

A convolution layer is a prevalent layer type in neural networks tailored for computer vision and image processing tasks. Multiple small two-dimensional filters or feature detectors characterize

this layer. Unlike the manually designed features in traditional machine learning methods, the filters within deep networks are refined during training. This approach, where feature and classifier training happen concurrently, is known as representation learning.

These adaptive filters engage in convolution operations with the input image, and the resulting feature outcomes are relayed to the subsequent layer [123]. Networks that sequence multiple convolution layers (often alternated with pooling layers) are called DCNNs.

Extending the AlexNet design philosophy, contemporary advancements lean towards architecting networks predominantly around convolution layers. A case in point is the ResNet [93], a 152-layer deep residual learning network (which excelled at ILSVRC 2015). This structure is predominantly built from convolution layers.

Filter Traditional machine-learning approaches in image processing often involve utilizing predefined feature detectors. These detectors produce responses, which are then passed through a classifier for label determination. In deep networks, the initial convolution layers, positioned closer to the input, act as these low-level feature detectors. Unlike classic methods, the parameters of these detectors, primarily the weights of the kernels, are fine-tuned during the training phase. Commonly, DCNNs deploy a cascading series of convolution layers, where the output from one layer, say C_n, feeds into the subsequent layer, C_{n+1}. This architecture facilitates extracting foundational (low-level) and intricate (high-level) features. In essence, high-level features emerge as a compounded blend of their foundational counterparts, allowing for the representation of sophisticated image patterns. There's been a surge in recent studies focusing on visualizing these feature levels in neural networks [237, 233, 244].

Stride The convolution filters, typically compact in dimension (ranging around 3×3 to 11×11 pixels), capitalize on the strong spatial correlations inherent within close-knit pixel clusters in images. The filter's size, called the 'receptive field', is a chosen parameter. It defines how many neurons from a preceding layer (or pixels if it's an input layer) connect to a neuron in the present layer. A defining trait of convolution layers is the concept of 'shared weights'. Regardless of the filter's position on an image, its weights remain consistent, ensuring weight adjustments are uniform across the entire image. This approach notably reduces the network's weight dimensionality, hence, convolution networks often carry fewer weights compared to their fully connected counterparts.

Another essential factor in convolution layers is the 'stride', dictating the horizontal or vertical movement of the filter after each convolution operation. For a visual, consider Figure 3.2. Here, a 5×5 kernel interacts with an image. A stride set to three pixels means, post convolution, the kernel will be moved three pixels rightward.

3.2.2 Pooling layer

Pooling layers are commonly utilized after convolution layers. They serve the function of subsampling feature responses from a prior convolution layer, propagating only the most essential feature response (based on the pooling type) to subsequent layers. Such layers also infuse a level of translation invariance into the neural networks.

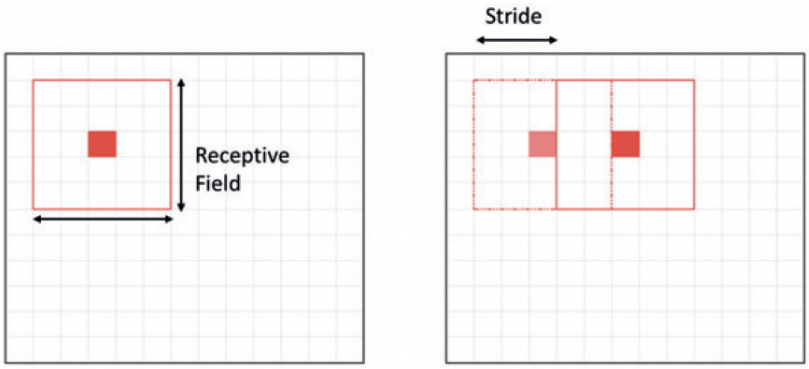


Figure 3.2: Illustration of the convolution layer's receptive field and stride: The receptive field covers a 5×5 area. In this particular example, a stride of three is employed, indicating that the convolution filter is displaced by a distance of three pixels after each convolution operation.

Three main parameters characterize the pooling layer: size, stride, and type. The size denotes the dimensions (both horizontally and vertically) of the region over which the pooling action happens. Stride, similarly to its role in convolution layers, determines the subsequent pooling area's location. By adjusting the size and stride, one can establish overlapping (if the stride is less than the size) or non-overlapping (if the stride is equal to or larger than the size) pooling areas.

There are two prevalent forms of pooling: max pooling and average (or mean) pooling. Max pooling takes and forwards only the highest feature response within the pooling region to the next layer. Conversely, average pooling calculates and sends forward the average of all feature responses within that pooling window. For instance, Figure 3.3 depicts the outcomes from both max and average pooling layers.

Notably, this subsampling process contributes to a reduced count of network parameters, simplifying the optimization process.

3.2.3 Fully connected layer

A fully connected layer is characterized by the fact that every neuron within it is linked to every element of the preceding (downstream) layer, leading to a very dense interconnected pattern. This contrasts convolution layers, where a neuron only connects to prior elements within its receptive field. Given the connectivity of the fully connected layers, even a limited number of them can account for a significant portion of the network's weights.

In most neural networks, these fully connected layers are found towards the end, just before the network computes its loss. It's also frequent to see an activation function, such as the Rectified Linear Unit (ReLU) [74], placed between two consecutive fully connected layers. However, recent trends have started to challenge the indispensability of these layers. There's a growing inclination to swap out the terminal fully connected layer with a support vector machine classifier. According to findings in [73], networks with fewer fully connected layers demonstrated better generalization on testing data.

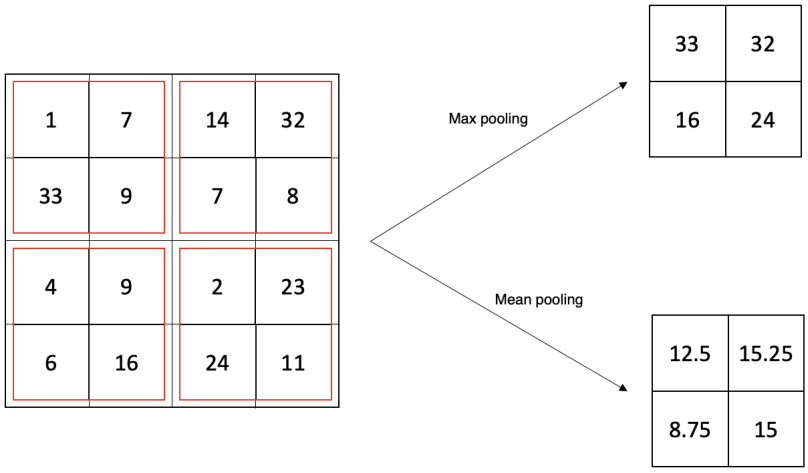


Figure 3.3: Illustration of max and mean pooling techniques. In this example, a pooling size of two pixels is applied in horizontal and vertical directions, with a stride of two pixels, ensuring that pooling neighborhoods do not overlap.

3.2.4 Dropout layer

Introduced by Hinton and his team in 2012 [99], dropout layers have since become an integral part of many deep learning architectures. These layers can be inserted anywhere within a neural network. Interestingly, in some experiments from [99], including a dropout layer right at the input (alongside those within hidden layers) led to even more significant reductions in recognition errors. It's worth noting that dropout layers are always used to support another functional layer, such as a convolution or a fully connected layer. This means a dropout layer is always paired with a specific functional layer.

The dropout probability, specified by the user, is a crucial parameter for each dropout layer. This probability determines how often elements from its paired functional layer are turned 'on' or 'off' during a training cycle. For instance, a 0.5 dropout probability means that half the neurons are randomly deactivated in any given iteration. As a result, these neurons do not participate in the forward or backward passes for that iteration. However, suppose a neuron is deactivated for a particular iteration (say iteration k), even though it remains inactive for that cycle. In that case, its last known weights (from iteration $k-1$) are preserved for potential use in the succeeding iteration ($k+1$), assuming the neuron is reactivated. During the testing phase, every neuron is active, and their weights are typically adjusted by a factor (for instance, halved for a dropout probability of 0.5).

Dropout layers play a crucial role in regulating deep networks. They work to reduce overfitting by effectively mitigating the model's variance. By periodically deactivating specific neurons, the network's structure gets modified. When using a dropout probability of 0.5, this behavior can be likened to sampling from an ensemble of 2^n distinct networks, where n represents the total neurons in the paired functional layer.

3.2.5 Attention layer

In addition to the previously discussed layers, another pivotal component that merits discussion is the attention layer [76]. This layer represents a significant advancement in the field of deep learning, especially in the context of CNNs. The attention layer works by selectively focusing on certain parts of the input data, much like the human visual attention mechanism. It identifies which features are more important in the given context, allowing the model to concentrate more 'attention' on those relevant features while processing information. This mechanism enhances the model's ability to learn complex patterns and relationships within the data, leading to more accurate and efficient performance. The integration of attention layers in CNNs has shown remarkable improvements in various tasks, particularly those involving intricate visual recognition challenges. This layer's ability to refine the focus of the network aligns seamlessly with the research field of the current thesis, where precision and detail-oriented analysis are crucial.

3.2.6 Activation functions

Rectified linear unit (ReLU) The ReLU is an activation function that can be integrated with other layers. Since its debut in [74], it has become the predominant activation function in deep learning, surpassing hyperbolic tangent and logistic sigmoid functions [114]. The ReLU can be described as:

$$f(x) = max(0, x), \tag{3.2}$$

where x is the ReLU's input. It operates similarly to a half-wave rectifier by setting all negative inputs to zero. When paired with a batch normalization layer, only about 50% of the ReLUs are active at any moment, leading to sparse activations.

One significant advantage of the ReLU over sigmoid-based activation functions is its gradient doesn't diminish as input increases. Instead, its value remains steady at one, provided the input stays positive. To illustrate, if a large positive input is fed into a sigmoid, the output is capped at one. Hence, any increment in input translates to only a minimal increase in the output and, subsequently, the gradient—giving rise to the "vanishing gradients" problem. The gradient for ReLU is defined as:

$$\frac{\partial f}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

However, the ReLU isn't without its issues. It's possible to end up with weights that cause the neuron to remain inactive for all training examples in the dataset. This phenomenon is known as "dead neurons". Imagine, for instance, a bias set too low during training. To counteract this, various ReLU variants have emerged, like leaky ReLU, parametric rectified ReLU, and randomized leaky ReLU. For an in-depth analysis of these activation functions, readers can refer to this study [228]. It's worth noting, however, that many of these ReLU alternatives are relatively new and require further investigation to solidify opinions about their effectiveness and adaptability to diverse datasets and network designs.

Last layer activation function The activation function used in the final fully connected layer often differs from those used in earlier layers. The choice of the activation function depends on the specific task at hand. For example, for tasks involving multiclass classification, the softmax function [55] is commonly used. This function converts the real-valued outputs from the last fully connected layer into probabilities for each target class. Each output value lies between 0 and 1, with the sum of all these values equaling 1. When it comes to binary classification, the most common one is the sigmoid function [89].

3.3 Training a network

Training a neural network involves identifying the optimal kernels in convolution layers and weights in fully connected layers to minimize the differences between the network's output predictions and the actual ground truth labels in the training dataset. The backpropagation algorithm is a widely-used method for training these networks, in which several components and methods are in use. The performance of a model with specific kernels and weights is evaluated by a loss function during forward propagation on the training dataset. The learnable parameters, which are the kernels and weights, are then adjusted based on the calculated loss value. This adjustment is done through backpropagation and optimization algorithms.

3.3.1 Batch normalization

Introduced in 2015 [106], the batch normalization (BN) layer rapidly gained traction, proving especially valuable for training deep neural networks. Positioned typically before a non-linearity, like the ReLU function, this layer normalizes each set of feature responses from the minibatch that enters the BN layer. Essentially, BN takes the feature responses corresponding to each feature detector and transforms them into a standard normal distribution. After this initial compulsory normalization, it provides the flexibility to learn and then apply a reverse transformation, which involves rescaling and re-translating the normalized features. The parameters of this reverse transformation are determined during network training.

At first glance, one might perceive normalization as trivial, especially when the data entering the network's input layer might already be normalized. Yet, in practice, batch normalization plays a vital role in the learning process. It ensures that only about half of the feature responses surpass the subsequent non-linear activation's activation threshold. Hence, post-BN, a feature response's contribution to the subsequent non-linear activation hinges on its relative value, not its absolute value. To illustrate, if all feature responses funneled into a ReLU exhibit high positive values, every ReLU unit becomes activated. Conversely, entirely negative input values could deactivate all non-linearities. Integrating a BN layer transforms this strictly positive or negative spectrum into a distribution with zero mean and unit standard deviation. This guarantees that around half of the feature responses consistently influence the gradient, potentially sidestepping the "dead neurons" scenario. In such a scenario, with ReLU's negative inputs, no weight updates might occur.

3.3.2 Loss functions

In supervised training of deep networks for classification tasks, data is often represented as input and target pairs (x_i, t_i), where x_i is an individual data sample and t_i is its corresponding class

label. Imagine the neural network as a function f, which comprises iterative applications of linear operations followed by non-linear activations. Broadly speaking, for a given input x_i, the output y_i can be denoted as:

$$y_i = f(x_i, W) \tag{3.4}$$

where W represents the set of weights that serve as parameters to be optimized.

The primary objective of training the network is to minimize the discrepancy between the predicted output y_i and the actual target t_i across all data points. This difference is quantified through a loss function, and various loss functions like Softmax, Hinge, and Euclidean loss exist in the literature [76].

One of the most prevalent loss functions is Softmax loss, often used for classification problems involving multiple classes. For better clarity, let's enhance our notation: y_{ij} represents the output and t_{ij} the target, where i indexes a data point and j a class. For instance, in a binary classification problem, each data point x_i yields two values y_{i1} and y_{i2}, interpreted as unnormalized probabilities for the two classes. The target vector t_i also has two elements: t_{i1} and t_{i2}, where the element corresponding to the correct class is set to one and the other to zero.

With this refined notation, the Softmax loss for a specific data point i can be represented as follows:

$$L_i = -log\left(\frac{e^{y_{ik}}}{\sum_j e^{y_{ij}}}\right), \tag{3.5}$$

where y_{ik} is the output corresponding to class k, often involving normalization based on the complete output vector y_{ij} for that data point. Usually, neural networks operate on minibatches containing n examples, and the total loss is an average of the individual losses computed for each of these n examples, as shown below.

$$L = \frac{\sum_{i=1}^{n} L_i}{n} \tag{3.6}$$

In addition to the primary loss term L, including a regularization term R is commonplace, often weighted by a coefficient λ. Typically, R is calculated over the network weight W, as follows:

$$R = \sum_p W_p^2, \tag{3.7}$$

and the final loss L_{final} is:

$$L_{final} = L + R. \tag{3.8}$$

The gradients of the final loss function are calculated to update the network's weights. These gradients are then propagated backward through the network, updating weights layer-by-layer using analytical gradient computation and the chain rule of calculus. This approach using the L and R components was explicitly used in Chapter 5.

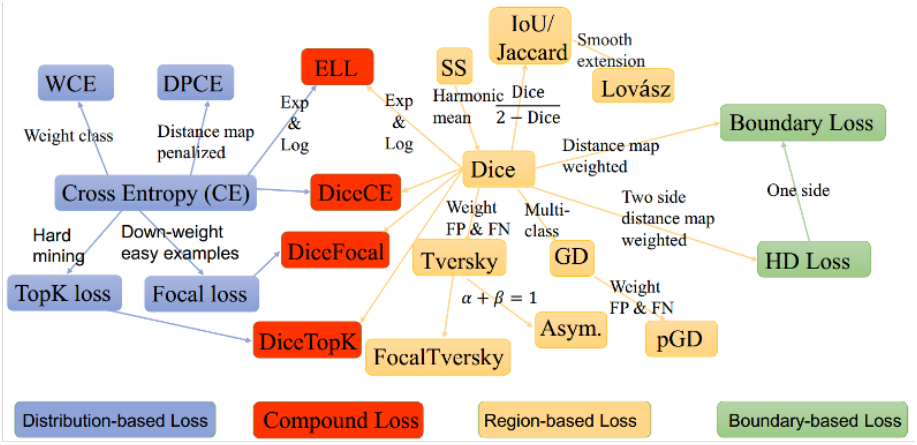


Figure 3.4: Tree representation of loss functions families.

The family of loss functions is extensive. Many of them were studied and used within this thesis for different purposes. More information on the specific selections of loss functions can be found in the corresponding sections. Here, in Figure 3.4[1], there is just a graphic representation of some of them, while below there is a short presentation of the main ones.

3.3.2.1 Cross-entropy loss

Cross-entropy loss [76] is used to measure the performance of a classification model whose output is a probability value between 0 and 1. It's effective for multi-class problems and is commonly used in a pixel-wise manner for segmentation. For a single data point, the Cross-Entropy Loss is given by equation 3.9, where M is the number of classes, y is a binary indicator (0 or 1) if class label c is the correct classification for observation o, and p is the predicted probability observation o is of class c. Special case is the binary cross-entropy loss, which is suitable for binary segmentation tasks.

$$loss_{CE} = -\sum_{c=1}^{M} y_{o,c} log(p_{o,c}) \tag{3.9}$$

3.3.2.2 Dice loss

Dice loss [46] is based on the Dice coefficient, a statistic used to gauge the similarity of two samples. It's particularly useful in medical image segmentation where the class imbalance issue is prevalent. Dice Loss is calculated as equation 3.10, where X and Y are the predicted and ground truth binary segmentations, respectively.

$$loss_{Dice} = 1 - \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{3.10}$$

[1] https://github.com/JunMa11/SegLoss

3.3.2.3 Jaccard loss

Jaccard (or IoU) loss [56], based on the Jaccard index, measures the overlap between the predicted segmentation and the ground truth. It's particularly effective for class imbalance issues. The equation 3.11 is below, with *X* being the predicted segmentation and *Y* the ground truth.

$$loss_{Jaccard} = 1 - \frac{|X \cap Y|}{|X \cup Y|} \tag{3.11}$$

3.3.2.4 Focal loss

Focal loss [130] is designed to address the class imbalance problem by focusing more on hard, misclassified examples. The Focal Loss is defined as in equation 3.12, where p_t is the model's estimated probability for each class, α is a weighting factor for the rare class, and γ is a focusing parameter.

$$loss_{Focal} = -\alpha(1 - p_t)^{\gamma} log(p_t) \tag{3.12}$$

3.3.2.5 Tversky loss

Tversky loss [187] allows more control over false positives and false negatives, making it suitable for imbalanced datasets. It's a generalization of the Dice coefficient. It is presented in equation 3.13, where α and β control the relative weight of false negatives and false positives, respectively.

$$loss_{Tversky} = 1 - \frac{|X \cap Y|}{|X \cap Y| + \alpha|X \backslash Y| + \beta|Y \backslash X|} \tag{3.13}$$

3.3.3 Optimization

The process of training a neural network is iterative and focuses on optimization. The final loss function, L_{final}, relies on two key factors: (1) The collection of network weights, denoted as W, and (2) The batch of input samples, referred to as X. Among these, the weight set W is considered the variable to be optimized. In other words, given a specific batch X, the weights W are adjusted to best classify the training data into their respective categories.

The literature contains a variety of optimization algorithms aimed at reducing network loss. Most of these algorithms are based on the principle of gradient descent and operate under a local search framework. Some commonly employed optimization algorithms include Gradient Descent (GD), Minibatch Gradient Descent (MBGD), Stochastic Gradient Descent (SGD), AdaDelta (which utilizes an adaptive learning rate), Adaptive Gradient, Adam and Nesterov's Accelerated Gradient [116, 154, 236, 53].

Despite the vast array of available optimization algorithms, Adam [116] remains a popular choice in practice. Specifically, in the experiments described in this thesis, Adam was the chosen optimization method.

Adam is an optimization technique that offers an alternative to the traditional Stochastic Gradient Descent (SGD) for iteratively updating network weights during training. Introduced in 2015, Adam sets itself apart from traditional SGD methods in several ways. Unlike SGD, which utilizes a fixed learning rate, commonly denoted as α, for all weight updates throughout the training process, Adam employs individual learning rates for each network weight. These rates are adaptively updated as the training progresses. Adam synthesizes the advantages of both AdaGrad and RMSProp. However, it goes a step further by not just using the average of the first moments (the mean) of the gradients, as seen in RMSProp, but also by leveraging the average of the second moments (the uncentered variance) of the gradients. The algorithm computes exponential moving averages for both the gradient and the squared gradient, controlled by decay rate parameters β_1 and β_2. When initialized, the moving averages and the β values (typically close to 1.0 as recommended) are biased toward zero. Adam corrects this bias by obtaining biased moment estimates and adjusting them to produce bias-corrected estimates. This multi-faceted approach allows Adam to be highly effective in various machine-learning tasks.

3.4 Architectures

3.4.1 Popular architectures

Over the past years, numerous network architectures have emerged. Some have set new performance standards on widely recognized image datasets. AlexNet [118] clinched the ILSVRC 2012 classification challenge. The Network in Network (NiN) [129] showcased an early modular neural network design built by repetitively aligning similar components. The Inception network (or GoogLeNet) became the benchmark in the ILSVRC 2014 detection challenge [202], with a 22-layer deep structure that used 12 times fewer parameters than AlexNet did in 2014. The Visual Geometry Group (VGG) at Oxford introduced models of various depths, with their 16 and 19-layer models [34] nearly matching GoogLeNet's performance at the ILSVRC 2014 challenge. Residual Networks (ResNets) [93], victors of the ILSVRC 2015 classification challenge, presented a unique design for ultra-deep networks with reduced parameters thanks to small 3×3 convolution filters. It's worth noting the foundational LeNet design [124], predating the others but influential for subsequent deeper models. Last, the most famous network for biomedical image segmentation has been proved to be U-Net [181], which is presented thoroughly later. A concise overview of the three architectures—AlexNet, NiN, and LeNet—is also provided.

3.4.1.1 LeNet

The LeNet architecture is often considered the forerunner of contemporary neural networks. It represents a groundbreaking approach to design, introducing a sequence of convolutional layers interspersed with downsampling layers, succeeded by several fully connected layers. LeCun et al. [124] have proposed various architectures under the 'LeNet' label. Specifically, the LeNet-5 features three convolutional layers and two fully connected layers, utilizing hyperbolic tangent activation functions and an Euclidean loss function.

3.4.1.2 AlexNet

The AlexNet employs a conventional convolutional architecture, consisting of five consecutive convolution layers with filter sizes of 11×11, 5×5, 3×3, 3×3, and 3×3, respectively. Three fully connected layers then follow these. Interestingly, these five convolution layers, also known as the convolution stack, are duplicated into two parallel columns. Due to the random initialization, the filters in these columns evolve to capture distinct features. For instance, in the original study, while one convolution stack learned edge detection, the other discerned color patterns characteristic of low-frequency backgrounds.

3.4.1.3 Network in Network

The NiN (Network in Network) architecture adopts a modular approach. This modular design philosophy is also foundational to the Inception model, which influenced the creation of GoogLeNet. At its core, NiN comprises several identical micro-networks that can be linked sequentially, creating networks with customizable depths. The underlying belief behind NiN is that if a single micro-network can deliver satisfactory performance, chaining multiple such micro-networks should enhance the results. However, this principle doesn't always translate seamlessly in practice. Introducing too many micro-networks might lead to common deep network challenges, like overfitting and heightened optimization expenses. In reality, Lin et al. [129] utilized only three of these micro-networks, each consisting of a convolutional layer followed by two fully connected layers, to assemble their holistic classification model.

3.4.2 U-Net

U-Net is a specialized convolutional neural network tailored for biomedical image segmentation, originating from the University of Freiburg's Computer Science Department [181]. Rooted in the fully convolutional network structure, U-Net's design was enhanced to perform optimally with limited training images while achieving more refined segmentation. The success of U-Net's design revolves around enhancing a conventional contracting network with layers where upsampling operators take the place of pooling operations. This results in an improved resolution of the output. Following this, a convolutional layer generates a detailed output leveraging this refined data.

A standout feature of U-Net is its abundant feature channels in the upsampling segment, enabling the network to relay contextual data to layers of higher resolution. This design choice results in the expansive path mirroring its contracting counterpart, creating a U-shaped outline. Without relying on any fully connected layers, the network employs only valid convolutions. To address the prediction of border pixels in images, U-Net extrapolates missing contexts by mirroring the image. This tiling strategy, crucial for processing large images, sidesteps potential GPU memory-related resolution constraints.

U-Net's structure comprises two main paths: contracting and expansive. The contracting path, a standard convolutional network, repetitively applies convolutions followed by a rectified linear unit (ReLU) and a max-pooling operation. This contraction process diminishes spatial data while augmenting feature information. Conversely, the expansive route melds feature and spatial details via a series of up-convolutions and concatenations, integrating high-resolution features from the contracting path.

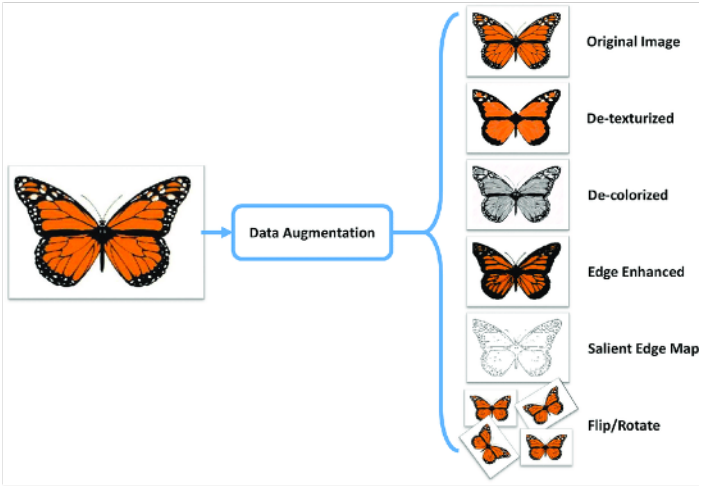


Figure 3.5: An example of data augmentation to a butterfly picture. The transformations found here contain de-texturization, de-colorization, enhancement of edges, mapping of salient edges, flips, and rotations.

U-Net has found numerous applications in biomedical image segmentation, notable examples being brain image segmentation and liver image segmentation [8, 149]. Additionally, its utility extends to protein binding site predictions [221]. The architecture's flexibility has given rise to variants, such as those used for medical image reconstructions [238]. Some noteworthy U-Net variants and applications include (i) pixel-wise regression with U-Net, utilized in pansharpening [231], (ii) 3D U-Net, designed for dense volumetric segmentation from sparse annotations [37], (iii) TernausNet, a U-Net variant with a VGG11 Encoder pre-trained on ImageNet, tailored for image segmentation [105].

3.5 Data augmentation

Depending on the specific task, deep learning methodologies can demand more training data than readily accessible. A shortage of such data can lead to overfitting, where performance on the training set significantly outpaces that on the validation and test sets. This data scarcity issue can be counteracted using data augmentation during preprocessing. Data augmentation essentially generates supplementary training data through appropriate transformations of the existing dataset.

The nature of these transformations hinges on the task. The oocytes in our cumulus-oocyte complexes segmentation framework don't exhibit a predominant direction. Thus, an oocyte rotated by 180 degrees remains as valid as its unaltered counterpart. Contrastingly, such rotations might be deemed inappropriate in standard computer vision datasets. For instance, rotating images of cars is not typical since upside-down cars are rarely encountered in conventional photography, barring exceptional scenarios like junkyards. Standard augmentation techniques include flips, translations, rotations, shearing, rescaling, contrast adjustments, and modifications to illumination in natural scenarios. Some of these augmentative strategies, especially for image recognition, have been deliberated in [227]. For data augmentation to be beneficial, the transformations should align with the nature of the data and the task's requirements. Some sample transformations can be found in Figure 3.5 and Figure 3.6.

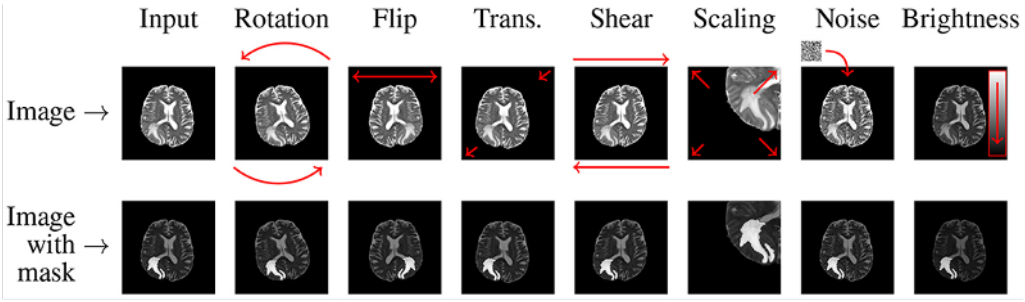


Figure 3.6: An example of data augmentation to a brain CT scan. The transformations here contain rotations, flips, transitions, shear, scaling, noise addition, and brightness changes.

