

A weak-labelling and deep learning approach for in-focus object segmentation in 3D widefield microscopy

Rui Li

Helmholtz-Zentrum Dresden- Rossendorf e. V. (HZDR)

Mikhail Kudryashev

Max Delbrück Center for Molecular Medicine in the Helmholtz Association

Artur Yakimovich (✉ a.yakimovoich@hzdr.de)

Helmholtz-Zentrum Dresden- Rossendorf e. V. (HZDR)

Article

Keywords: weak-labeling, deep neural network, widefield microscopy, surrogate model

Posted Date: December 20th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2362531/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Three-dimensional information is crucial to our understanding of biological phenomena. The vast majority of biological microscopy specimens are inherently three-dimensional. However, conventional light microscopy is largely geared towards 2D images, while 3D microscopy and image reconstruction remain feasible only with specialized equipment and techniques. Inspired by the working principles of one such technique - confocal microscopy, we propose a novel approach to 3D widefield microscopy reconstruction through semantic segmentation of in-focus and out-of-focus pixels. For this, we explore a number of rule-based algorithms commonly used for software-based autofocus and apply them to a dataset of widefield focal stacks. We propose a computation scheme allowing the calculation of lateral focus score maps of the slices of each stack using these algorithms. Furthermore, we identify algorithms preferable for obtaining such maps. Finally, to ensure the practicality of our approach, we propose a surrogate model based on a deep neural network, capable of segmenting in-focus pixels from the out-of-focus background in a fast and reliable fashion. The deep-neural-network-based approach allows a major speedup for data processing making it usable for online data processing.

Introduction

Gaining insights into biological processes in three dimensions (3D) is vital for understanding biological mechanisms, as well as improving translation between *in vitro* and *in vivo*¹. However, following the historical concept of microscopy, the vast majority of common techniques used in laboratories remain focused on acquiring 2D images. Among other existing techniques, confocal laser scanning microscopy (CLSM)² remains the most widely used to capture 3D information about biological entities. During CLSM imaging, the pinhole present in the optical path filters out the scattered light, ensuring all the captured intensities are in-focus. This process is repeated for each focal plane as the acquisition moves along the axial axis³. In this way, CLSM reconstructs the clear 3D models of biological entities slice-by-slice. However, long imaging time, high requirements for trained personnel or facility, as well as high equipment costs make CLSM less accessible than traditional widefield microscopy. At the same time, widefield microscopes and binoculars using transmission light are inexpensive, abundant in laboratories across the world, and require minimal training or specimen labeling.

The stepwise acquisition of a larger translucent specimen in 3D may also be performed using widefield microscopy through sequential alteration of the focal plane. Unlike optical sectioning in CLSM³, when widefield microscopy is directly applied to 3D specimens, all light - both in-focus and scattered - contributes to the formation of an image⁴. This introduces the noise from the other focal planes to the recorded images; reduces the contrast information and decreases the quality of 3D reconstruction. An ability to separate in-focus and out-of-focus parts of each focal slice would not only allow for precise 3D reconstruction of the specimen but also make such imaging modality quantitative. This would be possible through a clear separation of background and foreground pixels.

Separation of in-focus and out-of-focus microscopy images without reference may be achieved algorithmically using image-based (passive) autofocusing⁵. In-focus images of specimens are often rich with context patterns which makes neighbouring pixels in the in-focus image less autocorrelative compared to the out-of-focus image. This, in turn, leads to greater contrast, wider ranges of intensity, and sharper contour information in the in-focus images, making it possible to evaluate images' in-focus status. Many passive autofocusing algorithms, for example, Vollath⁶, Brenner⁷, and Variance⁸ are designed based on this concept. Yet, most of them are used to evaluate the in-focus status of the whole images or slices in a focal stack. Furthermore, most of these algorithms are relatively slow to compute. This may be addressed using surrogate machine learning models (ML). For example, models based on deep neural networks (DNNs) are recently showing great promise in a plethora of microscopy applications⁹⁻¹¹. With the development of DNNs has gained more popularity in various computer vision tasks (image classification (Simonyan & Zisserman 2015), segmentation¹² and object detection, etc.). Through convolution operations, the DNN models extract features from images on multiple scales. These diverse features enhance the accuracy of vision tasks. Specifically, Waller et al. have reviewed the potential of DNNs in 3D microscopy. Chen et al. proposed a 3D convolutional DNN and validated the algorithm for medical image segmentation. Yet, DNN models are known to perform best when used in a supervised ML setting, which would require manual data annotation.

To eliminate the need for manual annotation, here we created a novel DNN model for widefield focal sectioning through in-focus pixel segmentation, trained using algorithmically derived Ground Truth (GT). The widefield microscopy image dataset we employed contains *in vivo* transmission light focal stack micrographs of *Danio rerio* (zebrafish) larva's head^{13,14}. To obtain the GT we investigated 8 algorithms commonly used in autofocusing tasks including Brenner¹⁵, Variance¹⁶, Tenengrad⁵, etc. These algorithms obtained focus-score maps of each slice using a sliding window approach and maxima Z-projection. Next, we compared the sensitivity of the focus measurement algorithms using the output of focus measurement algorithms as a focus score. To ensure that the output of these algorithms represents a good proxy for in-focus pixels we compared these outputs to a manually annotated test subset (manual segmented GT). We concluded that five detectors - Variance, Vollath, Standard Deviation¹⁶, Brenner, and Laplacian - were superior to others in detecting changes of focus planes. After the assessment of the target segmentation qualities, we concluded that the Standard Deviation (std) detector outperformed the others in evaluating the focus status of images. Next, we adopted a DNN model with the U-Net¹⁷ architecture to obtain a surrogate model speeding up the previous focal score computation process.

Our results suggest that using conventional algorithms as weak labels, DNN may be employed as a surrogate model for the detection of in-focus pixels in stable quality. This solution separates the in-focus pixels from image stacks of widefield microscopy, enables the optical sectioning in a digital manner, and reveals the 3D information of the specimen. This, in turn, can make *in vivo* 3D imaging widely accessible for laboratories with modest funding.

Methods

Dataset source and ethics declaration

The dataset of this work comes from the observation of *in vivo* zebrafish (*Danio Rerio*) larvae heads recorded as focal stacks using a stereomicroscope (Leica M205FA; Leica Microsystems, Nussloch GmbH, Nussloch, Germany). All images were obtained at x130 magnification with a 1x objective. The lateral resolution was 0.79 μm per pixel. To obtain a focal stack, twenty Z-planes were captured covering a total axial distance of 171 μm at 8.55- μm intervals and saved as TIFF stacks¹³. In each file, the target is in the middle of the view field. As stated in¹³, the animal experiments were performed according to the Animals (Scientific Procedures) Act of 1986 and approved by the Home Office (project licenses PPL P84A89400 and P4E664E3C).

