FollicleFinder: automated three-dimensional segmentation of human ovarian follicles

Kevin Yamauchi[1,2,+], Marieke Biniasch[3,+], Leopold Franz[1,2], Harold Gómez[1,2], Christian De Geyter[3], and Dagmar Iber[1,2,*]

[1]*Department of Biosystems Science and Engineering, ETH Zürich, Mattenstrasse 26, 4058 Basel, Switzerland*
[2]*Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*
[3]*Reproductive Medicine and Gynecological Endocrinology (RME), University Hospital, University of Basel, Vogesenstr. 134, 4031 Basel, Switzerland*
[+]*contributed equally*

Summary: In vitro fertilization (IVF) treatment protocols require frequent monitoring of the ovarian follicle growth process. We report FollicleFinder, an open source pipeline for the automated, 3D segmentation of ovarian follicles. FollicleFinder also accurately measures clinically-relevant morphological properties such as diameter, surface area, and volume.

Availability: The FollicleFinder pipeline is available at `https://git.bsse.ethz.ch/iber/ovary-analysis` and the graphical user interface is available at `https://git.bsse.ethz.ch/iber/follicle-tracker`.

[*]**Correspondence:** iberd@ethz.ch

1 Introduction

Assisted reproductive technology (ART) offers a range of treatment options to couples having trouble conceiving. In vitro fertilization (IVF) treatment protocols include the collection of gametes. Successful IVF requires a set of multiple oocytes and ovarian hyper stimulation is therefore used to support the maturation of a larger number of ovarian follicles. During hyperstimulation, the ovarian response is monitored by repeated transvaginal ultrasound (TVUS) examination, and clinical decisions regarding the dosage and timing of hormonal stimulation are based on the follicle number and size.

Previously, 2D slices were imaged in clinical settings, but 3D segmentation algorithms have since been developed [1], and several commercial software packages now offer the 3D segmentation of ovarian follicles [2]. General Electric (GE) offers two 3D segmentation algorithms, VOCAL and SonoAVC [3], directly with its ultrasound scanner. Both software packages measure the volume and surface area of follicles [4]. Unfortunately, there are too many segmentation errors that require correction by hand as that either package could be used routinely in a clinical setting [5]. Measurements by hand are, however, slow and introduce a user-dependent variability. Accordingly, there is a need for better algorithms that enable the automatic, accurate detection and 3D segmentation of ovarian follicles.

Deep learning approaches have improved image segmentation across scientific domains, and have recently been explored in the context of ovarian follicle detection and segmentation. CR-UNet, which combines a spatial Recurrent Neural Network (RNN) with a standard 2D U-Net, yields promising results with 2D TVUS images [2, 1]. A subsequent study used an S-Net architecture for the simultaneous segmentation of ovary and follicles in 3D TVUS volumes [6]. These deep learning approaches provide more accurate segmentation of follicles in ultrasound images. However, the source code and models were not made available, making it difficult to build upon and reproduce this work.

To facilitate the translation of image-derived features into the clinic and the development of new models of follicular selection, there is a need for open source software that will serve as the basis for these innovations. To fill this gap, we have developed FollicleFinder, an open source platform for the detection and measurement of follicles from transvaginal ultrasound images.

2 Approach

FollicleFinder performs instance segmentation of the follicles from transvaginal ultrasound images using a 3D Unet via a two step segmentation pipeline: first the ovary is segmented and then the ovary is extracted and the follicles are segmented within the extracted ovary volume (Figure 1A). See the Methods section in the Supplementary Information for details on image acquisition and network training. We constrain the segmentation to the biologically plausible region by first segmenting the ovary. Since obtaining ground truth labels is laborious, we have implemented a self-supervised denoising preprocessing step. Denoising reduces the amount of training data required because denoising and semantic segmentation are related tasks [7]. Indeed our segmentations are on par with state of the art results while training on only 4% of the data (Supplemental Information, section 2). Morphological properties such as the volume, surface area, and effective diameter are directly measured from the segmented follicles. Thus, follicle growth dynamics can be tracked over the course of the menstrual cycle (Figure 1C)

To facilitate usage of the FollicleFinder pipeline in research and clinical settings, we have created a graphical user interface as a napari plugin called FollicleTracker (Figure 1B). The FollicleTracker plugin loads images from standard image formats (e.g., TIFF, DICOM) and renders them in 2D and 3D in the napari viewer. Via the graphical user interface, the user can use FollicleFinder to segment the follicles and measure their properties. Since FollicleTracker is written as a napari plugin with standard python libraries, we expect that others will be able to extend it for usage with other analyses.