Dealing with restricted datasets, where each data entry is invaluable, poses a challenge in reserving a subset exclusively for testing. A remedy is the k-fold data sampling and cross-validation method. Here, data is partitioned into k segments. Training occurs on k-1 segments, leaving one for testing. In our experiment setup, k-2 segments were utilized for training, with the remaining allocated for validation and testing. This train/test procedure recurs k times, with a distinct segment earmarked for testing in each iteration. Ultimately, the result emerges as an average of scores from the k iterations, ensuring that a specific data subset doesn't skew the outcome. One viable strategy for deployment on new, unseen data is interpreting the k differently-trained networks as components of an ensemble, averaging their predictions before class label assignment.

Moreover, tactics like pretraining on a large external image set or leveraging an already pre-trained network with subsequent fine-tuning on the original dataset have proven to boost performance in certain instances [60].

3.6 Transfer learning

Transfer learning (TL) is a machine learning strategy where knowledge acquired from one task is reused to enhance performance on a similar job. An illustration is in image classification, where insights from recognizing cars can assist in identifying trucks. This concept is akin to the psychological idea of transfer of learning, though connections between the two are scarce. Leveraging information from past tasks for new ones can notably elevate learning efficiency.

In 1976, Bozinovski and Fulgosi published a paper addressing transfer learning in neural network training [21]. The paper gives a mathematical and geometrical model of the topic. In 1981, a report considered the application of transfer learning to a dataset of images representing letters of computer terminals, experimentally demonstrating positive and negative transfer learning [20]. In 1993, Pratt formulated the discriminability-based transfer (DBT) algorithm [175]. In 1997, Pratt and Thrun guest-edited a special issue of Machine Learning devoted to transfer learning [174]. By 1998, the field had advanced to include multi-task learning [33], along with more formal theoretical foundations [16]. Learning to Learn [210], edited by Thrun and Pratt, is a 1998 review of the subject. Transfer learning has been applied in cognitive science as well. Pratt guest-edited an issue of Connection Science on the reuse of neural networks through transfer in 1996 [173]. Ng said in his NIPS 2016 tutorial [155] that TL would become the next driver of machine learning commercial success after supervised learning. In the 2020 paper "Rethinking Pretraining and self-training" [245], Zoph et al. reported that pretraining can hurt accuracy and advocate self-training

instead.

Algorithms are available for transfer learning in Markov logic networks [152] and Bayesian networks [157]. Transfer learning has been applied to cancer subtype discovery [85], building utilization [3], general game playing [14], text classification [49], digit recognition [138], medical imaging and spam filtering [18].

In 2020, it was discovered that, due to their similar physical natures, transfer learning is possible between electromyographic (EMG) signals from the muscles and classifying the behaviors of electroencephalographic (EEG) brainwaves from the gesture recognition domain to the mental state recognition domain. It was noted that this relationship worked in both directions, showing that electroencephalographics can likewise be used to classify EMG [19]. The experiments noted that the accuracy of neural networks and convolutional neural networks were improved [138], through transfer learning before any learning, i.e., compared to standard random weight distribution. That is, results are improved by exposure to another domain. Moreover, the end-user of a pre-trained model can change the structure of fully-connected layers to improve performance.

3.7 Hyperparameter tuning

Several factors influence the efficacy of neural networks in reaching an optimal solution. These include the architecture of the network (its depth, width, and chosen activation functions), the selected loss function, the optimizer used, the initial learning rate, and how this rate is adjusted as training progresses. With the notable advancements in computational capabilities, mainly due to GPU utilization, it has become feasible to fine-tune a selected range of these parameters.

A growing body of evidence suggests hyperparameter optimization techniques, like the Bayesian method proposed by Snoek et al. [196], can enhance the performance of these networks. However, another popular method is the grid search, which involves systematically testing various combinations of parameters to identify the optimal set. Two of the most frequently fine-tuned hyperparameters are the learning rate and the loss function.

In real-world scenarios, many practitioners opt for a more pragmatic approach. Rather than undertaking comprehensive grid searches or employing Bayesian hyperparameter optimization, they prefer to hone network performance by tweaking a select range of parameters based on prior experience and intuition. Although seemingly less systematic, this heuristic methodology often yields effective results by leveraging hands-on experience and domain-specific insights.

3.8 Conclusion

In concluding this segment, we've embarked on an in-depth exploration of CNNs, dissecting their intricate designs, operational mechanisms, and the vast capabilities they offer for analyzing complex visual datasets. This exploration has not been a mere academic exercise; rather, it has laid a robust practical foundation that holds the promise of ushering in transformative advancements across a myriad of sectors. As we prepare to transition into the subsequent chapter, our focus shifts towards the cutting-edge intersection of CNN technology and the domain of assisted reproduction, a field ripe for innovation. In the forthcoming discussions, we aim to unveil the innovative ways through which CNN principles and methodologies are being tailored to tackle some of the

most daunting challenges within reproductive medicine. This fusion of advanced computational techniques with the intricacies of human reproduction is poised to not only broaden our scientific comprehension but also redefine the methodologies employed in assisted reproduction, setting a new paradigm for addressing and solving real-world challenges. Through this journey, we aspire to illuminate the transformative potential that lies at the confluence of technology and medicine, heralding a new era of solutions that are as impactful as they are revolutionary.

Chapter 4

Cumulus-Oocyte Complex Quality

In this chapter, we delve into a series of important research questions, each shedding light on a distinct part of our investigation:

- **Q1:** What is the optimal method for computing the area of cumulus-oocyte complexes (COCs)?
- **Q2:** Can we develop a deep learning framework that achieves human-level performance using the optimal method for calculating COCs' area?
- **Q3:** Is it feasible to identify the expansion of COCs from an immature to a mature stage using the optimal method?
- **Q4:** How does the optimal method for measuring COCs' expansion impact healthy embryo development in mammalian species?
- **Q5:** What influence does the density of COCs, as determined by the optimal method, have on healthy embryo development in mammalian species?

Research questions **Q1**, **Q3**, and **Q4** presented below are an expansion of the paper: Raes, Annelies, Athanasiou, Georgios, et al. "Manual versus deep learning measurements to evaluate cumulus expansion of bovine oocytes and its relationship with embryo development in vitro." Computers in Biology and Medicine (2023): 107785. [176].

Research question **Q2** presented below is an expansion of the conference paper: Athanasiou, Georgios, et al. "Detecting the Area of Bovine Cumulus Oocyte Complexes Using Deep Learning and Semantic Segmentation." [7]

Research question **Q5** presented below is an expansion of the paper: Athanasiou, Georgios, et al. "Measuring cumulus density of mammalian oocytes: AI methods on determining the importance of cells density to a healthy oocyte development." [submitted and under review at *AI Communications*].

4.1 Introduction

As has already been described in the previous chapters, infertility, defined as the failure to attain a clinical pregnancy following 12 months or more of regular, unprotected sexual activity, presents considerable hurdles for both medical practice and societal well-being. When infertility is diagnosed, the treatment pathways frequently involve the use of Assisted Reproductive Technology (ART). These techniques consist of multiple sub-procedures, notably the essential step of identifying the morphological traits of oocytes and key elements of embryo biology.

Cumulus oocyte complexes in immature and mature stages play a crucial role in evaluating the quality of mammalian oocytes, particularly for their utilization in in-vitro fertilization (IVF) procedures. Their role has been extensively explained in Chapter 2. The area of the COC can provide us with information valuable for predicting embryo development. Among others, the main focus of this dissertation is the size of these areas, cell density within these areas, and the relative change of the size (expansion) and cell density.

Numerous methods for identifying cumulus areas have been described in the literature, including studies by Chen et al. [35] and Ploutarchou et al. [170]. However, these existing methods suffer from limitations such as being time-consuming and relying heavily on human subjectivity, as the annotations can vary among different experts. Some of the current approaches involve assessing the area of the cumulus, including the oocyte, to evaluate several characteristics of the cumulus cells. This research proposes a pipeline designed explicitly for segmenting the cumulus-oocyte complex to address these challenges and facilitate automation. Once the COC is accurately segmented, measuring its size and the density of its components becomes a straightforward task. By automating the segmentation process, our pipeline aims to streamline and simplify the assessment of studying cumulus characteristics, reducing reliance on subjective human judgments and potentially saving time in evaluating oocyte quality for IVF procedures.

In recent years, medical image segmentation has witnessed significant advancements, thanks to the emergence of deep learning techniques, particularly Convolutional Neural Networks. These developments have brought about positive impacts in medicine and healthcare. Image segmentation, a fundamental task in this domain, involves dividing an image into smaller parts to create a more meaningful representation that machines can process effectively. This research employs image segmentation to segment bright-field microscopy images of cumulus-oocyte complexes in immature and mature oocytes. The aim is to accurately delineate the boundaries of the cumulus region for further analysis. To accomplish this, a U-Net network architecture [181], well-suited for medical image segmentation tasks, is utilized. More information on U-Net is available in Chapter 3. By leveraging deep learning and the U-Net architecture, our approach enables precise and automated segmentation of COCs in bright-field microscopy images. This segmentation process is crucial in analyzing the information laid on the cumulus area, let alone the size of the area itself, for assessing the maturity of oocytes. The ability to automate this process not only reduces the reliance on manual annotations, but also offers the potential for increased accuracy and efficiency in evaluating this information.

The literature specifically focusing on image segmentation techniques for oocyte microscopy is somewhat narrow. For instance, a study by Firuzinia et al. [64] centered on the segmentation of human metaphase II mature oocytes, exploring various morphological features pertinent to this developmental stage. Although their dataset included 1009 images, details about the involvement of multiple specialists in annotating the oocytes are not provided. Another notable study by Tar-

gosz et al. [208] applied image segmentation to human oocytes at various stages, including MII, MI, and others. Their dataset had 334 images, each featuring one or more oocytes. As with the work by Firuzinia et al., Targosz et al. do not explicitly mention the use of multiple specialists for annotation. Both studies used pre-trained neural networks like ResNet [93] and MobileNet [101] and incorporated various data augmentation methods to improve their models' performance.

Image segmentation techniques have also found broader applications in Assisted Reproductive Technology (ART), particularly concerning early-stage human embryos and their morphological characterization. For example, Fukunaga et al. [69] devised a system for automated detection of pronuclei in 900 embryos, leveraging image segmentation to identify these crucial structures. Studies by Khan et al. [115] and Leahy et al. [122] targeted cell counting in early-stage human embryos, utilizing segmentation methods to determine cell numbers accurately. These investigations aimed to offer vital data for evaluating embryo quality and predicting developmental prospects. Moreover, various research efforts have been directed toward identifying embryos' developmental stages. Works by Dirvanauskas et al. [47], Liu et al. [135], Malmsten et al. [140, 139, 141], Lau et al. [121], Gingold et al. [72], and Meseguer et al. [151] have all employed segmentation-based approaches for this purpose, aiming to contribute valuable insights for enhancing embryo selection and optimizing ART procedures.

This research represents the first investigation to our knowledge into exploring the optimal method for computing cumulus expansion. It is also the first investigation into image segmentation specifically applied to bovine oocytes in bright-field microscopy images. The dataset utilized in this later part comprises only 100 oocytes, which is considerably smaller than the previously mentioned studies. Another novelty of this research is the segmentation of the areas used to perform analyses of the significance of cumulus expansion and density for healthy embryo development.

4.2 Best measuring method

In this section, we aim to approach and answer the first research question:

Q1: *What is the optimal method for computing the area of cumulus-oocyte complexes (COCs)?*

4.2.1 Introduction

Beginning with cumulus expansion, there exist several techniques for evaluating it, spanning both invasive and non-invasive methods. Measuring hyaluronic acid is the most accurate, though it can hinder embryo growth and pose health risks to the user because it involves radioactive material. A common method is classifying cumulus expansion by its overall appearance or counting cumulus cell layers. This method is efficient and only requires microscopy during the oocyte's in vitro processing, eliminating the need for extra tools and preserving the oocyte's future development. However, its reliability hinges on the embryologist's skill, making it subjective.

Advances in imaging technologies have introduced methods like measuring the cumulus cells' area and length, aiming to reduce subjectivity. However, these visual techniques can still not capture the three-dimensional form of the cumulus-oocyte complex and are time-intensive. There's no comparative study on these methods' precision and repeatability. In the following, we design and perform our own study to identify which method is the most promising.

4.2.2 Resources

This study did not require ethical approval as the ovaries were sourced from cows post-mortem at a standard slaughterhouse.

4.2.2.1 Media and reagents

Tissue culture medium (TCM-199), gentamicin, and phosphate-buffered saline were procured from Life Technologies Europe, based in Ghent, Belgium. Unless otherwise mentioned, all additional chemicals were sourced from Sigma-Aldrich in Overijse, Belgium. Before utilization, all media underwent filtration using a 0.22 μm filter from GE Healthcare-Whatman, Diegem, Belgium.

4.2.2.2 Oocytes collection and in-vitro maturation

The maturation process of bovine COCs was in line with the methodology presented by Azari-Dolatabad et al. [9]. Briefly, bovine ovaries were obtained from a nearby slaughterhouse and processed within a two-hour window. The ovaries were thrice washed in warm physiological saline (37 $^{\circ}C$) that contained kanamycin (25 mg/mL) and were sterilized using 90% ethanol. Cumulus-oocyte complexes were extracted from follicles measuring 4 to 8 mm, using an 18-gauge needle connected to a 10 mL syringe. They were then shifted to a 15 mL container with 2.5 mL of HEPES Tyrode's albumin–pyruvate–lactate (HEPES-TALP) solution. Each of these complexes was subsequently cultured in 20 μL droplets of the maturation medium (infused with 20 ng/mL of epidermal growth factor and 50 μg/mL gentamicin) and overlayed with 7.5 mL of paraffin oil. This setup was maintained for 22 hours at 38.5 $^{\circ}C$ in a humid environment with 5% CO_2.

4.2.2.3 Image acquisition

Images of COCs were taken both at the initiation of in vitro maturation and 22 hours post-maturation using an inverted Olympus microscope paired with a ToupCam camera and ToupView software (version 3.7.13270.20181102). Images, all taken with the same magnification level (56×), focused on the zona pellucida of a centrally placed oocyte. These images were saved in PNG format with a resolution of 2592 × 1944 pixels. Out of the lot, 68 images were discarded due to issues with clarity or focus, leaving 232 paired images of COCs which were then shown to three evaluators.

4.2.2.4 Cumulus expansion measurement

Three well-trained independent experts measured the cumulus expansion in bovine COCs before and after the in vitro maturation process. These COCs were rated by three distinct methodologies (by area, 3-distance method, and a scoring approach). These evaluations were conducted twice, spaced over different durations, and in no specific sequence.

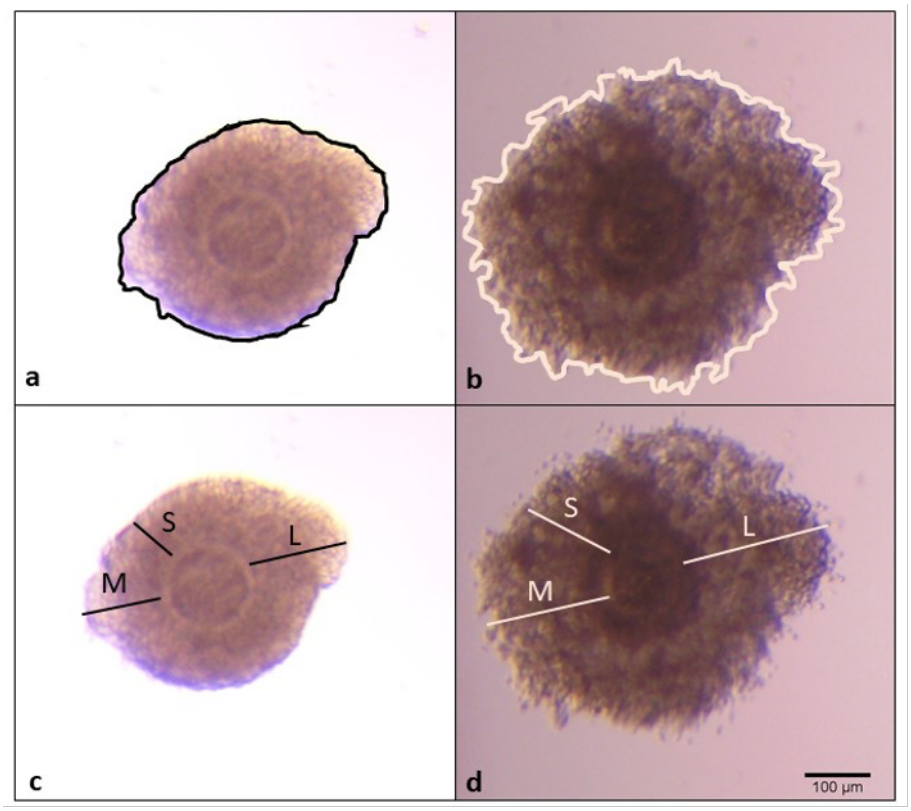


Figure 4.1: For measuring cumulus expansion, two methods were employed: the area method and the 3-distance method. In the area approach, the boundary of the cumulus cells was delineated to compute the area pre-maturation (a) and post-maturation (b). Using the 3-distance technique, three distances were identified: the shortest (S), medium (M), and the longest (L). These measurements were taken from the zona pellucida to the outermost edge of the cumulus. The arithmetic mean of these distances was determined both pre-maturation (c) and post-maturation (d). The provided scale bar is applicable to all the images.

Area ImageJ software was utilized to measure the areas of pre- and post-IVM COCs, by outlining it by hand. The growth area was determined by subtracting the initial from the final measurements and dividing by the initial. The cumulus expansion percentage was then determined.

3-distance In pre-IVM COCs, the straight lines tool in ImageJ was employed to measure the shortest, median, and longest distances from the zona pellucida to the outermost cumulus cells. Subsequently, an average of these three measurements was computed. This procedure was replicated for the post-IVM COCs. The differential between the post-IVM and pre-IVM COC average distances represented the absolute expansion. To calculate the percentage of cumulus expansion, this absolute expansion was divided by the pre-IVM COC's average distance (Figure 4.1).

Scoring To assess the expansion of cumulus cells, images of pre-IVM and their corresponding post-IVM COCs were juxtaposed. A 5-point Likert scale, previously outlined by Downs [52], was used for scoring. In essence, scores spanned from 0 to 4, where "0" indicated no expansion; "1" signified separation in only the peripheral layers of cumulus cells; "2" represented expansion

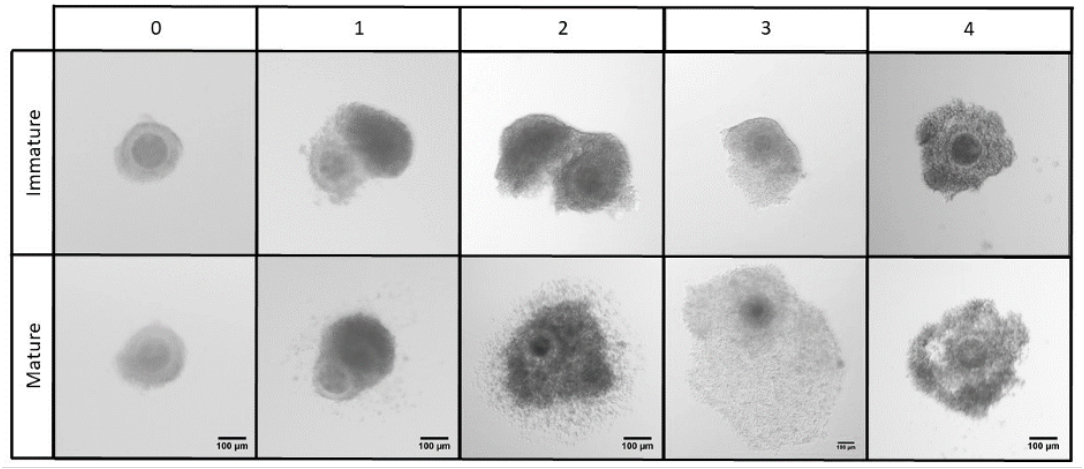


Figure 4.2: Using the scoring method for cumulus expansion assessment, cumulus-oocyte complexes were evaluated pre- and post-maturation (arranged in rows) and given scores (presented in columns) as outlined: ”0” for no noticeable expansion; ”1” for separation limited to the external-most layers of cumulus cells; ”2” for extended expansion reaching but not surpassing the halfway point of the cumulus oophorous; ”3” for expansion that extends to the corona radiata but stops short ofit; and ”4” for total expansion, encompassing the inmost cells of the corona radiata. The scale bar present in the bottom image applies to the image directly above it as well.

covering the external half of the cumulus oophorous; ”3” showed expansion nearing but not encompassing the corona radiata; and ”4” denoted full expansion involving the innermost corona radiata cells (Figure 4.2).

4.2.3 Proposal

In evaluating the different techniques for measuring cumulus expansion, the study proposes the examination of several factors: (1) The average cumulus expansion across all COCs assessed, (2) inter-observer agreement, which is the variability in cumulus expansion scores given by different observers, (3) overall inter-observer agreement, representing the mean value from two repeated measurements of inter-observer agreement since the measurements were done twice, and (4) intra-observer agreement, which refers to the incompatibility in cumulus expansion scores when an observer repeats the measurement.

Python (version 3.10.6) was used for data analysis. The inter-observer agreement was appraised using a two-way random effects model, while a one-way random effects model was employed to assess the intra-observer agreement of each expert. The ICC and its 95% confidence interval were then derived from the *intraclass_corr* function in the Pingouin [211] Python statistical package, version 0.5.2. The ICC values were interpreted based on the criteria set by Landis and Koch [120]: $<$0.20 indicates poor agreement; 0.20–0.39 signifies fair agreement; 0.40–0.59 represents moderate agreement; 0.60–0.79 is a good agreement, and anything $>$0.80 is viewed as very good agreement. Results are presented as both ICC and their 95% confidence intervals.

Method	Working principle	Equipment	Time
Area	Measuring the area by drawing the COC contour	++	++
3-distance	Measuring 3 distances between zona pellucida and outer cumulus	++	+++
Scoring	5-point Likert scale	+	+

Table 4.1: Comparison of methods to measure cumulus expansion. Methods were compared by three observers and evaluated for equipment- and time requirements and the level of subjectivity. +, easy or low; ++, moderate; +++, complicated or high.

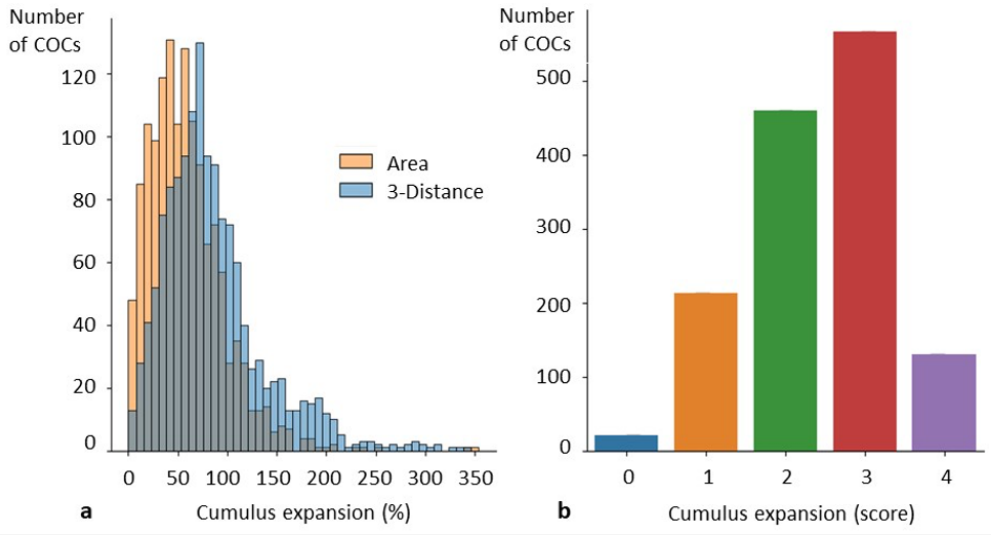


Figure 4.3: This figure illustrates the cumulus expansion for three distinct measurement techniques. Part (a) showcases the distribution of cumulus expansion percentages obtained through the area and 3-distance methods. In contrast, part (b) displays the score distribution on the Likert scale (the scoring method) for 232 cumulus-oocyte complexes (COCs), which were assessed twice by three different observers.

4.2.4 Results

Three observers assessed three distinct techniques to evaluate cumulus expansion. Table 4.1 provides a summary of these methods, detailing their operational principle, ease of use (considering the equipment and time needed), and efficacy (judged by the degree of subjectivity involved).

4.2.4.1 Cumulus expansion in numbers

The area and 3-distance techniques produced numerical results, whereas the cumulus expansion scoring method gave ordinal outcomes. As such, direct comparisons between the scoring method and the other two techniques were not conducted. When evaluating cumulus expansion using the area method, observations ranged from 0.06% to 346.3% with a median increase of 74.28% (IQR: 123.19%, Figure 4.3a). For the 3-distance approach, the range was 0.15% to 346.9% with a median expansion of 52.68% (IQR: 90.91%, Figure 4.3a). The scoring technique resulted in a median score of 3.0 (IQR: 2.0, Figure 4.3b).

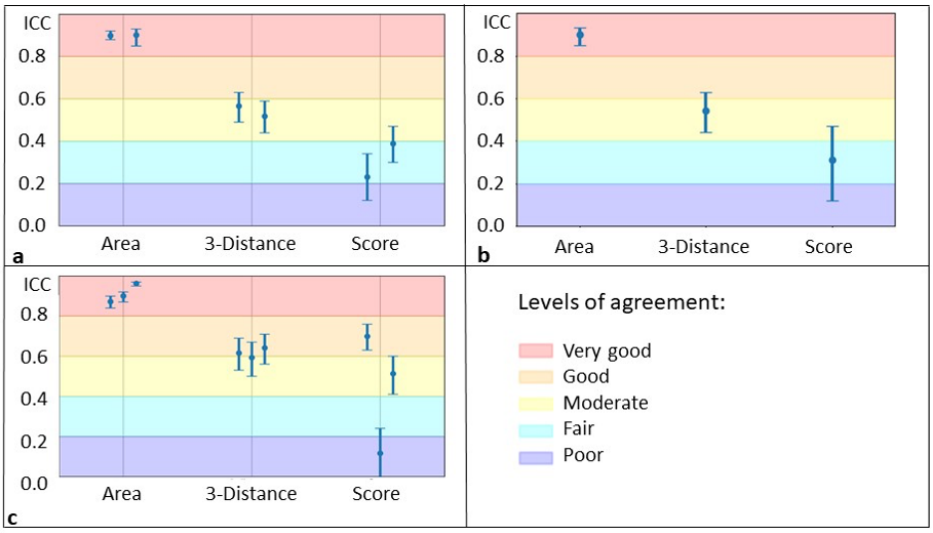


Figure 4.4: The Intraclass Correlation Coefficients (ICC) were computed and presented based on their respective agreement levels for (a) inter-observer consistency, (b) overall inter-observer agreement, and (c) intra-observer agreement. Each cumulus-oocyte complex's expansion was assessed twice for every method, leading to two ICC values for inter-observer agreement for each method.

4.2.4.2 The superiority of the area method

Inter-observer agreement The observer agreement for all three methods was assessed using the respective ICC, as depicted in Figure 4.4a and Table 4.2. Each observer carried out measurements twice for every method, leading to the ICC being computed in duplicate. For the area method, the inter-observer ICCs indicated a very high degree of agreement. In contrast, the 3-distance method had a moderate level of agreement. The scoring method's inter-observer agreement was categorized as fair.

Overall inter-observer agreement For each method, an aggregate agreement level was determined, considering both instances of inter-observer concurrence. The outcomes revealed a very good overall concurrence for the area method, a moderate consensus for the 3-distance method, and a subpar agreement for the scoring method, as displayed in Figure 4.4b and Table 4.2.

Intra-observer agreement The consistency within each observer was assessed using ICC calculations for every method and individual observer, as seen in Figure 4.4c and Table 4.2. When breaking it down, observers 1, 2, and 3 showed very good intra-observer agreement for the area method. For the 3-distance method, the agreement levels ranged from moderate to good. However, for the scoring method, the level of agreement differed for each observer, spanning from poor to moderate and reaching good in just a few instances.

Method	Inter-observer agreement		Overall agreement	Intra-observer agreement		
	Repet 1	Repet 2		Obs/er 1	Obs/er 2	Obs/er 3
Area	0.89 (0.88–0.92)	0.90 (0.85–0.93)	0.89 (0.85–0.93)	0.87 (0.84–0.90)	0.90 (0.87–0.92)	0.96 (0.95–0.97)
3-distance	0.56 (0.49–0.63)	0.51 (0.44–0.59)	0.54 (0.44–0.63)	0.61 (0.53–0.69)	0.59 (0.50–0.67)	0.64 (0.56–0.71)
Scoring	0.23 (0.12–0.34)	0.38 (0.3–0.47)	0.30 (0.12–0.47)	0.69 (0.63–0.76)	0.11 (-0.01–0.24)	0.51 (0.42–0.6)

Table 4.2: Intraclass correlation coefficients for three cumulus expansion measurement methods. Data are reported as intraclass correlation coefficients with their respective 95% confidence intervals.

