Supplementary Material

Jiaxing Huang[1,2], Yanfeng Zhou[1,2], Yaoru Luo[1,2], Guole Liu[1,2], Heng Guo[3,4], and Ge Yang[1,2]

[1] Institute of Automation, Chinese Academy of Sciences, 100190, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences
[3] DAMO Academy, Alibaba Group
[4] Hupan Lab, 310023, Hangzhou, China
{huangjiaxing2021, ge.yang}@ia.ac.cn

1 Box-counting Method

We provide a summary of the flow of the box-counting method in Algorithm 1 and present a schematic illustration of the algorithm in Fig. 1 to facilitate a better understanding of this algorithm among readers.

Algorithm 1: Box-counting.

Input: Image I of size $M \times N$.
Output: Fractal dimension (FD) of image.

Considering I as 3D spaces (x, y, z), (z) denotes gray value;
Assign $F \leftarrow \min(M, N)$;
Assign $L \leftarrow$ max gray level of I, typically $L = 255$;
$PointList_{NR} \leftarrow [\ \ ]$;
for $k \leftarrow 2$ **to** $F/2$ **do**
    Split image I to grids G_s size of $k \times k$;
    Calculate the height of the box by $h = \frac{(L-1)\times k}{F}$;
    $N_r \leftarrow 0$;
    $r \leftarrow k/F$;
    foreach $G_{i,j}$ *in* G_s **do**
        Size of each box is $k \times k \times h$;
        $l = \lceil \max G_{i,j}(z)/h \rceil$ ←Index of box contains $\max G_{i,j}(z)$;
        $m = \lceil \min G_{i,j}(z)/h \rceil$ ←Index of box contains $\min G_{i,j}(z)$;
        $n_r(i, j) = l - m + 1$;
        $N_r \leftarrow N_r + n_r(i, j)$;
    end
    $PointList_{NR}$ append $(\log N_r, \log 1/r)$;
end
$FD \leftarrow$ Linear Regression $(PointList_{NR})$;

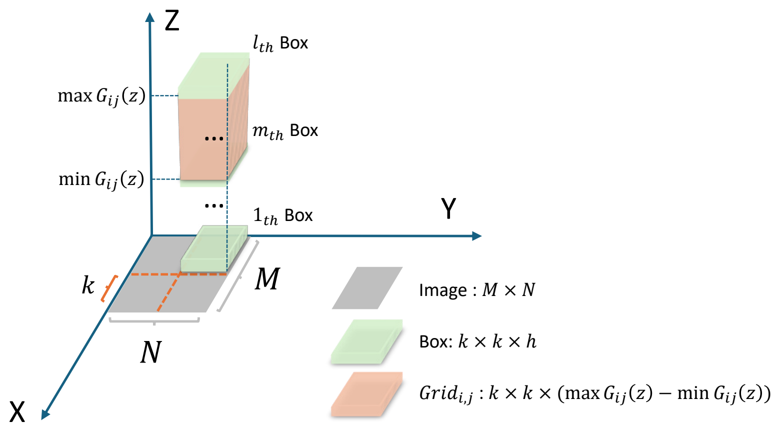

Fig. 1: Schematic Diagram of the Box-counting Method.

2 Dataset Details

ER [7] : The ER dataset comprises 2D fluorescence microscopy images representing the endoplasmic reticulum (ER) network within cultured live cells, which were acquired via spinning disk confocal microscopy. For ER dataset, the training, validation and testing sets consist of 157, 28 and 38 images, respectively. Each image has a resolution of 256×256. For training, we use the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm [11] to pre-process the images to adjust the image's contrast as employed in [2].

MITO [7]: The MITO dataset includes 2D fluorescence microscopy images that depict the mitochondrial (MITO) network in cultured live cells, acquired through widefield microscopy. For MITO dataset, the training, validation and testing sets consist of 165, 8 and 10 images, respectively. Each image has a resolution of 256×256. For training, we use the CLAHE algorithm [11] followed by normalization to pre-process the images as employed in [2].

