

# HaTU-Net: Harmonic Attention Network for Automated Ovarian Ultrasound Quantification in Assisted Pregnancy

Vivek Kumar Singh <sup>1,†</sup> , Elham Yousef Kalafi <sup>1,†</sup>, Eugene Cheah <sup>1</sup>, Shuhang Wang <sup>1</sup>, Jingchao Wang <sup>3</sup>, Arinc Ozturk <sup>1</sup>, Qian Li <sup>1</sup>, Yonina C. Eldar <sup>2</sup>, Anthony E. Samir <sup>1</sup>, and Viksit Kumar <sup>1</sup>

<sup>1</sup> Center for Ultrasound Research & Translation at the Massachusetts General Hospital, Department of Radiology, Harvard Medical School, Boston, 02114, MA USA; vivekkr.singh90@gmail.com

<sup>2</sup> Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, 7610001, Israel

<sup>3</sup> Department of Ultrasound, The Third Hospital of Hebei Medical University, Shijiazhuang 050051, China

\* Correspondence: vkumar14@mgh.harvard.edu

† These authors contributed equally to this work.

**Abstract:** Antral follicle Count (AFC) is a non-invasive biomarker used to assess ovarian reserves through transvaginal ultrasound (TVUS) imaging. Antral follicles' diameter is usually in the range of 2-10 mm. The primary aim of ovarian reserve monitoring is to measure the size of ovarian follicles and the number of antral follicles. Manual follicle measurement is inhibited by operator time, expertise and the subjectivity of delineating the two axes of the follicles. This necessitates an automated framework capable of quantifying follicle size and count in a clinical setting. This paper proposes a novel Harmonic Attention-based U-Net network, HaTU-Net, to precisely segment the ovary and follicles in ultrasound images. We replace the standard convolution operation with a harmonic block that convolves the features with a window-based discrete cosine transform (DCT). Additionally, we proposed a harmonic attention mechanism that helps to promote the extraction of rich features. The suggested technique allows for capturing the most relevant features, such as boundaries, shape, and textural patterns, in the presence of various noise sources (i.e., shadows, poor contrast between tissues, and speckle noise). We evaluated the proposed model on our in-house private dataset of 197 patients undergoing TransVaginal UltraSound (TVUS) exam. The experimental results on an independent test set confirms that HaTU-Net achieved a Dice coefficient score of 90% for ovaries and 81% for antral follicles, an improvement of 2% and 10%, respectively, when compared to a standard U-Net. Further, we accurately measure the follicle size, yielding the recall, and precision rates of 91.01% and 76.49%, respectively.

**Keywords:** Follicle Monitoring; Deep Learning; Antral Follicle Count; Harmonic Attention; Ultrasound Imaging; Pelvic Ultrasound

**Citation:** Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* 2022, 1, 0. <https://doi.org/>

Received:

Accepted:

Published:

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2022 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Patients with infertility have shown a correlation between predictors of functional ovarian reserves and ovarian responses to pregnancy outcomes [1]. Antral follicle count (AFC) and size, obtained using TransVaginal UltraSound (TVUS) images, are non-invasive imaging biomarkers used to assess and quantify ovarian reserve [1] [2]. The primary aim of ovarian reserve monitoring is to measure the number and size of ovarian follicles and the number of antral follicles, which are, on average, 2-10 mm in diameter [3]. Follicle size is measured by taking the average of each follicle's two largest orthogonal diameters [4]. There are limitations to manually estimating the size and count of follicles; the process is time-consuming, inconsistent [3], and highly variable depending upon the actual shape of primarily non-spherical follicles [2]. An accurate, automated method to segment ovaries and follicles and count the follicles could optimize the clinical flow and reduce subjectivity.

Developing an automated solution for ovary and follicle segmentation incorporates numerous challenges. Ultrasound imaging artifacts impede the performance of deep

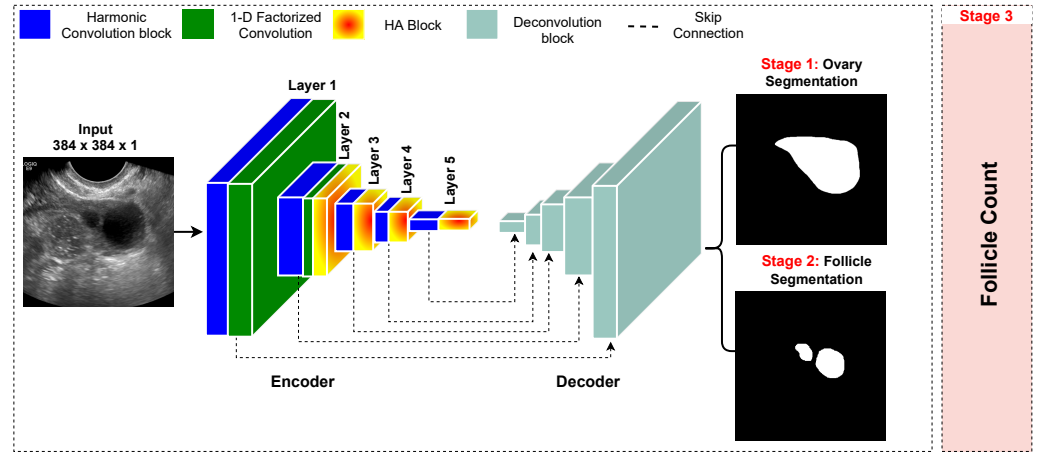
learning-based segmentation methods. Blurred ambiguous boundaries further compound challenges in delineating tissue boundaries and the presence of acoustic shadowing [5]. Many image processing and computer vision-based methods are suggested to overcome these challenges that involve geometric features [6] and watershed [7]. Active contours-based [8] approaches have been used to segment the ovary and follicles. Traditional ovary and follicle monitoring methods have been frequently explored with large and distinctly visible follicles [2], [9]. Boundary ambiguity is noticeable in ovarian and follicular images. The traditional methods have some limitations, such as watershed or thresholding approaches generating discontinuities and variances of intensity in the ovarian ultrasound images. Their slow speed creates challenges to adopt in actual practice clinical settings.