An algorithm for segmentation of in-focus pixels

The focus measurement algorithms⁸ evaluate the in-focus status through the pixel value patterns in images. Such algorithms give the highest focus score for in-focus pixel intensities. The focus score decreases when the focal plane changes. While a great number of focus detection algorithms have been proposed in the literature (reviewed in previous paper¹⁶, including an autofocus algorithm selection). In this work, we investigate the 8 most widely used algorithms. The algorithms can be classified into three categories based on their design

a) Derivative-based algorithms

These algorithms assume that in-focus images contain more high-frequency content. Therefore, the pixel intensity changes stronger than in out-focus images. These intensity variations can be recognized by computing the derivatives of pixel values. We selected the five most promising algorithms below.

1) Brenner gradient⁷. This algorithm computes the first-order derivation between the target pixel and its neighbours. Eq. 1 is presented below with $(i(x+1, y) - i(x, y))^2 \geq \theta$. Here the θ is a manually defined threshold.

$$F_{Brenner} = \sum_{Height} \sum_{Width} (i(x+2, y) - i(x, y))^2$$

1

2) Tenengrad¹⁸. This algorithm derives from the Sobel operator by detecting the contour in both horizontal and vertical directions ($S_x(x, y)$ and $S_y(x, y)$).

$$F_{Tenengrad} = \sum_{Height} \sum_{Width} S_x(x, y)^2 + S_y(x, y)^2$$

2

3) Laplacian¹⁹. This algorithm convolves the image with Laplacian operators and sums the values.

$$F_{SML} = \sum_{Height} \sum_{Width} |L_x(x, y)| + |L_y(x, y)|$$

3

4) Sum Modulus Difference (SMD) ²⁰. This algorithm calculates the first-order derivation between pixels and neighbours.

$$SMD_x = \sum_x \sum_y |I(x, y) - I(x, y - 1)| \quad SMD_y = \sum_x \sum_y |I(x, y) - I(x + 1, y)|$$

$$F_{SMD} = SMD_x + SMD_y \quad (4)$$

5) Vollath ⁶. The Vollath algorithm computes the derivation between pixel intensity in both horizontal and vertical directions.

$$F_{Vollath} = \sum_{i=1}^{M-1} \sum_{j=1}^N g(i, j) * g(i + 1, j) - \sum_{i=1}^{M-1} \sum_{j=1}^N g(i, j) * g(i + 2, j)$$

5

b) Statistic-based algorithms

These algorithms distinguish the in-focus status by statistical features of images (variance, standard derivation, correlation, etc.). Compared to the derivative-based algorithms, such algorithms are more stable to noise. The candidates in this work are below.

6) Standard deviation ²¹. When the images are in-focus, the contrast of pixel values is high. This can be detected by calculating the standard deviation.

$$F_{corr-stddev} = \sum_{Height} \sum_{Width} i(x, y) * i(x + 1, y) - H * W * \mu^2$$

6

7) Variance ⁸. The power operation amplifies the variation differences from pixel values.

$$F_{corr-stddev} = \frac{1}{H * W} \sum_{Height} \sum_{Width} (i(x, y) - \mu)^2$$

7

c). Histogram-based algorithm.

These algorithms assess the patterns of intensity distributions. This work inspects one histogram-based algorithm outlined below.

8) entropy algorithm ²². This method assumes the in-focus images contain more information about the target. Thus, it shows higher entropy scores. p_i is the probability of pixels with intensity i .

$$F_{entropy} = - \sum_{Intensities} p_i * \log_2 (p_i)$$

8

Sliding window scheme

To obtain the focus-score on fractions of high-resolution micrographs, we employed a scanning scheme^{23,24}. This algorithmic scheme - illustrated in Fig. 1a below - changes the scanning window sizes, slides windows in various step widths (strides), and segments the in-focus pixels based on the previously proposed focus score. Specifically, we computed the focus score of pixels within the scanning window. A perception window slides across the images with various stride parameters and extracts the pixel intensities within windows. The focus measurement algorithm then evaluates the intensities and outputs their focus scores.

Ranking of the focus score algorithms

Each widefield microscopy dataset contained twenty Z-slices - ranging from in-focus to out-of-focus. Besides the changes of focus status in the lateral plane (Fig. 1a), the focal plane differs also along the axial direction. We proposed the scanning scheme in Fig. 1b. Applied to the different slices, the eight in-focus segmentation algorithms above evaluated the pixels and outputted a focus score for every slice. Since each slice corresponds to different focal planes, the focus scores varied from each other. This makes it possible to distinguish the in-focus slice from the stacks in the axial direction. The higher the focus scores difference between slices, the better the algorithms can recognize the out-of-focus slices.

Image-wide focus map

These scores served for the generation of the focus-score patch (the same size as the scanning window). This process was repeated on every slice of image stacks in both stride directions (vertical and horizontal). It produces a focus-score map of the scanned image. By changing the stride parameters, this scanning generates multiple focus-score maps with different perceptions. Since the in-focus pixels gain higher scores in every scanning, their focus scores peak in the stacks. To obtain the distinct changes for each scanning, we amplify the difference between in-focus and out-focus pixels through the maxima Z-projections of the focus-score stack. This operation preserves the highest focus score for patches. As the scanning repeats, the focus score of in-focus pixels increases more steeply than the out-focus pixels. This makes it possible to distinguish the in-focus pixels from the images. To diversify the perceptions, the strides parameters and size of windows can be assigned with multiple values as presented in Fig. 2. This compresses the bias introduced by the scanning parameters configurations.

Deep neural network architecture

This work proposes a 7-layers symmetric U-Net model with a 3-layer encoder and decoder structure (C256-C128-C64-C32-DC64-DC128-DC256). This U-Net segments the in-focus pixels from the widefield microscopy images end-to-end. This bypasses calculating the computationally expensive focus score acting as a surrogate model. For the training dataset, we use the previous focus score maps as GT masks. The widefield microscopy image paired with the corresponding GT masks served as input. The U-Net model learns the transfer between raw widefield images and GT masks directly. These GT masks

serve as references for in-focus segmentation. After training, the model translates the widefield microscopy image stacks into corresponding 3D pixel information.

Hardware used and the training time

The GPU calculations for this work were performed on a Tesla V100, and the 8 rule-based algorithms were run on an AMD Rome core. The training speed for our DNN-based solution is 1s/epoch on average resulting in a training time of 8.5 min for 500 epochs. This training process can be around 20 min for the GPU of a consumer GPU (Nvidia GTX 30xx).

