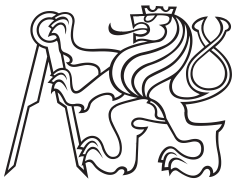
Master Thesis



Czech Technical University in Prague

F3

Faculty of Electrical Engineering
Department of Computer Science

Follicle segmentation in 2D ultrasound image sequences of ovaries

Bc. Lucie Borovičková

Supervisor: prof. Dr. Ing. Jan Kybic
Field of study: Open Informatics
Subfield: Artificial Intelligence
May 2024

MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Borovičková Lucie** Personal ID number: **483685**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Computer Science**

Study program: **Open Informatics**

Specialisation: **Artificial Intelligence**

II. Master’s thesis details

Master’s thesis title in English:

Follicle segmentation in 2D ultrasound image sequences of ovaries

Master’s thesis title in Czech:

Segementace folikul z 2D ultrazvukových sekvencí obrázků vaječníků

Guidelines:

An increasing number of people suffer from infertility problems. One of the important steps to determine a suitable treatment or to adjust its parameters is to examine the ovaries by ultrasound imaging. It is recommended to evaluate the number and size of the follicles, which is however rather time consuming. The goal of this work is to create and test an image processing pipeline to perform this evaluation automatically.
The data consisting of 2D ultrasound images and sequences will be provided to the student by the company Leeaf through collaboration with the General University Hospital in Prague (VFN). Annotated data from 94 patients is already available.
1. Get acquainted with existing methods.
2. Assemble, organize, preprocess and verify the quality of the available data.
3. Implement and evaluate a method for automatic follicle segmentation from individual images based on region growing (ref [1], Part I).
4. Implement and evaluate an improved version of the previous method by using spatial coherence within an image sequence (ref [1], Part II),
5. Implement and evaluate a deep learning based method for the same task (ref [2])
6. Based on the experimental evaluation, determine the most promissing method and its parameters. Suggest improvements and test them.
7. [Optional] If 3D ultrasound data is available, attempt to develop a method for follicle segmentation from this data.

Bibliography / sources:

1. Potočnik, Božidar, and Damjan Zazula. 2002. “Automated Analysis of a Sequence of Ovarian Ultrasound Images. Part II: Prediction-Based Object Recognition from a Sequence of Images.” Image and Vision Computing 20 (3): 227–35. https://doi.org/10.1016/s0262-8856(01)00097-x.
2. Li, Haoming, Jinghui Fang, Shengfeng Liu, Xiaowen Liang, Xin Yang, Zixin Mai, Tianfu Wang, Zhiyi Chen, and Dong Ni. 2020. “CR-Unet: A Composite Network for Ovary and Follicle Segmentation in Ultrasound Images.” IEEE Journal of Biomedical and Health Informatics 24 (4): 974–83. https://doi.org/10.1109/jbhi.2019.2946092.
3. Nixon, Mark. 2008. Feature Extraction & Image Processing. Amazon. 2nd edition. Amsterdam: Academic Press. https://www.amazon.com/Feature-Extraction-Image-Processing-Nixon/dp/0123725380.
4. Nahlawi, Suraya, and Nedi Gari. 2021. “Sonography Transvaginal Assessment, Protocols, and Interpretation.” PubMed. Treasure Island (FL): StatPearls

Publishing. 2021. https://www.ncbi.nlm.nih.gov/books/NBK572084/.

Name and workplace of master's thesis supervisor:

prof. Dr. Ing. Jan Kybic Biomedical imaging algorithms FEE

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **11.08.2023** Deadline for master's thesis submission: **24.05.2024**

Assignment valid until: **16.02.2025**

prof. Dr. Ing. Jan Kybic Supervisor's signature	Head of department's signature	prof. Mgr. Petr Páta, Ph.D. Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

.	
Date of assignment receipt	Student's signature

Acknowledgements

I would like to express my gratitude to my supervisor, prof. Dr. Ing. Jan Kybic for his patience with me and his guidance through the topic of this thesis. I would also like to thank my friends and family for their unlimited support during my studies at FEE CTU.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, 24. May 2024

Abstract

We implement and compare three methods with custom modifications for follicle segmentation and recognition in 2D ultrasound images and videos of ovaries. The first method employs a classical region growing algorithm, further refined by incorporating the second method - the Kalman filter. The third method utilizes deep learning techniques, specifically U-Net architecture. Our dataset comprises 110 individual ultrasound images and 82 videos cut into almost 1400 images. For training the neural networks, we applied data augmentations to extend the dataset profusely. The results of our experiments indicate the superiority of the deep learning methods over classical approaches. The region growing achieved an average $r_1 = 0.792$ and $r_2 = 0.804$ on the best-quality videos, whereas the U-Net reached an average $r_1 = 0.821$ and $r_2 = 0.839$ across all image qualities.

Keywords: follicle segmentation, ovarian 2D ultrasound, assisted reproduction, region growing, Kalman filter, U-Net

Supervisor: prof. Dr. Ing. Jan Kybic

Abstrakt

V této práci implementujeme a porovnáváme tři metody s vlastními úpravami pro segmentaci a rozpoznávání folikul v 2D ultrazvukových snímcích a videích vaječníků. První metoda využívá klasickou metodu narůstání oblastí, která je dále vylepšena použitím druhé metody - Kalmanova filtru. Třetí metoda využívá technik hlubokého učení, konkrétně architekturu U-Net. Náš dataset se skládá ze 110 samostatných ultrazvukových snímků a 82 videí, které byly rozřezány na téměř 1400 snímků. Pro trénování neuronových sítí jsme použili augmentaci dat, abychom dostatečně rozšířili dataset. Výsledky našich experimentů jasně ukazují převahu metod hlubokého učení nad klasickými přístupy. Metoda narůstání oblastí dosáhla průměrného $r_1 = 0.792$ a $r_2 = 0.804$ na videích nejlepší kvality, zatímco U-Net dosáhl průměrných hodnot $r_1 = 0.821$ a $r_2 = 0.839$ napříč všemi kvalitami snímků.

Klíčová slova: segmentace folikul, 2D ultrazvuk vaječníků, asistovaná reprodukce, metoda narůstání oblastí, Kalmanův filter, U-Net

Překlad názvu: Segementace folikul z 2D ultrazvukových sekvencí obrázků vaječníků

Contents

Figures

Tables

Chapter 1

Introduction

Between 48 million couples and 186 million individuals struggle with infertility globally [30]. Estimates of costs of successful treatment outcomes (delivery or ongoing pregnancy by 18 months) can climb up to 61, 377 USD in some countries [17]. An integral part of the treatment is a correct assessment of the couple's health, including an ovarian ultrasound examination. Automation and streamlining some of these tasks could help reduce the physician's time spent on repeated tasks and, therefore, bring down the immense price tag of becoming a parent.

This work replicates two methods done by Potočnik et al. [33] using region growing for segmentation and identification of follicles in 2D ultrasonography (USG) images and [34], which utilizes the previous method and employs Kalman filter (KF) for follicle segmentation in a video (image sequence). We implemented several adjustments to the method to fit the particulates of our data better. Furthermore, we reproduced a third method [22], which utilizes deep learning. Finally, we evaluated the results of each method and compared their performance.

1.1 Assisted reproduction process

Infertility is described by World Health Organization (WHO) as a disease of the male or female reproductive system defined by the failure to achieve a pregnancy after 12 months or more of regular unprotected sexual intercourse

[30].

When a couple struggling with infertility visits an In Vitro Fertilization (IVF) centre, both partners undergo copious amounts of tests to determine the reason for their struggles. The most common treatment, otherwise also called assisted reproductive technology (ART), in such a situation is an IVF, meaning that both the female and male gamete cells are combined outside the woman's body in a lab setting. To obtain the woman's gamete - an oocyte - it is necessary to stimulate her body with hormones, also called Controlled Ovarian Stimulation (COS), so a mature oocyte(s) can be collected. A mature oocyte is found inside a follicle sack collected from women's ovaries. There should ideally be more than one follicle. The normal response to the stimulation is 8 to 17 collected follicles [12]. The combined cells are reinserted into the woman's uterus as a 3 to 5 day old embryo [29] for further development.

1.2 Motivation

The prescribed treatment is, among other things, dependent on the woman's ovarian reserve. During the examination of the ovarian reserve, two parameters are considered the golden standard - Antral Follicle Count (AFC) and levels of anti-Müllerian hormone (AMH) [24]. While AMH can be obtained via a blood test, which a laboratory does, AFC has to be manually counted by the doctor. The counting takes around 10 minutes, and with the enormous pressure on doctors, it is no wonder that they resort to estimation with a verbal description such as the ovarian reserve is normal, poor, good, excellent and so on. Although the words used for description can be ordered to some extent, the intervals behind them differ for each doctor, which then causes discrepancies in the data.

Apart from AFC as the ovarian reserve marker, it is also considered good practice to count and measure the follicles during COS. The results are used for personalizing the treatment - adjusting the dosages of medications to control the growth of the follicles better and help prevent ovarian hyperstimulation syndrome (OHSS). Another utilization of follicle monitoring is the timing of the trigger (a very high dose of hormones administered to release the follicles) to obtain a higher number of follicles [7] [12].

This work focuses on images and videos obtained using a 2D ultrasound machine. Although 3D ultrasound machines that can measure the follicles'

volume are currently state-of-the-art, their cost can climb up to 200,000 USD, depending on the age, brand, model, portability and more [41]. This price tag means that only big hospitals can afford such machines, and the rest of the IVF centres still use 2D machines.

The dataset was obtained in cooperation with Cognitive IVF [1]. Cognitive IVF is a software company aiming to digitalize the assisted reproduction process, easing the doctor's workload and simplifying the patient journey. One of their products is Leeaf Physician Portal, an Electronic Health Records (EHR) system. According to their research, one of the features that doctors consider very useful is the automated monitoring of follicles using ultrasound images or videos. Hence their motivation for developing this feature.

[1] `http://www.leeaf.life`

Notation

$\mathbf{k}_{i,j}$	$k_i = \{\mathbf{k}_{i,j}, j = 0, \cdots, n-1\}$ where n is the number of pixels in the curve and $\mathbf{k}_{i,j} = [k_{x_{i,j}}, k_{y_{i,j}}]$
$\mathbf{p}$	pixel, vector $\mathbf{p} = [p_x, p_y]$
$\mathbf{z}_{i,j}$	$z_i = \{\mathbf{z}_{i,j}, j = 0, \cdots, m-1\}$ where m is the number of pixels in the curve and $\mathbf{z}_{i,j} = [z_{x_{i,j}}, z_{y_{i,j}}]$
$\sigma()$	standard deviation operator
Hom_i	ith homogeneous region
I	grey-level image
k_i	best follicle approximation in the ith image
$m()$	mean value operator
M_i	ith auxiliary matrix
O	outer boundary of region
S_i	ith image from the series of images cut from each video
T_i	various thresholds used in the algorithm
z_i	measurement (obtained by region growing algorithm) of selected follicle in the ith image
Filt	two times smoothed image

Acronyms

3D-US 3D ultrasound

AFC Antral Follicle Count

AMH anti-Müllerian hormone

ART assisted reproductive technology

COS Controlled Ovarian Stimulation

EHR Electronic Health Records

FPS frames per second

IVF In Vitro Fertilization

KF Kalman filter

MSR Misidentification Rate

NN neural network

OHSS ovarian hyperstimulation syndrome

PCOS polycystic ovary syndrom

RG region growing

RNN recurrent neural network

RR Recognition Rate

SVM Support Vector Machines

TVUS Transvaginal ultrasound

USG ultrasonography

WHO World Health Organization

Part I

Theoretical Part

Chapter 2

Related work

Transvaginal ultrasound (TVUS) is a safe and essential diagnostic tool for women dealing with infertility [27]. During ART, it allows the physician to observe and assess the women's reproductive system, especially the development of ovaries and follicles, monitor ovulatory time, and guide the timing of clinical embryo transfer [11]. The first efforts for automatic follicle detection in the ultrasound images were done in 1997 by Potocnik et al. [35]. They used thresholding to segment the regions and clinical rules, e.g. size and proportions of the regions, to accept or refute the region as follicle [4]. They followed up on their work with [33] and [34], where they refined their first method, used region growing and upgraded from single image to image sequence. This work follows their later work, which obtained Recognition Rate (RR)= 0.78 and Misidentification Rate (MSR)= 0.29.

Advanced methods used to tackle this subject include Support Vector Machines (SVM) [5], [21], K-means clustering [19] and neural networks [22]. RR range from 0.6 (MSR= 0.3) to 0.894 (MSR= 0.074), depending on the method and research paper. A non-exhaustive list of the methods and their performance can be found in [22].

Another valuable application of ovarian ultrasound image processing is the detection of polycystic ovary syndrom (PCOS) - a condition which makes conception more difficult [39]. One of the signs is a large number of follicular cysts in the ovary. Works dealing with this topic use filtering, watershed algorithm and some clinical criteria [6] and in combination with SVM and other methods [10].

Today, many bigger hospitals have a 3D ultrasound (3D-US) machine at their disposal and can, therefore, observe the follicles from an extra angle and estimate their volume. Those machines are pretty expensive and not common among medical facilities, and they are usually used to observe fetuses during pregnancy. However, methods segmenting the follicles in 3D-US can achieve better results than 2D. State-of-the-art works use unsupervised [38], semi-supervised [46], and supervised deep learning methods [32]. More works using 3D data were done using the public database USOVA3D [1].

[1] https://usova3d.um.si/wordpress/

Chapter 3

Image segmentation and identification methods

Below, we lay out the key terms for image segmentation and identification techniques, emphasizing the three methods used in this work: region growing, Kalman filter, and UNet-based techniques. We differentiate classical computer vision techniques and techniques based on artificial neural networks (deep learning).

3.1 Classical methods

Based on Kang et al. [16], the classical methods can be classified as edge-based and region-based. On the other hand, Raut et al. [36] divide segmentation algorithms into threshold-based, histogram-based, edge detection, region-based, and watershed Transformation techniques. Finally, Kaur and Kaur [18] add along edge-, threshold-, region-, and watershed-based methods as well as clustering and PDE-based methods. A recurring theme in the above classifications is the region-based techniques, with the most prominent example being region growing.

3.1.1 Region Growing

Region growing [13] is a traditional computer vision method for colour segmentation. It groups sub-regions into larger regions, starting with seed points and merging them with neighbouring pixels with similar properties based on the chosen criterion [16]. The algorithm stops when all pixels are assigned to a region. Multiple versions of the algorithm were developed through the years, varying, e.g., in the seeding method [1, 25], or in speed [42].

3.2 Kalman filter

The Kalman filter [15] is a linear quadratic estimation method originally used in the navigation domain [44] for estimating unknown variables from measurements. In image processing, the filter can be used, e.g., for noise filtering and image restoration [8], or tracking a segmented object on a series of images/video [45].

3.3 Deep learning methods

As with other computer vision tasks, deep learning has dominated the field of image processing, including image segmentation. Both Minaee et al. [26] and Ghosh et al. [9] divide the deep learning segmentation algorithms based on the type of neural network architecture into 10 categories, including methods using fully convolutional networks, encoder-decoder models, or recurrent neural networks.

One of the most frequented categories in image segmentation is the combination of encoder-decoder-based models with convolutional neural networks, embodied by models using the U-Net architecture [37]. This U-shaped neural network combines three deep learning concepts: convolutional layers, bottleneck autoencoders, and residual connections, and allows for a natural transformation of input images into segmentation masks.

Different improvements over the original U-Net were proposed for image segmentation. Gradual progress was made with UNet++ [47] and UNet

3+ [14]. Authors replaced the CNN modules with vision transformers in the Swin-UNet architecture [2]. Finally, other works added more complex operations to the residual connections: CR-UNet [22] introduced spatial RNN modules, while its later namesake [23] used channel attention modules.

Part II

Practical Part

Chapter 4

Data preparation

As data is a vital part of any work, a substantial part of our work was dedicated to consistent collection and annotation. The following chapter will present the challenges of ultrasound imaging, our data sources, the dataset description, annotation methods and the overall description of the preprocessing. Examples of ultrasound images and their annotation are provided in 4.1.