Convolutional neural networks (CNNs) have shown substantial performance, and accuracy advancements over conventional methods [10]. With the great success of CNNs, multiple popular segmentation methods have been developed such as FCN [11], U-Net [12], SegNet [13], Attention-UNet [14], DeepLabv3+ [15], ERFNet [16], and BiseNetv2 [17] that segment the objects or anatomies. These methods achieved state-of-the-art results for various semantic segmentation tasks.

Recently, many segmentation methods have been developed for segmenting biomedical images [18]. Meng et al. [19] proposed an instinctive deep learning-based contour regression model for biomedical image segmentation. The authors aggregated multi-level and multi-stage networks to regress the contour coordinates in an end-to-end manner rather than pixel-wise dense predictions. The authors used this method to segment the fetal head in ultrasound images and the optic disc and optic cup in color fundus images. Valanarasu et al. [20] presented a network architecture called KiU-Net, which projects data onto higher dimensions and picks finer details when compared to a standard U-Net. The suggested method addressed the performance failures when segmenting smaller anatomical structures with blurred, noisy boundaries. The authors performed the brain anatomy segmentation from 2D ultrasound (US). Singh et al. [21] proposed an automated solution to segment the breast lesion from the US images. The recommended method used generative adversarial learning (GAN) networks. The introduced method efficiently extracts spatial features such as texture, edge, shape, intensity, and global information. The authors used an attention mechanism that highlights the most relevant features and ignores the background ones. However, the GAN-based method has limitations due to its computational complexity and fails to delineate if the lesion shape is not complete. Further, Yang et al. [22] incorporated the multi-directional recurrent neural network (RNN) with a customized CNN to extract spatial intensity concurrencies to eliminate boundary ambiguities. The author employed semantic segmentation methods in prenatal ultrasound volumes that potentially encourage fetal health monitoring.

Various deep learning-based methods have been used to detect and segment ovary, and antral follicles [5], [23]. Li, Haoming et al. [5] proposed an ovary and follicle segmentation model called CR-UNet, consisting of spatially recurrent neural networks incorporated into a standard U-Net. The recommended network has limitations in correctly delineating and detecting the follicles that are joined with each other [23]. Gupta et al. [24] developed a deep learning-based framework for ovarian volume computation that utilizes 3D US volumes and the axial orientation. The authors evaluated their methods on 20 3D ovarian US volumes that enhanced the grade of the 3D rendering of the ovary and addressed the issue of combined follicles in segmentation. Yang et al. [25] introduced ovary and follicles segmentation using the contrastive rendering (C-Rend) framework. The authors employed the semi-supervised learning approach with C-Rend leveraging unlabeled 3D ultrasound for better performance. However, this study has some limitations during inference due to its hyperparameter default value which might not be the most suitable setting for each 3D US.

The main aim of this study is to develop an automated method for efficient ovary and follicles segmentation in ovarian TVUS images to facilitate measuring the size of the follicle. Figure 1 shows the schematic view of our proposed framework. The framework incor-



**Figure 1.** A general framework of ovarian ultrasound quantification. This includes the three stages: ovary segmentation, follicles segmentation, and follicle count. Note, HA refers to harmonic attention block.

porates three stages, i.e., ovary segmentation, follicle segmentation, and follicle counting. We designed a new segmentation method that replaces the standard 2D convolution layer with a harmonic convolution. In contrast, [26] harmonic convolution combines the learned kernels with predefined filters for feature learning. This weighted combination reduces overfitting and computational complexity. The proposed HaTU-Net method effectively extracts the features that allow precise segmentation of the ovary and follicles from the US images. Moreover, we developed a new attention block that helps to improve the segmentation performance by encouraging the feature discriminability between the pixels and ignoring US imaging artifacts. In summary, our main contributions are in four folds:

- We propose a segmentation network called HaTU-Net to segment ovaries and follicles with detailed, refined boundaries from TVUS images.
- We propose using harmonic convolution [26] to replace the standard convolutional filter. The input image is first decomposed using the discrete cosine transform (DCT); these transformed signals are combined using learned weights.
- We developed harmonic attention (HA) block to improve feature discriminability between the target and background pixels in the segmentation stage. The HA block encourages the features by avoiding the artifacts, and support for the HaTU-Net leads to improved segmentation results.
- Our experimental results confirm HaTU-Net has shown significant improvement compared to the various state-of-the-art segmentation methods (U-Net [12], AttentionU-Net [14], R2U-Net [27], U-Net++ [28], and DeepLabv3+ [15]).

The remainder of this paper is organized as follows: Section 2 describes the presented methodology. Section 3 describes our experimental results and highlights the limitations of the work. Section 4 completes our study and suggests some future lines of research.

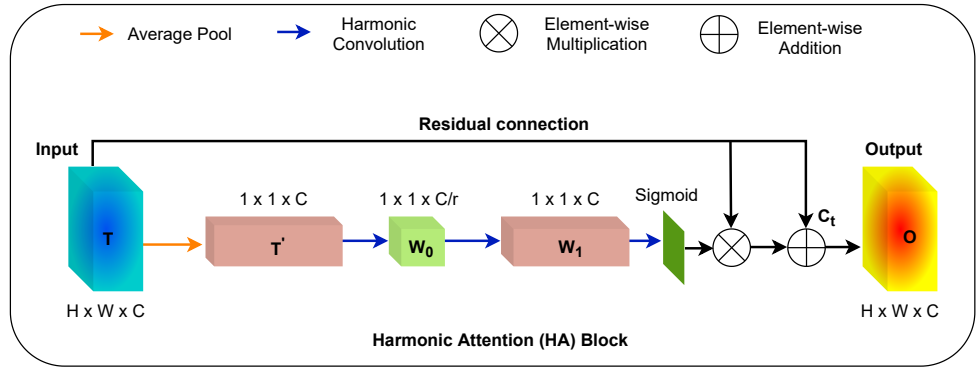
## 2. Methods

This section presents the detailed architecture description of the proposed HaTU-Net as depicted in Fig 1. The proposed method incorporated two main parts: harmonic convolution and harmonic attention (HA) blocks.