3 Results

3.1 Dataset

To help others further improve follicle segmentation algorithms, we are releasing our image dataset of raw images plus ground
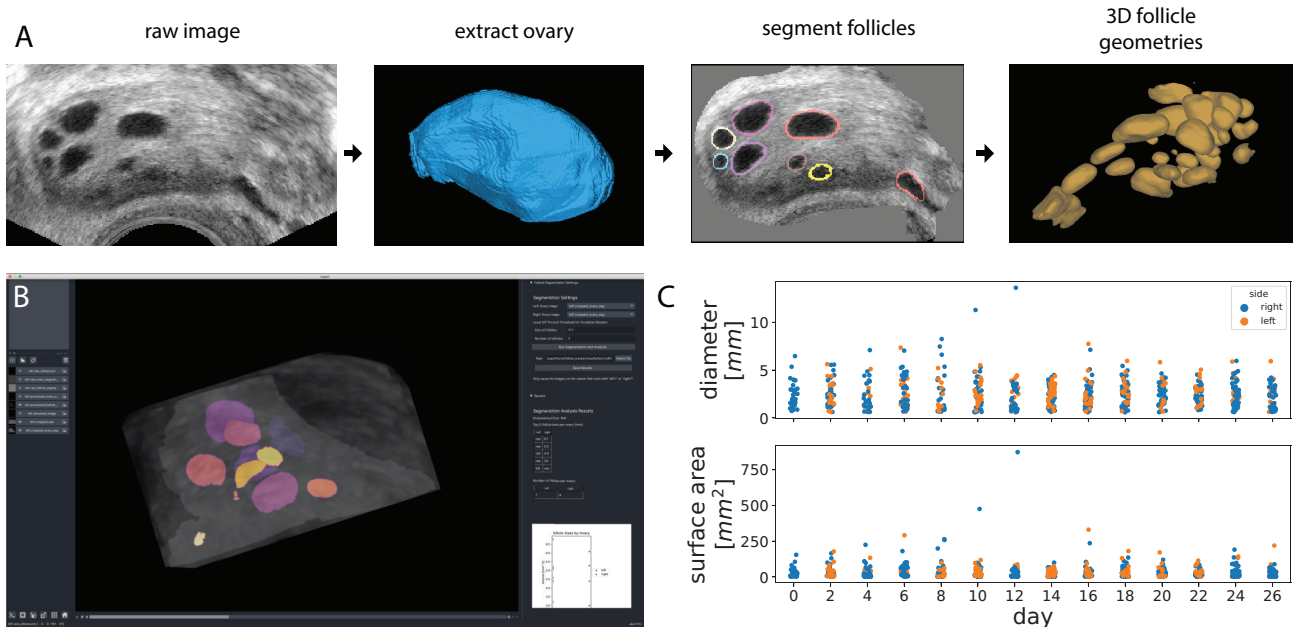
Figure 1: (A) Overview of the segmentation pipeline. **(B)** Screenshot of the FollicleTracker graphical user interface. **(C)** Example logitudinal morphometry of ovarian follicles.

truth masks. The dataset contains 94 3D transvaginal ultrasound images annotated by a medical doctor. These data were collected by vaginosonographic examination as part of the Bicycle study [8]. The data can be accessed via the ETH openBIS instance.

3.2 Measurement of clinically relevant metrics

First, we compared the number of detected follicles to the number of follicles in the ground truth. Follicle count is a clinical measure that can be used to guide IVF treatment. In particular, the antral follicle count, or the count of follicles 2-5 mm in diameter can be used to assess ovarian reserve. FollicleFinder counts are well correlated with the ground truth and have low error (1.2 follicle mean error, Figure S2). Next, we considered how accurately the size of follicles was measured. To do so, we compared the effective diameter and volumes measured on the FollicleFinder predictions to those measured in the ground truth. For both effective diameter and volume, the measurements were highly correlated (Pearson's correlation coefficient, r=0.99, Figure S3). Finally, we calculated the localization error when comparing the centroid detected by FollicleFinder to the centroid calculated from the ground truth. FollicleFinder had a median localization error of 0.15 mm, which is within one voxel (Figure S4).