4.1 Raw dataset and Challenges

Any task concerning detection and segmentation in ultrasound images is challenging for several reasons; the most noticeable is probably the low quality of the images due to the speckle noise and acoustic shadows [3]. Another limitation of the images is imaging artifacts [20].

Ovarian follicles are round or oval sacks filled with fluid and potentially an oocyte. They are located in the ovaries and are usually observed via a TVUS. They can also appear slightly irregular if the ovary contains a lot of them and are squished together. They manifest as dark spots on the ultrasound images and vary greatly in size and count. There can be anywhere from 4 to 24 follicles for normofollicular and even more than 25 for multifollicular, with sizes ranging from 2 to 30 mm [28]. Therefore, another challenge is distinguishing the smaller ones from other bigger veins (which are manifesting in a similar manner) and the bigger ones from smaller cysts [22]. Illustration

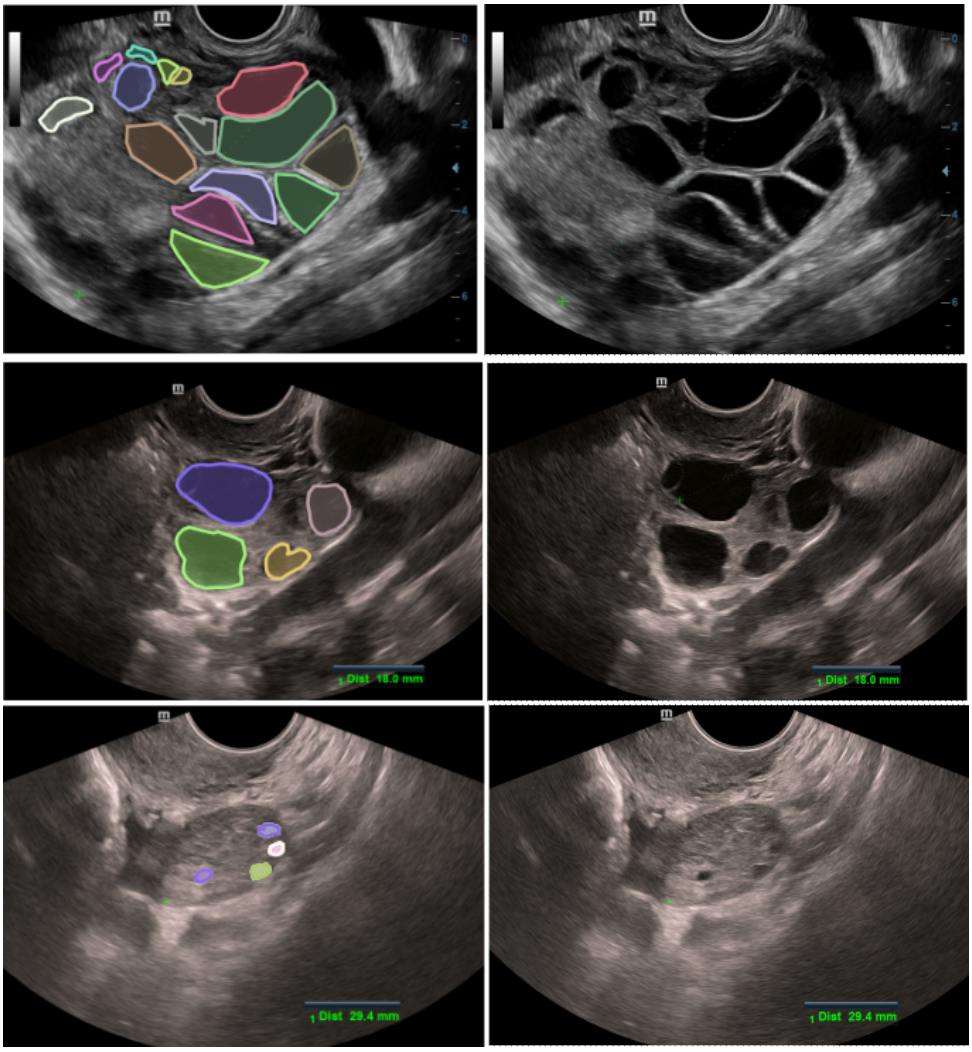


Figure 4.1: Example of differences in follicle counts and sizes

of different counts and sizes of follicles in ultrasound images can be seen in 4.1. Those parameters mainly depend on the woman's age and other health factors.

Lastly, the annotation can be quite challenging due to all of the factors mentioned above. Standardization among physicians is only fair, and consensus can be lacking in some situations.

video count	total imgs	avg imgs	min imgs	max imgs
82	1396	17.3	4	49

Table 4.1: Statistics regarding the number of images from videos.

4.2 Data sources

There are two data sources - first, an IVF centre in the Czech Republic, which provided only individual images and no videos. The dataset consists of images from 94 different patients, each having between 1 to 24 ultrasound images from one or multiple stimulation cycles; the total count is 582 images. The women visiting the clinic were struggling with infertility and therefore undergoing a COS. The images were taken before or during a medical check-up or before the oocyte pick-up procedure. The time span of the image taking was two years, from September 2019 to September 2022, using Mindray's DC-N3 PRO and DC-40 machines with iClear 2 or 3 Technology. All images were in BMP format, with resolutions of 800×600 and 1260×910 pixels for DC-N3 PRO and DC-40, respectively.

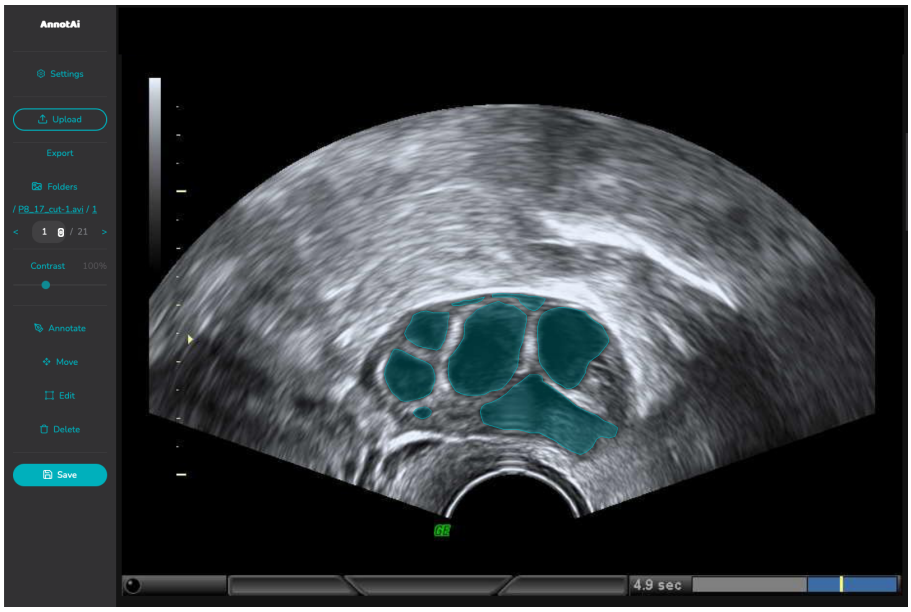
During the annotation of the images, we found that 110 contained one of the ovaries containing at least one follicle. The rest were primarily images of the uterus, endometrium, cysts or free spaces after the cyst puncture.

The second dataset, consisting of videos, was obtained in cooperation with the General University Hospital in Prague. The ultrasounds were conducted by prof. MUDr. Jaromír Mašata, CSc. with ethical approval by the hospital's internal ethical committee. In total, we obtained 78 videos from 20 patients who visited the hospital after failed attempts for conception. So, it can be presumed they were struggling with some degree of infertility. The videos were done using the Voluson E8 ultrasound machine manufactured by GE Healthcare and were between 1.5 and 10 seconds long. The videos were collected in the span of a year from March 2023 to March 2024. The videos often depicted the same clusters of follicles multiple times, and some of them did not contain follicles but endometrium. Therefore, we subdivided the videos into smaller segments to ensure each cluster was shown only once in an individual video. Some clusters were shown in multiple videos. This process resulted in 82 videos.

The resulting videos were split into individual sequential images with 7 to 10 frames per second (FPS). Altogether this resulted in 1396 number of images. The statistics regarding the number of images from videos are explained in Table 4.1 in detail.

4.3 Annotation of the dataset

For security reasons, we used an annotation tool that ran on the Leeaf server. The access was further restricted with a virtual private network. The annotation tool was custom-developed by a Leeaf employee in coordination with us to be user-friendly for the annotators and to be connected to the Leeaf database. A preview is included in 4.2



The image contains visualized annotation - green shapes. According to the doctor, there are also other dark structures, which are probably veins or intestinal contents.

Figure 4.2: Custom-made annotation tool: Annotai

Medical students in their fifth and sixth year did the first round of annotation. A gynaecologist with more than 20 years of practice then checked them and, if necessary, edited the annotations. The annotations were exported in the COCO JSON format [1].

[1] `https://cocodataset.org`

Chapter 5

Region growing method

In the following chapter, we will describe the algorithm for retrieving and segmenting the follicles in ultrasound images. This work closely follows [33], so we will often refer to their work in the following chapter. They divided their algorithm into three parts: 1. Homogeneous region identification, 2. Region growing, and 3. Follicle extraction. Throughout this work, we will use the notation described in Notation. Illustrations of differences between the original method and custom modifications are provided in Section 8.1.

The complete source code for all methods is available at the GitHub repository at `https://github.com/crimsoncress/diploma_thesis` and as an attachment to this thesis.

This region growing (RG) method was used as a baseline for evaluating more advanced methods and obtaining the initial measurements for Kalman filter 6. A measured follicle is denoted by z_i, where i represents the image index in the sequence in which the measurement was taken.

5.1 Image preprocessing

After loading the BMP image and its annotation in the COCO JSON format, several actions needed to be done to proceed with the algorithm itself. Firstly, we cropped the image so that there would not be any extra parts that could

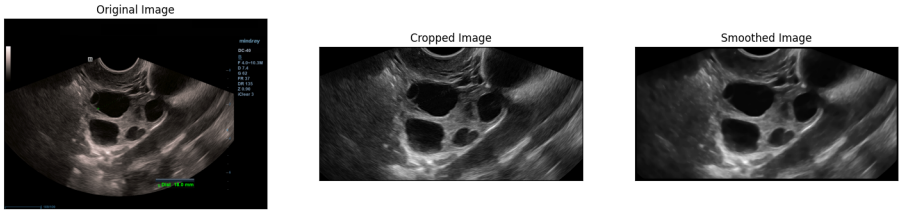


Figure 5.1: Results of preprocessing part.

potentially undesirably influence the grey scale 5.1. We recognized that there are four different image types, which require different cropping. The first partition is according to the position of the narrower part. It is either at the top or bottom of the image. Both of the image types ("top" and "bottom") then have an additional image type - "old top" and "old bottom" - those are images taken with DC-N3 PRO technology and thus have lower resolution. Secondly, we smoothed the image using the adaptive neighbourhood median filter with two kernel sizes based on comparing the grey scale of the evaluated pixel with the selected threshold T_1. We set the T_1 to the mean of grey-level values in the original image I_k. A kernel 11×11 was used for pixels below the threshold T_1, and a kernel 5×5 was used for pixels above this threshold. As stated, in ultrasound images, ovarian follicles are dark oval shapes, so the pixels representing them should have higher grey-scale values than the average pixel. Thus, this procedure smoothed them more thoroughly than their neighbourhood. This procedure was repeated twice, and the smoothed image was denoted as Filt.

5.2 First Part: Homogeneous Region Identification

This first part focuses on finding homogeneous regions of similar grey-level values in the image, and so obtaining a rough estimation of the follicles. We introduce three methods for obtaining homogeneous regions using different thresholds and calculations.

Obtaining the first homogeneous region, denoted as Hom_1, constitutes of two parts. First, we create a binary matrix M_1 as follows in 5.1.

$$M_1(\mathbf{p}) = \begin{cases} 1, & \mathrm{Filt}(\mathbf{p}) \leq T_2 \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

The original paper sets the threshold T_2 was set to the $m(\mathrm{Filt}) - \sigma(\mathrm{Filt})$. Nevertheless, after running the finished algorithm on our dataset, we decided to set the threshold to $m(\mathrm{Filt}) - \sigma(\mathrm{Filt})/1.2$ as the homogeneous region

highlights the follicles more precisely. To avoid merging adjacent follicles, we employ a second part - binary watershed segmentation using a Euclidean distance map of M_1. The result was converted back to a binary matrix and denoted as Hom_1.

For the second homogeneous region Hom_2, we introduced a new auxiliary matrix M_2 by calculating one standard deviation of grey levels in the 11×11 neighbourhood for every pixel in the original image I. The Hom_2 is described in 5.2

$$\text{Hom}_2(\mathbf{p}) = \begin{cases} 1, & M_2(\mathbf{p}) \leq T_3 \wedge I(\mathbf{p}) \leq T_1 \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

For the third region Hom_3, we constructed an auxiliary matrix M_3 as 5.3 and removed all regions touching the image's border.

$$M_3(\mathbf{p}) = \begin{cases} 1, & \text{Filt}(\mathbf{p}) \leq T_1 \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

Potocnik et al. merge all three homogeneous regions under Hom in the original algorithm. However, we empirically found out that due to the nature of our images, it is better to merge only the first and the third homogeneous regions to obtain better performance on our dataset. We assumed that the ovary is located somewhere in the middle of the image and thus deleted all the regions touching the image border from Hom. Subsequently, we defined a new threshold T_4 and removed all regions smaller than T_4. This threshold corresponded to the minimal size of a follicle of around 220 square pixels.

5.3 Second Part: Region Growing

In the following section, we grow the obtained coarse estimation of the follicles in Hom. We process the bigger regions first as they are more likely to be the centres of the follicles. The region-growing procedure consists of iteratively assessing the outer boundary O of a region R and adding the pixels from the outer boundary to the region if they fulfil some criteria. We denote the initial region R_0 and then R_i for the ith iteration. This process stops after the nth iteration when no pixel is added, that is, $R_{n-1} = R_n$.

The outer boundary O is a set of pixels adjacent to the region. However, for a pixel $\mathbf{p}$ to be considered a potential candidate for merging, it needs to

have at least 4 of its neighbours (3×3 neighbourhood) from the region we want to grow to control the compactness of the region. Furthermore, $\mathbf{p}$ needs to satisfy two criteria. First - 5.4 assures that the grey levels of the potential candidate are statistically close to the region.

$$|I_k(\mathbf{p}) - m(R_0)| \leq \alpha\sigma(R_0) \tag{5.4}$$

Second - 5.7, where $\text{grad}(\mathbf{p})$ is calculated using 5.6 and texture statistics $\text{tex}(\mathbf{p})$ as in 5.5 where $n_{11}(\mathbf{p})$ are the grey levels of 11×11 neighbourhood of pixel.

$$\text{tex}(\mathbf{p}) = m(n_{11}(\mathbf{p}))/\sigma(n_{11}(\mathbf{p})) \tag{5.5}$$

$$\text{grad}(\mathbf{p}) = ||\nabla I_k(\mathbf{p})||(e^{G/\text{tex}(\mathbf{p})} - 1) \tag{5.6}$$

$$|\text{grad}(\mathbf{p}) - m(\text{grad}(R_0))| \leq \alpha\sigma(\text{grad}(R_0)) \tag{5.7}$$

If the pixel satisfies all the criteria, we add it to the region. If it is already a member of another region P_m, we may merge the two regions if they are statistically similar. This hypothesis is checked using two critical terms 5.8, 5.9 and a threshold T_5.

$$\text{CT}_1 = \frac{|m(R_0) - m(P_0)|}{\sigma(R_0 \cup P_0)} \tag{5.8}$$

$$\text{CT}_2 = \frac{|m(\text{grad}(R_0)) - m(\text{grad}(P_0))|}{\sigma(\text{grad}(R_0 \cup P_0))} \tag{5.9}$$

As it can be shown, CT_1 and CT_2 follow the Student's distribution t, so the value of T_5 is determined by the confidence interval. In our case, it is 0.05 and $T_5 = 1.66$. The regions are merged if both CT_1 and CT_2 are below T_5. The regions are then checked for any potential holes that have been filled.