### 2.1. HaTU-Net Architecture

Figure 1 illustrates the architecture of the proposed HaTU-Net. It consists of encoder and decoder networks. Each network consists of five layers with added skip connections. The encoder network utilized a harmonic convolutional layer with kernel size  $3 \times 3$  instead of standard 2D convolutional layers. Each layer within the encoder employs batch normalization followed by the ReLU activation function. The first encoder layer combines

harmonic convolution with a 1-D kernel factorization [16], allowing feature extraction with low computational costs. The second layer uses a variety of harmonic convolution, 1-D kernel factorization, and harmonic attention (HA) blocks. The attention mechanism boosts the feature discriminability between the target and background pixels. The last three encoder layers use a harmonic convolution with an HA block to enable channel interdependencies, and highlight features specific to the ovaries and follicles. After each layer, dimensionality reduction is achieved using a max-pooling operation with a kernel size of  $2 \times 2$ . In the decoder, feature upsampling is performed through the *conv-transpose2D* operation. Each encoder layer's features are concatenated with the corresponding features in the decoding layer (skip connections). A threshold of 0.5 is used to generate the final predicted mask for ovary and follicles segmentation.



**Figure 2.** Illustration of harmonic attention (HA) block.

## 2.2. Feature Extraction with Harmonic Convolution

Motivated by [26], we replaced the standard convolution with a harmonic convolution, i.e., a weighted sum of the responses to a discrete cosine transform (DCT) filter bank to pull the harmonics from lower-level features to decrease the burden of overfitting. The DCT is a separable transform that converts a signal from the temporal domain to the spectral domain. The DCT of a 2D image  $I$  of size  $H \times W$  with a one-pixel discretization step can be formulated as follows [26]:

$$T_{u,v} = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} \sqrt{\frac{\beta_u}{H}} \sqrt{\frac{\beta_v}{W}} I_{(i,j)} \times \cos \left[ \frac{\pi}{H} \left( i + \frac{1}{2} \right) u \right] \cos \left[ \frac{\pi}{W} \left( j + \frac{1}{2} \right) v \right], \quad (1)$$

where  $T_{u,v}$  is the coefficient corresponding to a sinusoidal frequency of  $u$  and  $v$  in the two orthogonal directions. Here  $\beta_0 = 1$  and  $\beta_u = 2$  are scaling factors used to normalize the value of the basis function.

The input image features are decomposed using the DCT transform to perform the convolution operation. A kernel size of  $f \times f$  and depth of  $f^2$  in the filter bank is used. Let  $\gamma_{(u,v)}$  indicate the  $u, v$  frequency selective DCT filter with kernel size  $f \times f$ . The feature map  $\mathbb{F}_m$  at depth  $d$  is represented as a weighted linear aggregation of DCT coefficients on all input channels  $C$  as follows:

$$\mathbb{F}_m^d = \sum_{n=0}^{C-1} \sum_{u=0}^{f-1} \sum_{v=0}^{f-1} w_{n,u,v}^d \gamma_{(u,v)} * \mathbb{F}_{m(n)}^{d-1}, \quad (2)$$

where  $w_{n,u,v}^d$  represented the learned weight for the  $n^{th}$  feature at frequency  $u, v$  and  $**$  denoted as the 2-D convolution operator. The transformation process allows the input feature to undergo harmonic decomposition, allowing learned weights to be used for combining the transformed signals.

### 2.3. Harmonic Attention Block

148

To advance the feature discriminability between the small targeted region and background pixels, Fig. 2 presents the proposed harmonic attention (HA) block details. The block makes use of harmonic convolutions to extract feature maps. It is designed to promote feature discriminability between the target objects and their background, i.e., ovary and follicles, in our application. In this block, an input feature map  $T \in \mathbb{R}^{C \times H \times W}$  is average pooled to aggregate  $C$ , which is the channel statistics. This generates  $T' \in \mathbb{R}^{C' \times 1 \times 1}$  which is then passed to the two  $1 \times 1$  harmonic convolution layers to extract non-linear inter-channel relationships with the help of spectral DCT filters. If  $W_0 \in \mathbb{R}^{C \times \frac{C}{r}}$  and  $W_1 \in \mathbb{R}^{C \times \frac{C}{r}}$  are weights of two harmonic convolutional layers (where  $r$  refers to the reduction ratio), then the channel attention map can be formulated as:

$$attn(T') = \sigma(W_1 * (ReLU(W_0 * T'))), \quad (3)$$

where  $\sigma(\cdot)$  is the sigmoid activation function. Finally, the channel attention map can be generated as follows:

$$C_t = T * T'. \quad (4)$$

The use of skip connections between the input feature map  $T$  and  $C_t$  helps to narrow the semantic gap and results in the final output map  $O$ :

$$O = C_t + T. \quad (5)$$

### 2.4. Cost Function

149

We use a weighted sum of the binary cross-entropy (BCE) and focal loss  $\mathcal{L}^F$  which are defined as:

$$\mathcal{L}^F(y, \hat{y}) = -y(1 - \hat{y})^\gamma \cdot \log(\hat{y}) - (1 - y)\hat{y}^\gamma \cdot \log(1 - \hat{y}). \quad (6)$$

$$\mathcal{L}^{BCE}(y, \hat{y}) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})), \quad (7)$$

The final loss is expressed as follows:

$$Loss_{final} = \mathcal{L}^{BCE}(y, \hat{y}) + \eta * \mathcal{L}^F(y, \hat{y}), \quad (8)$$

where,  $y$  is the target binary mask, and  $\hat{y}$  is the predicted mask obtained by the segmentation model. We used  $\eta$  equal to 0.6 as an empirical weighting factor.

