vU-net: accurate cell edge segmentation in time-lapse fluorescence live cell images based on convolutional neural network

Chuangqi Wang[1*], Xitong Zhang[2*], Yenyu Chen[1], Kwonmoo Lee[1]

[1]Department of Biomedical Engineering

[2]Department of Computer Science

Worcester Polytechnic Institute, Massachusetts, 01609, USA

*These authors equally contributed to this work.

Corresponding Author: Kwonmoo Lee, Email: klee@wpi.edu

Abstract

Time-lapse fluorescence live cell imaging has been widely used to study various dynamical processes in cell biology. However, fluorescence live cell images often have low contrast, noises, and uneven illumination, preventing accurate cell segmentation. The convolutional neural network has been successfully applied in natural image classification and segmentation by extracting hierarchical features, which could be transferred into other fields for image segmentation. Moreover, the temporal coherence in time-lapse images can allow us to extract sufficient features from a limited number of image frames to segment the entire time-lapse movies. In this paper, we propose a novel framework called vU-net, which integrates the VGG-16[1] pretrained model as an encoder and a U-net[2] derived simplified convolutional structure to reconstruct cell edge with a higher accuracy using limited training images. We evaluated our framework on the high-resolution images of paxillin, a canonical adhesion marker in migrating PtK1 cells acquired by a Total Internal Reflection Fluorescence (TIRF) microscope, and achieved higher accuracy of cell segmentation than conventional U-net. We also validated our framework on noisy confocal fluorescence live images of GFP-mDia1 in PtK1 cells. We demonstrated that vU-net could be practically applied to challenging live cell movies since it required limited training sets and achieved highly accurate segmentation.

Introduction:

Over the past decades, time-lapse fluorescence microscopy has been successfully applied to obtain a massive amount of live cell movies with high spatiotemporal resolution, which opens up a new avenue for quantitative analyses on cellular dyanmics[3-5]. However, live cell images have numerous challenges regarding image analysis. Their characteristics include uneven illumination, optical noises, fluorescence background, photobleaching, complex cellular shapes, and continually changing intensity contrast due to migration at different rates[6,7]. These challenges fail the typical segmentation methods like intensity thresholding, watershed transform and energy-based segmentation methods to detect cell boundaries accurately. Therefore, as an initial image analysis step, there is still an unmet need for robust cell boundary segmentation with high accuracy, which is required for reliable quantitative analysis on high-resolution microscopy video.

Image segmentation can be divided into two groups: unsupervised[8] and supervised methods. In unsupervised methods, each image is segmented individually and independently based on extracted local or global features. Popular unsupervised segmentation methods which are also widely used in biological fields contains Otsu method[9], Canny Detector[10], active contour or snake-based method[11] and even recently PMI method based on mutual Information[12]. Even though the unsupervised methods are easy to handle and have the less computational burden, it is challenging to define specific local features, and the extracted edges are often disconnected, which requires human manual assistance. On the other side, many supervised methods achieve a higher accuracy in natural image segmentation since many images could be collected, and the features related to edges could be learned. Typically, the image segmentation problem is transformed into a binary classification issue to identify whether each pixel is on edge or not based on the features extracted from the training set. Popular methods like gPb[13], Sketch tokens[14,15] belong to this category. However, the biological dataset is always limited and small, which makes training impractical. Therefore, it is necessary to transfer features extracted from nature images dataset and reuse them in biological images

In last decade, deep learning has achieved great success in image processing fields [16-19]. DeepEdge[20], DeepContour[21] based on Alexnet[22] or VGG-16[1] and FCN[16] are proposed for natural image segmentation. The advantage of deep learning based segmentation is that the useful features can be learned hierarchically and easy to transfer to other fields. In cell biological image fields, several works such as U-Net[2] and deepcell[17] are proposed successfully for cell boundary segmentation.

Particularly, U-net can directly learn a nonlinear mapping from raw image patch to the labeled boundary by integrating low-level edge information and high-level reconstructed information. However, these methods were only evaluated in the low-resolution microscopy images, and it is unknown whether it can produce highly accurate edge information in high-resolution images.

To segment high-resolution time-lapse movies for subcellular analysis, we propose a novel framework called vU-net (integration of VGG-16 and U-net), which uses limited images for the training set. The framework could be divided into two parts: an encoder which extracts image features and decoder which identify the edge location using the extracted features. In order to use a limited number of training images, the pretrained VGG-16 model is reused in the encoder and extracts features in a different level, which is widely used in various frameworks[23]. The second difference from U-net is that in the decoder, a simpler and asymmetric framework is incorporated to detect the edge, which substantially reduces the model complexity. In comparison to U-net, our vU-net includes less number of trainable parameters, which will use much less training set. In our experiments, Total Internal Reflection Fluorescence (TIRF) microscopy image and noisy confocal microscopy images were used to evaluate our framework.

Materials and Methods:

To detect the cell boundary accurately, we proposed a novel framework called vU-net (Fig. 2i) and the related workflow including preparing the training set, training/testing process, is demonstrated in Fig 2. At first, several frames (Fig. 2b) are evenly selected from the entire image frames (Fig. 2a) for manual segmentation. Then, three cropping strategies: evenly cropping (white), randomly cropping along the edge(red) and randomly cropping region from the background and cell foreground(yellow), are applied to crop images into image patches as demonstrated in Fig. 2c, where the ratio in these three strategies is shown in Fig. 2d. Then, the cropped patches can be grouped into three clusters with the labels: edge-persevered patches, patches containing only background and patches containing only cell foreground shown in Fig. 2e. Finally, the cropped patches are augmented by the following strategies: x-flip, y-flip, x\y-flip, rotation with different angles and shown (Fig. 2f). In the following step, these augmented patches are divided into training and validation dataset with the ratio 8 to 2. The training dataset is used to train vU-net to map the raw image patch into the corresponding foreground/background

patches directly (Fig. 2g). During the testing step, each frame is evenly cropped with the same size (similar to the evenly cropped shown as white in Fig. 2c) and then the binary foreground/background information of each patch is predicted by the trained vU-net and then all the continuous patches are integrated together to generate the predicted segmentation (Fig. 2h).