ROSE [8] : The ROSE dataset is a 2D retinal OCTA (Optical Coherence Tomography Angiography) segmentation dataset. We use ROSE-1 (SVC) in this work. It has a predetermined split of 30 train and 9 test samples, with each sample having a resolution of 304×304. For training, we use normalization to pre-process the images as described in [3].

STARE [4] : The STARE dataset is a dataset for retinal vessel segmentation. It contains 20 equal-sized (700×605) color fundus images. The first ten images (im0001 to im0139) were used as training images, while the subsequent ten images (im0162 to im0324) were reserved for testing. During training, images were randomly sampled to a standard resolution of 256×256. Image pre-processing was conducted utilizing the CLAHE algorithm [11], as described in [10].

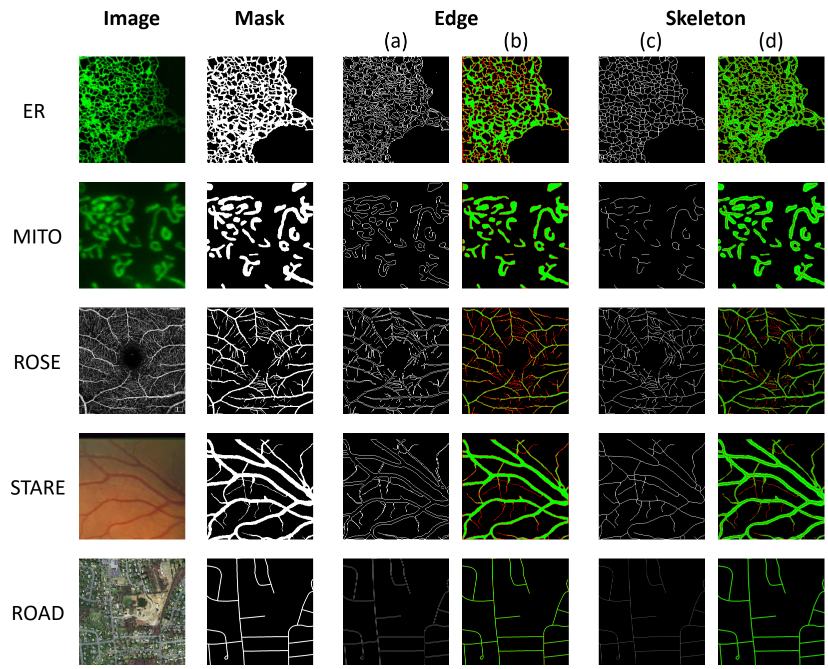

Fig. 2: Visualizations of the extracted edges and skeletons. Columns (a) and (c) display the extracted edges and skeletons, respectively. Columns (b) and (d) illustrate the visual comparison by overlaying the extracted edges and skeletons onto their corresponding masks. Red: true positive. Green: false negative. Blue: false positive.

ROAD [9] : The ROAD dataset is a large, non-medical dataset containing 1171 aerial images (1108 train, 14 validation, and 49 test), each of 1500×1500 resolution. A crop size of 256×256 is used randomly during training and regularly during testing for the ROAD dataset as in [10]. For training, we use the CLAHE algorithm [11] followed by normalization to pre-process the images.

NUCLEUS [1] : The NUCLEUS dataset contains a large number of segmented nuclei images. These images form a diverse collection of biological images collectively containing tens of thousands of nuclei. The dataset from stage 1, comprising 670 samples, was partitioned into training, validation, and test sets at a ratio of $7:1:2$, respectively. During training, random cropping was employed to obtain input images with a resolution of 256×256. For the validation and test phases, a sliding window approach was utilized, with each window 256×256 in size. We use the CLAHE algorithm [11] followed by normalization to pre-process the images.

3 Ground Truths for Edge and Skeleton

The ground truths for edges and skeletons are derived from the annotated masks by employing the findContours function from the OpenCV library and the skeletonization algorithms in the scikit-image library. We created visualizations of the obtained edges and skeletons to substantiate the accuracy of results. Fig. 2 presents these visualizations, with columns (a) and (c) displaying the extracted edges and skeletons, respectively. Conversely, columns (b) and (d) illustrate the visual comparison by overlaying the extracted edges and skeletons onto their corresponding masks. The color scheme is defined as follows: red indicates true positives, green signifies false negatives, and blue denotes false positives. Fig. 2 indicates that the regions marked in red correspond precisely to the target perimeters and skeletal structures, while the absence of blue regions corroborates the method's high level of accuracy.

