

Supplementary File for “Potential-guided Connected Network for Tiny Structure Segmentation in Medical Images”

Chouyu Chen, Yaotong Song, Junyan Yi, Lijun Guo, Zhenyu Lei, *Member, IEEE*
and Shangce Gao, *Senior Member, IEEE*

THE supplementary file is structured into three main parts for clarity and comprehensiveness. Specifically, Section I presents all the intermediate segmentation maps of IG module and gives figures to show the variation tendency of segmentation potential metrics. Furthermore, Section II presents the results and detailed analysis of comparison experiments, focusing on the evaluation of the proposed method against state-of-the-art (SOTA) approaches. Finally, Section III provides an in-depth network explorations of the ablation study, hyper-parameter selection, and structural composition, offering insights into the design choices and their impact on performance.

To support the analyses in these sections, detailed information about the datasets employed is provided below. A summary of their characteristics, including dataset size, image resolutions, and primary segmentation challenges, is provided in Table S1.

DRIVE: The DRIVE dataset [1] includes 40 retinal images with binary annotations at a resolution of 584×565 pixels, divided into 20 images each for training and testing. In this study, each training image is augmented into 12 samples (via rotation, noise addition, and warping) and resized to 512×512 pixels.

MoNuSeg: The MoNuSeg dataset [2] consists of 51 H&E stained images at 40x magnification, with 37 images in the training set and 14 in the test set. Each image has binary segmentation annotations and is resized to 512×512 pixels.

CoNIC: The CoNIC dataset [3] contains 4,981 histology images (256×256 pixels) stained with H&E at 20x magnification (~0.5μm/pixel) from six sources. Instance segmentation and classification masks are provided for each image. Blank images are excluded, and the dataset is split into training and testing subsets in an 8:2 ratio.

FIVES: The FIVES dataset [4] includes 800 high-resolution color fundus images (2048×2048 pixels) with manual annotations standardized through expert crowdsourcing. The dataset is divided into 600 training and 200 testing images, with 20 test images randomly selected for evaluation. Training images are cropped to 512×512 patches, and test images follow a similar process.

GlaS: The GlaS dataset [5], from the Colon Histology Images Challenge, contains 165 images (775×522 pixels) from 16 histological sections of stage T3 or T42 colorectal adenocarcinoma. Each image, resized to 512×512 pixels for training and testing, exhibits significant inter-subject variability in

stain distribution and tissue structure, allowing for multi-scale performance assessment of the proposed method.

Moreover, the evaluation metrics employed in this paper are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP}, \quad (1)$$

$$Rec = \frac{TP}{TP + FN}, \quad (2)$$

$$Pre = \frac{TP}{TP + FP}, \quad (3)$$

$$F1 = \frac{2 \times Rec \times Pre}{Rec + Pre}, \quad (4)$$

$$IoU = \frac{TP}{TP + FP + FN}. \quad (5)$$

Here, true positive (TP) represents correctly segmented target pixels, true negative (TN) indicates correctly identified background pixels, while false positive (FP) and false negative (FN) refer to incorrectly segmented target and background pixels, respectively.

I. RESULTS OF IG MODULE

To elucidate the internal mechanisms of the IG module, we present intermediate visualizations obtained from multiple benchmark datasets. These representations reveal the progressive refinement of features as they propagate through successive stages of the module. By tracing the evolution of these features, we aim to illuminate the correspondence between internal transformations and resultant segmentation performance. Consistent visual patterns observed across datasets underscore the robustness and efficacy of the IG module in facilitating discriminative feature learning. Representative outputs of the IG module are provided in Fig. S1, S2, S3, and S4 while the corresponding trajectories of the segmentation potential metric is depicted in Fig. S5.

II. COMPARISON EXPERIMENTS

The comparison methods are summarized in Table S2, classified into two main groups: single networks and serially or progressively connected models. The single networks include UNet [6], UNet++ [7], CANet [8], MISSFormer [9], Swin-PANet [10], BRAUNet++ [11], and EMCAD [12]. Among these, UNet and UNet++ represent classical segmentation

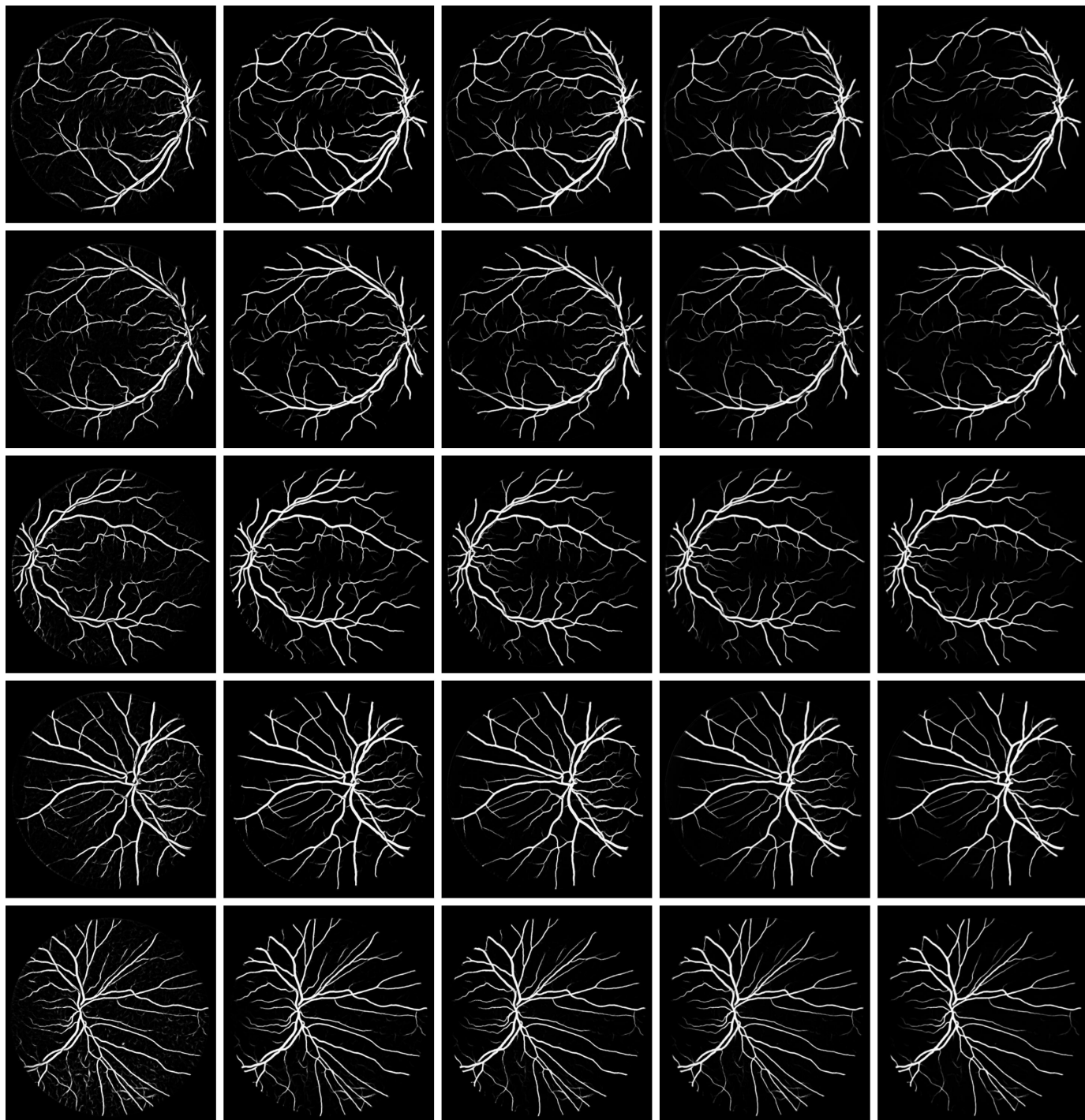


Fig. S1: Part of the results of the IG module on the DRIVE dataset.

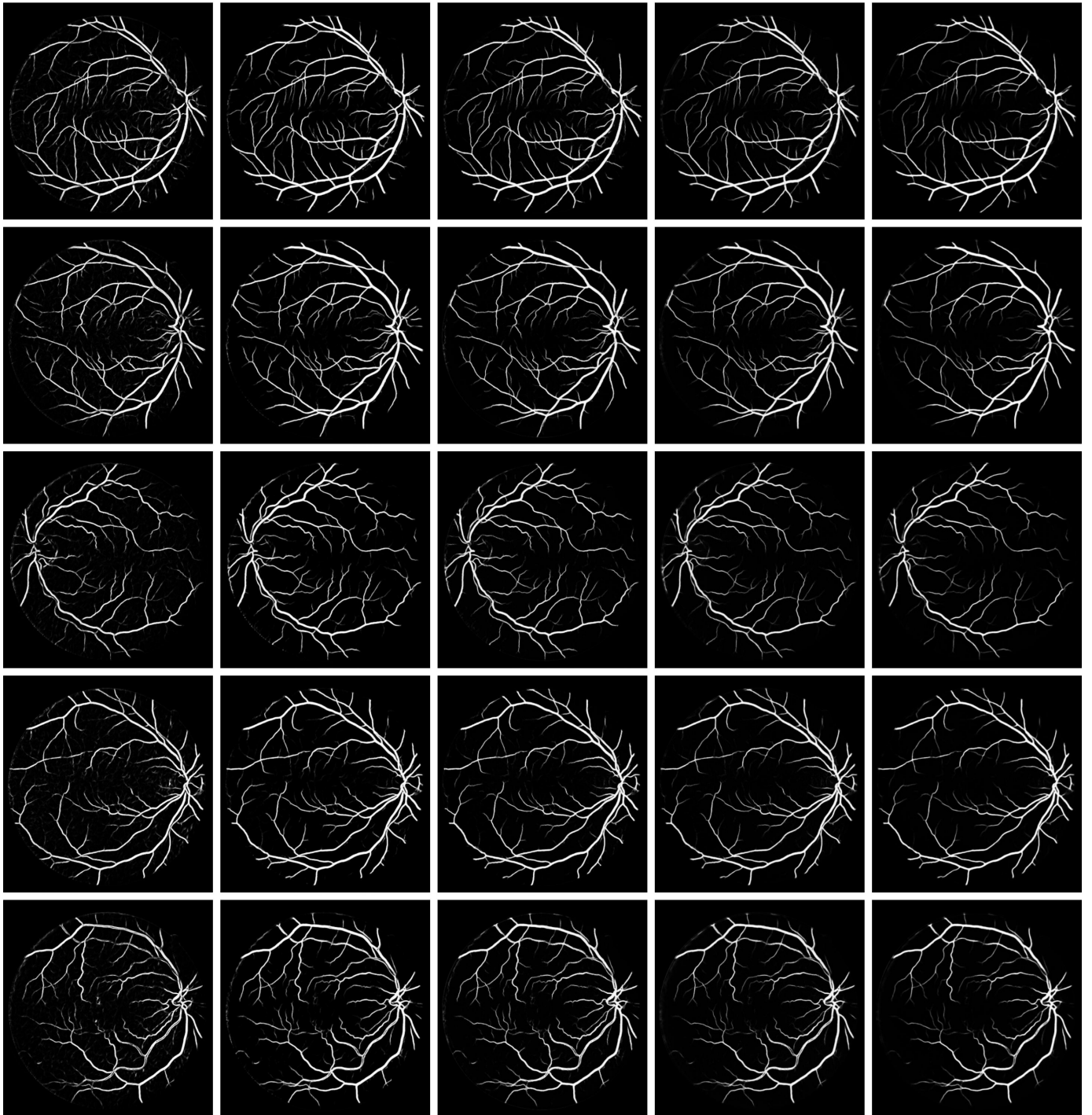


Fig. S2: Part of the results of the IG module on the DRIVE dataset.