The details of our novel framework vU-net are shown here (Fig. 2i). The framework can be divided into two parts: encoder path and decoder path. In the encoder, the same structure of VGG-16 is applied, which contains five convolutional cells, each of which contains a different number of convolution and max-pooling operation with the depth of 32-64-128-256-512. Also, to reduce the number of parameters to fit the smaller dataset, the weights of VGG-16 are transferred and fixed. In the decoder path, different from VGG-16, only two convolution cells comprising convolution and up-pooling operations with the depth of 128-64 are set up to integrate the edge information directly extracted from the image as lower level and reconstruction from region information as a higher level. Globally, our framework is asymmetric to take the benefits of VGG-16 to extract the useful features and reduce the number of trainable parameters to make it suitable for the small dataset. Since during the convolution operation, the zero-padding strategy is applied to make the size the same for convenience but make the boundary region of each patch less accurate. For higher accuracy, only central region of the predicted patch is used to generate the entire foreground/background prediction by cropped operation at the end of the structure. The size of input patch is 128x128 while the size of prediction patch is 68x68. The size of the convolutional filter is 3x3, and the size of max-pooling is 2x2 while that of up-pooling is 4x4 to fit the size of convolution. Besides, for comparison, we also implemented U-net with the similar strategies with two differences. The structure of the encoder part is the same with the original U-net and parameters of the encoder part is trainable while the structure of encoder and decoder are symmetric shown in SFig. 1.

During the training process, the binary cross-entropy between prediction and manually labeled mask for each patch is used as a loss function, and the dice coefficient is used to compare the performance of U-net and the proposed vU-net. Adam is used as an optimizer, and the initial learning rate is 1e-5, and other parameters are default values in the Keras with Theano backend. To save the training time and avoid overfitting, early stopping which the validation loss does not decrease 0.0001 in continuous three epochs is set, and the model can be trained until 30 epochs as a maximum iteration.

Results:

Challenges in segmenting cell edges in TIRF (Total Internal Reflectance Fluorescence) images.

TIRF microscopy is a popular imaging method to visualize cell-matrix adhesions since it can excite fluorescence molecule near cell-substrate interfaces. However, TIRF images usually suffer from non-uniform illumination, and there are strong fluorescence signals from focal adhesions. These make it difficult to choose a single intensity threshold value for edge segmentation (Fig. 1a, 1b). To better visualize the cell image, we first fit two Gaussian mixture model (GMM)[26] with the intensity histogram and the intensities outside two standard deviations were trimmed (Fig. 1b) and then the remaining intensity values are mapped into the range between 0 and 255 (Fig. 1d-e). To assess the intensity variability, we overlaid the manually labeled edge dilated with the shape disk of size 5 (Fig. 1c) on the transformed image (Fig. 1e) and the intensity histogram along the edge is quantified (Fig. 1f). This shows that the edge intensity is highly variable even though the contrast between cell and background is visually acceptable, indicating that a single thresholding value will not be sufficient to extract the entire edge boundary. Therefore, various segmentation methods applied to the transformed images blurred by the Gaussian filter (Fig. 1g) produced less accurate and broken edges or regions (Fig. 1h-l).

vU-net improves cell boundary segmentation using limited training images.

We trained our proposed vU-net and U-net using the augmented datasets from 16 training frames. The dice coefficient and loss curves in training and validation sets are shown in Fig. 3a. From the dice coefficient and loss curves, we can see that vU-net can achieve the better performance in both training and validation with much less training epochs than U-net. Also, to save the training time, early stopping was used in the vU-net training. The training cost time is much less in vU-net than U-net in Fig. 3b where the training times were recorded to achieve the dice coefficient 0.95 shown in Fig. 3a. To assess the robustness of training, we randomized the training and testing images and repeated training three times. As demonstrated in Fig. 3c, vU-net showed more robust performance regarding accuracy, recall and dice coefficient in the testing set (Fig. 3c). Finally, we demonstrated the testing results using the frame #20 (Fig. 3d-h). First, we generated the original prediction results of vU-net (Fig. 3e) and U-net (Fig. 3f), where we predicted the boundaries binarized by the threshold value 0.8 together for vU-net (Fig. 3g) and U-net (Fig. 3h). We overlaid the manually segmented boundary (Fig. 3d) with these predicted boundaries.

As shown in the zoomed regions of Fig. 3g-h, the predicted boundary from vU-net is better overlapped with the manual segmentation than that of U-net. Taken together, we showed that vU-net could improve the accuracy of cell segmentation of challenging live cell images with limited training images.

vU-net requires much less training sets than U-net.

Since our framework requires manually segmented training images within the same time-lapse movies, we tested how many training images should be prepared to achieve accurate segmentation results in both vU-net and U-net. We increased the number of the training frames from 2 to 24. In each condition, we randomly selected the training frames and repeated the training three times. As shown in Fig. 4a-b, there is a large margin of dice coefficient and loss between vU-net and U-net. With the same number of training frames, vU-net was more accurate than U-net. The variability of vU-net performance is smaller than that of U-net, which means that vU-net is more robust that U-net. When the number of training frames reaches to 16, the performance of vU-net reached a plateau, while the performance of U-net kept increasing with the increasing number of frames, which means more labeling works are required to achieve the good performance of U-net.

To demonstrate the performance with the increasing number of labeled frames, we visualized the segmentation results of frame #50 (Fig. 4c-h). We also confirmed that vU-net produced the visually better results than U-net in each number of training frame. Particularly, in the gray rectangular region in Fig. 4e, 16 frames were required for training to achieve good performance in vU-net (Fig. 4e) whereas U-net still did not produce accurate boundary (Fig. 4h). Finally, we also plotted time evolution of cell edges in vU-net and U-net. In comparison with the result of U-net (Fig. 4j), vU-net produced more smooth spatiotemporal edge changes (Fig. 4i) in comparison with the result of U-net (Fig. 4i).

vU-net showed good performance of cell segmentation in noisy confocal microscopic images

We also evaluated the vU-net in other types of microscopy images. We used the confocal fluorescence images visualizing GFP-mDia1 in a PtK1 cell (Fig. 5 a-c). The Signal to Noise Ratio (SNR) of this image is so low that it is difficult to identify the cell membrane visually (Fig. 5a-b). We applied global Ostu, local Ostu and canny method to this noisy image, and they failed to produce correct segmentation (Fig. 5 e-g). We manually segmented the cell boundary (Fig. 5c) to build up a training dataset using 16 frames. As in Fig. 5h, vU-net was training well to achieve a high

dice coefficient and low loss value in both training and validation sets. We confirmed visually that the edge prediction was successful (Fig. 5i) We also randomly chose the training/testing sets and repeated the training three times to quantify the accuracy, recall, and dice coefficient (Fig. 5j). We quantitatively confirmed that vU-net model accomplished very high accuracy and recall. Finally, the edge evolution plot demonstrated the smooth changes of edge motions (Fig. 5k).