5.4 Third Part: Follicle Extraction

After the second part, we should obtain all the darker round structures in the image, and we need to decide which of them are follicles based on domain knowledge. We will remove all regions touching the image's border or having an area smaller than 220 pixels - the threshold is calculated from the minimal size of a follicle. Next, we calculate the ratio b for each follicle as the ratio between its area and bounding box. If b is higher than 0.5, we remove this region as we do not consider it compact or round enough. After this step, we need to determine the most likely follicle. We sort all potential follicles by their area in descending order and by the b ratio in descending order. We then add the follicle that scores the best in both categories as the actual follicles set. Next, we iteratively calculate the centre of gravity of all actual follicles and add a new one closest to the centre of the actual follicles set. This process stops when no other potential follicles or the closest follicle is further than $0.25 * \max(\text{image width, image height})$.

Chapter 6

Kalman filter predictor-corrector application

In the previous chapter, we showed how using the RG method for follicle segmentation in independent ultrasound images can yield accurate results. However, most ultrasound machines are able to capture a short ultrasound video (a sequence of images), which could possibly provide more precision in determining the follicle positions. Therefore, in this chapter, we will segment follicles from ultrasound videos and subsequently compare the results with the independent images method.

After converting the videos to image sequences as described in 4.3, we need to exploit the property of the image sequence. The consecutive images vary only slightly as the examiner moves the ultrasound probe around the ovary, so the follicles appear to be either moving or changing in size. This alternation will be captured by applying the KF predictor-corrector scheme as similarly done in [34]. We will apply the filter independently on each of the follicles.

In short, the process adjusts the measurement of an object z_i, in our case, a follicle, according to the best approximation of its position k_{i-1} in the previous image S_{i-1} and estimated shifts along the x and y axes.

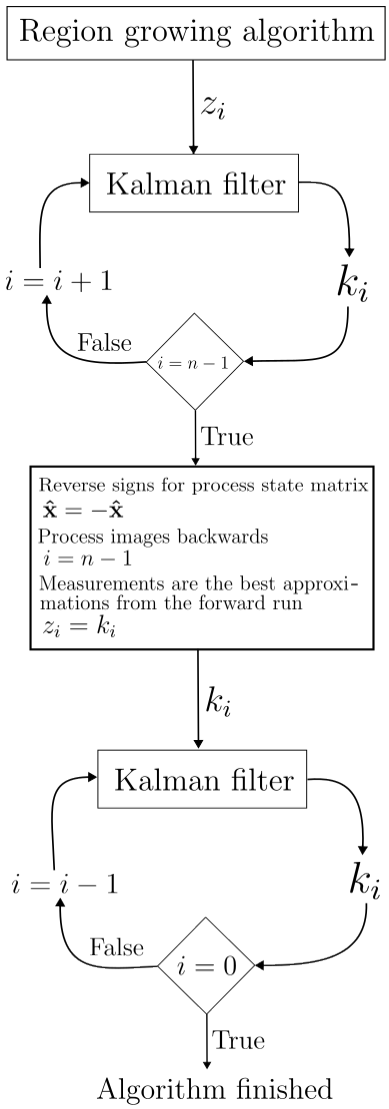


Figure 6.1: Overview of the full algorithm run. Detailed Kalman filter is described in 6.2

6.1 Tracking of a single follicle

After obtaining the images with annotations done by the RG algorithm, the first step was to follow the follicles throughout the sequence of images and determine their corresponding follicle matches in subsequent images. A follicle j from image $i-1$ is matched with a follicle k_{max} from subsequent image i if they form the biggest (among all follicles in image i) non-zero intersection with each other.

$$k_{max} = \max_{k} ||z_{i-1}^{j} \cap z_{i}^{k}|| \tag{6.1}$$

Moreover, the follicle has to appear in at least two consecutive images to be considered. It is otherwise discarded as a misidentification by the RG algorithm. Each follicle then belongs to a so-called one-follicle-sequence, a set of follicles (following the conditions defined above) found in a subset of sequential images (one follicle in one image). As one image can contain more follicles, it will belong to the same number of one-follicle-sequences.

6.2 Kalman filter application

In this work, we applied the Discrete Kalman filter [40], meaning that the follicles are not tracked as a whole object by tracking their centre of gravity but by tracking each pixel of the follicle curve separately. After running the algorithm we will get a curve $\mathbf{k}_0 = \{\mathbf{k}_{0,j}, j = 0, \ldots, n-1\}$ where n is the number of pixels in the curve. We will call this curve the best follicle approximation, noting that the "best" notion is with regard to our algorithm. Each of the pixels corresponds to a process state vector of KF $\mathbf{x}_{i,j} = [x_{i,j}, y_{i,j}, \Delta x_{i,j}, \Delta y_{i,j}]^T$ where Δx and Δy correspond to the shifts of a pixel over x and y axis, respectively. Horizontal stacking those vectors creates the process state matrix $\hat{\mathbf{x}}$ of KF.

The process is depicted in 6.1 with a detailed description of KF in 6.2. The application starts with initializing some variables and then loops the measurement and time update until the selected follicle's last image is found. We save all the best follicle approximations in each image and also $\hat{\mathbf{x}}$ and $\mathbf{P}_0^-$ after the last iteration. Those values will serve as the initial values in the reverse run of the filter. In the reverse run, the measurements are no longer obtained from the RG algorithm as they were in the forward run, but they are substituted with the best approximations from the forward run from respective images.

6.2.1 Algorithm Initialization

We start with the first image in the sequence S_0 and select one of the follicles in the image. Since we do not have any previous position and shifts, in other words, the process state matrix, we declare the measurement of the selected follicle z_0 as our best approximation of this follicle k_0 in S_0. As initialization of our algorithm, we set the shifts in process vectors to zeros $\hat{\mathbf{x}}_{0,j} = [k_{0,x_j}, k_{0,y_j}, 0, 0]^T$ and the a priori estimate error covariance matrix

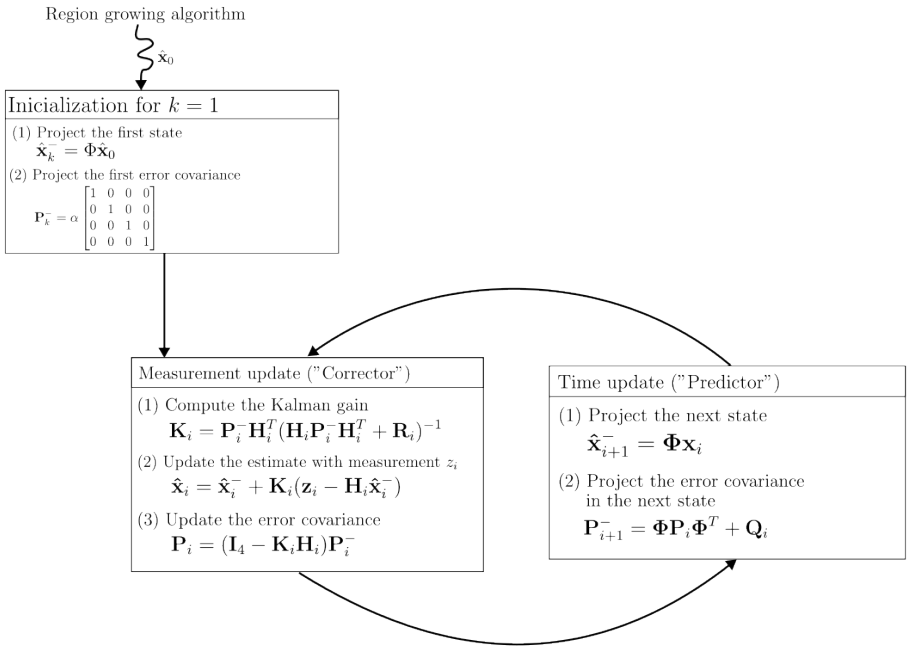


Figure 6.2: Overview of Kalman filter [40].

$\mathbf{P}_0^- = \alpha I_4$. As done in the [33], we set the α to 1000 and note that any value higher than 100 should not make a significant difference.

Since we want the model to predict the coordinates of pixels in image i by adding the shifts from the image $i-1$ to the coordinates in the image $i-1$, the $\mathbf{\Phi}$ matrix for this operation is as follows:

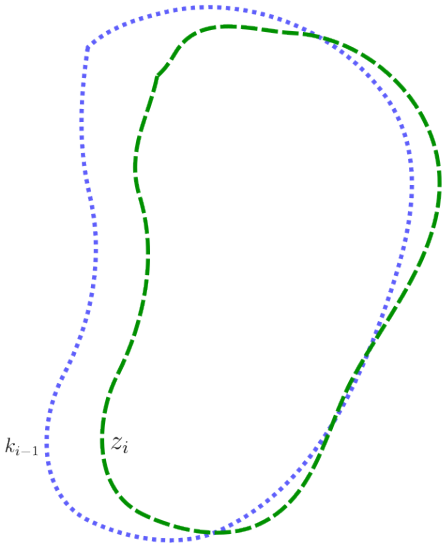
$$\mathbf{\Phi} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.2}$$

According to the KF theory [34]

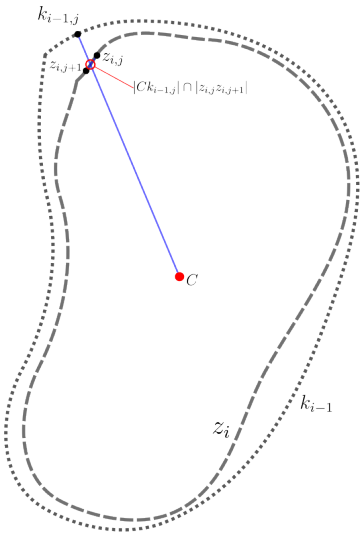
$$\mathbf{x}_{i+1,j} = \mathbf{\Phi}\mathbf{x}_{i,j} + \mathbf{w}_i, \tag{6.3}$$

where $\mathbf{w}_i$ is the noise in the system. We will estimate the initial position for the same follicle in the following image with

$$\hat{\mathbf{x}}_{1,j}^- = \mathbf{\Phi}\mathbf{x}_{0,j} \tag{6.4}$$



(a) : Best follicle approximation from the previous image k_{i-1} and its corresponding measurement from following image z_i



(b) : Closest intersection point $|Ck_{i-1,j}| \cap |z_{i,j}z_{i,j+1}|$

Figure 6.3: Matching points from corresponding curves

6.2.2 Measurement update ("Corrector")

We iteratively go over the one-follicle-sequence (a subset of the complete sequence of images 6.1), and for each $i, i = 1, \ldots, m$ where m is the number of images in the sub-sequence, execute the following list of operations/steps. The equation for this step was resourced from [40].

Firstly, we retrieve the best follicle approximation from the previous image in the sequence k_{i-1} and for every pixel $\mathbf{k}_{i-1,j}$ of this curve find a corresponding point from the obtained measurement of the currently-processed follicle $\mathbf{z}_{i,j} = [z_{x_{i,j}}, z_{y_{i,j}}]$. Such match is found by firstly aligning centres of mass of the two follicles and secondly determining intersections of a half-line defined by the centre of mass $C = [c_x, c_y] = [\sum_{j=1}^{n} z_{x_{i,j}}/n, \sum_{j=1}^{n} z_{y_{i,j}}/n]$ and the matching pixel $\mathbf{k}_{i-1,j}$ with lines connecting every two neighbouring pixels from the current curve. And choose the closest intersection point as the match. The match is then moved back using the negative value of the translation vector used for the follicle alignment. Now, the boundary pixels are paired with their respective matches. Empirically, we found out that it is necessary at this point to decide the correctness of the follicle annotation from RG algorithm and, if necessary, exclude it from the one-follicle sequence so it does not confuse the KF further on. The decision process is described in 6.2.3. After this check, we calculate the measurement update, starting with

the Kalman gain

$$\mathbf{K}_i = \mathbf{P}_i^- \mathbf{H}_i^T (\mathbf{H}_i \mathbf{P}_i^- \mathbf{H}_i^T + \mathbf{R}_i)^{-1} \tag{6.5}$$

where $\mathbf{R}_0 = \mathbf{I}_2$ and $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$. The next step is updating the time estimation with measurement update by

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_i^-). \tag{6.6}$$

Lastly, in the measurement update, we calculate the posterior estimate error covariance

$$\mathbf{P}_i = (\mathbf{I}_4 - \mathbf{K}_i \mathbf{H}_i)\mathbf{P}_i^- \tag{6.7}$$

6.2.3 Follicle exclusion

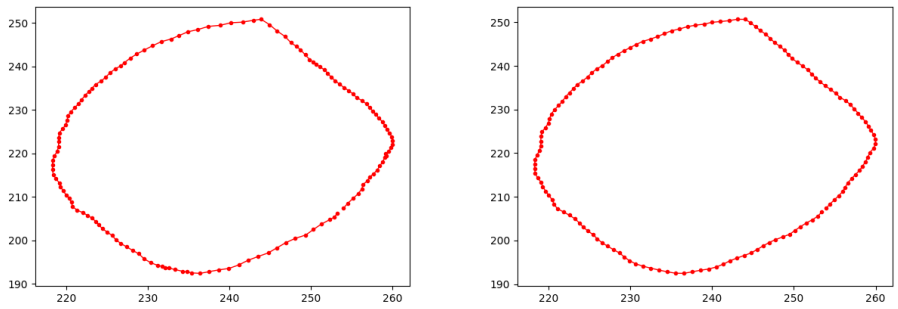
As mentioned at the end of 6.2.2, we detected that KF was sometimes misguided by the RG, and this mystification was carried over to the next follicle in the one-follicle sequence due to the nature of this method. Therefore, we decided to check the soundness of the follicle measurement from RG by comparing the minimal and maximal distance between the pair matches. If this difference exceeds some threshold, we disregard the follicle and continue with the next one in the sequence. The rationale behind this criterion is that the follicle grows very disproportionately. We experimentally set the threshold to 50. One issue arises when we are processing the next follicle in sequence and need a time update from the previous follicle and also the previous follicle itself to compute the KF. When such situation occurs we try to go back to $i-2$ follicle and if no such follicle exists we disregard the time update and only consider the measurement and noise in the current image.

6.2.4 Resampling curve

We resample the curve to have a defined number (k) of points to ensure they are well-spaced. Let p and q be two neighbouring points in the curve, and E be a set of all such pairs from a curve. Firstly, we calculate the distance between every two pairs $d(p,q) = |p-q|$, where $E_i = (p,q)$ (the length of their edge) and sum those distances to get the full length of the curve

$$D = \left(\sum_{i=0}^{k-1} d(c_i, c_{i+1}) \right) + d(c_n, c_0), \tag{6.8}$$

where n is the number of all points and edges and $c = (\hat{\mathbf{x}}_i)_{r,c}$, where $r = 0, \cdots, n$ and $c = 0, 1$ representing the first two columns in $\hat{\mathbf{x}}_i$ i.e. only the



(a) : Follicle before resampling **(b) :** Follicle after resampling

Figure 6.4: Illustration of follicle points resampling

coordinates of the points. By dividing the curve length by k, we get the new edge length $s_{new} = D/k$. For each of the edge (p, q) we determine the new point(s) v using linear interpolation

$$v = [(1-t)p_x + tq_x, (1-t)p_y + tq_y], \tag{6.9}$$

where t is a parameter that helps determine the number of new points for each edge. We assign new points the shifts from the closer of the points $E_i[0]$ or $E_i[1]$. Vertical stacking the points from V, vertical stacking the shifts from S and horizontal stacking those two matrices creates the new state process vector $\mathbf{x}_{new}$. Therefore, the best follicle approximation in image i is $(\hat{\mathbf{x}}_i)_{r,c}$, where $r = 0, \cdots, n$ and $c = 0, 1$ the first two rows in the process state matrix. A pseudocode is provided in 1.

