Potential-guided Connected Network for Tiny Structure Segmentation in Medical Images

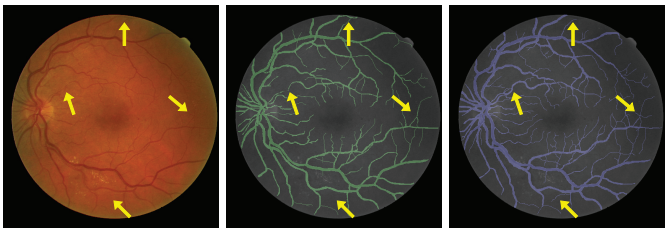
Chouyu Chen, Yaotong Song, Junyan Yi, Lijun Guo, Zhenyu Lei, *Member, IEEE* and Shangce Gao, *Senior Member, IEEE*

***Abstract*—Medical images provide essential information for diagnosing and monitoring various diseases and systemic disorders. With advancements in deep learning and neural networks, numerous methods have been proposed to achieve high-level medical image segmentation results. However, the variability of tiny structures and their high similarity to the background often lead to mis-segmentation in existing methods. To mitigate these challenges, we propose a potential-guided connected network (PCNet) that integrates an innovative dual soft-hard constraint strategy, combining two different progressive supervisions. This strategy modulates the ability of network to differentiate between well-defined and ambiguous structures through a hyper-parameter, thereby enhancing its capability to detect tiny structures. Furthermore, PCNet is composed of two key modules, including the intermediate generation (IG) module and the progressive inference (PI) module. The IG module produces a range of outputs with varying segmentation potentials using a novel serial architecture, which serves as the foundational input for progressive reasoning in the PI module. The PI module, leveraging the outputs of the IG module, is designed to progressively extract comprehensive contextual information, ultimately producing refined segmentation results. PCNet is evaluated on several publicly available datasets, including DRIVE, MoNuSeg, CoNIC, FIVES, and GlaS, achieving accuracy of 96.92%, 90.29%, 93.93%, 98.82%, and 92.00%, respectively. Extensive experiments demonstrate that our model outperforms the current state-of-the-art methods for tiny structure segmentation in medical image.**

***Index Terms*—medical image, segmentation, connected network, potential-guided, progressive inference**

I. Introduction

MEDICAL images are essential in clinical diagnosis, supporting applications like retinal disease detection [1], cell pathology identification [2], and other medical employs [3]–[5]. Detecting subtle changes in tiny structures, such as microhemangiomas, hemorrhages, and neovascularizations in fundus images, could indicate a high risk for diabetic retinopathy (DR) [6]. However, segmentation in medical imaging remains challenging due to the intricate and variable nature of anatomical structures. Traditional manual segmentation, though accurate, is labor-intensive, demands substantial clinical expertise, and is susceptible to inter-observer variability, particularly in complex cases, as illustrated in Fig. 1. In contrast, deep learning-based computer vision methods provide a promising alternative by automating segmentation tasks. These models are capable of understanding intricate patterns and features from large datasets, making them particularly well-suited for detecting specific structures within the medical images across large and diverse populations [7]. This automation not only reduces time requirements but also minimizes dependency on specialized expertise, thereby improving the efficiency and consistency of medical image segmentation.



(a) Original (b) Professor 1 (c) Professor 2

Fig. 1. Comparison of different manual segmentation results. (a) displays the original image, while (b) and (c) show manual segmentation results by professor 1 and 2. Despite their expertise, professionals may produce varying segmentation results due to the anatomical complexity and variability, especially in cases involving intricate pathological conditions.

Over the past decade, significant research has focused on enhancing the efficiency and accuracy of medical image segmentation, spurring the development of numerous automatic methods. These approaches generally fall into two categories based on their connection strategies, including the single networks and serially connected networks. A pivotal advancement in single networks is the introduction of UNet in 2015 [8], whose U-shaped architecture effectively captures contextual features and enables precise pixel-level classification. Since then, most models have adopted this U-like structure, incorporating advanced attention mechanisms [9], [10] and dynamic convolution kernels [11], [12] to expand receptive fields in convolutional operations. Conversely, serially connected networks leverage multiple base networks to decompose the segmentation task into sub-tasks, with each sub-network concentrating on specific aspects and features [13]. By decomposing

This research was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grants JP25K21298 and JP25K03179, and Japan Science and Technology Agency (JST) Support for Pioneering Research Initiated by the Next Generation (SPRING) under Grant JPMJSP2145.

C. Chen, Y. Song, Z. Lei and S. Gao are with the Faculty of Engineering, University of Toyama, Toyama-shi, 930-8555, Japan (e-mail: chenchouyu1998@163.com; sytforwork@gmail.com; leizystu@outlook.com; gaosc@eng.u-toyama.ac.jp).

J. Yi is with the Department of Computer Science & Technology, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China (e-mail: yijunyan@bucea.edu.cn).

L. Guo is with the Faculty of Electrical Engineering and Computer Sciences, Ningbo University, Zhejiang Province, China (e-mail: guolijun@nbu.edu.cn).

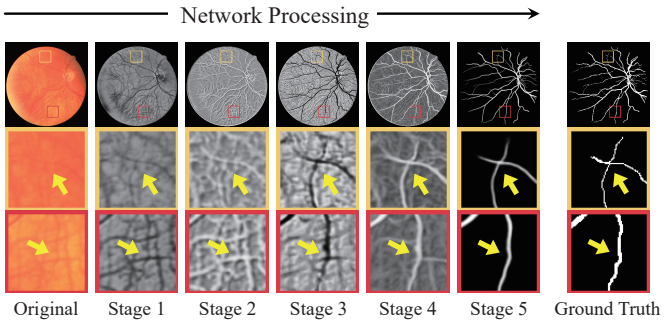


Fig. 2. The intermediate segmentation maps from a serially connected network. As the images are transmitted through the network, the intermediate segmentation maps progressively enhance the visibility of tiny, low-contrast structures within challenging regions, step by step.

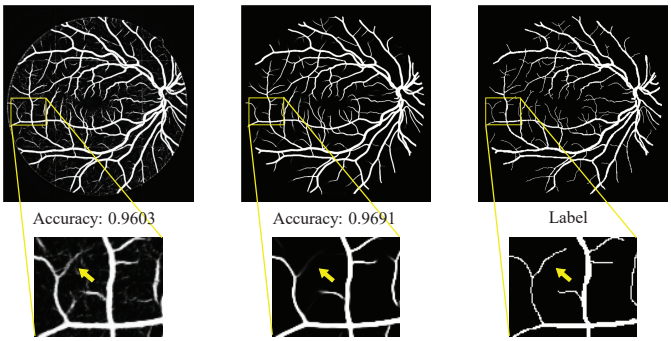


Fig. 3. Example samples from a segmentation process. As indicated by the arrow in the enlargement of yellow box, the lower accuracy segmentation result shows a greater potential for identifying tiny vessels against the background compared to the higher accuracy result.

complex segmentation processes into manageable steps or progressively refining the extracted information, these networks effectively enhance the accuracy and robustness of the final segmentation results. This hierarchical and task-specific focus enables serially connected networks to address challenges posed by intricate and heterogeneous medical image data.

Despite these advancements, notable challenges remain. The inherent limitations of convolution-based networks, particularly in terms of receptive field size and down-sampling, can impede the accurate detection of tiny structures within medical images [14]. Although the locality and translation invariance of convolutional networks enable effective modeling of local features [15], they often fail to capture long-range dependencies, resulting in segmentation inaccuracies, especially for tiny structures. The architecture of serially connected networks provides a promising approach to mitigate these challenges. By employing iterative feature extraction, these networks progressively enhance segmentation accuracy, particularly for tiny structures. As illustrated in Fig. 2, intermediate segmentation maps generated at various depths within a serially connected network reveal a progressive refinement of ambiguous regions, including those containing tiny structures. This observation underscores the potential of such architectures to iteratively refine segmentation potential into precise final outputs. However, conventional serially connected networks often emphasize final outputs at the expense of intermediate segmentation maps [16]. This focus can result in the loss of critical contextual information necessary for accurately delineating tiny structures. We argue that intermediate segmentation maps, despite their comparatively lower accuracy, encode vital structural details crucial for capturing fine-grained features. Preserving and strategically leveraging these intermediate outputs can significantly enhance segmentation outcomes. As highlighted in the yellow box of Fig. 3, lower-accuracy segmentation results often reveal critical structural information that should not be dismissed. For instance, in contrast to the middle sub-figure, the left sub-figure with less accurate segmentation better captures fine-grained details, emphasizing the importance of preserving and leveraging these intermediate outputs.

To alleviate these issues, we introduce the potential-guided connected network (PCNet), which enhances segmentation through two key components, including the intermediate generation (IG) module and the progressive inference (PI) module. The IG module, using a serially connected architecture, generates intermediate segmentation maps with varying potentials, maximizing information reuse. These outputs are then processed by the PI module, which incorporates progressive context transform (PCT) blocks to refine segmentation progressively. By leveraging a dual soft-hard constraint strategy, the PI module maximizes the utility of intermediate maps, improving accuracy for tiny and complex structures. The main contributions are summarized as follows:

1) We introduce PCNet, a novel framework for segmenting tiny structures in medical images, leveraging the IG and PI modules to refine segmentation through distinct intermediate maps.
2) We propose a novel IG module, designed to generate a series of intermediate segmentation maps with varying segmentation potential using a serially connected architecture.
3) We develop the PI module, which incorporates newly designed PCT blocks to progressively capture detailed contextual information across the varying segmentation potentials produced by the IG module, thereby enhancing global feature representations.
4) We design a dual soft-hard constraint strategy, which optimizes the balance between the plasticity and stability of feature fusion, effectively improving segmentation in complex regions by stimulating segmentation potentials from intermediate maps.

The structure of this paper is organized as follows: Section II introduces existing methods for medical image segmentation. Section III provides a detailed description of PCNet. The experimental settings and a comprehensive analysis of results are presented in Section IV. Finally, conclusions are given in Section V.

II. Related Work

Over past decades, a lot of approaches are proposed to alleviate the abovementioned challenges. According to the connection mode, they could be classified into two categories, including the single and serially connected models.

A. Single Models

Single models often leverage sophisticated blocks, such as attention mechanisms and feature fusion modules, to effectively extract high-level semantic features. For instance, Yu et al. [11] and Qi et al. [17] enhance the UNet framework by incorporating dilated convolutions and dynamic snakes, thereby improving segmentation accuracy, respectively. Meanwhile, Wang et al. [18] introduce deformable convolutions that dynamically adjust to the input data. Similarly, Chen et al. [19] employ dynamic convolutions to increase model complexity without expanding the network dimensions, utilizing input-dependent attention to aggregate multiple convolution kernels. To improve feature representation, many approaches focus on refining skip connections within U-like architectures. Zhou et al. [20] implemented nested, dense skip connections to bridge semantic gaps between the encoder and decoder pathways. Building on this, Lan et al. [21] incorporate the bi-level routing attention within skip connections, improving the modeling capacity for multi-scale features.

Moreover, Huang et al. [22] utilize full-scale skip connections combined with deep supervision. To mitigate information loss caused by down-sampling, M-Net proposed by Fu et al. [23] incorporates multi-scale pooling inputs, and Yi et al. [24] fuse wavelet-processed images with original inputs to enrich information capture. Other strategies blend traditional techniques with modern deep learning approaches, for instance, Shu et al. [25] integrated level set frameworks with deep learning for improved efficiency, while SwinPANet proposed by Du et al. [26] employs multi-scale feature fusion modules to accommodate varying lesion sizes. Additionally, Rahman et al. [27] propose EMCAD, which combines multi-scale depth-wise convolution with large-kernel gated attention to enhance spatial relationships while emphasizing salient regions.

Despite these advancements, the limited receptive fields and constrained multi-scale utilization of single models often impede performance, particularly with tiny, low-contrast structures. In contrast, serially connected models have emerged to better leverage contextual information across all stages of the network. It is exemplified by our proposed IG module, which integrates multiple U-like structures to facilitate detailed structural extraction.

B. Serially Connected Models

To enhance network performance, serially connected architectures frequently integrate features from multiple streams, thereby expanding the information flow pathways to capture a wider range of data characteristics. For instance, Valanarasu et al. [28] introduce KiUNet, which employs two branches to better capture fine details and accurately delineate edges of specific structures. To better reconstruct the tiny architectures in the original datas, Karlsson et al. [16] propose a series of serially connected networks for detailed artery-vein classification in fundus images, utilizing several shallow U-like networks to extract rich contextual information across all network stages.

Furthermore, Zhao et al. [29] present a triple UNet architecture for nuclei segmentation, where the first branch detects nuclear edges, the second identifies the main nuclear components, and the final branch integrates these features to produce high-quality segmentation results. Han et al. [30] propose a convolution-inspired architecture that uses convolution for down-sampling while incorporating a lightweight attention mechanism to suppress irrelevant features and filter out noise in low-level semantic information. Shu et al. [31] introduce a multi-stream encoder-based segmentation framework, which fuses attention-driven data to create meaningful connections between structural and semantic features. To support progressive learning, Eppenhof et al. [32] propose an incremental training strategy that starts with smaller network versions on lower-resolution images and deformation fields, gradually scaling up to address higher-resolution data. Zhou et al. [33] introduce a dual-path model that enhances generalization by incorporating progressive heatmaps to represent lesion density alongside original ultrasound images.