Discussion

Live cell imaging becomes a fundamental tool to study dynamic biological processes such as cell migration. Since segmentation is the initial step of image analysis, accurate and effective segmentation of live cell images is crucial. In this paper, based on the temporal coherence of time-lapse image sequences, a novel framework, vU-net was designed to accurately segment the cell boundary from high-resolution fluorescence microscopy images[25]. In this framework, we evaluated the applicability of VGG-16 feature extractor to high-resolution microscopy images. VGG-16 was trained on natural images in ImageNet database to extract generic image descriptors. We confirmed that the image descriptors from VGG-16 were highly effective in predicting accurate cell boundary. We also incorporated a simple decoder of VGG-16 extracted features to produce cell segmentation using limited training set. As in U-net, this decoder still integrates low-level and high-level information to identify the cell edges. To our best knowledge, it is the first study to evaluate how small dataset is sufficient to train deep learning framework in cell segmentation.

Although 16 frame training set produced the best result, the model trained with only four frames still performed well on segmenting the whole 200 frames. To reduce manual labor, we can utilize prior knowledge from experience or other image segmentation algorithms to identify challenging regions and prepare more specific training set. Moreover, the following iterative segmentation framework can be proposed. First, we can train the model with very limited frames such as four frames, and then evaluate the performance on the entire dataset to identify the challenging regions. Then, we can manually label these regions and include them into the training set. The model can be retrained iteratively until a satisfied performance is achieved.

In this paper, the model has to be trained for individually time-lapse movies. This still requires significant human labor, preventing the high-throughput application. Building up the training set across various movies with different cell types, molecules, and experimental/imaging conditions will require tremendous human

efforts. Nonetheless, since only limited image frames were sufficient for entire image frames in our study, it is plausible that small frames extracted from each image sequence can be collected to set up a training set for a more generic vU-net model.

Acknowledgements: We thank NVIDIA for providing us with TITAN X GPU cards (NVIDIA Hardware Grant Program), Microsoft for providing us with Azure cloud computing resources (Microsoft Azure Research Award), and Boston Scientific for providing us with the gift for deep learning research. This work was supported by the WPI Start-up Fund for new faculty.

References

1 Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
2 Ronneberger, O., Fischer, P. & Brox, T. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 234-241 (Springer).
3 Ettinger, A. & Wittmann, T. Fluorescence live cell imaging. *Methods in cell biology* **123**, 77 (2014).
4 Frigault, M. M., Lacoste, J., Swift, J. L. & Brown, C. M. Live-cell microscopy–tips and tools. *J Cell Sci* **122**, 753-767 (2009).
5 Waters, J. C. Live-cell fluorescence imaging. *Methods in cell biology* **114**, 125-150 (2013).
6 Wu, K., Gauthier, D. & Levine, M. D. Live cell image segmentation. *IEEE Transactions on biomedical engineering* **42**, 1-12 (1995).
7 Stephens, D. J. & Allan, V. J. Light microscopy techniques for live cell imaging. *Science* **300**, 82-86 (2003).
8 Zhang, H., Fritts, J. E. & Goldman, S. A. Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding* **110**, 260-280 (2008).
9 Otsu, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* **9**, 62-66 (1979).
10 Canny, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 679-698 (1986).
11 Chan, T. F. & Vese, L. A. Active contours without edges. *IEEE Transactions on image processing* **10**, 266-277 (2001).

12 Isola, P., Zoran, D., Krishnan, D. & Adelson, E. H. in *European Conference on Computer Vision*. 799-814 (Springer).
13 Arbelaez, P., Maire, M., Fowlkes, C. & Malik, J. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **33**, 898-916 (2011).
14 Lim, J. J., Zitnick, C. L. & Dollár, P. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3158-3165.
15 Dollár, P. & Zitnick, C. L. in *Proceedings of the IEEE International Conference on Computer Vision*. 1841-1848.
16 Long, J., Shelhamer, E. & Darrell, T. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3431-3440.
17 Van Valen, D. A. *et al.* Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology* **12**, e1005177 (2016).
18 Badrinarayanan, V., Handa, A. & Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293* (2015).
19 Turaga, S. C. *et al.* Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation* **22**, 511-538 (2010).
20 Bertasius, G., Shi, J. & Torresani, L. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4380-4389.
21 Shen, W., Wang, X., Wang, Y., Bai, X. & Zhang, Z. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3982-3991.
22 Krizhevsky, A., Sutskever, I. & Hinton, G. E. in *Advances in neural information processing systems*. 1097-1105.
23 Hernández, C. X. & Sultan, M. M. Using Deep Learning for Segmentation and Counting within Microscopy Data. (2017).
24 Lee, K. *et al.* Functional hierarchy of redundant actin assembly factors revealed by fine-grained registration of intrinsic image fluctuations. *Cell systems* **1**, 37-50 (2015).
25 Wang, C., Choi, H. J., Kim, S.-J., Bae, Y. & Lee, K. Deconvolution Of Subcellular Protrusion Heterogeneity And The Underlying Actin Regulator Dynamics From Live Cell Imaging. *bioRxiv*, 144238 (2017).
26 Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. Speaker verification using adapted Gaussian mixture models. *Digital signal processing* **10**, 19-41 (2000).
27 Wang, Y., Zhao, X. & Huang, K. Deep Crisp Boundaries.
28 Pinheiro, P. O., Lin, T.-Y., Collobert, R. & Dollár, P. in *European Conference on Computer Vision*. 75-91 (Springer).
29 Shi, W. *et al.* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1874-1883.

30 Boykov, Y. & Funka-Lea, G. Graph cuts and efficient ND image segmentation. *International journal of computer vision* **70**, 109-131 (2006).

31 Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* **22**, 888-905 (2000).

32 Felzenszwalb, P. F. & Huttenlocher, D. P. Efficient graph-based image segmentation. *International journal of computer vision* **59**, 167-181 (2004).

Figures Legends

Figure 1: The segmentation challenging in Total Internal Reflection Fluorescence (TIRF) microscopy image. (**a**) The heatmap of a PtK1 cell expressing fluorescent paxillin. (**b**) The intensity histogram of the image in (a) and fitting with two Gaussian Mixed Model (GMM). The intensity interval with two variances is shown as dark dots. (**c**) The manually labeled cell boundary in (a). (**d**) The intensity histogram after enhancing the boundary contrast. (**e**) The contrast-enhanced image of (a). (**f**) The intensity histogram on the manually labeled cell boundary. (**g**) Gaussian blurred image of (c). (**h**) The result of Canny edge detector. (**i**) The result of global Ostu threshold. (**j**) The result of Graph cut [30]. (**k**) The result of the Normalized cut (N-cut) method [31]. (**l**) The result of Advanced Graph-based segmentation[32].

Figure 2: The schematic of vU-net. (**a**) The time-lapse microscopy movies. (**b**) Manual segmentation. (**c**) Three cropped strategies: even cropping(white), random cropping along the labeled edge (red) and randomly cropping beyond the labeled edge (yellow). (**d**) The ratio of three cropping strategies. (**e**) Categorized patches: edge patches, foreground patches, and background patches. (**f**) Image augmentation. (**g**) Neural network training. (**h**) The predicted segmentation. (**i**) The asymmetric vU-net framework.