4 Additional Quantitative Results

To assess the statistical significance of the improvements introduced by FFM, we perform t-tests comparing MD-Net with MD-Net* across various metrics on the ER, ROSE, and STARE datasets. As shown in Table 1, all metrics, except for Hausdorff Distance (HD) on ER and STARE, exhibited p-values below 0.05, underscoring the performance improvements because of FFM.

Further, t-tests between MD-Net* and contemporary SOTA methods revealed our model achieved a significance rate of 90.27%, succeeding in 65 out of 72 t-tests. Failures, predominantly occurring in the HD metric (6 out of 7 instances), highlight our method's significant uplift across evaluated metrics while indicating potential areas for refinement in boundary precision.

Table 1: Results of t-test. Red numbers indicate $p < 0.05$.

Dataset	**Model Compare**	IoU	clDice	ACC	AUC	β Error	HD
ER	DSC-Net vs MD-Net*	0.00024	0.00154	0.00052	1.17E-08	0.00461	0.00045
	Dconn-Net vs MD-Net*	0.00026	0.00060	2.72E-06	1.76E-26	0.00048	0.00042
	AF-Net vs MD-Net*	0.00021	2.05E-07	0.00126	2.56E-06	7.08E-06	0.13543
	GT-DLA vs MD-Net*	1.06E-06	2.90E-05	0.00010	1.09E-07	0.00077	0.01539
	MD-Net vs MD-Net*	0.00385	0.00234	0.04513	4.09E-06	0.03452	0.21833
ROSE	DSC-Net vs MD-Net*	0.00021	0.00155	0.01685	0.00138	0.01310	0.04272
	Dconn-Net vs MD-Net*	2.45E-06	0.00261	4.95E-08	2.58E-10	0.00049	0.00050
	AF-Net vs MD-Net*	3.59E-05	0.00019	4.44E-05	0.00028	0.00055	0.15576
	GT-DLA vs MD-Net*	0.01274	0.00049	0.88988	0.00343	9.09E-05	0.61038
	MD-Net vs MD-Net*	1.09E-05	0.00062	3.90E-06	0.01861	0.02387	0.04788
STARE	DSC-Net vs MD-Net*	0.00125	3.17E-05	0.03435	4.21E-09	0.04995	0.63767
	Dconn-Net vs MD-Net*	2.00E-11	6.97E-06	2.81E-08	2.76E-27	0.00407	0.22611
	AF-Net vs MD-Net*	0.00034	2.59E-05	0.00162	3.50E-10	0.02008	0.16595
	GT-DLA vs MD-Net*	1.45E-06	7.29E-07	6.45E-05	8.52E-17	0.04880	0.00667
	MD-Net vs MD-Net*	3.21E-08	1.50E-07	5.41E-07	3.16E-10	4.73E-07	0.15050

5 Additional Qualitative Results

In this section, we provide supplementary qualitative analysis of our methods as depicted in Fig. 3. We showcase representative images from various datasets, displaying the global and local results in the upper and lower sections respectively. Moreover, for the ROAD dataset, we present additional segmentation outcomes of images. The observed results further validate that the integration of FFM yields improved segmentation outcomes, effectively enhancing both edge accuracy and topological continuity.

6 Ablation Study Results with All Metrics

We provide the ablation study results with all metrics, intersection-over-union (IoU), accuracy (ACC), centerlineDice (clDice) [14], AUC, the Betti Error [5] β

Table 2: Ablation study of FFM in U-Net.