4 Conclusion

In conclusion, we have developed a segmentation algorithm and software that achieves clinically-relevant and state of the art performance while only using 4% of the amount of training data. We expect that model performance can be further improved by increasing the amount of training data. By releasing both our model and dataset as an open resource, we anticipate other researchers will be able to build upon and improve segmentation algorithms and integrate the rich 3D information into clinical practice.

5 Acknowledgements

6 Funding

References

[1] H. Li, J. Fang, S. Liu, X. Liang, X. Yang, Z. Mai, M. T. Van, T. Wang, Z. Chen, and D. Ni. Cr-unet: A composite network for ovary and follicle segmentation in ultrasound images. *IEEE J Biomed Health Inform*, 24(4):974–983, 2020. ISSN 2168-2208 (Electronic) 2168-2194 (Linking). doi: 10.1109/JBHI.2019.2946092.

[2] X. Liang, J. Fang, H. Li, X. Yang, D. Ni, F. Zeng, and Z. Chen. Cr-unet-based ultrasonic follicle monitoring to reduce diameter variability and generate area automatically as a novel biomarker for follicular maturity. *Ultrasound Med Biol*, 46 (11):3125–3134, 2020. ISSN 1879-291X (Electronic) 0301-5629 (Linking). doi: 10.1016/j.ultrasmedbio.2020.07.020.

[3] N. Raine-Fenning, K. Jayaprakasan, and J. Clewes. Automated follicle tracking facilitates standardization and may improve work flow. *Ultrasound Obstet Gynecol*, 30(7):1015–8, 2007. ISSN 0960-7692 (Print) 0960-7692 (Linking). doi: 10.1002/uog.5222.

[4] N. Noor, C. P. Vignarajan, N. Malhotra, and P. Vanamail. Three-dimensional automated volume calculation (sonography-based automated volume count) versus two-dimensional manual ultrasonography for follicular tracking and oocyte retrieval in women undergoing in vitro fertilization-embryo transfer: A randomized controlled trial. *J Hum Reprod Sci*, 13(4):296–302, 2020. ISSN 0974-1208 (Print) 1998-4766 (Linking). doi: 10.4103/jhrs.JHRS_91_20.

2

[5] M. A. Coelho Neto, A. Ludwin, A. Borrell, B. Benacerraf, D. Dewailly, F. da Silva Costa, G. Condous, J. L. Alcazar, L. Jokubkiene, S. Guerriero, T. Van den Bosch, and W. P. Martins. Counting ovarian antral follicles by ultrasound: a practical guide. *Ultrasound Obstet Gynecol*, 51(1):10–20, 2018. ISSN 1469-0705 (Electronic) 0960-7692 (Linking). doi: 10.1002/uog.18945.

[6] P. Mathur, K. Kakwani, Diplav, S. Kudavelly, and G. A. Ramaraju. Deep learning based quantification of ovary and follicles using 3d transvaginal ultrasound in assisted reproduction. *42nd Annual International Conferences of the Ieee Engineering in Medicine and Biology Society: Enabling Innovative Technologies for Global Healthcare Embc'20*, pages 2109–2112, 2020. ISSN 1557-170x.

[7] Mangal Prakash, Tim-Oliver Buchholz, Manan Lalit, Pavel Tomancak, Florian Jug, and Alexander Krull. Leveraging self-supervised denoising for image segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 428–432, 2020. doi: 10.1109/ISBI45749.2020.9 098559.

[8] Marieke Biniasch, Ruediger Paul Laubender, Martin Hund, Katharina Buck, and Christian De Geyter. Intra- and inter-cycle variability of anti-mÜllerian hormone (amh) levels in healthy women measured with elecsys amh plus immunoassay: The bicycle study. *Fertility and Sterility*, 114(3), 2020. ISSN 00150282. doi: 10.1016/j.fertnstert.2020.08.624.