Algorithm 1 Curve resampling pseudocode

```
V ← [E_0[0]]                                   ▷ V is a list of the new points
S ← [(x̂_i)_{0,{0,1}}]      ▷ S is a list of shifts corresponding to the new points
for i = 1, 2, ..., k do
    f_t ← d(E_i)/s_new
    while t + f_t ≥ |V| < k do
        t ← (|V| − t)/f_t
        v ← [(1 − t)p_x + tq_x, (1 − t)p_y + tq_y]
        V ← V + v
        u ← arg min_j d(v, E_i[j])
        S ← S + (x̂_i)[u]                       ▷ Add shifts corresponding to u
    end while
    t ← t + f_t
end for
```

6.2.5 Time update ("Predictor")

As the last part of a single loop, we calculate the time estimate update for the next one by estimating the position of the follicle in image $i+1$

$$\hat{\mathbf{x}}^{-}_{i+1,j} = \mathbf{\Phi}\mathbf{x}_{i,j}, \tag{6.10}$$

and projecting the error covariance matrix

$$\mathbf{P}^{-}_{i+1} = \mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}^T + \mathbf{Q}_i, \tag{6.11}$$

where $\mathbf{Q}$ is set to unit matrix $\mathbf{Q}_0 = I_4$ and $\mathbf{w}_i$ is supposedly distributed according to Gaussian distribution $\mathbf{w}_i \sim N(0, \mathbf{Q}_i)$. The $\mathbf{Q}_i$ is used as a noise source. Its choice can improve the performance of the filter. The tuning is usually performed with the help of another (distinct) Kalman filter. In our case, $\mathbf{Q}_i$ is a constant, but it can be changed during the filter operation to account for different dynamics in the system [40].

6.2.6 Backward run

After finishing the forward run, we initialize the variables for the reverse one by taking the last update of the process state matrix and reversing the signs of the shifts for each pixel. So when the follicle is moving to the right or shrinking, it will do the opposite: moving to the left or increasing in size. Such modified $\hat{\mathbf{x}}$ and lastly updated $\mathbf{P}^{-}$ or $\mathbf{P}^{-}_{m-1}$ will be the initialization. The algorithm behaves almost the same, except the iterator i goes from $m-2$ to 0. After this last part, the algorithm is finished.

Chapter 7

Deep Learning - UNet-based architectures

The following chapter describes the last and most advanced of the methods used for follicle recognition in this work. The initial approach was implementing a CR-Unet - a composite network described in [22]. During the implementation, we discovered that stripping the method of its spatial recurrent neural network (RNN) modules and returning to the original, simpler version - plain U-Net - still yielded great results on our dataset. Nevertheless, there were some differences in the predictions of the two methods.

7.1 Dataset and Augmentation

The original dataset has 1396 images, for each of which a true annotation mask was generated; for the train and test set, 80% (1128) and 20% (268) of the dataset were allocated, respectively. We generated up to 10 augmented counterparts for each training image and its mask using `torchvision` image augmentation transformations and their compositions. We used four different types of random augmentations - rotation ($-30°$ to $+30°$), horizontal and vertical flip - both with probability $p = 0.5$ of flipping - and colour jitter (changing brightness and contrast) with a parameter 0.5, meaning that the jitter level will be chosen randomly from a uniform distribution $[0.5, 1.5]$ for both brightness and contrast.

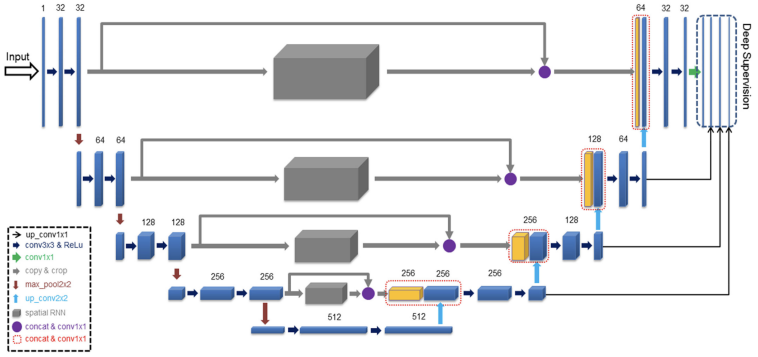


Figure 7.1: The illustration of the proposed pipeline for ovarian follicle segmentation in CR-UNet. The backbone is a standard U-Net, of which some customized spatial RNN modules are embedded between the encoder-decoder. When there are, in total, four spatial RNN modules, the proposed network is named CR-UNet. Numbers on each module indicate the number of channels [22].

7.2 CR-UNet

Haoming et al. propose the CR-Unet as a composite network incorporating the spatial recurrent neural network (RNN) into a plain U-Net. Their solution is supposed to learn multi-scale and long-range spatial contexts effectively [22]. The architecture of the neural network (NN) is illustrated in 7.1, and a spatial RNN module is depicted in 7.2. Unfortunately, the authors did not provide access to their source code and the exact setting of the NN. Therefore, our implementation is only based on the information in the research article. Moreover, they also segmented ovaries as well as follicles, so we altered their method to segment only follicles.

7.2.1 Experimental setup

As the first model, we implemented the CR-UNet with the base of PyTorch UNet[1] (with fewer channels, matching the CR-Unet paper). We added the RNN modules and upsampling in between the layers as depicted in 7.1. For the *loss function*, we followed the paper and implemented Logarithmic Dice (LD)

$$\mathrm{LD} = -\log\left(2\sum_{c=1}^{C}\sum_{i=0}^{N}\hat{y}_i^c y_i^c\right) + \log\left(\sum_{c=1}^{C}\sum_{i=0}^{N}\hat{y}_i^c + \sum_{c=1}^{C}\sum_{i=0}^{N} y_i^c\right), \tag{7.1}$$

[1] `https://github.com/milesial/Pytorch-UNet/`

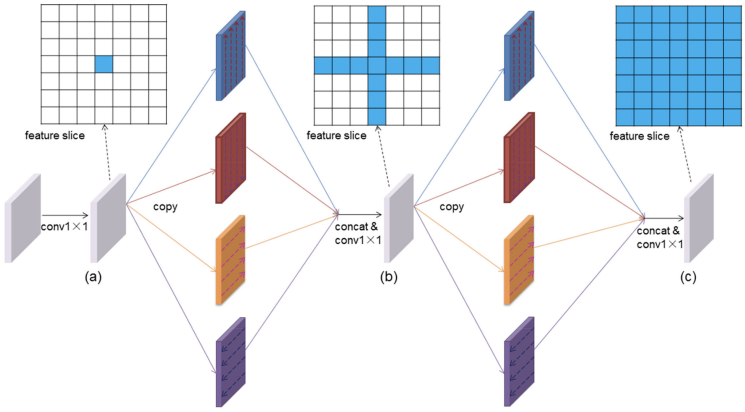


Figure 7.2: The Spatial RNN module. Each feature slice of a certain layer in the encoder is convoluted with 1×1 kernel to be taken as input of the spatial RNN, on which four directional operations (up, down, left and right) are implemented. Their outputs are concatenated and then convoluted with 1×1 kernel. The number of channels is, therefore, the same as the input of the RNN. The process is repeated once to generate final feature map, as a result each pixel integrates the global spatial information [22].

where C is the number of classes, N is the number of samples, $\hat{y}_i^c$ denotes the prediction for a pixel belonging to a class c and y_i^c is the true label of the pixel. The paper had three classes - background, ovary and follicle - but we found it sufficient and straightforward to have only one class for our task (no ovary recognition).

For loss computation, Haoming et al. used the deep supervision technique [43] to improve the gradient flow through the network. Outputs of each layer in the right part of the UNet were up-sampled to the same size as the final model output and aggregated by a weighted sum to create the final loss. The loss weights were set the same as in the paper [22] - from top to bottom as $[1.0, 0.8, 0.4, 0.1]$. After running the first few experiments, the model favoured labelling bigger blobs as follicles instead of dissecting them into separate ovals. We identified this might be happening due to the imbalance in the extent of background and follicles. We added binary cross-entropy (BCE) 7.2 from PyTorch [31] to address some of these limitations.

$$\text{BCE}(x, y) = \{l_1, \cdots, l_N\}^T,\ l_n = y_n \log x_n + (1 - y_n) \log(1 - x_n) \qquad (7.2)$$

where x and y and input and target, respectively, and N is the batch size. We tested several *batch sizes* - $1, 4, 8,$ and16 - with no discernible influence on the outcome. The *learning rate* was set to 1×10^{-5} and later decreased to 1×10^{-6} as we sometimes encountered problems with gradient explosion. Depending on the type of training method, 3% or 10% of the train set was used for validation. The train and validation losses convergence plots can be seen in 7.3.

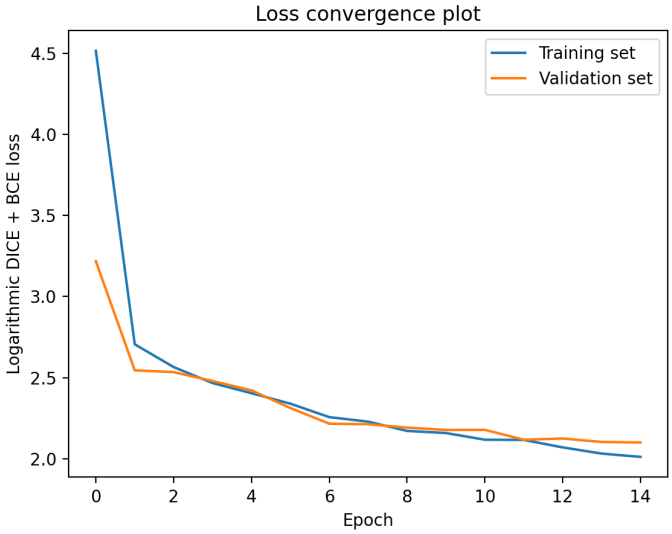


Figure 7.3: Loss convergence plot for the original CR-UNet model with 3387 images, including augmentations, validation was set to 10% of the train set.

As the second model, we implemented the plain PyTorch UNet and also changed the *loss function* only to encompass the loss from the last layer output (with weight of 1) and kept the BCE addition. The rest of the settings remained the same 7.4, but to boost the performance, we also tried randomly sampling 2000 from a set of 12408 images for each training epoch 7.5. Both plots show that the models' training and validation losses are nicely converging but not over-fitting.

7.2.2 Resource and time specification

The models were trained on a MacBook Air M2 with Apple Silicon chip, and the average speed was around 1.2s/img, so it took around one hour to train an epoch or 15 to 20 hours for our specified number of epochs.

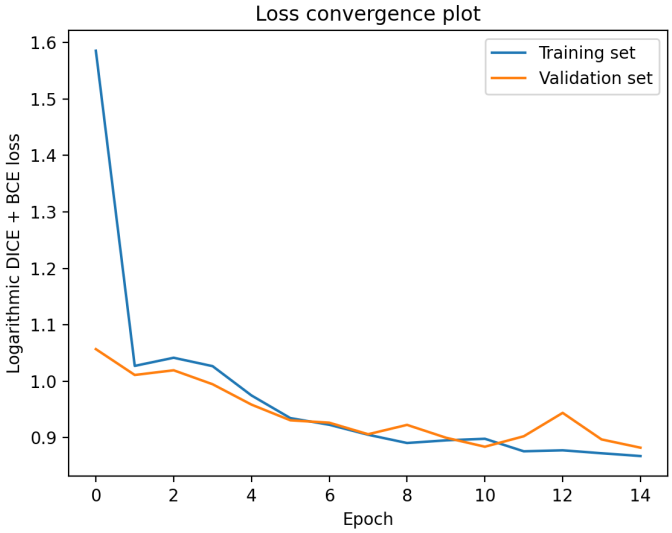


Figure 7.4: Loss convergence plot for the UNet model with 3387 images, including augmentations. The validation set size was 10% of the train set.

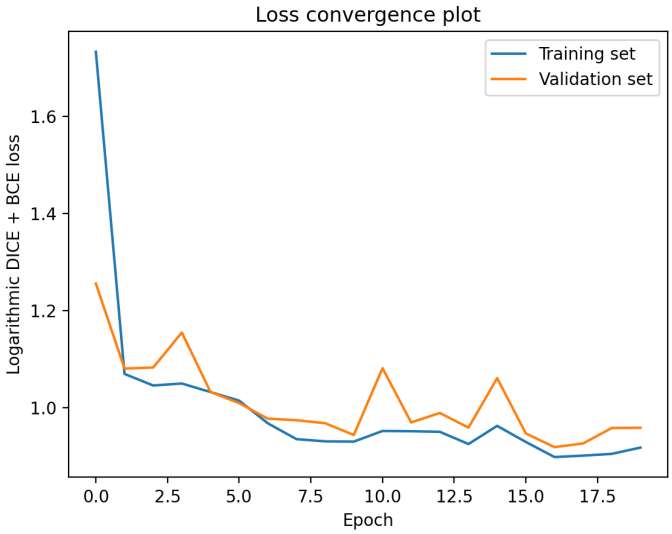


Figure 7.5: Loss convergence plot for the UNet model with a random sampling of 2000 from a set of 12408 images (including augmentations) for each epoch. The validation set size was 3% of the train set

Chapter 8

Performance of Algorithmic Methods

The following chapter will be dedicated to the evaluation of the RG method described in 5 and KF method as an extension of this method described in 6. The first part will show how our modification improved the RG method to perform better for our data, and then we will report on the performance of both methods on the second part of the data - the video dataset. Finally, we will compare the methods and discuss their performance differences.

8.1 Images Dataset Predictions Evaluation

We ran the RG algorithm for all of the obtained images, even though the quality of some of them was deficient, and the doctor was not entirely sure whether there was a follicle in the image. The comparison between the results yielded by the original algorithm and the modified we produce can be seen in 8.2 - modified algorithm vs. 8.3 - original algorithm and 8.4 vs. 8.5. It can be observed how the Hom_2 detects many regions which are not follicles. This confuses region growing and leads to worse results.

After examining all the images annotated by the algorithm, we concluded that four cases happened and clustered the images accordingly. The groups were created based on two criteria: how well the algorithm performed on the image and the quality of the image. The first group contained images on which the algorithm performed well, and their quality was also good 8.2, 8.4. In the second group are images on which the region growing performed poorly,

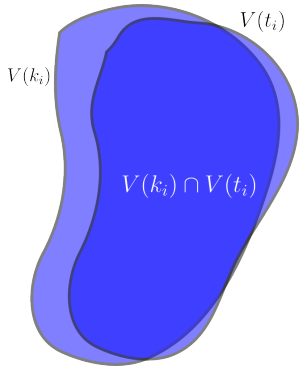


Figure 8.1: Illustration of r_1 and r_2 metrics

but their quality was good, so the result is still acceptable 8.6, 8.7. The third group consists of images where the homogeneous region identification part of the algorithm identified the regions well. Still, the region growing part resulted in some misidentifications, so the result is of poor quality 8.8, 8.9. The last group comprises low-quality images, which also yielded poor results 8.10, 8.11.

We also calculated two ratios for each image - r_1 and r_2. r_1 is the ratio between the area of the intersection of the detected follicles and actual follicles and the area of the actual follicles

$$r_1 = \frac{V(k_i) \cap V(t_i)}{V(t_i)}, \tag{8.1}$$

where V notes the ares of the curve. r_2 is the ratio between the area of the intersection and the area of the detected follicles

$$r_2 = \frac{V(k_i) \cap V(t_i)}{V(k_i)}. \tag{8.2}$$

An illustration for those metrics is provided in 8.1.

Due to the varying nature of the image quality, we provide results regarding those four groups. In Table 8.1 columns "50%", "75%", "90%" denote the percentage of images from the group where the algorithm recognized at least 50%, 75% and 90% of follicles, respectively. Note that we compare only the numbers in this part, so the reader has to evaluate this result together with the rations. When the percentage is high, and the ratios are low, not only does the algorithm not find the actual follicles, but it also incorrectly marks other dark structures as follicles.