Moreover, ScribFormer proposed by Li et al. [34] offers an innovative three-branch architecture, combining a convolution branch, a Transformer branch, and an attention-guided class activation map branch. This synergy of convolution-captured local details and Transformer-derived global context addresses the constraints of traditional segmentation methods, leading to improved accuracy in complex scenarios.

While serially connected models mitigate challenges such as information loss and restricted receptive fields, naive concatenation of features from various branches may not fully harness the underlying information. In this study, we propose a PI module, a novel iterative framework that sequentially accumulates feature maps generated by prior serially connected modules to iteratively refine segmentation outcomes. Additionally, our newly developed dual soft-hard constraint strategy dynamically adjusts the contribution of intermediate feature maps, leading to enhanced adaptability and segmentation performance.

III. Method

As illustrated in Fig. 4, our proposed potential-guided connected network (PCNet) is designed as a two-module framework, aiming to achieve high-precision segmentation of tiny structures in medical images. The PCNet involves two key components, including the IG and PI module. The IG module generates intermediate segmentation maps, which reflect varying levels of potential and accuracy. This is achieved through a serially connected architecture, where each stage contributes to improving segmentation accuracy and consuming the segmentation potential progressively, forming a state-changing process of mutual gain and loss.

Subsequently, the PI module processes these intermediate maps through the newly designed PCT block. This block leverages the segmentation potentials to further contextualize and refine the outputs at each stage. In addition, a dual soft-hard constraint strategy is applied to optimize this process, in which the soft constraints guide the network in accurately estimating potential regions. In contrast, hard constraints ensure precise corrections for smaller structures. This complementary approach enables the network to refine its segmentation iteratively, resulting in highly accurate outputs, with a marked

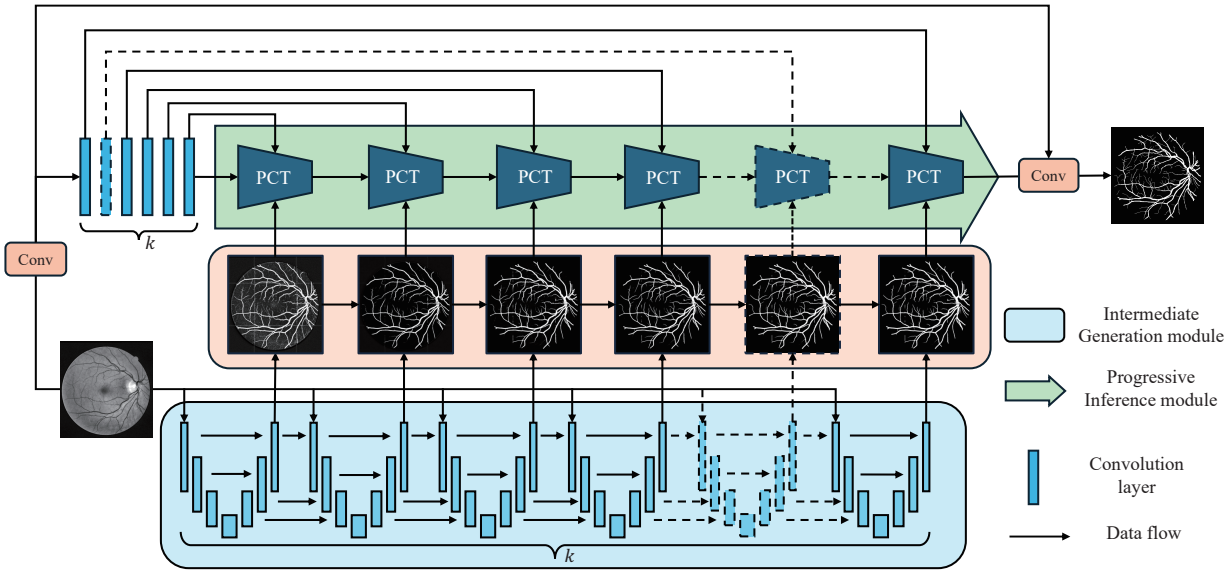


Fig. 4. The structure of PCNet, which consists of two main modules, progressively inferring the final results.

improvement in segmenting the tiny and complex structures in medical images.

A. Quantization of Segmentation Potential

According to the description and analyze at the end of Section I, we propose a quantitative method of segmentation potential. Effective concept-based segmentation relies on the intermediate maps capturing both meaningful semantic content and structural detail. However, the direct evaluation of such intermediate representations remains inherently challenging due to their abstract and often ambiguous nature. To address this limitation, we propose the segmentation potential (SP), a composite metric designed to quantitatively and interpretably assess the quality of concept maps.

Formally, the segmentation potential for the i-th intermediate map, SP_i, is defined as:

$$SP_i = CL_i + EN_i + SR_i^{GT}, \tag{1}$$

where the constituent terms represent distinct, complementary facets of segmentation quality:

- **Clarity** (CL_i): quantifies the sharpness and focus of the segmentation map;
- **Entropy** (EN_i): measures the diversity and richness of visual information;
- **Structural Richness** (SR_i^{GT}): evaluates the alignment of edge structures with ground-truth semantic boundaries.

This metric is motivated by three core principles. First, an effective concept map should exhibit pronounced clarity, characterized by crisp edges and localized features. We operationalize this via the variance of the Laplacian operator applied to the intermediate map,

$$CL_i = \mathrm{Var}(\nabla^2 I_i), \tag{2}$$

where $\mathrm{Var}(\cdot)$ denotes the variance and ∇^2 the Laplacian operator applied to the intermediate representation I_i.

Second, a high-quality representation must encode rich, informative content, captured here through Shannon entropy [35] over the pixel intensity distribution,

$$EN_i = -\sum_{j=0}^{255} p_j \log_2 p_j, \tag{3}$$

where p_j equals to $\frac{\{(x,y)|I_i(x,y)=j\}}{HW}$, representing the normalized histogram value at gray level j, computed as the proportion of pixels with intensity j in the binary segmentation map I_i.

Third, and critically, the metric emphasizes semantic alignment by assessing structural richness along object boundaries defined by ground-truth annotations. This is quantified as

$$SR_i^{GT} = \frac{1}{|\mathcal{G}|} \sum_{(x,y)\in\mathcal{G}} |Sobel(I_i)_{(x,y)}|, \tag{4}$$

where $\mathcal{G}$ denotes the set of pixel coordinates within a dilated ground-truth boundary region, and $Sobel(\cdot)$ is the Sobel edge detector applied to the intermediate representation.

By integrating these complementary components, the segmentation potential offers a principled and holistic framework for the quantitative evaluation of intermediate concept maps, reflecting their semantic and structural fidelity critical to the success of downstream segmentation tasks.

B. Intermediate Generation Module

To generate a serially intermediate segmented pictures with varying accuracy, we design a serially connected module, named the IG module, consisting of several basic networks with the similar structure, supervised by multiple loss functions with different weight. The structure of IG module is shown as the Fig. 5.

Each basic network exhibits a symmetric architecture reminiscent of UNet [8], comprising contracting and expanding

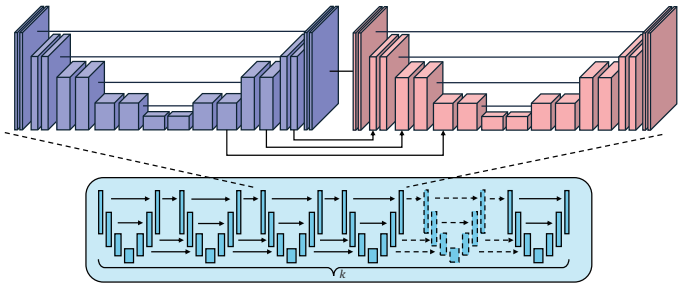


Fig. 5. The structure of IG module.

Algorithm 1 Interconnection Strategy

Input: The input data x, the number of basic networks k.

$c_i \Leftarrow$ The output of i-th basic network.
$\hat{x}_m^i \Leftarrow$ The output of m-th decoder stage in i-th basic network.
for $i = 0$ to k **do**
  if $i = 0$ **then**
    $c_1, \hat{x}_1^1, \hat{x}_2^1, \hat{x}_3^1 \leftarrow Basic_0(x)$
  else if $0 < i < k$ **then**
    $c_i, \hat{x}_1^i, \hat{x}_2^i, \hat{x}_3^i \leftarrow Basic_i(c_{i-1}, \hat{x}_1^{i-1}, \hat{x}_2^{i-1}, \hat{x}_3^{i-1})$
  else
    $c_i \leftarrow Basic_i(c_{i-1}, \hat{x}_1^{i-1}, \hat{x}_2^{i-1}, \hat{x}_3^{i-1})$
  end if
end for

paths. These paths are organized into four distinct stages, ensuring a clear hierarchical progression through the network. At each stage, the input data x is processed by two convolution layers, both employing a 3×3 kernel to capture spatial features at multiple levels, contributing to the capability of models to learn intricate patterns in the input data. Furthermore, each convolution layer within the network is followed by a batch normalization operation, ensuring that the distribution of feature values remains stable throughout the layers. Lastly, a LeakyReLU [36] activation function is applied after each normalization step, enhancing the nonlinear ability of the network.

The contracting path at each stage employs down-sampling operations, typically using max-pooling, to reduce the spatial dimensions of the feature maps, thereby emphasizing abstract, high-level features. Conversely, the expanding path utilizes up-sampling operations to progressively restore the spatial resolution, enabling the network to generate detailed output.

To promote interaction between base networks within the IG module, a novel interconnection strategy is employed. Specifically, the output of decoder stage from each basic network is fed into the corresponding encoder stage of the subsequent network, with the exception of the bottleneck stage. The process of the IG module is formalized in Algorithm 1.

This interconnection strategy facilitates the generation and transfer of hierarchical feature representations across successive base networks, enabling progressive refinement of feature extraction and integration. To mitigate information loss inherent in U-shaped architectures due to down-sampling and to retain critical details from the original medical images,

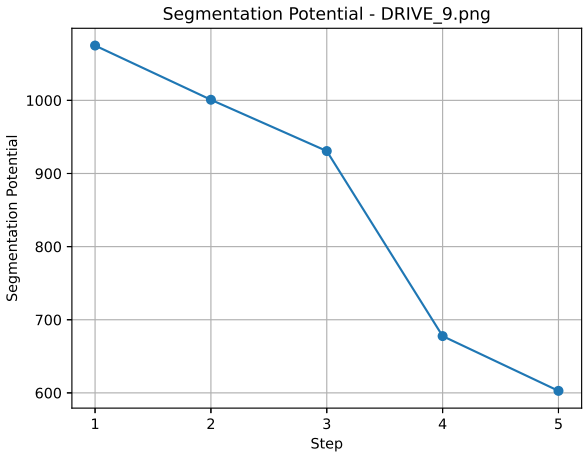


Fig. 6. Trend of the segmentation potential metric for a DRIVE sample image over different stages of IG module output.

the original image is concatenated with the output of each basic network (c_i) and passed into the next network. This methodology effectively preserves essential image details, minimizes downstream information degradation, and enhances the precision of intermediate segmentation outputs. Consequently, each base network iteratively produces higher-quality segmentation maps, progressively improving segmentation potential and overall performance. To validate the effectiveness of the proposed interconnection architecture and IG module, the segmentation potential metric described earlier is applied to evaluate the quality of intermediate segmentation maps generated by the IG module on an image from the DRIVE dataset. As illustrated in Fig. 6, the segmentation potential values exhibit a consistent decline over successive stages. This trend supports the previous analysis and indicates a gradual reduction of fine-grained, potentially noisy structures, in favor of more coherent and semantically meaningful representations. The downward trajectory reflects a deliberate trade-off, where less relevant microstructures are progressively filtered out to improve the overall clarity and consistency of the segmentation. Additional qualitative results, including representative visualizations of the IG module outputs, are provided in the Supplementary File to further substantiate these observations.

C. Progressive Inference Module

To fully leverage the rich contextual information across varying potential intermediate segmentation maps for inferring the final results, we propose a novel PI module, guided by the varying segmentation potential, constructed by the proposed dual soft-hard constraint strategy, as illustrated in Fig. 7. This module comprises several PCT blocks, where images with different segmentation potential are processed to fine-tune the overall feature representation.

Before entering the PI module, the original data is processed by a feature extractor, which consists of several convolution layers, to produce a series of middle-cross maps $(y_1, y_2, \ldots y_i, \ldots y_k)$, where k represents the total number of PCT blocks in the PI module. Additionally, a high-level

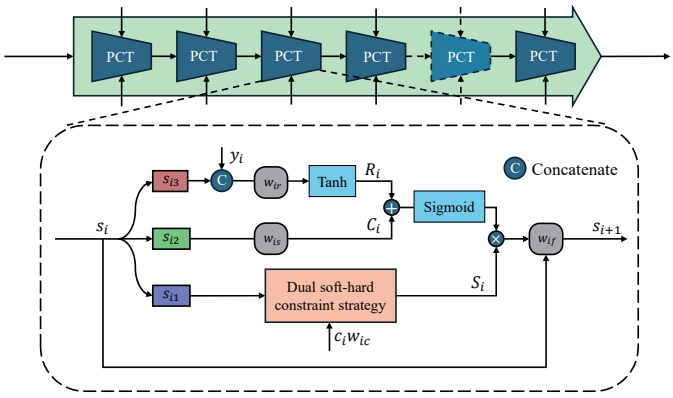


Fig. 7. The structure of PI module which consists of several PCT blocks.

semantic information map (s_1) is produced and serves as input to the first PCT block. In subsequent operations, parameter s_i represents the high-level semantic map, which is transferred into the ith PCT block.