Results

Widefield focal stack dataset

To develop an approach for in-focus region detection, we have employed a published dataset of *Danio rerio* (zebrafish) *in vivo* widefield microscopy¹³. In this dataset, the fraction of the head of the zebrafish is located in the middle of the field of view. Each observation consists of a stack of 20 images taken in different focal planes (focal stack). The last slice (No. 19) of stacks contains mostly in-focus pixels, while most of the pixels in the first slice (No. 0) are out-of-focus. The remaining slices contain a mixture of in-focus and out-of-focus signals (see Methods for details).

Rule-based in-focus region detection

The widefield microscopy datasets of bulk objects, similar to the one we employed here, often contain in-focus and out-of-focus lateral regions in each slice of the focal stack. These regions change from slice to slice as the focal plane goes through the bulk of the specimen. To distinguish the regions of the slice which are in-focus from those that are out-of-focus, we have explored algorithms typically used for focal plane detection in the axial direction. For each subregion (see sliding window scheme in the Methods section) in each slice of the focal stack we have computed a score corresponding to the focal plane detection algorithm (Fig. 1). Specifically, we compared the following 8 algorithms: Brenner, Tenengrad, Laplacian, SMD, Vollath, Std, Variance, Entropy. Figure 1a illustrates the scanning results from one image (slice). To preserve the homogeneity of the original image, this work uses a square perception window (see Methods section). As presented in the left part of the panel, the perception window slid in the same step size in both horizontal and vertical directions. We tested the following windows sizes: 64, 128, and 256. We tested the strides (step sizes) 16, 32, 64, and 128 in this experiment. We obtained a focus-score map for each slice of the focal stack by reassembling them into a stack and maxima Z-projecting them (see Methods).

As presented in the focus-score map in the right part of Fig. 1a, the brightness indicates a high focus score. We noted that the smaller the perception window was, the more detailed the scanning result was. Yet, a small perception window failed to show the low-frequency information (the global features). For example, the first row extracted only contour information, while the other preserved more global

information. Conversely, the large window lost the high-frequency signal (details of images) during scanning leading to undesired results. Notably, between the second and third rows, the third row failed to capture the detailed contour information. Therefore, the balance between low- and high-frequency signals, the window size 128 proved a more appropriate choice for both global features and local details. Next, we examined the stride parameters for this optimal perception window size (Fig. 1a, second row). We noted that the smaller the strides correspond to smoother the final focus scores in the map. A smooth focus-score map indicates the structure details of the zebrafish (eye contour, body components). Conversely, bigger stride steps allow for retaining more global information. This prevents the focus-score map scanning from turning into a simple contour detection method. We further noted that all stride parameters contributed valuable detailed information at various levels. Therefore, a better scanning process should contain multiple stride values to preserve both high-frequency information and local details.

To determine the appropriate focus metric for each region, we applied the eight described rule-based algorithms on the widefield microscopy dataset. The best focus metric was expected to distinguish images on varied focal planes continuously in the axial direction. The in-focus region should score the highest value, while the out-of-focus region should rank at the bottom. With this in mind, we have measured the outputs of each algorithm in comparison to the distance from the perfect focus of a region (defocus). Figure 1b illustrates the sensitivity detection to the focal plane changes.

We noted that five (Variance, Vollath, Std, Brenner, and Laplacian) out of the eight algorithms detected the focal plane changes successfully - from slice No. 0 to slice No. 19. Interestingly, the Vollath algorithm recognized the difference between in-focus images and out-focus images. However, it failed to detect the changes continuously in the middle slices of the stacks. In these slices, the amount of in-focus pixels was visually comparable to the amount of out-of-focus pixels (mixed-focus slice). However, the Vollath score varied only slightly since the fifth slice (see Fig. 1b). In Table 1, we evaluated the time consumption for all eight candidates. Compared to the other candidates, the Brenner and Vollath algorithms were much more computationally expensive for the same images (16 minutes/ slice vs. less than 1 minute from other candidates). Nonetheless, the results were marginally better than Laplacian. Therefore, we concluded that the three algorithms - Variance, Std, and Laplacian outperform other algorithms in sensitivity along the axial direction, as well as in processing time.

Table 1

The time consumption comparison between 9 candidates for in-focus pixels segmentation. To obtain measurements, we evaluated the computation time for segmentation on the whole stack (all 20 slices of one example stack).

Candidates	Computation time (s)
Standard deviation (std)	203
Variance	199
Laplacian	270
Tenengrad	1212
SMD	$> 2.4 \cdot 10^4$
SMD2	$> 2.4 \cdot 10^4$
Brenner	$1.92 \cdot 10^4$
Vollath	$1.2 \cdot 10^4$
DNN	0.062

Unlike focal plane detection in the axial direction, the task of detecting in-focus parts of the specimen requires focus measurement to distinguish both the in-focus status and the contour information in the lateral directions. The first row of Fig. 1c presents the in-focus status of two images. Zooming into the same patch of these images, the pixel intensities indicate varied contour information. The optimal algorithm should preserve the correct image content during the focus status detection. In the second row of Figure. 1c, we calculated the focus score map from the mixed-focus images (middle slice, mixture from both in-focus and out-of-focus pixels) with the three algorithms above. The detected contour information using the Laplacian differs from the other two. To validate the differences, we merged the focus score map with the corresponding microscopy image in the third row of Fig. 1c. The segmented contour from Laplacian appears to show relatively less detail, compared with Std and Variance.

Thresholding the focus score map with the Otsu algorithm²⁵, we obtained the focus score masks as references for in-focus pixels. This mask preserves the pixels only from the target focal plane and filters out pixels from other focal planes. In Figure. 2a, we compared the segmentations of the three focus algorithms (Laplacian, Variance, and Std) to the manual segmented GT for validation. The Laplacian shows severe inconsistency with the manual GT. Thus, we concluded that it is inferior to the other two in preserving the correct image content. To be noticed, the Variance marks the in-focus pixels in a more conservative way. This yields the loss of specimen information. We assessed the information loss of two algorithms (Variance and Std) on whole stacks in Figure. 2b. Compared to the manual GT, the Variance barely preserved the complete contour information. The Std, however, showed consistency with the

manual GT. This makes the Std the focus algorithm of choice satisfying focus sensitivity and the ability to detect the morphology of specimens.