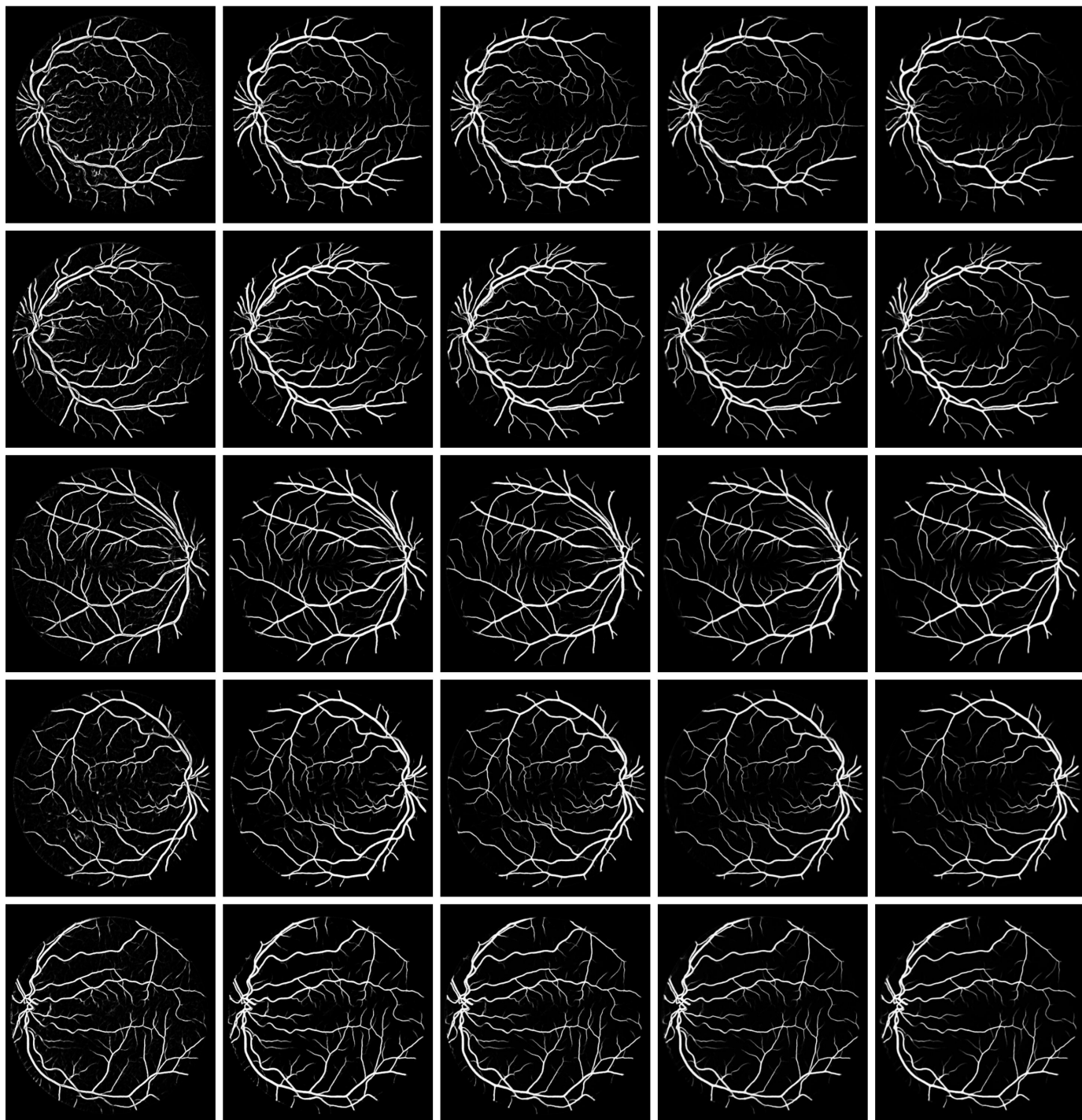


Fig. S3: Part of the results of the IG module on the DRIVE dataset.

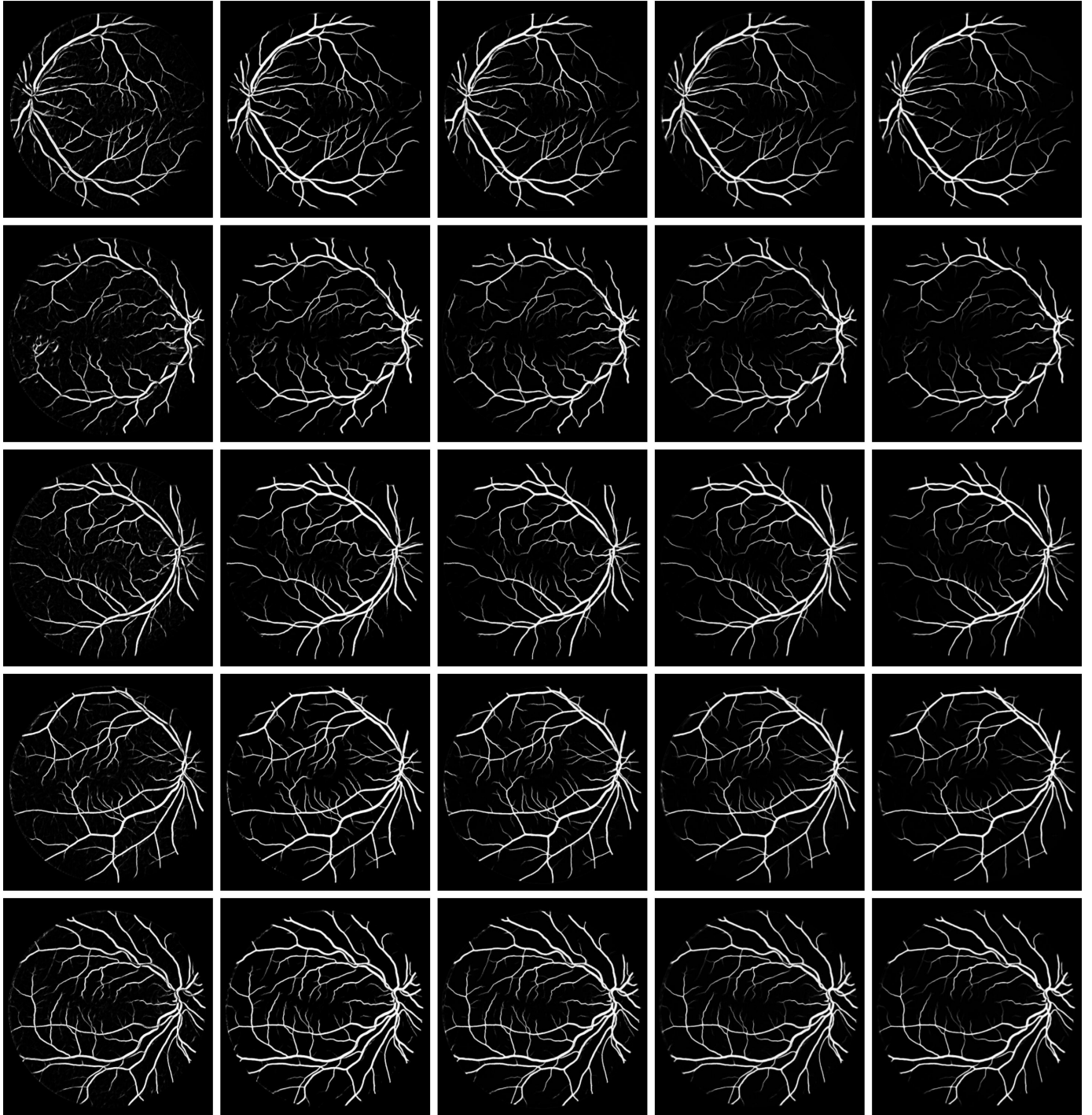


Fig. S4: Part of the results of the IG module on the DRIVE dataset.

TABLE S1: The summarize of datasets used in this paper.

Dataset	Size	Resolution	Challenges
DRIVE	40 images (20 training, 20 testing)	584×565 (resized to 512×512)	Small size limits generalization
MoNuSeg	51 images (37 training, 14 testing)	512×512	Complex overlapping nuclei structures
CoNIC	4,981 images (3984 training, 997 testing)	256×256	High diversity from six sources complicates domain adaptation; blank images need exclusion
FIVES	800 images (600 training, 200 testing)	2048×2048 (cropped to 512×512)	Large image size and high resolution present computational challenges; variability in fundus appearance due to age, lighting, and disease stages
GlaS	165 images (85 training, 80 testing)	775×522 (resized to 512×512)	Images show high inter-subject and inter-slide variability, requiring multi-scale methods for accurate segmentation

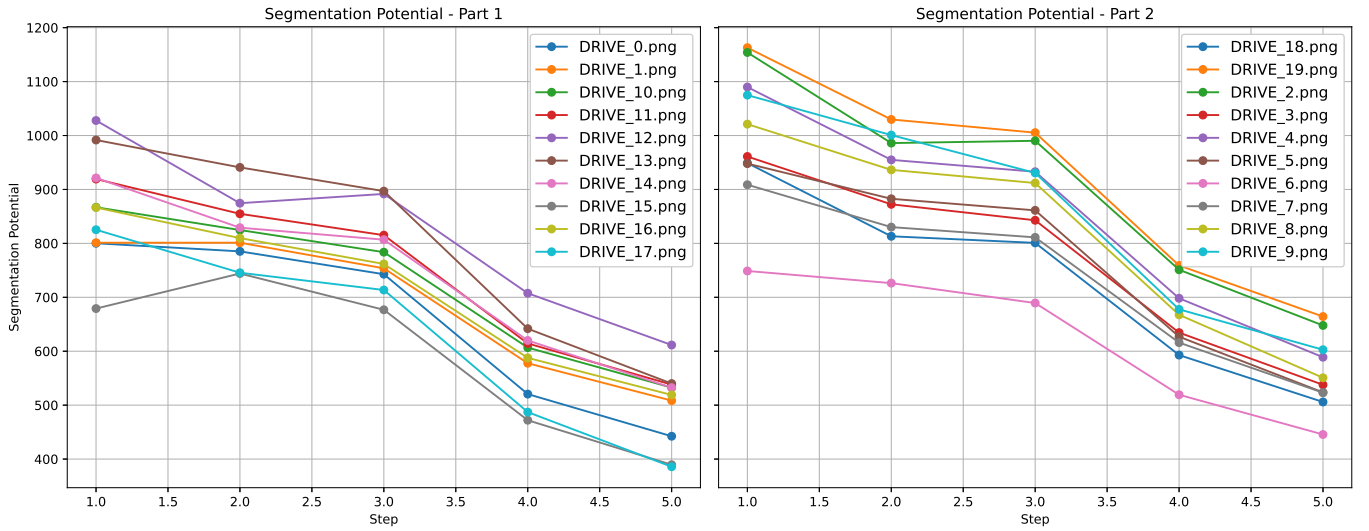


Fig. S5: Trend of the segmentation potential metric for all DRIVE dataset images over different stages of IG module output.

techniques that serve as benchmarks, while the other models incorporate more advanced methodologies, such as improved attention mechanisms, to enhance segmentation performance.

The serially or progressively connected models comprise SerialUnet [13], ConvUNeXt [14], MCDAUNet [15], and ScribFormer [16]. SerialUnet is a widely recognized example of a serially connected architecture, whereas newer models like MCDAUNet focus on addressing specific challenges in medical image segmentation through innovative strategies. In terms of target applications, MISSFormer, MCDAUNet, and SerialUnet are particularly designed for retinal vessel segmentation, while SwinPANet and ScribFormer are specialized in identifying tiny and intricate structures. The remaining models are general-purpose segmentation methods, capable of handling a wide range of medical image segmentation tasks. The comparison between different methods are split into two groups according to the usage of datasets, which consist of three traditional and two challenging datasets.

A. Traditional Datasets

This supplementary mainly provides the detailed results and analysis for the rest two datasets in the traditional datasets. For nuclear segmentation, with results presented in the second and third parts of Table S3, where the best metrics are highlighted in bold and the second-best are underlined. Nuclear segmentation in digital microscopic tissue images is particularly challenging, as cells exhibit complex shapes, slender extensions, irregular boundaries, and diverse morphologies. For MoNuSeg dataset, our proposed method achieves improvements of 0.33%, 0.88%, 0.03%, 0.15%, and 1.29% in the Acc, F1, Rec, Pre, and IoU metrics compared to the second-best results, respectively, underscoring the superior performance of PCNet. The highest Pre and Rec values illustrate that PCNet could effectively capture and segment intricate details without overlooking tiny cell components. Moreover, the highest F1 score indicates that the proposed method accurately detects nuclei while ensuring that identified regions are correct. For CoNIC dataset, the proposed method achieves the highest scores across nearly all evaluated metrics, including Acc,

TABLE S2: The summarize of comparison methods in this paper.

Methods	Year	Source	Category
UNet [6]	2015	MICCAI 2015	Single network
UNet++ [7]	2018	IEEE Trans. Med. Imag.	Single network
SerialUnet [13]	2021	Comput. Meth. Prog. Biomed.	Serially connected network
MISSFormer [9]	2022	IEEE Trans. Med. Imag.	Single network
ConvUNeXt [14]	2022	Knowl.-Based Syst.	Serially connected network
CANet [8]	2023	Biomed. Signal Process. Control	Single network
MCDAUNet [15]	2023	Comput. Biol. Med.	Serially connected network
BRAUNet++ [11]	2024	arxiv	Single network
EMCAD [12]	2024	CVPR 2024	Single network
SwinPANet [10]	2024	IEEE Trans. Neural Networks Learn. Syst.	Single network
ScribFormer [16]	2024	IEEE Trans. Med. Imag.	Serially connected network

F1, Pre, and IoU. Compared to SOTA results, our method shows improvements of 0.10%, 0.57%, 0.97%, and 0.46%, demonstrating the superior capability of PCNet. Although the proposed method performs slightly below the SOTA method MCDAUNet on the Pre metric, the best F1 score demonstrates its balanced segmentation performance. Moreover, the ROC and PR curves of the above two datasets in Fig. S6 further illustrate these findings, highlighting the superior performance of the proposed method.