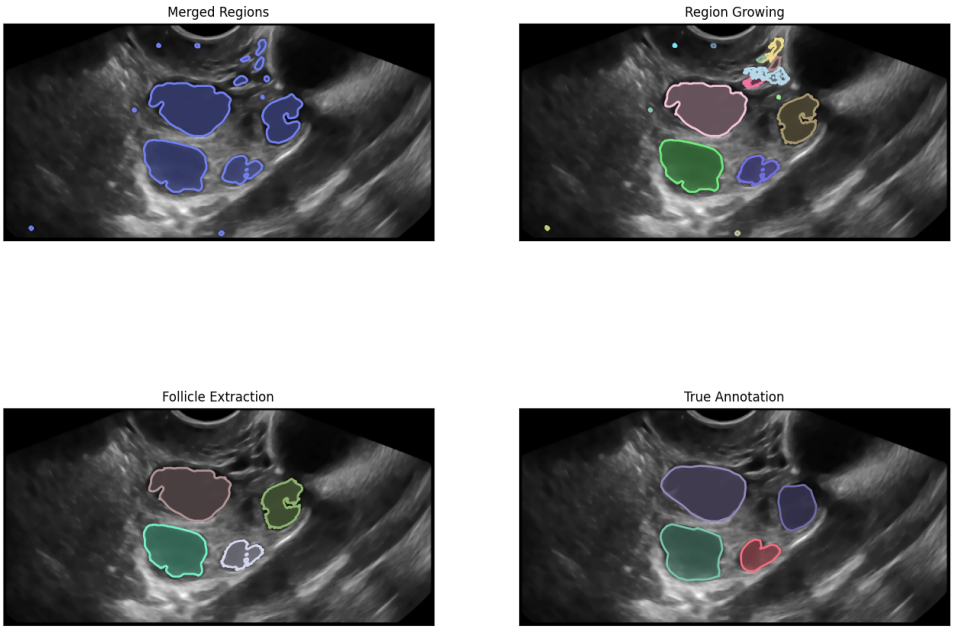


Figure 8.2: Annotation of an image 413 from group one. $r_1 = 0.87$, $r_2 = 0.96$

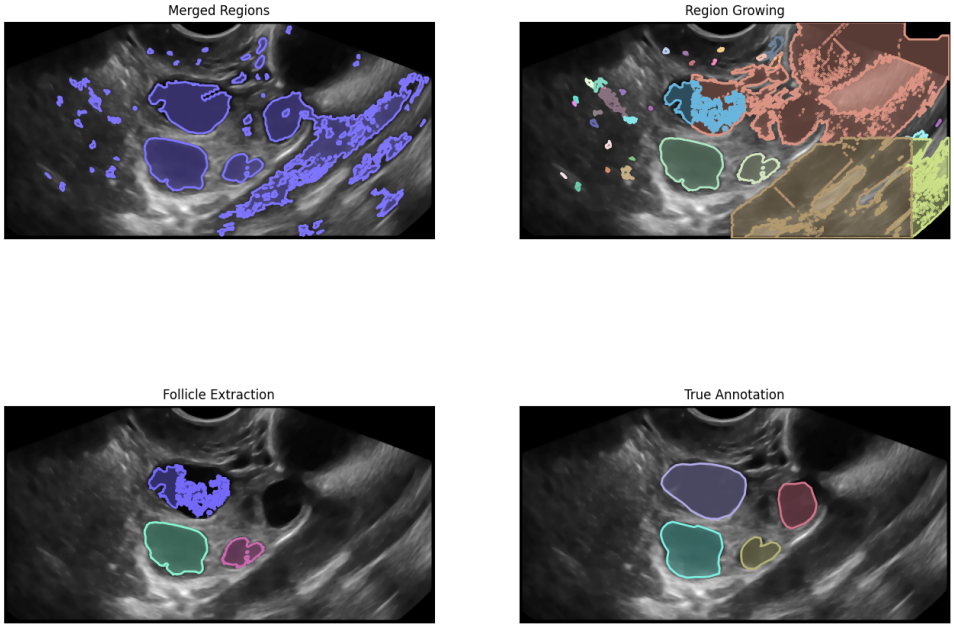


Figure 8.3: Annotation of an image 413 using original method.

group	images	avg r_1	avg r_2	50%	75%	90%
1	20	0.731	0.816	80%	75%	75%
2	18	0.604	0.833	61%	33%	28%
3	14	0.073	0.285	39%	39%	39%
4	36	0.072	0.176	47%	25%	22%

Table 8.1: Results divided into the four groups

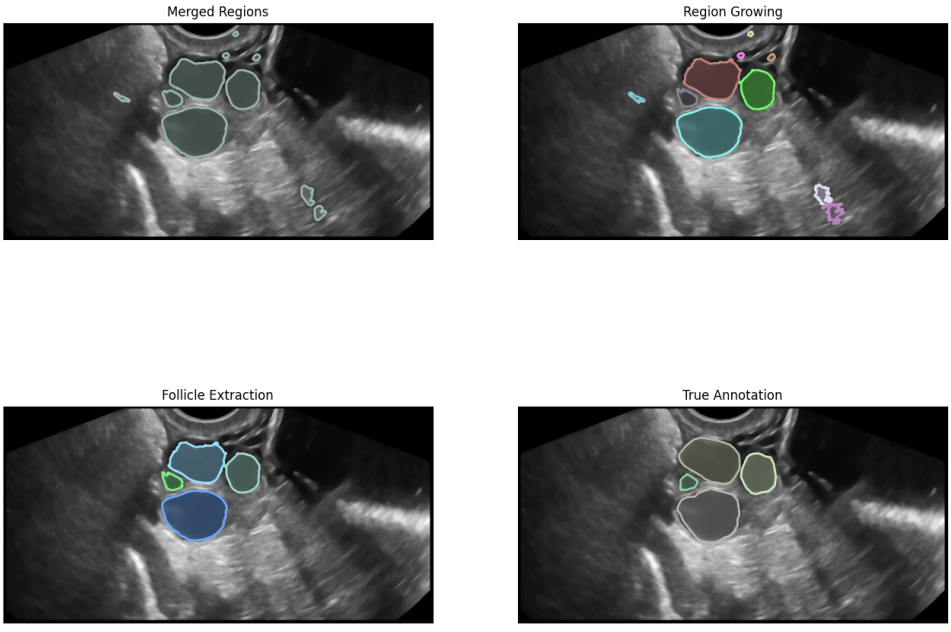


Figure 8.4: Annotation of an image 414 from group one. $r_1 = 0.90$, $r_2 = 0.96$

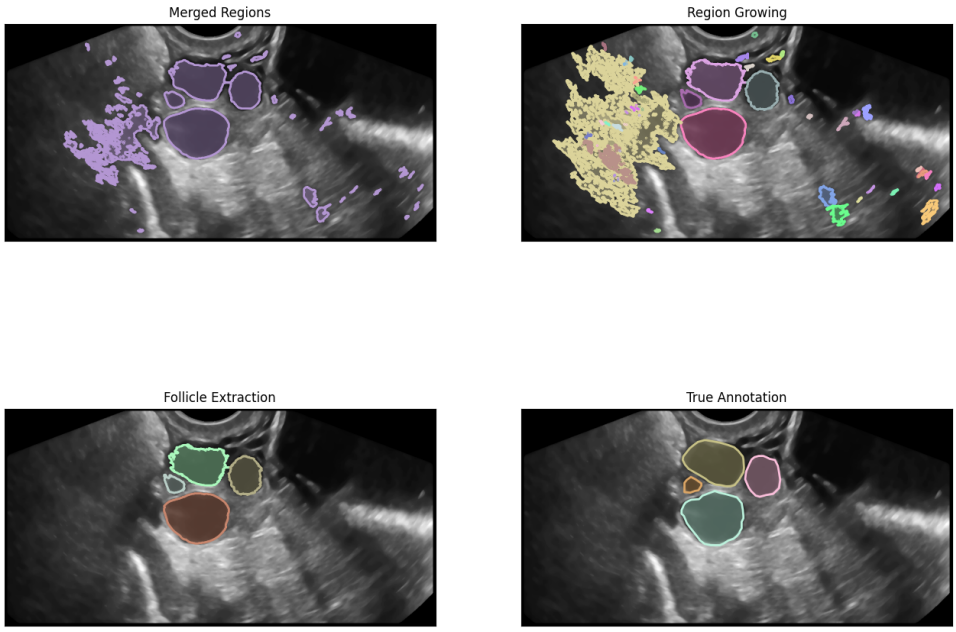


Figure 8.5: Annotation of an image 414 using original method.

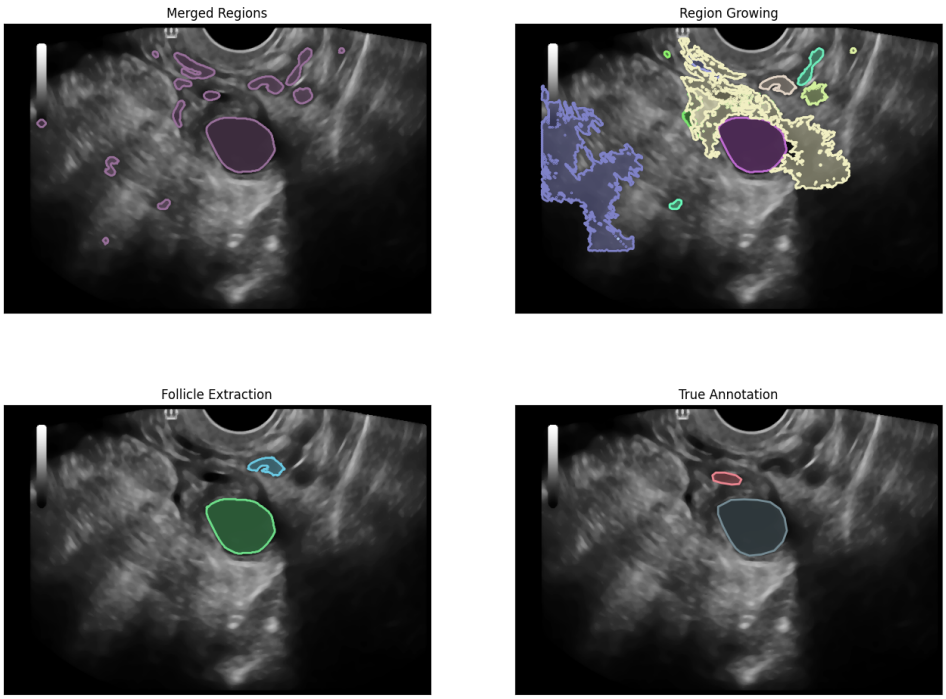


Figure 8.6: Annotation of an image 253 from group two. $r_1 = 0.87$, $r_2 = 0.88$

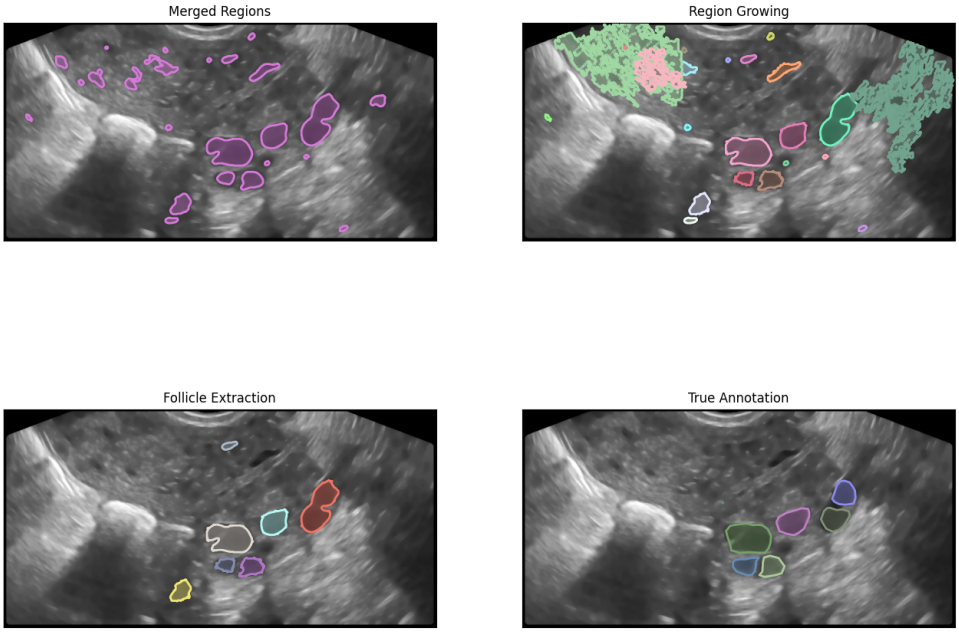


Figure 8.7: Annotation of an image 553 from group two. $r_1 = 0.84$, $r_2 = 0.83$

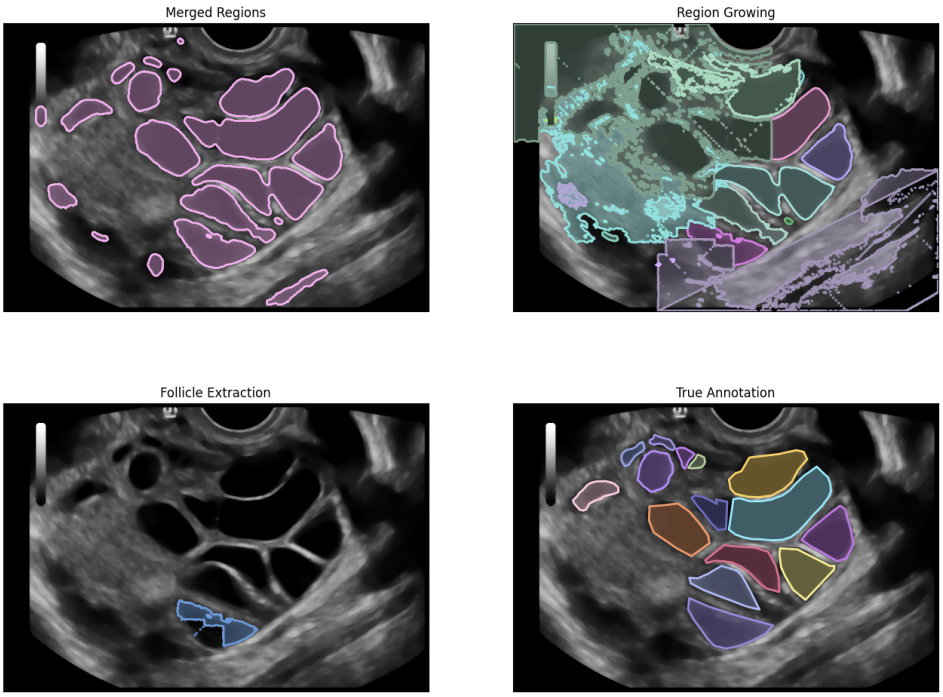


Figure 8.8: Annotation of an image 270 from group three. $r_1 = 0.06$, $r_2 = 0.98$