4.3 AI-xpansion

In this section, we continue with the second and third research questions:

Q2: *Can we develop a deep learning framework that achieves human-level performance using the optimal method for calculating COCs' area?*

Q3: *Is it feasible to identify the expansion of COCs from an immature to a mature stage using the optimal method?*

4.3.1 Introduction

Previously, we established that the best method for measuring COC expansion is the area method. Regardless of the method, human judgment introduces variability. A potential remedy is deep learning, the branch of artificial intelligence that learns from data patterns rather than pre-set rules. Traditional non-invasive cumulus expansion evaluations largely rely on human judgment and are time-consuming. Creating an automated tool for this purpose could enhance the consistency between cumulus expansion and oocyte potential, simplifying the process for researchers and embryologists.

Cumulus expansion is a key indicator of effective COC maturation, crucial in reproductive biotechnologies. However, current non-invasive techniques aren't definitive, consume considerable time, and involve subjectivity. This section introduces a proposed DL-based tool, named AI-xpansion, for automating cumulus expansion assessment.

4.3.2 Resources

The study incorporated two datasets: (i) a pre-existing dataset for pretraining the segmentation model and (ii) a cumulus-oocyte dataset selected specifically for the segmentation task.

The pre-existing training dataset is a collection of images previously employed for segmentation model training, sourced from a distinct domain's existing dataset. More insights on this dataset are available in the referenced literature for this context.

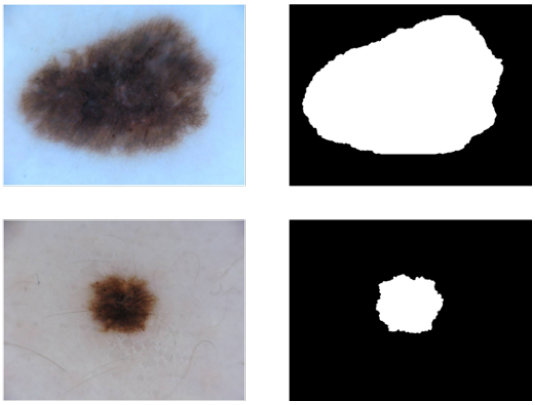


Figure 4.5: A sample representation of the melanoma dataset along with the corresponding masks provided by the experts.

Additionally, a specialized cumulus oocyte dataset was curated specifically in alignment with the segmentation goal. This in-house developed dataset was meticulously crafted to meet the specific goals and criteria of the research. Detailed aspects, including its makeup, features, and annotation methodologies for the cumulus-oocyte dataset, are further detailed in the ensuing sections.

By integrating the preliminary dataset and the specific cumulus-oocyte dataset, this research sought to harness pre-existing expertise and adapt it for cumulus-oocyte segmentation, aiming to enhance the segmentation model's performance and efficiency.

4.3.2.1 Melanoma

In this research, the pretraining dataset was sourced from the ISIC 2017 Challenge dataset[1] dedicated to Skin Lesion Analysis, with a primary emphasis on melanoma detection. This collection includes 2,000 RGB images, each meticulously segmented by medical professionals, leading to binary masks corresponding to every image (Figure 4.5). To align with the CNN's input specifications, both images and masks were transformed to grayscale and resized to dimensions of 192×240 pixels.

4.3.2.2 Cumulus Oocyte Complexes

A specialized dataset featuring bovine cumulus-oocyte complexes was created (Figure 4.6). The collection includes images of 100 oocytes. Information on how these COCs were treated and on how these images were captured are available in previous Sections 4.2.2.2 and 4.2.2.3.

To produce annotations for the COC imagery, three experts (A1, A2, A3) undertook manual segmentation of the images employing the ImageJ software. This resulted in masks identical in dimension to the original photos. Owing to minor discrepancies in the annotations by the individual experts, a consensus segmentation was derived based on mutual agreement among the annotators. Every pixel within this consensus segmentation was tagged as part of the COC if identified as such by at least two of the three annotators. Adopting this majority voting system assured annotations

[1] https://challenge.isic-archive.com/data/

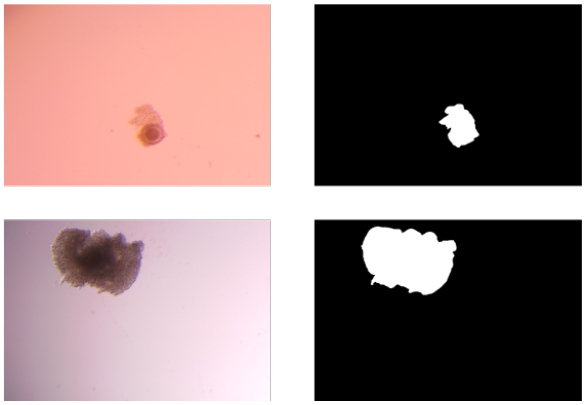


Figure 4.6: A sample representation of the COC dataset and the masks constructed using the majority-vote technique. In this case, three different experts provided masks outlining the COC region of interest. The majority-vote technique was employed to reconcile any discrepancies between the expert annotations.

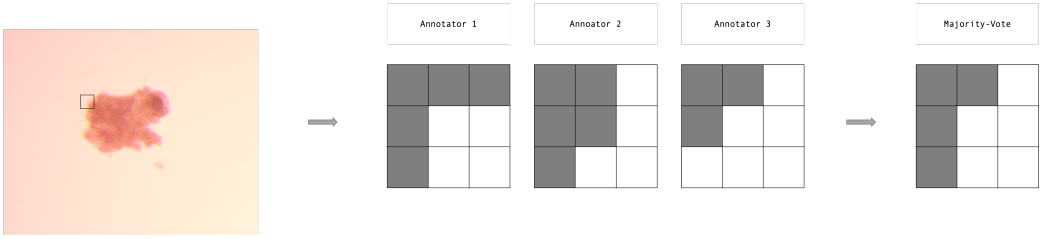


Figure 4.7: The concept of the majority vote approach using a simplified example of a 3x3-pixel sub-part of a sample image. The image consists of a grid representing the individual pixels, each containing a binary value indicating the presence or absence of a specific feature of interest

that were both consistent and dependable. An illustration of this majority voting mechanism is presented in Figure 4.7. Here, the definitive classification of each pixel is ascertained according to the predominant vote within a sample grid of 9×9 pixels.

The resulting consensus annotations were then deemed the benchmark for instructing the segmentation model, henceforth referenced as the majority-vote dataset. To correspond with the dimensions of the melanoma dataset, all images, and their associated masks, were transformed to grayscale and subsequently adjusted to a size of 192×240 pixels, utilizing OpenCV's area-interpolation method. These masks were enlarged to their initial dimensions (1944×2592 pixels) using OpenCV's cubic-interpolation technique for the conclusive assessment. It's important to emphasize that the resizing process did not impact the eventual outcomes or findings. While alternative methodologies were contemplated, like using probability-based pixels instead of definitive ones, such approaches have been earmarked for exploration in subsequent research.

4.3.3 Proposal

As it has already been stated earlier, due to its superior inter- and intra-observer agreement, the area method was chosen for further automation. This led to proposing the development of AI-xpansion, a DL algorithm designed to automatically identify the COC's area, facilitating cumulus

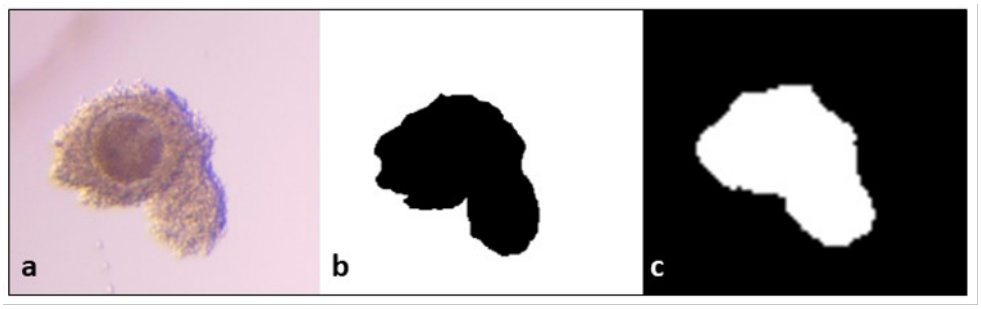


Figure 4.8: Assessment through visual segmentation masks. Starting with the original image of the cumulus-oocyte complex (COC) (a), human experts drew a segmentation mask around the COC's outline (b). This manually drawn mask was then juxtaposed with the segmentation executed by the deep learning algorithm, AI-xpansion (c).

expansion measurements. An in-depth technical explanation of AI-xpansion, which employs DL to gauge cumulus expansion autonomously, is presented below (Section 4.3.3). The DL model behind AI-xpansion is a result of combining transfer learning, image pre-processing, and a U-Net network approach inspired by Ronneberger et al. [181]. The enhancement of AI-xpansion's efficacy was based on transfer learning, leveraging a publicly available melanoma image dataset available from the ISIC archive[2]. This dataset established the foundational pre-trained DL model, which subsequently trained AI-xpansion using pre- and post-IVM COC images. From the data gathered (details in Section 4.3.2), a random selection of 100 COCs was made, primarily due to the cost implications of image annotation. This translates to a collection of 200 annotated images. Experts crafted segmentation masks by hand, outlining the area of both pre- and post-IVM COCs. These images and their associated segmentation masks formed the training dataset for AI-xpansion.

For evaluating the segmentation models, specifically AI-xpansion's capability to pinpoint cumulus boundaries, a 10-fold cross-validation was adopted. This decision was influenced by the dataset's limited size, and the method's known stability. In each iteration, 90 COCs (both pre- and post-IVM, equating to 180 images) were designated for training, while 10 COCs (or 20 images) were set aside for validation purposes. Training was performed in mini-batches of 32 spanning 200 epochs. Each validation set, across all folds, yielded a segmentation mask for its 20 images, culminating in 200 masks. These generated masks were juxtaposed against expert-provided masks for the identical COCs (illustrated in Figure 4.8). The dice coefficient [46] metric was employed to gauge the resemblance between the two masks. The average dice score of the majority-vote model consistently hovered around 95%.

Within the proposed framework, as showcased in Figure 4.9, our principal segmentation model consists of convolutional neural networks, specifically capitalizing on the U-Net architecture [181]. Figure 4.9 also illuminates the strategy of utilizing transfer learning to circumvent the issue of scant training data in our domain. On the left side, the illustration showcases the preliminary training phase wherein the U-Net model is initially trained using a universally accessible dataset from a neighboring sector. This dataset consists of bright-field microscopy images of melanoma cells, allowing the model to decipher pertinent features and patterns. Our COC segmentation endeavor is divided into two phases within the framework. Commencing with the first stage, the methodology utilizes local entropy to accomplish an initial segmentation. This initial segmenta-

[2]https://challenge.isic-archive.com/data/

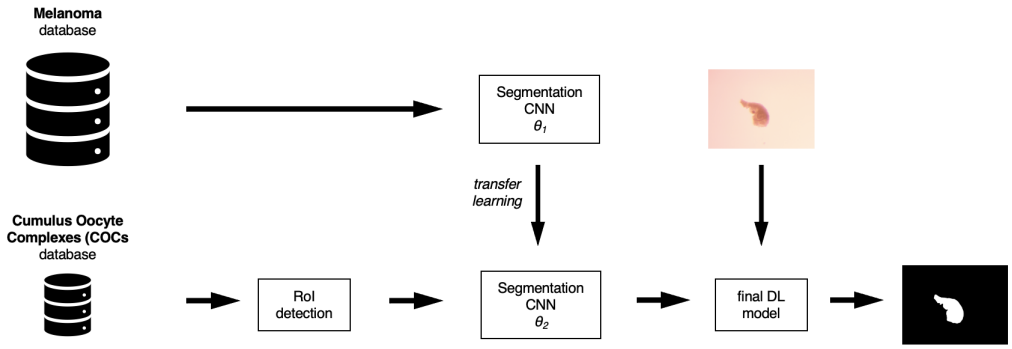


Figure 4.9: The procedure for constructing the segmentation model is outlined as follows. The procedure starts with deploying two discrete datasets: the first containing melanoma images and the second containing COC images. As the next step, transfer learning takes place, using the model pre-educated with melanoma data to lay the foundation for the segmentation model. Thanks to this transfer learning technique, the model learns crucial characteristics and patterns inherent in the melanoma dataset, rendering them invaluable for segmenting COC visuals. Ultimately, the proficiently tutored segmentation model stands primed to deftly segment a presented COC visual, efficiently identifying the cumulus area from the ambient background. This figure outlines the steps in constructing the segmentation model, culminating in its adept execution for COC visual segmentation.

tion then aids in defining a region of interest (RoI), which is visually seen as a bounding rectangle containing the COC. Subsequently, the RoI snapshot is introduced into the U-Net model in the latter phase, which undertakes segmentation to enhance the initial pre-trained segmentation outcome. By harnessing the insights gained from the features and the intricacies of the U-Net, this phase achieves an enhanced contouring of the COC. The integration of transfer learning, a two-phase segmentation technique, and the U-Net framework within the proposed schema contribute to COC segmentation's robustness and high performance in bright-field microscopy images.

4.3.3.1 Architecture

The U-Net model, introduced by Ronneberger et al. [181], was the chosen architecture for the entirety of the experiments carried out in this research. The decision to use U-Net is grounded in its well-documented history of producing commendable results when segmenting biomedical images [8]. This model is particularly adept at handling varying image sizes and effectively differentiating intricate structures, which makes it especially suitable for biomedical segmentation challenges. As demonstrated in Figure 4.10, this architecture is characterized by its symmetrical U-shape, comprising a contracting pathway that captures context and an expansive pathway that refines localization. Our experiments harnessed the U-Net architecture's inherent strengths to achieve precise and accurate segmentation results.

The U-Net model's contracting phase in our design encompasses four blocks. Each block is structured with dual 3×3 convolutional layers, succeeded by a ReLU activation function and a 2×2 max-pooling operation with a stride set at 2. We've incorporated an extra block that mirrors the earlier ones but augments them with a dropout layer, set to a 0.5 probability.

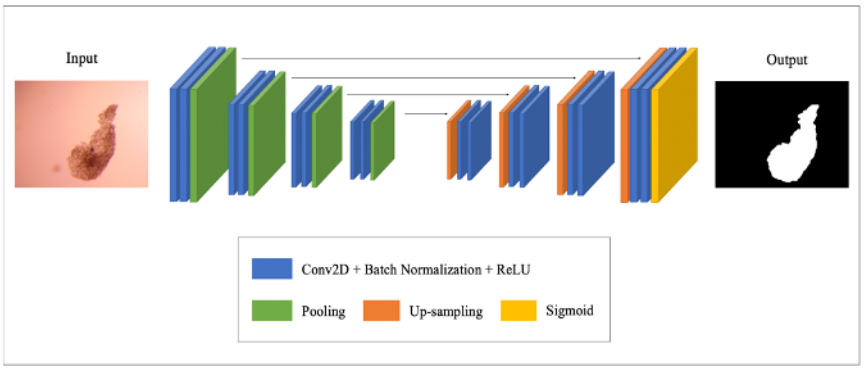


Figure 4.10: The architecture of a U-Net model specifically designed for learning to segment COCs. The U-Net architecture, depicted in the figure, consists of four levels of the contracting path and four levels of the expanding path, followed by a sigmoid function.

Analogously, the expansion phase is constructed with four blocks. Each of these blocks integrates an up-sampling mechanism using a transposed convolutional layer. They also incorporate a layer for concatenation, which merges feature maps from the equivalent depth in the contraction phase, followed by twin 3×3 convolutional layers and a ReLU activation function. Culminating this design is the last convolutional layer, producing the segmentation's final output.

To implement the proposed model, we turned to the Keras[3] open-source package with TensorFlow[4] serving as the backend platform. Keras provides a user-friendly interface for designing and training deep learning models, while TensorFlow offers efficient computational support and optimization for training and inference processes.

4.3.3.2 Loss function and metrics

The performance of the suggested segmentation technique is evaluated with the Dice coefficient, as mentioned in [46]. The Dice coefficient measures spatial overlap between two areas, ranging from 0 to 1. A score of 0 implies no common area, whereas a score of 1 denotes complete overlap. The formula of the Dice coefficient is provided in Equation 4.1.

$$dice(f, x, y) = \frac{2 \sum_{ij} f(x)_{ij} y_{ij}}{\sum_{ij} f(x)_{ij} + \sum_{ij} y_{ij}} \tag{4.1}$$

In the given equation, y stands for the ground truth, x denotes the input image, and $f(x)$ signifies the model's prediction. For training the weights of the U-Net architecture, we picked the Dice loss metric derived from the Dice coefficient. The formula for the Dice loss is as follows:

$$\mathcal{L}_{dice}(f, x, y) = 1 - dice(f, x, y) \tag{4.2}$$

[3]https://keras.io/
[4]https://www.tensorflow.org/

By employing the $\mathcal{L}_{dice}$ metric, the model aims to reduce the divergence between its predicted segmentation and the actual ground truth. This fosters greater Dice coefficient scores and subsequently enhances the performance of the segmentation.

Other loss functions and metrics Various loss functions, suitable for segmentation tasks, were explored alongside dice loss during our experiments in this chapter. However, several were ultimately not adopted for specific reasons. Details on these loss functions are explicitly described in Chapter 3.3.2. We recommend consulting that chapter for more in-depth information about these loss functions.

4.3.3.3 Transfer learning

Transfer learning, as introduced in Section 3.6, is a method in machine learning where knowledge from a model trained on one task is leveraged to enhance performance on a different but related task. This technique can reduce the amount of data needed for training new models. This is especially beneficial when data is scarce or annotating it is time-consuming or costly.

In our research, annotating images for the specific goal of segmenting cumulus-oocyte complexes is a laborious task that demands the expertise of domain specialists. Yet, in this dataset, these specific annotations are not always available. We used an open-source melanoma image dataset from the ISIC Archive[5] to navigate this obstacle. The models, already trained on this dataset, are further refined using a limited batch of images and their annotations tuned toward segmenting cumulus-oocyte complexes.

Through transfer learning, our method uses the knowledge gained by the original model trained on melanoma images, reshaping it for our specific task. This strategy paves the way for creating proficient models for segmenting cumulus-oocyte complexes. It minimizes the need for large datasets and expedites the annotation task, saving domain experts' knowledge and time.

4.3.3.4 Statistical analysis

To juxtapose the outcomes of the DL method against those of human evaluators, the agreement among each expert's annotations was calculated and contrasted with the agreement between the DL method and each expert. Several metrics gauged the consistency between AI-xpansion and human evaluators, including average rank, bias, and variances. The primary objective was determining how frequently AI-xpansion's assessments mirrored those of the experts. In this regard, the average rank metric was utilized. In this context, when taking Observer 1 as a benchmark, the ratings of Observer 2, observer 3, and AI-xpansion are compared against Observer 1. Rankings range from 1 to 3 based on which assessment is closest to the benchmark. This procedure is replicated for all evaluators. A lower value denotes a more accurate estimation, per Equation 4.3, where r is the rank with values, and N is the total size of the dataset. Further, biases and variances between AI-xpansion's estimates and those of the experts were scrutinized, referencing Equations 4.4 and Equation 4.5, where $q_{ij} = ln(expansion_{ij})$, i being the distribution of interest, j being the distribution of reference, n being the sample variable, and N being the total size of the dataset.

[5] https://challenge.isic-archive.com/data/

$$average\ rank = \sum_{i=1}^{N} r_i \tag{4.3}$$

$$bias_{ij} = \sum_{n=1}^{N} \frac{q_n^i - q_n^j}{N} \tag{4.4}$$

$$variance_{ij} = \sum_{n=1}^{N} \frac{(q_n^i - bias_{ij} - q_n^j)^2}{N} \tag{4.5}$$

4.3.4 Results

4.3.4.1 Processor capacity

Given that the 'Area' method yielded the highest ICC values for manual measurement of cumulus expansion, a deep learning (DL) model named AI-xpansion was developed based on this approach. In the pre-processing phase, where the region of interest was identified, the model successfully segmented 98% of the intended regions. Detection failures occurred due to the very low signal of the COC and the interference of an oil droplet in the image. Consequently, these 2 instances were omitted from subsequent analyses and image enhancement. Finally, a total of 100 COCs were harnessed to train the model, detailed further in Section 4.3.2.

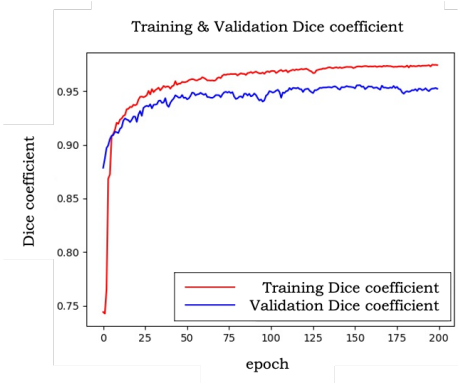
4.3.4.2 Evaluation procedure

To evaluate the segmentation models, we employed a 10-fold cross-validation technique. This was chosen due to the dataset's limited size, ensuring robust results. In each fold, 90 oocytes (equivalent to 180 images) were selected as the training set, while 10 oocytes (or 20 images) were set aside for validation. The model was trained using minibatches of 32 images across 200 epochs. Each fold generated masks for the 20 validation images, reaching 200 masks once all folds were processed. These resultant masks were then juxtaposed against the expert-provided masks. The dice coefficient was utilized to quantify the similarity between the two masks.

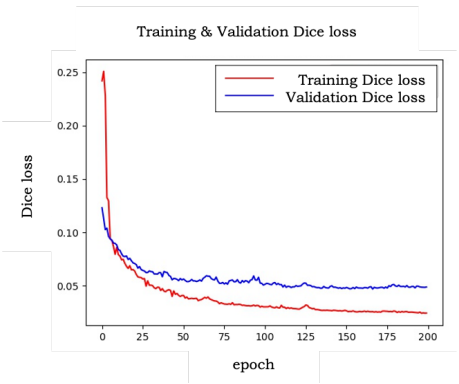
Figure 4.11 visually represents the evolution of the mean dice coefficient alongside the mean dice loss over the ten folds. The mean dice for the majority-vote model exhibited high performance, settling around a convergence point of approximately 95%.

A comparison of our method with human annotators is crucial. To achieve this, we assessed the annotations' similarity by each of the three experts and then matched these findings against the similarity between each human expert and our recommended approach. Since there was a non-normal distribution of dice values, we employed the dice coefficient's median as the metric to evaluate the similarity between two sets of annotations. For a perfect similarity, this median should touch 100%, while it should plummet to 0% for a complete disagreement among annotators.

In Table 4.3, the initial three rows illustrate the degree of similarity in the cumulus oocyte complex segmentations among different pairs of human annotators. The agreement levels span from



(a) Dice coefficient convergence to 0.95.



(b) Dice loss convergence to 0.05.

Figure 4.11: The evaluation of the model using the mean Dice coefficient and mean Dice loss as performance metrics. The figure showcases the convergence of these metrics during the training process.

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	-	95.15%	95.49%
Annotator 2	95.15%	-	95.63%
Annotator 3	95.49%	95.63%	-
Our proposal	**96.32%**	**95.99%**	**96.48%**
without RoI	95.95%	95.61%	95.97%
without RT	13.17%	13.55%	14.00%
without TL	no convergence		

Table 4.3: Comparison of the median of dice coefficients of different models with the ones of human specialists.

95.15% to 95.63%. Contrastingly, our proposed deep learning model achieves similarity levels between 95.99% to 96.48%, surpassing human experts' consistency. This suggests that our model's outcomes are nearly indistinguishable from human expert's segmentations. Given the time and cost associated with human annotation, our technique offers a cost-effective and efficient alternative to relying solely on human experts.

To discern the contribution of each step in our workflow – specifically, ROI determination and transfer learning – we executed an ablation study by omitting each component.

The first modification entails using our model without the preliminary stage of detecting the region of interest (ROI) but using transfer learning from a melanoma dataset. The performance from this setup ($withoutROI$, as listed in Table 4.3) remains reasonable, mirroring the consistency seen among human experts, with values between 95.61%-95.97%; it is just slightly lower than the proposed model. On the other hand, solely utilizing the model trained on the melanoma dataset without subsequent retraining to the COC dataset yielded considerably poorer results – between 13.17% to 14.00% ($withoutRT$ in Table 4.3). Removing transfer learning and initializing with random weights resulted in a non-convergent model, underscoring the crucial role of transfer learning in this scenario, likely due to our restricted set of annotated images.

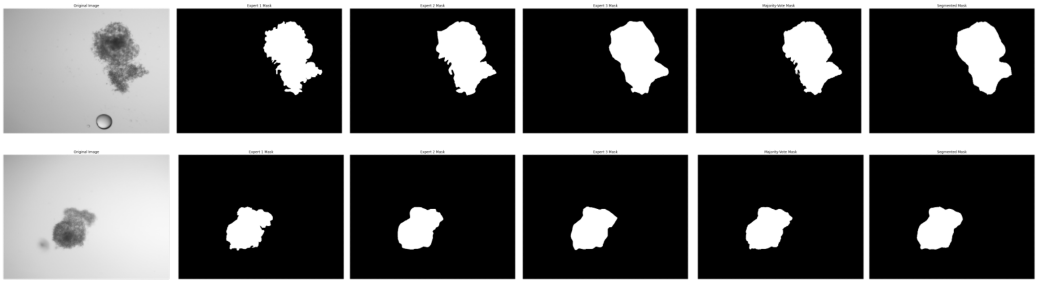


Figure 4.12: A visual comparison between the experts' masks and the model's predictions for segmenting COCs. Side-by-side subplots are displayed, each containing two images, the annotators' individual masks, the majority-vote mask, and the segmented mask.

In Figure 4.12, we illustrate the outcomes of the segmentation procedure. This figure showcases a sample of cumulus-oocyte complexes as the baseline image. Following this, the next column presents the three expert-generated masks. A clear observation is that these masks don't perfectly match, especially at the oocyte's borders. While some annotations present a loose outline, others suggest a more strict boundary. Such discrepancies can make model training challenging when the foundational annotations are controversial. To mitigate this, the subsequent column offers the masks formed by the majority consensus from the three expert annotations. This method seeks to bridge the interpretation gap and deliver a unified representation. Lastly, the final column presents the masks produced by our proposed model. Visually, it aligns closely with the majority-vote mask while recognizing and integrating key details. Notably, the model's approach provides a smoother boundary than that of annotator 1, yet retains vital features. Through these visual representations, we gain insights into the segmentation workflow and witness the ability of the model to identify essential features of the region of interest.

4.3.4.3 Performance

The comparative ranking of scores among the observers is detailed in Table 4.4. In two of the three scenarios, AI-xpansion outperformed the other observers. The efficiency of AI-xpansion in measuring the area of the COC was similar to that of human observers ($p = 0.15$).

Differences in bias and variance among human observers, and between them and AI-xpansion, are outlined in Table 4.5 and Table 4.6. When assessing cumulus expansion using AI-xpansion, the system displayed reduced bias and variance one-third of the time compared to human evaluators. In the other two scenarios, AI-xpansion's measurements for bias and variance were comparable to those of human observers. This demonstrates that AI-xpansion achieves performance levels akin to human capabilities.

4.4 Embryo development

Finally, in this section, we present the fourth and fifth research questions:

Q4: *How does the optimal method for measuring COCs' expansion impact healthy embryo development in mammalian species?*

	$\mathbf{O}_1$	$\mathbf{O}_2$	$\mathbf{O}_3$
$\mathbf{O}_1$		2.04	1.93
$\mathbf{O}_2$	2.04		2.15
$\mathbf{O}_3$	1.97	2.11	
AI-xpansion	1.99	1.85	1.93

Table 4.4: Average rank calculated in the comparison among the different human observers and between human observers and AI-xpansion. This table shows the similarity between the average ranking of three observers (O_1, O_2, O_3) and the deep learning method (AI-xpansion). Scores closer to zero indicate that the performance is closer to the column observer.

	$\mathbf{O}_1$	$\mathbf{O}_2$	$\mathbf{O}_3$
$\mathbf{O}_1$		0.0475	-0.0097
$\mathbf{O}_2$	-0.0475		-0.0572
$\mathbf{O}_3$	0.0097	0.0572	
AI-xpansion	-0.0214	0.0260	-0.0312

Table 4.5: Bias calculated in the comparison between the different human observers and between human observers and AI-xpansion. It shows the level of simplifying assumptions made to approximate the reference better. Scores closer to zero indicate that the performance is closer to the reference.