The visualization of comparative experiments on traditional datasets is presented in Fig. S7. For the DRIVE dataset, as shown in the second row of the first part of the figure, the enlarged region marked by the red box highlights the superior performance of the proposed PCNet in extracting tiny capillaries compared to other methods. Similarly, for the MoNuSeg and CoNIC datasets, particularly in the regions indicated by the yellow arrows in the fourth row, the proposed network demonstrates enhanced sensitivity to tiny targets. In contrast to comparison approaches, PCNet effectively captures more segmented targets while minimizing false detections.

B. Challenge Datasets

The challenge datasets include the FIVES and GlaS datasets. The FIVES dataset is characterized by high-resolution images, which pose challenges for traditional methods in accurately identifying tiny structures within ambiguous regions. Meanwhile, the GlaS dataset features multi-scale targets, creating difficulties for traditional approaches in distinguishing glandular structures of varying sizes within a single medical image.

For the FIVES dataset, the proposed PCNet achieves superior performance across all evaluation metrics, with scores of 98.82%, 84.29%, 83.71%, 87.04%, and 75.87% for Acc, F1, Rec, Pre, and IoU, respectively, surpassing the second-best SOTA method, MCDAUNet [15]. These results demonstrate that even when applied to high-resolution datasets, the proposed PCNet can deliver high-precision segmentation by progressively exploring and unlocking the segmentation potential of the original medical images. Compared to recent methods such as EMCAD, SwinPANet, and ScribFormer, PCNet achieves the highest scores across all metrics, underscor-

ing its robustness and adaptability to complex segmentation tasks. This performance advantage stems primarily from the innovative integration of the dual soft-hard constraint strategy in PCNet, which enhances its ability to capture fine details and segment intricate structures with exceptional accuracy, robustness, and generalization capabilities. The ROC and PR curves shown in the right part of Fig. S8 demonstrating the superior performance of our proposed PCNet.

For gland segmentation, the GlaS dataset, with its multi-scale challenges arising from variable gland sizes and shapes, presented a distinct segmentation task. As shown in the second part of Table S4, PCNet outperforms the second-best method, EMCAD, across all evaluation metrics, demonstrating the superior effectiveness of the proposed approach. Specifically, the Acc and Rec metrics, which assess overall segmentation precision and sensitivity in detecting glands regions, show improvements of 1.40% and 2.04%, respectively, compared to EMCAD [12]. These results highlight the capability of PCNet to accurately capture intricate glandular structures. Additionally, PCNet achieves significant gains in F1 and IoU, with improvements of 1.45% and 2.42%, respectively, compared to EMCAD. Notably, PCNet also achieves a slightly higher Pre metric (91.35%) compared to EMCAD (90.74%), underscoring its robust and consistent performance in balancing precision and recall. These results collectively demonstrate the effectiveness of innovative dual soft-hard constraint strategy in PCNet, enabling it to surpass SOTA method in achieving accurate, robust, and comprehensive segmentation outcomes. Furthermore, the ROC and PR curves, shown in the left part of Fig. S8, provide further evidence of the superior performance of PCNet. The ROC curve demonstrates a larger AUC compared to EMCAD, highlighting the capability of PCNet to distinguish between glandular structures and background regions. Similarly, the PR curve shows a balance point closer to the upper-right corner, indicating the ability of PCNet to achieve high recall while maintaining superior precision—crucial for the accurate segmentation of complex structures in the GlaS dataset.

Additionally, the sample results of comparison experiments on FIVES and GlaS datasets, including UNet [6],

TABLE S3: The comparison experiment results on DRIVE, MoNuSeg, and CoNIC datasets.

Dataset	Methods	Year	Evaluation Metrics (Mean $\pm$ Std)				
			Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
DRIVE	UNet [6]	2015	96.69 $\pm$ 0.01	79.84 $\pm$ 0.29	75.36 $\pm$ 1.26	84.46 $\pm$ 0.99	66.50 $\pm$ 0.39
	UNet++ [7]	2018	96.56 $\pm$ 0.03	79.28 $\pm$ 0.40	75.59 $\pm$ 1.44	83.90 $\pm$ 1.01	65.72 $\pm$ 0.54
	SerialUnet [13]	2021	96.81 $\pm$ 0.01	79.93 $\pm$ 0.22	77.93 $\pm$ 1.14	84.70 $\pm$ 0.90	67.01 $\pm$ 0.31
	MISSFormer [9]	2022	96.51 $\pm$ 0.01	78.72 $\pm$ 0.27	74.09 $\pm$ 1.37	84.54 $\pm$ 1.17	64.94 $\pm$ 0.36
	ConvUNeXt [14]	2022	96.57 $\pm$ 0.02	79.12 $\pm$ 0.19	74.68 $\pm$ 0.62	84.63 $\pm$ 0.47	65.50 $\pm$ 0.25
	CANet [8]	2023	96.68 $\pm$ 0.03	80.03 $\pm$ 0.36	76.35 $\pm$ 1.24	84.94 $\pm$ 0.87	66.74 $\pm$ 0.49
	MCDAUNet [15]	2023	96.78 $\pm$ 0.06	80.70 $\pm$ 0.63	77.24 $\pm$ 2.02	85.03 $\pm$ 1.22	67.69 $\pm$ 0.88
	BRAUNet++ [11]	2024	96.52 $\pm$ 0.03	78.44 $\pm$ 0.51	72.85 $\pm$ 1.73	85.29 $\pm$ 1.25	64.57 $\pm$ 0.69
	EMCAD [12]	2024	96.30 $\pm$ 0.01	77.88 $\pm$ 0.18	74.84 $\pm$ 0.90	81.68 $\pm$ 0.70	63.80 $\pm$ 0.24
	SwinPANet [10]	2024	96.36 $\pm$ 0.04	79.47 $\pm$ 0.14	77.01 $\pm$ 0.19	79.73 $\pm$ 0.30	65.11 $\pm$ 0.18
	ScribFormer [16]	2024	96.34 $\pm$ 0.03	76.92 $\pm$ 0.36	70.12 $\pm$ 1.34	85.69 $\pm$ 1.23	62.55 $\pm$ 0.47
	PCNet (Ours)	-	96.92 $\pm$ 0.01	81.67 $\pm$ 0.24	78.68 $\pm$ 1.25	85.39 $\pm$ 0.96	69.07 $\pm$ 0.34
MoNuSeg	UNet [6]	2015	88.52 $\pm$ 0.15	78.91 $\pm$ 0.25	85.29 $\pm$ 0.92	74.03 $\pm$ 0.64	65.34 $\pm$ 0.31
	UNet++ [7]	2018	88.30 $\pm$ 0.13	78.38 $\pm$ 0.13	85.23 $\pm$ 0.55	73.52 $\pm$ 0.52	64.78 $\pm$ 0.15
	SerialUnet [13]	2021	89.59 $\pm$ 0.17	79.75 $\pm$ 0.26	81.00 $\pm$ 1.17	79.08 $\pm$ 0.98	66.52 $\pm$ 0.34
	MISSFormer [9]	2022	87.91 $\pm$ 0.02	77.70 $\pm$ 0.08	84.12 $\pm$ 0.38	74.07 $\pm$ 0.33	64.16 $\pm$ 0.13
	ConvUNeXt [14]	2022	88.63 $\pm$ 0.16	78.38 $\pm$ 0.18	82.74 $\pm$ 0.73	75.40 $\pm$ 0.63	64.67 $\pm$ 0.21
	CANet [8]	2023	89.82 $\pm$ 0.26	80.71 $\pm$ 0.10	84.67 $\pm$ 1.39	77.84 $\pm$ 1.50	67.80 $\pm$ 0.13
	MCDAUNet [15]	2023	89.61 $\pm$ 0.06	79.74 $\pm$ 0.32	81.71 $\pm$ 1.39	78.32 $\pm$ 0.73	66.49 $\pm$ 0.45
	BRAUNet++ [11]	2024	89.32 $\pm$ 0.46	79.25 $\pm$ 0.83	81.88 $\pm$ 2.30	78.09 $\pm$ 2.03	65.86 $\pm$ 1.07
	EMCAD [12]	2024	89.96 $\pm$ 0.06	81.01 $\pm$ 0.14	84.65 $\pm$ 0.33	78.07 $\pm$ 0.09	68.16 $\pm$ 0.20
	SwinPANet [10]	2024	86.62 $\pm$ 0.30	79.37 $\pm$ 0.35	81.53 $\pm$ 0.08	78.87 $\pm$ 0.36	65.54 $\pm$ 0.39
	ScribFormer [16]	2024	89.42 $\pm$ 0.05	79.34 $\pm$ 0.74	72.89 $\pm$ 3.06	79.05 $\pm$ 1.87	65.87 $\pm$ 0.97
	PCNet (Ours)	-	90.29 $\pm$ 0.35	81.89 $\pm$ 0.48	85.32 $\pm$ 0.34	79.23 $\pm$ 1.12	69.45 $\pm$ 0.68
CoNIC	UNet [6]	2015	92.92 $\pm$ 0.05	75.28 $\pm$ 0.41	73.43 $\pm$ 1.86	78.32 $\pm$ 1.25	61.18 $\pm$ 0.52
	UNet++ [7]	2018	92.95 $\pm$ 0.04	75.45 $\pm$ 0.35	74.37 $\pm$ 1.03	77.63 $\pm$ 0.58	61.46 $\pm$ 0.39
	SerialUnet [13]	2021	93.01 $\pm$ 0.04	75.24 $\pm$ 0.41	73.05 $\pm$ 1.32	78.77 $\pm$ 0.66	61.18 $\pm$ 0.51
	MISSFormer [9]	2022	92.77 $\pm$ 0.06	75.27 $\pm$ 0.29	74.91 $\pm$ 1.64	77.05 $\pm$ 1.24	61.16 $\pm$ 0.37
	ConvUNeXt [14]	2022	91.89 $\pm$ 0.02	71.68 $\pm$ 0.25	70.75 $\pm$ 1.08	73.92 $\pm$ 0.75	56.84 $\pm$ 0.33
	CANet [8]	2023	93.37 $\pm$ 0.02	77.34 $\pm$ 0.30	77.04 $\pm$ 1.45	78.37 $\pm$ 0.89	63.79 $\pm$ 0.40
	MCDAUNet [15]	2023	93.83 $\pm$ 0.06	78.53 $\pm$ 0.33	78.04 $\pm$ 0.96	80.10 $\pm$ 0.75	65.70 $\pm$ 0.37
	BRAUNet++ [11]	2024	92.08 $\pm$ 0.02	73.04 $\pm$ 0.29	73.73 $\pm$ 1.21	73.53 $\pm$ 0.63	58.37 $\pm$ 0.36
	EMCAD [12]	2024	93.22 $\pm$ 0.01	76.90 $\pm$ 0.10	75.72 $\pm$ 0.42	78.72 $\pm$ 0.27	63.07 $\pm$ 0.13
	SwinPANet [10]	2024	92.70 $\pm$ 0.01	75.76 $\pm$ 0.15	75.56 $\pm$ 0.66	77.52 $\pm$ 0.39	62.43 $\pm$ 0.19
	ScribFormer [16]	2024	92.96 $\pm$ 0.04	74.66 $\pm$ 0.26	70.54 $\pm$ 1.26	80.65 $\pm$ 1.09	60.34 $\pm$ 0.31
	PCNet (Ours)	-	93.93 $\pm$ 0.02	79.10 $\pm$ 0.29	77.95 $\pm$ 1.13	81.07 $\pm$ 0.74	66.16 $\pm$ 0.36

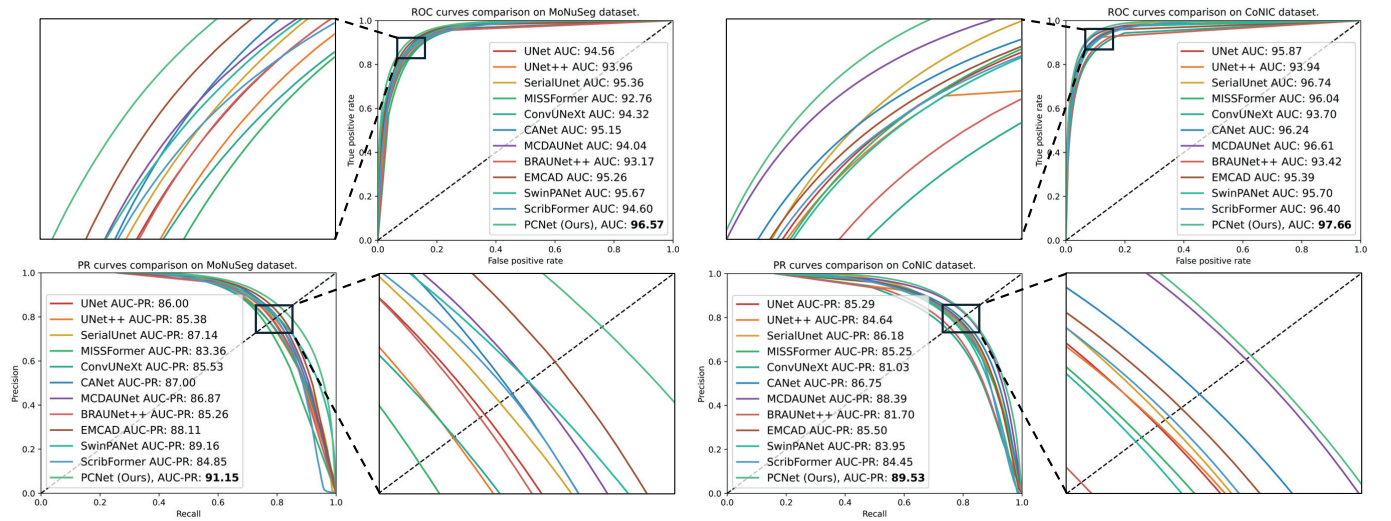


Fig. S6: The visualization of ROC and PR curves of comparison experiments on MoNuSeg and CoNIC datasets.