Deep neural network surrogate for the rule-based in-focus region detection

We have shown that the focus score pipeline with the Std algorithm may generate reliable focus score maps (Fig. 3a). Furthermore, accompanied by automated thresholding algorithms, these maps allow to obtain focus masks which are comparable to the manual GT. However, this rule-based pipeline is relatively computationally expensive and requires a long time to process image stacks). To achieve similar results in an end-to-end single-step fashion, we proposed a DNN surrogate model with the U-Net-like structure¹⁷ Illustrated in Fig. 3b. To obtain the U-Net-based surrogate model, we used the binary masks captured from the output of the rule-based focus score algorithms as weak labels²⁶. This way, the model learns directly the transformation between focus score masks and raw images.

We opted for U-Net architecture as it is commonly used in biomedical image segmentation tasks. U-Net combines the convolutional neural network (CNN) and the Autoencoder (AE) like structures²⁷. As a representation learning model, the first several convolutional layers of U-Net (the encoder part) enhance channel numbers of the input images and extract the features in the AE structure. The middle convolutional layer (the bottleneck part) encodes the previous features as embedding vectors in the latent space. The last multiple de-convolutional layers (the decoder part) upsamples the embedding vectors back into the original images. Optimizing the loss between reconstructed images and inputs, the encoder and decoder learn jointly the manifold structures²⁸ of given datasets. In contrast to the traditional AE structures, the U-Net concatenates the up-sampled embedding code with the feature maps from the corresponding layers in the encoder part²⁹. This operation casts constraints on the outputs and gives the U-Net an advantage in the supervised learning tasks. This model segmented the in-focus pixels directly from the widefield microscopy images.

As illustrated in Fig. 3b, this model contains 7 convolutional layers – 3 for the encoder; 1 for the bottleneck; 3 up-sampling layers for the decoder. The loss function we chose consisted of focal loss, dice loss, and binary cross entropy summed up into a total loss. To evaluate the performance of the model, this work adopted the IoU score³⁰ as the metric. As the optimizer, we used Adam with a learning rate of 0.001. After 400 epochs of learning with a batch size of 8, the model converged to a stable value both for IoU scores and the loss. The final IoU was 0.98 and the loss was 0.05 (Supplementary Fig. 1). As presented in Table 1, this DNN model speeds up the segmentation process with 0.062 seconds for one stack. Even with the training time of 8.5 minutes in one shot, this solution is still superior to other candidates by accelerating the segmentation to at least ~ 10000 times.

Surrogate model evaluation

The trained U-Net surrogate model can directly predict the focus-score mask with the widefield microscopy images as inputs. This approach significantly simplified the segmentation compared to the previous focus-score pipeline. Figure 4a illustrates part of the segmentation results. Notably, from the completely out-of-focus slice (slice 0) to the completely in-focus slice (slice 19), the DNN model distinguished the in-focus pixels correctly. In slice 0, the model labelled the whole image as out-focus. The results were consistent with the GT mask. As the focal plane changed during optical sectioning, the image contained more in-focus pixels. The predictions of the DNN model stayed reliable. In slice 19, the model labeled correctly the whole target as in-focus. However, the prediction varied from the GT masks in certain cases (slice 9), which highlights opportunities for future improvements of this DNN model in the future.

The focus-score masks obtained from the surrogate model may be employed for in-focus pixel segmentation from the widefield microscopy images. As shown in Figure. 4b, these masks allow retaining only the in-focus part of the image. These, in turn, may be assembled into a 3D model of a specimen. Notably, this is possible by employing images obtained using widefield *in vivo* microscopy, in which unlike in CLSM, both in-focus and out-focus light contribute to the formation of the image in every image plane.

Conclusion & Discussion

Inspired by the mechanism of CLSM and software-based autofocusing algorithms, this paper proposed an approach to filter in-focus from out-of-focus regions of the image in the focal stack obtained in widefield microscopy. The latter can be obtained by altering the focal plane and scanning through bulk specimens like live zebrafish (*D. Rerio*). However, focal stacks obtained by widefield microscopy in such a manner contain a mixture of information produced by both in-focus and out-of-focus light. To select the optimal focus measurement algorithm, this work investigated eight candidates widely used in software-based focal plane detection (Vollath, Brenner, std, etc.). Our experiments showed that for the purpose of in-focus region detection and content information preservation, standard-deviation-based pipelines were optimal.

To overcome the computational costs of the rule-based pipeline, we proposed a DNN surrogate model based on U-Net architecture for in-focus pixel segmentation. This model was trained to adopt the previous rule-based segmentation results as GT. The resulting DNN model filtered out the out-of-focus signals digitally without a complex and expensive confocal setup. The segmentation results on the zebrafish dataset showed consistency with the manual segmentation GT. Compared to the previous eight candidates, the DNN model overwhelms others by the calculation speed with at least ~ 10000 times faster by in-focus segmentations, still noting that the rule-based processing was performed without GPU acceleration. Nonetheless, the speedup is impressive making virtual optical sectioning possible for *in vivo* widefield microscopy. Upon segmentation of the in-focus pixels, our DNN allows us to reconstruct 3D models of the specimen obtained from widefield imaging.

While multiple tasks in machine learning and deep learning for microscopy have been proposed in the past^{9,31–33}, no tasks for the separation of in-focus from out-of-focus images have been explored until now. We argue that our work opens an avenue to advanced image protocols, such as 3D in vivo imaging using simple and inexpensive hardware. Widefield microscopes are abundant in research and education facilities and may find new applications using approaches akin to ours. Remarkably, as the GT for our surrogate model was obtained purely programmatically, it is tempting to speculate that this approach may be useful in weak labeling and self-supervised learning^{26,34,35}.

As a possible extension of this work, a better focus-score pipeline could combine multiple focus measurement algorithms instead of only one. This could possibly enhance the quality of the segmentation. This serves as a better GT for training the DNN models. Besides, other DNN structures (pix2pix GAN, transfer learning, 3D U-Net, etc.) might bring better performance regarding segmentation accuracy.

Declarations

Acknowledgments

This work was partially funded by the Center for Advanced Systems Understanding (CASUS) which is financed by Germany's Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Culture, and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament. MK acknowledges the support of the Helmholtz Association and the Heisenberg Award from the German Research Foundation (DFG). MK was supported by Heisenberg award from the DFG (KU 3222/2-1), as well as funding from the Helmholtz Association.

Author contributions

AY, MK, and RL conceived the project and planned the experiment. RL wrote program code and performed the computational experiments. AY, MK, and RL wrote the manuscript.

Code and Data availability statement

The program code used in this work is available for use and re-use under an open-source license and can be accessed via GitHub (<https://github.com/casus/deepfocus>). The Dataset of *D. Rerio* focal stacks was previously published in¹³ and raw or additional data is available upon request from the corresponding authors¹³. As stated in¹³, the animal experiments were performed according to the Animals (Scientific Procedures) Act of 1986 and approved by the Home Office (project licenses PPL P84A89400 and P4E664E3C). Processed data necessary to reproduce this work is available via GitHub repository.