FollicleFinder: automated three-dimensional segmentation of human ovarian follicles
Supplementary Information

Kevin Yamauchi[1, 2, +], Marieke Biniasch[3, +], Leopold Franz[1, 2], Harold Gómez[1, 2], Christian De Geyter[3], and Dagmar Iber[1, 2, *]

[1]*Department of Biosystems Science and Engineering, ETH Zürich, Mattenstrasse 26, 4058 Basel, Switzerland*
[2]*Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*
[3]*Reproductive Medicine and Gynecological Endocrinology (RME), University Hospital, University of Basel, Vogesenstr. 134, 4031 Basel, Switzerland*
[+]*contributed equally*

1 Methods

1.1 Ethical Statement

The study was presented to and approved by the local ethics review board (EKNZ ID 2016-01824). All activities were conducted in accordance with the Declaration of Helsinki of 1975 (revised 2013), the International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Good Clinical Practice guidelines and current national guidelines. All participants provided written informed consent.

1.2 Image Acquisition

Transvaginal ultrasound images of both ovaries were acquired in 50 women every second day during two natural cycles by a medical doctor using a GE Healthcare Voluson E8 + RIC5-9-D ultrasound imaging system.

1.3 Image Export and Training Dataset

The 3D images were exported in the DICOM format using GE 4D View software, and converted to tiff images. For training and testing, a clinical expert hand-segmented 94 ovaries using 3D slicer (`https://www.slicer.org`, [9]).

1.4 Pre-processing

To prepare the ultrasound images for processing, we first rescaled the images to 0.157288 mm x 0.157288 mm x 0.157288 mm voxels. Then we performed self-supervised denoising using the aydin Noise2SelfFGR restoration function [10]. Finally, we replaced the padding voxels with the mean intensity of the image. The padding voxels are the voxels added to yield a rectangular image from cone shaped detection volume.

1.5 3D UNet

We segmented the ovary using a 3D UNet [11] and extracted the voxels corresponding to the ovary. Then, we used a second 3D UNet to segment the follicles from the extracted image of the ovary. Finally, we identified individual follicles via connected component analysis.

1.6 Training ovary segmentation

We trained the ovary segmentation network using the ADAM optimizer. We set the initial learning rate to 0.0004 and the weight decay to 0.00001. We reduced the learning rate by a factor of 5 after no improvement of 10 validation runs. We set the batch size to 1 and validated after 1000 iterations. We used the intersection over union as our validation metric. To augment our training data, we applied random flips, random 90 degree rotations, random rotations ($\pm$ 30 degrees), and random elastic deformations. We stopped training when the learning rate was reduced to below 1E-6.

1.7 Training follicle segmentation

We trained the follicle segmentation network using the ADAM optimizer. As an auxillary task, the network predicted the follicle boundaries in addition to the follicles themselves. We set the initial learning rate to 0.0002 and the weight decay to 0.00001. We reduced the learning rate by a factor of 5 after no improvement of 10 validation runs. We used intersection over union as our validation metric. We set the batch size to 1 and validated after 1000 iterations. To augment our training data, we applied random flips, random 90 degree rotations, random rotations ($\pm$ 30 degrees), and random elastic deformations. We stopped training when the learning rate was reduced to below 1E-6.

1.8 Volume measurement and surface area measurements

We measured the volume of the segmented follicles with the scikit-image[12] regionprops_table function. To measure the surface area of the of the follicles, we generated meshes from the segmented objects. First, we generated a mesh using the scikit-image marching cubes algorithm. Then we cleaned up the mesh (i.e, made it watertight) using pymeshfix [13]. Finally, we smoothed the mesh using the trimesh[14] filter_taubin function (a Laplacian filter). Finally, we calculated the surface area using the Trimesh.area method().

1.9 FollicleTracker Graphical User Interface

We developed the graphical user interface (GUI) as a plugin for napari [15], a GPU-accelerated nD data viewer. Within the plugin, users can explore the data in both 2D and 3D. Our plugin is structured in three parts: IO[1], GUI and Model (Figure S1).