Figure 3: The performance of vU-net on TIRF microscopy images. (**a**) The training and validation performance of vU-net and U-net. (**b**) Training time comparison between vU-net and U-net. (**c**) The quantification of foreground/background accuracy, recall and dice coefficient of vU-net and U-net. (**d**) the manual segmentation (**e**) the segmentation result of vU-net (**f**) the segmentation of U-net. (**g-h**) The predicted edges from vU-net (g) and U-net (h) are overlaid with the manual segmentation.

Figure 4: Evaluation of the number of training images. (**a-b**) The dice coefficient (a) and loss (b) on the testing set with a different number of training images. (**c-h**) The predicted edges from vU-net and U-net are overlaid with the manual segmentation with a different number of training images. (**i-j**) The edge evolution predicted by vU-net (i) and U-net (j).

Figure 5: The performance of vU-net in a noisy confocal fluorescence image. (**a**) A confocal microscopy image of a PtK1 cell expression GFP-mDia1. (**b**) The heatmap of (a). (**c**) The manual cell segmentation of (a). (**d**) Global Ostu thresholding fails to identify the cell membrane boundary. (**e**) Local Ostu thresholding. (**f**) Canny edge detection. (**g**) The training performance of vU-net using 16 training images. (**h**) The predicted segmentation. (**i**) The quantification of foreground/background accuracy, recall and dice coefficient of vU-net. (**j**) The edge evolution predicted by vU-net.

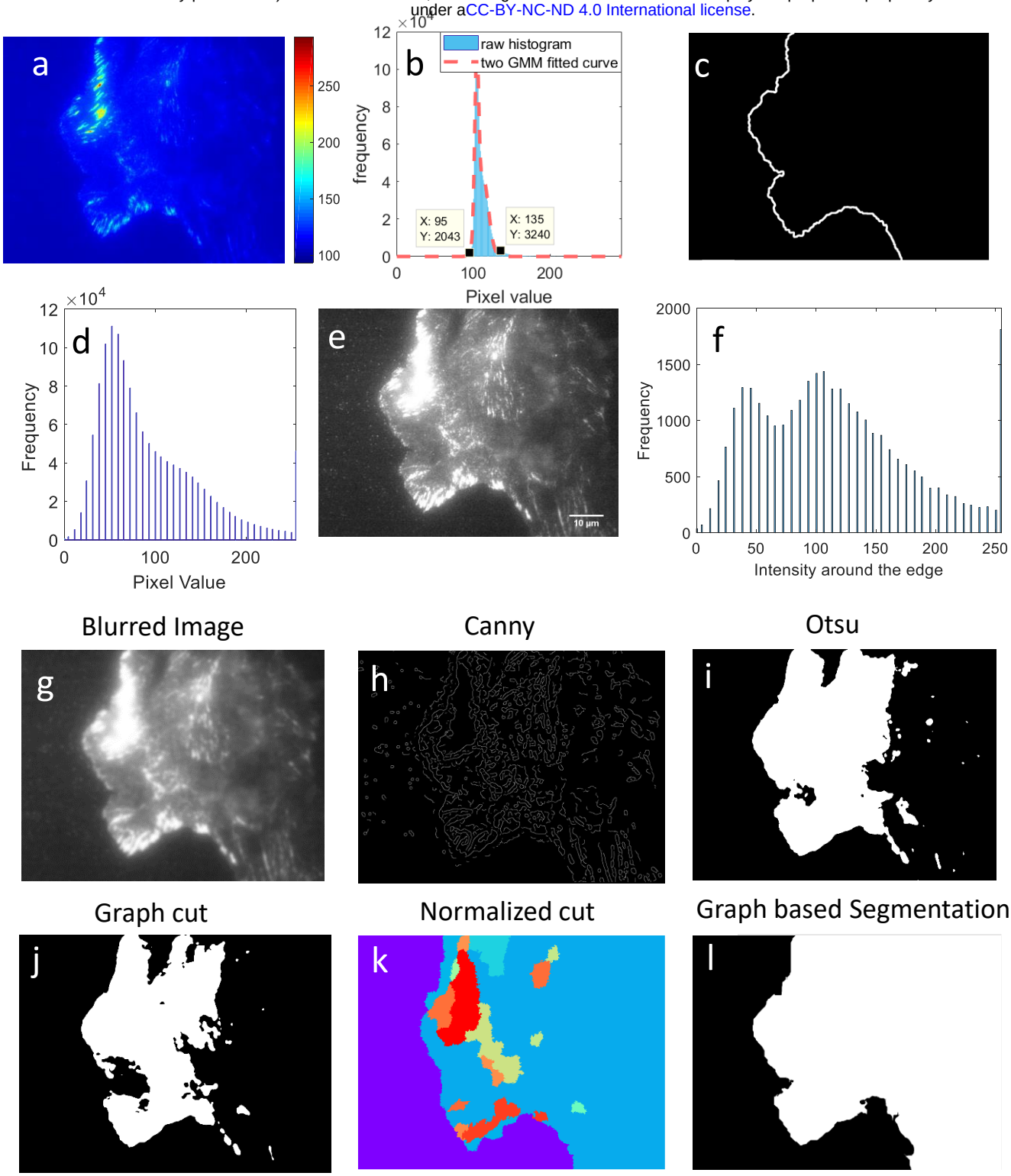


Figure 1: The segmentation challenging in Total Internal Reflection Fluorescence (TIRF) microscopy image.