Model	Dataset	Input	IoU↑	clDice↑	ACC↑	AUC↑	β Error↓	HD↓
U-Net	ER	(image)	75.44	94.63	91.82	97.35	28.72	6.87
		(image, image)	75.61	94.56	91.82	97.35	29.81	6.82
		(image, HF)	*75.89*	94.77	91.89	97.43	*26.86*	6.83
		(image, MF)	75.86	*94.92*	*91.93*	*97.45*	26.97	**6.77**
		(image, CF)	75.10	94.38	91.63	97.27	33.39	6.91
		(image, FFM_{image})	**76.59**	**95.43**	**92.02**	**97.56**	**20.78**	*6.81*
	MITO	(image)	79.77	96.91	98.07	99.61	2.80	4.56
		(image, image)	80.28	97.39	98.17	99.61	3.00	4.32
		(image, HF)	*80.51*	*97.41*	98.19	*99.62*	**2.60**	**4.23**
		(image, MF)	80.23	97.46	98.14	99.54	3.50	*4.26*
		(image, CF)	80.38	97.13	*98.20*	99.61	3.50	4.38
		(image, FFM_{image})	**80.71**	**97.42**	**98.21**	**99.63**	*2.70*	4.27
	ROSE	(image)	61.52	*67.53*	91.33	94.04	*8.22*	7.42
		(image, image)	62.57	67.45	91.80	*94.31*	9.11	7.32
		(image, HF)	*62.61*	67.26	*91.82*	94.20	9.44	7.29
		(image, MF)	62.58	67.15	91.77	94.17	9.67	*7.28*
		(image, CF)	62.37	67.42	91.74	94.30	9.11	*7.28*
		(image, FFM_{image})	**64.07**	**67.95**	**92.25**	**94.47**	**7.88**	**7.10**
	STARE	(image)	66.15	76.05	94.68	96.39	3.22	6.67
		(image, image)	66.08	76.11	94.66	*96.40*	3.31	6.56
		(image, HF)	65.54	75.14	94.64	96.01	3.28	6.47
		(image, MF)	66.14	75.82	94.79	96.29	*3.00*	6.64
		(image, CF)	*66.83*	*76.16*	*94.95*	96.25	3.18	**6.32**
		(image, FFM_{image})	**68.07**	**77.39**	**95.15**	**96.63**	**2.77**	*6.33*
	ROAD	(image)	62.47	86.87	*97.97*	*98.29*	*2.61*	8.11
		(image, image)	63.44	85.75	97.23	97.57	3.35	7.15
		(image, HF)	*64.93*	*87.01*	97.29	97.78	2.66	*7.05*
		(image, MF)	64.58	86.89	97.30	97.74	3.17	7.09
		(image, CF)	64.25	86.52	97.33	97.70	3.02	*7.05*
		(image, FFM_{image})	**65.74**	**87.74**	**98.38**	**98.72**	**2.48**	**6.98**