Napari allows users to import 3D images in various formats by drag-and-dropping them into the GUI. This includes images saved in the TIFF or standard DICOM format. Note, that some

---

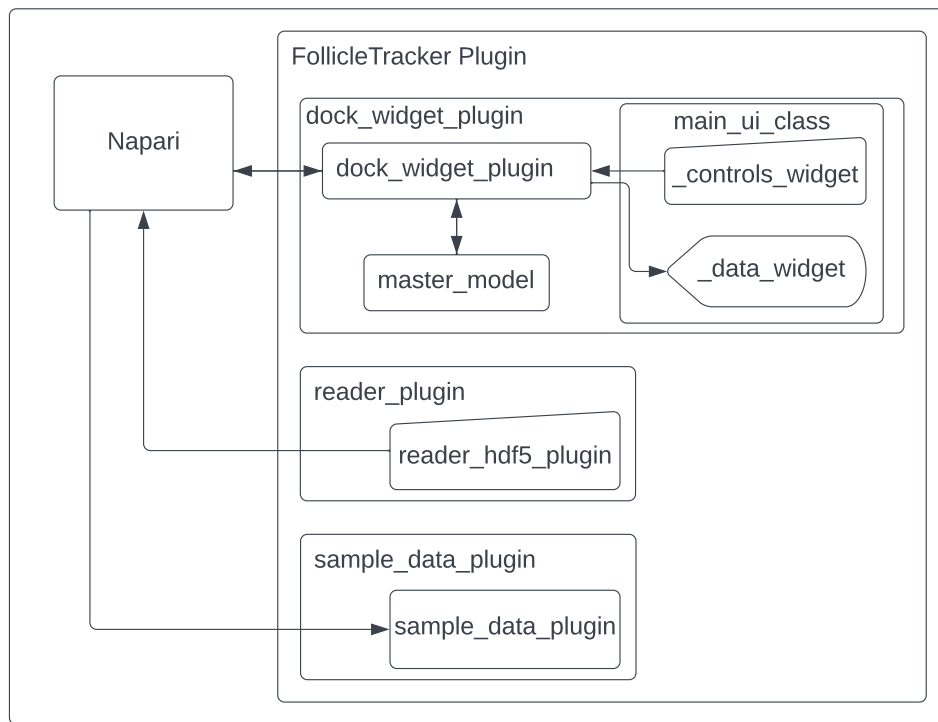[1]Input-Output: Functions defining how data is loaded and saved.

Figure S1: FollicleTracker is implemented as a napari plugin and has a modular architecture.

GE ultrasound machines have a feature to save 3D ultrasound images as "DICOM", but then do not save them using the universal DICOM standard but a proprietary image file format [16]. The plugin allows users to save images as hdf5 files with separate datasets for the "left" and "right" ovaries. The plugin saves the follicle segmentations and all intermediate results from each side. These files can then be opened in napari using the plugin.

Through the controls on the napari plugin GUI, users can start the segmentation of ovarian ultrasound images and morphological feature analysis, such as volume and surface area of all detected follicles. The morphological measurements of the follicles in each ovary are then displayed in table similar to existing reproductive electronic health record programs. Users can select a morphological feature to plot so they can understand the distribution of follicle sizes. The default morphological feature is volume.

The model that processes and analyses the input ovary images is implemented separately from the GUI to ensure all variables, images and settings only have one state. The model computation is run in a separate thread than the GUI, so that the viewer remains responsive during segmentation and analysis. Any intermediate results are directly send back to the user interface thread and added to the viewer. The intermediate results are the de-noised image, the initial ovary segmentation, the processed ovary segmentation, the initial follicle segmentation and the processed follicle segmentation. The final results are the morphological follicle measurements.

1.10 Data Availability

All images and corresponding ground truth annotations are available via the ETH openBIS instance.

1.11 Installation and Code Availability

All source code used to produce the quantitative plots is released under the 3-clause BSD license, and is publicly available as a git repository. Installation instructions are available on the readme of each code repository.

- FollicleFinder: `https://git.bsse.ethz.ch/iber/ovary-analysis`

- FollicleTracker: `https://git.bsse.ethz.ch/iber/follicle-tracker`

- Source code for publication figures: `https://git.bsse.ethz.ch/iber/Publications/2021_yamauchi_follicle_analysis`

2 Quantitative Benchmarking of Ovarian Follicle Segmentation

To validate our segmentation algorithm, we performed a 10-fold cross-validation with each fold using a 80-10-10 train-validate-test split. We evaluated the precision (true positives / (true positives + false positives)) and recall (true positives / (true positives + false negatives)). We classify the error modes as described in Greenwald *et al.*[17]. True positive is assigned when the predicted object has an intersection-over-union greater than 0.5 with a matching object in the ground truth. A false positive is when a predicted object does not match any ground truth objects. A false negative is when a ground truth object does not match any predicted objects. A merge error is when one prediction object matches multiple ground truth objects. A split error is when multiple predicted objects match to a single ground truth object.