Within the PCT block, the high-level semantic information map (s_i) is evenly divided into three components, denoted as s_{i1}, s_{i2}, and s_{i3}, each serving a distinct purpose in segmentation optimization. Leveraging the proposed dual soft-hard constraint strategy, s_{i1} interacts with the intermediate map (c_i), generated by the corresponding base network in the IG module, to construct a dynamic search space. This search space facilitates the seamless fusion of the segmentation potential encapsulated in c_i with the high-level semantic information, thereby refining and adapting features. This process enhances the capacity of network to capture contextual relationships essential for accurate and detailed segmentation. Meanwhile, s_{i2} retains the original semantic information from s_i, acting as a stable content source to support the transformation of segmentation potential. Finally, s_{i3} integrates information from the skip connections linking the PCT block with the preceding feature extractor, serving as a reference point for segmentation potential transformation. This tripartite division ensures a balanced and effective utilization of semantic features for enhanced segmentation outcomes.

The intermediate maps (c_i), generated by the IG module, are first passed through a convolution layer (w_{ic}) to further refine the feature representation of s_{i1}. This process is guided by the proposed dual soft-hard constraint strategy, which enhances feature learning and representation, resulting in the generation of a vector termed the Search (S_i). This process is calculated as follows:

$$S_i = f_{dual}(s_{i1}, w_{ic}c_i), \tag{5}$$

where $f_{dual}(\cdot)$ refers to the proposed dual soft-hard constraint strategy.

Subsequently, s_{i2} is transferred to generate a vector through $w_i s$ termed Content (C_i), respectively. Similarly, s_{i3} is concatenated with the middle-cross maps (y_i), processing through a convolution layer ($w_i r$) and a $tanh$ activation function, to produce a special vector Reference (R_i). These process could be described as following.

$$C_i = s_{i2}w_{is}. \tag{6}$$

$$R_i = Tanh(Concatenate(s_{i3}, y_i)w_{ir}), \tag{7}$$

To facilitate the learning and interpretation of long-range contextual information, PCT block applies the Search (S_i) which is refined by the intermediate maps (c_i) from IG module via the proposed dual soft-hard constraint strategy, to guide the Reference (R_i) in identifying and inferring potential tiny structure information within the Content (C_i). These operations enhance the transformation of segmentation potential into improved overall accuracy. Specifically, the sum of C_i and R_i is multiplied with S_i, processed by a convolution layer (w_{if}) to generate the result of PCT block and transferred into the following blocks, which is calculated by:

$$s_{i+1} = w_{if}(Sigmoid(R_i + C_i) \cdot S_i). \tag{8}$$

The proposed PI module is designed to optimize the segmentation process by leveraging a dual soft-hard constraint strategy to harness a diverse range of segmentation potentials. This module incorporates a sequence of PCT blocks, iteratively refining feature representations to enhance the identification of intricate details and complex structures in medical images. At the core of this methodology is the dual soft-hard constraint strategy, which delicately balances two complementary learning pathways, enabling more precise feature refinement. This approach is crucial for accurately identifying tiny and challenging structures, as it influences the way intermediate feature maps are adjusted to refine high-level semantic information. By progressively refining features through successive PCT blocks, the PI module significantly improves the delineation of fine anatomical details, leading to superior segmentation outcomes.

D. Dual Soft-hard Constraint Strategy

In the PI module, the intermediate segmentation maps (c_i) are utilized to guide the process in which the subset of high-level semantic information component (s_{i1}) generates the Search (R_i) through a progressive connection operation. This operation aims to effectively enhance the expression of segmentation potentials in the intermediate maps. Conventional progressive modes typically include concatenation, single addition, and multiplication operations. As illustrated in Fig. 8, these progressive modes can be described as the following equations.

$$f_{concate}(s_{i1}, w_{ic}c_i) = Concatenate(s_{i1}, w_{ic}c_i) \tag{9}$$

$$f_{soft}(s_{i1}, w_{ic}c_i) = s_{i1} + w_{ic}c_i \tag{10}$$

$$f_{hard}(s_{i1}, w_{ic}c_i) = s_{i1} \times w_{ic}c_i \tag{11}$$

In this study, we believe that the single addition operation, which is referred to as a soft connection, preserves the relative sizes of pixel values between different structures. This feature facilitates the capability of networks to explore and segment the tiny structures with low contrast in ambiguous regions.

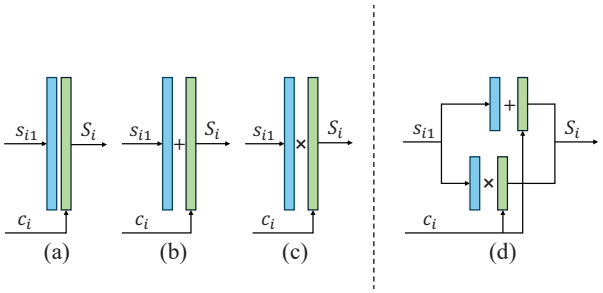


Fig. 8. Different connection mode. (a) Traditional concatenation operation. (b) The single soft connection. (c) The single hard connection. (d) The proposed dual soft-hard constraint strategy.

However, this operation is also subject to subtle changes that tend to blur the boundaries between distinct structures, especially in areas of low contrast or blurriness. Consequently, tiny or complex structures that may have initially been overlooked become more observable. Thus, this connection mode leads to a segmentation outcome that is characterized by increased fuzziness while exhibiting greater plasticity, particularly in the representation of tiny structures. Conversely, the multiplication operation, known as a hard connection, effectively emphasizes pixel differences between distinct structures. By amplifying these differences, the hard connection produces clearer and more stable segmentation results, especially in regions where precise demarcation of structures is crucial. However, this stability may lead to the oversight of smaller structures, which can be inadvertently dismissed during the segmentation process.

To leverage the advancements of both single hard and soft connections, we propose a dual soft-hard constraint strategy, as shown in the last part of Fig. 8, which not only enhances the contrast between background and target pixels but also significantly improves the distinction between tiny, intricate structures and the surrounding tissue during the inference process. In this strategy, c_i is first combined with a subset of the high-level semantic information (s_{i1}) using assigned weights (α) in one hand. Meanwhile, it multiplied with s_{i1} via $(1-\alpha)$ to produce R_i. This process can be described as:

$$f_{dual}(s_{i1}, w_{ic}c_i) = \alpha s_{i1} \cdot w_{ic}c_i + (1-\alpha)(s_{i1} + w_{ic}c_i). \quad (12)$$

Through this way, the intermediate segmented maps, which capture the varying segmentation potential, can subtly enhance the expression of the original information, considering more possible structural pixels to infer potential tiny structures. This approach differs from the traditional method of concatenating the original information with intermediate feature maps, which typically combines both relevant and irrelevant information before passing it to the next layer of the network. By tuning the hyper-parameter α, the dual constraint strategy ensures more precise, reliable, and well-defined segmentation, particularly in cases involving tiny or low-contrast structures, which are often the most challenging to identify in medical images.

In Section IV-F, we investigate the impact of this approach on the final results compared to the traditional concatenation method, single hard, and soft connection.

E. Loss Function

In this paper, two traditional loss functions, binary cross-entropy (L^{bce}) loss and Dice similarity coefficient (L^{dice}) loss, are employed to supervise the whole network, i.e.,

$$L^{bce} = \sum -[y_i log(p_i) + (1-y_i) log(1-p_i)], \quad (13)$$

$$L^{dice} = 1 - \frac{2\sum y_i p_i + \epsilon}{\sum y_i + \sum p_i + \epsilon}, \quad (14)$$

where y_i and p_i refer to the truth label and perdition for each pixel. ϵ is a smoothing parameter, which is utilized to prevent potential computational inaccuracies, setting as 10^{-6} in this study.

The outputs of all base networks of IG module are employed to supervise this module using a progressive supervision, ensuring it can produce results with varying accurate outputs from different base network that have the potential to improve the final result. The criterion is calculated as:

$$L_{IG} = \sum_{i=0}^{k-1} \frac{1}{k-i} (L_i^{bce} + 0.1 \cdot L_i^{dice}), \quad (15)$$

where k denotes the total number of base networks comprising the IG module. L_i represents the loss associated with the ith network. By employing this composite loss, the outputs of each base network would obtain a higher segmentation accuracy with the increase of depth. In the experimental section, we explore the impact of network depths k on overall performance.

For PI module, the final progressive inference results are conducted to calculate the segmentation loss, i.e.,

$$L_{PI} = L^{bce} + 0.1 L^{dice}. \quad (16)$$

The total loss for the proposed network is defined as:

$$L = \frac{L_{IG} + L_{PI}}{2}. \quad (17)$$

F. The Advantages of the Proposed PCNet

To underscore the advantages of our proposed PCNet, we highlight its strengths across multiple dimensions, including its capabilities in image segmentation, innovative modules, and novel architecture.

1) Unlike traditional image segmentation methods, which usually based on single structure or serially connected architectures, PCNet leverages all intermediate feature maps to construct the final results. This approach mitigates the information loss commonly encountered in network processing, thereby enhancing segmentation performance.
2) PCNet is composed of two key modules: IG and PI module. The former IG module introduces a novel connection strategy to generate intermediate graphs with varying potentials, while the PI module employs a novel dual soft-hard constraint strategy to further refine segmentation outputs.
3) By decomposing the segmentation task into intermediate image generation and progressive feature transformation, PCNet adopts a stepwise architecture. This design

enables gradual excitation of segmentation potentials in intermediate maps, ensuring a more accurate and efficient segmentation process.

IV. Experiments

A. Evaluation Metrics

To comprehensively evaluate our method, we utilize several segmentation metrics, including Accuracy (Acc), Recall (Rec), Precision (Pre), F1-score (F1), Intersection over Union (IoU), Receiver Operating Characteristic (ROC) curve, and Precision-Recall (PR) curve. Acc evaluates the ability of a model to segment both target and background pixels. Rec measures sensitivity by identifying target pixels, while Pre assesses the correctness of the segmented target pixels. F1 combines Pre and Rec, providing an overall sensitivity measure. IoU quantifies the overlap between predictions and ground truth, capturing both false positives and negatives. The PR curve illustrates the trade-off between precision and recall, while the ROC curve assesses performance across thresholds, with the Area Under the Curve (AUC) representing overall classification performance. The calculation of these metrics are introduced in the Supplementary File.

B. Datasets and Experiments Details

To comprehensively evaluate the proposed method, we describe the experimental setup and the benchmark datasets used in this study. All experiments, including comparative and ablation studies, are conducted using standardized hardware and software configurations: an AMD Ryzen 7 5700X 8-Core Processor (3.40 GHz), an NVIDIA GeForce RTX 3090 GPU, and PyTorch 2.2.1 as the computational framework. Each method is evaluated across five independent runs with identical random seeds, and we report the mean and standard deviation to ensure statistical reliability. The proposed method incorporates a pre-trained IG module trained for 10 epochs, followed by the full training of PCNet. To evaluate the segmentation performance of PCNet in medical imaging, we utilize a range of widely-used public datasets, including DRIVE [37], MoNuSeg [38], CoNIC [39], FIVES [40], and GlaS [41]. Detailed descriptions of these datasets are provided in the Supplementary File. Moreover, to assess the statistical significance of the observed performance differences between our method and the baselines, we employ the Wilcoxon signed-rank test on F1 scores across test samples. This non-parametric test is suitable for paired comparisons and does not assume normality in the distribution of performance differences.

C. Comparison Experiments

In this section, we evaluate the proposed PCNet against a diverse range of medical image segmentation models across multiple datasets. The detailed methods introduction and categorization are available in the Supplementary File. We first present a comparative analysis of the number of parameters and floating-point operations (FLOPs) across all evaluated methods to underscore the computational efficiency and compactness of the proposed framework. As illustrated in Fig. 9, although the FLOPs of the proposed model are comparable to those of existing approaches, the substantial reduction in parameter count highlights its lightweight architecture and suitability for resource-constrained deployment scenarios.

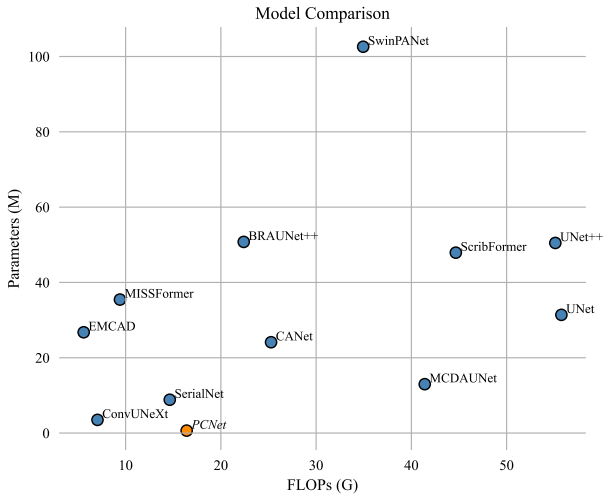


Fig. 9. Comparison of different segmentation models in terms of computational cost and model size. PCNet demonstrates a significantly lower parameter count while maintaining competitive FLOPs, indicating its efficiency.