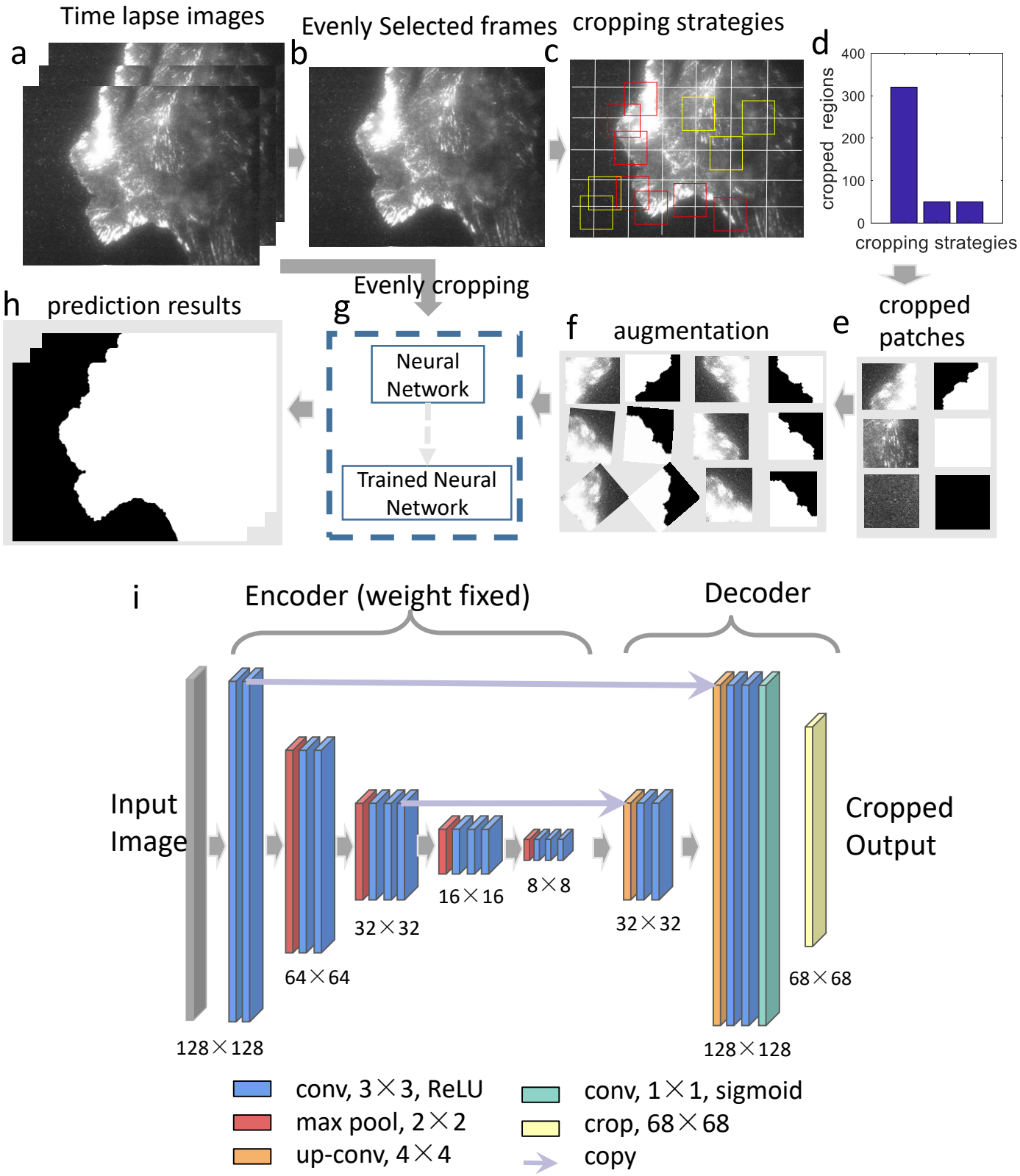
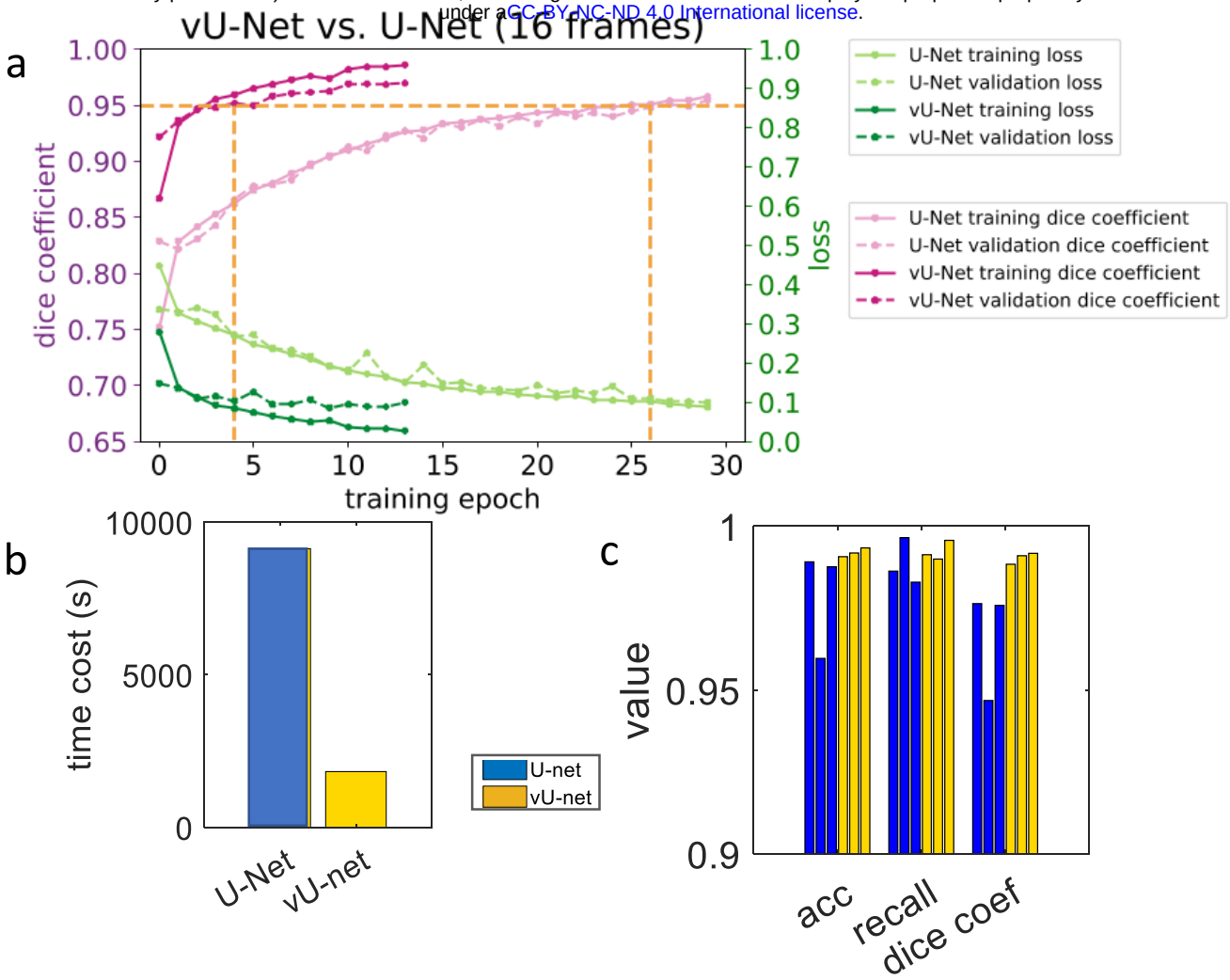


Figure 2: The schematic of vU-net.