	$\mathbf{O}_1$	$\mathbf{O}_2$	$\mathbf{O}_3$
$\mathbf{O}_1$		0.0092	0.0047
$\mathbf{O}_2$	0.0092		0.0115
$\mathbf{O}_3$	0.0047	0.0115	
AI-xpansion	0.0076	0.0070	0.0082

Table 4.6: Variance calculated in the comparison between the different human observers and human observers vs AI-xpansion. It shows how much change the performance will have when it comes across different data. Scores closer to zero indicate that the performance is closer to the reference.

Q5: *What influence does the density of COCs, as determined by the optimal method, have on healthy embryo development in mammalian species?*

4.4.1 Resources

Bovine COCs (totaling 427 across 14 replicates) were collected and underwent in vitro maturation per the procedure outlined in Section 4.2.2.2. The additional criteria for selection included oocytes encircled by a dense cumulus consisting of a minimum of 5 layers, along with a uniformly dark or slightly grainy ooplasm. Frozen semen from a bull with confirmed fertility was defrosted in a water bath set at 38 °C. It was then processed through a Percoll gradient (from GE Healthcare Biosciences, Uppsala, Sweden) to isolate viable sperm cells. These COCs were then paired with the chosen spermatozoa in separate droplets (each holding 1 COC and measured at 20 μL). This was done in IVF-Tyrode's Albumin Lactate Pyruvate (TALP) medium that had added bovine serum albumin (BSA; Sigma A8806 at 6 mg/ml) and heparin (concentration: 20 $\mu g/mL$), up to a cumulative concentration of 1 $\times$ 106 spermatozoa/mL. The droplets were sealed with paraffin oil and were incubated for 22 hours at 38.5 $^{\circ}C$ in an environment with 5% CO_2 humidified air.

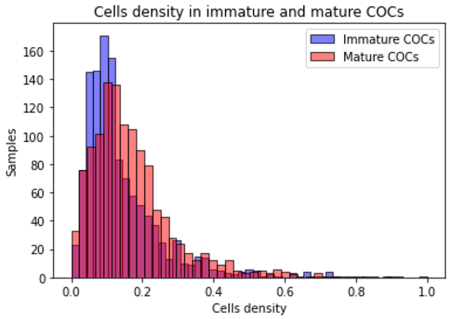
Post in vitro fertilization, the cumulus cells were eliminated using delicate pipetting (using a 140 μm EZ-Tip®, sourced from CooperSurgical, Malov, Denmark), and surplus sperm cells were discarded by immersing the presumed zygotes in the HEPES-TALP medium. The following step involved transferring the zygotes individually into droplets (1 zygote per 20 μL droplet) containing synthetic oviductal fluid medium, which was enriched with 0.4% BSA (Sigma A9647) and ITS (consisting of 5 $\mu g/mL$ insulin, 5 $\mu g/mL$ transferrin, and 5 ng/mL selenium). The embryos were then nurtured up to the eighth day post-fertilization, in conditions set at 38.5 $^{\circ}C$, with 5% CO_2, 5% O_2, and 90% N_2. On the eighth day, the progress of embryonic development was documented based on the blastocyst rate, which signifies the proportion of embryos that matured to the blastocyst stage (identified by the emergence of a blastocoel or cavity) to the total count of presumed zygotes.

COC images were captured before and after the maturation process, as earlier mentioned in Section 4.2.2.3. The assessment of cumulus expansion was conducted by a single observer employing the area, 3-distances, and scoring technique as detailed in Section 4.2.2.4, and also through the AI-xpansion method outlined in the current section.

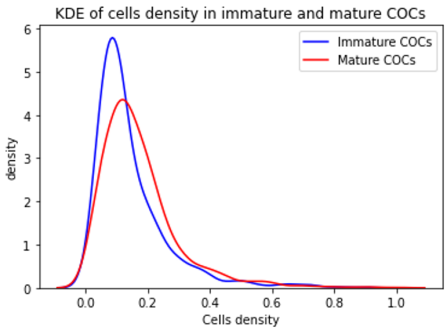
4.4.2 Proposal

4.4.2.1 COC expansion and embryo development

The relationship between cumulus expansion and embryonic development was studied using the Mann-Whitney U test [146], with a significance threshold set at $\alpha = 0.05$. This test juxtaposed the cumulus expansion of COCs that either failed or succeeded in development. The base assumption, H_0, posits that both sample sets hail from an identical distribution. Conversely, the alternative hypothesis suggests one sample exhibits greater expansions than its counterpart.



(a) Histogram showing the distributions of cell density for immature and mature COCs.



(b) Kernel density estimation (KDE) of cell density for immature and mature COCs.

Figure 4.13: Cell density distribution for immature and mature COCs.

4.4.2.2 COC density and embryo development

In Figure 4.13a and Figure 4.13, two graphs illustrate the cell density distribution across the given samples. A noticeable trend emerges, indicating a subtle increase in cell density during the transition from immature to mature stages.

We employed the Mann-Whitney U tests to further interpret these findings, comparing cell density across varied situations. We considered two separate scenarios: a set comprising 331 oocytes and a more extensive collection of 922 oocytes. It's worth highlighting that the eventual sample count shrunk by 15%, necessitated by data anomalies and the subsequent removal of specific cases.

4.4.3 Results

4.4.3.1 COC expansion and embryo development

The correlation between embryo development and cumulus expansion –measured via the area method, 3-distance method, scoring method, and AI-xpansion – was investigated to identify the accuracy of these methods. Successful embryo development was identified when embryos progressed to the blastocyst stage on day eight after fertilization. Out of 427 anticipated zygotes, 118 matured into blastocysts (27.6%), while 309 (72.4%) halted at a preceding developmental stage.

Notably, the median cumulus expansion was significantly less in embryos that didn't reach the blastocyst stage compared to those that did when assessed using (i) AI-xpansion (64.95 (IQR: 48.30) and 77.88 (IQR: 55.96) respectively, with p = 0.011), (ii) area method (72.15 (IQR: 58.55) and 80.75 (IQR: 62.81) respectively, with p = 0.024), and (iii) 3-distance method (46.83 (IQR: 49.74) and 55.20 (IQR: 50.50) respectively, with p = 0.046).

However, when evaluated by the scoring method, the median cumulus expansion showed no significant difference between the halted embryos and the blastocysts, with values of 2.00 (IQR: 2.0) and 2.00 (IQR: 1.0), respectively, and p = 0.256. These findings are illustrated in Figure 4.14.

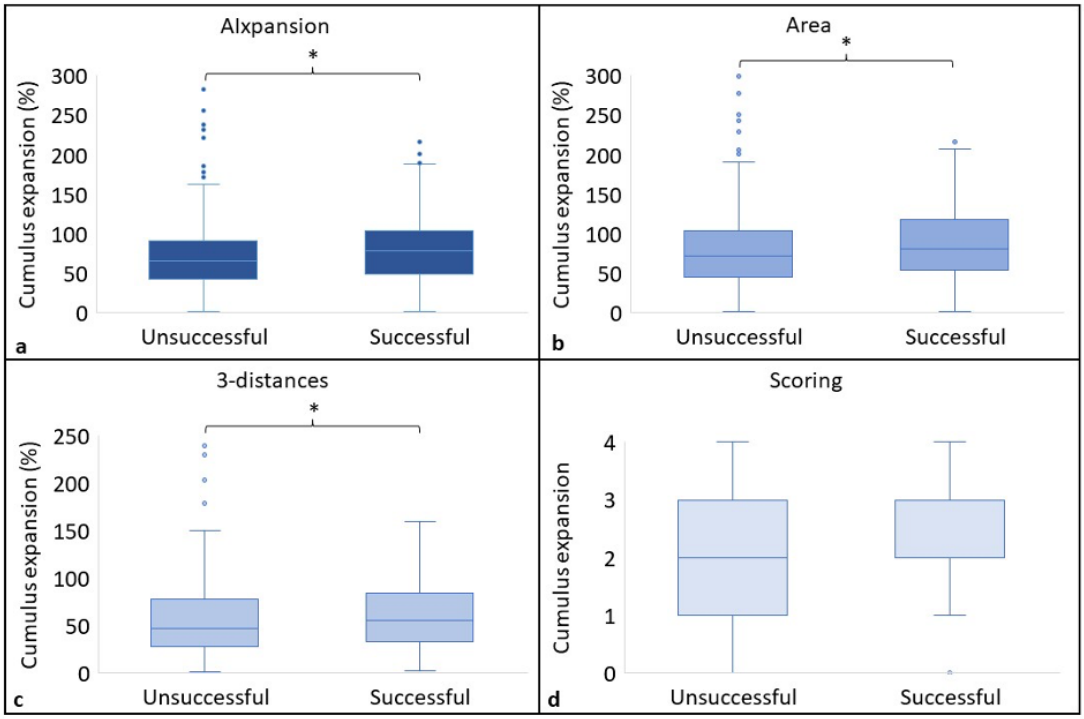


Figure 4.14: Distribution of cumulus expansion in cumulus-oocyte complexes that resulted in unsuccessful and successful embryo development. Cumulus expansion was measured using (a)AI-xpansion, (b)the area method, (c)the 3-distance method, and (d)the scoring method. Cumulus expansion significantly differed between unsuccessful and successful COCs ($p \leq 0.05$)) when being measured by AI-xpansion, area method, or 3-distance method.

4.4.3.2 COC density and embryo development

The outcomes for the previously described comparisons are gathered in Table 4.7. A statistically significant difference in cell density between immature and mature oocytes was observed from the reduced dataset. The significance level was larger for mature oocytes, whereas the immature ones approached the defined significance threshold. Yet, when relative cell density was evaluated, it failed to exhibit a significant difference, implying its potentially absent role in ensuring embryo health.

Interestingly, on augmenting the dataset by nearly 2.5 times, the results proved to have a higher significant difference. Our analysis revealed a strong correlation between cell density and healthy embryo development, as reflected by p-values of 0.0032 and 0.0003 for immature and mature COCs, respectively. Furthermore, the p-value tied to the relative density change reached a number of 0.0029, indicating its role as a significant factor in most instances.

Drawing from these observations and the evidence from the enlarged dataset, we conclude that cell density, especially within mature cumulus-oocyte complexes and, to an extent, in the immature variants, holds considerable influence over healthy embryo development.

	Cell density (I)	Cell density (M)	Relative change
$p - value_{331}$	0.0155	0.0063	0.1498
$p - value_{922}$	0.0032	0.0003	0.0029

Table 4.7: The p-values for the cell density parameters in two separate datasets: one comprising 331 oocytes and the other containing 922 oocytes. The p-values indicate the level of statistical significance for each parameter in relation to healthy embryo development. The parameters studied here are the cell density (I) in immature COCs, the cell density (M) in mature COCs, and the relative cell density change.

4.5 Discussion

This chapter raises the following research questions. Firstly, it is sought to determine the most effective method for assessing the COC area. Furthermore, it is questioned whether evaluating the COC area can be automated. Whether a model can be established to calculate COC expansion in mammalian species is also pondered. Last, the significance of cumulus expansion in achieving healthy embryo development is examined, followed by the impact of cumulus density on forming a healthy blastocyst.

4.5.1 Best measurement method

This research examined the reliability of three commonly employed techniques for measuring cumulus area and expansion using the Intraclass Correlation Coefficient (ICC). While the Likert scale is often the go-to method for such assessments, it displayed high subjectivity in our study. Comparing the area before and after in vitro maturation yielded the most consistent results among different observers and repeated measurements by the same observer (inter-observer and intra-observer agreement). As a result, the area method was chosen as the most accurate and the foundation for creating a deep learning algorithm for future evaluations.

4.5.2 AI in ART

Recent advances in Assisted Reproductive Technology have witnessed an escalating adoption of deep learning and image segmentation techniques. While some strides have been made in identifying morphological characteristics from human oocyte and embryo bright-field microscopy images, applications in other mammalian species and how several key components affect healthy embryo development are sparse. Segmenting bovine cumulus-oocyte complexes and linking them with cumulus expansion and density remained uncharted territory.

4.5.3 Area method and deep learning

Following, a supervised detection technique for the cumulus was employed, leveraging transfer learning from a closely related domain – melanoma images to assist the fairly small COC dataset. The Dice coefficient scores revealed that the top-performing model, which was trained

using majority-vote annotations, holds promise as it achieved scores comparable to human annotations. Upon closely examining the complexities of the task, it was observed that experts show slight variations in their annotation or evaluation of cumulus-oocyte image datasets. Some experts are highly strict, striving for maximum accuracy without concern for the time invested. In contrast, others aim for quicker results, which may lead to less detailed and potentially less accurate annotations. To address the issue of minor disagreement among multiple annotators, we introduced a majority-vote model. This model determines whether a pixel belongs to the cumulus-oocyte based on the consensus among the majority of experts. According to the median Dice coefficient metrics, our proposed deep learning model outperforms human capabilities, as demonstrated in previously. Notably, despite the small dataset size and the inconsistencies among experts on what should be categorized as a part of the cumulus-oocyte, the deep learning model still delivers high and consistent performance. This offers a more accurate and time-efficient solution.

4.5.4 Expansion and blastocyst

Based on the merits of this area method and the deep-learning model built, a deep-learning pipeline named AI-xpansion was developed. The pipeline, designed to measure the COC area and then the expansion from immature to mature oocyte, was performed with reliability comparable to human observers. Importantly, the findings revealed that cumulus expansion, as measured by AI-xpansion, the area method, and the 3-distance method, could be pivotal in predicting oocyte developmental competence. All three methods exhibited a significant association between embryo development and the cumulus expansion of the oocyte.

4.5.5 Density and blastocyst

Last, in this study, the importance of cell density in the maturation of oocytes was studied. Our analysis found marked differences in cell density between oocytes that successfully matured into healthy embryos and those that did not, indicating that cell density is a key factor in influencing the developmental viability of oocytes. Along with cumulus expansion, cumulus density is another important feature in healthy embryo development.

4.5.6 Future

These analyses were performed under the specific frame of bovine cumulus-oocyte complexes. Having proved its functionality and worthiness, the pipeline provided, including the AI-xpansion framework and the study of expansion and density to healthy embryo development, could be expanded to other related species. Starting with human [44], and moving on to other mammalian species, such as horses [195] and wildlife [40], where in-vitro maturation is performed, the algorithms could be developed to be a crucial tool for assisted reproduction, for studying the insights of infertility, and for preserving endangered species with better understanding and use of assisted reproductive techniques.

Additionally, these models described here could be the starting point for a more general platform that segments COC areas and be used to study more features than the one described here, such as the position of the oocyte and the surrounding cells and the link to healthy development. Last,

generalizing the current pipeline to other microscopy settings or different image-capturing equipment could be an important step for automating the process and saving a great amount of time for the experts in the field to devote their time to more sophisticated tasks.

4.6 Conclusions

In this section, it was established that the area method is the most reliable for measuring cumulus expansion, while the commonly referenced scoring method was found to be the least reliable. Subsequently, the area method inspired the development of an objective alternative, the AI-xpansion DL algorithm, which was created following the establishment of a framework for segmenting COC areas. AI-xpansion is poised to be a valuable asset for embryologists and researchers in in-vitro embryo development laboratories for accurately measuring cumulus expansion with accuracy comparable to human judgment. The importance of measuring cumulus expansion was validated, as it was demonstrated through the use of AI-xpansion, the area, and the 3-distance method that median cumulus expansion is significantly greater in competent COCs than in COCs that do not develop into blastocysts. Additionally, the study explored the impact of COC density on the potential of an oocyte, and then an embryo, into a healthy blastocyst. The upcoming chapter will explore more advanced methods of extracting ground truth in medical images to enhance the realism of DL model performance in segmentation tasks.

Chapter 5

Consensus Segmentation

In this chapter, we delve into a research question that is important to overcome some obstacles in our investigation:

- **Q6:** How can expert biases in medical image segmentation be effectively overcome?

Research question **Q6** presented below expands the paper: Athanasiou, Georgios, Josep Lluis Arcos, and Jesus Cerquides. "Enhancing Medical Image Segmentation: Ground Truth Optimization through Evaluating Uncertainty in Expert Annotations." *Mathematics* 11.17 (2023): 3771.

5.1 Introduction

Assessing the efficiency of image segmentation methods remains a consistent hurdle. While expert annotators often handle interactive segmentation, it results in differences in their interpretations. Automated systems strive to remove these inconsistencies, but measuring their success with medical images is challenging because there isn't a clear "gold standard" for segmentation.

Various methods are suggested for determining annotator expertise and establishing ground truth (GT) labels. These techniques fall into two primary categories: a *two-stage approach* and a *simultaneous* approach.

The *two-stage* methods involve two distinct steps: first, the aggregation of labels, and then, training a supervised learning model. Initially, noisy labels are compiled using a probabilistic model that treats annotator proficiency and GT labels as hidden variables to be deciphered. Afterward, a machine learning model is trained using the GT labels to execute the desired task. Notably, these methods often overlook details about the raw inputs, X, in the generative model for noisy labels during the aggregation phase, which can adversely affect the determination of accurate labels.

The *simultaneous* methods tackle this shortcoming by incorporating the predictions of the supervised learning model, specifically the distribution $p(Y \mid X)$, into the probabilistic model for noisy labels. Studies have indicated that these methods enhance predictive accuracy. Such techniques utilize versions of the expectation-maximization (EM) algorithm during their training process and necessitate a significant number of labels for every sample. Yet, in many real-world scenarios, gathering an extensive number of labels for each instance is practically unfeasible, constraining the applicability of these methods.

5.2 Recent literature

5.2.1 Two-step approach

Label aggregation and supervised learning model training are carried out independently in the two-stage methodology. The noisy labels are initially consolidated using a probabilistic model that considers annotator skills and ground truth labels as unknown variables to be calculated. Subsequently, a machine learning model is trained on the ground truth labels to perform the task of interest. Regrettably, these methods often overlook the raw input information during label aggregation, adversely affecting the estimated true labels.

In 2002, Warfield et al. [222] initially proposed the STAPLE algorithm, which offers a unique way to characterize each annotator w via their confusion matrix $\theta_w \in \mathbb{R}^{L \times L}$, where L is the number of classes, and $\theta_{w,c',c}$ denotes the probability that expert w labels a pixel as c' given c in consensus. Here, the confusion matrix denotes the probability $\theta_{w,c',c} = P(S_{w_{i,j}} = c' | C_{i,j} = c)$ that an expert w labels a pixel as c', given that the same pixel is labeled as c in the consensus. The STAPLE algorithm harnesses the power of the Expectation Maximization technique [223] to derive the maximum likelihood consensus segmentation. STAPLE has since witnessed several enhancements. Asman and Landman [5] extended STAPLE into Spatial STAPLE. This algorithm accounts for spatially varying performance by augmenting the performance level parameters to generate a smooth, pixel-wise performance level field unique to each annotator. Concurrently, Commowick et al. [41] introduced Local MAP STAPLE, an approach that tackles spatially variant performance using a sliding window technique. Furthermore, STAPLE has been improved by incorporating the image's intensity in the consensus computation. Asman and Landman [6] proposed a novel statistical fusion algorithm, Non-Local STAPLE (NLS), which reformulates the STAPLE framework from a non-local means perspective.

However, Hamzaoui et al. [88] noted that STAPLE is highly sensitive to the size of the background and to the consensus prior; to mitigate this issue, they introduced MOJITO, an approach that defines consensus segmentation as one that minimizes the Fréchet variance with respect to the set of annotators. In Carass et al. [31], the STAPLE algorithm and a wide range of evaluation metrics were implemented as part of their challenge. Within this challenge, a Consensus Delineation was selected by employing the STAPLE algorithm, which integrated information provided by two experts and the 14 distinct participating groups, each with their algorithms. This Consensus Delineation facilitated a direct comparison not only between the experts themselves but also among the algorithms proposed by the different groups, but the study concluded that the experts still outperformed the algorithms in their challenge.

5.2.2 Simultaneous approach

Simultaneous approaches aim to address this issue by integrating the prediction of the supervised learning model with a model that handles uncertain or noisy labels. This combination has been demonstrated to improve prediction accuracy. These methods utilize variants of the expectation-maximization algorithm during their learning process, and they require an adequate quantity of labels for each example. However, collecting many labels for each example is often impractical in real-life situations, which limits their use.

The models in this category strive to simultaneously curate labels and train a supervised model in a synchronized, end-to-end manner. This strategy allows the two components to complement and enhance each other. Although current evidence relates primarily to basic classification tasks, these concurrent approaches have demonstrated encouraging enhancements over the methods in the first category in both the predictive capabilities of the supervised model and the sample efficiency (i.e., fewer labels per input are needed). However, up to this point, the same issue has received minimal attention in more complex, structured prediction tasks, where the outputs have high dimensionality.

Zhang et al. [240] present an unprecedented end-to-end supervised segmentation method. This method uniquely estimates the reliability of multiple human annotators and the true segmentation labels using solely noisy labels. While they evaluate their methodology in several synthetic scenarios, it has not been tested on actual medical data. In subsequent work, Zhang et al. [239] continue to advance this approach, optimizing specific parameters, but only with crowdsourcing data. On the other hand, Liu et al. [133] shift their focus to learning dynamics — a concept previously studied in classification problems but scarcely explored in the realm of image segmentation. They propose the ADELE model, but the results are not grounded in the core issue of multiple annotators with varying annotation methods.

Several other approaches have been proposed [82, 183, 220, 172], but none of them fulfill the requirements of actual medical data from multiple expert sources.

5.3 Materials and Methods

5.3.1 Dataset

Images of 100 Cumulus Oocyte Complexes (COCs) were meticulously captured at two critical stages: the immature stage (0 hours of maturation) and the mature stage (22 hours post-maturation). This endeavor was achieved utilizing an advanced Olympus stereomicroscope seamlessly integrated with a high-definition ToupCam camera. The entire process was efficiently streamlined with the aid of the ToupView software (version 3.7.13270.20181102). To ensure consistency and clarity, all images were captured with a uniform magnification of 56×, strategically positioning a single oocyte's zona pellucida at the epicenter of the field of focus. Each image was saved in the .png format, boasting a crisp resolution of 2592 $\times$ 1944 pixels and presented in vibrant RGB color. Three experts were then entrusted with meticulously annotating the COC area from 100 paired images using the renowned ImageJ software (version 1.53j). This effort resulted in a comprehensive collection of 200 images and 200 $\times$ three individual masks. The images and masks underwent downscaling for optimization, achieving final dimensions of 192 $\times$ 240 $\times$ 1.

The esteemed method of majority vote, as delineated by Nguyen et al. (2022) [156], was employed to derive a Ground Truth (GT) for evaluation objectives. For a visual representation, a sample of the data — provided by the Department of Internal Medicine, Reproduction, and Population Medicine at Ghent University (accessible at https://www.ugent.be/di/irp/en, as of 09 November 2021) — is presented in Figure 5.1.

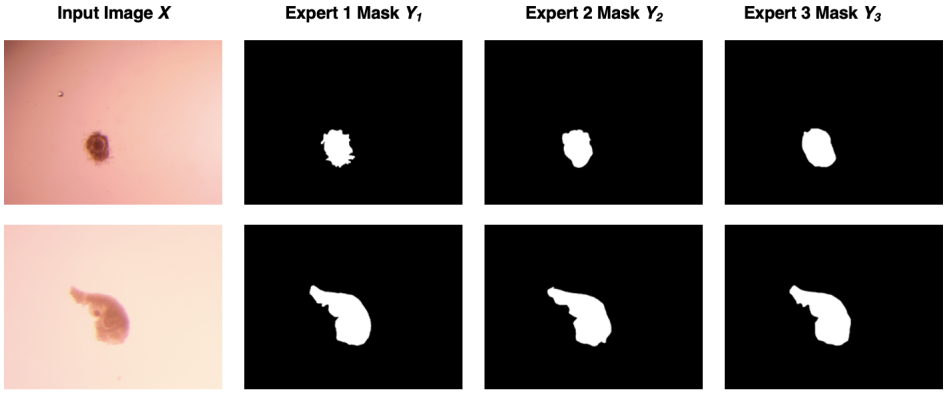


Figure 5.1: A sample of the dataset. The first column contains a COC in both immature and mature stages. The subsequent three columns represent the corresponding masks provided by different experts. It is obvious that there is no complete agreement among the experts for each case.

5.3.2 Architecture

The method was developed by using the PyTorch framework. For the segmentation part, the network is designed based on the UNet architecture, which comprises five layers. The channel count in these layers increases in a sequence, starting from 32 and doubling each time until it reaches 512, then decreases again. The regularization network uses a Softmax function [75] for the basic global CM scenario. However, when dealing with the more intricate local CM scenario, it adopts a double convolutional block, comprising a convolution operation followed by a ReLU function, as illustrated in Figure 5.2.

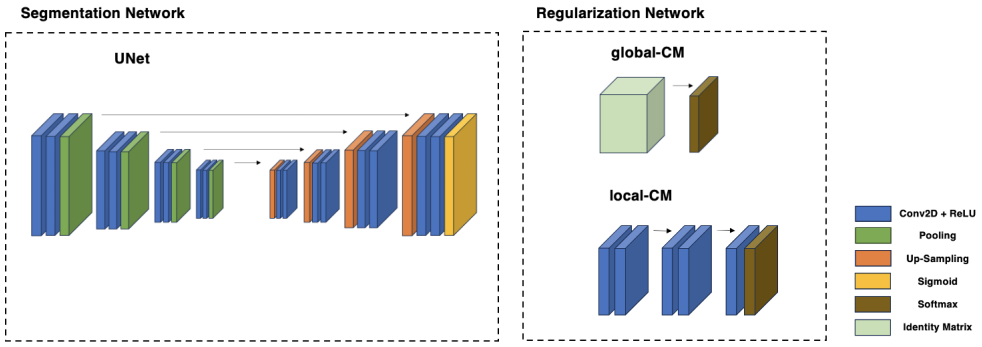


Figure 5.2: The segmentation network structure consists of a UNet architecture parameterized by θ and the regularization networks parameterized by ϕ. The UNet has a depth of 5 layers, with the number of channels moving progressively from 32 to 64, 128, 256, and finally, to 512. The maxpooling layer has a padding and stride of 2, while the upsampling layer has a kernel size and a stride of 2. The regularization networks contain a simple network to compute the global confusion matrices and a CNN to compute the local confusion matrices.

The available dataset contains 200 images, which we split into different segments: 80% was allocated for training, 10% for validation, and the remaining 10% for testing. The training spanned

100 epochs, and the learning rate was initialized at 1e-3, gradually reducing it to 1e-6. A batch size of 16 was used during training, and the Adam optimizer [116] was employed for the process.

5.3.2.1 Segmentation network

The segmentation model is driven by a UNet architecture [181]. This model, characterized by the parameter θ, yields a predicted probability distribution expressed as $p_\theta(\mathbf{X}) \in [0,1]^{WxHxL}$, where W represents the image's width, H its height, and L the total number of classes. This outputted distribution conveys the probability of a pixel being associated with a particular class. For this specific application, the classification consists of just two classes (i.e., $L = 2$): one for the COC area and the other for the background.

UNet is renowned for its unique U-shaped design, making it a preferred choice for medical image segmentation. This is because the architecture captures both low-level and high-level features, thereby improving segmentation precision. The human-level performance of UNet in tasks related to COC has been documented previously [7]. Moreover, UNet showcases optimal results even when trained on limited datasets, rendering it an ideal fit for our current challenge. The UNet's visual representation can be found on the left side of Figure 5.2, which is connected to the subsequent regularization network.

5.3.2.2 Regularization network

Based on [240], the regularization network, as depicted in Figure 5.2, is characterized by the parameter ϕ. Its main objective is to ascertain the genuine probability distribution of segmentations. This can be accomplished by replicating the unique tendencies of the three annotators illustrated in Figure 5.3. Here, 'tendencies' indicate the individual propensities of annotators to gravitate toward certain erroneous segmentation patterns, for example, presuming pixels belong to the COC zone rather than the background. Such tendencies are represented through confusion matrices (CMs). Assuming that the true label is previously known, then the CM is constructed as follows:

$$a_{ij}^{(m)}(\mathbf{x}) = p(y^{(m)} = i \mid y_{GT} = j, \mathbf{x}) \tag{5.1}$$

where $a_{ij}^{(m)}$ stands for the element in the (i, j) cell of CM for expert m. In an image of size WxH, for every pixel $w{\in}W$, $h{\in}H$, the CM could be the same among all pixels (global-CM case), or change for each pixel (w, h) of each image $\mathbf{x}$ (local-CM case). Consequently, in the global-CM case, only one CM represents the whole image, while in the local-CM case, there is a CM for each pixel. The y_{GT} refers to the Ground Truth (GT) mask, which was utilized in this procedure, while $y^{(m)}$ represents the mask annotated by the m^{th} annotator. It's worth noting that the GT is not directly observable in this context. Thus, the model undergoes a disentangling process to deduce its version of the GT. As a result, directly computing the Confusion Matrix (CM) isn't a straightforward process.