## 3. Experimental Design and Results

152

### 3.1. Dataset

153

The Institutional Review Board approved the retrospective study (IRB), and the requirement for informed consent was waived. The radiology reports of adult patients who underwent transvaginal ultrasound (TVUS) exams between 2005 and 2019 in a single institution were reviewed, and a total of 197 eligible patients were identified and selected. The inclusion criteria were: (1) premenopausal female adult patients, (2) underwent TVUS exams, (3) normal ovary on the pelvic ultrasound, and (4) available B-mode static images of the ovary. Patients with the following criteria were excluded: (1) low-quality US images with obscure boundaries of the ovary, (2) abnormal ovary findings on the ultrasound, or (3) known ovarian pathologies or ovarian surgical history.

All TVUS exams were performed using GE LOGIQ E9 (General Electric Healthcare, Waukesha, WI, USA) ultrasonic system equipped with a transvaginal IC5-9 transducer. US exams were reviewed in a picture archiving and communication system (PACS) to extract the DICOM images of ovaries in sagittal and coronal planes. Using the MicroDicom Viewer tool (Version 3.2.7, Sofia, Bulgaria), a radiologist who had ten years of experience in pelvic ultrasound annotated the contours of the ovary and follicles following a specific protocol, (1) annotated all follicles of 2-28mm [29] in diameter within each ovary, (2) annotate ovary

and follicles in different colors, and (3) avoid overlaps among annotations within the same ovary. Another senior radiologist with more than ten of experience reviewed the annotated images as part of the quality control process. The final dataset consisted of 767 images with qualified annotations, split into training, validation, and testing sets with 466, 160, and 141 US images.

### 3.2. Implementation Details

All the methods are implemented in PyTorch [30]. We use an NVIDIA GeForce RTX 2080Ti GPU with 11GB RAM. The images are resized to  $384 \times 384$ , and the pixel values are normalized between 0 and 1. The training dataset is augmented with random 15-degree rotations and horizontal flips. An ADAM optimizer with  $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ , and an initial learning rate of 0.0002 is used to optimize the model better. Step decay learning is activated if the Dice coefficient score for the validation set plateaus for two consecutive epochs. We used a batch size of four images and trained the model for 50 epochs since the model was optimized completely.

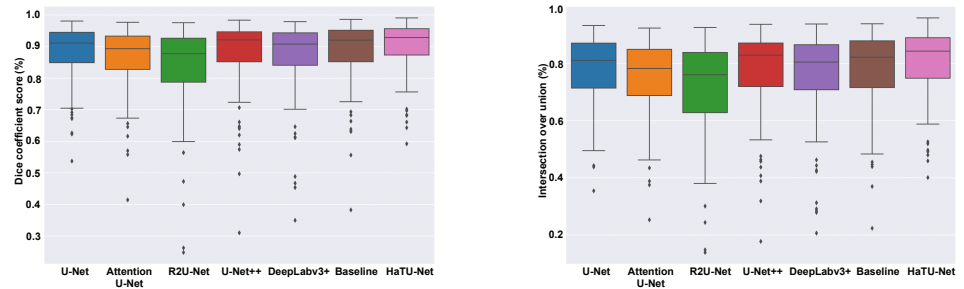
This study organizes our experiments into three key stages: 1. ovary segmentation, 2. follicle segmentation, and 3. follicle counting. To measure the effectiveness of the proposed model on segmentation stages, five evaluation metrics are used, i.e., accuracy, dice similarity coefficient (Dice), intersection over union (IoU), sensitivity, and specificity [31]. The precision and recall metrics are used to evaluate the follicle counts.

### 3.3. Ovary Segmentation

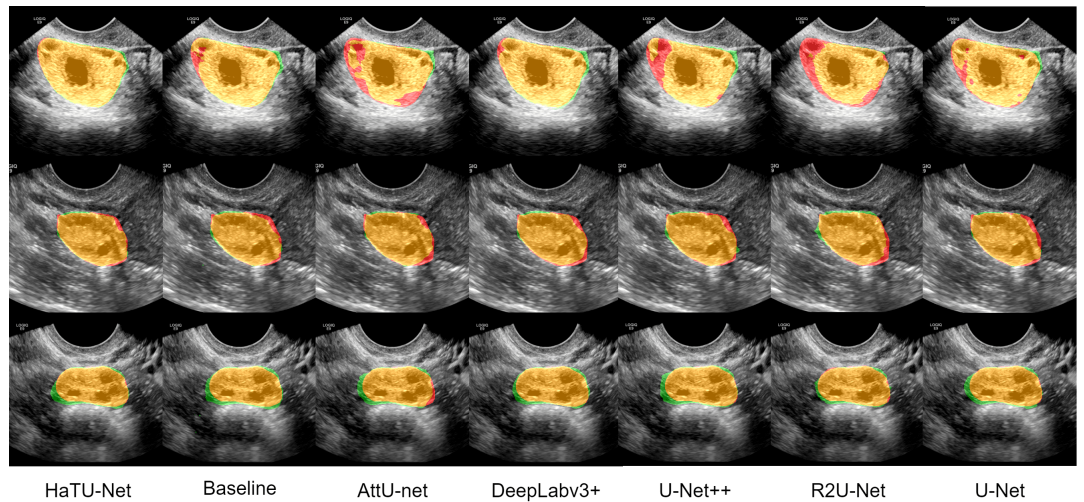
**Table 1.** Ovary segmentation results in five state-of-the-art segmentation methods. The best significant results are highlighted in bold.