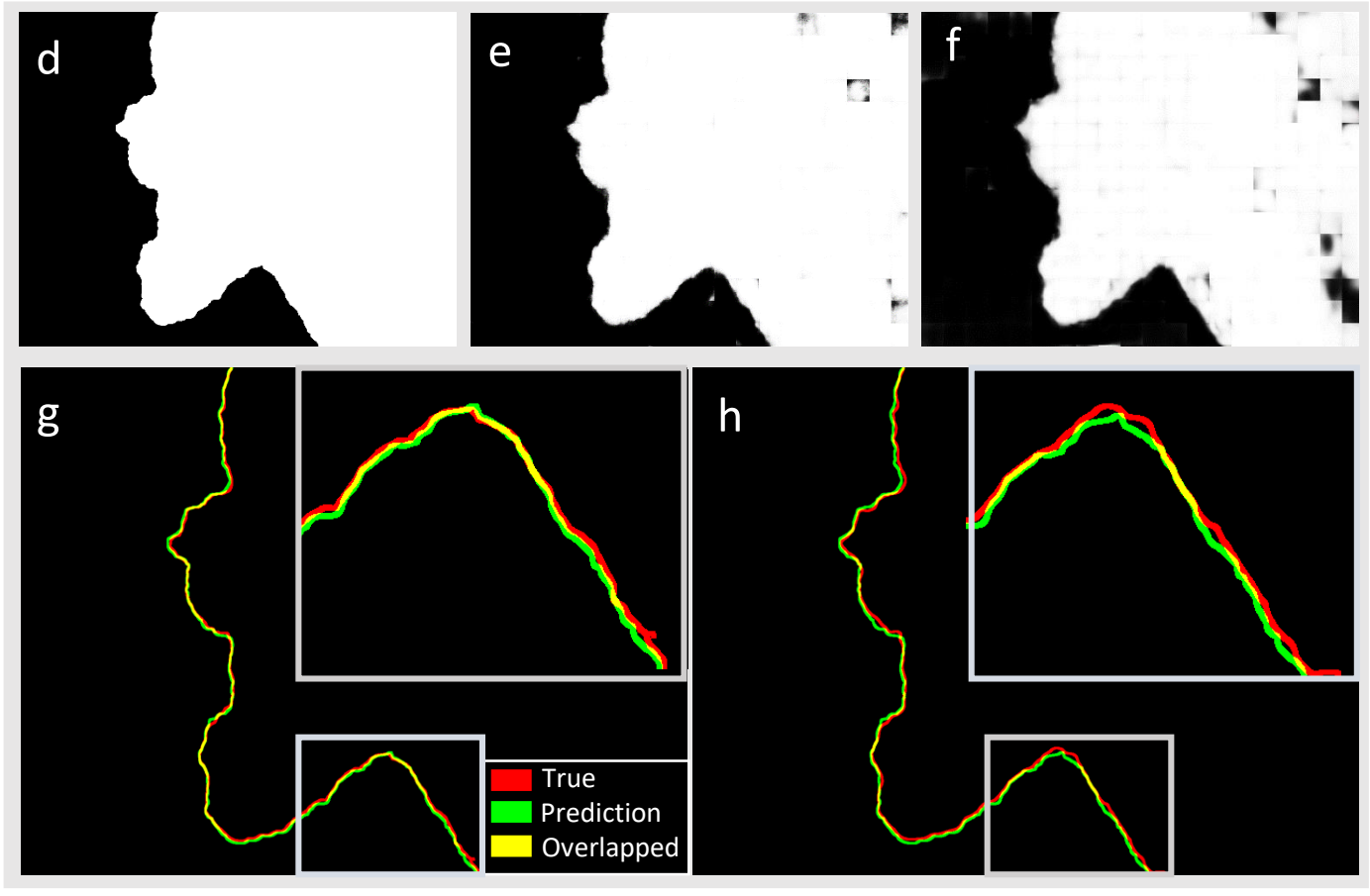
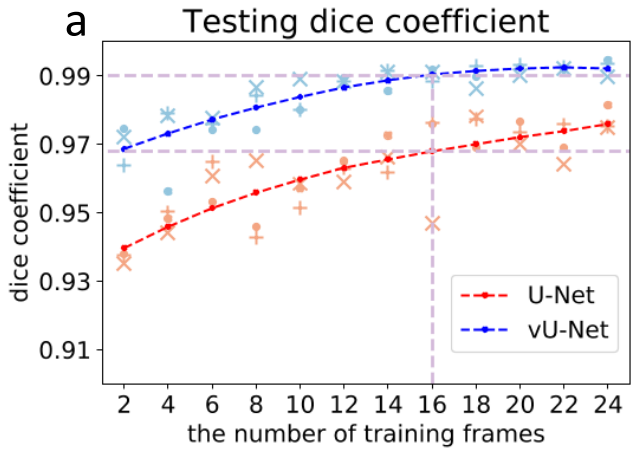
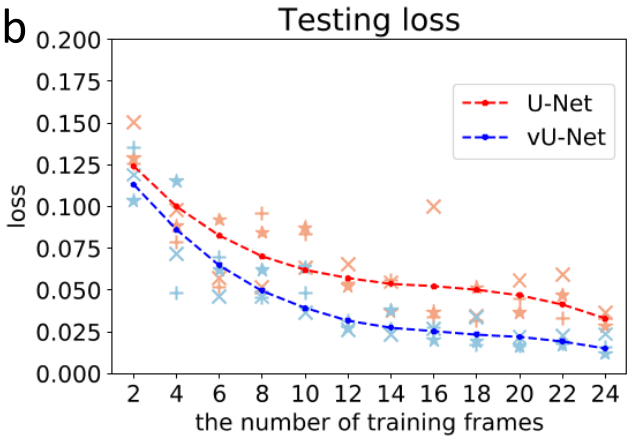
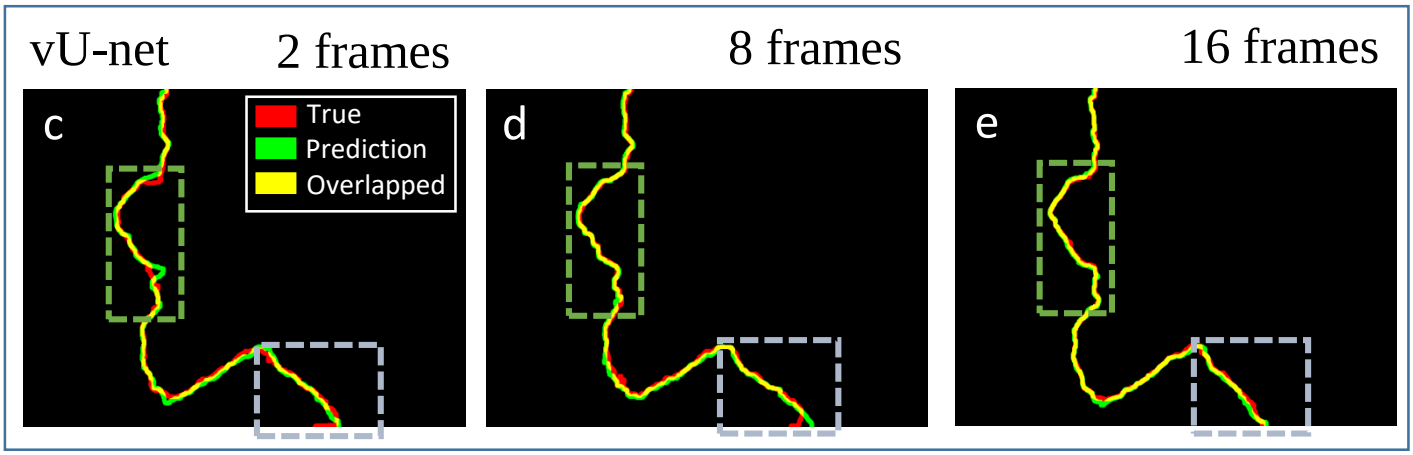
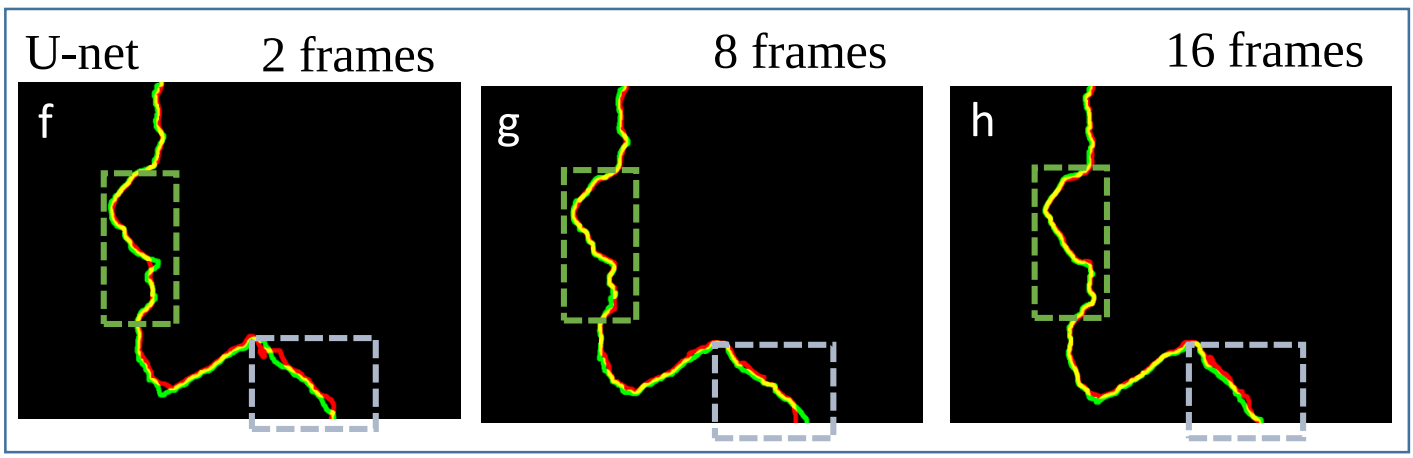