The comparison experiments are structured into two components, which the first focuses on traditional datasets, including DRIVE, MoNuSeg, and CoNIC, while the second focus on challenging datasets, including the FIVES and GlaS.

For traditional datasets, the results are all summarized in Table I, the best evaluation metrics are highlighted in bold, while the second-best are underlined. Notably, for DRIVE dataset, PCNet surpasses almost all of the other methods across every evaluation metric, underscoring the superior effectiveness of the proposed approach. Specifically, the Acc and Rec metrics, which assesses the overall precision of the segmentation task, shows an improvement of 0.11% and 0.75% compared to the second-best SerialUnet [16], indicating a greater sensitivity in detecting positive components of the segmentation task. Furthermore, the proposed PCNet shows an improvement of 0.97% in the F1 and 1.38% in the IoU metric over the second-best MCDAUNet [33], demonstrating that our network effectively balances the detection of both target and background pixels. Although PCNet demonstrates slightly lower performance in the Pre metric, nearly matching the top-performing ScribFormer [34], it achieves the highest F1, underscoring a superior balance in accurately detecting both positive and negative pixels. These results highlight the robustness of PCNet in achieving precise and comprehensive segmentation outcomes. These results collectively indicate that the proposed network delivers outstanding performance across all aspects for the DRIVE dataset. In addition, the ROC and PR curves, shown in Fig. 10, provide further insights, with a larger AUC indicating enhanced differentiation between target and background pixels. Moreover, a balance point closer to

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

Datasets	Methods	Year	Evaluation Metrics (Mean ± Std)					p-value
			Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)	
DRIVE	UNet [8]	2015	96.69 ± 0.01	79.84 ± 0.29	75.36 ± 1.26	84.46 ± 0.99	66.50 ± 0.39	<0.0001
	UNet++ [20]	2018	96.56 ± 0.03	79.28 ± 0.40	75.59 ± 1.44	83.90 ± 1.01	65.72 ± 0.54	<0.0001
	SerialUnet [16]	2021	96.81 ± 0.01	79.93 ± 0.22	77.93 ± 1.14	84.70 ± 0.90	67.01 ± 0.31	<0.0001
	MISSFormer [42]	2022	96.51 ± 0.01	78.72 ± 0.27	74.09 ± 1.37	84.54 ± 1.17	64.94 ± 0.36	<0.0001
	ConvUNeXt [30]	2022	96.57 ± 0.02	79.12 ± 0.19	74.68 ± 0.62	84.63 ± 0.47	65.50 ± 0.25	<0.0001
	CANet [10]	2023	96.68 ± 0.03	80.03 ± 0.36	76.35 ± 1.24	84.94 ± 0.87	66.74 ± 0.49	<0.0001
	MCDAUNet [33]	2023	96.78 ± 0.06	80.70 ± 0.63	77.24 ± 2.02	85.03 ± 1.22	67.69 ± 0.88	0.0016
	BRAUNet++ [21]	2024	96.52 ± 0.03	78.44 ± 0.51	72.85 ± 1.73	85.29 ± 1.25	64.57 ± 0.69	<0.0001
	EMCAD [27]	2024	96.30 ± 0.01	77.88 ± 0.18	74.84 ± 0.90	81.68 ± 0.70	63.80 ± 0.24	<0.0001
	SwinPANet [26]	2024	96.36 ± 0.04	79.47 ± 0.14	77.01 ± 0.19	79.73 ± 0.30	65.11 ± 0.18	<0.0001
	ScribFormer [34]	2024	96.34 ± 0.03	76.92 ± 0.36	70.12 ± 1.34	**85.69 ± 1.23**	62.55 ± 0.47	<0.0001
	PCNet (Ours)	-	**96.92 ± 0.01**	**81.67 ± 0.24**	**78.68 ± 1.25**	85.39 ± 0.96	**69.07 ± 0.34**	-
MoNuSeg	UNet [8]	2015	88.52 ± 0.15	78.91 ± 0.25	85.29 ± 0.92	74.03 ± 0.64	65.34 ± 0.31	<0.0005
	UNet++ [20]	2018	88.30 ± 0.13	78.38 ± 0.13	85.23 ± 0.55	73.52 ± 0.52	64.78 ± 0.15	<0.0005
	SerialUnet [16]	2021	89.59 ± 0.17	79.75 ± 0.26	81.00 ± 1.17	79.08 ± 0.98	66.52 ± 0.34	<0.0005
	MISSFormer [42]	2022	87.91 ± 0.02	77.70 ± 0.08	84.12 ± 0.38	74.07 ± 0.33	64.16 ± 0.13	<0.0005
	ConvUNeXt [30]	2022	88.63 ± 0.16	78.38 ± 0.18	82.74 ± 0.73	75.40 ± 0.63	64.67 ± 0.21	<0.0005
	CANet [10]	2023	89.82 ± 0.26	80.71 ± 0.10	84.67 ± 1.39	77.84 ± 1.50	67.80 ± 0.13	<0.0005
	MCDAUNet [33]	2023	89.61 ± 0.06	79.74 ± 0.32	81.71 ± 1.39	78.32 ± 0.73	66.49 ± 0.45	<0.0005
	BRAUNet++ [21]	2024	89.32 ± 0.46	79.25 ± 0.83	81.88 ± 2.30	78.09 ± 2.03	65.86 ± 1.07	<0.0005
	EMCAD [27]	2024	89.96 ± 0.06	81.01 ± 0.14	84.65 ± 0.33	78.07 ± 0.09	68.16 ± 0.20	0.0013
	SwinPANet [26]	2024	86.62 ± 0.30	79.37 ± 0.35	81.53 ± 0.08	78.87 ± 0.36	65.54 ± 0.39	<0.0005
	ScribFormer [34]	2024	89.42 ± 0.05	79.34 ± 0.74	72.89 ± 3.06	79.05 ± 1.87	65.87 ± 0.97	<0.0005
	PCNet (Ours)	-	**90.29 ± 0.35**	**81.89 ± 0.48**	**85.32 ± 0.34**	**79.23 ± 1.12**	**69.45 ± 0.68**	-
CoNIC	UNet [8]	2015	92.92 ± 0.05	75.28 ± 0.41	73.43 ± 1.86	78.32 ± 1.25	61.18 ± 0.52	<0.0001
	UNet++ [20]	2018	92.95 ± 0.04	75.45 ± 0.35	74.37 ± 1.03	77.63 ± 0.58	61.46 ± 0.39	<0.0001
	SerialUnet [16]	2021	93.01 ± 0.04	75.24 ± 0.41	73.05 ± 1.32	78.77 ± 0.66	61.18 ± 0.51	<0.0001
	MISSFormer [42]	2022	92.77 ± 0.06	75.27 ± 0.29	74.91 ± 1.64	77.05 ± 1.24	61.16 ± 0.37	<0.0001
	ConvUNeXt [30]	2022	91.89 ± 0.02	71.68 ± 0.25	70.75 ± 1.08	73.92 ± 0.75	56.84 ± 0.33	<0.0001
	CANet [10]	2023	93.37 ± 0.02	77.34 ± 0.30	77.04 ± 1.45	78.37 ± 0.89	63.79 ± 0.40	<0.0001
	MCDAUNet [33]	2023	93.83 ± 0.06	78.53 ± 0.33	**78.04 ± 0.96**	80.10 ± 0.75	65.70 ± 0.37	0.0014
	BRAUNet++ [21]	2024	92.08 ± 0.02	73.04 ± 0.29	73.73 ± 1.21	73.53 ± 0.63	58.37 ± 0.36	<0.0001
	EMCAD [27]	2024	93.22 ± 0.01	76.90 ± 0.10	75.72 ± 0.42	78.72 ± 0.27	63.07 ± 0.13	<0.0001
	SwinPANet [26]	2024	92.70 ± 0.01	75.76 ± 0.15	75.56 ± 0.66	77.52 ± 0.39	62.43 ± 0.19	<0.0001
	ScribFormer [34]	2024	92.96 ± 0.04	74.66 ± 0.26	70.54 ± 1.26	80.65 ± 1.09	60.34 ± 0.31	<0.0001
	PCNet (Ours)	-	**93.93 ± 0.02**	**79.10 ± 0.29**	77.95 ± 1.13	**81.07 ± 0.74**	**66.16 ± 0.36**	-

the upper right corner of the figure in the PR curve implies the effective capability of PCNet to identify most target pixels with high accuracy. In comparison with recent methods such as EMCAD, SwinPANet, and ScribFormer, PCNet achieves the highest scores across all metrics, emphasizing its robustness and adaptability to complex segmentation tasks. This performance advantage is largely due to the innovative integration of the dual soft-hard constraint strategy in PCNet, which collectively improves the ability of PCNet to capture fine details and segment intricate structures with high accuracy, robustness, and generalization capabilities. The similar results for nuclear segmentation in MoNuSeg and CoNIC datasets could be found in the Supplementary File. To further validate the statistical significance of the observed improvements, we conduct Wilcoxon signed-rank tests on the F1 scores between PCNet and each comparison method. The resulting p-values are all below 0.005 across datasets, indicating that the performance gains of PCNet are statistically significant. Moreover, we observed that failure cases are primarily concentrated around ambiguous boundaries and in regions with dense or complex structures, where precise localization becomes challenging. As shown in Fig. 12, the top row, which presents examples from the DRIVE dataset, illustrates misclassification of faint or vessel-like structures, which are often confused with the background, leading to false positives. This typically occurs in low-contrast areas or where vessels exhibit discontinuities. In contrast, the bottom row of the figure, which samples from the MuNoSeg dataset, highlights the difficulty of exists methods in segmenting nuclei with blurry contours or overlapping boundaries, particularly in densely packed regions. These conditions often result in false negatives, where nuclei are partially or entirely missed. While our method demonstrates relatively better performance compared to other comparison methods in these challenging scenarios, it still faces limitations in accurately resolving highly ambiguous structures. These observations demonstrate the potential directions for further improvement, such as adaptive attention to enhance discrimination in complex regions.

In addition, to evaluate the effectiveness of PCNet in handling challenging data with tiny structures, we conducted comparative experiments on the FIVES and GlaS datasets. The results are shown in Table II. For the FIVES dataset, comprising high-resolution fundus images that demand precise feature extraction, PCNet achieved top performance with an Acc of 98.82% and an F1 score of 84.29%, as well as the highest scores across metrics including Rec, Pre, and IoU.

TABLE II. The comparison experiment results on FIVES and GlaS datasets.

Datasets	Methods	Year	Evaluation Metrics (Mean ± Std)					p-value
			Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)	
FIVES	UNet [8]	2015	98.61 ± 0.02	82.05 ± 0.27	80.02 ± 0.63	86.36 ± 0.76	72.32 ± 0.20	<0.0001
	UNet++ [20]	2018	98.68 ± 0.01	82.13 ± 0.16	80.65 ± 0.45	84.87 ± 1.17	73.00 ± 0.16	<0.0001
	SerialUnet [16]	2021	98.47 ± 0.03	80.90 ± 0.45	80.21 ± 1.63	85.16 ± 1.45	70.82 ± 0.60	<0.0001
	MISSFormer [42]	2022	98.65 ± 0.01	83.17 ± 0.13	81.66 ± 0.74	86.29 ± 0.98	73.50 ± 0.18	<0.0001
	ConvUNeXt [30]	2022	98.44 ± 0.03	79.24 ± 0.38	74.67 ± 1.21	86.42 ± 1.83	68.79 ± 0.48	<0.0001
	CANet [10]	2023	98.71 ± 0.00	82.40 ± 0.19	80.70 ± 0.71	85.17 ± 1.98	73.45 ± 0.20	<0.0001
	MCDAUNet [33]	2023	98.76 ± 0.01	84.00 ± 0.20	83.02 ± 0.60	86.59 ± 0.67	74.98 ± 0.19	<0.0001
	BRAUNet++ [21]	2024	98.62 ± 0.00	82.32 ± 0.20	80.64 ± 0.23	86.86 ± 0.32	72.64 ± 0.13	<0.0001
	EMCAD [27]	2024	98.40 ± 0.00	82.02 ± 0.14	81.04 ± 0.27	83.46 ± 0.23	71.14 ± 0.11	<0.0001
	SwinPANet [26]	2024	97.58 ± 0.13	82.20 ± 0.25	81.12 ± 0.25	83.42 ± 0.56	68.59 ± 0.44	<0.0001
	ScribFormer [34]	2024	98.12 ± 0.03	79.63 ± 0.97	71.57 ± 2.23	86.96 ± 2.98	65.94 ± 0.98	<0.0001
	PCNet (Ours)	-	**98.82 ± 0.04**	**84.29 ± 0.70**	**83.71 ± 0.11**	**87.04 ± 1.83**	**75.87 ± 0.83**	-
GlaS	UNet [8]	2015	89.02 ± 0.06	88.24 ± 0.09	88.44 ± 0.59	89.05 ± 0.48	79.78 ± 0.14	<0.0001
	UNet++ [20]	2018	89.76 ± 0.08	89.14 ± 0.09	89.02 ± 0.57	90.35 ± 0.63	81.37 ± 0.15	<0.0001
	SerialUnet [16]	2021	88.33 ± 0.18	87.51 ± 0.22	88.00 ± 1.15	88.28 ± 0.82	78.84 ± 0.29	<0.0001
	MISSFormer [42]	2022	87.37 ± 0.20	86.92 ± 0.28	87.86 ± 0.45	87.73 ± 0.34	78.33 ± 0.36	<0.0001
	ConvUNeXt [30]	2022	87.42 ± 0.04	86.67 ± 0.10	87.25 ± 0.43	87.36 ± 0.37	77.43 ± 0.13	<0.0001
	CANet [10]	2023	90.43 ± 0.08	89.90 ± 0.12	89.93 ± 0.82	89.90 ± 0.70	82.59 ± 0.16	0.0002
	MCDAUNet [33]	2023	89.11 ± 0.10	88.36 ± 0.16	88.83 ± 0.61	89.45 ± 0.77	80.40 ± 0.22	<0.0001
	BRAUNet++ [21]	2024	89.23 ± 0.09	88.84 ± 0.12	90.34 ± 0.41	88.40 ± 0.26	80.73 ± 0.15	<0.0001
	EMCAD [27]	2024	90.60 ± 0.27	90.32 ± 0.31	90.90 ± 1.14	90.74 ± 1.12	83.10 ± 0.43	<0.0001
	SwinPANet [26]	2024	88.18 ± 0.15	87.39 ± 0.24	88.61 ± 0.70	87.58 ± 0.56	78.70 ± 0.30	<0.0001
	ScribFormer [34]	2024	89.29 ± 0.15	89.13 ± 0.15	90.51 ± 0.87	88.79 ± 0.85	81.21 ± 0.20	<0.0001
	PCNet (Ours)	-	**92.00 ± 0.18**	**91.77 ± 0.21**	**92.94 ± 0.67**	**91.35 ± 0.65**	**85.52 ± 0.30**	-