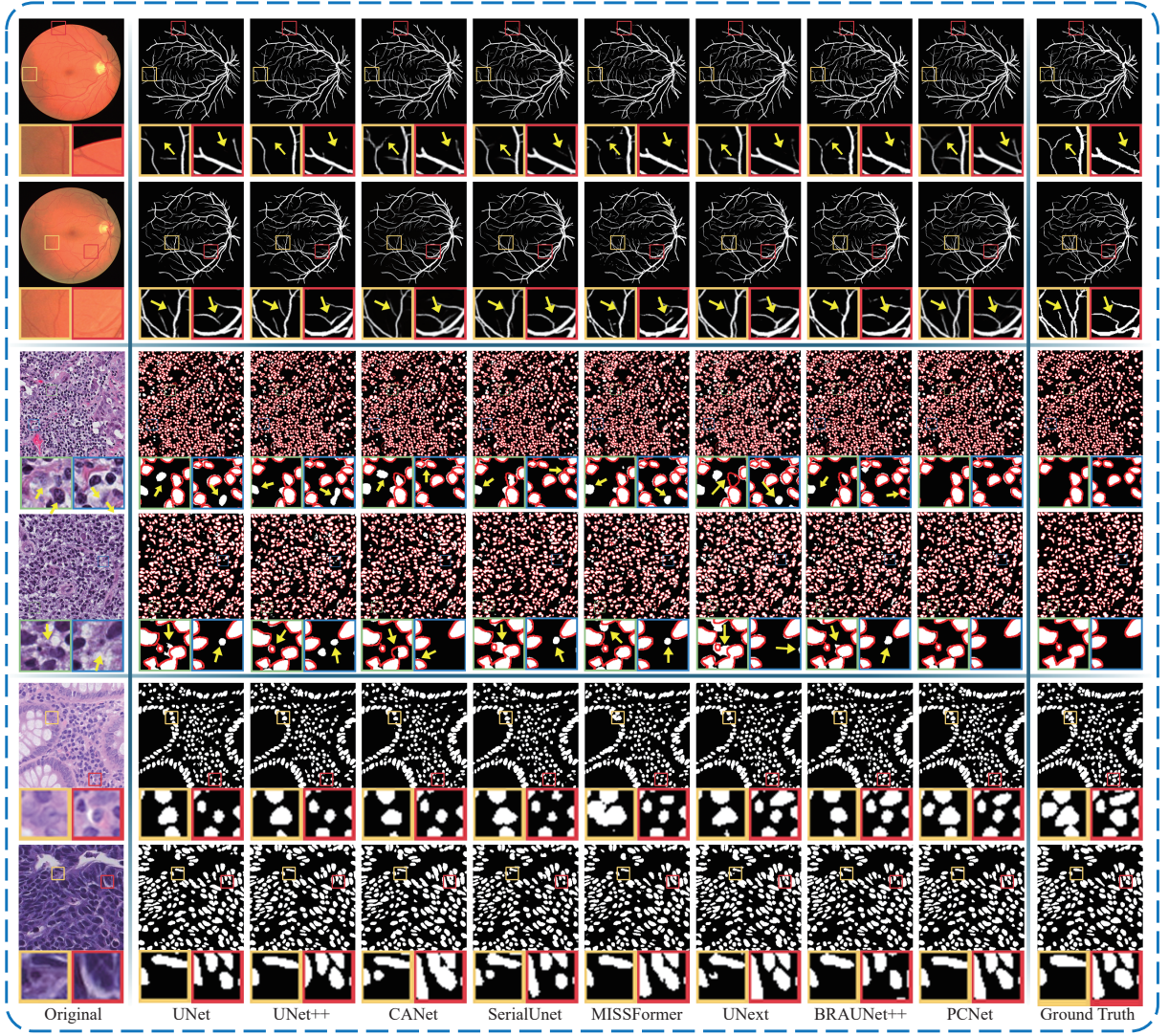


Fig. S7: Samples from the DRIVE, MoNuSeg, and CoNIC datasets. First and second rows show sample from the DRIVE dataset, third and fourth rows show image from the MoNuSeg dataset, while the last two rows show sample from the CoNIC dataset.

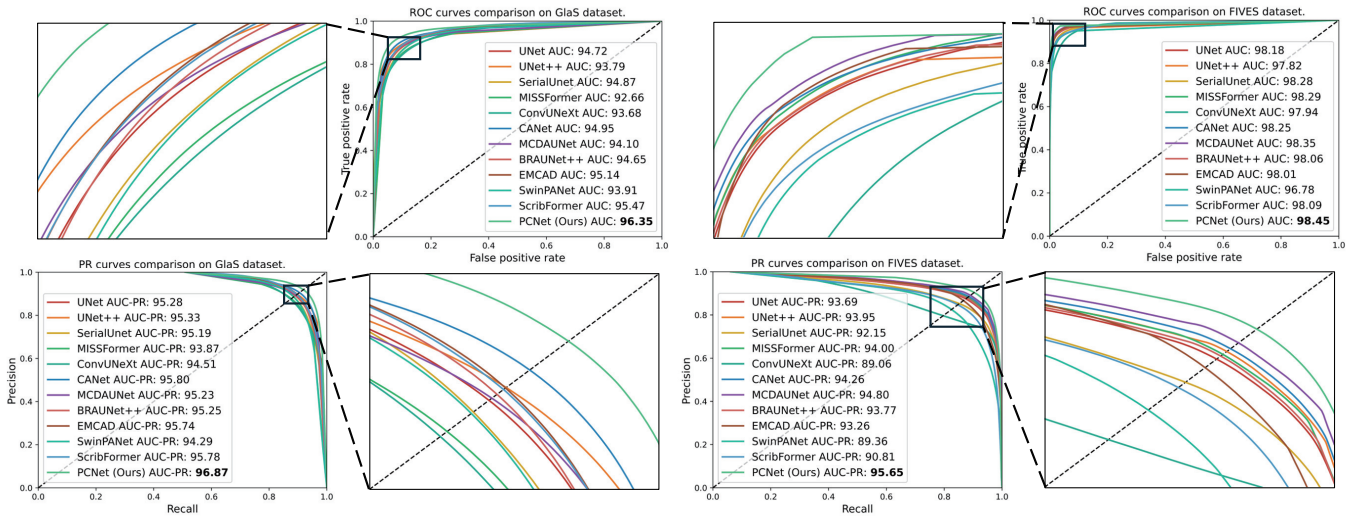


Fig. S8: The visualization of ROC and PR curves of comparison experiments on GlaS and FIVES datasets.

TABLE S4: The comparison experiment results on FIVES and GlaS datasets.

Datasets	Methods	Year	Evaluation Metrics (Mean $\pm$ Std)				
			Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
FIVES	UNet [6]	2015	98.61 $\pm$ 0.02	82.05 $\pm$ 0.27	80.02 $\pm$ 0.63	86.36 $\pm$ 0.76	72.32 $\pm$ 0.20
	UNet++ [7]	2018	98.68 $\pm$ 0.01	82.13 $\pm$ 0.16	80.65 $\pm$ 0.45	84.87 $\pm$ 1.17	73.00 $\pm$ 0.16
	SerialUnet [13]	2021	98.47 $\pm$ 0.03	80.90 $\pm$ 0.45	80.21 $\pm$ 1.63	85.16 $\pm$ 1.45	70.82 $\pm$ 0.60
	MISSFormer [9]	2022	98.65 $\pm$ 0.01	83.17 $\pm$ 0.13	81.66 $\pm$ 0.74	86.29 $\pm$ 0.98	73.50 $\pm$ 0.18
	ConvUNeXt [14]	2022	98.44 $\pm$ 0.03	79.24 $\pm$ 0.38	74.67 $\pm$ 1.21	86.42 $\pm$ 1.83	68.79 $\pm$ 0.48
	CANet [8]	2023	98.71 $\pm$ 0.00	82.40 $\pm$ 0.19	80.70 $\pm$ 0.71	85.17 $\pm$ 1.98	73.45 $\pm$ 0.20
	MCDAUNet [15]	2023	98.76 $\pm$ 0.01	84.00 $\pm$ 0.20	83.02 $\pm$ 0.60	86.59 $\pm$ 0.67	74.98 $\pm$ 0.19
	BRAUNet++ [11]	2024	98.62 $\pm$ 0.00	82.32 $\pm$ 0.20	80.64 $\pm$ 0.23	86.86 $\pm$ 0.32	72.64 $\pm$ 0.13
	EMCAD [12]	2024	98.40 $\pm$ 0.00	82.02 $\pm$ 0.14	81.04 $\pm$ 0.27	83.46 $\pm$ 0.23	71.14 $\pm$ 0.11
	SwinPANet [10]	2024	97.58 $\pm$ 0.13	82.20 $\pm$ 0.25	81.12 $\pm$ 0.25	83.42 $\pm$ 0.56	68.59 $\pm$ 0.44
	ScribFormer [16]	2024	98.12 $\pm$ 0.03	79.63 $\pm$ 0.97	71.57 $\pm$ 2.23	86.96 $\pm$ 2.98	65.94 $\pm$ 0.98
	PCNet (Ours)	-	98.82 $\pm$ 0.04	84.29 $\pm$ 0.70	83.71 $\pm$ 0.11	87.04 $\pm$ 1.83	75.87 $\pm$ 0.83
GlaS	UNet [6]	2015	89.02 $\pm$ 0.06	88.24 $\pm$ 0.09	88.44 $\pm$ 0.59	89.05 $\pm$ 0.48	79.78 $\pm$ 0.14
	UNet++ [7]	2018	89.76 $\pm$ 0.08	89.14 $\pm$ 0.09	89.02 $\pm$ 0.57	90.35 $\pm$ 0.63	81.37 $\pm$ 0.15
	SerialUnet [13]	2021	88.33 $\pm$ 0.18	87.51 $\pm$ 0.22	88.00 $\pm$ 1.15	88.28 $\pm$ 0.82	78.84 $\pm$ 0.29
	MISSFormer [9]	2022	87.37 $\pm$ 0.20	86.92 $\pm$ 0.28	87.86 $\pm$ 0.45	87.73 $\pm$ 0.34	78.33 $\pm$ 0.36
	ConvUNeXt [14]	2022	87.42 $\pm$ 0.04	86.67 $\pm$ 0.10	87.25 $\pm$ 0.43	87.36 $\pm$ 0.37	77.43 $\pm$ 0.13
	CANet [8]	2023	90.43 $\pm$ 0.08	89.90 $\pm$ 0.12	89.93 $\pm$ 0.82	89.90 $\pm$ 0.70	82.59 $\pm$ 0.16
	MCDAUNet [15]	2023	89.11 $\pm$ 0.10	88.36 $\pm$ 0.16	88.83 $\pm$ 0.61	89.45 $\pm$ 0.77	80.40 $\pm$ 0.22
	BRAUNet++ [11]	2024	89.23 $\pm$ 0.09	88.84 $\pm$ 0.12	90.34 $\pm$ 0.41	88.40 $\pm$ 0.26	80.73 $\pm$ 0.15
	EMCAD [12]	2024	90.60 $\pm$ 0.27	90.32 $\pm$ 0.31	90.90 $\pm$ 1.14	90.74 $\pm$ 1.12	83.10 $\pm$ 0.43
	SwinPANet [10]	2024	88.18 $\pm$ 0.15	87.39 $\pm$ 0.24	88.61 $\pm$ 0.70	87.58 $\pm$ 0.56	78.70 $\pm$ 0.30
	ScribFormer [16]	2024	89.29 $\pm$ 0.15	89.13 $\pm$ 0.15	90.51 $\pm$ 0.87	88.79 $\pm$ 0.85	81.21 $\pm$ 0.20
	PCNet (Ours)	-	92.00 $\pm$ 0.18	91.77 $\pm$ 0.21	92.94 $\pm$ 0.67	91.35 $\pm$ 0.65	85.52 $\pm$ 0.30