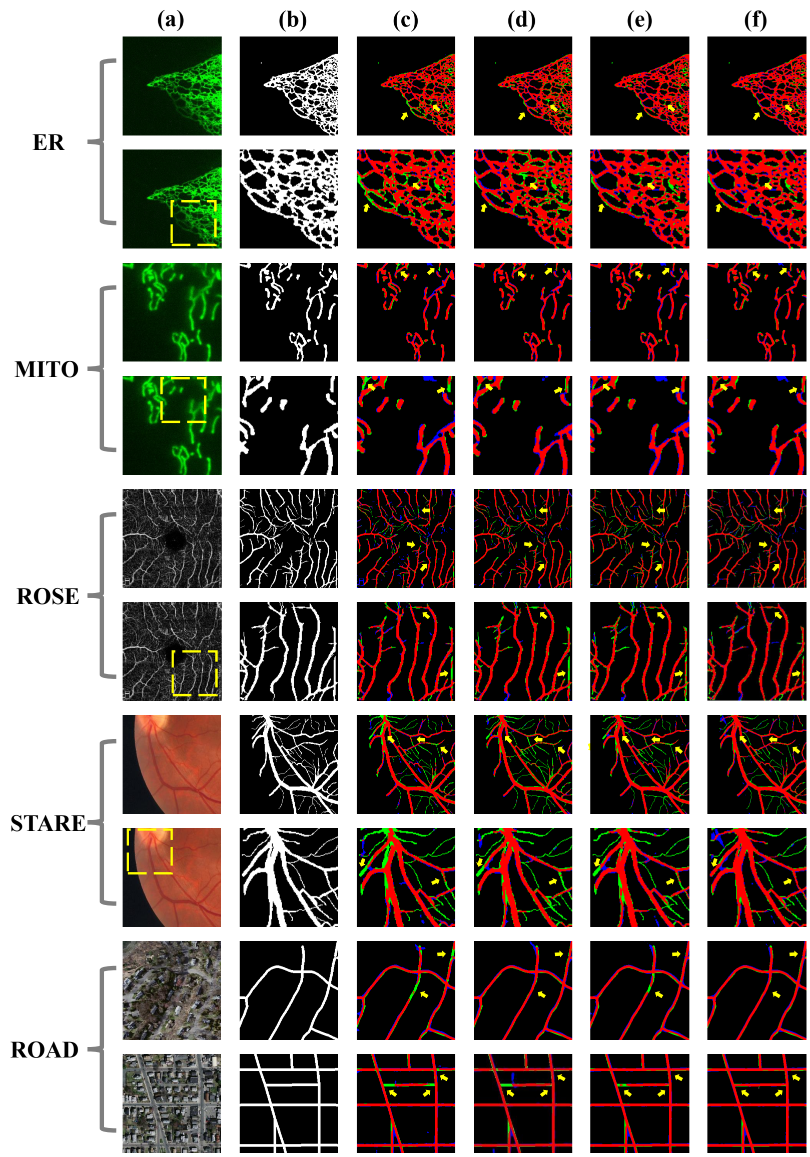

Fig. 3: Comparison of segmentation results. (a) Image. (b) Label. (c) Results of U-Net [12]. (d) Results of existing SOTA approaches. From top to bottom, it's AF-Net [13], AF-Net [13], GT-DLA [6], AF-Net [13], and Dconn-Net [16]. (e) Results of U-Net*. (f) Results of MD-Net*. Red: true positive. Green: false negative. Blue: false positive.

for the sum of Betti Numbers β_0 and β_1, and Hausdorff Distance (HD) [15], as shown in Tab. 2, Tab. 3 and Tab. 4. The evaluation metrics IoU, ACC, AUC,

Table 3: Ablation study of FFM in MD-Net.

Model	Dataset	Input	IoU↑	clDice↑	ACC↑	AUC↑	β Error↓	HD↓
MD-Net	ER	(image)	76.28	94.68	91.52	97.40	24.89	6.93
		(image, image)	*76.34*	*95.17*	91.85	*97.44*	24.16	6.81
		(image, HF)	76.24	95.06	*91.97*	97.42	*22.79*	**6.76**
		(image, MF)	75.97	94.94	91.92	97.40	25.53	6.80
		(image, CF)	75.52	94.68	91.52	97.40	24.89	6.93
		(image, FFM_{image})	**77.01**	**95.78**	**92.06**	**97.59**	**19.10**	*6.77*
	MITO	(image)	80.28	97.68	98.17	99.63	3.20	**4.16**
		(image, image)	80.31	*97.74*	98.17	*99.64*	3.30	*4.17*
		(image, HF)	79.97	97.39	*98.18*	99.52	*2.40*	4.27
		(image, MF)	*80.41*	**97.77**	**98.25**	99.60	3.20	*4.17*
		(image, CF)	80.32	97.15	98.16	99.60	3.40	4.26
		(image, FFM_{image})	**81.11**	97.72	**98.25**	**99.66**	**2.20**	**4.16**
	ROSE	(image)	*63.31*	68.24	91.90	*94.55*	8.67	*7.26*
		(image, image)	63.30	68.22	*91.92*	94.54	8.56	*7.26*
		(image, HF)	62.77	67.82	91.80	94.36	7.67	7.28
		(image, MF)	62.89	*68.41*	91.68	94.41	*7.22*	7.30
		(image, CF)	62.90	67.93	91.79	94.41	8.00	7.29
		(image, FFM_{image})	**65.07**	**69.78**	**92.36**	**94.88**	**4.22**	**7.10**
	STARE	(image)	66.46	76.54	94.76	96.42	3.25	6.48
		(image, image)	66.57	76.30	94.80	96.44	3.18	6.56
		(image, HF)	66.43	76.25	94.82	96.35	*2.85*	6.53
		(image, MF)	67.03	*76.96*	94.90	*96.51*	3.28	6.45
		(image, CF)	*67.25*	76.59	*95.05*	96.49	3.60	*6.39*
		(image, FFM_{image})	**68.49**	**77.79**	**95.20**	**96.91**	**2.57**	**6.29**
	ROAD	(image)	64.79	86.87	97.31	97.70	3.26	7.05
		(image, image)	65.04	87.28	97.34	97.85	2.63	*6.97*
		(image, HF)	*65.23*	*87.36*	*97.35*	97.81	2.71	7.05
		(image, MF)	65.12	87.12	97.32	97.85	2.78	7.02
		(image, CF)	65.15	87.19	97.31	*97.88*	*2.56*	7.03
		(image, FFM_{image})	**66.07**	**88.08**	**98.43**	**98.80**	**2.19**	**6.92**