Figure 3: The performance of vU-net on TIRF microscopy images. .









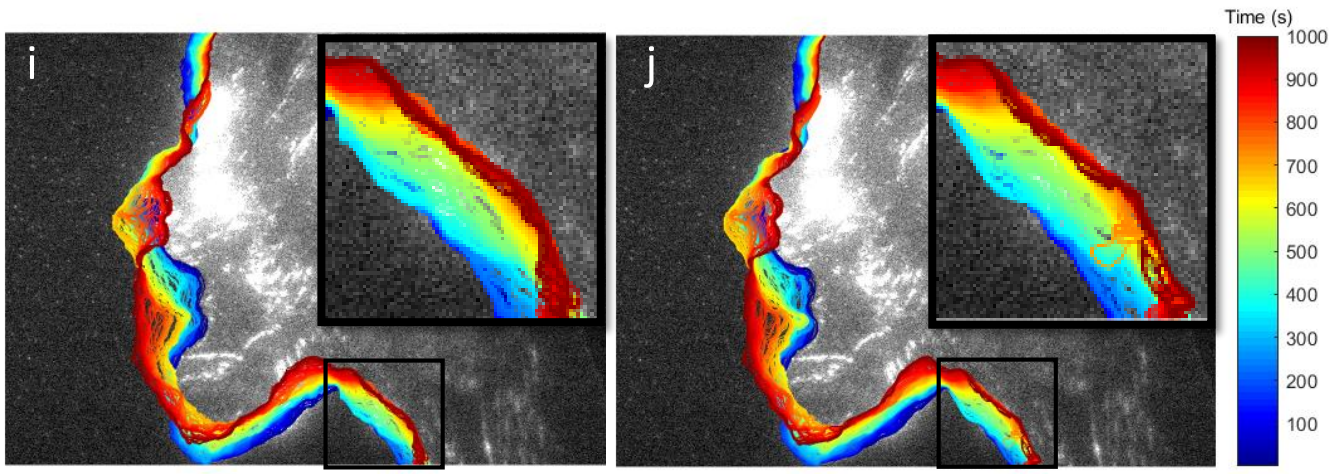


Figure 4: Evaluation of the number of training images.