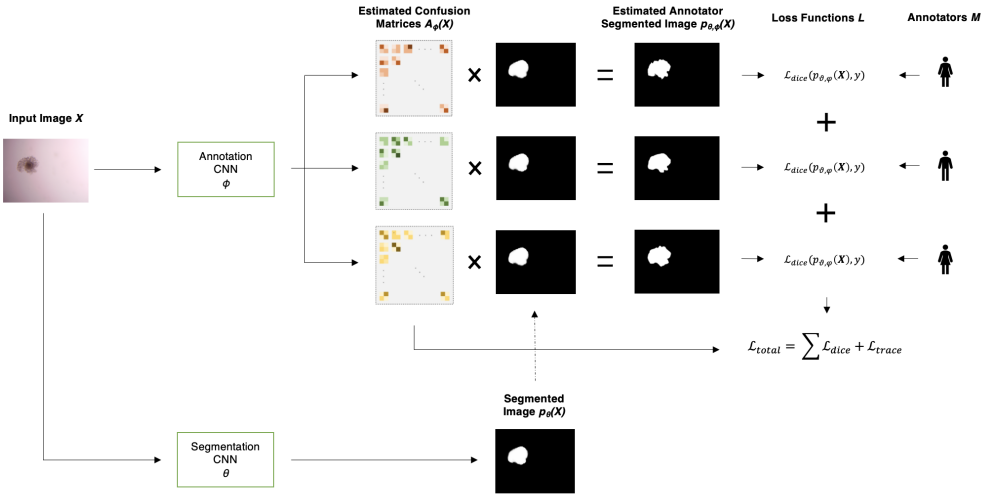


Figure 5.3: The architecture consists of two components: (a) a segmentation network, characterized by the parameter θ, which produces a probability distribution p_θ for segmentation; and (b) a regularization module consisting of a CNN, parameterized by ϕ, which utilizes the input image to generate three pixel-wise confusion matrices A_ϕ at the local (pixel) level. During the training process, the parameters (θ, ϕ) are learned simultaneously by optimizing the overall loss function.

Thus, a regularization network is employed to simulate the confusion matrices. Each of these matrices emerges as an output from a Convolutional Neural Network (CNN) that uses the COC image, denoted as $\mathbf{x}$ as its input. This CNN, defined by $\phi(m)$, maintains a consistent architecture across all CM branches but is independently fine-tuned after a random initialization. The segmentation network supplies the probability distribution prediction, denoted as $p_\phi(\mathbf{x})$, which is evident from the diagrams in Figure 5.3 and Figure 5.4. We obtain the anticipated probability distributions that align with each annotator's patterns by carrying out an element-wise multiplication between the CMs and this projected probability distribution.

$$p^{(m)}_{\theta,\phi}(\mathbf{x}) = A^{(m)}_{\phi}(\mathbf{x}) \cdot p_\theta(\mathbf{x}) \tag{5.2}$$

where $A^{(m)}_{\phi}$ is the total of the CMs, with $m \in \{1, 2, 3\}$ and '$\cdot$' operator denotes the element-wise matrix multiplication in (w, h), where needed.

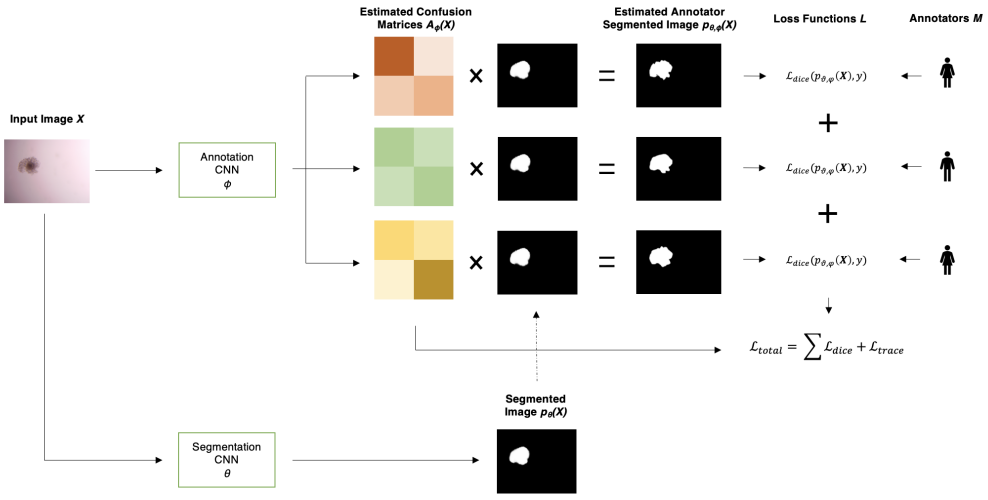


Figure 5.4: The architecture consists of two components: (a) a segmentation network, characterized by the parameter θ, which produces a probability distribution p_θ for segmentation, and (b) a regularization module consisting of a CNN, parameterized by ϕ, which utilizes the input image to generate three image-wise confusion matrices A_ϕ at the global (image) level. During the training process, the parameters (θ, ϕ) are learned simultaneously by optimizing the overall loss function.

The formulation of the optimal confusion matrix and its associated predicted probability distribution is the outcome of a combined optimization procedure encompassing both the segmentation and regularization networks.

5.3.3 Loss function and evaluation metrics

In the following segment, we delve into a comprehensive overview of the method we employ to jointly optimize the parameters of the segmentation network, represented by θ, and those of the annotator network, symbolized by ϕ. Additionally, we will elucidate the metrics used for evaluation in our research.

To assess the performance of the proposed segmentation, the Dice Coefficient [46] is selected. The Dice Coefficient quantifies the spatial concurrence between two regions, providing a range from 0 to 1, where 0 implies no overlap, while 1 indicates complete agreement. The corresponding equation is given below (5.3):

$$dice(f, \mathbf{x}, y) = \frac{2\sum_{ij} f(\mathbf{x})_{ij} y_{ij}}{\sum_{ij} f(\mathbf{x})_{ij} + \sum_{ij} y_{ij}} \tag{5.3}$$

with y being the ground truth, $\mathbf{x}$ being the input image, and $f(\mathbf{x})$ the prediction of the model.

To guarantee that the predicted probability distribution, which encapsulates the tendencies of the annotators, aligns accurately with the real annotations, we make use of the dice loss. The dice loss, when applied to binary segmentation, is articulated subsequently:

$$\mathcal{L}_{dice}(f, \mathbf{x}, y) = 1 - dice(f, \mathbf{x}, y) \tag{5.4}$$

Subsequently, we incorporate a regularization term derived from the trace norm theorem as outlined in [207]. According to this theorem, when presented with a prominent trace for the confusion matrix, the procedure denoted by 5.2 aspires to curtail this trace. This reduction aids in aligning the projected confusion matrices corresponding to individual annotators with their genuine counterparts. The fine-tuning of the trace of the CM is achieved using the following loss function:

$$\mathcal{L}_{trace}(f, \mathbf{x}, y) = \sum_{m=1}^{3} tr(A_{\phi}^{(m)}(\mathbf{x})) \tag{5.5}$$

with $tr(A)$ denoting the trace of A.

By minimizing the trace, we're essentially encouraging the maximum uncertainty in the predictions of the annotators. On the other hand, by minimizing the dice loss, we ensure that the predictions stay true and loyal to the annotations we have observed. The comprehensive loss function is then designed by merging these two components: the dice loss and the trace loss.

$$\mathcal{L}_{total}^{\theta,\phi} = \mathcal{L}_{dice}^{\theta} + \gamma \mathcal{L}_{trace}^{\phi} \tag{5.6}$$

where γ corresponds to a regularization term. To establish the optimal γ, several options are tested, followed by a quantitative comparison, from γ = 0.0 to γ = 1.0, with a step of 0.1, while some options above 1.0, to examine the case of assigning higher importance to the trace loss. The model learns the parameters θ and ϕ by minimizing the combined loss function.

5.3.4 Methodology

In our research, we structure our approach into three key stages.

Firstly, we examine the strategies documented in existing scholarly works that delve into training a supervised segmentation model using labels generated by multiple human specialists (Section 5.3.4.1). We specifically evaluate this in the backdrop of a constrained dataset of COC images paired with their binary masks annotated by three experts. The primary objective at this juncture is to extract the accurate segmentation deriving solely from this ensemble of imperfect labels.

However, due to poor results, we switch our attention towards regions marked by heightened uncertainty (Section 5.3.4.2). As we delve deeper, we sharpen our analysis by discerning unique annotating tendencies for each specialist.

In the conclusive phase (Section 5.3.4.3), we aim to perfect the ground truth. We highlight the regions with significant uncertainty, enhancing our segmentation model's accuracy and reliability.

The code for this part of the thesis is available on the following GitHub page: maximum-likelihood-gt.

5.3.4.1 Jointly-learning

In this research, we implement the suggested methodology [240] comprising a pair of Convolutional Neural Networks (CNNs). The unique function of these CNNs is to predict the confusion matrix for each expert, which is subsequently used for segmenting the area of the COC, eliminating the need for prior GT.

Coupled CNN training The initial approach was implementing a local-CM model, where each image pixel was assigned its individual CM. Given the high dimensionality involved with assigning a CM to each pixel – effectively a $H \times W \times 2 \times 2$ dimensional problem, we decided to revise our strategy, with H being the height of the image, W the width of it, and a 2×2 confusion matrix for each pixel. (Figure 5.3)

To manage the high dimensionality, we adopted a different approach, reducing the multiple CMs to a single one (global-CM) to capture the behavior across the entire image. This effectively condensed the problem from a $H \times W \times 2 \times 2$ dimension to a more manageable 2×2 CM. (Figure 5.4)

Coupled CNN training with transfer learning To assist the initial models, examining further options was necessitated. We adopted the method proposed by Athanasiou et al. [7] and trained a segmenting CNN model to segment the COC area proficiently. Subsequently, the model weights were saved to serve as a starting point for disentangling the process from the ground truth, negating the need to train both CNNs simultaneously and offering a promising starting point for the training.

Upon revisiting the approach, two primary concepts stood out. The first entailed training with the pre-trained weights, allowing the models to optimize the weights for both CNNs. The second concept involved training with the pre-trained weights, freezing the segmentation CNN, and enabling the models to train the annotating CNN, thus learning the CMs.

5.3.4.2 Confusion Matrices on Uncertainty

To derive the confusion matrices on areas of uncertainty, a pre-trained deep learning model that has achieved a high dice score (Athanasiou et al. [7]) for segmenting the COC area is employed. This model is used to pinpoint areas of uncertainty by setting a threshold at 0.05. In this context, every pixel that falls between $[0.05 - 0.95]$ is considered uncertain. Conversely, any pixel outside this range is deemed a segmentation where experts agree.

The pixels of interest for each image are identified and directly compared with the GT. In this instance, the GT is defined as the majority-voted masks created by combining the masks provided by experts. Confusion matrices are constructed by comparing the expert annotations with the majority-vote annotation solely on uncertain pixels.

Specifically, for each pixel in the uncertainty range, if an expert annotates it as '1' and the GT indicates '0', it is recorded as a False Positive in the confusion matrix. If the expert marks it as '0' and the GT indicates '1', it is a False Negative. Conversely, if the expert's annotation aligns with the GT, it is recorded as a True Positive or True Negative. This process is applied to all pixels in the uncertain areas of an image.

Following this, we normalize across the rows of actual values to obtain a confusion matrix for each annotator per image. To attain a total confusion matrix for each annotator, we compute the mean of their confusion matrices across all images.

5.3.4.3 Maximum Likelihood Ground Truth

After obtaining estimates for the confusion matrices of each annotator, we can then deduce the ground truth using a maximum likelihood approach. Below, we illustrate how to amalgamate

the labels provided by every annotator and their respective confusion matrices to compute the Maximum Likelihood Ground Truth.

Consider a specific pixel for which we have annotations. Let M denote the ensemble of all annotators, and each individual annotator $m \in M$ assigns a label $y^{(m)}$ to this pixel. The confusion matrix for annotator m is given by $A^{(m)} = (a_{ij}^{(m)})$. Here, $a_{ij}^{(m)}$ signifies the probability that annotator m designates the pixel with label i when the genuine label for that pixel is j.

Let's also assume we have a prior distribution $\rho = (\rho_1, \ldots, \rho_L)$. This prior provides initial beliefs about the class distribution of the pixel. Specifically, ρ_j indicates the predisposition or initial likelihood that this pixel belongs to class j, even before any annotations have been considered.

With all the components above, our objective now is to determine:

$$p(y_{GT} = j | y^{(M)}, A^{(M)}, \rho) = \frac{\alpha_j \rho_j}{\sum_{l \in L} \alpha_l \rho_l}, \tag{5.7}$$

where $\alpha_j = \prod_{m \in M} a_{y^{(m)} j}^{(m)}$.

This represents the postulated probability that the pixel belongs to class j, given all the information we have at hand: the annotations from all annotators ($y^{(M)}$) and their corresponding confusion matrices ($A^{(M)}$).

For a straightforward application, we suggest defining ρ_j as the frequency of label j in the set of annotations.

5.4 Results

In the upcoming section, we explore the outcomes derived from implementing the methodology detailed earlier. This includes an examination of the performance of the coupled CNN framework, which is grounded in the existing literature and further augmented by our enhancements. Additionally, we delve into the identification and understanding of the unique annotation patterns demonstrated by each expert. Finally, we reach the progress achieved in establishing a more refined ground truth, with a particular focus on the areas marked by uncertainty. Through this analysis, our objective is to highlight the proficiency of our comprehensive approach in tackling the intricacies of segmentation tasks.

5.4.1 Performance of Coupled CNNs

The performance of the coupled CNN structure failed to match the anticipated outcomes based on its earlier applications to synthetic data. This trend persisted across all cases, whether using global or local CMs and even when applying transfer learning. There were several reasons for this gap between the expected potential and the final results.

We identified two main challenges during our research. The first challenge was the inherent limitation in the data size due to the nature of real-world medical problems, such as the COC study we conducted. Our dataset consisted of a maximum of only 200 COC images. The second challenge was the high accuracy exhibited by medical experts in identifying COC regions. Such precision

led to a strong agreement among them, making it difficult to segment and recognize different expert behaviors.

Our attempts to train the coupled CNN model for COC area segmentation were not successful using either the local or global CM strategies. The model started deviating after just a few training rounds in both instances. We made numerous adjustments, including tuning various hyperparameters and experimenting with different γ values to control the trace loss better. However, the model's performance still lagged, highlighting the difficulties in adapting it to this particular task.

A previous research by Athanasiou et al. [7] suggested transfer learning as a crucial strategy for effective COC area segmentation. Even with a promising start, our initial approach couldn't maintain a high performance for either CM strategy. The subsequent approach aimed at locking the segmenting CNN's performance and shifted attention to teaching the CMs on the annotating CNN.

Interestingly, although the model learned three distinct CMs, their values were remarkably similar across different experts. Both the True Positive (TP) and True Negative (TN) values hovered around 1.0, while the False Positive (FP) and False Negative (FN) values were close to 0.0. A closer inspection revealed this consistency was expected, given the high level of agreement among experts regarding the pixels in both the image background and the COC area.

Our results remained disappointing despite starting with a pre-trained network from Athanasiou et al. [7]. Even with its initial advantage, the network couldn't sustain a robust performance under either the local or global CM setups. In a subsequent strategy, we decided to 'freeze' the segmentation CNN to preserve its optimal performance, focusing instead on training the CMs in the annotation CNN. While this approach did differentiate among three CMs, their values were surprisingly consistent across all experts. The True Positive (TP) and True Negative (TN) values were close to 1.0, and the False Positive (FP) and False Negative (FN) values hovered around 0.0.

This outcome was anticipated, given the considerable agreement among experts on most pixels. Ultimately, we recognized that the intended approach requires further refinement to realize its creators' vision, prompting us to reconsider our emphasis on the CMs.

5.4.2 Performance on CMs—Learning

In our current research phase, our main objective is to understand the distinct CMs of each expert within areas of high uncertainty, as shown in Figure 5.5. Given the broad consensus among experts, the only practical way to distinguish their individual annotation styles is within these disputed zones. To start, we identify these areas of high uncertainty using our deep learning model, which pinpoints zones with significant ambiguity in pixel classification, as detailed in Section 5.3.4. We then contrast each pixel in an annotator's mask with the majority-voted ground truth. After iterating this comparison five times, we generate an average confusion matrix for each annotator based on the entire dataset. The resulting matrices, which highlight the unique characteristics of each expert, are depicted in Figure 5.6.

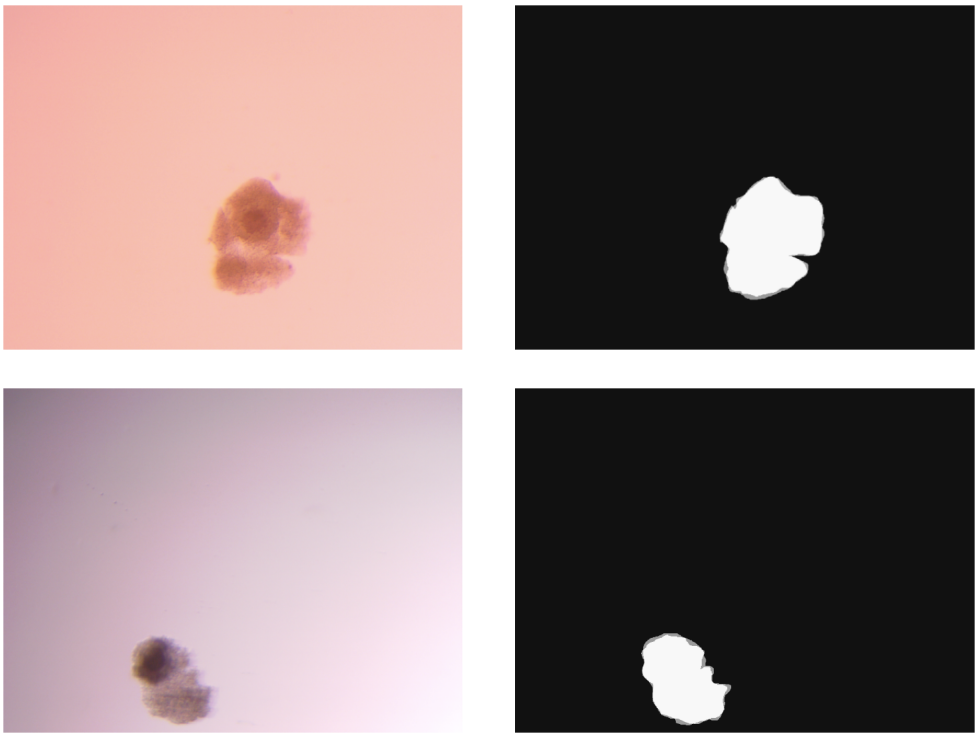


Figure 5.5: Visualization of uncertainty regions in the segmentation process: On the left-hand side, a sample of the original microscopy images of COC is shown. On the right-hand side, the uncertainty regions corresponding to the sample on the left-hand side are displayed. As is evident, areas of high uncertainty are concentrated along the borders of the cumulus oocyte complexes.

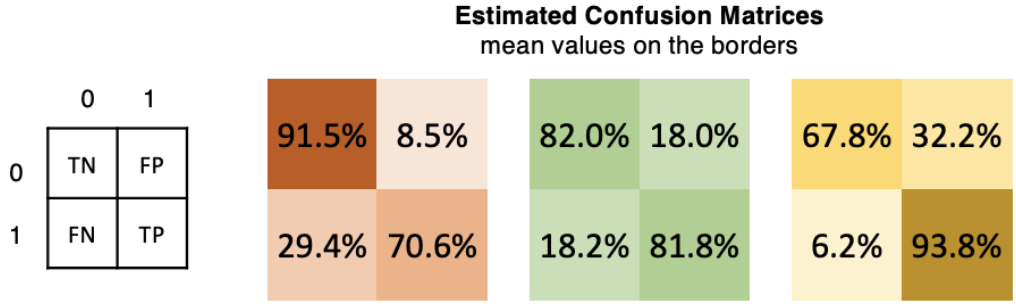


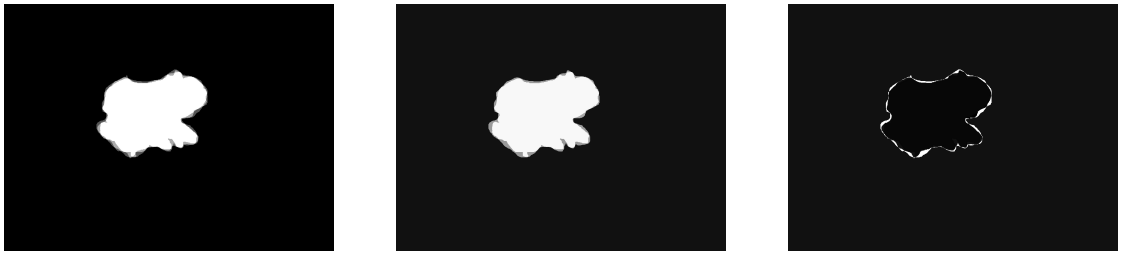
Figure 5.6: Visualisation of the confusion matrices for each annotator, focusing on the areas of high uncertainty. This representation shows a clear behavior of each expert on the most difficult areas to identify.

This information shows the experts' approaches to the most ambiguous areas, where disagreements are most probable, are evident. The first expert stands out for a high True Negative score but only an average True Positive score. This suggests a preference for classifying pixels as background and a stricter standard for deeming a pixel part of the COC. In contrast, the third expert tends to be more liberal in classifying pixels as belonging to the COC, even if it means mistakenly including background pixels within the COC region. This tendency is reflected in his high True

Positive and average True Negative rates. Meanwhile, the second expert strikes a middle ground between these two, showcasing a more balanced style. He exhibits good accuracy in correctly recognizing COC pixels and distinguishing background pixels.

5.4.3 Ground Truth

As detailed in Section 5.3.4.3, we propose a refined method for determining the ground truth, leveraging the maximum likelihood approach and incorporating individual expert annotations. An example of this advanced ground truth computation is illustrated in Figure 5.7.



(a) The average mask and areas of disagreement.

(b) The maximum likelihood mask.

(c) The area of disagreement between (**a**,**b**).

Figure 5.7: A comparison between the majority-vote ground truth and the maximum likelihood ground truth, which focuses on the areas of uncertainty. In (**a**), there is the majority-vote mask, with a gray zone on the borders, for the pixels of disagreement. In (**b**), there is the maximum likelihood mask, which can vary within the range of (0.0–1.0) since it is calculated using the confusion matrix identity of each expert. In (**c**), there is the zone of disagreement and alterations between case (**a**) and case (**b**).

Initially, the method relies on a majority vote strategy (as seen in Figure 5.7a). When three experts participate, an uncertainty level of 0.67 is reached if two experts concur while one dissent. However, the second method (shown in Figure 5.7b) capitalizes on the confusion matrices obtained from the annotation processes. This provides insights into the unique behaviors of the annotators. Using this knowledge, the certainty level for individual pixels can surpass 0.67, ranging between 0.67 and 1. The end result is a more robust ground truth, capable of detecting potential errors that might go unnoticed in the majority vote technique. The differences between the two methods, specifically where they diverge, are showcased in Figure 5.7c.

The superiority of the maximum likelihood method over prior techniques is evident. Our approach uniquely addresses the challenges faced by contemporary techniques. While existing methods perform somehow well on synthetic data or annotations from non-experts, they fail when used on real-world medical annotation data from assisted reproduction.

Our technique focuses on the specific annotation behaviors of each expert, generating tailored profiles based on their annotations. We can pinpoint areas where experts display inconsistencies or uncertainties by extracting these profiles using confusion matrices. This is achieved by harnessing real-world annotation data that genuinely reflects expert disagreements.

Additionally, understanding each expert's unique annotation behavior is pivotal. It facilitates the reapplication of these behaviors for subsequent annotations, eliminating redundant experimenta-

tion. This approach not only saves time and resources but also enhances model training performance. Importantly, our method eliminates the need for an odd number of experts, a requirement in many current techniques, such as the majority-vote one.

A key advantage of our methodology is its easy integration into clinical and laboratory settings. This compatibility ensures the retention and utilization of individual annotation profiles, simplifying the generation of enhanced results without the complexities of compiling multiple annotations.

In summation, our pioneering approach revolutionizes ground truth computation. It adeptly navigates the pitfalls of existing methods, offering pronounced accuracy and operational efficacy in medical image analysis.

5.5 Conclusions

This study centered on investigating multi-annotator segmentation using a COC dataset. The research followed a three-step process: first, the application of a coupled CNN architecture to real medical data, as suggested in the existing literature; second, a specific focus on ambiguous regions to extract individual expert annotation profiles; and third, an effort to enhance the accuracy of the ground truth by emphasizing areas of high uncertainty.

A primary discovery of this investigation was the significant contrast between results obtained from artificial data, commonly used in prior studies, and results from real-life medical datasets. Strategies that had shown promise with artificial data proved ineffective when applied to actual medical data. This underscored real-world medical annotation data's complexities and unique attributes, emphasizing that solutions developed in idealized environments may not directly apply to practical scenarios. Attempts to boost model performance using various strategies, such as transfer learning and adjustments to CMs, did not yield significant improvements.

Another challenge encountered was the difficulty in achieving consensus among the experts. While high agreement among experts is generally desirable, it posed a challenge in improving the coupled CNN's segmentation model using individual CMs, as minimal insight could be gained in an environment with high agreement. This implies that models designed for scenarios with significant annotation discrepancies may struggle when discrepancies are minimal.

In response to these challenges, the research shifted its focus toward uncertain regions within the data. An opportunity arose to delve deeper into understanding the unique behaviors of each expert, particularly in ambiguous areas. This introspection proved fruitful, revealing distinct tendencies and preferences for each expert.

Armed with this newfound understanding of annotator behaviors, the research transitioned to redefine the ground truth (GT) from a more informed perspective. This was not an attempt to disconnect the GT from expert annotations but rather a sophisticated effort to better integrate expert preferences into GT construction. A novel GT was proposed and formulated based on maximum likelihood, specifically on areas of uncertainty identified through deep learning. This new approach surpassed the simplistic majority-vote strategy, providing a more nuanced reflection of expert knowledge in the resulting GT.

The proposed method offers several advantages, primarily addressing challenges in multi-labeling by focusing on uncertain regions. It also creates personalized annotation profiles for experts, improving and assessing their performance in future tasks. Moreover, these methods hold practical

value in real-world settings, saving time and resources while overcoming limitations associated with an odd number of experts for accurate ground truth determination.

In conclusion, this research highlighted the limitations of transferring methodologies developed from artificial data to real-world applications. However, it also showcased the value of focusing on uncertain areas and understanding individual expert behaviors. These insights led to a more sophisticated construction of ground truth, emphasizing the potential for future research in multi-annotator segmentation.

Chapter 6

Side Projects and Future Work

In this chapter, we delve deeper into additional projects undertaken during the course of this PhD thesis. Owing to our comprehensive engagement, these projects are currently in an ongoing phase, revealing preliminary research findings that indicate the necessity for additional work to achieve their intended outcomes. Among these, we explored the ratio between the oocyte area and follicle area, aiming to understand its impact on the healthy maturation of oocytes (Section 6.1). Additionally, we investigated the development of healthy mammalian embryos, focusing on identifying critical stages in their growth process, such as the formation of pronuclei before cleavage (Section 6.2). Lastly, this chapter outlines prospective research directions at the intersection of deep learning and assisted reproduction (Section 6.3). This forward-looking discussion ties in with the work presented in the previous chapter and extends beyond, paving the way for future innovations in the field.

6.1 Follicle maturation quality

Ovarian tissue cryopreservation (OTC) followed by transplantation (OTT) has become the established method for preserving fertility in young girls and women facing the risk of premature ovarian insufficiency. While OTT has demonstrated its efficacy and safety, it remains unsuitable for certain patients due to the potential for the original disease to reoccur after transplantation. This limitation applies to specific cancer types such as leukemia, neuroblastoma, and ovarian cancer. In-vitro folliculogenesis emerges as a promising alternative to OTT for fertility preservation, particularly in cases where there is a risk of cancer relapse. Culturing ovarian follicles is a valuable model for exploring the fundamental aspects of oogenesis and folliculogenesis, both in healthy and pathological contexts.

In this section, we embark on a study to assess the following research question:

- **Q7:** How does the ratio between oocyte and follicle area affect the development of healthy oocytes?

6.1.1 Resources

The code for this work can be found in the following Github page: follicle_oocyte_prediction.

Figure 6.1: A representative image of a human ovarian follicle, meticulously captured by specialists at LRB using a Leica MZ12.5 stereomicroscope.

The dataset for this research project was collected and annotated by our skilled colleagues at RigsHospitalet in Copenhagen, specifically in the Laboratory of Reproductive Biology[1] (LRB).