and clDice, are expressed as percentages (%). The Hausdorff Distance (HD) is measured in pixels (px). In the evaluation of ROSE and STARE, β Error represents the β_1 only.

6.1 Effectiveness of Fractal Feature Maps

The results in Tab. 2 and Tab. 3 confirm that the observed performance improvement of the model is attributed to the FFM. The U-Net and MD-Net with input (image, FFM_{image}) performs better compared with (image, image), (image, Hurst Feature (HF)), (image, Mean Feature (MF)), (image, Contrast Feature (CF)) in five tubular datasets. With the incorporation of the FFM_{image}, both U-Net and MD-Net achieved either the best or the second-best results across six evaluation metrics. The results presented in the Tab. 2 and Tab. 3 validate that FFM, which is based on the self-similarity attributes of images and utilizes fractal geometry and fractal dimension, provides information that more closely

Table 4: Ablation of step size and window size in U-Net and MD-Net.

Model	Dataset	Step Size	Window Size	IoU↑	clDice↑	ACC↑	AUC↑	β Error↓	HD↓
U-Net*	ER	1	11	**76.75**	95.43	**92.12**	97.47	20.37	**6.70**
		1	9	76.73	95.35	92.03	97.46	23.97	6.71
		1	7	76.65	95.19	92.00	97.51	24.65	6.74
		1	5	76.59	95.43	92.02	**97.56**	20.78	6.81
		2	5	76.49	95.30	92.09	97.29	21.35	6.73
		3	5	76.68	**95.44**	92.09	97.55	**19.39**	6.74
		4	5	76.61	95.32	92.00	97.49	22.36	6.76
	STARE	1	11	67.87	77.31	95.08	96.71	3.52	6.34
		1	9	67.91	**77.61**	95.06	**96.78**	**2.65**	6.35
		1	7	67.89	76.87	**95.25**	96.71	3.55	6.36
		1	5	**68.07**	77.39	95.15	96.63	2.77	**6.33**
		2	5	67.87	77.11	95.14	**96.78**	3.52	6.34
		3	5	67.78	76.90	95.21	96.76	3.20	6.41
		4	5	67.78	77.12	95.13	96.74	2.75	6.36
MD-Net*	ER	1	11	**77.03**	95.72	92.06	97.58	20.13	6.71
		1	9	77.01	95.58	92.02	97.57	20.44	6.78
		1	7	76.92	95.58	92.00	97.57	20.36	6.82
		1	5	77.01	**95.78**	92.06	**97.59**	19.10	6.77
		2	5	76.91	95.54	**92.12**	97.57	19.78	6.70
		3	5	76.88	95.57	92.02	**97.59**	20.68	6.77
		4	5	76.90	95.58	92.01	97.54	**18.97**	**6.69**
	STARE	1	11	68.41	77.56	95.16	96.82	2.73	6.34
		1	9	68.21	77.81	95.15	96.87	2.77	6.42
		1	7	**68.50**	77.78	95.20	96.90	2.77	6.41
		1	5	68.49	**77.79**	95.20	**96.91**	**2.57**	**6.29**
		2	5	68.46	77.78	95.19	96.83	2.72	6.39
		3	5	68.32	77.55	95.13	96.86	2.65	6.36
		4	5	68.30	77.66	**95.22**	96.82	2.75	6.37

corresponds to the characteristics of tubular structures in comparison to other features.