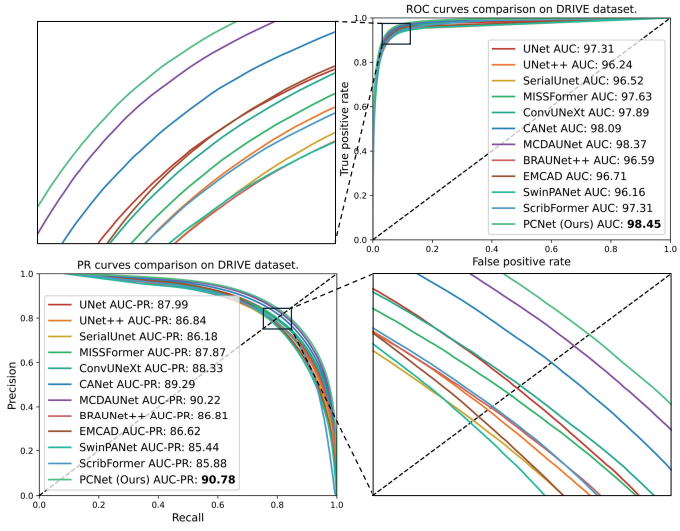


Fig. 10. The visualization of ROC and PR curves of comparison experiments on DRIVE dataset.

These results underscore the ability of PCNet to maintain high precision and capture subtle boundary details, essential for high-resolution image segmentation. In comparison, while earlier models like UNet and UNet++ also demonstrated robust accuracy which around 98.6%, they showed limitations in finer-grained metrics such as F1 and IoU, suggesting potential constraints when addressing high-resolution segmentation demands. The similar finding and total analysis for GlaS dataset can be found in the Supplementary File.

Furthermore, we present the visual comparisons of the segmentation results of DRIVE, MoNuseg, and CoNIC datasets which are obtained using various methods as illustrated in Fig. 11. It is evident that PCNet effectively captures the relevant structures in medical images, achieving a close resemblance to the original ground truth. For retinal vessel segmentation, highlighted in the yellow boxes in the first row, PCNet demonstrates superior performance by detecting more thin vessel pixels with low contrast compared to other methods. For nuclear segmentation, specific regions are presented in the second and third rows of Fig. 11. In these regions, highlighted with green and blue boxes, we apply red lines to indicate nuclear boundaries in the label for each comparison algorithm, facilitating a clearer comparison. Notably, PCNet demonstrates superior accuracy in delineating nuclear boundaries compared to other methods. As indicated by the yellow arrow within the green box in the second row, two ambiguous structures that are prone to mis-segmentation by other approaches are successfully identified by PCNet, which accurately avoids detecting these problematic structures. These observations indicate that PCNet enhances the ability to capture finer structural details and supports the notion that leveraging intermediate map potentials through a well-designed fusion approach can achieve more precise detail segmentation. Other visualizations are available at the Supplementary File.

D. Ablation Study

To evaluate the impact of the newly designed IG and PI modules on the overall performance of PCNet, we conduct a series of experiments across five datasets. In these experiments, the IG and PI modules are substituted with SerialUnet [16], a baseline network with an equivalent convolutional depth and a similar serially connected architecture. The results obtained on the DRIVE dataset are summarized in Table III.

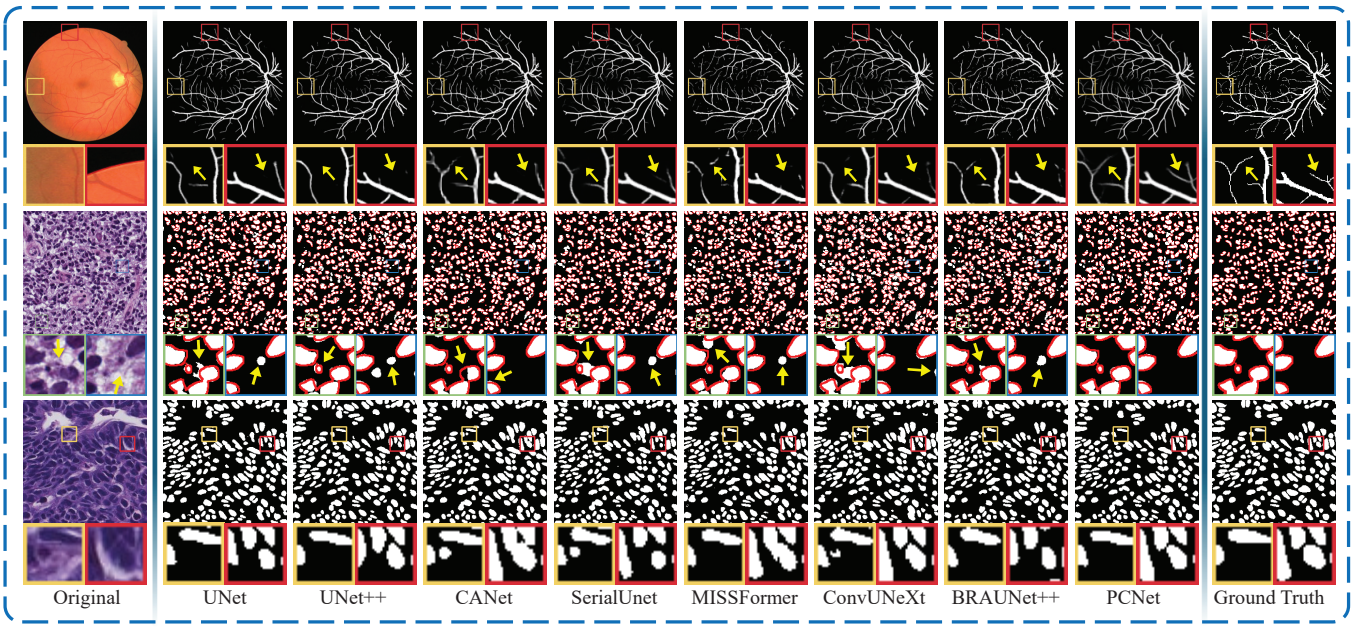


Fig. 11. Samples from the DRIVE, MoNuSeg, and CoNIC datasets. First row shows sample from the DRIVE dataset, second row shows image from the MoNuSeg dataset, while the last row shows sample from the CoNIC dataset.

TABLE III. The results of experiments on ablation study.

Modules		Evaluation Metrics (Mean±Std)				
IG	PI	Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
-	-	96.62±0.10	80.86±0.57	77.04±1.57	84.47±1.10	67.89±0.81
✓	-	96.71±0.03	81.56±0.14	78.54±0.62	85.03±0.52	68.87±0.20
-	✓	96.75±0.01	81.30±0.23	78.24±1.06	85.12±0.69	68.53±0.32
✓	✓	**96.92±0.01**	**81.67±0.24**	**78.68±1.25**	**85.39±0.96**	**69.07±0.34**

The comparison of Rows 1 and 2, as well as Rows 3 and 4, demonstrates that substituting SerialUnet with the IG module markedly improves evaluation metrics, with the balanced metric F1 score showing particularly substantial improvements. These results underscore the efficacy of the proposed interconnection strategy and weighted loss function in enhancing feature extraction and segmentation accuracy. The IG module enables the generation of intermediate segmentation maps with progressively higher fidelity, preserving critical details and optimizing hierarchical feature representations, which is an advantage particularly suited for addressing complex segmentation tasks. Similarly, the comparison of Rows 1 and 3, along with Rows 2 and 4, highlights significant performance improvements when the PI module is integrated into the baseline network. Metrics such as IoU reveal notable enhancements in segmentation accuracy, particularly in delineating precise boundaries, demonstrating the superior capability of PI module for accurate spatial representation. These findings validate the effectiveness of the PCT block in capturing both spatial and contextual information, surpassing the baseline architecture. Moreover, through a dual soft-hard constraint strategy, the PCT block further amplifies the ability of network to transfer the segmentation potential into the overall accuracy, delivering superior results for tiny and intricate structures.

TABLE IV. The experiment results of different α on DRIVE dataset.

α	Evaluation Metrics (Mean±Std)				
	Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
0.1	96.82±0.03	81.36±0.09	77.64±1.05	83.65±0.97	68.61±0.12
0.2	96.85±0.02	81.56±0.05	78.08±0.76	83.56±0.70	68.88±0.07
0.3	96.84±0.03	81.52±0.33	78.10±1.48	83.52±1.06	68.85±0.46
0.4	96.87±0.03	81.70±0.05	78.38±0.65	84.52±0.76	69.10±0.07
0.5	**96.92±0.01**	81.67±0.24	78.68±1.25	**85.39±0.96**	69.07±0.34
0.6	96.86±0.02	**81.78±0.04**	78.85±0.51	84.20±0.52	**69.21±0.05**
0.7	96.86±0.02	81.75±0.14	**78.87±0.76**	83.12±0.58	69.16±0.20
0.8	96.85±0.04	81.68±0.03	78.75±0.83	83.11±0.91	69.06±0.05
0.9	96.84±0.04	81.74±0.18	78.18±1.36	82.81±1.16	69.15±0.26

TABLE V. The experiment results of varying Depth (k) on DRIVE dataset.

k	Evaluation Metrics (Mean±Std)				
	Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
3	96.87±0.01	81.36±0.14	77.62±0.56	**85.98±0.40**	68.63±0.20
4	96.89±0.02	81.47±0.44	**78.73±1.70**	84.95±1.15	68.78±0.60
5	**96.92±0.01**	**81.67±0.24**	78.68±1.25	85.39±0.96	**69.07±0.34**
6	96.88±0.01	81.41±0.21	78.45±0.97	85.09±0.73	68.69±0.29
7	96.81±0.03	81.38±0.15	78.28±0.37	84.97±0.18	68.50±0.20
8	96.71±0.04	81.23±0.13	78.22±0.05	84.71±0.03	68.48±0.18

E. Hyper-parameter Analysis

For PCNet, two primary hyper-parameters play a crucial role in determining network performance, including 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. 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. All experiments are executed utilizing the all five datasets.

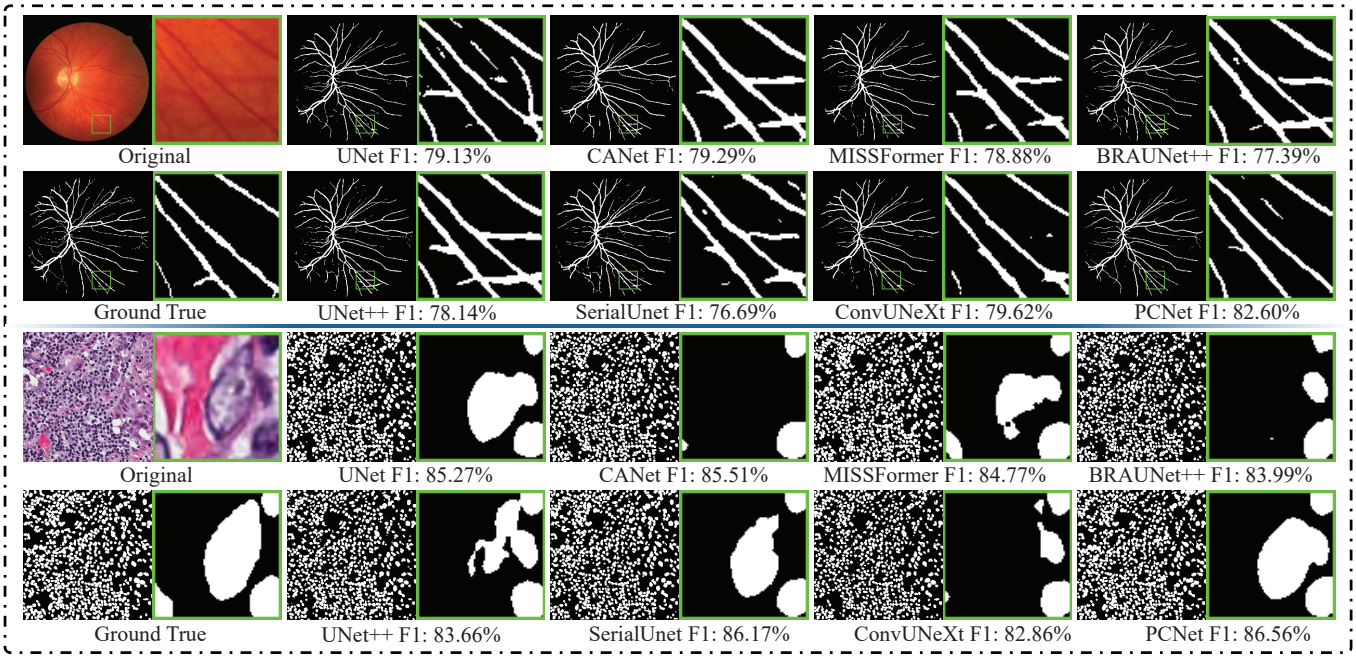


Fig. 12. Error maps of several compared methods across two different datasets.