In Figure 6.1, an example of a human ovarian follicle, as it was captured, is presented, while in Figure 6.2, a human ovarian follicle with its corresponding masks, one for the follicle area and one for the oocyte area, is shown. The dataset comprises 75 detailed images of human ovarian follicles, each accompanied by their respective follicle and oocyte masks. Follicles were carefully extracted from the ovaries using a process of mechanical isolation. This procedure was conducted under the precision optics of a Leica MZ12.5 stereomicroscope. 23-gauge needles, affixed to 1 mL syringes, were utilized for the extraction. The isolation process took place in a specially prepared, pre-warmed medium. This medium comprised McCoy's 5α, enhanced with 25 mM HEPES (Gibco), and further supplemented with 5% FBS (Fetal Bovine Serum, Gibco) and 1% penicillin/streptomycin at a 100$\times$ concentration. Additionally, a comprehensive track record indicates which ovarian follicles evolved into mature oocytes.

6.1.2 Methods

The primary objective of this study was to investigate the influence of the ratio between the oocyte and follicle areas on the development of healthy oocytes, as illustrated in Figure 6.3.

The initial idea for the pipeline to succeed in this task is shown in Figure 6.4. To accomplish this, we designed an advanced deep learning method to accurately segment the areas of oocytes and follicles in the ovarian follicle image dataset. This involved employing a pair of sophisticated deep learning Convolutional Neural Networks for segmentation, enhanced by a series of refined pre-processing and post-processing techniques to optimize the results. Furthermore, we incorporated the segmented area data and additional relevant information pertaining to these follicles to

[1] https://www.rigshospitalet.dk/english/departments/juliane-marie-centre/fertility-department/laboratory-of-reproductive-biology/Sider/default.aspx

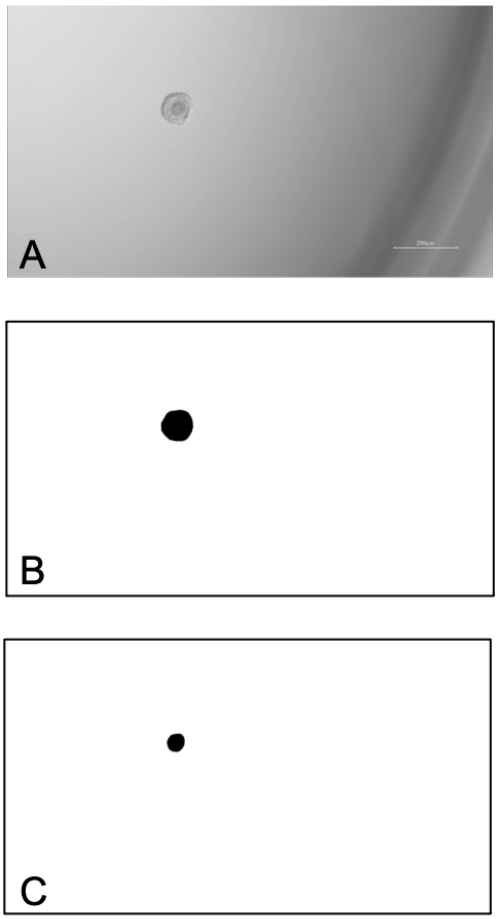


Figure 6.2: Displayed is an image of a human ovarian follicle (A) accompanied by its respective mask delineating the follicle area (B), and another mask highlighting the oocyte area (C).

develop a predictive model. This model aims to identify the most promising ovarian follicles for maturation, thereby providing experts with an automated and reliable tool for follicle selection in the context of human treatment in assisted reproductive technologies.

We implemented a Region Of Interest (ROI) extraction algorithm before the training phase. This was crucial for eliminating extraneous areas surrounding the follicle, thereby preventing the inclusion of artifacts present in the original environment, such as the Petri dish borders, or those introduced during the image acquisition process, like water droplets. For the segmentation process, we experimented with UNet architectures of varying specifications. Initially, the model was trained on a domain-similar task with a more extensive dataset[2]. Subsequent training was conducted on the specific dataset of interest. Of the total dataset, 65 images were utilized for the final model training, while 10 were reserved for validation and testing purposes. We opted for the Dice loss function, conducting the training over 200 epochs with a batch size of 8.

[2]https://zenodo.org/records/7152850

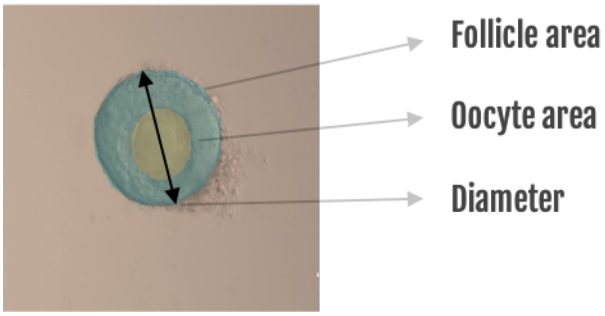


Figure 6.3: This image illustrates an oocyte situated within a follicle, marked with annotations essential for our study, specifically (i) the follicle's surface area and (ii) the oocyte's surface area.

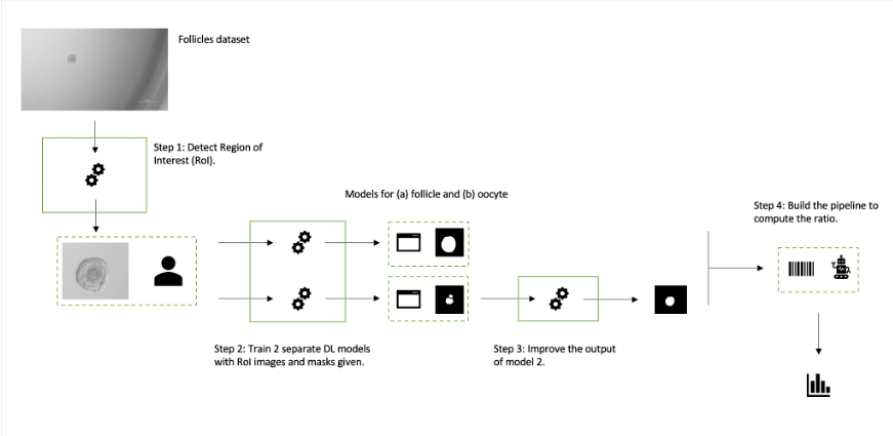


Figure 6.4: The pipeline proposed to solve this research question: First, the ROI algorithm is applied to determine the area of interest. Then, there are two segmentation models to segment the follicle and oocyte areas in different steps. Due to the issues faced in the oocyte segmentation, additional advancements are required to reach a more robust result. Last, we automatically compute the ratio for different cases and study its impact on healthy oocyte development.

6.1.3 Preliminary results

The follicle area segmentation model achieved a satisfying performance, registering a 92% Dice score. However, the model for segmenting the oocyte area faced challenges, peaking at an 88% Dice score, but exhibiting sensitivity to new data inputs, while it faced overfitting. Despite these advancements, the limited dataset size hindered our ability to validate the results and derive statistically significant conclusions. Some sample results of the second phase, which was the phase that had the most difficulties, are presented in Figure 6.5. Nonetheless, these preliminary findings lay a solid foundation for future research. This groundwork paves the way for further exploration into the automated analysis of the relationship between oocyte and follicle area ratios and the development of healthy oocytes using deep learning methodologies.

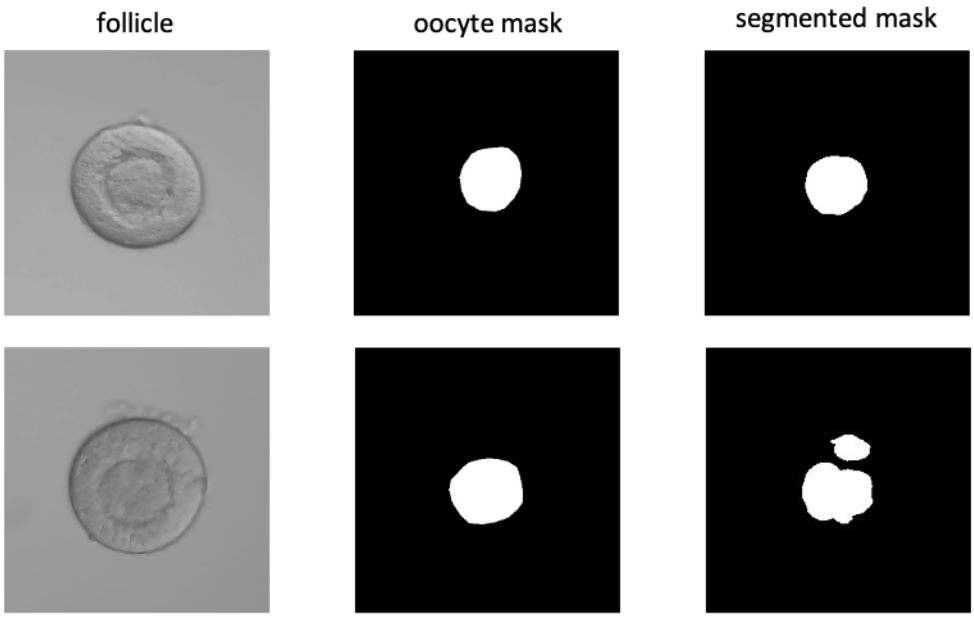


Figure 6.5: The following set of images showcases results from the secondary model dedicated to segmenting the oocyte area. The initial column displays the ROI area of the ovarian follicle, the middle column presents the ground truth masks, and the final column contains the areas segmented by our proposed model. Notably, there were numerous instances, such as in the second example, where the segmented area appeared flawed and evidently unrealistic, marked by discontinuities.

6.2 Embryo development

Mammalian cleavage has posed considerable challenges for scientific investigation. Mammalian eggs are among the smallest in the animal kingdom, making them exceptionally challenging to manipulate in experimental settings. For instance, the human zygote is a mere 100 μm in diameter, barely discernible to the naked eye, and less than one-thousandth the size of a Xenopus egg. Moreover, unlike sea urchin or frog zygotes, mammalian zygotes are not produced in large quantities, making it arduous to acquire sufficient material for biochemical studies. Typically, a female ovulates fewer than ten eggs at a time. Adding to the complexity, mammalian embryo development occurs within another organism rather than in an external environment. Only in recent times have we managed to replicate some of these internal conditions and observe mammalian embryo development in vitro.

When it comes to endangered species, the challenges are further amplified, as obtaining available data is intricate, and using it requires utmost sensitivity. In this section, we embark on an endeavor to explore deep learning methods for transferring knowledge across early-stage embryos of different mammalian species, encompassing humans, mice, and certain endangered species. The primary focus was the study of pronuclei formation, a crucial step before reaching the cleavage stage. The attempted research question to be assessed is structured as follows:

- **Q8:** Can deep learning methods facilitate the transfer of knowledge about pronuclei formation across early-stage embryos of different mammalian species?

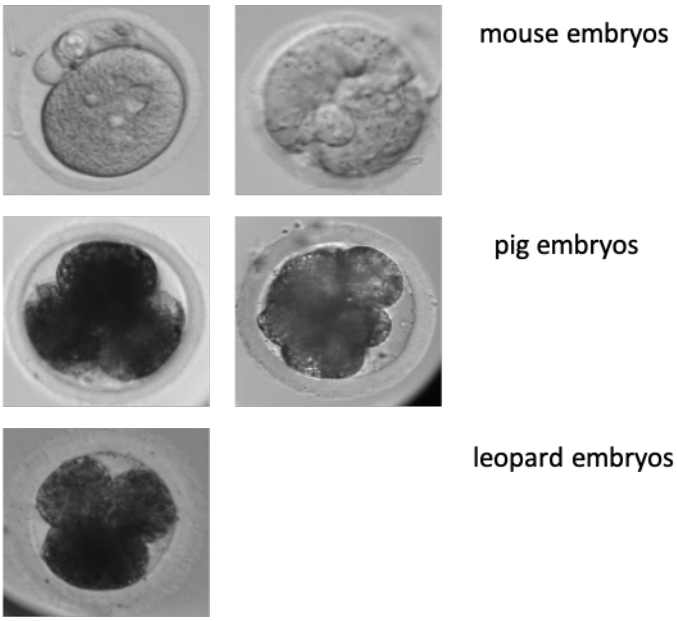


Figure 6.6: This collection showcases a diverse array of mammalian embryos, with mouse, pig, and leopard species, meticulously procured by our colleagues at the IZW in Berlin. Each sample in this array represents a unique embodiment of embryonic development across different mammalian species, highlighting the intricacies of biological diversity and their similarities as members of the mammalian family.

6.2.1 Resources

The research dataset utilized in this study was meticulously gathered by the proficient team at the Leibniz Institute for Zoo and Wildlife Research[3] (IZW) in Berlin. This dataset encompasses an extensive collection of early-stage embryonic development time-lapse videos from various mammalian species, predominantly featuring mouse embryos. A snapshot of some mammalian embryos in an early stage can be found in Figure 6.6. A significant challenge presented by this dataset is the scarcity of annotations and masks, of which we managed to acquire only a limited quantity.

During the experimental phase, post-intracytoplasmic sperm injection (ICSI), the prospective zygotes were placed into embryo culture dishes containing pre-equilibrated media overlaid with light mineral oil. The experimentation involved different culture media, notably (1) the commercially available EmbryoMax KSOM medium, enriched with amino acids, and (2) a Synthetic oviduct fluid (mSOF) formulation based on the work of Tervit et al. [102]. Detailed methodologies and conditions utilized for embryo procurement are documented in the corresponding publications from IZW laboratory [194]. After in-vitro fertilization, the embryos underwent development in an incubator (Geri[4], EmbryoScope[5]), facilitating the acquisition of time-lapse data.

The core objective of this project was to evaluate three primary stages in mammalian embryonic development: (i) pronuclei formation, (ii) progression to the cleavage stage, and (iii) quantifica-

[3]https://www.izw-berlin.de/en/home.html
[4]https://www.hamiltonthorne.com/product/geri-incubator/
[5]https://www.vitrolife.com/why-vitrolife/the-patient-ivf-journey/embryoscope-time-lapse-system/

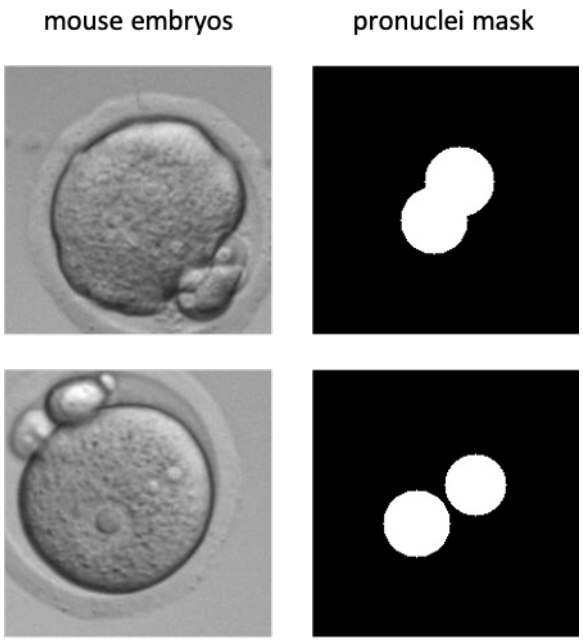


Figure 6.7: Featured in this set are early-stage mouse embryos (first column), each accompanied by their respective masks that accurately highlight the formation of two pronuclei at this developmental phase (second column).

tion of fragmentation rates. However, the study's focus was narrowed to solely pronuclei formation due to constraints in the available data. This shift in focus brought forth two additional challenges. Initially, the dataset suffered from a severe dearth of annotated data (less than 10 samples, an example available in Figure 6.7). Furthermore, a significant issue with the majority of mammalian embryos, barring human and mouse specimens, is their opaque nature, rendering internal structures like pronuclei undetectable in embryos from species such as pigs or leopards, as it is also visible in Figure 6.6.

6.2.2 Preliminary results

In an attempt to address these challenges, a deep learning model was developed to identify pronuclei in the available mouse embryos. Unfortunately, the model demonstrated unsatisfactory performance, marred by significant overfitting. Consequently, the project was paused, pending future continuation when field experts can provide adequate annotations. Despite the current halt, the project remains a compelling challenge with the potential to revolutionize the automated assessment of mammalian embryonic development. This holds especially true for endangered species, where data availability is critically limited, presenting practical and conservation-oriented applications.

6.3 Future directions

6.3.1 Deep learning application in Cumulus Oocyte Complexes

The research presented in Chapter 4 offers several promising avenues for future exploration:

- *Expansion of research to Other COC characteristics:* The application of deep learning to study COCs can be expanded to include a more comprehensive analysis of oocyte characteristics. This could involve examining the position of the oocyte and investigating the transformations occurring within shorter time frames. Such an expansion would not only enhance our understanding of oocyte biology but also contribute to the development of predictive models capable of identifying oocytes with the highest potential for developing into healthy embryos.
- *Adaptation to other mammalian oocytes and microscopic variabilities:* The deep learning models utilized in this research should be tested and adapted for applicability to other mammalian oocytes. This adaptation is crucial, as it was not covered in the current study. Additionally, these models should be versatile enough to accommodate variations in microscopy equipment and settings, ensuring broader applicability and relevance across different research and clinical environments.
- *Correlation between embryo development stages and COC development:* There is a potential link between the different stages of embryo development and the development of the COC and oocyte. Initial indications in this research suggest that the expansion of the COC plays a more crucial role in the stages following cleavage in embryo development, implying a shift in influence from the sperm to the COC at this stage. Future research should focus on elucidating these relationships in greater detail, providing invaluable insights into embryonic development processes.
- *Commercialization potential:* Given the considerable acceptance and interest this tool has garnered, its commercial potential is worth exploring. Future work could involve assessing the viability of introducing this deep learning tool into the market, especially considering its applications in reproductive technology and research.

6.3.2 Deep learning in consensus segmentation

The proposed deep learning method for consensus segmentation, discussed in Chapter 5, also presents exciting opportunities for further research:

- *Application of expert annotation profiles in practical settings:* Future research should explore the implementation of expert annotating profiles in laboratory and clinical settings. This involves studying the effectiveness of applying these profiles to automatically enhance the ground truth in segmentation tasks. Investigating the real-world application of these enhanced profiles could revolutionize current practices in both research labs and clinical environments.

- *Potential for standardized commercial application:* The feasibility of integrating this consensus segmentation tool into standard clinical procedures is another promising area for future research. There is a need to evaluate its efficiency and effectiveness in scenarios with limited expert annotations. Such a study could pave the way for its standardized use in clinical settings, potentially transforming the landscape of clinical diagnostics and treatment planning.

6.3.3 Deep learning in tackling general problems in ART field

The opportunities coming with deep learning can tackle side problems appeared also in assisted reproduction:

- *The use of Generative Adversarial Networks (GANs) in assisted reproduction:* GANs have shown significant promise in various fields, including assisted reproduction [48, 92]. GANs could revolutionize this area by generating synthetic but realistic images or data patterns. This capability is particularly beneficial in scenarios where data is scarce or privacy concerns limit data availability. In assisted reproduction, GANs could be used to generate high-quality, realistic images of embryos or oocytes, aiding in the research and development of reproductive technologies. By enhancing the quality and quantity of available data, GANs can potentially improve predictive models and decision-making processes in assisted reproduction, leading to better outcomes in fertility treatments. This becomes stronger, considering the limitations we faced in the projects presented previously in this chapter, but also in some cases in Chapters 4 and 5.
- *The use of Active Learning in assisted reproduction:* Active learning, a subset of machine learning, involves algorithms that selectively identify the most informative data points for training. Budd et al. [27] has shown a massive use of active learning in medical image analysis in his survey presented some years ago. In assisted reproduction, active learning can significantly enhance the efficiency and accuracy of machine learning models. By focusing on the most informative and relevant data, these models can become more efficient in predicting outcomes such as embryo viability, optimal fertilization techniques, or success rates of various reproductive procedures. This approach could lead to more personalized and effective treatments in assisted reproduction, improving success rates and patient outcomes. One of the initial ideas was to dive into the active learning opportunities in ART, however, due to other opportunities, we decided to leave that for future exploration.

Chapter 7

Concluding Remarks

In this thesis, a comprehensive exploration of reproductive biology and medical imaging was undertaken, leading to significant advancements in these fields.

Developing and integrating a deep learning pipeline for cumulus-oocyte complexes marked a key achievement, offering precision in segmenting COC areas that rival or even surpass human expertise. This allowed for a detailed investigation into COC expansion and its role in healthy embryo development, including studying cell density around the oocyte. This tool exemplifies the effectiveness of deep learning in biological research.

Furthermore, the thesis addressed the challenge of expert bias in medical image segmentation. A novel consensus segmentation method was introduced, focusing on areas of disagreement among experts, thus advancing toward more accurate and unbiased medical image interpretations. Finally, this approach can create the experts' annotating profiles with many more real-world applications.

In addition to the major contributions, the thesis also explored two other important cases. Firstly, the ratio between the oocyte's area and the follicle's area was explored, providing insights into how this ratio could affect the healthy development of the oocyte. Secondly, the thesis studied early-stage embryos of different mammalian species. This approach could potentially bridge the knowledge gap and foster a more comprehensive understanding of endangered mammalian species. These minor attempts further enrich the thesis, adding breadth to its scope and depth to its scientific contributions.

Wrapping it up, in this thesis, we have tackled several complex questions in the realm of embryology and deep learning. Here's a summary of our findings:

- **Q1:** What is the optimal method for computing the area of COCs?
  We established that the area method is optimal in terms of objectivity, time, and resources (see Chapter 4).
- **Q2:** Can we develop a deep learning framework that achieves human-level performance using the optimal method for calculating COCs' area?
  A UNet-based deep learning model was proposed and successfully implemented to segment COC area with human-level performance (see Chapter 4).
- **Q3:** Is it feasible to identify the expansion of COCs from an immature to a mature stage using the optimal method?

We designed a deep learning framework to compute areas in both immature and mature COCs, enabling us to identify COC expansion, which is vital for healthy oocyte and embryo development (see Chapter 4).

- **Q4:** How does the optimal method for measuring COCs' expansion impact healthy embryo development in mammalian species?
  Through the Mann-Whitney U test, it was demonstrated that COC expansion significantly influences healthy embryo development in mammalian species (see Chapter 4).
- **Q5:** What influence does the density of COCs, as determined by the optimal method, have on healthy embryo development in mammalian species?
  Our research indicated that the density of COCs, estimated by the brightness of the segmented areas, is another crucial factor in healthy embryo development (see Chapter 4).
- **Q6:** How can expert biases in medical image segmentation be effectively overcome?
  We proposed a sophisticated method for computing the ground truth in medical image masks, focusing on uncertain areas among experts, which is a step toward fully disentangling biases in medical image annotations (see Chapter 5).
- **Q7:** How does the ratio between oocyte and follicle area affect the development of healthy oocytes?
  This project is ongoing, and conclusive results are yet to be presented (see Chapter 6).
- **Q8:** Can deep learning methods facilitate the transfer of knowledge about pronuclei formation across early-stage embryos of different mammalian species?
  This project is ongoing, and conclusive results are yet to be presented (see Chapter 6).

Future research directions also include expanding deep learning applications to other mammalian oocytes, exploring the commercial viability of these tools, and investigating active learning methods. Last, the potential of GANs in assisted reproduction is also highlighted, indicating new possibilities in the field.

Overall, the thesis marks a convergence of technology and biology, where deep learning becomes crucial in understanding embryo development and improving medical image analysis. The findings and methodologies established provide a solid base for future research and practical applications in reproductive biology.

Bibliography

[1] Patrick Allard, Qin Yang, William F Marzluff, and Hugh J Clarke. “The stem-loop binding protein regulates translation of histone mRNA during mammalian oogenesis”. In: *Developmental biology* 286.1 (2005), pp. 195–206 (cit. on p. 18).

[2] Everett Anderson and David F Albertini. “Gap junctions between the oocyte and companion follicle cells in the mammalian ovary.” In: *The Journal of cell biology* 71.2 (1976), pp. 680–686 (cit. on p. 12).

[3] Irvan B Arief-Ang, Flora D Salim, and Margaret Hamilton. “Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co2 sensor data”. In: *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. 2017, pp. 1–10 (cit. on p. 39).

[4] Daniel R Arnold, Patricia Françon, James Zhang, Kyle Martin, and Hugh J Clarke. “Stem–loop binding protein expressed in growing oocytes is required for accumulation of mRNAs encoding histones H3 and H4 and for early embryonic development in the mouse”. In: *Developmental biology* 313.1 (2008), pp. 347–358 (cit. on p. 18).

[5] Andrew J. Asman and Bennett A. Landman. “Formulating spatially varying performance in the statistical fusion framework”. In: *IEEE transactions on medical imaging* 31.6 (2012). Publisher: IEEE, pp. 1326–1336 (cit. on p. 67).

[6] Andrew J. Asman and Bennett A. Landman. “Non-local statistical label fusion for multi-atlas segmentation”. en. In: *Medical Image Analysis* 17.2 (2013), pp. 194–208. DOI: 10.1016/j.media.2012.10.002 (cit. on p. 67).

[7] Georgios Athanasiou, Jesus Cerquides, and Josep-Lluis Arcos. “Detecting the Area of Bovine Cumulus Oocyte Complexes Using Deep Learning and Semantic Segmentation”. In: CCIA 2022 : 24th International Conference of the Catalan Association for Artificial Intelligence, 2022, pp. 249–258. DOI: https://doi.org/10.3233/FAIA220346 (cit. on pp. 41, 70, 74, 76).

[8] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. *Medical Image Segmentation Review: The success of U-Net*. 2022 (cit. on pp. 37, 53).

[9] Nima Azari-Dolatabad, Annelies Raes, Krishna Chaitanya Pavani, Anise Asaadi, Daniel Angel-Velez, Petra Van Damme, Jo L. M. R. Leroy, Ann Van Soom, and Osvaldo Bogado Pascottini. “Follicular fluid during individual oocyte maturation enhances cumulus expansion and improves embryo development and quality in a dose-specific manner”. en. In: *Theriogenology* 166 (2021), pp. 38–45. DOI: 10.1016/j.theriogenology.2021.02.016 (cit. on p. 44).

[10] R Bachvarova, V De Leon, and I Spiegelman. “Mouse egg ribosomes: evidence for storage in lattices”. In: *Development* 62.1 (1981), pp. 153–164 (cit. on p. 13).

[11] Rosemary Bachvarova, Victor De Leon, Andrew Johnson, Gail Kaplan, and Barbara V Paynton. “Changes in total RNA, polyadenylated RNA, and actin mRNA during meiotic maturation of mouse oocytes”. In: *Developmental biology* 108.2 (1985), pp. 325–331 (cit. on p. 16).

[12] TG Baker. “A quantitative and cytological study of germ cells in human ovaries”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 158.972 (1963), pp. 417–433 (cit. on p. 10).

[13] Basak Balaban, Bulent Urman, Aycan Sertac, Cengiz Alatas, Senai Aksoy, and Ramazan Mercan. “Oocyte morphology does not affect fertilization rate, embryo quality and implantation rate after intracytoplasmic sperm injection.” In: *Human reproduction (Oxford, England)* 13.12 (1998), pp. 3431–3433 (cit. on p. 15).

[14] Bikramjit Banerjee and Peter Stone. “General Game Learning Using Knowledge Transfer.” In: *IJCAI*. Citeseer. 2007, pp. 672–677 (cit. on p. 39).

[15] Jason A Barritt, Carol A Brenner, Henry E Malter, and Jacques Cohen. “Mitochondria in human offspring derived from ooplasmic transplantation: Brief communication”. In: *Human Reproduction* 16.3 (2001), pp. 513–516 (cit. on p. 23).

[16] Jonathan Baxter. “Theoretical models of learning to learn”. In: *Learning to learn*. Springer, 1998, pp. 71–94 (cit. on p. 38).

[17] Michael J Berridge. “Inositol trisphosphate and calcium signalling”. In: *Nature* 361.6410 (1993), pp. 315–325 (cit. on p. 22).

[18] Steffen Bickel. “ECML-PKDD discovery challenge 2006 overview”. In: *ECML-PKDD Discovery Challenge Workshop*. 2006, pp. 1–9 (cit. on p. 39).

[19] Jordan J Bird, Jhonatan Kobylarz, Diego R Faria, Anikó Ekárt, and Eduardo P Ribeiro. “Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG”. In: *IEEE Access* 8 (2020), pp. 54789–54801 (cit. on p. 39).

[20] S Bozinovski. “Teaching space: A representation concept for adaptive pattern classification”. In: *COINS Technical Report No. 81-28* (1981) (cit. on p. 38).

[21] Stevo Bozinovski and Ante Fulgosi. “The influence of pattern similarity and transfer learning upon training of a base perceptron b2”. In: *Proceedings of Symposium Informatica*. Vol. 3. 1976, pp. 121–126 (cit. on p. 38).