Methods	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [12]	$96.89 \pm 0.02$	$87.89 \pm 0.08$	$77.54 \pm 0.12$	$90.51 \pm 0.09$	$98.27 \pm 0.01$
Attention U-net [14]	$96.41 \pm 0.02$	$86.02 \pm 0.09$	$74.89 \pm 0.13$	$87.96 \pm 0.11$	$98.16 \pm 0.02$
R2U-Net [27]	$95.73 \pm 0.03$	$83.31 \pm 0.12$	$71.49 \pm 0.15$	$86.31 \pm 0.13$	$97.53 \pm 0.02$
U-Net++ [28]	$97.14 \pm 0.02$	$87.78 \pm 0.10$	$77.72 \pm 0.13$	$89.42 \pm 0.13$	$98.19 \pm 0.01$
DeepLabv3+ [15]	$96.99 \pm 0.02$	$86.66 \pm 0.11$	$76.21 \pm 0.14$	$87.35 \pm 0.15$	$98.25 \pm 0.01$
Baseline	$97.26 \pm 0.01$	$88.37 \pm 0.09$	$78.51 \pm 0.13$	$90.46 \pm 0.12$	$98.18 \pm 0.01$
<b>HaTU-Net</b>	<b><math>97.55 \pm 0.01</math></b>	<b><math>90.01 \pm 0.07</math></b>	<b><math>80.72 \pm 0.11</math></b>	<b><math>90.86 \pm 0.10</math></b>	<b><math>98.57 \pm 0.01</math></b>

Table 1, demonstrates the quantitative result of proposed HaTU-Net (i.e., Baseline (BL) plus harmonic attention block) compared to five state-of-the-art segmentation methods including U-Net [12], Attention U-Net [14], R2U-Net [27], U-Net++ [28], and DeepLabv3+ [15]. We also demonstrate the result of the Baseline (BL) method consisting of a standard U-Net network with harmonic convolution without an attention block. Experimental results confirm that the HaTU-Net performed significantly better than the second-highest U-Net method in DSC, and IoU metrics with 2%, and 3%, respectively. We observed that the DCT-based spectral kernel aids in learning shape, boundary, and texture mapping from noisy ultrasound images. These noisy images include poor contrast, shadows, speckle variation, and poor signal-to-noise ratio. To reach the optimal version of the proposed method, we performed an ablation study to determine and quantify the perceptiveness of each employed block to segmentation results. We define our baseline model as a U-Net where it replaces the standard 2D convolutions layers with harmonic convolutions. Since the ovary occupies a large part of an ultrasound image, we enforce fine boundary segmentation while retaining shape information through an HA block (BL+ HA). Therefore, adding the harmonic attention block to BL leads to better segmentation results. Finally, HaTU-Net improves the DSC and IoU scores by approximately 1.5% in both metrics compared to our BL model. The characteristics of the attention mechanism allow highlighting the most relevant feature of the hypoechoic ovary region and ignoring the background or acoustic shadows in the US images.



**Figure 3.** Boxplots of Dice and IoU scores on ovarian images. Different color boxes indicate the score range of different methods; the black line inside each box represents the median value, box limits include interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are computed as 1.5 times the distance of upper and lower limits of the box. All values outside the whiskers are considered outliers.



**Figure 4.** Illustration of three qualitative example results generated by HaTU-Net against five state-of-the-art methods for ovary segmentation. Colors maps represent the following description: Orange (true positive), Green (false positive) and Red (false negative).

Figure. 3 shows the box-plot analysis of Dice and IoU scores on the ovarian dataset. The HaTU-Net has generated fewer outliers compared to other segmentation methods. For instance, we can see that the proposed method has a lower standard deviation than other methods, showcasing its robustness.

Further, qualitative analysis plays a crucial role in visually determining segmentation results. Figure. 4 exhibits the three qualitative examples results generated by state-of-the-art segmentation methods compared to the HaTU-Net. We provide the color maps that help to identify the true positive (orange), false positive (green), false negative (red), and true negative, including the background. Visual inspection confirms that HaTU-Net precisely segments the ovary boundaries, whereas other methods have produced many false positives.

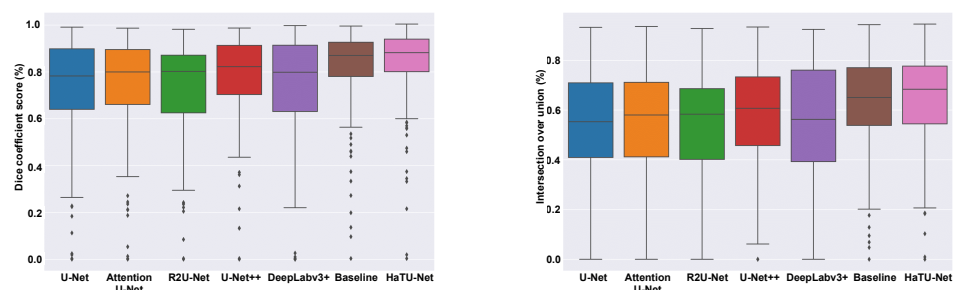
### 3.4. Follicle Segmentation

Table 2 exhibits the follicle segmentation results. The proposed HaTU-Net improved follicle segmentation results with 10% DSC and IoU scores compared to U-Net. Our BL methods demonstrated more remarkable results than other state-of-the-art segmentation methods. We observed that incorporating the HA block leads to better segmentation results. Especially, HA block can capture many small follicles' shape features and provide separation between them. This allows delineating of the follicle boundaries efficiently.