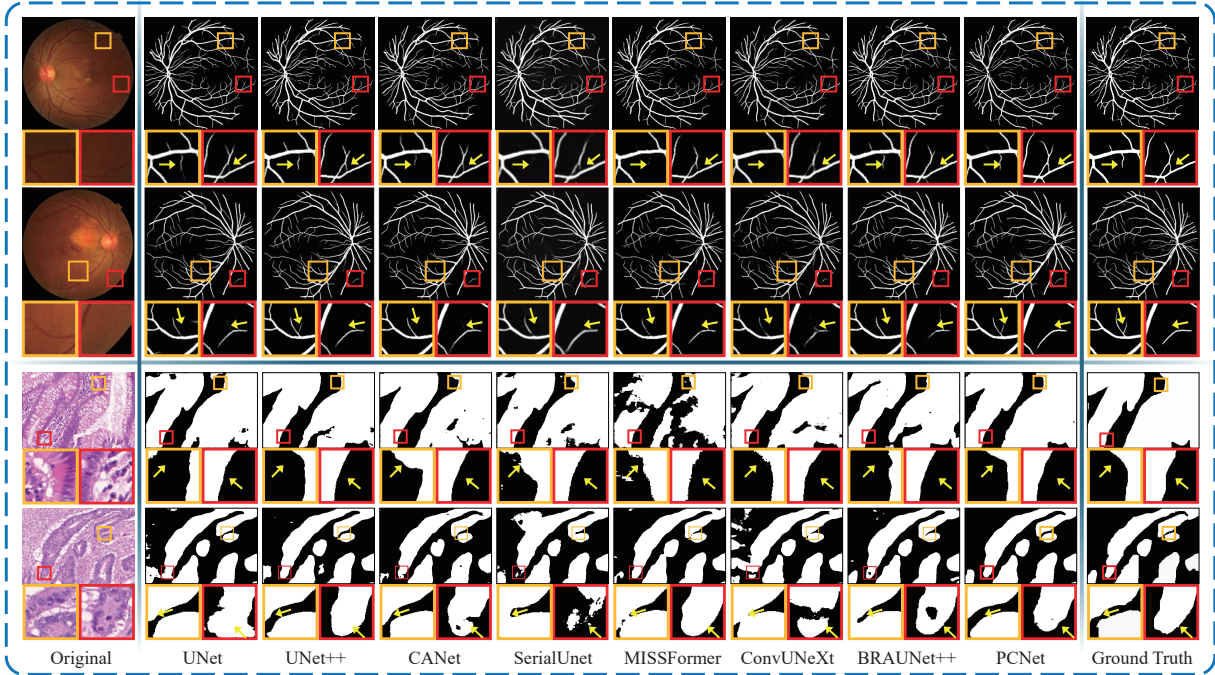


Fig. S9: Visualization of the changing trends across all evaluation metrics for four different datasets. All values are normalized, and the curves are smoothed using an interpolation technique for clarity.

UNet++ [7], CANet [8], SerialUnet [13], MISSFormer [9], ConvUNet [14], BRAUNet++ [11], and the proposed PCNet, are illustrated in Fig. S9. It is evident that PCNet effectively captures relevant structures in medical images, achieving results that closely resemble the original ground truth. For retinal vessel segmentation, when operating on high-resolution data, the proposed PCNet demonstrates superior performance, particularly in detecting capillaries that are challenging to identify in high-resolution images. As highlighted by the yellow arrow within the red-boxed region in the first row of Fig. S9, PCNet effectively extracts the capillary structures at the terminal ends of retinal vessels, outperforming other comparison approaches. For gland segmentation, the presence of multi-scale targets in the images makes segmentation a challenging task. As indicated by the yellow arrow in the yellow-boxed regions in Fig. S9, the proposed PCNet efficiently detects gland edges, showcasing its ability to handle complex and varying scales in gland structures more effectively than the comparison methods.

III. NETWORK EXPLORATION

A. Ablation Study

The proposed PCNet architecture incorporates two novel modules, termed the IG and PI modules. To evaluate their contributions to the overall performance, a series of ablation experiments are conducted on five datasets. This section presents the results and analyses for the remaining four datasets. In these experiments, the IG and PI modules are replaced with SerialUnet [13], a baseline network with similar convolutional depth and a serially connected structure.

The results for the MoNuSeg dataset are summarized in the first part of Table S5. A comparison between Rows 1 and 2 reveals that replacing SerialUnet with the IG module leads to substantial improvements, particularly in the F1 score. These findings highlight the crucial role of the IG module in enhancing feature extraction and segmentation precision. By generating progressively refined intermediate maps, the IG module proves to be well-suited for detailed segmentation tasks. Similarly, comparing Rows 1 and 3 shows that the inclusion of the PI module significantly enhances performance when substituting SerialUnet. Notable improvements in segmentation accuracy, especially IoU, emphasize the effectiveness of PI module in capturing spatial context and refining boundaries. These gains can be attributed to the dual soft-hard constraint strategy employed in the PCT blocks, which enables precise delineation of complex structures. For the CoNIC dataset, similar trends are observed when IG and PI modules replace SerialUnet. As shown in the second part of Table S5, both modules markedly improve performance. By refining intermediate segmentation maps, the IG module ensures comprehensive feature representation and higher segmentation quality, addressing the challenge of varying data sources in the CoNIC dataset. Additionally, replacing SerialUnet with the PI module results in significant performance gains, further demonstrating the robustness and versatility of these modules in diverse and challenging datasets.

For the FIVES dataset, the inclusion of IG and PI modules shows a consistent trend of improved performance compared

to SerialUnet. As presented in the third part of Table S5, the PI module significantly enhances segmentation metrics, particularly the balanced F1 score. This improvement reflects the capability of PI module to progressively refine segmentation maps, ensuring the accurate capture of fine vessel details, which is a critical requirement for this high-resolution dataset. The IG module also achieves notable improvements compared to SerialUnet, demonstrating its robustness in accurately segmenting vessel boundaries. Together, the IG and PI modules enable superior segmentation performance, effectively addressing the high-resolution challenges posed by the FIVES dataset. Ablation experiments on the GlaS dataset further validate the efficacy of the IG and PI modules. The IG module demonstrates significant improvements in evaluation metrics, reflecting its ability to enhance feature extraction and segmentation quality for intricate gland structures. The progressive refinement of intermediate segmentation maps ensures the preservation of essential features, helping to overcome the unique challenges of this dataset. Similarly, replacing the baseline network with the PI module yields marked improvements, highlighting its capability to accurately segment complex structures. The dual soft-hard constraints strategy facilitates iterative refinement, enabling precise delineation of tiny and detailed structures. These results confirm the critical role of the PI module in improving segmentation accuracy, particularly for the challenge of multi-scale segmentation posed by the GlaS dataset.

Across all datasets, the proposed IG and PI modules demonstrate significant improvements over the baseline SerialUnet. These findings highlight the superior ability of modules.

B. Hyper-parameters Selection

The hyper-parameters in PCNet include the parameter α , which adjusts the proportion of soft and hard connection in dual soft-hard constraint strategy, and the depth parameter k , which specifies the depth of the IG and PI modules. Also, to ensure the robustness of our experiments, we have fixed k at 5 while assessing the impact of α , and set α to 0.5 during the evaluation of k .

1) *Experiments and Analyses for α* : The detailed results of hyper-parameter experiments for α on the traditional datasets are illustrated in Table S6. Consistent with earlier observations, the MoNuSeg and CoNIC datasets exhibit a clear trend where performance metrics initially improve with increasing α before declining. For the MoNuSeg dataset, a progressive improvement in evaluation metrics is observed as α increases, reaching optimal performance at $\alpha = 0.5$. Specifically, Acc attains 90.15%, F1 peaks at 80.76%, Rec achieves 85.83%, and IoU reaches a maximum of 69.26%. Beyond this threshold, these metrics decline, suggesting that a balanced integration of soft and hard connections allows the network to capture a more comprehensive range of specific features or constraints, thereby enhancing its overall performance. On the CoNIC dataset, characterized by its low resolution and large dataset size, a similar pattern of performance enhancement followed by decline is observed, though it is less pronounced compared to MoNuSeg. Performance fluctuations across all metrics are

TABLE S5: The results of experiments on ablation study for four different datasets.

Dataset	Modules		Evaluation Metrics (Mean $\pm$ Std)				
	IG	PI	Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
MoNuSeg	-	-	89.86 $\pm$ 0.25	80.48 $\pm$ 0.57	83.57 $\pm$ 2.67	77.72 $\pm$ 1.46	67.86 $\pm$ 0.80
	✓	-	90.04 $\pm$ 0.11	81.00 $\pm$ 0.25	84.27 $\pm$ 0.67	77.98 $\pm$ 0.21	68.60 $\pm$ 0.35
	-	✓	90.13 $\pm$ 0.22	81.81 $\pm$ 0.57	84.33 $\pm$ 1.37	78.51 $\pm$ 0.93	68.37 $\pm$ 0.73
	✓	✓	90.29 $\pm$ 0.35	81.89 $\pm$ 0.48	85.32 $\pm$ 0.34	79.23 $\pm$ 1.12	69.45 $\pm$ 0.68
CoNIC	-	-	93.12 $\pm$ 0.06	77.00 $\pm$ 0.65	76.82 $\pm$ 2.75	79.28 $\pm$ 1.62	63.43 $\pm$ 0.77
	✓	-	93.49 $\pm$ 0.10	78.29 $\pm$ 0.03	77.13 $\pm$ 1.13	80.65 $\pm$ 0.99	65.29 $\pm$ 0.05
	-	✓	93.33 $\pm$ 0.19	78.27 $\pm$ 0.46	77.75 $\pm$ 2.83	80.08 $\pm$ 2.16	65.74 $\pm$ 0.58
	✓	✓	93.93 $\pm$ 0.02	79.10 $\pm$ 0.29	77.95 $\pm$ 1.13	81.07 $\pm$ 0.74	66.16 $\pm$ 0.36
FIVES	-	-	98.57 $\pm$ 0.05	81.88 $\pm$ 0.68	81.24 $\pm$ 2.29	84.77 $\pm$ 2.71	72.96 $\pm$ 0.60
	✓	-	98.66 $\pm$ 0.02	83.23 $\pm$ 0.23	82.68 $\pm$ 0.73	85.87 $\pm$ 2.65	73.01 $\pm$ 0.29
	-	✓	98.71 $\pm$ 0.02	83.57 $\pm$ 0.14	82.19 $\pm$ 1.21	85.19 $\pm$ 2.25	73.09 $\pm$ 0.18
	✓	✓	98.82 $\pm$ 0.04	84.29 $\pm$ 0.70	83.71 $\pm$ 0.11	87.04 $\pm$ 1.83	75.87 $\pm$ 0.83
GlaS	-	-	90.45 $\pm$ 0.09	90.22 $\pm$ 0.16	91.10 $\pm$ 1.46	90.11 $\pm$ 1.33	82.95 $\pm$ 0.18
	✓	-	91.70 $\pm$ 0.16	91.53 $\pm$ 0.16	92.11 $\pm$ 0.62	90.70 $\pm$ 0.89	85.05 $\pm$ 0.27
	-	✓	90.74 $\pm$ 0.19	90.77 $\pm$ 0.23	91.95 $\pm$ 0.58	90.47 $\pm$ 0.53	85.33 $\pm$ 0.35
	✓	✓	92.00 $\pm$ 0.18	91.77 $\pm$ 0.21	92.94 $\pm$ 0.67	91.35 $\pm$ 0.65	85.52 $\pm$ 0.30

TABLE S6: The evaluation metrics for different α on traditional datasets.