To investigate the impact of α on medical image segmentation performance, we conduct experiments varying α from 0.1 to 0.9, where larger α values denote a higher ratio of hard connections. In fact, a higher α reflects a more conservative network exploration approach according to the argument in Section III. Results are summarized in Tables IV, with optimal values emphasized in bold.

In retinal vessel segmentation, we observe that as α increases, the evaluation metrics Acc and Pre improve initially, peaking at $\alpha = 0.5$ before declining. This trend suggests that the dual path soft and hard progressive mode optimally balances plasticity and stability, enhancing the segmentation potential in intermediate maps, particularly for delineating tiny structures. The subsequent decrease in performance likely reflects a loss of critical structural information or missed detection of tiny features under overly strict conditions. In contrast, Rec and IoU show limited fluctuation with changes in α, demonstrating that this parameter has minimal impact on positive pixel segmentation. The results on the rest four datasets are available in the Supplementary File.

The results regarding varying depth parameters (k) are summarized in Table V. For the DRIVE dataset, PCNet achieves optimal evaluation metrics, including Acc, F1, and IoU, at $k = 5$. Additionally, Rec and Pre reached their second-best performance, respectively. Most metrics initially increase 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. The similar trend and analysis of the rest four datasets can be found in the Supplementary File.

F. Progressive Mode Analysis

In Section III, we discuss the impact of progressive connected mode on overall segmentation accuracy and the utilization ratio of intermediate segmentation maps. Specifically, we define the single multiplication as the hard connection constraint and the single addition as the soft one. In this section, we conduct a series of experiments to demonstrate the superior performance of the newly proposed dual soft-hard constraint strategy. To ensure comprehensive evaluation, we compare five different connection strategies: feature concatenation, single soft, single hard connection, the proposed dual soft-hard constraint, and an additional attention-based scheme using Squeeze-and-Excitation (SE) [43] modules. All experiments are conducted on the DRIVE dataset.

To further explore the influence of different progressive strategies on segmentation performance, we employ SerialUnet [16] and the proposed PCNet as backbone architectures. Both models utilize a serial progressive design and are evaluated combining with several different strategies, including SE mode, traditional concatenation mode, single soft, hard connection, and the proposed novel dual soft-hard constraint approach. The experimental results are summarized in Table VI, where the best metrics are marked in bold. Notably, across both SerialUnet and PCNet, replacing the conventional concatenation method with the dual soft-hard constraint consistently enhances evaluation metrics, including Acc, F1, Rec, Pre, and IoU, underscoring the effectiveness of the proposed strategy. However, for SerialUnet, replacing the traditional concatenation with either hard or soft connection resulted in a slight decrease in F1, Rec, and Pre. This decline could be attributed to inadequate feature fusion in the single progressive mode, which may have limited segmentation accuracy. In contrast, the performance drop is mitigated in the

TABLE VI. The results of experiments on different progressive mode for DRIVE dataset.

Backbones	Modes					Evaluation Metrics (Mean ± Std)				
	SE	Concate	Hard	Soft	Dual (Proposed)	Acc (%)	F1 (%)	Rec (%)	Pre (%)	IoU (%)
SerialUnet [16]	✓					96.59 ± 0.03	80.00 ± 0.13	77.80 ± 1.08	82.45 ± 1.04	66.56 ± 0.02
		✓				96.81 ± 0.01	80.93 ± 0.22	77.93 ± 1.14	84.30 ± 0.90	67.01 ± 0.31
			✓			96.84 ± 0.01	80.90 ± 0.16	76.89 ± 0.63	83.88 ± 0.46	67.97 ± 0.22
				✓		96.84 ± 0.00	80.87 ± 0.10	76.63 ± 0.42	83.12 ± 0.33	67.93 ± 0.14
					✓	**96.88 ± 0.02**	**80.98 ± 0.14**	**78.18 ± 1.02**	**84.52 ± 0.93**	**68.08 ± 0.18**
PCNet (Ours)	✓					96.80 ± 0.03	80.51 ± 0.04	76.03 ± 1.32	84.19 ± 1.01	67.42 ± 0.05
		✓				96.89 ± 0.01	81.41 ± 0.15	78.35 ± 0.78	85.21 ± 0.62	68.70 ± 0.21
			✓			96.89 ± 0.01	81.40 ± 0.17	78.20 ± 0.90	85.35 ± 0.73	68.68 ± 0.24
				✓		96.89 ± 0.01	81.55 ± 0.18	**79.00 ± 0.92**	84.75 ± 0.73	68.88 ± 0.25
					✓	**96.92 ± 0.01**	**81.67 ± 0.24**	78.68 ± 1.25	**85.39 ± 0.96**	**69.07 ± 0.34**

proposed PCNet, and even showed marginal improvement with soft connection, reflecting its superior feature extraction and fusion capabilities compared to traditional serial architectures. However, introducing the SE connection results in a consistent decline in performance across both two backbones. This may stem from the fact that SE modules primarily focus on global channel-wise attention, which could weaken the spatial and hierarchical information flow essential for effective feature fusion in progressive structures.

In addition to the quantitative results, the qualitative analysis of segmentation outputs, as illustrated in Fig. 13, further demonstrates that PCNet, utilizing the proposed dual soft-hard constraint, yields more coherent and accurate segmentation results compared to traditional concatenation method. The visual evidence emphasizes the efficacy of the proposed architecture in enhancing segmentation quality. As highlighted in the red and yellow boxes in the first row, it is evident that PCNet with a single soft connection tends to misclassify ambiguous structures as target pixels, resulting in over-segmentation in regions where caution is warranted. This tendency may arise from the inherent flexibility of the soft connection, which allows the model to explore uncertain areas liberally, leading to the misidentification of background structures as target objects. In contrast, when employing the single hard connection, the model adopts a more conservative strategy. While this approach is beneficial for minimizing false positives, it may inadvertently result in the misclassification of similar structures as background pixels. Such conservative behavior can restrict the ability of model to capture all relevant features, particularly when target structures are closely like with background elements. These observations corroborate the analysis presented in Section III regarding the influence of soft and hard connections on segmentation outcomes. The comparative assessment underscores the advantages of the dual soft-hard constraint in achieving balanced and precise segmentation. This approach effectively addresses the limitations encountered with single soft or hard connections by harnessing the strengths of both methods to enhance overall performance. Ultimately, the qualitative findings complement the quantitative results, reinforcing the conclusion that the dual constraint mechanism significantly improves the robustness and accuracy of the PCNet model.

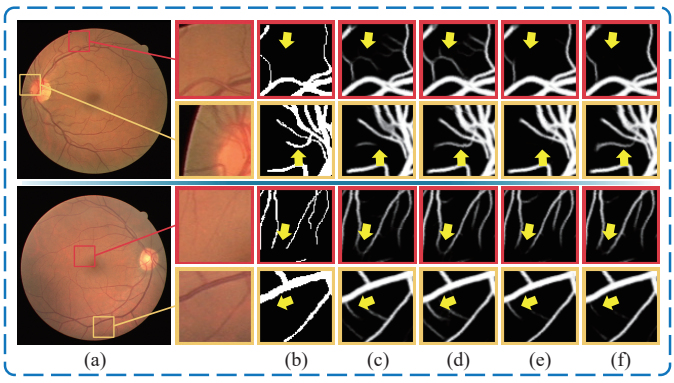


Fig. 13. The visualization of ablation studies on different progressive modes. (a) refers the original images. (b) represents the ground truth of certain regions. (c), (d), (e), and (f) illustrate the results produced by the PCNet with progressive mode of traditional concatenate, single soft, single hard connection, and the newly proposed dual soft-hard constraint strategy respectively.

G. Generalization and Robustness Study

To evaluate the generalization ability of our method under domain shifts, we conducted a cross-dataset experiment between the DRIVE and FIVES datasets. Specifically, we trained models on DRIVE and tested on FIVES, and vice versa. We compared three architectures in this setting, including the baseline UNet, SerialUNet, and our proposed PCNet. As shown in Table VII, while all models experienced a performance drop due to domain differences, PCNet consistently achieved better results than the other methods in all comparison metrics, demonstrating its stronger robustness and generalization across datasets.

We attribute this improved generalization to the progressive feature integration strategy and the segmentation potential activation mechanism introduced in PCNet. These components allow the network to gradually incorporate multi-scale semantic cues while selectively enhancing features with high segmentation relevance, thereby enabling more stable and accurate performance under varying data distributions.

V. Conclusion

In this study, we propose a novel potential-guided connected network (PCNet) to address the challenge of accurately identifying tiny structures, thereby achieving high-quality segmentation results. The network is composed of two core

TABLE VII. The results of generalization and robustness study.

Datasets	Methods	Evaluation Metrics		
		Acc (%)	F1 (%)	Rec (%)
DRIVE –>FIVES	UNet [8]	96.72	54.87	54.46
	SerialUnet [16]	96.55	54.24	64.75
	PCNet (Ours)	**97.66**	**69.13**	**77.46**
FIVES –>DRIVE	UNet [8]	96.65	80.44	79.07
	SerialUnet [16]	96.61	79.34	74.65
	PCNet (Ours)	**96.67**	**80.53**	**82.27**

modules, including the intermediate generation (IG) module and the progressive inference (PI) module. The IG module utilizes a newly designed serially connected UNet to generate a series of intermediate segmentation maps with varying levels of segmentation accuracy. In contrast to previous studies, we explicitly consider segmentation potential within these intermediate maps. The progressive context transform (PCT) block, integrated into the PI module, progressively reveals these potentials, significantly enhancing the final segmentation performance. To explore more efficient feature detection architectures, we experimented with multiple progressive connected modes. Ablation studies highlight that the newly proposed dual soft-hard constraint strategy effectively combines the advantages of both soft and hard connections, further optimizing segmentation results. Experimental results on three publicly available datasets, DRIVE, MoNuSeg, and CoNIC, all of which contain numerous tiny structures, demonstrate that the proposed network could accurately detect even the smallest structures compared to SOTA methods. Additionally, experiments conducted on a multi-scale dataset, GlaS, reveal that the PCNet could also effectively handle multi-scale medical image segmentation tasks, underscoring its robustness and versatility in medical image analysis.

REFERENCES

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[2] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *IEEE Signal Process. Mag.*, vol. 29, pp. 140–145, 2012.

[3] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. Van Ginneken, J. C. Folk, and M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 30, pp. 1941–1950, 2011.

[4] N. F. Greenwald, G. Miller, E. Moen, A. Kong, A. Kagel, T. Dougherty, C. C. Fullaway, B. J. McIntosh, K. X. Leow, M. S. Schwartz *et al.*, "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning," *Nat. Biotechnol.*, vol. 40, pp. 555–565, 2022.

[5] S. K. Singh, I. D. Clarke, M. Terasaki, V. E. Bonn, C. Hawkins, J. Squire, and P. B. Dirks, "Identification of a cancer stem cell in human brain tumors," *Cancer Res.*, vol. 63, pp. 5821–5828, 2003.

[6] L. Fang and H. Qiao, "Diabetic retinopathy classification using a novel DAG network based on multi-feature of fundus images," *Biomed. Signal Process. Control*, vol. 77, p. 103810, 2022.

[7] Y. Tian and S. Fu, "A descriptive framework for the field of deep learning applications in medical images," *Knowl.-Based Syst.*, vol. 210, p. 106445, 2020.

[8] 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.

[9] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 10 012–10 022.

[10] 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.

[11] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Pro. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 472–480.

[12] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 764–773.

[13] J. Yi, C. Chen, Q. Wei, D. Ding, and G. Yang, "MMF-Net: A novel multimodal multiscale fusion network for artery/vein segmentation in retinal fundus," in *Proc. Int. Conf. Syst. Man Cybern. (SMC)*, 2022, pp. 1192–1197.

[14] P. Tang, P. Yang, D. Nie, X. Wu, J. Zhou, and Y. Wang, "Unified medical image segmentation by learning from uncertainty in an end-to-end manner," *Knowl.-Based Syst.*, vol. 241, p. 108215, 2022.

[15] X. Huang, H. Gong, and J. Zhang, "HST-MRF: Heterogeneous swin transformer with multi-receptive field for medical image segmentation," *IEEE J. Biomed. Health Inform.*, vol. 28, pp. 4048–4061, 2024.

[16] 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.

[17] Y. Qi, Y. He, X. Qi, Y. Zhang, and G. Yang, "Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 6070–6079.