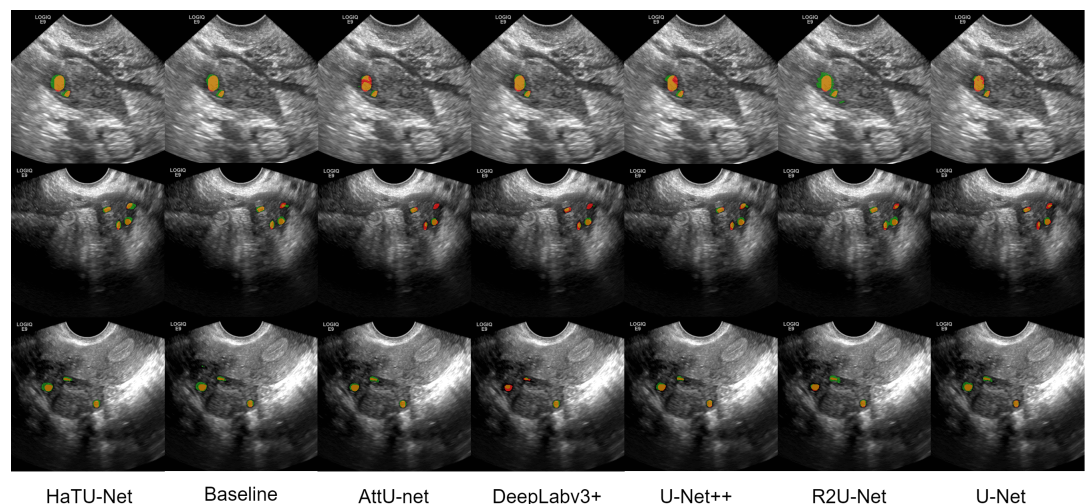
**Table 2.** Follicles segmentation results in five state-of-the-art segmentation methods. The best significant results are highlighted in bold.

Methods	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [12]	99.20 $\pm$ 0.01	71.77 $\pm$ 0.22	54.18 $\pm$ 0.23	74.55 $\pm$ 0.25	99.51 $\pm$ 0.01
Attention U-net [14]	99.28 $\pm$ 0.01	73.23 $\pm$ 0.22	55.56 $\pm$ 0.22	73.92 $\pm$ 0.24	99.62 $\pm$ 0.01
R2U-Net [27]	98.99 $\pm$ 0.01	73.28 $\pm$ 0.2	56.07 $\pm$ 0.21	<b>84.91 <math>\pm</math> 0.19</b>	99.17 $\pm$ 0.01
U-Net++ [28]	99.35 $\pm$ 0.01	75.48 $\pm$ 0.20	57.61 $\pm$ 0.22	75.58 $\pm$ 0.23	99.68 $\pm$ 0.01
DeepLabv3+ [15]	99.35 $\pm$ 0.01	69.37 $\pm$ 0.26	52.36 $\pm$ 0.26	66.43 $\pm$ 0.28	99.77 $\pm$ 0.01
Baseline	99.45 $\pm$ 0.01	79.80 $\pm$ 0.19	62.51 $\pm$ 0.21	81.64 $\pm$ 0.20	99.70 $\pm$ 0.01
<b>HaTU-Net</b>	<b>99.51 <math>\pm</math> 0.01</b>	<b>81.40 <math>\pm</math> 0.18</b>	<b>64.11 <math>\pm</math> 0.20</b>	82.24 $\pm$ 0.19	<b>99.78 <math>\pm</math> 0.01</b>

Furthermore, Fig. 5 displays the boxplot analysis of Dice and IoU scores. The results confirm that HaTU-Net achieves significantly high mean segmentation results with a lower standard deviation. Besides, it is seen that it produces fewer outliers than other methods. Note that all values outside the whiskers are considered outliers.



**Figure 5.** Boxplots of Dice and IoU scores on follicle dataset. Different color boxes indicate the score range of different methods; the black line inside each box represents the median value, box limits include interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are computed as 1.5 times the distance of upper and lower limits of the box. All values outside the whiskers are considered outliers.



**Figure 6.** Illustration of three qualitative example results generated by HaTU-Net against five state-of-the-art methods for follicle segmentation. Colors maps represent the following description: Orange (true positive), Green (false positive), and Red (false negative).

Figure 6 represents the three examples for qualitative assessment of follicle segmentation. The harmonic attention block helps in refining the boundaries between follicles, as

**Table 3.** Ablation study of the loss function. The highest significant results are highlighted in bold.

Dataset	Loss Function	Accuracy	Dice	IoU	Sensitivity	Specificity
Ovary	BCE	<b>97.67 <math>\pm</math> 0.01</b>	89.28 $\pm$ 0.10	80.06 $\pm$ 0.13	88.94 $\pm$ 0.14	<b>98.86 <math>\pm</math> 0.01</b>
	Dice Loss	97.46 $\pm$ 0.02	89.45 $\pm$ 0.08	79.97 $\pm$ 0.12	89.69 $\pm$ 0.12	98.7 $\pm$ 0.01
	BCE + Dice	97.44 $\pm$ 0.0	89.13 $\pm$ 0.09	79.64 $\pm$ 0.13	90.76 $\pm$ 0.12	98.45 $\pm$ 0.01
	BCE + Focal	97.55 $\pm$ 0.01	<b>90.01 <math>\pm</math> 0.07</b>	<b>80.72 <math>\pm</math> 0.11</b>	<b>90.86 <math>\pm</math> 0.10</b>	98.57 $\pm$ 0.01
Follicle	BCE	99.11 $\pm$ 0.00	68.1 $\pm$ 0.22	55.6 $\pm$ 0.23	62.95 $\pm$ 0.26	99.76 $\pm$ 0.01
	Dice Loss	99.14 $\pm$ 0.01	72.72 $\pm$ 0.19	60.26 $\pm$ 0.21	72.02 $\pm$ 0.20	99.71 $\pm$ 0.01
	BCE + Dice	99.18 $\pm$ 0.01	74.56 $\pm$ 0.18	62.4 $\pm$ 0.20	77.44 $\pm$ 0.19	99.64 $\pm$ 0.0
	BCE + Focal	<b>99.51 <math>\pm</math> 0.01</b>	<b>81.40 <math>\pm</math> 0.18</b>	<b>64.11 <math>\pm</math> 0.20</b>	<b>82.24 <math>\pm</math> 0.19</b>	<b>99.76 <math>\pm</math> 0.01</b>

shown by the performance of HaTU-Net in separating follicles for automating the follicle count process.