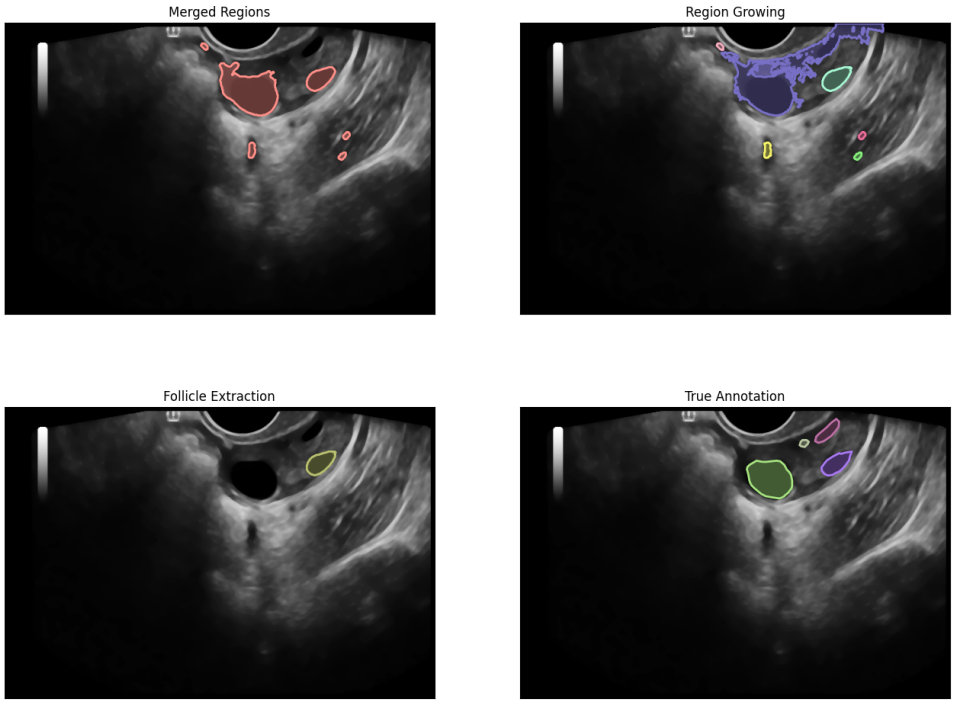


Figure 8.9: Annotation of an image 373 from group three. $r_1 = 0.2$, $r_2 = 0.9$

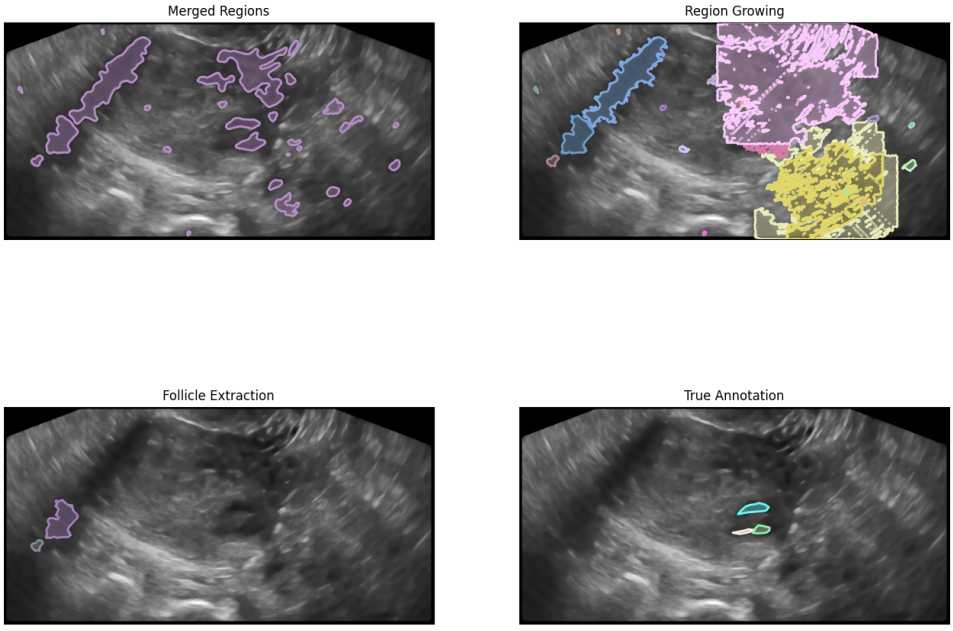


Figure 8.10: Annotation of an image 584 from group four. $r_1 = 0$, $r_2 = 0$

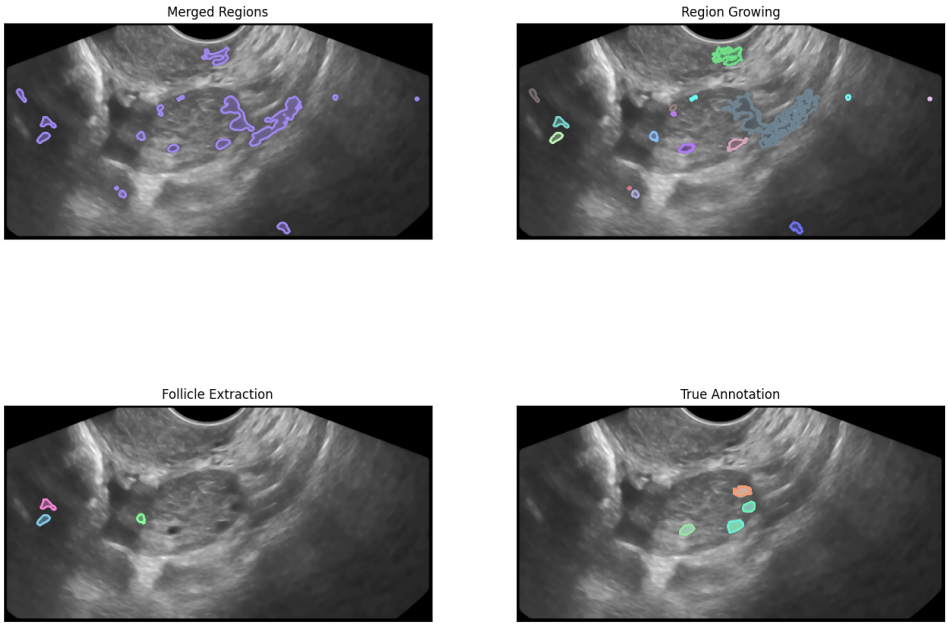


Figure 8.11: Annotation of an image 590 from group four. $r_1 = 0$, $r_2 = 0$

8.1.1 Conclusion

It can be seen in Table 8.1 that if the image was of good quality, we found over 90% of follicles in 75% of images. If the quality was at least moderate (groups one and two), we could identify 75% of follicles in a third of images. Those results suggest that we managed to recreate the algorithm satisfactorily with slight modifications. Although the dataset we acquired was not ideal, we sorted through the images and reported performance, which was acceptable for group one images.

8.2 Results on Videos Dataset

This section will provide results of both the RG and KF methods on the sequences of images, which were the result of automatic cuts of the videos into a specified number of FPS.

8.2.1 Region growing efficiency

In the previous chapter, we divided the images into four groups based on the image quality and the algorithm's performance. This was possible due to a relatively smaller number of images in the dataset, but the same tedious approach is not feasible for the video dataset. Therefore, we decided to split the images into groups based on their recognition rates - r_1 and r_2. The thresholds were chosen so the group performances are similar to the previous distribution.

Table 8.2 shows the different groups of images and their performance. The criteria for groups 1 to 4 $G_j, j = 1, 2, 3$ were

$$G_j = \{S_i \in G_j \mid r_1(S_i) \geq T_j \wedge r_2(S_i) \geq T_j\}. \tag{8.3}$$

The thresholds were $T_1 = 0.7$, $T_2 = 0.5$ and $T_3 = 0.3$. Group 4 G_4 comprises the rest of the images. A similar approach for diving into four groups was taken for the videos as well, the only difference being that the r_1 and r_2 are averages over all images in the video.

group	images	avg r_1	avg r_2	50%	75%	90%
1	550	0.842	0.826	64%	42%	27%
2	373	0.609	0.890	79%	43%	23%
3	214	0.404	0.832	64%	25%	15%
4	259	0.098	0.360	35%	15%	10%

Table 8.2: Results of RG on individual images divided into four groups based on the performance.

The video results, composed of the individual images, are shown in table 8.3.

Almost 40% of images were almost perfectly classified as both the recognition rates were over 80%. Over 65% of images had r_1 over 50% while maintaining still high r_2 - on average over 85%. This means that the algorithm recognized at least 50% of follicle area and did not misidentify more than 15% of the area. Differences in the number of actual and predicted follicles are demonstrated in 8.12. We already established that the RG method performs satisfactorily on better-quality images, and those results also support this conclusion.

group	videos	avg r_1	avg r_2	50%	75%	90%
1	14	0.792	0.804	66%	42%	26%
2	43	0.601	0.819	68%	37%	21%
3	23	0.412	0.626	54%	27%	18%
4	2	0.161	0.514	17%	6%	0%

Table 8.3: Results of RG on video dataset divided into four groups based on the performance.

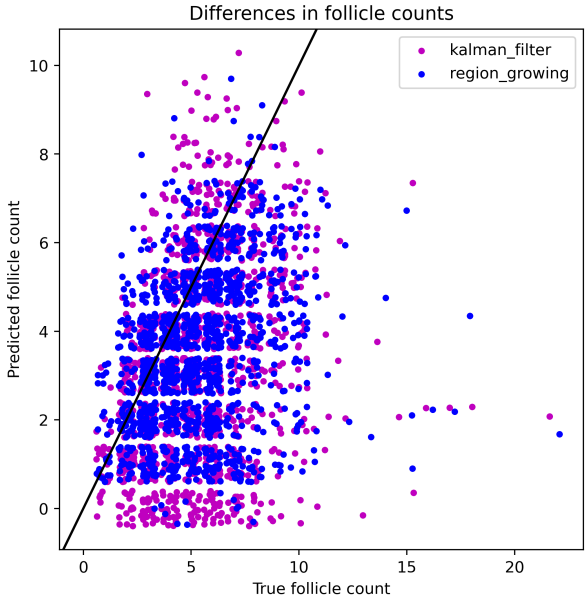


Figure 8.12: Differences in true and predicted follicle count for Kalman filter and region growing methods.

8.2.2 Kalman filter addition evaluation

As this method should serve as an enhancement of RG, the results are presented in a comparative way to highlight the contribution of KF. We compared the recognition rates r_1 and r_2 for each image in each video and calculated different statistical indicators to illustrate the results. The comprehensive results are in Table 8.4. We evaluated r_1 and r_2 of each image based on whether it was better or worse with Kalman (in contrast to only region growing), hence the naming in Table 8.4. We then calculated statistical indicators - total count, average, minimum, maximum and median - from sets of images belonging to a single video. We further aggregated those results and calculated and reported the average and median from all of the videos. We can see from the table that a better r_2 in images annotated with Kalman was on average in 8.22 images per video, and the average and median of an

			count	min	max	avg	median
r_1	better	avg	1.74	0.005	0.14	0.009	0.009
		median	1	0.001	0.004	0.003	0.003
	worse	avg	13.7	0.009	0.573	0.182	0.123
		median	12.5	0.005	0.552	0.154	0.083
r_2	better	avg	8.22	0.006	0.218	0.066	0.038
		median	7	0.002	0.159	0.049	0.025
	worse	avg	6.86	0.045	0.474	0.186	0.146
		median	6	0.003	0.523	0.11	0.032

Table 8.4: Statistics regarding the comparison of r_1 and r_2 rates in region growing and Kalman filter.

average increase in r_2 was 0.066 and 0.049, respectively.

The number of images where RG gives better r_1 than KF was higher for all videos. However, the r_2 was better for over 62% of videos; in almost 5% of videos, the majority of images showed no change.

To better understand and make sense of the somewhat ambiguous results, we will illustrate the most common mistakes KF makes and explain how those mistakes impact the recognition rates. The first case is when RG already performs quite well on the image, and therefore, there is not a lot of space for improvement to be done by KF. In those cases, we can observe marginal increase or decrease in either r_1 - images 8.13 and 8.14 or r_2 - images 8.15 and 8.16.

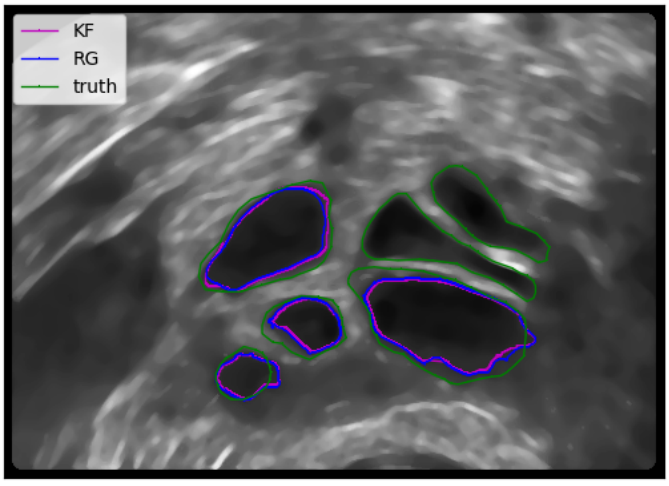


Figure 8.13: Marginal decline in $r1$ when using Kalman filter. $r_1(\text{RG}) = 0.5718$, $r_1(\text{KF}) = 0.5681$, difference of 0.0037.

The second case is when RG misidentifies a follicle. We have taken some

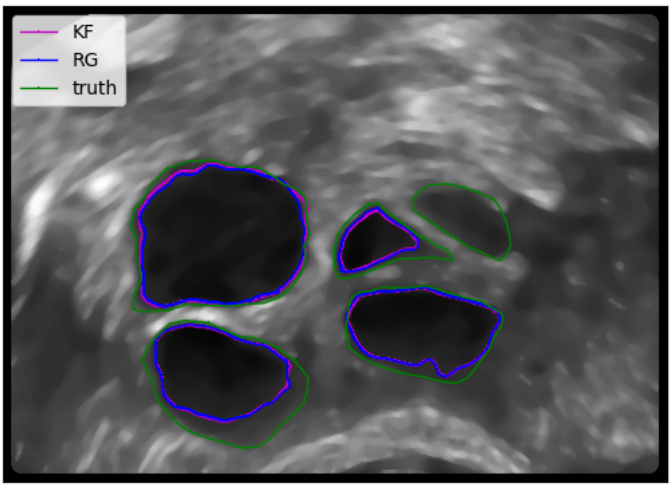


Figure 8.14: Marginal increase in $r1$ when using Kalman filter. $r_1(\mathrm{RG}) = 0.7301$, $r_1(\mathrm{KF}) = 0.7443$ difference of 0.0142.

precautions to check the correctness in 6.2.3. If the exclusion is correct, it will result in a much higher r_2 score 8.17. Nevertheless, if misidentification happens over a region in true follicle annotations, the result is a significant decrease in r_1 8.18.

However, to be sure we do not exclude a valid follicle, the exclusion criteria we introduced were relatively mild. The result is that sometimes a follicle is not excluded, even though it is not valid 8.19. If no other cases are present, the recognition rates will only have marginal differences again.

The third case is RG correctly identifying a follicle in the ith image but omitting it in both images $i-1$ and $i+1$. This will cause KF to skip the follicle in image i and result in a considerable decrease in r_1. This instance is illustrated in 8.20 on the bottom right follicle. If a follicle is recognized in $i+2$ again, we could avert this by interpolating the follicle in images $i-1$ and $i+1$. Nevertheless, this could create a new mistake when the follicle in $i+2$ is, in fact, a different one than the one in i.

8.2.3 Conclusion

In conclusion, we have shown how, in certain situations, the Kalman filter might advance the region growing method, but also how it is hugely reliant on the measurements from it. Further improvements might comprise more cooperation between the methods, as suggested in the third case of the previous section. A potential increase in the performance of RG on image

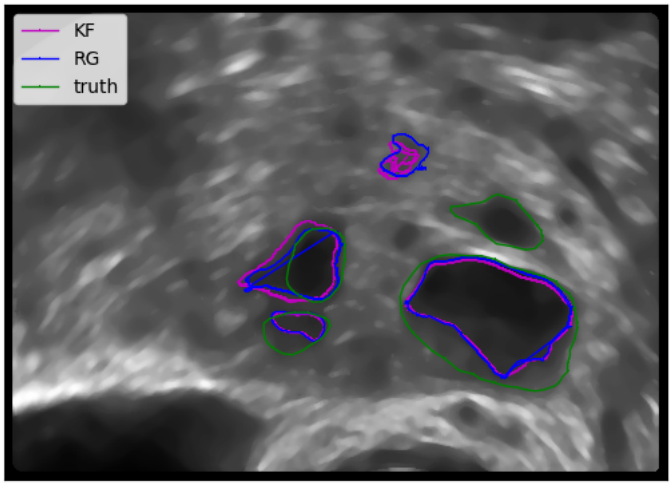


Figure 8.15: Marginal decline in $r2$ when using Kalman filter. $r_2(\mathrm{RG}) = 0.8654$, $r_2(\mathrm{KF}) = 0.8402$, difference of 0.0253.

sequences could be gained by initiating the RG with the previous follicle segmentation and slightly adjusting for the changes.