[18] Y. Wang, L. Gu, T. Jiang, and F. Gao, "MDE-UNet: A multitask deformable UNet combined enhancement network for farmland boundary segmentation," *IEEE Geosci. Remot. S.*, vol. 20, pp. 1–5, 2023.

[19] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Pro. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11 030–11 039.

[20] 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.

[21] 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.

[22] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *Proc.-ICASSP IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 1055–1059.

[23] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, pp. 1597–1605, 2018.

[24] J. Yi, C. Chen, and G. Yang, "Retinal artery/vein classification by multi-channel multi-scale fusion network," *Appl. Intell.*, vol. 53, pp. 26 400–26 417, 2023.

[25] X. Shu, Y. Yang, and B. Wu, "Adaptive segmentation model for liver CT images based on neural network and level set method," *Neurocomputing*, vol. 453, pp. 438–452, 2021.

[26] 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.

[27] 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.

[28] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, "KiU-Net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation," *IEEE Trans. Med. Imag.*, vol. 41, pp. 965–976, 2021.

[29] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple U-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Med. Imag. Anal.*, vol. 65, p. 101786, 2020.

[30] 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.

[31] Y. Shu, J. Zhang, B. Xiao, and W. Li, "Medical image segmentation based on active fusion-transduction of multi-stream features," *Knowl.-Based Syst.*, vol. 220, p. 106950, 2021.

[32] K. A. Eppenhof, M. W. Lafarge, M. Veta, and J. P. Pluim, "Progressively trained convolutional neural networks for deformable image registration," *IEEE Trans. Med. Imag.*, vol. 39, pp. 1594–1604, 2019.

[33] 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.

[34] 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.

[35] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[36] A. L. Maas, A. Y. Hannun, A. Y. Ng *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, p. 3.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

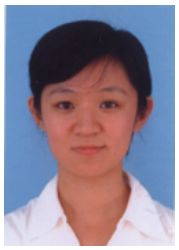
[41] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Imag. Anal.*, vol. 35, pp. 489–502, 2017.

[42] 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.

[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Pro. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

Chouyu Chen received the B.S. degree in Automation from the University of Science and Technology Beijing, China, in 2020. And M.S. degree in Computer Science and Technology from Beijing University of Civil Engineering and Architecture, China, in 2023. He is currently pursuing the Ph.D. degree with the University of Toyama, Toyama, Japan. His research interests include artificial intelligence, neural networks, deep learning, pattern recognition, and image processing.

Yaotong Song received the B.S. degree from Beijing University of Civil Engineering and Architecture, Beijing, China. He is currently pursuing the M.E. degree from the University of Toyama, Toyama, Japan. His current interests include computational intelligence and neural networks for real-world applications.



Junyan Yi Junyan Yi received the B.S. degree and M.S. degree from Shandong University, Shandong, China in 2001 and 2005 respectively, and received the Ph.D degree from University of Toyama, Toyama, Japan in 2009, all in computer science. Between 2007 and 2009, she worked as a Research Assistant at University of Toyama. She joined the College of Computer Science & Technology, Zhejiang University of Technology, China in 2009 as an Associate Professor. She is currently an Associate Professor with the Department of Computer Science & Technology, Beijing University of Civil Engineering and Architecture, China. Her main research interests include intellectual information technology, neural networks, data mining, image processing and optimizations problems.

Lijun Guo received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2011. He is currently a Full Professor with Ningbo University, Ningbo, China. His research focuses on intelligent computing, computer vision, neuromorphic computing, and medical image analysis. He serves as a committee member for the Cognitive Computing Special Committee of the Chinese Association for Artificial Intelligence.

Zhenyu Lei (Member, IEEE) received the Ph.D. degree in Science and Engineering from the University of Toyama, Toyama, Japan, in 2023. He is currently an Assistant Professor at the Faculty of Engineering, University of Toyama, Japan. His current research interests include evolutionary computation, machine learning, and neural networks for real-world applications and optimization problems.

Shangce Gao (Senior Member, IEEE) received his Ph.D. degree in Innovative Life Science from University of Toyama, Toyama, Japan in 2011. He is currently a Professor with the Faculty of Engineering, University of Toyama, Japan. His current research interests include nature-inspired technologies, machine learning, and neural networks for real-world applications. He serves as an Associate Editor for many international journals such as IEEE Transactions on Neural Networks and Learning Systems, and IEEE/CAA Journal of Automatica Sinica.

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 S.I.

TABLE S.I: 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

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:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP}, \tag{1}$$

$$Rec = \frac{TP}{TP + FN}, \tag{2}$$

$$Pre = \frac{TP}{TP + FP}, \tag{3}$$

$$F1 = \frac{2 \times Rec \times Pre}{Rec + Pre}, \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN}. \tag{5}$$

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. S.I, S.II, S.III, and S.IV while the corresponding trajectories of the segmentation potential metric is depicted in Fig. S.V.

II. Comparison Experiments

The comparison methods are summarized in Table S.II, 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], SwinPANet [10], BRAUNet++ [11], and EMCAD [12]. Among these, UNet and UNet++ represent classical segmentation 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.

TABLE S.II: 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

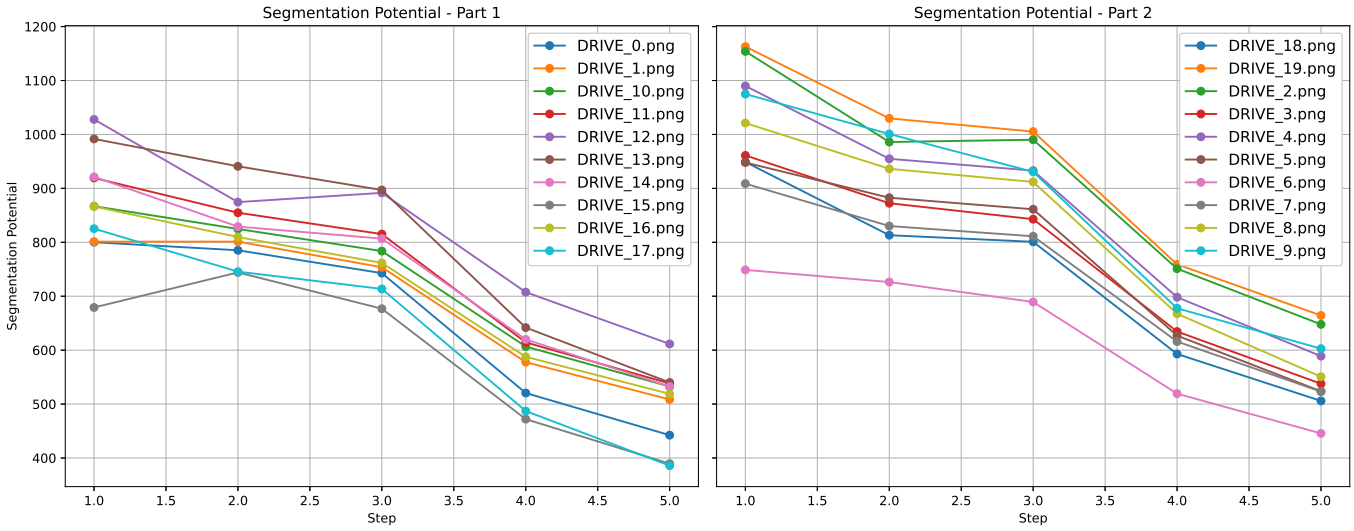


Fig. S.V: Trend of the segmentation potential metric for all DRIVE dataset images over different stages of IG module output.

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

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

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 S.III, where the best metrics are highlighted in

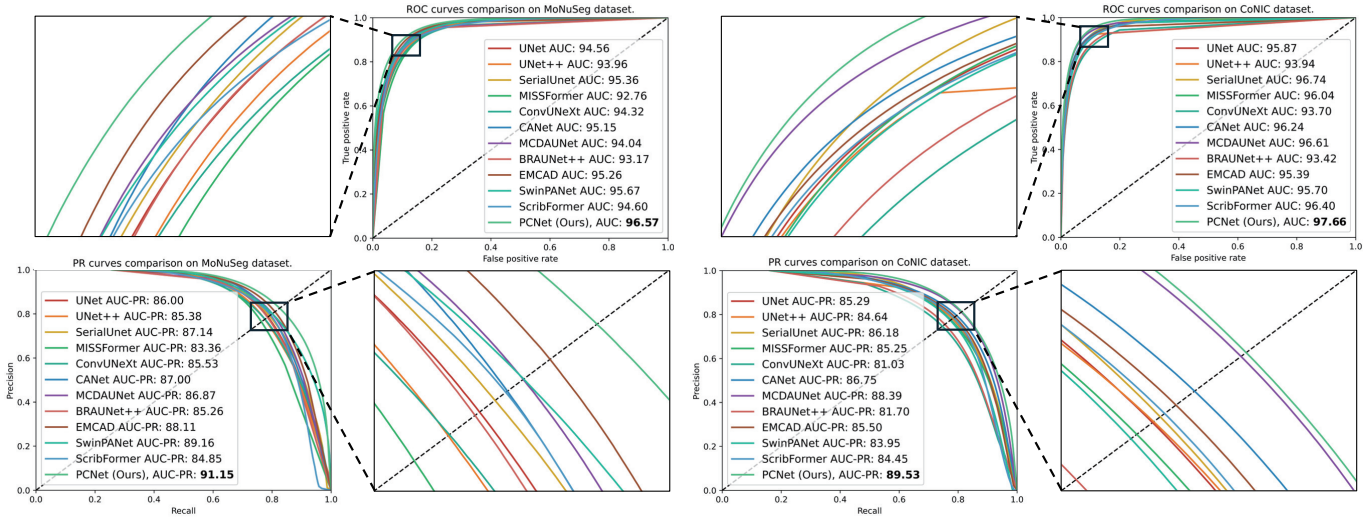


Fig. S.VI: The visualization of ROC and PR curves of comparison experiments on MoNuSeg and CoNIC datasets.

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, 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. S.VI further illustrate these findings, highlighting the superior performance of the proposed method.

The visualization of comparative experiments on traditional datasets is presented in Fig. S.VII. 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, underscoring 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. S.VIII 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 S.IV, PCNet outperforms the second-best method, EMCAD, across all evaluation metrics, demonstrating the superior effectiveness of the proposed approach. Specifically, the Acc

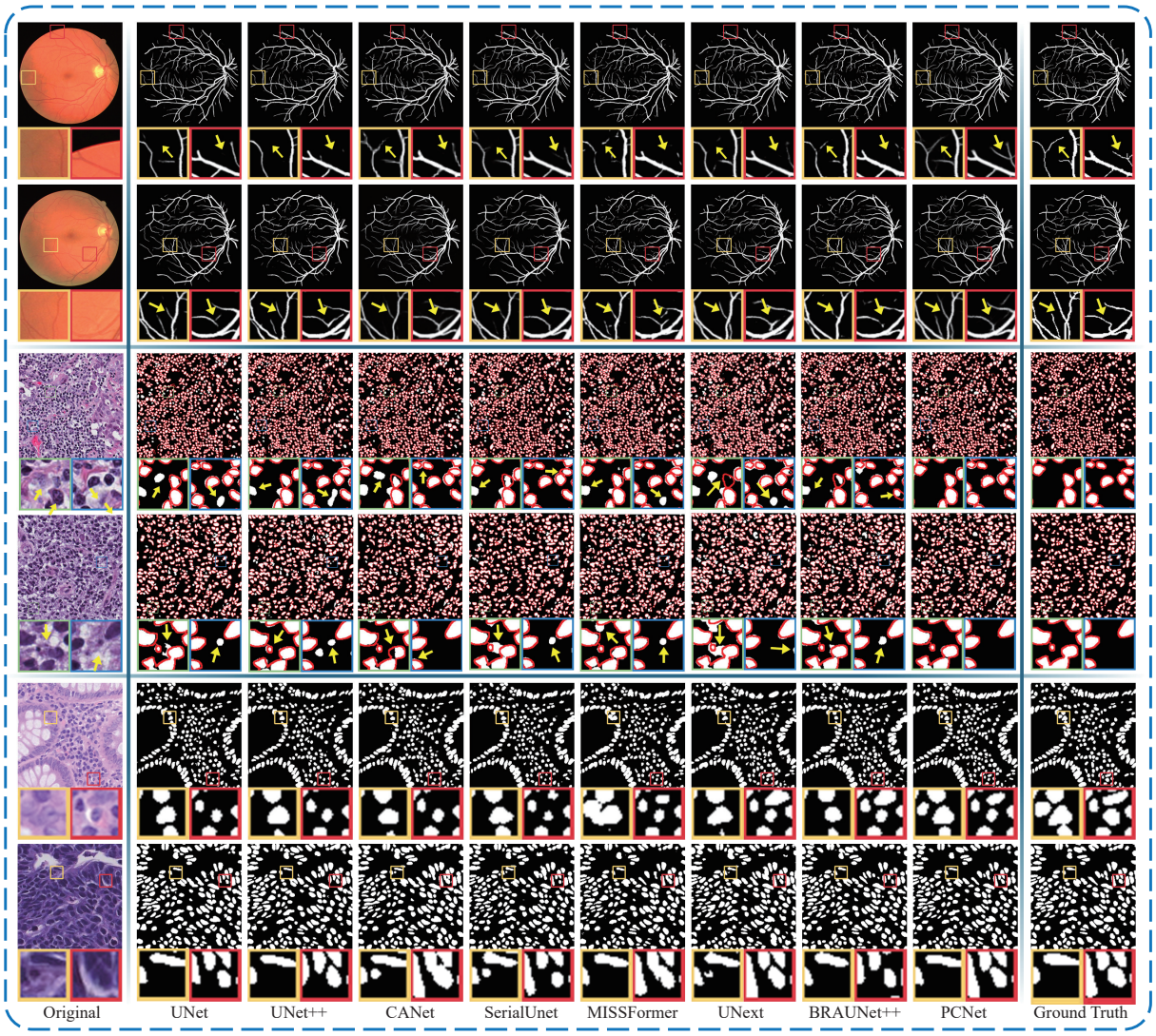


Fig. S.VII: Samples from the DRIVE, MoNuSeg, and CoNIC datasets. First and second rows show sample from the DRIVE dataset, third and fourth rows show image from the MoNuSeg dataset, while the last two rows show sample from the CoNIC dataset.