Datasets	α	Evaluation Metrics (Mean $\pm$ Std)				
		Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
MoNuSeg	0.1	89.44 $\pm$ 0.18	79.98 $\pm$ 0.40	81.29 $\pm$ 1.27	77.56 $\pm$ 0.44	67.48 $\pm$ 0.59
	0.2	89.52 $\pm$ 0.05	80.06 $\pm$ 0.57	81.72 $\pm$ 1.41	79.03 $\pm$ 0.63	67.86 $\pm$ 0.55
	0.3	89.56 $\pm$ 0.09	80.03 $\pm$ 0.50	81.60 $\pm$ 1.94	79.46 $\pm$ 1.51	67.95 $\pm$ 0.66
	0.4	89.65 $\pm$ 0.24	80.38 $\pm$ 0.53	83.86 $\pm$ 2.56	77.09 $\pm$ 1.68	68.02 $\pm$ 0.75
	0.5	90.15 $\pm$ 0.03	80.76 $\pm$ 0.14	85.83 $\pm$ 0.84	78.47 $\pm$ 0.48	69.26 $\pm$ 0.19
	0.6	89.37 $\pm$ 0.36	80.46 $\pm$ 0.53	85.32 $\pm$ 1.33	76.61 $\pm$ 1.27	68.52 $\pm$ 0.72
	0.7	88.99 $\pm$ 0.29	80.08 $\pm$ 0.36	84.22 $\pm$ 0.48	74.79 $\pm$ 0.32	67.78 $\pm$ 0.26
	0.8	88.83 $\pm$ 0.18	80.02 $\pm$ 0.39	84.11 $\pm$ 1.02	75.57 $\pm$ 1.45	67.89 $\pm$ 0.53
	0.9	88.33 $\pm$ 0.23	79.91 $\pm$ 0.20	83.28 $\pm$ 0.68	77.05 $\pm$ 0.28	68.02 $\pm$ 0.49
CoNIC	0.1	93.49 $\pm$ 0.08	79.10 $\pm$ 0.18	75.06 $\pm$ 0.67	80.23 $\pm$ 0.56	66.05 $\pm$ 0.22
	0.2	93.48 $\pm$ 0.12	78.95 $\pm$ 0.27	75.61 $\pm$ 0.83	80.33 $\pm$ 0.71	66.03 $\pm$ 0.34
	0.3	93.57 $\pm$ 0.09	79.08 $\pm$ 0.14	75.71 $\pm$ 0.73	80.57 $\pm$ 0.75	66.06 $\pm$ 0.19
	0.4	93.70 $\pm$ 0.12	79.03 $\pm$ 0.25	76.70 $\pm$ 1.45	80.32 $\pm$ 1.89	66.09 $\pm$ 0.28
	0.5	93.93 $\pm$ 0.02	79.10 $\pm$ 0.29	77.95 $\pm$ 1.13	81.07 $\pm$ 0.74	66.16 $\pm$ 0.36
	0.6	93.72 $\pm$ 0.06	78.98 $\pm$ 0.21	77.60 $\pm$ 0.65	80.64 $\pm$ 1.01	65.91 $\pm$ 0.24
	0.7	93.66 $\pm$ 0.13	79.05 $\pm$ 0.19	76.82 $\pm$ 1.04	80.49 $\pm$ 1.28	66.03 $\pm$ 0.22
	0.8	93.51 $\pm$ 0.09	78.98 $\pm$ 0.01	75.93 $\pm$ 0.76	80.79 $\pm$ 0.98	65.89 $\pm$ 0.12
	0.9	93.42 $\pm$ 0.13	78.93 $\pm$ 0.23	75.69 $\pm$ 1.12	80.42 $\pm$ 1.48	65.82 $\pm$ 0.29

TABLE S7: The evaluation metrics for different α on challenging datasets.

Dataset	α	Evaluation Metrics (Mean $\pm$ Std)				
		Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
FIVES	0.1	98.58 $\pm$ 0.01	83.47 $\pm$ 0.46	83.40 $\pm$ 0.47	86.24 $\pm$ 2.10	73.42 $\pm$ 0.36
	0.2	98.58 $\pm$ 0.02	83.53 $\pm$ 0.26	83.51 $\pm$ 0.81	86.39 $\pm$ 1.59	73.55 $\pm$ 0.14
	0.3	98.67 $\pm$ 0.02	83.76 $\pm$ 0.42	83.25 $\pm$ 0.92	86.65 $\pm$ 1.00	73.61 $\pm$ 0.27
	0.4	98.77 $\pm$ 0.01	84.35 $\pm$ 0.52	83.49 $\pm$ 0.88	86.63 $\pm$ 2.08	73.24 $\pm$ 0.36
	0.5	98.82 $\pm$ 0.04	84.29 $\pm$ 0.70	83.71 $\pm$ 0.11	87.04 $\pm$ 1.83	75.87 $\pm$ 0.83
	0.6	98.76 $\pm$ 0.01	84.14 $\pm$ 0.04	83.29 $\pm$ 0.70	86.80 $\pm$ 1.81	73.20 $\pm$ 0.33
	0.7	98.68 $\pm$ 0.02	83.74 $\pm$ 0.12	83.38 $\pm$ 0.76	86.32 $\pm$ 1.06	73.19 $\pm$ 0.13
	0.8	98.58 $\pm$ 0.02	83.54 $\pm$ 0.31	83.33 $\pm$ 0.78	86.30 $\pm$ 0.77	73.64 $\pm$ 0.26
	0.9	98.56 $\pm$ 0.02	83.44 $\pm$ 0.51	83.05 $\pm$ 0.63	86.16 $\pm$ 2.27	73.19 $\pm$ 0.46
GlaS	0.1	91.05 $\pm$ 0.14	90.79 $\pm$ 0.14	92.11 $\pm$ 0.44	90.31 $\pm$ 0.42	85.61 $\pm$ 0.21
	0.2	91.78 $\pm$ 0.19	91.52 $\pm$ 0.31	92.98 $\pm$ 0.62	90.89 $\pm$ 0.79	85.10 $\pm$ 0.47
	0.3	92.17 $\pm$ 0.14	90.99 $\pm$ 0.12	92.88 $\pm$ 0.43	90.81 $\pm$ 0.59	85.87 $\pm$ 0.21
	0.4	92.01 $\pm$ 0.08	90.87 $\pm$ 0.13	93.64 $\pm$ 0.50	90.82 $\pm$ 0.59	85.60 $\pm$ 0.19
	0.5	92.00 $\pm$ 0.18	91.77 $\pm$ 0.21	92.94 $\pm$ 0.67	91.35 $\pm$ 0.65	85.52 $\pm$ 0.30
	0.6	91.88 $\pm$ 0.19	90.69 $\pm$ 0.25	92.90 $\pm$ 0.31	90.33 $\pm$ 0.66	85.44 $\pm$ 0.37
	0.7	91.91 $\pm$ 0.09	90.67 $\pm$ 0.02	93.11 $\pm$ 0.65	90.15 $\pm$ 0.43	85.40 $\pm$ 0.29
	0.8	91.96 $\pm$ 0.09	90.84 $\pm$ 0.09	93.34 $\pm$ 0.45	90.07 $\pm$ 0.52	85.60 $\pm$ 0.15
	0.9	92.00 $\pm$ 0.23	90.91 $\pm$ 0.29	93.29 $\pm$ 0.63	90.30 $\pm$ 0.59	85.74 $\pm$ 0.42

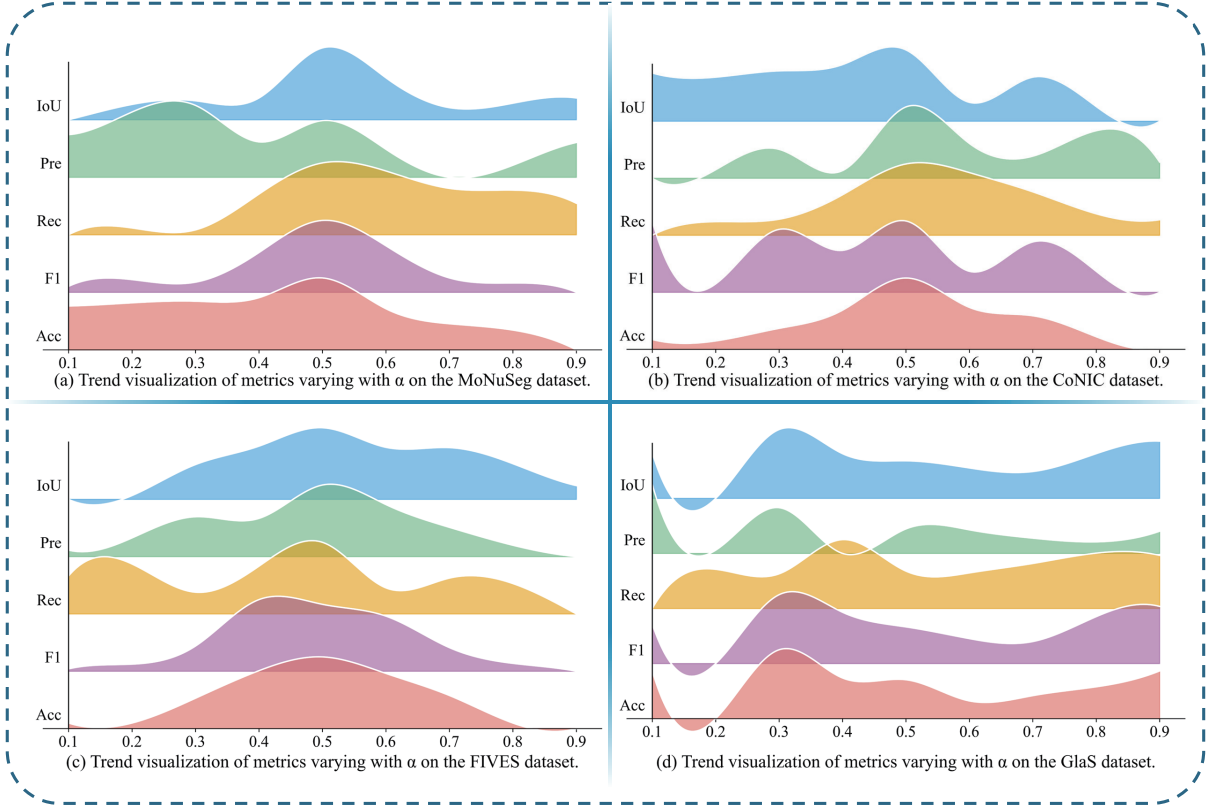


Fig. S10: Visualization of the trends across all evaluation metrics for four different datasets as α changes. All values are normalized, and the curves are smoothed using an interpolation technique for better clarity.

TABLE S8: The results of hyper-parameter k for four different datasets.