References

1. Pampaloni, F., Reynaud, E. G. & Stelzer, E. H. K. The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* **8**, 839–845 (2007).
2. Nwaneshiudu, A. *et al.* Introduction to Confocal Microscopy. *J. Invest. Dermatol.* **132**, 1–5 (2012).
3. Wilson, T. Resolution and optical sectioning in the confocal microscope. *J. Microsc.* **244**, 113–121 (2011).
4. Conchello, J.-A. & Lichtman, J. W. Optical sectioning microscopy. *Nat. Methods* **2**, 920–931 (2005).
5. Pech-Pacheco, J. L., Cristobal, G., Chamorro-Martinez, J. & Fernandez-Valdivia, J. Diatom autofocusing in brightfield microscopy: a comparative study. in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* vol. 3 314–317 vol.3 (2000).
6. Yousefi, S., Rahman, M. & Kehtarnavaz, N. A New Auto-Focus Sharpness Function for Digital and Smart-Phone Cameras. *IEEE Trans. Consum. Electron. - IEEE TRANS CONSUM ELECTRON* **57**, 1003–1009 (2011).
7. Brenner, J. F. *et al.* An automated microscope for cytologic research a preliminary evaluation. *J. Histochem. Cytochem.* **24**, 100–111 (1976).
8. Groen, F. C. A., Young, I. T. & Ligthart, G. A comparison of different focus functions for use in autofocus algorithms. *Cytometry* **6**, 81–91 (1985).
9. Buchholz, T.-O., Jordan, M., Pigino, G. & Jug, F. Cryo-CARE: Content-Aware Image Restoration for Cryo-Transmission Electron Microscopy Data. *ArXiv181005420 Cs* (2018).
10. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-convex Polygons. in vol. 11071 265–273 (2018).
11. Waller, L. & Tian, L. Machine learning for 3D microscopy. *Nature* **523**, 416–417 (2015).
12. Havaei, M. *et al.* Brain Tumor Segmentation with Deep Neural Networks. *ArXiv150503540 Cs* (2016) doi:10.1016/j.media.2016.05.004.
13. Yakimovich, A. *et al.* Mimicry Embedding Facilitates Advanced Neural Network Training for Image-Based Pathogen Detection. *mSphere* **5**, e00836-20 (2020).
14. Yoshida, N. *et al.* The zebrafish as a novel model for the *in vivo* study of *Toxoplasma gondii* replication and interaction with macrophages. *Dis. Model. Mech.* **13**, dmm043091 (2020).
15. Yazdanfar, S. *et al.* Simple and robust image-based autofocusing for digital microscopy. *Opt. Express* **16**, 8670 (2008).
16. Yu Sun, Duthaler, S. & Nelson, B. J. Autofocusing algorithm selection in computer microscopy. in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* 70–76 (IEEE, 2005). doi:10.1109/IROS.2005.1545017.
17. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) vol. 9351 234–241 (Springer International Publishing, 2015).

18. Yeo, T., Ong, S., Jayasooriah & Sinniah, R. Autofocusing for tissue microscopy. *Image Vis. Comput.* **11**, 629–639 (1993).
19. Nayar, S. K. & Nakagawa, Y. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 824–831 (1994).
20. Yao, Y., Abidi, B., Doggaz, N. & Abidi, M. Evaluation of sharpness measures and search algorithms for the auto focusing of high-magnification images. in (eds. Rahman, Z., Reichenbach, S. E. & Neifeld, M. A.) 62460G (2006). doi:10.1117/12.664751.
21. Yang, G., Gaines, J. & Nelson, B. A Supervisory Wafer-Level 3D Microassembly System for Hybrid MEMS Fabrications. *J. Intell. Robot. Syst.* **37**, 43–68 (2003).
22. Firestone, L., Cook, K., Culp, K., Talsania, N. & Preston Jr., K. Comparison of autofocus methods for automated microscopy. *Cytometry* **12**, 195–206 (1991).
23. Arasu, A. & Manku, G. S. Approximate counts and quantiles over sliding windows. in *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '04* 286 (ACM Press, 2004). doi:10.1145/1055558.1055598.
24. KrishnaKumar, A., Amrita, D. & Priya, N. S. Mining Association Rules between Sets of Items in Large Databases. **1**, 4 (2013).
25. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
26. Ratner, A. *et al.* Snorkel: Rapid Training Data Creation with Weak Supervision. *Proc. VLDB Endow. Int. Conf. Very Large Data Bases* **11**, 269–282 (2017).
27. Doersch, C. Tutorial on Variational Autoencoders. Preprint at <http://arxiv.org/abs/1606.05908> (2021).
28. Liu, H. *et al.* Latent Space Optimal Transport for Generative Models. Preprint at <http://arxiv.org/abs/1809.05964> (2018).
29. Hu, X., Naiel, M. A., Wong, A., Lamm, M. & Fieguth, P. RUNet: A Robust UNet Architecture for Image Super-Resolution. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 505–507 (IEEE, 2019). doi:10.1109/CVPRW.2019.00073.
30. Zitnick, C. L. & Dollár, P. Edge Boxes: Locating Object Proposals from Edges. in *Computer Vision – ECCV 2014* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) vol. 8693 391–405 (Springer International Publishing, 2014).
31. Arganda-Carreras, I. *et al.* Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–2426 (2017).
32. Das, D. K., Ghosh, M., Pal, M., Maiti, A. K. & Chakraborty, C. Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron* **45**, 97–106 (2013).
33. Kan, A. Machine learning applications in cell image analysis. *Immunol. Cell Biol.* **95**, 525–530 (2017).
34. Bai, W. *et al.* Self-Supervised Learning for Cardiac MR Image Segmentation by Anatomical Position Prediction. Preprint at <http://arxiv.org/abs/1907.02757> (2019).

35. Bringer, E., Israeli, A., Shoham, Y., Ratner, A. & Ré, C. Osprey: Weak Supervision of Imbalanced Extraction Problems without Code. in *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning - DEEM'19* 1–11 (ACM Press, 2019). doi:10.1145/3329486.3329492.