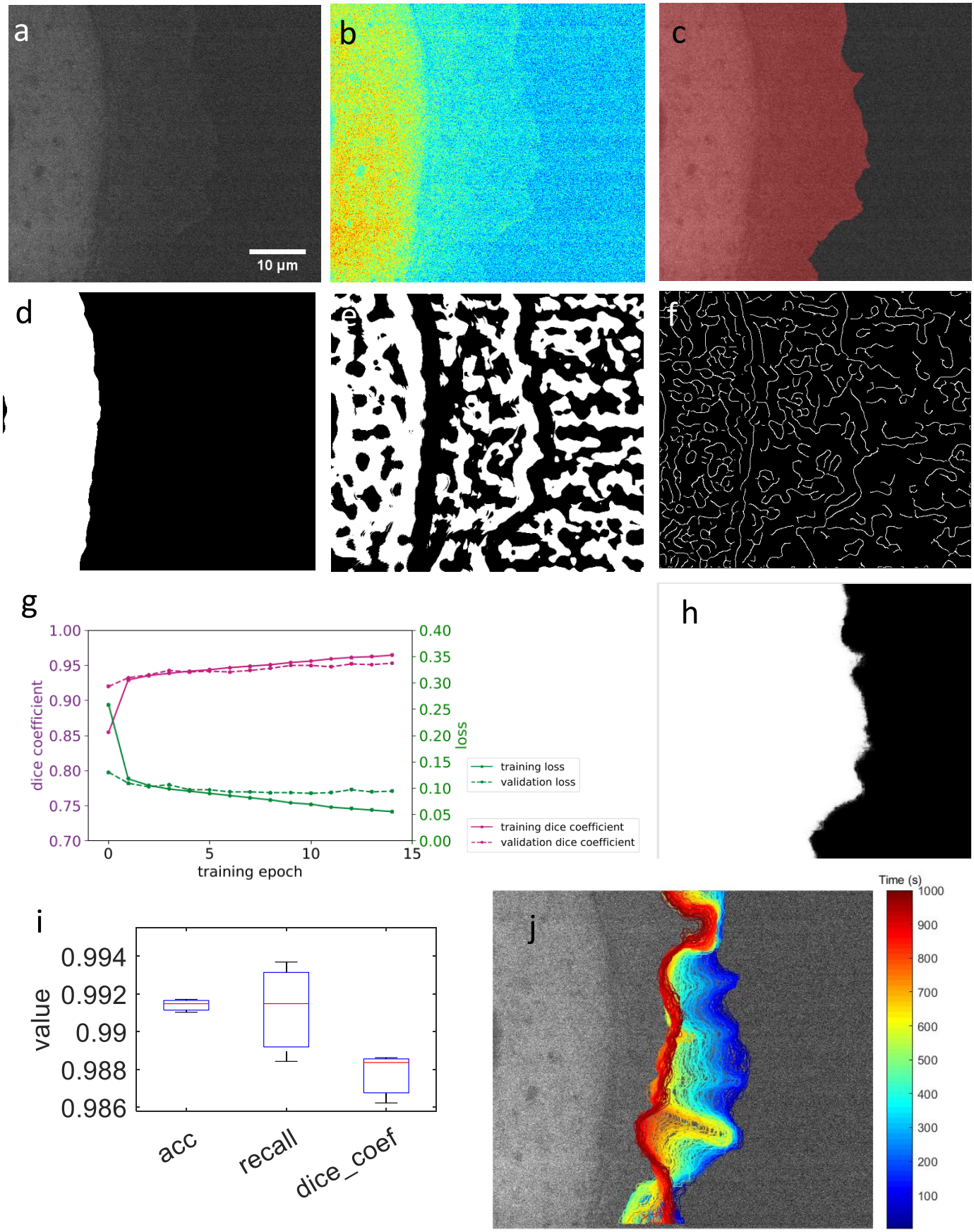
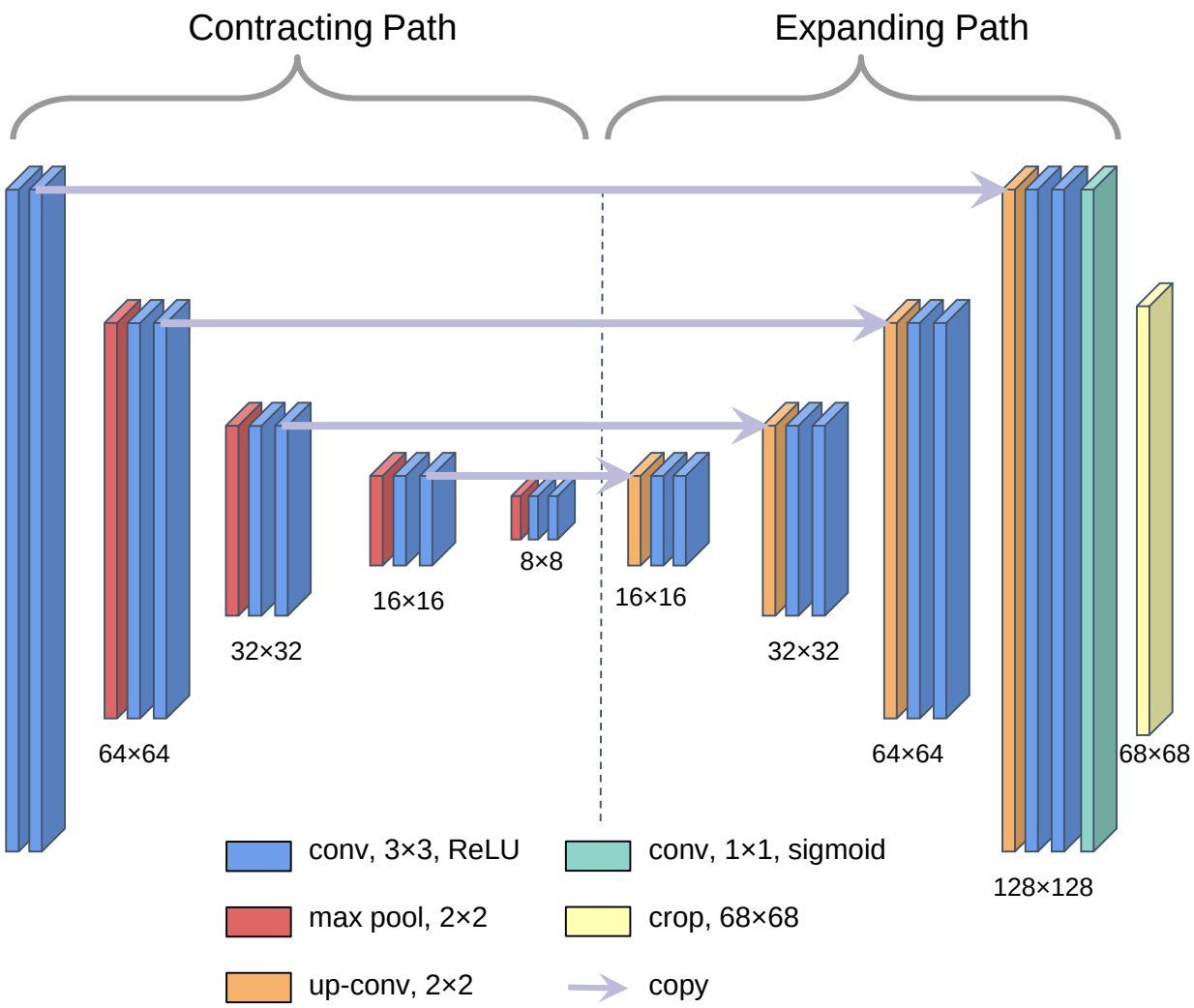


Figure 5: The performance of vU-net in a noisy confocal fluorescence image.



Supplementary Figure 1. The schematic framework of U-net.