Dataset	Depth (k)	Evaluation Metrics (Mean $\pm$ Std)				
		Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
MoNuSeg	3	89.42 $\pm$ 0.30	79.55 $\pm$ 0.21	81.76 $\pm$ 1.02	78.63 $\pm$ 1.01	67.36 $\pm$ 0.25
	4	89.71 $\pm$ 0.09	80.71 $\pm$ 0.15	83.99 $\pm$ 1.04	78.36 $\pm$ 0.72	68.86 $\pm$ 0.19
	5	90.15 $\pm$ 0.03	80.76 $\pm$ 0.14	85.83 $\pm$ 0.84	78.47 $\pm$ 0.48	69.26 $\pm$ 0.19
	6	89.59 $\pm$ 0.21	80.48 $\pm$ 0.26	84.65 $\pm$ 1.89	78.54 $\pm$ 1.32	69.11 $\pm$ 0.37
	7	89.39 $\pm$ 0.06	80.37 $\pm$ 0.07	83.81 $\pm$ 0.42	77.93 $\pm$ 0.41	68.13 $\pm$ 0.09
	8	89.27 $\pm$ 0.06	80.17 $\pm$ 0.18	83.19 $\pm$ 0.59	76.97 $\pm$ 0.22	67.29 $\pm$ 0.26
CoNIC	3	93.33 $\pm$ 0.11	78.25 $\pm$ 0.17	78.17 $\pm$ 1.54	80.42 $\pm$ 1.24	65.48 $\pm$ 0.20
	4	93.55 $\pm$ 0.15	78.62 $\pm$ 1.18	78.92 $\pm$ 0.91	80.87 $\pm$ 3.02	65.97 $\pm$ 1.54
	5	93.93 $\pm$ 0.02	79.10 $\pm$ 0.29	77.95 $\pm$ 1.13	81.07 $\pm$ 0.74	66.16 $\pm$ 0.36
	6	93.57 $\pm$ 0.10	78.16 $\pm$ 0.49	76.80 $\pm$ 1.75	80.89 $\pm$ 0.93	65.83 $\pm$ 0.63
	7	93.45 $\pm$ 0.09	78.05 $\pm$ 0.54	76.54 $\pm$ 2.23	80.02 $\pm$ 1.09	65.67 $\pm$ 0.69
	8	93.20 $\pm$ 0.13	78.31 $\pm$ 0.35	77.01 $\pm$ 1.56	79.93 $\pm$ 1.13	64.96 $\pm$ 0.45
FIVES	3	98.76 $\pm$ 0.01	83.52 $\pm$ 0.37	82.42 $\pm$ 0.96	85.64 $\pm$ 1.66	74.12 $\pm$ 0.27
	4	98.87 $\pm$ 0.01	83.57 $\pm$ 0.15	82.97 $\pm$ 0.62	86.32 $\pm$ 1.02	75.29 $\pm$ 0.07
	5	98.82 $\pm$ 0.04	84.29 $\pm$ 0.70	83.71 $\pm$ 0.11	87.04 $\pm$ 1.83	75.87 $\pm$ 0.83
	6	98.75 $\pm$ 0.01	84.33 $\pm$ 0.47	83.19 $\pm$ 1.14	87.01 $\pm$ 1.29	75.02 $\pm$ 0.42
	7	98.68 $\pm$ 0.02	83.75 $\pm$ 0.27	82.19 $\pm$ 0.89	86.86 $\pm$ 1.98	74.46 $\pm$ 0.22
	8	98.57 $\pm$ 0.01	82.86 $\pm$ 0.36	82.09 $\pm$ 0.41	86.27 $\pm$ 1.76	73.58 $\pm$ 0.19
GlaS	3	91.02 $\pm$ 0.15	90.98 $\pm$ 0.16	92.28 $\pm$ 0.41	90.36 $\pm$ 0.45	84.74 $\pm$ 0.28
	4	91.40 $\pm$ 0.09	91.14 $\pm$ 0.18	92.86 $\pm$ 0.58	91.20 $\pm$ 0.28	85.47 $\pm$ 0.27
	5	92.00 $\pm$ 0.18	91.77 $\pm$ 0.21	92.94 $\pm$ 0.67	91.35 $\pm$ 0.65	85.52 $\pm$ 0.03
	6	91.73 $\pm$ 0.05	91.69 $\pm$ 0.12	92.42 $\pm$ 1.61	91.77 $\pm$ 1.42	85.39 $\pm$ 0.17
	7	91.12 $\pm$ 0.04	91.03 $\pm$ 0.10	92.19 $\pm$ 0.04	91.52 $\pm$ 0.43	84.87 $\pm$ 0.16
	8	91.01 $\pm$ 0.15	90.70 $\pm$ 0.15	91.93 $\pm$ 0.34	90.70 $\pm$ 0.34	84.37 $\pm$ 0.02

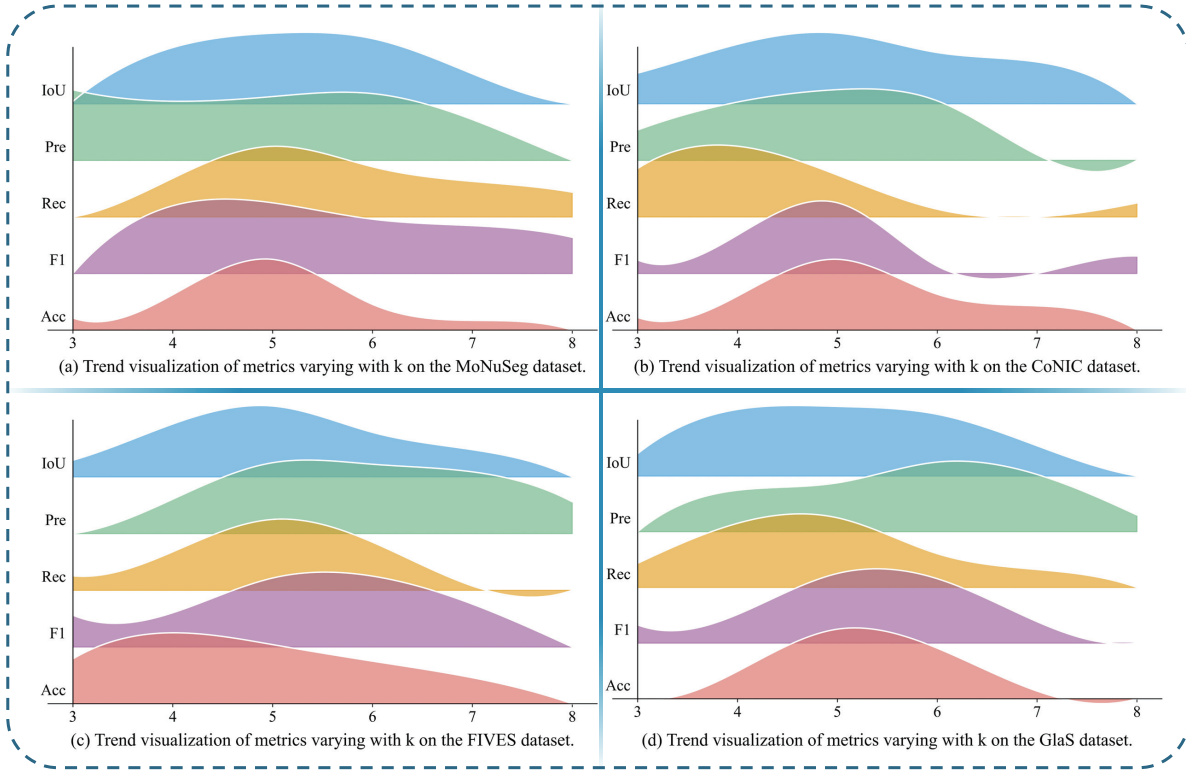


Fig. S11: Visualization of the trends across all evaluation metrics for four different datasets as k changes. All values are normalized, and the curves are smoothed using an interpolation technique for clarity.

relatively minor, indicating that low-resolution data exhibits reduced sensitivity to variations in α , potentially due to its lower reliance on fine-grained parameter adjustments for feature extraction. These findings highlight that while an initial increase in the depth of progressive reasoning enhances the feature extraction and representation of PCNet, excessively deep architectures may negatively impact performance.

The results for the challenge datasets are presented in Table S7. On the FIVES dataset, which focuses on retinal vessel segmentation and is characterized by high resolution and intricate capillary structures, a trend of initial increase followed by a subsequent decline is observed. All evaluation metrics reach their peak at $\alpha = 0.5$, with Acc achieving 98.82%, Rec attaining 83.71%, Pre peaking at 87.04%, and IoU achieving 75.87%. These findings, combined with prior observations, suggest that high-resolution data exhibit greater sensitivity to variations in α compared to low-resolution data. For the GlaS dataset, defined by its multi-scale target structures, performance values vary across cases, with optimal parameters observed at different α settings. However, the metrics generally peak around $\alpha = 0.5$.

The trends in evaluation metrics are visualized in Fig. S10. For multi-size and high-resolution datasets, the metrics consistently achieve their best values near $\alpha = 0.5$. Conversely, for the multi-scale dataset, the metrics exhibit limited variation, with peaks scattered across different α values. This suggests that for multi-scale datasets, neither the harder nor softer connection modes substantially affect the ability of network

to extract positive examples effectively.

2) *Experiments and Analyses for Depth k* : Moreover, a series of experiments are conducted to evaluate the impact of hyper-parameter Depth (k) on overall segmentation performance. A smaller k corresponds to fewer progressive layers within the network, which, on the one hand, reduces the excitation level of segmentation potential but also minimizes the risk of introducing excessive extraneous information. Conversely, increasing k amplifies the translation of segmentation potential into improved accuracy but comes at the expense of an increased risk of overloading the network with additional information.

The results for hyper-parameter depth (k) are summarized in Table S8, with the best results highlighted in bold. For cell nuclear segmentation on the MoNuSeg dataset, Acc and F1 demonstrate consistent improvement with increasing depth, reaching their highest values at $k = 5$ with Acc at 90.15% and F1 at 80.76%. This highlights the importance of moderate depth in enhancing nuclear segmentation by balancing structural information and feature extraction effectively. However, beyond $k = 5$, a decline in these metrics is observed, likely due to the introduction of excessive noisy information or optimization challenges as the network complexity increases. For the CoNIC dataset, the metrics exhibit a significant improvement trend up to $k = 5$, after which they either plateau or decline. F1, in particular, steadily improves, achieving its peak value of 79.10% at $k = 5$, while Rec reaches its maximum slightly earlier at $k = 4$. These observations suggest

TABLE S9: The results of experiments on different progressive modes

Datasets	Backbones	Progressive Mode	Evaluation Metrics (Mean $\pm$ Std)				
			Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
MoNuSeg	SerialUnet	Concat	89.11 $\pm$ 0.23	79.66 $\pm$ 0.61	82.01 $\pm$ 1.77	78.41 $\pm$ 0.79	66.32 $\pm$ 0.83
		Hard	88.89 $\pm$ 0.33	79.55 $\pm$ 0.52	83.24 $\pm$ 1.46	77.26 $\pm$ 1.38	66.21 $\pm$ 0.68
		Soft	88.68 $\pm$ 0.15	79.57 $\pm$ 0.36	84.68 $\pm$ 1.06	76.19 $\pm$ 0.87	66.20 $\pm$ 0.05
		Dual	89.50 $\pm$ 0.16	80.53 $\pm$ 0.24	83.56 $\pm$ 0.52	78.59 $\pm$ 0.49	67.52 $\pm$ 0.33
	PCNet	Concat	90.08 $\pm$ 0.22	81.89 $\pm$ 0.25	84.62 $\pm$ 0.82	79.06 $\pm$ 0.94	69.44 $\pm$ 0.36
		Hard	90.17 $\pm$ 0.15	82.18 $\pm$ 0.21	85.39 $\pm$ 0.73	79.11 $\pm$ 0.57	69.87 $\pm$ 0.29
		Soft	89.94 $\pm$ 0.11	81.83 $\pm$ 0.26	85.09 $\pm$ 0.84	78.80 $\pm$ 0.34	69.35 $\pm$ 0.37
		Dual	90.29 $\pm$ 0.35	81.89 $\pm$ 0.48	85.32 $\pm$ 0.34	79.23 $\pm$ 1.12	69.45 $\pm$ 0.68
CoNIC	SerialUnet	Concat	93.13 $\pm$ 0.04	76.12 $\pm$ 0.47	75.05 $\pm$ 1.32	78.22 $\pm$ 0.62	62.24 $\pm$ 0.52
		Hard	93.12 $\pm$ 0.04	76.36 $\pm$ 0.35	75.83 $\pm$ 1.96	77.90 $\pm$ 1.45	62.49 $\pm$ 0.45
		Soft	93.21 $\pm$ 0.02	76.34 $\pm$ 0.23	75.14 $\pm$ 1.03	78.58 $\pm$ 0.79	62.54 $\pm$ 0.28
		Dual	93.23 $\pm$ 0.01	76.56 $\pm$ 0.40	75.55 $\pm$ 0.15	78.51 $\pm$ 0.86	62.75 $\pm$ 0.46
	PCNet	Concat	93.44 $\pm$ 0.03	78.95 $\pm$ 0.29	76.07 $\pm$ 1.47	80.01 $\pm$ 0.93	65.87 $\pm$ 0.35
		Hard	93.45 $\pm$ 0.12	79.06 $\pm$ 0.06	76.88 $\pm$ 1.37	80.45 $\pm$ 1.15	66.02 $\pm$ 0.01
		Soft	93.42 $\pm$ 0.06	78.59 $\pm$ 0.72	77.42 $\pm$ 2.86	80.15 $\pm$ 1.74	65.47 $\pm$ 0.83
		Dual	93.93 $\pm$ 0.02	79.10 $\pm$ 0.29	77.95 $\pm$ 1.13	81.07 $\pm$ 0.74	66.16 $\pm$ 0.36
FIVES	SerialUnet	Concat	98.47 $\pm$ 0.03	80.90 $\pm$ 0.45	80.21 $\pm$ 1.63	85.16 $\pm$ 1.45	70.82 $\pm$ 0.60
		Hard	98.66 $\pm$ 0.01	80.82 $\pm$ 0.18	77.26 $\pm$ 0.58	88.73 $\pm$ 1.19	70.98 $\pm$ 0.01
		Soft	98.72 $\pm$ 0.01	81.62 $\pm$ 0.15	78.82 $\pm$ 0.47	87.70 $\pm$ 0.77	71.73 $\pm$ 0.14
		Dual	98.54 $\pm$ 0.01	81.32 $\pm$ 0.11	77.98 $\pm$ 0.37	88.58 $\pm$ 0.39	71.85 $\pm$ 0.13
	PCNet	Concat	98.57 $\pm$ 0.02	82.36 $\pm$ 0.12	81.48 $\pm$ 0.81	84.43 $\pm$ 0.76	73.17 $\pm$ 0.18
		Hard	98.56 $\pm$ 0.02	83.46 $\pm$ 0.19	82.41 $\pm$ 0.61	85.53 $\pm$ 1.54	74.30 $\pm$ 0.21
		Soft	98.67 $\pm$ 0.02	83.76 $\pm$ 0.62	82.91 $\pm$ 0.71	86.06 $\pm$ 1.37	74.33 $\pm$ 0.59
		Dual	98.82 $\pm$ 0.04	84.29 $\pm$ 0.70	83.71 $\pm$ 0.11	87.04 $\pm$ 1.83	75.87 $\pm$ 0.83
GlaS	SerialUnet	Concat	89.36 $\pm$ 0.01	88.81 $\pm$ 0.14	90.84 $\pm$ 0.49	87.87 $\pm$ 0.41	80.69 $\pm$ 0.22
		Hard	87.78 $\pm$ 0.04	87.27 $\pm$ 0.06	88.24 $\pm$ 0.37	87.62 $\pm$ 0.44	78.21 $\pm$ 0.09
		Soft	89.81 $\pm$ 0.03	89.25 $\pm$ 0.05	90.60 $\pm$ 0.47	88.93 $\pm$ 0.04	81.42 $\pm$ 0.07
		Dual	90.10 $\pm$ 0.02	89.52 $\pm$ 0.06	90.26 $\pm$ 0.26	88.77 $\pm$ 0.19	80.18 $\pm$ 0.07
	PCNet	Concat	91.58 $\pm$ 0.11	91.50 $\pm$ 0.16	92.79 $\pm$ 0.61	90.95 $\pm$ 0.53	84.96 $\pm$ 0.27
		Hard	91.30 $\pm$ 0.01	91.13 $\pm$ 0.05	92.77 $\pm$ 0.66	90.28 $\pm$ 0.65	84.39 $\pm$ 0.01
		Soft	91.43 $\pm$ 0.05	91.35 $\pm$ 0.08	92.61 $\pm$ 0.52	90.81 $\pm$ 0.62	84.74 $\pm$ 0.15
		Dual	92.00 $\pm$ 0.18	91.77 $\pm$ 0.21	92.94 $\pm$ 0.67	91.35 $\pm$ 0.65	85.52 $\pm$ 0.30