6.2 Robustness of Fractal Feature Maps

In addition to IoU, the results of other evaluation metrics exhibit fluctuations within a narrow range, as illustrated in Tab. 4. Notably, the optimal value of each metric tends to occur across various step sizes and window sizes. This observation further reinforces that FFMs computed by different parameters can effectively enhance the model's performance.

6.3 Limitations of Fractal Feature Map

We have recorded the training and inference times for models on the ER dataset with 50 training epochs and 38 test samples, utilizing four Nvidia RTX 3090

Table 5: Comparison of training and inference time on ER dataset.

Stage	DSC-Net	Dconn-Net	AF-Net	GT-DLA	HR-Net	HR-Net*	MD-Net	MD-Net*
Training	48min 42s	24min 15s	61min 6s	25min 48s	20min 27s	20min 44s(+142.5s)	18min 10s	18min 37s(+142.5s)
Inference	4.78038s	3.08774s	5.16747s	2.81530s	2.54742s	2.57226s(+5.7s)	2.13556s	2.20907s(+5.7s)
Total	48min 47s	24min 18s	61min 12s	25min 51s	20min 29s	20min 47s(+148.2s)	18min 12s	18min 40s(+148.2s)

GPUs. Table 5 shows that integrating FFMs into our model only slightly increases the total time required. Herein, red numbers represent the extra time required for FFMs' computation. While our method demands more inference time compared to other models, the overall time does not markedly rise. This issue will be discussed in the main text as the limitations of our method. Efforts to reduce computation time by exploiting GPU acceleration are currently underway.

7 Codes

Codes are now available at `https://github.com/cbmi-group/FFM-Multi-Decoder-Network`. More implementation details and usage instructions can be found in the code files and the Readme file.

References

1. Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al.: Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nature Methods **16**(12), 1247–1253 (2019)
2. Guo, Y., Huang, J., Zhou, Y., Luo, Y., Li, W., Yang, G.: Segmentation of intracellular structures in fluorescence microscopy images by fusing low-level features. In: Pattern Recognition and Computer Vision: 4th Chinese Conference, PRCV 2021, Beijing, China, October 29–November 1, 2021, Proceedings, Part III 4. pp. 386–397. Springer (2021)
3. Gupta, S., Zhang, Y., Hu, X., Prasanna, P., Chen, C.: Topology-aware uncertainty for image segmentation. Advances in Neural Information Processing Systems **36** (2024)
4. Hoover, A., Goldbaum, M.: Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. IEEE Transactions on Medical Imaging **22**(8), 951–958 (2003)
5. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. Advances in neural information processing systems **32** (2019)
6. Li, Y., Zhang, Y., Liu, J.Y., Wang, K., Zhang, K., Zhang, G.S., Liao, X.F., Yang, G.: Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation. IEEE Transactions on Cybernetics **53**(9), 5826–5839 (2023)
7. Luo, Y., Guo, Y., Li, W., Liu, G., Yang, G.: Fluorescence microscopy image datasets for deep learning segmentation of intracellular orgenelle networks (2020). `https://doi.org/10.21227/t2he-zn97`, `https://dx.doi.org/10.21227/t2he-zn97`

8. Ma, Y., Hao, H., Xie, J., Fu, H., Zhang, J., Yang, J., Wang, Z., Liu, J., Zheng, Y., Zhao, Y.: Rose: a retinal oct-angiography vessel segmentation dataset and new model. IEEE Transactions on Medical Imaging **40**(3), 928–939 (2020)
9. Mnih, V.: Machine learning for aerial image labeling. University of Toronto (Canada) (2013)
10. Qi, Y., He, Y., Qi, X., Zhang, Y., Yang, G.: Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6070–6079 (2023)
11. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology **38**, 35–44 (2004)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
13. Shi, T., Ding, X., Zhou, W., Pan, F., Yan, Z., Bai, X., Yang, X.: Affinity feature strengthening for accurate, complete and robust vessel segmentation. IEEE Journal of Biomedical and Health Informatics **27**(8), 4006–4017 (2023)
14. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16560–16569 (2021)
15. Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. BMC Medical Imaging **15**(1), 1–28 (2015)
16. Yang, Z., Farsiu, S.: Directional connectivity-based segmentation of medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11525–11535 (2023)