[22] Peter Braude, Virginia Bolton, and Stephen Moore. “Human gene expression first occurs between the four-and eight-cell stages of preimplantation development”. In: *Nature* 332.6163 (1988), pp. 459–461 (cit. on p. 18).

[23] David Brawand, Walter Wahli, and Henrik Kaessmann. “Loss of egg yolk genes in mammals and the origin of lactation and placentation”. In: *PLoS biology* 6.3 (2008), e63 (cit. on p. 13).

[24] Stephane Brunet and Bernard Maro. “Cytoskeleton and cell cycle control during meiotic maturation of the mouse oocyte: integrating time and space”. In: *Reproduction* 130.6 (2005), pp. 801–811 (cit. on p. 20).

[25] Roberto Buccione, Barbara C Vanderhyden, Philip J Caron, and John J Eppig. "FSH-induced expansion of the mouse cumulus oophorus in vitro is dependent upon a specific factor (s) secreted by the oocyte". In: *Developmental biology* 138.1 (1990), pp. 16–25 (cit. on p. 11).

[26] William M Buckett, Ri-Cheng Chian, Hananel Holzer, Nicola Dean, Robert Usher, and Seang Lin Tan. "Obstetric outcomes and congenital abnormalities after in vitro maturation, in vitro fertilization, and intracytoplasmic sperm injection". In: *Obstetrics & Gynecology* 110.4 (2007), pp. 885–891 (cit. on p. 19).

[27] Samuel Budd, Emma C Robinson, and Bernhard Kainz. "A survey on active learning and human-in-the-loop deep learning for medical image analysis". In: *Medical Image Analysis* 71 (2021), p. 102062 (cit. on p. 89).

[28] Antonin Bukovsky, Michael R Caudle, Marta Svetlikova, Jay Wimalasena, Maria E Ayala, and Roberto Dominguez. "Oogenesis in adult mammals, including humans: a review". In: *Endocrine* 26 (2005), pp. 301–316 (cit. on p. 23).

[29] Kathleen H Burns, Maria M Viveiros, Yongsheng Ren, Pei Wang, Francesco J DeMayo, Donald E Frail, John J Eppig, and Martin M Matzuk. "Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos". In: *Science* 300.5619 (2003), pp. 633–636 (cit. on p. 22).

[30] Rita Canipari, Olga Epifano, Gregorio Siracusa, and Antonietta Salustri. "Mouse oocytes inhibit plasminogen activator production by ovarian cumulus and granulosa cells". In: *Developmental Biology* 167.1 (1995), pp. 371–378 (cit. on p. 11).

[31] Aaron Carass et al. "Longitudinal multiple sclerosis lesion segmentation: Resource and challenge". en. In: *NeuroImage* 148 (2017), pp. 77–102. DOI: 10.1016/j.neuroimage.2016.12.064 (cit. on p. 67).

[32] Inger B Carlsson, Jennifer E Scott, JA Visser, Olli Ritvos, APN Themmen, and Outi Hovatta. "Anti-Müllerian hormone inhibits initiation of growth of human primordial ovarian follicles in vitro". In: *Human reproduction* 21.9 (2006), pp. 2223–2227 (cit. on p. 10).

[33] Rich Caruana. "Multitask learning". In: *Machine learning* 28 (1997), pp. 41–75 (cit. on p. 38).

[34] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. "Return of the devil in the details: Delving deep into convolutional nets". In: *arXiv preprint arXiv:1405.3531* (2014) (cit. on p. 35).

[35] L. Chen, P. T. Russell, and W. J. Larsen. "Functional significance of cumulus expansion in the mouse: Roles for the preovulatory synthesis of hyaluronic acid within the cumulus mass". en. In: *Molecular Reproduction and Development* 34.1 (1993), pp. 87–93. DOI: 10.1002/mrd.1080340114 (cit. on p. 42).

[36] Ri-Cheng Chian, Jin-Ho Lim, and Seang-Lin Tan. "State of the art in in-vitro oocyte maturation". In: *Current Opinion in Obstetrics and Gynecology* 16.3 (2004), pp. 211–219 (cit. on p. 20).

[37] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer. 2016, pp. 424–432 (cit. on p. 37).

[38] Adam Coates and Andrew Y Ng. “Learning feature representations with k-means”. In: *Neural Networks: Tricks of the Trade: Second Edition*. Springer, 2012, pp. 561–580 (cit. on p. 25).

[39] Catherine MH Combelles, Mary Jo Carabatsos, T Rajendra Kumar, Martin M Matzuk, and David F Albertini. “Hormonal control of somatic cell oocyte interactions during ovarian follicle development”. In: *Molecular Reproduction and Development: Incorporating Gamete Research* 69.3 (2004), pp. 347–355 (cit. on p. 12).

[40] Pierre Comizzoli, Pascal Mermillod, and Robert Mauget. “Reproductive biotechnologies for endangered mammalian species”. In: *Reproduction Nutrition Development* 40.5 (2000), pp. 493–504 (cit. on p. 64).

[41] Olivier Commowick, Alireza Akhondi-Asl, and Simon K. Warfield. “Estimating a reference standard segmentation with spatially varying performance parameters: local MAP STAPLE”. eng. In: *IEEE transactions on medical imaging* 31.8 (2012), pp. 1593–1606. DOI: 10.1109/TMI.2012.2197406 (cit. on p. 67).

[42] SJ Conner, Linda Lefièvre, DC Hughes, and CLR Barratt. “Cracking the egg: increased complexity in the zona pellucida”. In: *Human Reproduction* 20.5 (2005), pp. 1148–1152 (cit. on p. 14).

[43] EH Davidson. “Gene activity during oogenesis”. In: *Gene activity in early development* (1986), pp. 305–407 (cit. on p. 12).

[44] Daniel G De Matos, Kathleen Miller, Richard Scott, Cam Anh Tran, David Kagan, Selva G Nataraja, Ann Clark, and Stephen Palmer. “Leukemia inhibitory factor induces cumulus expansion in immature human and mouse oocytes and improves mouse two-cell rate and delivery rates when it is present during mouse in vitro oocyte maturation”. In: *Fertility and sterility* 90.6 (2008), pp. 2367–2375 (cit. on p. 64).

[45] Jurrien Dean. *Oocyte-specific gene expression: roles in folliculogenesis, fertilization and early development*. Cambridge, United Kingdom: Cambridge University Press, 2003 (cit. on p. 14).

[46] Lee R. Dice. “Measures of the Amount of Ecologic Association Between Species”. en. In: *Ecology* 26.3 (1945), pp. 297–302. DOI: 10.2307/1932409 (cit. on pp. 33, 52, 54, 72).

[47] Darius Dirvanauskas, Rytis Maskeliunas, Vidas Raudonis, and Robertas Damasevicius. “Embryo development stage prediction algorithm for automated time lapse incubators”. en. In: *Computer Methods and Programs in Biomedicine* 177 (2019), pp. 161–174. DOI: 10.1016/j.cmpb.2019.05.027 (cit. on p. 43).

[48] Darius Dirvanauskas, Rytis Maskeliūnas, Vidas Raudonis, Robertas Damaševičius, and Rafal Scherer. “Hemigen: human embryo image generator based on generative adversarial networks”. In: *Sensors* 19.16 (2019), p. 3578 (cit. on p. 89).

[49] Chuong B Do and Andrew Y Ng. “Transfer learning for text classification”. In: *Advances in neural information processing systems* 18 (2005) (cit. on p. 39).

[50] Adam S Doherty, Mellissa RW Mann, Kimberly D Tremblay, Marisa S Bartolomei, and Richard M Schultz. “Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo”. In: *Biology of reproduction* 62.6 (2000), pp. 1526–1535 (cit. on p. 19).

[51] Jinwen Dong, David F Albertini, Katsuhiko Nishimori, T Rajendra Kumar, Naifang Lu, and Martin M Matzuk. “Growth differentiation factor-9 is required during early ovarian folliculogenesis”. In: *Nature* 383.6600 (1996), pp. 531–535 (cit. on p. 11).

[52] Stephen M Downs. “Specificity of epidermal growth factor action on maturation of the murine oocyte and cumulus oophorus in vitro”. In: *Biology of Reproduction* 41.2 (1989), pp. 371–379 (cit. on p. 45).

[53] John Duchi, Elad Hazan, and Yoram Singer. “Adaptive subgradient methods for online learning and stochastic optimization.” In: *Journal of machine learning research* 12.7 (2011) (cit. on p. 34).

[54] Tom Ducibella, Daniel Huneau, Elizabeth Angelichio, Zhe Xu, Richard M Schultz, Gregory S Kopf, Rafael Fissore, Stephane Madoux, and Jean-Pierre Ozil. “Egg-to-embryo transition is driven by differential responses to Ca2+ oscillation number”. In: *Developmental biology* 250.2 (2002), pp. 280–291 (cit. on p. 21).

[55] Rob A Dunne and Norm A Campbell. “On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function”. In: *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*. Vol. 181. Citeseer. 1997, p. 185 (cit. on p. 31).

[56] David Duque-Arias, Santiago Velasco-Forero, Jean-Emmanuel Deschaud, Francois Goulette, Andrés Serna, Etienne Decencière, and Beatriz Marcotegui. “On power Jaccard losses for semantic segmentation”. In: *VISAPP 2021: 16th International Conference on Computer Vision Theory and Applications*. 2021 (cit. on p. 34).

[57] Thomas Ebner, Marianne Moser, Michael Sommergruber, and Gernot Tews. “Selection based on morphological assessment of oocytes and embryos at different stages of preimplantation development: a review”. In: *Human reproduction update* 9.3 (2003), pp. 251–262 (cit. on p. 16).

[58] Mark A Edson, Ankur K Nagaraja, and Martin M Matzuk. “The mammalian ovary from genesis to revelation”. In: *Endocrine reviews* 30.6 (2009), pp. 624–712 (cit. on p. 12).

[59] Robert G Edwards, Barry D Bavister, and Patrick C Steptoe. “Early stages of fertilization in vitro of human oocytes matured in vitro”. In: *Nature* 221.5181 (1969), pp. 632–635 (cit. on p. 20).

[60] Dumitru Erhan, Aaron Courville, Yoshua Bengio, and Pascal Vincent. “Why does unsupervised pre-training help deep learning?” In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 201–208 (cit. on p. 38).

[61] G Esposito, AM Vitale, FPJ Leijten, AM Strik, AMCB Koonen-Reemst, P Yurttas, TJAA Robben, S Coonrod, and JA Gossen. “Peptidylarginine deiminase (PAD) 6 is essential for oocyte cytoskeletal sheet formation and female fertility”. In: *Molecular and cellular endocrinology* 273.1-2 (2007), pp. 25–31 (cit. on p. 13).

[62] Heng-Yu Fan and Qing-Yuan Sun. “Involvement of mitogen-activated protein kinase cascade during oocyte maturation and fertilization in mammals”. In: *Biology of reproduction* 70.3 (2004), pp. 535–547 (cit. on p. 20).

[63] Eleonora Inácio Fernandez, André Satoshi Ferreira, Matheus Henrique Miquelão Cecı'lio, Dóris Spinosa Chéles, Rebeca Colauto Milanezi de Souza, Marcelo Fábio Gouveia Nogueira, and José Celso Rocha. "Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data". In: *Journal of Assisted Reproduction and Genetics* 37 (2020), pp. 2359–2376 (cit. on p. 4).

[64] Sedighe Firuzinia, Seyed Mahmoodreza Afzali, Fatemeh Ghasemian, and Seyed Abolghasem Mirroshandel. "A robust deep learning-based multiclass segmentation method for analyzing human metaphase II oocyte images". en. In: *Computer Methods and Programs in Biomedicine* 201 (2021), p. 105946. DOI: 10.1016/j.cmpb.2021.105946 (cit. on p. 42).

[65] Rafael A Fissore, Frank J Longo, Everett Anderson, Jan B Parys, and Tom Ducibella. "Differential distribution of inositol trisphosphate receptor isoforms in mouse oocytes". In: *Biology of reproduction* 60.1 (1999), pp. 49–57 (cit. on p. 22).

[66] Greg FitzHarris, Petros Marangos, and John Carroll. "Changes in endoplasmic reticulum structure during mouse oocyte maturation are controlled by the cytoskeleton and cytoplasmic dynein". In: *Developmental biology* 305.1 (2007), pp. 133–144 (cit. on p. 20).

[67] Sophie D Fosså, Henriette Magelssen, Kari Melve, Anne B Jacobsen, Frøydis Langmark, and Rolv Skjærven. "Parenthood in survivors after adulthood cancer and perinatal health in their offspring: a preliminary report". In: *JNCI Monographs* 2005.34 (2005), pp. 77–82 (cit. on p. 19).

[68] Toshihiro Fujiwara, Ken Nakada, Hideki Shirakawa, and Shunichi Miyazaki. "Development of inositol trisphosphate-induced calcium release mechanism during maturation of hamster oocytes". In: *Developmental Biology* 156.1 (1993), pp. 69–79 (cit. on p. 20).

[69] Noritaka Fukunaga, Sho Sanami, Hiroya Kitasaka, Yuji Tsuzuki, Hiroyuki Watanabe, Yuta Kida, Seiji Takeda, and Yoshimasa Asada. "Development of an automated two pronuclei detection system on time-lapse embryo images using deep learning techniques". en. In: *Reproductive Medicine and Biology* 19.3 (2020), pp. 286–294. DOI: https://doi.org/10.1002/rmb2.12331 (cit. on p. 43).

[70] Barbara P Fulton and DG Whittingham. "Activation of mammalian oocytes by intracellular injection of calcium". In: *Nature* 273.5658 (1978), pp. 149–151 (cit. on p. 21).

[71] Fatima Gebauer, WENHAO Xu, Geoffrey M Cooper, and Joel D Richter. "Translational control by cytoplasmic polyadenylation of c-mos mRNA is necessary for oocyte maturation in the mouse." In: *The EMBO journal* 13.23 (1994), pp. 5712–5720 (cit. on p. 18).

[72] J. A. Gingold, N. H. Ng, J. McAuley, Z. Lipton, and N. Desai. "Predicting embryo morphokinetic annotations from time-lapse videos using convolutional neural networks". English. In: *Fertility and Sterility* 110.4 (2018). Publisher: Elsevier, e220. DOI: 10.1016/j.fertnstert.2018.07.634 (cit. on p. 43).

[73] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Region-based convolutional networks for accurate object detection and segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 38.1 (2015), pp. 142–158 (cit. on p. 28).

[74] X Glorot, A Bordes, and Y Bengio. "Deep sparse rectifier neural networks. InProceedings of the fourteenth international conference on artificial intelligence and statistics 2011 Jun 14 (pp. 315-323)". In: *JMLR Workshop and Conference Proceedings* (cit. on pp. 28, 30).

[75] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016 (cit. on p. 69).

[76] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016 (cit. on pp. 30, 32, 33).

[77] Ana Carla Gordo, Patricia Rodrigues, Manabu Kurokawa, Teru Jellerette, Ginger E Exley, Carol Warner, and Rafael Fissore. "Intracellular calcium oscillations signal apoptosis rather than activation in in vitro aged mouse eggs". In: *Biology of Reproduction* 66.6 (2002), pp. 1828–1837 (cit. on p. 20).

[78] Roger Gosden. "The vocabulary of the egg". In: *Nature* 383.6600 (1996), pp. 485–486 (cit. on p. 10).

[79] Roger Gosden, Jennifer Krapez, and David Briggs. "Growth and development of the mammalian oocyte". In: *Bioessays* 19.10 (1997), pp. 875–882 (cit. on p. 12).

[80] Roger Gosden, Bora Lee, et al. "Portrait of an oocyte: our obscure origin". In: *The Journal of clinical investigation* 120.4 (2010), pp. 973–983 (cit. on pp. 8, 14, 15).

[81] Roger G Gosden, Julie Mullan, Helen M Picton, Hang Yin, and Seang-Lin Tan. "Current perspective on primordial follicle cryopreservation and culture for reproductive medicine". In: *Human Reproduction Update* 8.2 (2002), pp. 105–110 (cit. on p. 23).

[82] Xutao Guo, Shang Lu, Yanwu Yang, Pengcheng Shi, Chenfei Ye, Yang Xiang, and Ting Ma. "Modeling Annotator Variation and Annotator Preference for Multiple Annotations Medical Image Segmentation". In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2022, pp. 977–984. DOI: 10.1109/BIBM55620.2022.9995619 (cit. on p. 68).

[83] Vanessa E Gurtu, Shelly Verma, Allie H Grossmann, R Michael Liskay, William C Skarnes, and Sean M Baker. "Maternal effect for DNA mismatch repair in the mouse". In: *Genetics* 160.1 (2002), pp. 271–277 (cit. on p. 19).

[84] Nasrin Haghighat and Lon J Van Winkle. "Developmental change in follicular cell-enhanced amino acid uptake into mouse oocytes that depends on intact gap junctions and transport system Gly". In: *Journal of Experimental Zoology* 253.1 (1990), pp. 71–82 (cit. on p. 11).

[85] Ehsan Hajiramezanali, Siamak Zamani Dadaneh, Alireza Karbalayghareh, Mingyuan Zhou, and Xiaoning Qian. "Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data". In: *Advances in Neural Information Processing Systems* 31 (2018) (cit. on p. 39).

[86] Toshio Hamatani, Mark G Carter, Alexei A Sharov, and Minoru SH Ko. "Dynamics of global gene expression changes during mouse preimplantation development". In: *Developmental cell* 6.1 (2004), pp. 117–131 (cit. on p. 16).

[87] Aleš Hampl and John J Eppig. "Analysis of the mechanism (s) of metaphase I arrest in maturing mouse oocytes". In: *Development* 121.4 (1995), pp. 925–933 (cit. on p. 20).

[88] Dimitri Hamzaoui, Sarah Montagne, Raphaële Renard-Penna, Nicholas Ayache, and Hervé Delingette. "MOrphologically-Aware Jaccard-Based ITerative Optimization (MOJITO) for Consensus Segmentation". en. In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*. Ed. by Carole H. Sudre, Christian F. Baumgartner, Adrian Dalca, Chen Qin, Ryutaro Tanno, Koen Van Leemput, and William M. Wells III. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022, pp. 3–13. DOI: 10.1007/978-3-031-16749-2_1 (cit. on p. 67).

[89] Jun Han and Claudio Moraga. “The influence of the sigmoid function parameters on the speed of backpropagation learning”. In: *International workshop on artificial neural networks*. Springer. 1995, pp. 195–201 (cit. on p. 31).

[90] Sarah E Harris, Henry J Leese, Roger G Gosden, and Helen M Picton. “Pyruvate and oxygen consumption throughout the growth and development of murine oocytes”. In: *Molecular Reproduction and Development: Incorporating Gamete Research* 76.3 (2009), pp. 231–238 (cit. on p. 11).

[91] Terry Hassold and Patricia Hunt. “To err (meiotically) is human: the genesis of human aneuploidy”. In: *Nature Reviews Genetics* 2.4 (2001), pp. 280–291 (cit. on p. 9).

[92] Chloe He, Neringa Karpaviciute, Cristina Hickman, and Nikica Zaninovic. “Artificial Intelligence (AI) in Gamete and Embryo Selection”. In: *Textbook of Assisted Reproductive Techniques* (2023), pp. 208–222 (cit. on p. 89).

[93] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*. 2015 (cit. on pp. 27, 35, 43).

[94] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology press, 2005 (cit. on p. 24).

[95] Björn Heindryckx, Stefanie De Gheselle, Jan Gerris, Marc Dhont, and Petra De Sutter. “Efficiency of assisted oocyte activation as a solution for failed intracytoplasmic sperm injection”. In: *Reproductive biomedicine online* 17.5 (2008), pp. 662–668 (cit. on p. 21).

[96] Javier Herrero and Marcos Meseguer. “Selection of high potential embryos using time-lapse imaging: the era of morphokinetics”. In: *Fertility and sterility* 99.4 (2013), pp. 1030–1034 (cit. on p. 1).

[97] Elke Heytens, J Parrington, K Coward, C Young, STEFANIE Lambrecht, S-Y Yoon, RA Fissore, R Hamer, CM Deane, M Ruas, et al. “Reduced amounts and abnormal forms of phospholipase C zeta (PLCζ) in spermatozoa from infertile men”. In: *Human Reproduction* 24.10 (2009), pp. 2417–2428 (cit. on p. 22).

[98] Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: *science* 313.5786 (2006), pp. 504–507 (cit. on p. 24).

[99] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. “Improving neural networks by preventing co-adaptation of feature detectors”. In: *arXiv preprint arXiv:1207.0580* (2012) (cit. on p. 29).

[100] Craig A Hodges, Ekaterina Revenkova, Rolf Jessberger, Terry J Hassold, and Patricia A Hunt. “SMC1β-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction”. In: *Nature genetics* 37.12 (2005), pp. 1351–1355 (cit. on p. 19).

[101] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017 (cit. on p. 43).

[102] TERVI HR. “Successful culture in vitro of sheep and cattle ova”. In: *J. Reprod. Fert.* 30 (1972), pp. 489–492 (cit. on p. 86).

[103] Joaquin Huarte, André Stutz, Marcia L O’Connell, Pascale Gubler, Dominique Belin, Andrew L Darrow, Sidney Strickland, and Jean-Dominique Vassalli. “Transient translational silencing by reversible mRNA deadenylation”. In: *Cell* 69.6 (1992), pp. 1021–1030 (cit. on p. 17).

[104] J Huntriss, R Gosden, M Hinkins, B Oliver, D Miller, AJ Rutherford, and HM Picton. "Isolation, characterization and expression of the human Factor In the Germline alpha (FIGLA) gene in ovarian follicles and oocytes". In: *Molecular Human Reproduction* 8.12 (2002), pp. 1087–1095 (cit. on p. 14).

[105] Vladimir Iglovikov and Alexey Shvets. "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation". In: *arXiv preprint arXiv:1801.05746* (2018) (cit. on p. 37).

[106] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. pmlr. 2015, pp. 448–456 (cit. on p. 31).

[107] ELVIN JA. "Molecular characterization of the follicle defects in the growth differentiation factor 9-deficient ovary". In: *Mol Endocrinol* 13 (1999), pp. 1035–1045 (cit. on p. 11).

[108] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. "Caffe: Convolutional architecture for fast feature embedding". In: *Proceedings of the 22nd ACM international conference on Multimedia*. 2014, pp. 675–678 (cit. on p. 26).

[109] Martin H Johnson. "From mouse egg to mouse embryo: polarities, axes, and tissues". In: *Annual Review of Cell and Developmental* 25 (2009), pp. 483–512 (cit. on p. 13).

[110] Keith T Jones. "Turning it on and off: M-phase promoting factor during meiotic maturation and fertilization". In: *MHR: Basic science of reproductive medicine* 10.1 (2004), pp. 1–5 (cit. on p. 20).

[111] Keith T Jones, John Carroll, and David G Whittingham. "Ionomycin, Thapsigargin, Ryanodine, and Sperm Induced Ca2+ Release Increase during Meiotic Maturation of Mouse Oocytes (*)". In: *Journal of Biological Chemistry* 270.12 (1995), pp. 6671–6677 (cit. on pp. 20, 22).

[112] H Joris, Z Nagy, H Van de Velde, A De Vos, and A Van Steirteghem. "Intracytoplasmic sperm injection: laboratory set-up and injection procedure". In: *Human reproduction* 13.suppl_1 (1998), pp. 76–86 (cit. on p. 1).

[113] S Kahraman, K Yakın, E Dönmez, H Şamlı, M Bahce, G Cengiz, S Sertyel, M Şamlı, and N İmirzalıoğlu. "Relationship between granular cytoplasm of oocytes and pregnancy outcome following intracytoplasmic sperm injection". In: *Human reproduction* 15.11 (2000), pp. 2390–2393 (cit. on p. 15).

[114] A Karpathy. *Commonly used activation functions* (cit. on p. 30).

[115] A. Khan, S. Gould, and M. Salzmann. "Segmentation of developing human embryo in time-lapse microscopy". In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. ISSN: 1945-8452. 2016, pp. 930–934. DOI: 10.1109/ISBI.2016.7493417 (cit. on p. 43).

[116] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on pp. 34, 70).

[117] Douglas Kline and Joanne T Kline. "Repetitive calcium transients and the role of calcium in exocytosis and cell cycle activation in the mouse egg". In: *Developmental biology* 149.1 (1992), pp. 80–89 (cit. on p. 21).

[118] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012) (cit. on pp. 25, 35).

[119] JZ Kubiak, M Weber, G Geraud, and B Maro. "Cell cycle modification during the transitions between meiotic M-phases in mouse oocytes". In: *Journal of cell science* 102.3 (1992), pp. 457–467 (cit. on p. 20).

[120] J Richard Landis and Gary G Koch. "The measurement of observer agreement for categorical data". In: *biometrics* (1977), pp. 159–174 (cit. on p. 46).

[121] Tingfung Lau, Nathan Ng, Julian Gingold, Nina Desai, Julian McAuley, and Zachary C. Lipton. "Embryo staging with weakly-supervised region selection and dynamically-decoded predictions". In: *arXiv:1904.04419 [cs]* (2019). arXiv: 1904.04419 (cit. on p. 43).

[122] Brian D. Leahy, Won-Dong Jang, Helen Y. Yang, Robbert Struyven, Donglai Wei, Zhe Sun, Kylie R. Lee, Charlotte Royston, Liz Cam, Yael Kalma, Foad Azem, Dalit Ben-Yosef, Hanspeter Pfister, and Daniel Needleman. "Automated Measurements of Key Morphological Features of Human Embryos for IVF". In: *arXiv:2006.00067 [cs, q-bio]* (2020). arXiv: 2006.00067 (cit. on p. 43).

[123] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. "Backpropagation applied to handwritten zip code recognition". In: *Neural computation* 1.4 (1989), pp. 541–551 (cit. on pp. 24, 27).

[124] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324 (cit. on pp. 24, 35).

[125] Bora Lee, Sook-Young Yoon, and Rafael A Fissore. "Regulation of fertilization-initiated [Ca2+] i oscillations in mammalian eggs: a multi-pronged approach". In: *Seminars in cell & developmental biology*. Vol. 17. 2. Elsevier. 2006, pp. 274–284 (cit. on pp. 20, 22).

[126] O Levinsohn-Tavor, S Friedler, M Schachter, A Raziel, D Strassburger, and R Ron-El. "Coasting—what is the best formula?" In: *Human Reproduction* 18.5 (2003), pp. 937–940 (cit. on p. 16).

[127] Lei Li, Boris Baibakov, and Jurrien Dean. "A subcortical maternal complex essential for preimplantation mouse embryogenesis". In: *Developmental cell* 15.3 (2008), pp. 416–425 (cit. on p. 19).

[128] Li-fang Liang, Selma M Soyal, and Jurrien Dean. "FIGα, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes". In: *Development* 124.24 (1997), pp. 4939–4947 (cit. on p. 14).

[129] Min Lin, Qiang Chen, and Shuicheng Yan. "Network in network". In: *arXiv preprint arXiv:1312.4400* (2013) (cit. on pp. 35, 36).

[130] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988 (cit. on p. 34).

[131] Lin Liu and David L Keefe. "Defective cohesin is associated with age-dependent misaligned chromosomes in oocytes". In: *Reproductive biomedicine online* 16.1 (2008), pp. 103–112 (cit. on p. 19).

[132] Shan Liu, Yuan Li, Xuan Gao, Jun-Hao Yan, and Zi-Jiang Chen. “Changes in the distribution of mitochondria before and after in vitro maturation of human oocytes and the effect of in vitro maturation on mitochondria distribution”. In: *Fertility and sterility* 93.5 (2010), pp. 1550–1555 (cit. on p. 20).

[133] Sheng Liu, Kangning Liu, Weicheng Zhu, Yiqiu Shen, and Carlos Fernandez-Granda. “Adaptive Early-Learning Correction for Segmentation From Noisy Annotations”. en. In: 2022, pp. 2606–2616 (cit. on p. 68).

[134] Xin-Yong Liu, Suo-Feng Mal, De-Qiang Miao, Dong-Jun Liu, Shorgan Bao, and Jing-He Tan. “Cortical granules behave differently in mouse oocytes matured under different conditions”. In: *Human Reproduction* 20.12 (2005), pp. 3402–3413 (cit. on p. 20).

[135] Z. Liu, B. Huang, Y. Cui, Y. Xu, B. Zhang, L. Zhu, Y. Wang, L. Jin, and D. Wu. “Multi-Task Deep Learning With Dynamic Programming for Embryo Early Development Stage Classification From Time-Lapse Videos”. In: *IEEE Access* 7 (2019). Conference Name: IEEE Access, pp. 122153–122163. DOI: 10.1109/ACCESS.2019.2937765 (cit. on p. 43).

[136] Alex Lopata. “History of the Egg in Embryology”. In: *Journal of Mammalian Ova Research* 26.1 (2009), pp. 2–9 (cit. on p. 8).