To benchmark the performance, we compared to the results of a previous study [18], which reports results from a neural

5

network architecture called CR-Unet. CR-Unet is network architecture that incorporates a spatial recurrent neural network unit between the encoder and decoder arms of a Unet and has top end performance of recently-published algorithms (Table 1). A more recent study has achieved a recall of 0.91 for 4-12 mm diameter follicles with a network architecture called S-Net [19], but the authors do not report the precision, so we did not include it in the benchmarking. Similarly, a direct comparison with CR-Unet is difficult because authors did not report the method used to call detection events for calculating precision and recall.

Table 1: Benchmarking of FollicleFinder performance. Prec. is precision and Rec. is recall. CR-Unet benchmarks are taken from [18].

	2-5 mm		5-10 mm		>10 mm	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
Ours	0.79	0.79	0.92	0.93	1.00	0.95
CR-Unet	0.96	0.85	0.99	0.85	0.88	0.95

For large follicles (>10 mm diameter), FollicleFinder achieves 14% higher precision and similar recall when compared to CR-Unet. For medium follicles (5-10 mm diameter), FollicleFinder achieves 9% higher recall and 7% lower precision. Finally, for small follicles, CR-Unet has higher precision and recall (18% and 7%). As we explore in the following section, in spite of the difference in precision, FollicleFinder is still able to make accurate clinical measurements (e.g., counts and sizes). In summary, when compared to CR-Unet, FollicleFinder achieves better performance on large follicles, similar performance on medium follicles, and slightly worse performance. Finally, we note that we have trained on far fewer images than CR-Unet (90 versus 2509 images) and we anticipate FollicleFinder's performance will be further improved as we incorporate more training data is collected. Taken together, in spite of training on less data than CR-Unet, FollicleFinder achieves similar performance.

3 Clinically-relevant Measurements

To characterize the performance of FollicleFinder on clinically-relevant metrics, we compared the values obtained from our predictions from fold 0 of our 10-fold cross validation to the ground truth values. In particular, we compared the follicle counts (Figure S2), follicle sizes (Figure S3), and follicle localization (Figure S4). We see that for all measurements, the FollicleFinder values closely match ground truth.
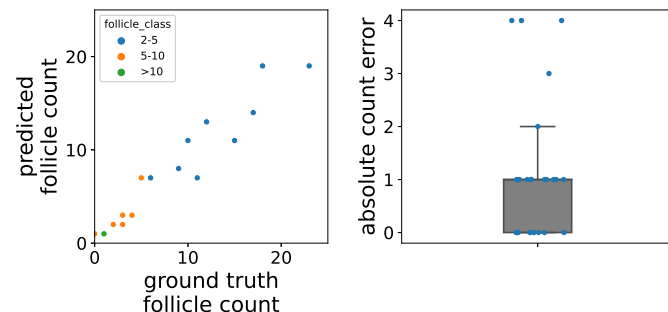
Figure S2: (left) Predicted follicle counts are well-correlated with ground truth for all follicle classes. **(right)** Mean absolute count error is 1.2 follicles across all follicle classes.
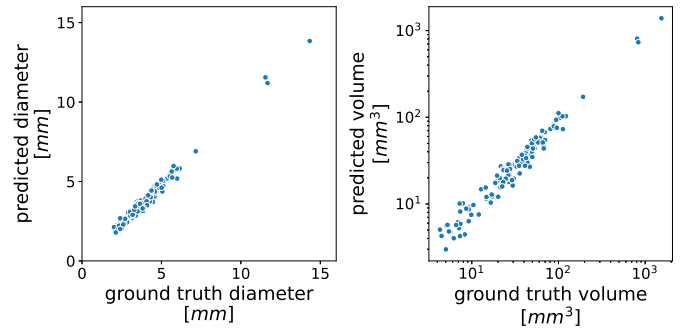
Figure S3: (left) Measured effective diameters from the predictions are well-correlated with ground truth measurements (Pearson's R=0.99) **(right)** Measured follicle volume from the predictions are well-correlated with ground truth measurements (Pearson's R=0.99). Horizontal and vertical axes are log scale.
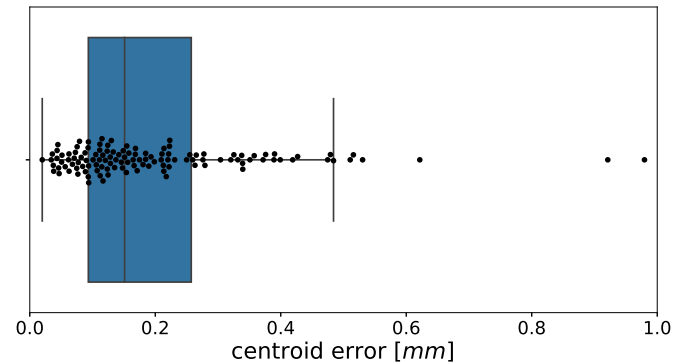
Figure S4: Median follicle localization error is 0.15 mm, which is less than one voxel.