Figures

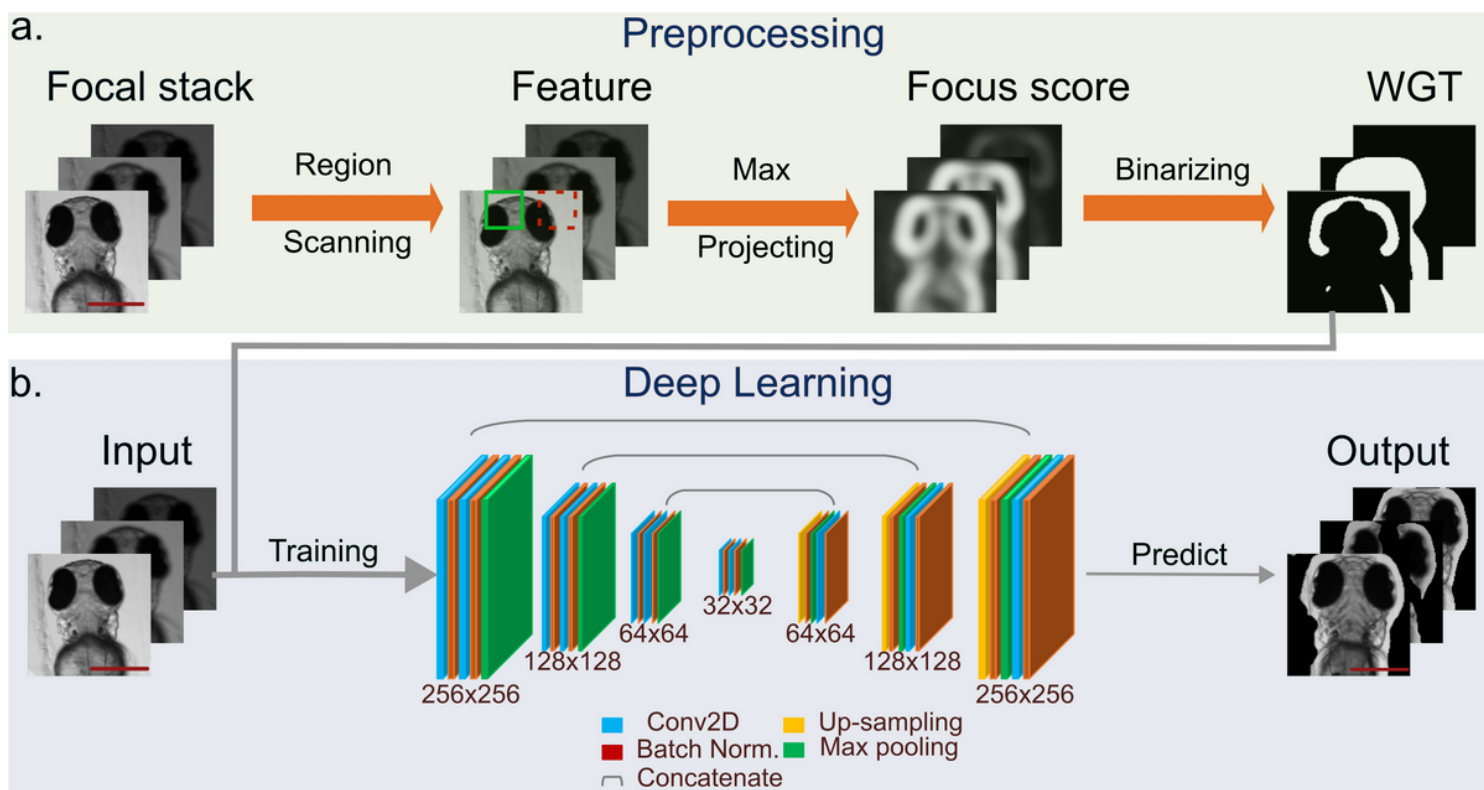


Figure 1

Sliding window scanning approach for focus score maps computation. (a) illustrates a perception window sliding across the widefield microscopy images in both x and y directions, while evaluating the in-focus status of pixels, and outputting the focus-score maps. The three rows show the scanning results with window sizes (64, 128, 256). The focus-score maps with bigger perception windows show fewer details. The three columns stand for different stride plans for specific window sizes. (b) shows the axial sensitivity of focus measurements. The image stack contains 20 images - from the out-of-focus slice 0 to the in-focus slice 19 (shown in the OX axis). Min-max-normalized focus score for each tested algorithm is shown in the OY axis. Figure legend names algorithms in the order of maximum focus score value reached. (c) presents the lateral sensitivity of focus measurements. The slices of one stack range from out-of-focus (left) to in-focus (right). (d) the top row shows an example of how the three detectors (Variance, std, and Laplacian) marked the in-focus pixels in the middle slice. This slice contains pixels from both in-focus and out-of-focus. The bottom row shows merges of the maps with microscopy images. The scale bar in all images is 500 μm .