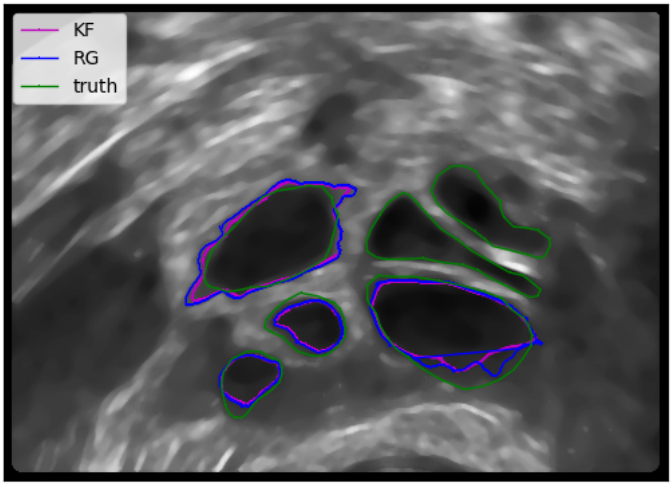


Figure 8.16: Marginal increase in $r2$ when using Kalman filter. $r_2(\mathrm{RG}) = 0.8867$, $r_2(\mathrm{KF}) = 0.9210$, difference of 0.0344.

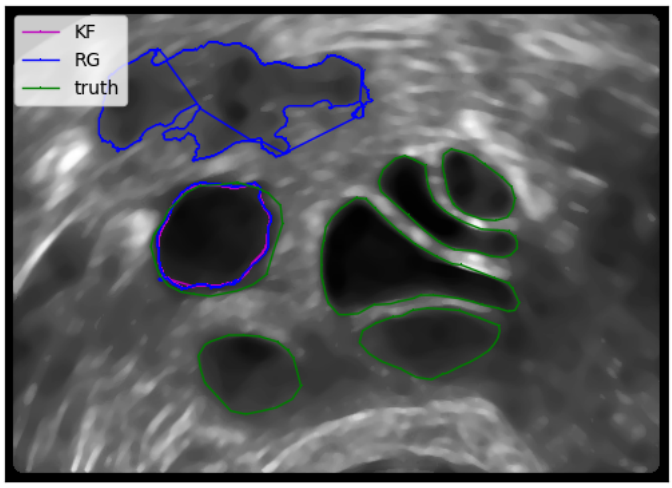


Figure 8.17: Substantial increase in $r2$ when using Kalman filter. $r_2(\mathrm{RG}) = 0.3257$, $r_2(\mathrm{KF}) = 0.9862$, difference of 0.6605.

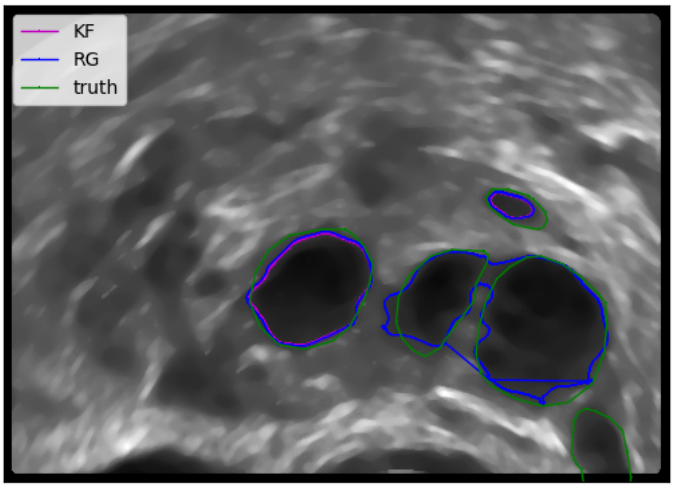


Figure 8.18: Substantial increase in $r2$ when using Kalman filter. $r_2(\mathrm{RG}) = 0.3257$, $r_2(\mathrm{KF}) = 0.9862$, difference of 0.6605.

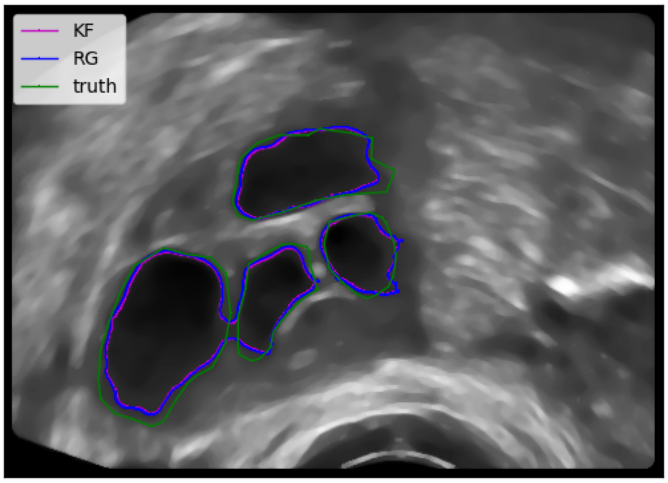


Figure 8.19: Misidentified follicle (two follicles merged) by RG, which is (wrongly) not excluded by KF.

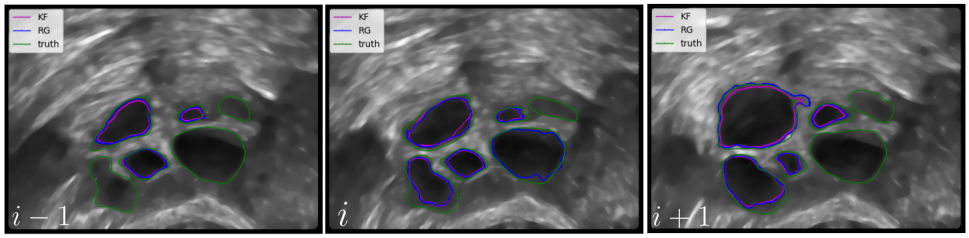


Figure 8.20: Follicle in the right bottom corner is skippingly recognized by RG. Results in a substantial decrease in r_1 in the ith image. $r_1(\mathrm{RG}) = 0.7267$, $r_2(\mathrm{KF}) = 0.4396$, difference of 0.2872.

Chapter 9

Results of Deep Learning Methods

The following chapter is dedicated to evaluating the deep learning methods for our task of follicle segmentation. We will evaluate three models - CR-UNet as done by Haoming et al., trained on 3387 images (including augmentations) and then two UNet models - one trained the same way as CR-UNet and the second one trained with randomly sampling a subset (2000 images) from a more extensive train set (12408 images) each epoch.

9.1 Evaluation metrics

Additionally to the recognition rates used in the evaluation of the algorithmic method, we used the Dice Similarity Coefficient (DSC) 9.1 for the model evaluation. Similar to recognition rates, it measures the overlap between two binary masks - the predicted and truth segmentation. A score of 1 indicates a perfect match, while a 0 signifies no intersection between the segmented follicles.

$$\mathrm{DSC} = \frac{2\sum_{i=0}^{N} \hat{y}_i y_i}{\sum_{i=0}^{N} y_i + \sum_{i=0}^{N} \hat{y}_i} \tag{9.1}$$

metric	model	train method	avg	min	max	median
DSC	CR-UNet	whole train set	0.834	0	0.952	0.865
	UNet	whole train set	0.853	0	0.966	0.890
	UNet	random sampler	0.813	0	0.957	0.857
r_1	CR-UNet	whole train set	0.835	0	0.993	0.875
	UNet	whole train set	0.864	0	0.995	0.906
	UNet	random sampler	0.821	0	0.998	0.855
r_2	CR-UNet	whole train set	0.851	0	1	0.879
	UNet	whole train set	0.866	0	1	0.911
	UNet	random sampler	0.839	0	0.999	0.891

Table 9.1: Performance of different deep learning models on the test dataset (no augmentation). The whole train set method uses the same train set for every epoch (3387 images), while a random sampler is randomly sampling a subset (2000 images) from a larger train set (12408 images).

9.2 Models Testing Outcomes

The statistical indicators regarding the models' performance are illustrated in 9.1. The superiority of deep learning over algorithmic methods in "darker regions" recognition is apparent despite the fact that the test images are not divided into groups based on image quality. See RG results on images in Table 8.2 or videos in Table 8.3 for comparison.

To illustrate the differences and similarities in the model performances we provide several image examples 9.2, 9.3, 9.4, 9.5, 9.6, 9.7. The results of those images are summarized in Table 9.2.

9.2.1 Follicle counts

However great the recognition rates and DSC were, frequently, the prediction (mostly plain UNets) included additional small regions, which drove up the number of follicles in images 9.2. The predicted number of follicles was higher than the truth for the UNet-whole and UNet-random models in 45% and 69% of images. On average, the increase was 2.5 and 3.5 follicles. In 27% and 18% of images where the difference was reversed (more true follicles than predicted), the average of this difference was 2.1 for both. Therefore, the model got right the exact number of follicles in 28% and 13% of images.

On the other hand, the mistake most common for the CR-UNet model

image	model	DSC	r_1	r_2
Fig. 9.2	CR-UNet	0.0.890	0.864	0.919
	UNet-whole	0.911	0.897	0.925
	UNet-random	0.829	0.750	0.927
Fig. 9.3	CR-UNet	0.857	0.888	0.828
	UNet-whole	0.914	0.917	0.911
	UNet-random	0.901	0.874	0.929
Fig. 9.4	CR-UNet	0.938	0.952	0.924
	UNet-whole	0.949	0.965	0.933
	UNet-random	0.949	0.924	0.976
Fig. 9.5	CR-UNet	0.904	0.875	0.935
	UNet-whole	0.947	0.918	0.978
	UNet-random	0.924	0.872	0.982
Fig. 9.6	CR-UNet	0.944	0.905	0.987
	UNet-whole	0.933	0.915	0.951
	UNet-random	0.912	0.860	0.971
Fig. 9.7	CR-UNet	0.906	0.929	0.883
	UNet-whole	0.899	0.971	0.837
	UNet-random	0.915	0.900	0.931

Table 9.2: Performance of deep learning models on specific images with illustrations of different error types.

was merging closely located regions, which resulted in less than the true number of follicles in 77% of images and only a higher number of follicles in 7%—leaving 16% of images with the correct count of follicles. The average decrease and increase were 2.5 and 1.4 follicles, respectively. An illustration of this merge is depicted in 9.3. A scatter plot demonstrating the differences in follicle counts can be seen at 9.1.

Those differences are not outrageous since the physicians themselves differ in their counts in ambiguous images. Nonetheless, those inaccuracies could hold greater value in borderline cases (very few or many follicles).

A well-chosen set of heuristics for follicle exclusion would at least partially fix the issue with many small misidentified regions. Such heuristics could be the ones used at the end of RG algorithm 5.4, that is, removing regions at the border of the image, regions smaller or bigger than some threshold and regions which are not correctly proportionate (not oval like). To lower the incentive of the CR-UNet model to segment larger regions, we could add more weight to the one type of mistake - confusing background for a follicle.

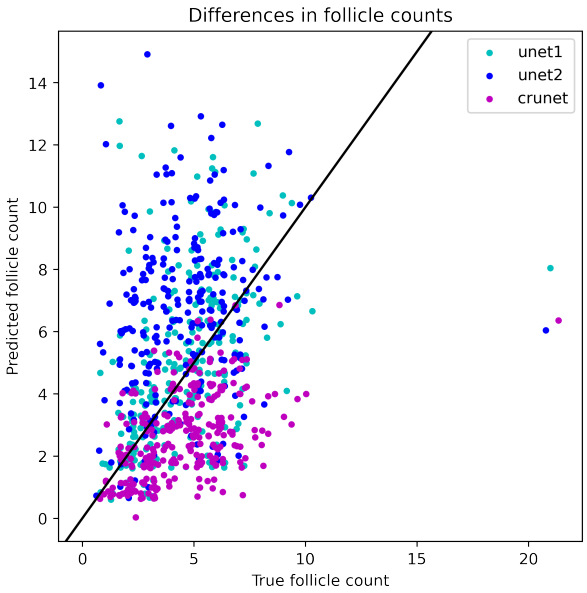


Figure 9.1: Differences in true and predicted follicle counts for deep learning models.

9.2.2 Stray lines and Roughness

There are sometimes lines across the follicle or some other disorganized set of lines in the performance illustration images. Those are the results of the unevenness of the recognized follicle. The predictions of the UNets usually have a rough border; on the contrary, CR-Unet predictions usually have very smooth edges 9.4. Another reason for the lines is the occasional hole inside the region, which was the case for all of the models 9.3. Such behaviour is expected due to the challenges of ultrasound images 4.1.

A suitable heuristic applied to the predictions could again correct those inaccuracies - holes as well as roughness - to some degree.

9.3 Conclusion

The deep learning models undeniably outperform the algorithmic methods regarding "darker regions" recognition. Although they still have some flaws, which could be minimized or even mitigated with appropriate heuristics,

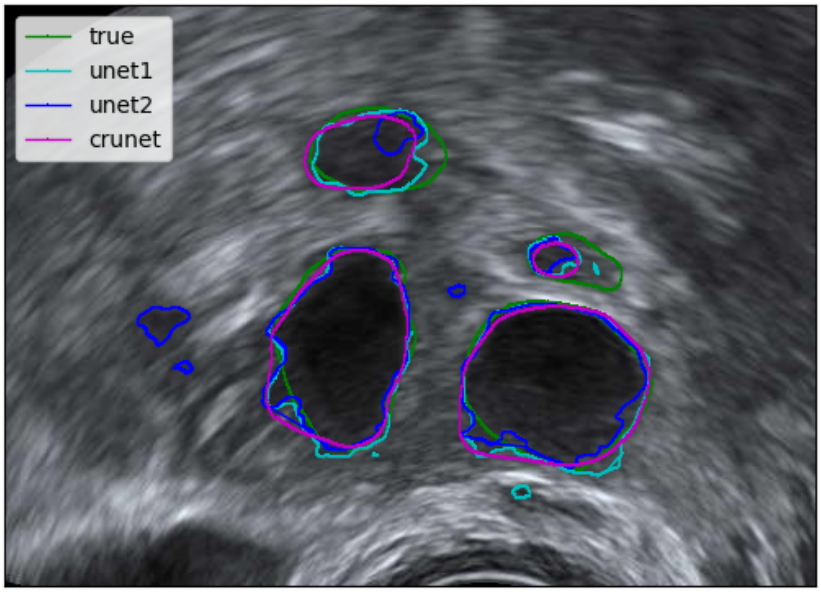


Figure 9.2: Illustration of multiple small misidentified follicles by UNet model. Results are in Table 9.2

they reach almost state-of-the-art recognition rates for 2D ultrasound images. Considering no image filtering based on quality was done, those results are excellent. After adding production finishing touches, the models could assist the physician (with proper training).