### 3.5. Ablation Study

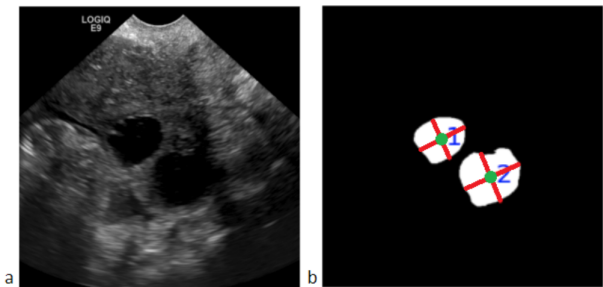
Table 3 demonstrates an ablation study to estimate the effect of loss function employing the proposed HaTU-Net method. Our experiments use various combinations of loss functions such as BCE, Dice loss, BCE+Dice, and BCE+Focal loss. BCE+Focal loss leads to 1% refinement over Dice loss in segmentation results. This combination of loss functions focuses on boundaries. Separately, employing the Dice loss gains better results versus BCE+Dice loss. On the follicle dataset, we also see that BCE+ Focal loss yields a 7% and 2% increment in DSC and IoU scores, respectively, compared to BCE+Dice loss. This suggested loss function help detach follicles that appear joined together, thus improving clinical outcome. However, employing only BCE leads to poor segmentation results achieving a 68.10% Dice score.

### 3.6. Follicle Counting

In the third stage, segmented follicles are used to measure the follicle size. Figure 7 shows an ovarian US image with oval-shaped follicles and its corresponding follicle segmentation mask generated by stage 2. The main steps of follicle counting are explained as follows:

- Input:** Single ground truth follicle segmentation mask and predicted follicle segmentation mask.
- Load ground truth mask and predicted mask images.
  - Measure follicle diameters on both images in pixels.
  - Convert pixel diameters to physical measurements.
  - Exclude follicles sized outside the recruitable range of 2-10 mm in diameter by converting pixels black, as counting antral follicles <2 mm in diameter might heighten the chances of counting small anechoic structures like vessels or artifacts; whereas counting dominant follicles >10 mm lack the evidence of clinical practicality [2].
  - Compute the dice similarity coefficient (DSC) between the ground truth mask and predicted mask images.
  - Calculate the number of correctly detected follicles from the predicted mask (follicles with >0.5 Dice coefficient score are considered).
  - Calculate the number of detected follicles from the predicted mask.
  - Calculate the number of actual follicles from the ground truth mask.
  - Evaluate the precision and recall of our predicted follicle counting with the formula from [32].

Table 4 confirms the results for follicle counting. Experimental results prove that HaTU-Net attained a very high rate of follicle counting (i.e., precision) against other state-of-the-art models. In addition, we observed that the 76.79% of HaTU-Net gained a 1.5% recall rate compared to R2U-Net. The HaTU-Net helps lower the false positive rate from segmentation artifacts, leading to higher follicle counting.



**Figure 7.** (a) shows an ovarian US image with two oval-shaped follicles; (b) shows the corresponding follicle segmentation mask. The follicle delineations are white with green dots denoting individual follicles’ centroids. These act as the starting point for calculating the major and minor axis lengths shown in red lines. The diameter is then calculated by multiplying the axis lengths by the pixel size in mm to obtain physical measurements. The follicle count is labeled in blue.

**Table 4.** Follicles counting results in five state-of-the-art segmentation methods. Follicle counts are measured by precision and recall. The best significant results are highlighted in bold.

Methods	U-Net	Attention U-Net	R2U-Net	U-Net++	DeepLabv3+	HaTU-Net
Total No. of Images	141					
No. of Real Follicles	378					
No. of Detected Follicles	447	438	482	453	488	448
No. of Correctly Detected Follicles	315	303	350	339	328	344
Precision (%)	70.47	69.18	72.61	74.83	67.21	<b>76.69</b>
Recall (%)	83.33	80.16	90.59	89.68	86.77	<b>91.01</b>

4. Conclusion

In this paper, we proposed an automated solution for ovarian ultrasound quantification. We have developed a novel method named HaT-UNet that provides accurate segmentation of ovaries and follicles from TVUS images. The proposed HaTU-Net employed harmonic convolution with discrete cosine transform (DCT) and enhanced feature discriminability through HA block to handle ambiguous boundaries. Experimental results proved that the HaTU-Net tackles the presence of imaging artifacts by achieving DSC score improvement of 2% and 10% for the ovary and follicles, respectively, compared to U-Net. Further, the proposed model verified its effectiveness in follicle counting and attained a recall of 91% and a precision of 76.69%. Conclusively, the experimental output demonstrated HaTU-Net’s outstanding ability and provided efficient segmentation results by outperforming other state-of-the-art algorithms. In the future, we will extend the potential of the proposed model to 3D segmentation tasks. We would explore investigating the proposed model’s ability on additional biomedical datasets.

**Author Contributions:** Conceptualization, V.K.S, E.Y.K, E.C., V.K; methodology, V.K.S, E.Y.K, E.C, V.K.; software, V.K.S, E.Y.K, E.C.; validation, V.K.S, E.Y.K, E.C.; formal analysis, V.K.S, E.Y.K, E.C, V.K.; investigation, V.K.S, E.Y.K.; resources, V.K, A.E.S.; data curation, A.O, J.W, Q.L.; writing—original draft preparation, V.K.S, E.Y.K, E.C., writing—review and editing, V.K.S, E.Y.K, E.C, S.W, J.W, A.O, Q.L, V.K, Y.C.E, A.E.S.; visualization, V.K.S, E.Y.K, E.C.; supervision, V.K, Y.C.E, A.E.S.; project administration, V.K, Y.C.E, A.E.S.; funding acquisition, Y.C.E., A.E.S. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** This research was partly funded by Jameel Clinic, MIT.

References

1. Rosen, M.P.; Johnstone, E.; Addaun-Andersen, C.; Cedars, M.I. A lower antral follicle count is associated with infertility. *Fertility and sterility* **2011**, *95*, 1950–1954.