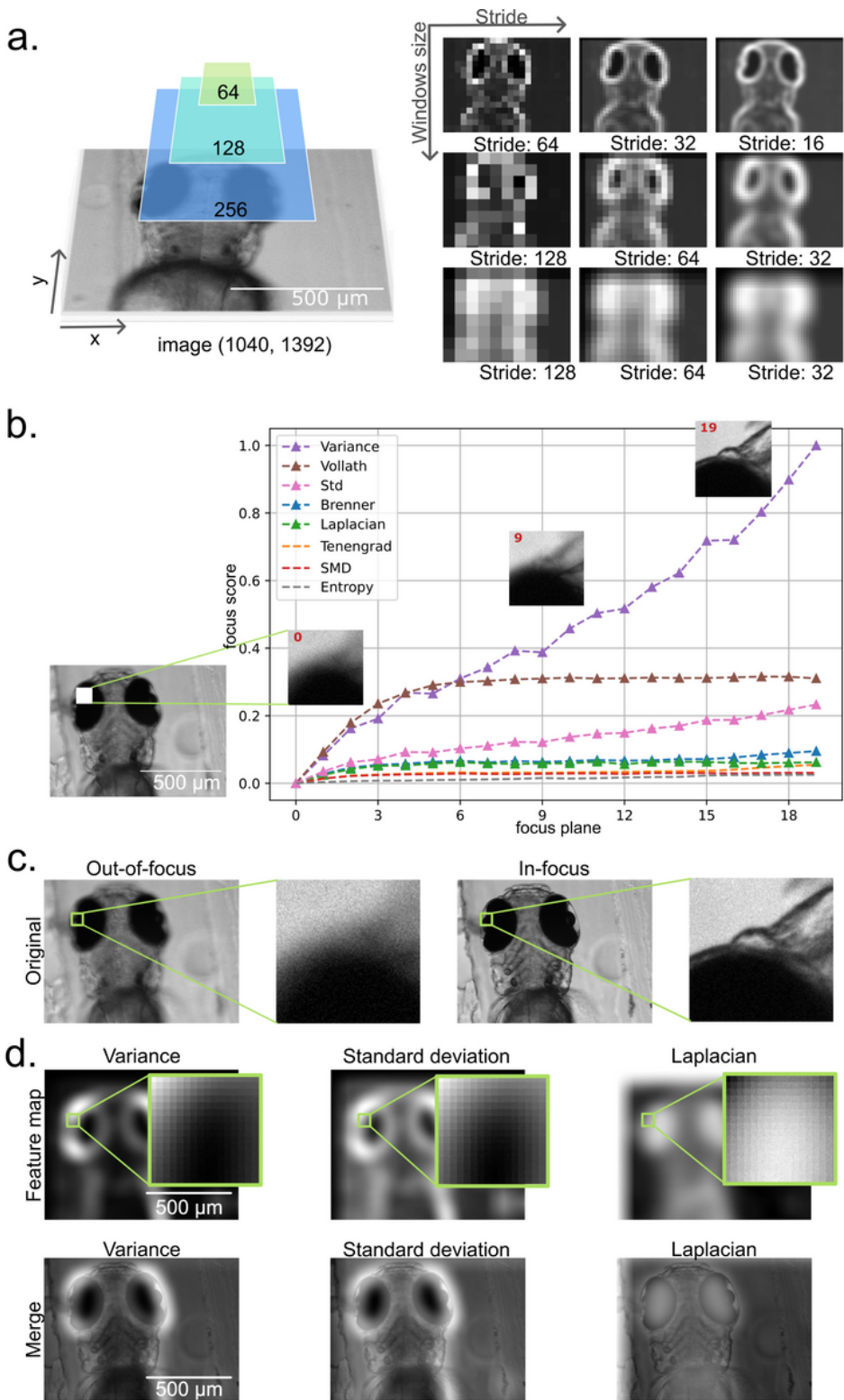


Figure 2

In-focus masks and pixel segmentation. (a) shows binary masks from the focus measurements corresponding to the respective algorithm. To obtain masks, the focus-score maps were binarized using the Otsu thresholding. Comparison with the manually marked ground truth (GT) masks is presented on the right-hand side. (b) shows in-focus masks of Variance and Standard deviation merged with

corresponding images compared to the manual ground truth (GT) in red. The scale bar in all images is 500 μm .

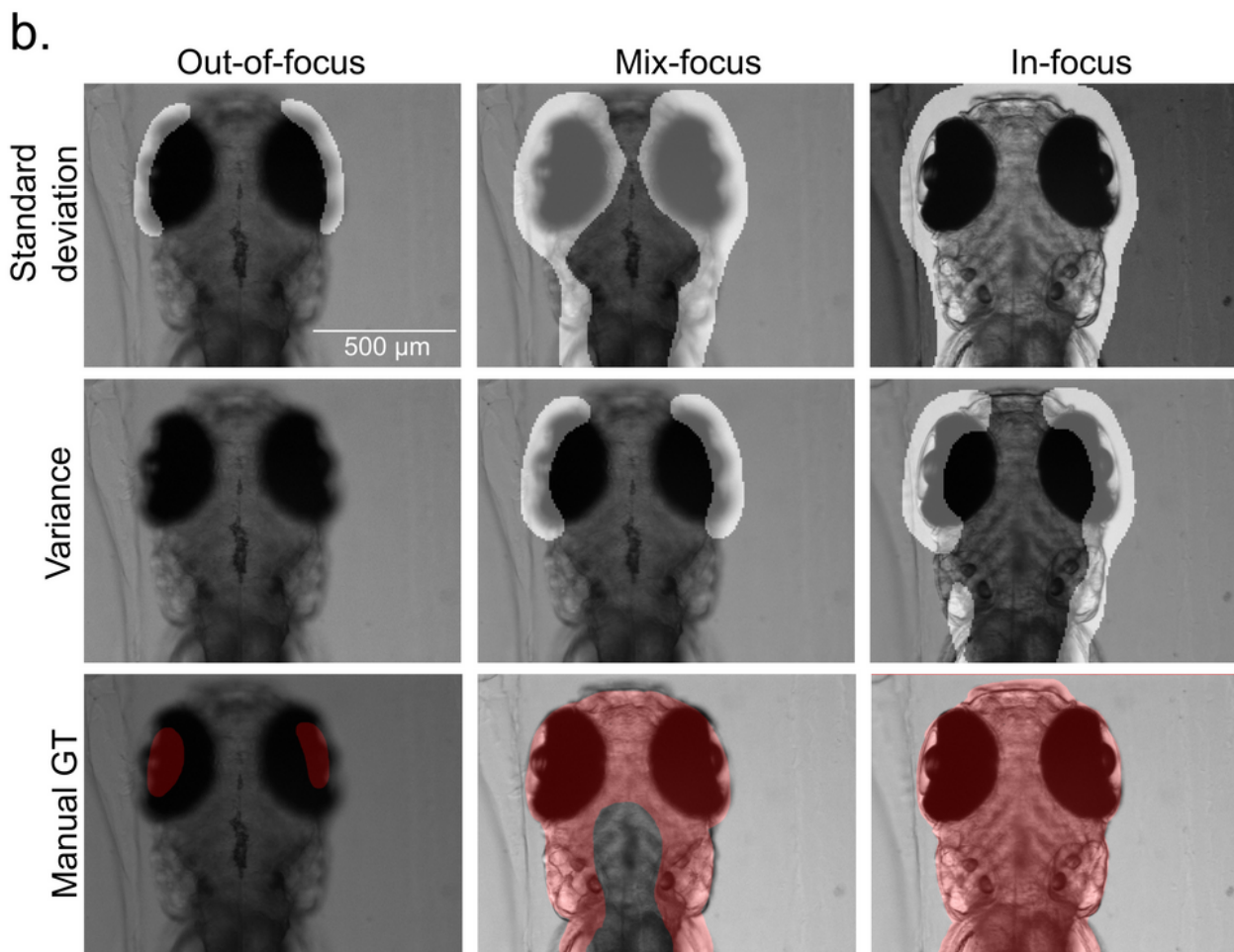
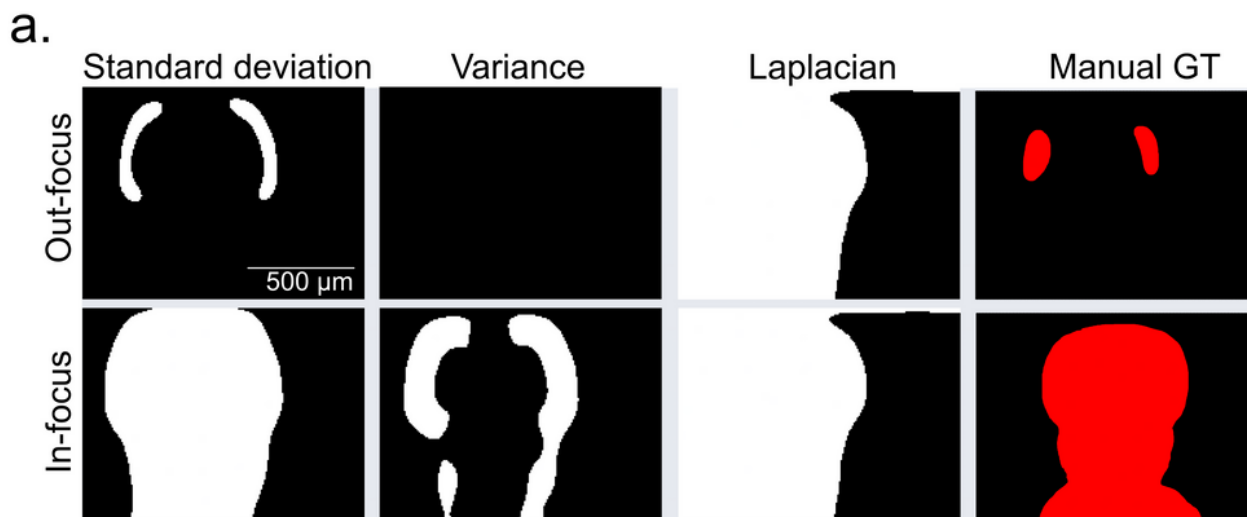