Figure 9.3: Illustration of a hole in a region and merging closely located regions by CR-UNet model. Results are in Table 9.2

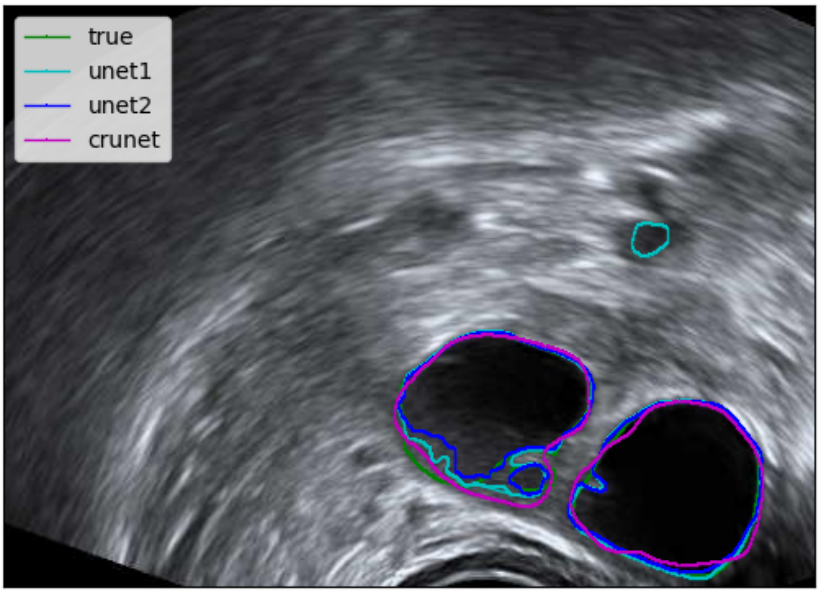


Figure 9.4: Illustration of border roughness in UNet predictions and relative smoothness in CR-UNet predictions. Results are in Table 9.2

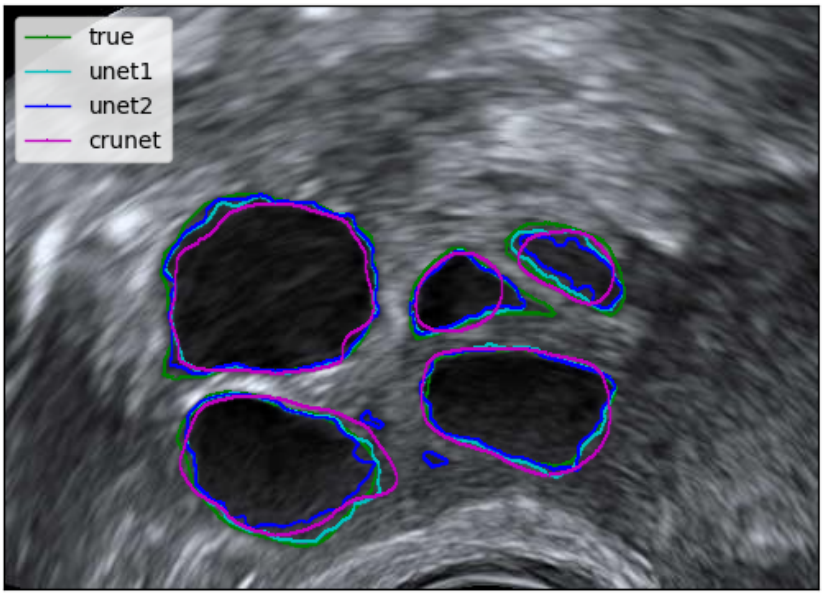


Figure 9.5: Illustration of an accordance of all the models. Results are in Table 9.2

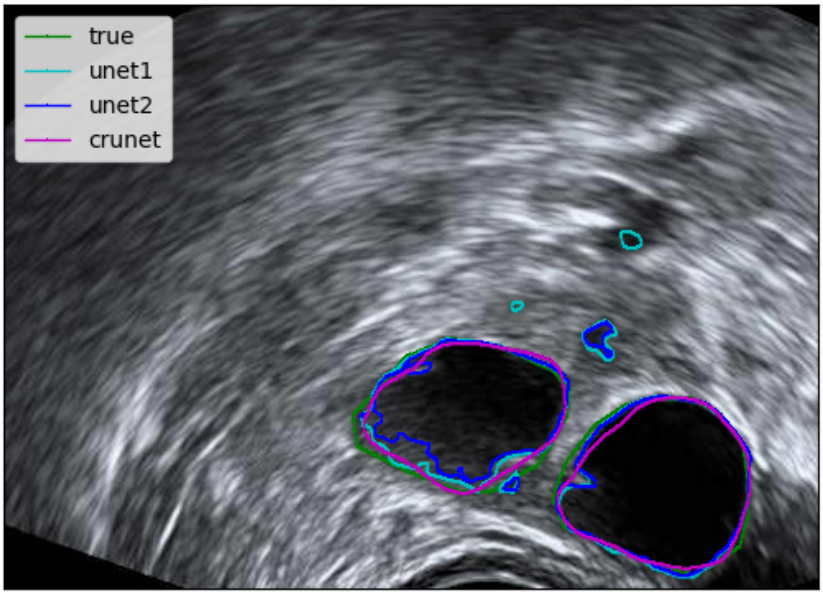


Figure 9.6: Illustration of an accordance of all the models. Results are in Table 9.2

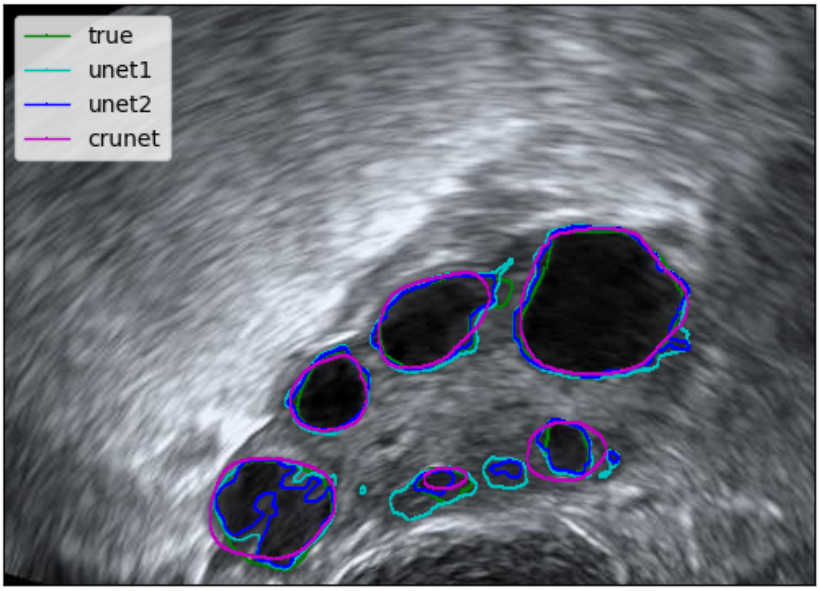


Figure 9.7: Illustration of an accordance of all the models. Results are in Table 9.2

Chapter 10

Conclusion

In summary, we have successfully implemented and evaluated three computer vision methods — region growing, Kalman filter, and U-Net-based deep learning models — for follicle recognition in 2D ultrasound images and videos. The algorithmic methods achieved average recognition rates between 0.6 and 0.8 for high-quality images. Following data augmentation, the deep learning models' predictions averaged over 0.83 across the entire test dataset. These findings highlight the potential of deep learning to improve, automate and streamline follicle segmentation in 2D ultrasound images and suggest its promising applicability in clinical practice.

10.1 Future Work

As the results of deep learning methods were preferable to the algorithmic ones, future works should take this line of research further. Starting with the suggested heuristics for follicle exclusion or implementing the additional ovaries recognition (but this would need new annotations). Using 3D-US images or videos could also increase the recognition performance, but as we mentioned in the intro of this work - the machines are not as widespread as 2D ultrasounds.

Bibliography

[1] Rolf Adams and Leanne Bischof. "Seeded region growing". In: *IEEE Transactions on pattern analysis and machine intelligence* 16.6 (1994), pp. 641–647.

[2] Hu Cao et al. "Swin-unet: Unet-like pure transformer for medical image segmentation". In: *European conference on computer vision*. Springer, 2022, pp. 205–218.

[3] Terrence Chen et al. "Automatic ovarian follicle quantification from 3D ultrasound data using global/local context with database guided segmentation". In: *2009 IEEE 12th International Conference on Computer Vision*. 2009, pp. 795–802. DOI: 10.1109/ICCV.2009.5459243.

[4] Zhiyi Chen et al. "Artificial Intelligence in the Assessment of Female Reproductive Function Using Ultrasound: A Review". In: *Journal of Ultrasound in Medicine* 41.6 (2022), pp. 1343–1353. DOI: 10.1002/jum.15827.

[5] Boris Cigale, Mitja Lenič, and Damjan Zazula. "Segmentation of ovarian ultrasound images using cellular neural networks trained by support vector machines". In: *Knowledge-Based Intelligent Information and Engineering Systems* (2006), pp. 515–522. DOI: 10.1007/11893011_66.

[6] Yinhui Deng, Yuanyuan Wang, and Ping Chen. "Automated detection of Polycystic Ovary Syndrome from ultrasound images". In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008, pp. 4772–4775. DOI: 10.1109/IEMBS.2008.4650280.

[7] Michael Fanton et al. “An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation”. In: *Fertility and Sterility* 118.1 (2022), pp. 101–108. DOI: 10.1016/j.fertnstert.2022.04.003.

[8] Katerina Fronckova and Antonin Slaby. “Kalman filter employment in image processing”. In: *Computational Science and Its Applications–ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part I 20*. Springer. 2020, pp. 833–844.

[9] Swarnendu Ghosh et al. “Understanding deep learning techniques for image segmentation”. In: *ACM computing surveys (CSUR)* 52.4 (2019), pp. 1–35.

[10] C. Gopalakrishnan and M. Iyapparaja. “Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning”. In: *International Journal of System Assurance Engineering and Management* (2021). DOI: 10.1007/s13198-021-01203-x.

[11] Maribel Grande et al. “Antral follicle count as a marker of ovarian biological age to reflect the background risk of fetal aneuploidy”. In: *Human Reproduction* 29.6 (2014), pp. 1337–1343. DOI: 10.1093/humrep/deu055.

[12] ESHRE Reproductive Endocrinology Guideline Group. *Ovarian Stimulation for IVF/ICSI. Guideline of the European Society of Human Reproduction and Embryology*. ESHRE. URL: https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Ovarian-Stimulation-in-IVF-ICSI.

[13] SA Hojjatoleslami and Josef Kittler. “Region growing: a new approach”. In: *IEEE Transactions on Image processing* 7.7 (1998), pp. 1079–1084.

[14] Huimin Huang et al. “Unet 3+: A full-scale connected unet for medical image segmentation”. In: *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2020, pp. 1055–1059.

[15] Rudolph Emil Kalman. “A new approach to linear filtering and prediction problems”. In: (1960).

[16] Wen-Xiong Kang, Qing-Qiang Yang, and Run-Peng Liang. “The comparative research on image segmentation algorithms”. In: *2009 First international workshop on education technology and computer science*. Vol. 2. IEEE, 2009, pp. 703–707.

[17] Patricia Katz et al. “Costs of infertility treatment: results from an 18-month prospective cohort study”. In: *Fertility and Sterility* 95 (3 2011). DOI: 10.1016/j.fertnstert.2010.11.026.

[18] Dilpreet Kaur and Yadwinder Kaur. “Various image segmentation techniques: a review”. In: *International Journal of Computer Science and Mobile Computing* 3.5 (2014), pp. 809–814.

[19] V. Kiruthika and M.M. Ramya. "Automatic Segmentation of Ovarian Follicle Using K-Means Clustering". In: *2014 Fifth International Conference on Signal and Image Processing*. 2014, pp. 137–141. DOI: 10.1109/ICSIP.2014.27.

[20] F. W. Kremkau and K. J. Taylor. "Artifacts in ultrasound imaging". In: *Journal of Ultrasound in Medicine* 5.4 (1986), pp. 183–237. DOI: 10.7863/jum.1986.5.4.227.

[21] Mitja Lenic, Damjan Zazula, and Boris Cigale. "Segmentation of ovarian ultrasound images using single template cellular neural networks trained with support vector machines". In: *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*. 2007, pp. 205–212. DOI: 10.1109/CBMS.2007.97.

[22] Haoming Li et al. "CR-Unet: A Composite Network for Ovary and Follicle Segmentation in Ultrasound Images". In: *IEEE journal of biomedical and health informatics* 24.3 (2020), pp. 974–983.

[23] Yang Li. "CRU-Net: A Deep Learning Network for Semantic Segmentation of Pathological Tissue Slices". In: *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*. IEEE. 2021, pp. 46–50.

[24] Antonio La Marca and Sesh Kamal Sunkara. "Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice". In: *Human Reproduction Update* 20.1 (2013), pp. 124–140. DOI: 10.1093/humupd/dmt037.

[25] Andrew Mehnert and Paul Jackway. "An improved seeded region growing algorithm". In: *Pattern Recognition Letters* 18.10 (1997), pp. 1065–1071.

[26] Shervin Minaee et al. "Image segmentation using deep learning: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021), pp. 3523–3542.

[27] Suraya Nahlawi and Nedi Gari. *Sonography Transvaginal Assessment, Protocols, and Interpretation*. Treasure Island (FL): StatPearls Publishing, 2022.

[28] M. A. Coelho Neto et al. "Counting ovarian antral follicles by ultrasound: a practical guide". In: *Ultrasound Obstetrics Gynecology* 51 (2018), pp. 10–20.

[29] ESHRE Guideline Group on the Number of Embryos to Transfer et al. *Evidence-based guideline: Number of embryos to transfer during IVF/ICSI*. ESHRE. URL: https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Embryo-transfer.

[30] World Health Organization. *Infertility*. https://www.who.int/news-room/fact-sheets/detail/infertility. Accessed: 2022-11-29. 2020.

[31] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (2017).

[32] Božidar Potočnik and Martin Šavc. "Deeply-Supervised 3D Convolutional Neural Networks for Automated Ovary and Follicle Detection from Ultrasound Volumes". In: *Applied Sciences* 12.3 (2022). ISSN: 2076-3417. DOI: 10.3390/app12031246. URL: https://www.mdpi.com/2076-3417/12/3/1246.

[33] Božidar Potočnik and Damjan Zazula. "Automated analysis of a sequence of ovarian ultrasound images. Part I: segmentation of single 2D images". In: *Image and Vision Computing* 20.3 (2002), pp. 217–225.

[34] Božidar Potočnik and Damjan Zazula. "Automated analysis of a sequence of ovarian ultrasound images. Part II: prediction-based object recognition from a sequence of images". In: *Image and Vision Computing* 20.3 (2002), pp. 227–235.

[35] Božidar Potočnik, Damjan Zazula, and Danilo Korže. "Automated Computer-Assisted Detection of Follicles in Ultrasound Images of Ovary". In: *Journal of Medical Systems* 21 (1997), pp. 445–457. DOI: 10.1023/A:1022832515369.

[36] Shital Adarsh Raut et al. "Image segmentation–a state-of-art survey for prediction". In: *2009 international conference on advanced computer control*. IEEE, 2009, pp. 420–424.

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer. 2015, pp. 234–241.

[38] Diplav Srivastava et al. "Unsupervised Deep Learning based Longitudinal Follicular Growth Tracking during IVF Cycle using 3D Transvaginal Ultrasound in Assisted Reproduction". In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2021, pp. 3209–3212. DOI: 10.1109/EMBC46164.2021.9630495.

[39] Helena Teede et al. *International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome 2023*. Monash University, 2023. ISBN: 978-0-6458209-0-4. DOI: 10.26180/24003834.v1.

[40] *The Discrete Kalman Filter*. https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/WELCH/kalman.1.html. Accessed: 24-04-10. 1999.

[41] National Ultrasound. *How Much Does a Sonogram Machine Cost?* https://www.nationalultrasound.com/how-much-does-an-ultrasound-machine-cost/. Accessed: 2022-12-14. 2021.

[42] Anh-Vu Vo et al. "Octree-based region growing for point cloud segmentation". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 104 (2015), pp. 88–100.

[43] Liwei Wang et al. “Training deeper convolutional networks with deep supervision”. In: *arXiv preprint arXiv:1505.02496* (2015).

[44] Greg Welch, Gary Bishop, et al. “An introduction to the Kalman filter”. In: (1995).

[45] Shiuh-Ku Weng, Chung-Ming Kuo, and Shu-Kang Tu. “Video object tracking using adaptive Kalman filter”. In: *Journal of Visual Communication and Image Representation* 17.6 (2006), pp. 1190–1208.

[46] Xin Yang et al. “Contrastive rendering with semi-supervised learning for ovary and follicle segmentation from 3D ultrasound”. In: *Medical Image Analysis* 73 (2021), pp. 102–134. ISSN: 1361-8415. DOI: 10.1016/j.media.2021.102134.

[47] Zongwei Zhou et al. “Unet++: A nested u-net architecture for medical image segmentation”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer. 2018, pp. 3–11.