that while greater depth enhances the ability of network to capture detailed nuclear features, excessive depth may lead to redundant information, thereby compromising the feature representation ability of model.

For the FIVES dataset, which focuses on retinal vessel segmentation, both Pre and IoU show significant improvement with deeper networks, peaking at $k = 5$ with Pre at 87.04% and IoU at 75.87%. This suggests that moderate depth optimally balances the ability of network to delineate vessel boundaries and connections. However, beyond $k = 5$, performance begins to decline, indicating that excessive complexity may impair segmentation accuracy. In the GlaS dataset, a stable trend is observed across all metrics as depth increases, with the best performance consistently occurring at $k = 5$. For instance, Acc reaches 92.00% and IoU achieves 85.52%, demonstrating the effective segmentation capability of model for identifying the gland structures at this depth. Further increases in k leading to the declines of evaluation metrics, highlighting the importance of the trade-off between introduction of additional noise and transformation of segmentation potential.

The trends of these evaluation metrics for hyper-parameter k are visualized in Fig. S11. Most metrics initially increased before declining with further increases in k , suggesting that a moderate increase in depth enhances feature extraction

and utilization. However, beyond a certain threshold, deeper networks may induce overfitting or exacerbate the vanishing gradient problem, hindering performance.

In summary, setting α near 0.5 and configuring network depth to $k = 5$ achieves an optimal balance between positive and negative detection in medical image segmentation. This configuration enhances the capacity of model to capture fine-grained structures while maintaining balanced sensitivity and specificity, which are critical for achieving accurate and comprehensive segmentation.

C. Structure Exploration

In this section, we conduct a serial experiments to prove the superior performance of the newly proposed dual soft-hard constraint strategy. All the experiments are conducted on the five datasets and the best results are marked in bold. To further explore the influence of different progressive strategies on segmentation performance, we employ SerialUnet [13] and the proposed PCNet as backbone architectures. Both models utilize a serial progressive design and are evaluated combining with several different strategies, including traditional concatenation mode, single soft, hard connection, and the proposed novel dual soft-hard constraint approach.

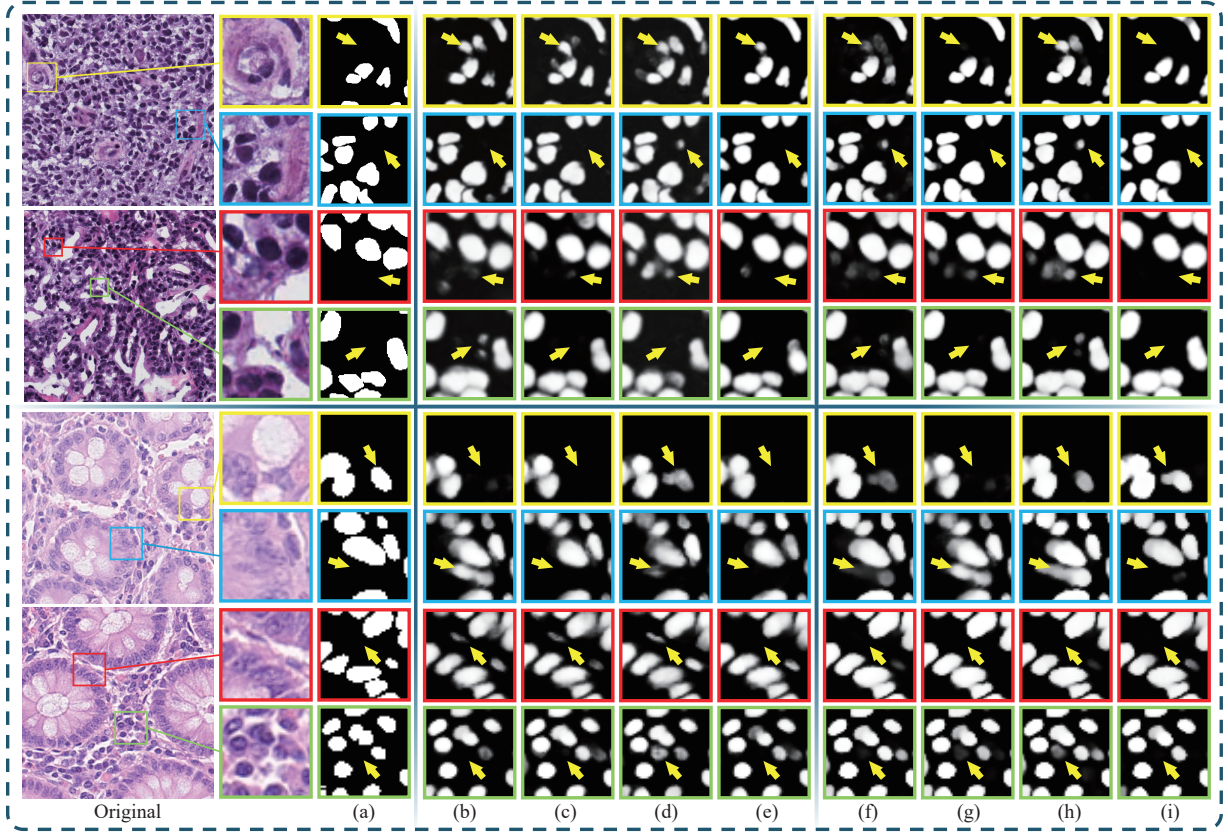


Fig. S12: Visualization results of structure exploration experiments. (a) Ground truth of the original images. (b) and (f) Results obtained using SerialUnet and PCNet with traditional concatenation connections. (c) and (g) Results obtained using single hard connections in the two baseline networks. (d) and (h) Results obtained using single soft connections. (e) and (i) Results obtained using the proposed dual soft-hard constraint strategy.

As shown in Table S9, which summarizes all experimental results with the best metrics highlighted in bold, replacing the traditional concatenation operation with single hard or soft connections in SerialUnet results in a slight decline in some evaluation metrics. In contrast, the proposed PCNet mitigates this performance drop and even demonstrates marginal improvements with soft connections, highlighting its superior feature extraction and fusion capabilities compared to traditional serial architectures. For cell nuclear, retinal vessel, and gland segmentation tasks, such as those on the MoNuSeg, CoNIC, FIVES, and GlaS datasets, neither single hard nor soft connections consistently outperform the traditional concatenation operation in both SerialUnet and PCNet. This limitation is attributed to the varying scales of segmentation target in different cells and glands, where the imbalance between exploration plasticity and stability inherent in these connection modes constrains overall performance. Furthermore, for SerialUnet, inconsistent experimental results are observed in the FIVES and GlaS datasets. This inconsistency may be due to insensitivity of SerialUnet to high-resolution and multi-scale datasets, further emphasizing the need for more advanced architectures like PCNet to address these challenges effectively.

Specifically, Fig. S12 visualizes the structure exploration experiments conducted using SerialUnet and PCNet with various connection methods, including traditional concatenation,

single hard, soft connections, and the proposed dual soft-hard constraint strategy. As shown in columns (d) and (h), compared to columns (c) and (g), the results demonstrate that soft connections excel in capturing subtle variations and accurately identifying target pixels for both SerialUnet and the proposed PCNet. However, relative to the ground truth, the single soft connection is prone to over-segmentation, leading to erroneous results. Conversely, hard connections effectively preserve structural integrity and delineate clear segmentation boundaries, but they are more prone to under-segmentation, missing finer details of the target structures. The proposed dual soft-hard constraint strategy successfully addresses these limitations by integrating the strengths of both approaches. This strategy achieves a balanced performance, mitigating over-segmentation while maintaining precise boundaries. This duality highlights the potential of our method to improve segmentation performance, particularly in challenging datasets.

REFERENCES

- [1] Q. Hu, M. D. Abràmoff, and M. K. Garvin, "Automated separation of binary overlapping trees in low-contrast color retinal images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, part 2, 2013, pp. 436–443.
- [2] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 39, pp. 1380–1391, 2019.

- [3] S. Graham, Q. D. Vu, M. Jahanifar, M. Weigert, U. Schmidt, W. Zhang, J. Zhang, S. Yang, J. Xiang, X. Wang *et al.*, “CoNIC challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting,” *Med. Image Anal.*, vol. 92, p. 103047, 2024.
- [4] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, “FIVES: A fundus image dataset for artificial intelligence based vessel segmentation,” *Sci. Data*, vol. 9, pp. 475–475, 2022.
- [5] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez *et al.*, “Gland segmentation in colon histology images: The glas challenge contest,” *Med. Imag. Anal.*, vol. 35, pp. 489–502, 2017.
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), part 3*, 2015, pp. 234–241.
- [7] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Trans. Med. Imag.*, vol. 39, pp. 1856–1867, 2019.
- [8] X. Xie, W. Zhang, X. Pan, L. Xie, F. Shao, W. Zhao, and J. An, “CANet: Context aware network with dual-stream pyramid for medical image segmentation,” *Biomed. Signal Process. Control*, vol. 81, p. 104437, 2023.
- [9] X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu, “MISSFormer: An effective transformer for 2d medical image segmentation,” *IEEE Trans. Med. Imag.*, vol. 42, pp. 1484–1494, 2022.
- [10] H. Du, J. Wang, M. Liu, Y. Wang, and E. Meijering, “SwinPA-Net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 35, pp. 5355–5366, 2022.
- [11] L. Lan, P. Cai, L. Jiang, X. Liu, Y. Li, and Y. Zhang, “BRAU-Net++: U-Shaped hybrid CNN-Transformer network for medical image segmentation,” *arXiv preprint arXiv:2401.00722*, 2024.
- [12] M. M. Rahman, M. Munir, and R. Marculescu, “EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation,” in *Pro. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024, pp. 11 769–11 779.
- [13] R. A. Karlsson and S. H. Hardarson, “Artery vein classification in fundus images using serially connected U-Nets,” *Comput. Meth. Prog. Biomed.*, vol. 216, p. 106650, 2022.
- [14] Z. Han, M. Jian, and G.-G. Wang, “ConvUNeXt: An efficient convolution neural network for medical image segmentation,” *Knowl.-Based Syst.*, vol. 253, p. 109512, 2022.
- [15] W. Zhou, W. Bai, J. Ji, Y. Yi, N. Zhang, and W. Cui, “Dual-path multi-scale context dense aggregation network for retinal vessel segmentation,” *Comput. Biol. Med.*, vol. 164, p. 107269, 2023.
- [16] Z. Li, Y. Zheng, D. Shan, S. Yang, Q. Li, B. Wang, Y. Zhang, Q. Hong, and D. Shen, “ScribFormer: Transformer makes cnn work better for scribble-based medical image segmentation,” *IEEE Trans. Med. Imag.*, pp. 2254–2265, 2024.