[137] Diana Lucifero, Carmen Mertineit, Hugh J Clarke, Timothy H Bestor, and Jacquetta M Trasler. “Methylation dynamics of imprinted genes in mouse germ cells”. In: *Genomics* 79.4 (2002), pp. 530–538 (cit. on p. 18).

[138] Durjoy Sen Maitra, Ujjwal Bhattacharya, and Swapan K Parui. “CNN based common approach to handwritten character recognition of multiple scripts”. In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE. 2015, pp. 1021–1025 (cit. on p. 39).

[139] J. Malmsten, N. Zaninovic, Q. Zhan, Z. Rosenwaks, and J. Shan. “Automated cell stage predictions in early mouse and human embryos using convolutional neural networks”. In: *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. ISSN: 2641-3604. 2019, pp. 1–4. DOI: 10.1109/BHI.2019.8834541 (cit. on p. 43).

[140] J. Malmsten, N. Zaninovic, Q. Zhan, M. Toschi, Z. Rosenwaks, and J. Shan. “Automatic prediction of embryo cell stages using artificial intelligence convolutional neural network”. English. In: *Fertility and Sterility* 110.4 (2018). Publisher: Elsevier, e360. DOI: 10.1016/j.fertnstert.2018.07.1005 (cit. on p. 43).

[141] Jonas Malmsten, Nikica Zaninovic, Qiansheng Zhan, Zev Rosenwaks, and Juan Shan. “Automated cell division classification in early mouse and human embryos using convolutional neural networks”. en. In: *Neural Computing and Applications* (2020). DOI: 10.1007/s00521-020-05127-8 (cit. on p. 43).

[142] AI Marques-Mari, Orly Lacham-Kaplan, JV Medrano, A Pellicer, and C Simón. “Differentiation of germ cells and gametes from stem cells”. In: *Human reproduction update* 15.3 (2009), pp. 379–390 (cit. on p. 23).

[143] Yoshio Masui and Clement L Markert. “Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes”. In: *Journal of Experimental Zoology* 177.2 (1971), pp. 129–145 (cit. on p. 20).

[144] Britta A Mattson and David F Albertini. “Oogenesis: chromatin and microtubule dynamics during meiotic prophase”. In: *Molecular reproduction and development* 25.4 (1990), pp. 374–383 (cit. on p. 12).

[145] Martin M Matzuk, Kathleen H Burns, Maria M Viveiros, and John J Eppig. “Intercellular communication in the mammalian ovary: oocytes carry the conversation”. In: *Science* 296.5576 (2002), pp. 2178–2180 (cit. on p. 11).

[146] Patrick E McKnight and Julius Najab. “Mann-Whitney U Test”. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1 (cit. on p. 60).

[147] Lisa M Mehlmann and Douglas Kline. “Regulation of intracellular calcium in the mouse egg: calcium release in response to sperm or inositol trisphosphate is enhanced after meiotic maturation”. In: *Biology of reproduction* 51.6 (1994), pp. 1088–1098 (cit. on p. 20).

[148] Lisa M Mehlmann, Katsuhiko Mikoshiba, and Douglas Kline. “Redistribution and increase in cortical inositol 1, 4, 5-trisphosphate receptors after meiotic maturation of the mouse oocyte”. In: *Developmental biology* 180.2 (1996), pp. 489–498 (cit. on p. 22).

[149] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. “The multimodal brain tumor image segmentation benchmark (BRATS)”. In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024 (cit. on p. 37).

[150] Funda Meric, Anjanette M Searfoss, Michael Wormington, and Alan P Wolffe. “Masking and unmasking maternal mRNA: The role of polyadenylation, transcription, splicing, and nuclear history”. In: *Journal of Biological Chemistry* 271.48 (1996), pp. 30804–30810 (cit. on p. 17).

[151] Marcos Meseguer, Javier Herrero, Alberto Tejera, Karen Marie Hilligsøe, Niels Birger Ramsing, and Jose Remohí. “The use of morphokinetics as a predictor of embryo implantation”. eng. In: *Human Reproduction (Oxford, England)* 26.10 (2011), pp. 2658–2671. DOI: 10.1093/humrep/der256 (cit. on p. 43).

[152] Lilyana Mihalkova, Tuyen Huynh, and Raymond J Mooney. “Mapping and revising markov logic networks for transfer learning”. In: *Aaai*. Vol. 7. 2007, pp. 608–614 (cit. on p. 39).

[153] Shun-ichi Miyazaki, Michisuke Yuzaki, Ken Nakada, Hideki Shirakawa, Setsuko Nakanishi, Shinji Nakade, and Katsuhiko Mikoshiba. “Block of Ca2+ wave and Ca2+ oscillation by antibody to the inositol 1, 4, 5-trisphosphate receptor in fertilized hamster eggs”. In: *Science* 257.5067 (1992), pp. 251–255 (cit. on p. 22).

[154] Yurii Evgen’evich Nesterov. “A method of solving a convex programming problem with convergence rate O\bigl(k^2\bigr)”. In: *Doklady Akademii Nauk*. Vol. 269. 3. Russian Academy of Sciences. 1983, pp. 543–547 (cit. on p. 34).

[155] Andrew Ng. “Nuts and bolts of building AI applications using Deep Learning”. In: *NIPS Keynote Talk* (2016) (cit. on p. 38).

[156] Nhan T T Nguyen and Phuoc Bao Le. “Topological Voting Method for Image Segmentation”. In: *J Imaging* 8.2 (18, 2022), p. 16. DOI: 10.3390/jimaging8020016 (cit. on p. 68).

[157] Alexandru Niculescu-Mizil and Rich Caruana. “Inductive transfer for Bayesian network structure learning”. In: *Artificial intelligence and statistics*. PMLR. 2007, pp. 339–346 (cit. on p. 39).

[158] Elena Oancea and Tobias Meyer. “Protein kinase C as a molecular machine for decoding calcium and diacylglycerol signals”. In: *Cell* 95.3 (1998), pp. 307–318 (cit. on p. 22).

[159] Kutluk Oktay, David Briggs, and Roger G Gosden. “Ontogeny of follicle-stimulating hormone receptor gene expression in isolated human ovarian follicles”. In: *The Journal of Clinical Endocrinology & Metabolism* 82.11 (1997), pp. 3748–3751 (cit. on p. 10).

[160] Fumio Otsuka and Shunichi Shimasaki. “A negative feedback system between oocyte bone morphogenetic protein 15 and granulosa cell kit ligand: its role in regulating granulosa cell mitosis”. In: *Proceedings of the National Academy of Sciences* 99.12 (2002), pp. 8060–8065 (cit. on p. 11).

[161] J Otsuki, A Okada, K Morimoto, Y Nagai, and H Kubo. “The relationship between pregnancy outcome and smooth endoplasmic reticulum clusters in MII human oocytes”. In: *Human Reproduction* 19.7 (2004), pp. 1591–1597 (cit. on p. 15).

[162] Jean-Pierre Ozil, Bernadette Banrezes, Szabolcs Tóth, Hua Pan, and Richard M Schultz. “Ca2+ oscillatory pattern in fertilized mouse eggs affects gene expression and development to term”. In: *Developmental biology* 300.2 (2006), pp. 534–544 (cit. on p. 21).

[163] Jean-Pierre Ozil and Daniel Huneau. “Activation of rabbit oocytes: the impact of the Ca2+ signal regime on development”. In: *Development* 128.6 (2000), pp. 917–928 (cit. on p. 21).

[164] Hua Pan, Pengpeng Ma, Wenting Zhu, and Richard M Schultz. “Age-associated increase in aneuploidy and changes in gene expression in mouse eggs”. In: *Developmental biology* 316.2 (2008), pp. 397–407 (cit. on p. 16).

[165] John Parrington, Sophie Brind, Humbert De Smedt, Ratthi Gangeswaran, F Anthony Lai, Richard Wojcikiewicz, and John Carroll. “Expression of inositol 1, 4, 5-trisphosphate receptors in mouse oocytes and early embryos: the type I isoform is upregulated in oocytes and downregulated after fertilization”. In: *Developmental Biology* 203.2 (1998), pp. 451–461 (cit. on p. 22).

[166] Franck Pellestor, Brigitte Andréo, Françoise Arnal, Claude Humeau, and Jacques Demaille. “Maternal aging and chromosomal abnormalities: new data drawn from in vitro unfertilized human oocytes”. In: *Human genetics* 112 (2003), pp. 195–203 (cit. on p. 19).

[167] Melissa E Pepling, James E Wilhelm, Ashley L O’Hara, Grant W Gephardt, and Allan C Spradling. “Mouse oocytes within germ cell cysts and primordial follicles contain a Balbiani body”. In: *Proceedings of the National Academy of Sciences* 104.1 (2007), pp. 187–192 (cit. on p. 12).

[168] Sally D Perreault, Robert A Wolff, and Barry R Zirkin. “The role of disulfide bond reduction during mammalian sperm nuclear decondensation in vivo”. In: *Developmental biology* 101.1 (1984), pp. 160–167 (cit. on p. 22).

[169] Lajos Piko and Kerry B Clegg. “Quantitative changes in total RNA, total poly (A), and ribosomes in early mouse embryos”. In: *Developmental biology* 89.2 (1982), pp. 362–378 (cit. on p. 16).

[170] Panayiota Ploutarchou, Pedro Melo, Anthony Day, Caroline Milner, and Suzannah Williams. “Molecular analysis of the cumulus matrix: Insights from mice with O-glycan-deficient oocytes”. In: *Reproduction* 149 (2015), pp. 533–43. DOI: 10.1530/REP-14-0503 (cit. on p. 42).

[171] Zbigniew Polanski, Emilie Ledan, Stéphane Brunet, Sophie Louvet, Marie-Hélène Verlhac, Jacek Z Kubiak, and Bernard Maro. “Cyclin synthesis controls the progression of meiotic maturation in mouse oocytes”. In: *Development* 125.24 (1998), pp. 4989–4997 (cit. on p. 20).

[172] Ferran Prados et al. “Spinal cord grey matter segmentation challenge”. en. In: *NeuroImage* 152 (2017), pp. 312–329. DOI: 10.1016/j.neuroimage.2017.03.010 (cit. on p. 68).

[173] Lorien Pratt. “Reuse of neural networks through transfer”. In: *Connection Science (Print)* 8.2 (1996) (cit. on p. 38).

[174] Lorien Pratt. *Special issue on inductive transfer*. Kluwer, 1997 (cit. on p. 38).

[175] Lorien Y Pratt. “Discriminability-based transfer between neural networks”. In: *Advances in neural information processing systems* 5 (1992) (cit. on p. 38).

[176] Annelies Raes, Georgios Athanasiou, Nima Azari-Dolatabad, Hafez Sadeghi, Sebastian Gonzalez Andueza, Josep Lluis Arcos, Jesus Cerquides, Krishna Chaitanya Pavani, Geert Opsomer, Osvaldo Bogado Pascottini, et al. “Manual versus deep learning measurements to evaluate cumulus expansion of bovine oocytes and its relationship with embryo development in vitro”. In: *Computers in Biology and Medicine* (2023), p. 107785 (cit. on p. 41).

[177] Tracy L Rankin, Zhi-Bin Tong, Philip E Castle, Eric Lee, Robert Gore-Langton, Lawrence M Nelson, and Jurrien Dean. “Human ZP3 restores fertility in Zp3 null mice without affecting order-specific sperm binding”. In: *Development* 125.13 (1998), pp. 2415–2424 (cit. on p. 15).

[178] Pradeep Reddy, Lian Liu, Deepak Adhikari, Krishna Jagarlamudi, Singareddy Rajareddy, Yan Shen, Chun Du, Wenli Tang, Tuula Hämäläinen, Stanford L Peng, et al. “Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool”. In: *Science* 319.5863 (2008), pp. 611–613 (cit. on p. 10).

[179] Joel D Richter. “CPEB: a life in translation”. In: *Trends in biochemical sciences* 32.6 (2007), pp. 279–285 (cit. on p. 17).

[180] P Rodrigues, D Limback, LK McGinnis, CE Plancha, and DF Albertini. “Oogenesis: prospects and challenges for the future”. In: *Journal of cellular physiology* 216.2 (2008), pp. 355–365 (cit. on p. 12).

[181] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv:1505.04597 [cs]* (2015). arXiv: 1505.04597 (cit. on pp. 35, 36, 42, 52, 53, 70).

[182] Frank Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” In: *Psychological review* 65.6 (1958), p. 386 (cit. on p. 24).

[183] Matthias Rottmann and Marco Reese. “Automated Detection of Label Errors in Semantic Segmentation Datasets via Deep Learning and Uncertainty Quantification”. en. In: 2023, pp. 3214–3223 (cit. on p. 68).

[184] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536 (cit. on p. 24).

[185] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115 (2015), pp. 211–252 (cit. on p. 25).

[186] Mitinori Saitou, Sheila C Barton, and M Azim Surani. “A molecular programme for the specification of germ cell fate in mice”. In: *Nature* 418.6895 (2002), pp. 293–300 (cit. on p. 10).

[187] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. “Tversky loss function for image segmentation using 3D fully convolutional deep networks”. In: *International workshop on machine learning in medical imaging*. Springer. 2017, pp. 379–387 (cit. on p. 34).

[188] Heide Schatten and Qing-Yuan Sun. “The role of centrosomes in mammalian fertilization and its significance for ICSI”. In: *Molecular human reproduction* 15.9 (2009), pp. 531–538 (cit. on p. 13).

[189] Jürgen Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural networks* 61 (2015), pp. 85–117 (cit. on p. 24).

[190] Richard M Schultz and Gregory S Kopf. “2 Molecular basis of mammalian egg activation”. In: *Current topics in developmental biology* 30 (1995), pp. 21–62 (cit. on p. 21).

[191] PF Serhal, DM Ranieri, A Kinis, S Marchant, M Davies, and IM Khadum. “Oocyte morphology predicts outcome of intracytoplasmic sperm injection.” In: *Human reproduction (Oxford, England)* 12.6 (1997), pp. 1267–1270 (cit. on p. 15).

[192] Geraldine Seydoux and Robert E Braun. “Pathway to totipotency: lessons from germ cells”. In: *Cell* 127.5 (2006), pp. 891–904 (cit. on p. 18).

[193] Alexander M Simon, Daniel A Goodenough, En Li, and David L Paul. “Female infertility in mice lacking connexin 37”. In: *Nature* 385.6616 (1997), pp. 525–529 (cit. on p. 12).

[194] Raffaella Simone, Daniel Čižmár, Susanne Holtze, Geert Michel, Anje Sporbert, Charlotte Okolo, and Thomas B Hildebrandt. “In vitro production of naked mole-rats’ blastocysts from non-breeding females using in vitro maturation and intracytoplasmic sperm injection”. In: *Scientific Reports* 13.1 (2023), p. 22355 (cit. on p. 86).

[195] Katrien Smits, Maarten Hoogewijs, H Woelders, Peter Daels, and Ann Van Soom. “Breeding or assisted reproduction? Relevance of the horse model applied to the conservation of endangered equids”. In: *Reproduction in Domestic Animals* 47 (2012), pp. 239–248 (cit. on p. 64).

[196] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. “Practical bayesian optimization of machine learning algorithms”. In: *Advances in neural information processing systems* 25 (2012) (cit. on p. 39).

[197] Nury Steuerwald, Jacques Cohen, Rene J Herrera, Mireia Sandalinas, and Carol A Brenner. “Association between spindle assembly checkpoint expression and maternal age in human oocytes”. In: *Molecular human reproduction* 7.1 (2001), pp. 49–55 (cit. on p. 16).

[198] Christelle Stouder, Samuel Deutsch, and Ariane Paoloni-Giacobino. “Superovulation in mice alters the methylation pattern of imprinted genes in the sperm of the offspring”. In: *Reproductive Toxicology* 28.4 (2009), pp. 536–541 (cit. on p. 18).

[199] You-Qiang Su, Koji Sugiura, and John J Eppig. “Mouse oocyte control of granulosa cell development and function: paracrine regulation of cumulus cell metabolism”. In: *Seminars in reproductive medicine*. Vol. 27. 01. © Thieme Medical Publishers. 2009, pp. 032–042 (cit. on p. 11).

[200] Koji Sugiura, You-Qiang Su, Francisco J Diaz, Stephanie A Pangas, Shweta Sharma, Karen Wigglesworth, Marilyn J O’Brien, Martin M Matzuk, Shunichi Shimasaki, and John J Eppig. “Oocyte-derived BMP15 and FGFs cooperate to promote glycolysis in cumulus cells”. In: (2007) (cit. on p. 11).

[201] Qing-Yuan Sun. “Cellular and molecular mechanisms leading to cortical reaction and polyspermy block in mammalian eggs”. In: *Microscopy Research and Technique* 61.4 (2003), pp. 342–348 (cit. on p. 13).

[202] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9 (cit. on p. 35).

[203] D Szollosi. "Changes of some cell organells during oogenesis in mammals". In: *Oogenesis* (1972) (cit. on p. 13).

[204] Masahito Tachibana, Michelle Sparman, Hathaitip Sritanaudomchai, Hong Ma, Lisa Clepper, Joy Woodward, Ying Li, Cathy Ramsey, Olena Kolotushkina, and Shoukhrat Mitalipov. "Mitochondrial gene replacement in primate offspring and embryonic stem cells". In: *Nature* 461.7262 (2009), pp. 367–372 (cit. on p. 23).

[205] Oliver H Tam, Alexei A Aravin, Paula Stein, Angelique Girard, Elizabeth P Murchison, Sihem Cheloufi, Emily Hodges, Martin Anger, Ravi Sachidanandam, Richard M Schultz, et al. "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes". In: *Nature* 453.7194 (2008), pp. 534–538 (cit. on p. 18).

[206] Fuchou Tang, Masahiro Kaneda, Dónal O'Carroll, Petra Hajkova, Sheila C Barton, Y Andrew Sun, Caroline Lee, Alexander Tarakhovsky, Kaiqin Lao, and M Azim Surani. "Maternal microRNAs are essential for mouse zygotic development". In: *Genes & development* 21.6 (2007), pp. 644–648 (cit. on p. 18).

[207] Ryutaro Tanno, Ardavan Saeedi, Swami Sankaranarayanan, Daniel C. Alexander, and Nathan Silberman. "Learning from noisy labels by regularized estimation of annotator confusion". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 11244–11253 (cit. on p. 73).

[208] Anna Targosz, Piotr Przystałka, Ryszard Wiaderkiewicz, and Grzegorz Mrugacz. "Semantic segmentation of human oocyte images using deep neural networks". en. In: *BioMedical Engineering OnLine* 20.1 (2021), p. 40. DOI: 10.1186/s12938-021-00864-w (cit. on p. 43).

[209] Fiona H Thomas, Jean-François Ethier, Shunichi Shimasaki, and Barbara C Vanderhyden. "Follicle-stimulating hormone regulates oocyte growth by modulation of expression of oocyte and granulosa cell factors". In: *Endocrinology* 146.2 (2005), pp. 941–949 (cit. on p. 11).

[210] Sebastian Thrun and Lorien Pratt. *Learning to learn*. Springer Science & Business Media, 2012 (cit. on p. 38).

[211] Raphael Vallat. "Pingouin: statistics in Python." In: *J. Open Source Softw.* 3.31 (2018), p. 1026 (cit. on p. 46).

[212] Barbara C Vanderfiyden, Evelyn E Telfer, and John J Eppig. "Mouse oocytes promote proliferation of granulosa cells from preantral and antral follicles in vitro". In: *Biology of reproduction* 46.6 (1992), pp. 1196–1204 (cit. on p. 11).

[213] BC Vanderhyden, JN Cohen, and P Morley. "Mouse oocytes regulate granulosa cell steroidogenesis". In: *Endocrinology* 133.1 (1993), pp. 423–426 (cit. on p. 11).

[214] Lucinda L Veeck. "An atlas of human gametes and conceptuses". In: *(No Title)* (1999) (cit. on p. 15).

[215] Arturo C Verrotti, Sunnie R Thompson, Christopher Wreden, Sidney Strickland, and Marvin Wickens. "Evolutionary conservation of sequence elements controlling cytoplasmic polyadenylylation." In: *Proceedings of the National Academy of Sciences* 93.17 (1996), pp. 9027–9032 (cit. on p. 17).

[216] Alejandra M Vitale, Meredith E Kennedy Calvert, Mallika Mallavarapu, Piraye Yurttas, Julie Perlin, John Herr, and Scott Coonrod. “Proteomic profiling of murine oocyte maturation”. In: *Molecular reproduction and development* 74.5 (2007), pp. 608–616 (cit. on p. 13).

[217] T Wakayama and R Yanagimachi. “Fertilisability and developmental ability of mouse oocytes with reduced amounts of cytoplasm”. In: *Zygote* 6.4 (1998), pp. 341–346 (cit. on p. 13).

[218] WHB Wallace, AB Thomson, and TW Kelsey. “The radiosensitivity of the human oocyte”. In: *Human reproduction* 18.1 (2003), pp. 117–121 (cit. on p. 19).

[219] Le-Ben Wan, Hua Pan, Sridhar Hannenhalli, Yong Cheng, Jun Ma, Andrew Fedoriw, Victor Lobanenkov, Keith E Latham, Richard M Schultz, and Marisa S Bartolomei. “Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development”. In: (2008) (cit. on p. 19).

[220] Chongyang Wang, Yuan Gao, Chenyou Fan, Junjie Hu, Tin Lun Lam, Nicholas D. Lane, and Nadia Bianchi-Berthouze. “AgreementLearning: An End-to-End Framework for Learning with Multiple Annotators without Groundtruth”. In: *arXiv preprint arXiv:2109.03596* (2021) (cit. on p. 68).

[221] Xinglong Wang, Beichen Zhao, Penghui Yang, Yameng Tan, Ruyi Ma, Shengqi Rao, Jianhui Du, Jian Chen, Jingwen Zhou, and Song Liu. “DUnet: A deep learning guided protein-ligand binding pocket prediction”. In: *bioRxiv* (2022), pp. 2022–08 (cit. on p. 37).

[222] Simon K. Warfield, Kelly H. Zou, and William M. Wells. “Validation of image segmentation and expert quality with an expectation-maximization algorithm”. In: *in Proceedings of Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part I*. 2002 (cit. on p. 67).

[223] Simon K. Warfield, Kelly H. Zou, and William M. Wells. “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation”. eng. In: *IEEE transactions on medical imaging* 23.7 (2004), pp. 903–921. DOI: 10.1109/TMI.2004.828354 (cit. on p. 67).

[224] Paul M Wassarman. “Zona pellucida glycoproteins”. In: *Annual review of biochemistry* 57.1 (1988), pp. 415–442 (cit. on p. 14).

[225] Paul M Wassarman and Wendy J Josefowicz. “Oocyte development in the mouse: an ultrastructural comparison of oocytes isolated at various stages of growth and meiotic competence”. In: *Journal of Morphology* 156.2 (1978), pp. 209–235 (cit. on p. 12).

[226] Toshiaki Watanabe, Yasushi Totoki, Atsushi Toyoda, Masahiro Kaneda, Satomi Kuramochi-Miyagawa, Yayoi Obata, Hatsune Chiba, Yuji Kohara, Tomohiro Kono, Toru Nakano, et al. “Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes”. In: *Nature* 453.7194 (2008), pp. 539–543 (cit. on p. 18).

[227] Ren Wu, Shengen Yan, Yi Shan, Qingqing Dang, and Gang Sun. “Deep image: Scaling up image recognition”. In: *arXiv preprint arXiv:1501.02876* 7.8 (2015), p. 4 (cit. on p. 37).

[228] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. “Empirical evaluation of rectified activations in convolutional network”. In: *arXiv preprint arXiv:1505.00853* (2015) (cit. on p. 30).

[229] Zhe Xu, Gregory S Kopf, and Richard M Schultz. "Involvement of inositol 1, 4, 5-trisphosphate-mediated Ca2+ release in early and late events of mouse egg activation". In: *Development* 120.7 (1994), pp. 1851–1859 (cit. on pp. 21, 22).

[230] Changning Yan, Pei Wang, Janet DeMayo, Francesco J DeMayo, Julia A Elvin, Cecilia Carino, Sarvamangala V Prasad, Sheri S Skinner, Bonnie S Dunbar, Jennifer L Dube, et al. "Synergistic roles of bone morphogenetic protein 15 and growth differentiation factor 9 in ovarian function". In: *Molecular endocrinology* 15.6 (2001), pp. 854–866 (cit. on p. 11).

[231] Wei Yao, Zhigang Zeng, Cheng Lian, and Huiming Tang. "Pixel-wise regression using U-Net and its application on pansharpening". In: *Neurocomputing* 312 (2018), pp. 364–371 (cit. on p. 37).

[232] Sook-Young Yoon, Teru Jellerette, Ana Maria Salicioni, Hoi Chang Lee, Myung-sik Yoo, Kevin Coward, John Parrington, Daniel Grow, Jose B Cibelli, Pablo E Visconti, et al. "Human sperm devoid of PLC, zeta 1 fail to induce Ca 2+ release and are unable to initiate the first step of embryo development". In: *The Journal of clinical investigation* 118.11 (2008), pp. 3671–3681 (cit. on p. 22).

[233] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. "Understanding neural networks through deep visualization". In: *arXiv preprint arXiv:1506.06579* (2015) (cit. on p. 27).

[234] Ling-Zhu Yu, Bo Xiong, Wen-Xue Gao, Chun-Min Wang, Zhi-Sheng Zhong, Li-Jun Huo, Qiang Wang, Yi Hou, Kui Liu, Johné Liu, et al. "MEK1/2 regulates microtubule organization, spindle pole tethering and asymmetric division during mouse oocyte meiotic maturation". In: *Cell Cycle* 6.3 (2007), pp. 330–338 (cit. on p. 20).

[235] Piraye Yurttas, Alejandra M Vitale, Robert J Fitzhenry, Leona Cohen-Gould, Wenzhu Wu, Jan A Gossen, and Scott A Coonrod. "Role for PADI6 and the cytoplasmic lattices in ribosomal storage in oocytes and translational control in the early mouse embryo". In: (2008) (cit. on p. 13).

[236] Matthew D Zeiler. "Adadelta: an adaptive learning rate method". In: *arXiv preprint arXiv:1212.5701* (2012) (cit. on p. 34).

[237] Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer. 2014, pp. 818–833 (cit. on p. 27).

[238] Hai-Miao Zhang and Bin Dong. "A review on deep learning in medical image reconstruction". In: *Journal of the Operations Research Society of China* 8 (2020), pp. 311–340 (cit. on p. 37).

[239] Jichang Zhang, Yuanjie Zheng, Wanchen Hou, and Wanzhen Jiao. "Leveraging non-expert crowdsourcing to segment the optic cup and disc of multicolor fundus images". EN. In: *Biomedical Optics Express* 13.7 (2022). Publisher: Optica Publishing Group, pp. 3967–3982. DOI: 10.1364/BOE.461775 (cit. on p. 68).

[240] Le Zhang, Ryutaro Tanno, Mou-Cheng Xu, Chen Jin, Joseph Jacob, Olga Ciccarelli, Frederik Barkhof, and Daniel C. Alexander. "Disentangling human error from the ground truth in segmentation of medical images". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS'20. Red Hook, NY, USA: Curran Associates Inc., 2020, pp. 15750–15762 (cit. on pp. 68, 70, 73).

[241] Xiao-yang Zhao, Wei Li, Zhuo Lv, Lei Liu, Man Tong, Tang Hai, Jie Hao, Chang-long Guo, Qing-wen Ma, Liu Wang, et al. “iPS cells produce viable mice through tetraploid complementation”. In: *nature* 461.7260 (2009), pp. 86–90 (cit. on p. 23).

[242] Ping Zheng and Jurrien Dean. “Role of Filia, a maternal effect gene, in maintaining euploidy during cleavage-stage mouse embryogenesis”. In: *Proceedings of the National Academy of Sciences* 106.18 (2009), pp. 7473–7478 (cit. on p. 19).

[243] Chang-Cheng ZHU and Richard JH Wojcikiewicz. “Ligand binding directly stimulates ubiquitination of the inositol 1, 4, 5-trisphosphate receptor”. In: *Biochemical Journal* 348.3 (2000), pp. 551–556 (cit. on p. 22).

[244] Luisa M Zintgraf, Taco S Cohen, and Max Welling. “A new method to visualize deep neural networks”. In: *arXiv preprint arXiv:1603.02518* (2016) (cit. on p. 27).

[245] Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin Dogus Cubuk, and Quoc Le. “Rethinking pre-training and self-training”. In: *Advances in neural information processing systems* 33 (2020), pp. 3833–3845 (cit. on p. 38).

[246] Kang Zou, Zhe Yuan, Zhaojuan Yang, Huacheng Luo, Kejing Sun, Li Zhou, Jie Xiang, Lingjun Shi, Qingsheng Yu, Yong Zhang, et al. “Production of offspring from a germline stem cell line derived from neonatal ovaries”. In: *Nature cell biology* 11.5 (2009), pp. 631–636 (cit. on p. 23).