Figure 3

The pipeline of the in-focus segmentation using deep neural networks. (a) shows the preprocessing part, which scans the regions of image stacks from widefield microscopy and outputs the feature map for the

maxima projecting. The focus-score map resulting from maxima projecting serves as the input for the binarizing step. This process marks the in-focus pixels in the focus-score maps as the Weak-label Ground Truth (WGT) masks. **(b)** the deep learning part adopts the WGT masks along with the widefield microscopy images for the surrogate deep neural network (DNN) training. The DNN segments directly the in-focus pixels from the original image stacks and presents the 3D information of targets. The scale bar in all images corresponds to 500 μm .

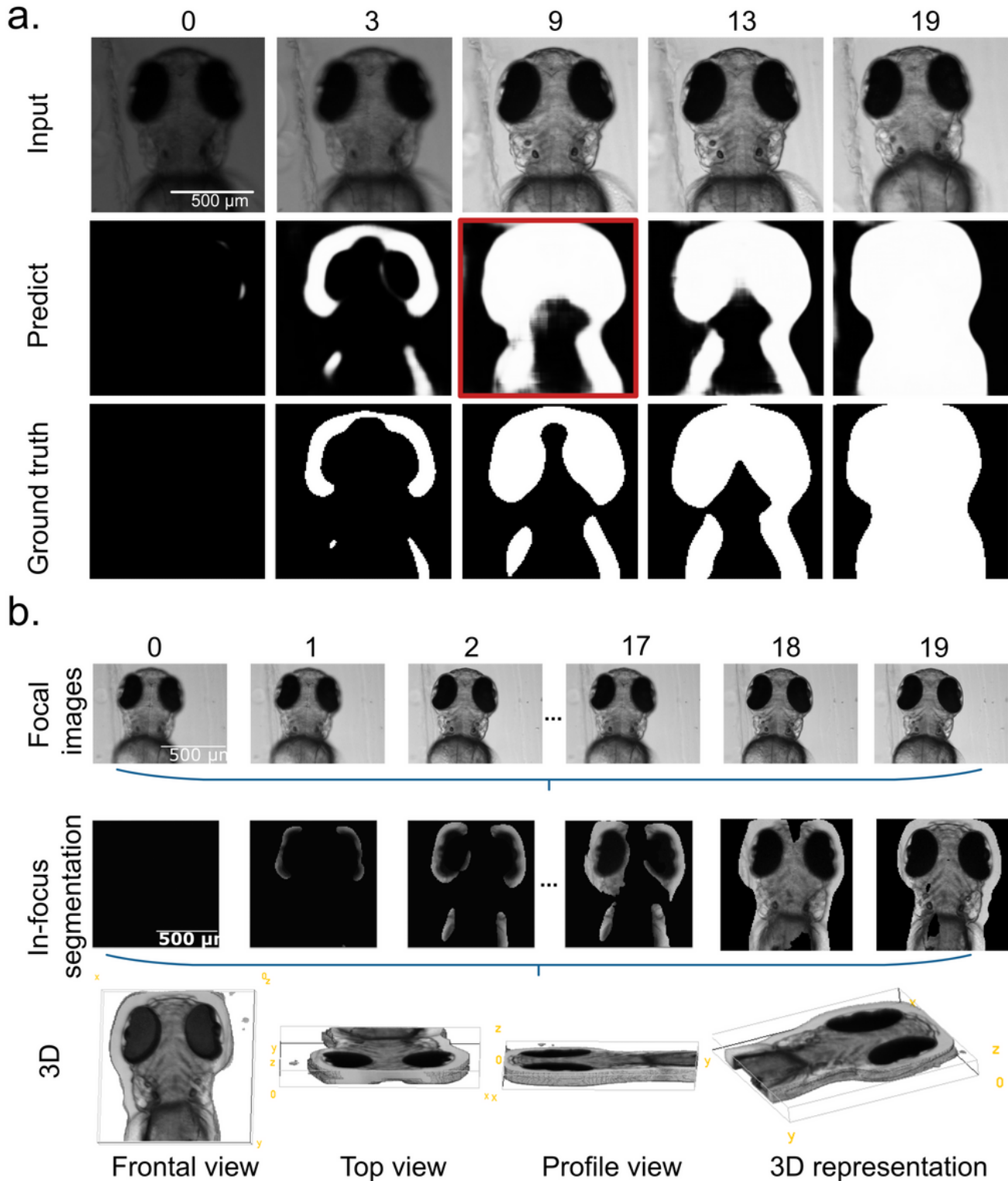


Figure 4

The end-to-end in-focus segmentation model using deep neural network (DNN). (a) shows input, prediction, and manual Ground Truth on slices ranging from 0 (out-of-focus) to slice 19 (in-focus) in one stack. But in some slices (slice 9 for example), the 7-layer DNN fails to keep consistent with the GT mask. This leaves the improvement space for future work. (b) as the end-to-end pipeline, the trained DNN segments the in-focus pixels from the image stacks of widefield images directly. These pixels represent the 3D information of targets. This enables the operations of optical sectioning in a digital way by using widefield microscopy.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.pdf](#)