Supplementary References

[9] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis. 3d slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*, 30 (9):1323–41, 2012. ISSN 1873-5894 (Electronic) 0730-725X (Linking). doi: 10.1016/j.mri.2012.05.001.

[10] Ahmet Can Solak, Loic A. Royer, Abdur-Rahmaan Janhangeer, and Hirofumi Kobayashi. royerlab/aydin: v0.1.12, January 2022. URL https://doi.org/10.5281/zenodo.5908198.

[11] A. Wolny, L. Cerrone, A. Vijayan, R. Tofanelli, A. V. Barro, M. Louveaux, C. Wenzl, S. Strauss, D. Wilson-Sanchez, R. Lymbouridou, S. S. Steigleder, C. Pape, A. Bailoni, S. Duran-Nebreda, G. W. Bassel, J. U. Lohmann, M. Tsiantis, F. A. Hamprecht, K. Schneitz, A. Maizel, and A. Kreshuk. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *Elife*, 9, 2020. ISSN 2050-084X (Electronic) 2050-084X (Linking). doi: 10.7554/eLife.57613.

[12] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

[13] Marco Attene. A lightweight approach to repairing digitized polygon meshes. *The visual computer*, 26(11):1393–1406, 2010.

[14] URL `https://github.com/mikedh/trimesh`.

[15] Nicholas Sofroniew, Talley Lambert, Kira Evans, Juan Nunez-Iglesias, Grzegorz Bokota, Gonzalo Peña-Castellanos, Philip Winston, Kevin Yamauchi, Matthias Bussonnier, Draga Doncila Pop, Ziyang Liu, ACS, Pam, alisterburt, Genevieve Buckley, Andy Sweet, Lorenzo Gaifas, Jaime Rodríguez-Guerra, Lukasz Migas, Volker Hilsenstein, Jordão Bragantini, Gregory R. Lee, Hector, Jeremy Freeman, Peter Boone, Alan R Lowe, Christoph Gohlke, Loic Royer, Andrea PIERRÉ, and Hagai Har-Gil. napari/napari: 0.4.12rc2, October 2021. URL `https://doi.org/10.5281/zenodo.3555620`.

[16] Pádraig Looney, Gordon Stevenson, and Sally Collins. 3d ultrasound file reading and coordinate transformations. *Journal of Open Source Software*, 4:1063, 01 2019. doi: 10.21105/joss.01063.

[17] Noah F Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Thomas Dougherty, Christine Camacho Fullaway, Brianna J McIntosh, Ke Xuan Leow, Morgan Sarah Schwartz, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology*, pages 1–11, 2021.

[18] Haoming Li, Jinghui Fang, Shengfeng Liu, Xiaowen Liang, Xin Yang, Zixin Mai, Manh The Van, Tianfu Wang, Zhiyi Chen, and Dong Ni. Cr-unet: A composite network for ovary and follicle segmentation in ultrasound images. *IEEE journal of biomedical and health informatics*, 24(4):974–983, 2019.

[19] Praful Mathur, Karan Kakwani, Srinivas Kudavelly, GA Ramaraju, et al. Deep learning based quantification of ovary and follicles using 3d transvaginal ultrasound in assisted reproduction. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2109–2112. IEEE, 2020.