2. Coelho Neto, M.A.; Ludwin, A.; Borrell, A.; Benacerraf, B.; Dewailly, D.; da Silva Costa, F.; Condous, G.; Alcazar, J.L.; Jokubkiene, L.; Guerriero, S.; et al. Counting ovarian antral follicles by ultrasound: a practical guide. *Ultrasound in Obstetrics & Gynecology* **2018**, *51*, 10–20.

3. Faghih, R.T.; Styer, A.K.; Brown, E.N. Automated ovarian follicular monitoring: a novel real-time approach. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 632–635. 302
4. Wertheimer, A.; Nagar, R.; Oron, G.; Meizner, I.; Fisch, B.; Ben-Haroush, A. Fertility Treatment Outcomes After Follicle Tracking With Standard 2-Dimensional Sonography Versus 3-Dimensional Sonography-Based Automated Volume Count: Prospective Study. *Journal of Ultrasound in Medicine* **2018**, *37*, 859–866. 303
5. Li, H.; Fang, J.; Liu, S.; Liang, X.; Yang, X.; Mai, Z.; Van, M.T.; Wang, T.; Chen, Z.; Ni, D. CR-Unet: a composite network for ovary and follicle segmentation in ultrasound images. *IEEE journal of biomedical and health informatics* **2019**, *24*, 974–983. 304
6. Hiremath, P.; Tegnoor, J.R. Recognition of follicles in ultrasound images of ovaries using geometric features. In Proceedings of the 2009 International Conference on Biomedical and Pharmaceutical Engineering. IEEE, 2009, pp. 1–8. 305
7. Deng, Y.; Wang, Y.; Chen, P. Automated detection of polycystic ovary syndrome from ultrasound images. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008, pp. 4772–4775. 306
8. Hiremath, P.; Tegnoor, J.R. Automatic detection of follicles in ultrasound images of ovaries using edge-based method. *IJCA, Special Issue on RTIPPR* **2010**, *2*, 120–125. 307
9. Potočnik, B.; Zazula, D.; Korže, D. Automated computer-assisted detection of follicles in ultrasound images of ovary. *Journal of medical systems* **1997**, *21*, 445–457. 308
10. Sultana, F.; Sufian, A.; Dutta, P. Evolution of image segmentation using deep convolutional neural network: a survey. *Knowledge-Based Systems* **2020**, *201*, 106062. 309
11. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440. 310
12. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241. 311
13. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *39*, 2481–2495. 312
14. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* **2018**. 313
15. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818. 314
16. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *19*, 263–272. 315
17. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *arXiv preprint arXiv:2004.02147* **2020**. 316
18. Liu, X.; Song, L.; Liu, S.; Zhang, Y. A review of deep-learning-based medical image segmentation methods. *Sustainability* **2021**, *13*, 1224. 317
19. Meng, Y.; Wei, M.; Gao, D.; Zhao, Y.; Yang, X.; Huang, X.; Zheng, Y. CNN-GCN aggregation enabled boundary regression for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2020, pp. 352–362. 318
20. Jose, J.M.; Sindagi, V.; Hacıhaliloglu, I.; Patel, V.M. KiU-Net: towards accurate segmentation of biomedical images using over-complete representations. *arXiv preprint arXiv:2006.04878* **2020**. 319
21. Singh, V.K.; Abdel-Nasser, M.; Akram, F.; Rashwan, H.A.; Sarker, M.M.K.; Pandey, N.; Romani, S.; Puig, D. Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. *Expert Systems with Applications* **2020**, *162*, 113870. 320
22. Yang, X.; Yu, L.; Li, S.; Wen, H.; Luo, D.; Bian, C.; Qin, J.; Ni, D.; Heng, P.A. Towards automated semantic segmentation in prenatal volumetric ultrasound. *IEEE transactions on medical imaging* **2018**, *38*, 180–193. 321
23. Mathur, P.; Kakwani, K.; Kudavelly, S.; Ramaraju, G.; et al. Deep Learning based Quantification of Ovary and Follicles using 3D Transvaginal Ultrasound in Assisted Reproduction. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 2109–2112. 322
24. Gupta, S.; Kudavelly, S.R.; Ramaraju, G.; et al. Ovarian assessment using deep learning based 3D ultrasound super resolution. In Proceedings of the Medical Imaging 2021: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2021, Vol. 11597, p. 115970K. 323
25. Yang, X.; Li, H.; Wang, Y.; Liang, X.; Chen, C.; Zhou, X.; Zeng, F.; Fang, J.; Frangi, A.; Chen, Z.; et al. Contrastive Rendering with Semi-supervised Learning for Ovary and Follicle Segmentation from 3D Ultrasound. *Medical Image Analysis* **2021**, p. 102134. 324
26. Ulicny, M.; Krylov, V.A.; Dahyot, R. Harmonic Convolutional Networks based on Discrete Cosine Transform. *arXiv preprint arXiv:2001.06570* **2020**. 325
27. Alom, M.Z.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In Proceedings of the NAECON 2018-IEEE National Aerospace and Electronics Conference. IEEE, 2018, pp. 228–233. 326
28. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*; Springer, 2018; pp. 3–11. 327

- 
29. Gougeon, A.; Lefèvre, B. Evolution of the diameters of the largest healthy and atretic follicles during the human menstrual cycle. *Reproduction* **1983**, *69*, 497–502. 361
30. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703* **2019**. 362
31. Lalande, A.; Garreau, M.; Frouin, F. Evaluation of cardiac structure segmentation in cine magnetic resonance imaging. *Multi-Modality Cardiac Imaging: Processing and Analysis* **2015**, pp. 169–215. 363
32. Sonigo, C.; Jankowski, S.; Yoo, O.; Trassard, O.; Bousquet, N.; Grynberg, M.; Beau, I.; Binart, N. High-throughput ovarian follicle counting by an innovative deep learning approach. *Scientific reports* **2018**, *8*, 1–9. 364
- 365
- 366
- 367
- 368