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. S.VIII, 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], UNet++ [7], CANet [8], SerialUnet [13], MISSFormer [9], ConvUNeXt [14], BRAUNet++ [11], and the proposed PCNet, are illustrated in Fig. S.IX. 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. S.IX, PCNet effectively extracts the

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

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

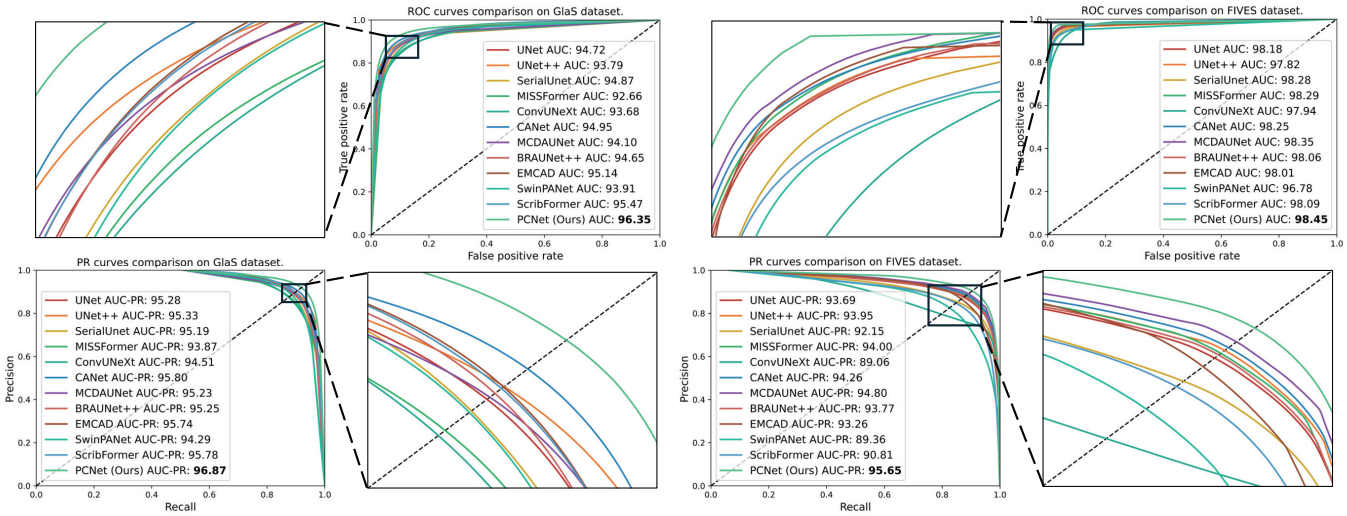


Fig. S.VIII: The visualization of ROC and PR curves of comparison experiments on GlaS and FIVES datasets.

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. S.IX, 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 S.V. 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

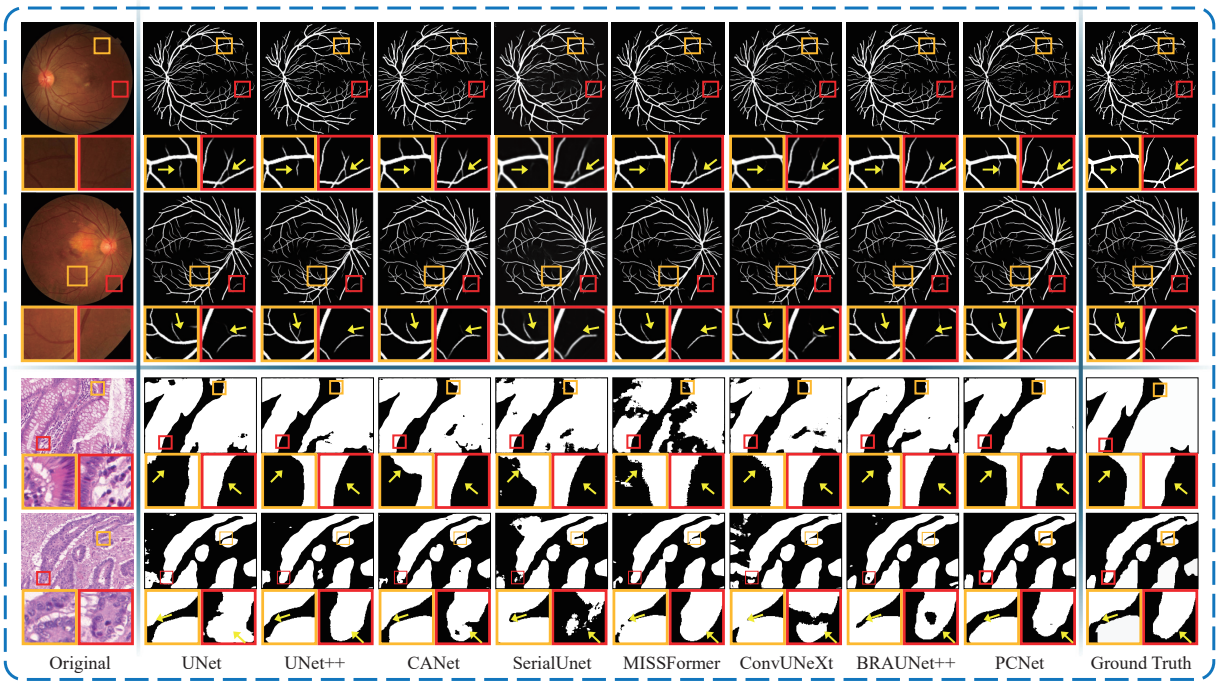


Fig. S.IX: Visualization of the changing trends across all evaluation metrics for four different datasets. All values are normalized, and the curves are smoothed using an interpolation technique for clarity.

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

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

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 S.V, 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

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

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

SerialUnet. As presented in the third part of Table S.V, 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 S.VI. 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 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 S.VII. 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

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

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

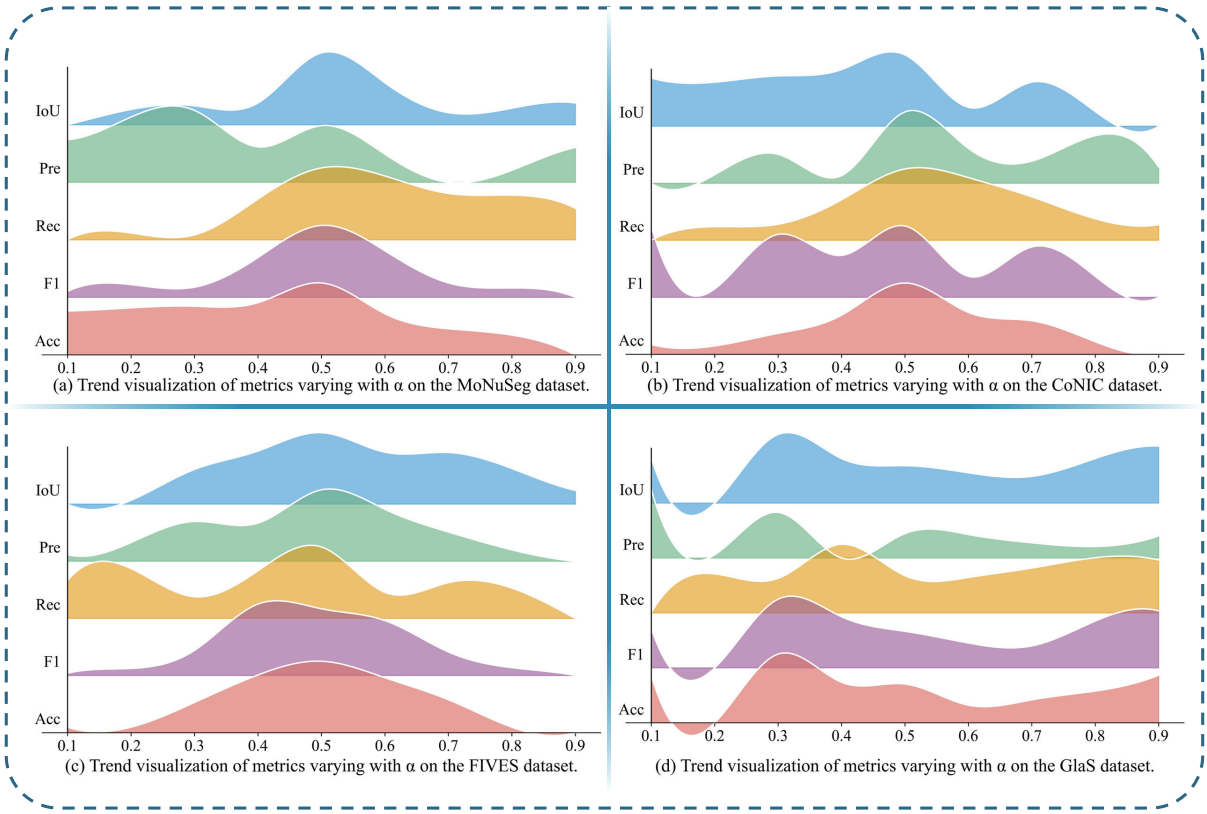


Fig. S.X: Visualization of the trends across all evaluation metrics for four different datasets as α changes. All values are normalized, and the curves are smoothed using an interpolation technique for better clarity.

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. S.X. 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.

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

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

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 S.VIII, 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 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. S.XI. 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.

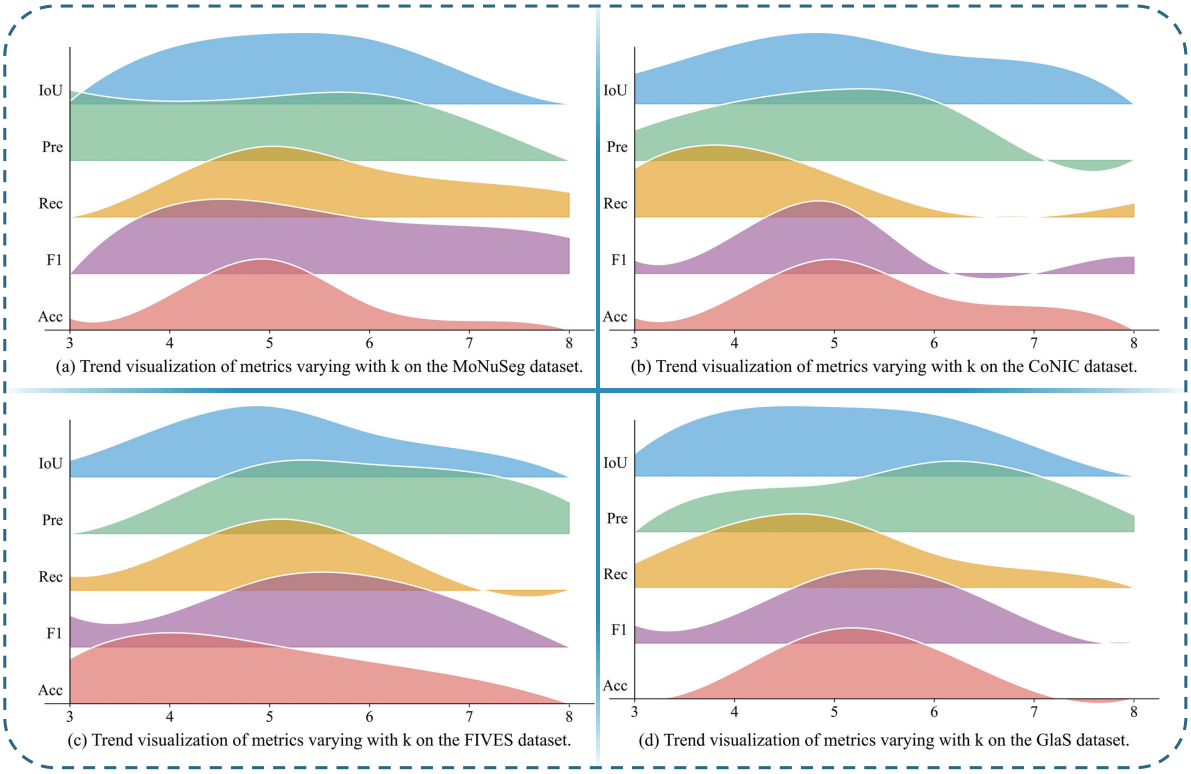


Fig. S.XI: Visualization of the trends across all evaluation metrics for four different datasets as k changes. All values are normalized, and the curves are smoothed using an interpolation technique for clarity.

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.

As shown in Table S.IX, 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. S.XII 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.

TABLE S.IX: The results of experiments on different progressive modes

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

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.

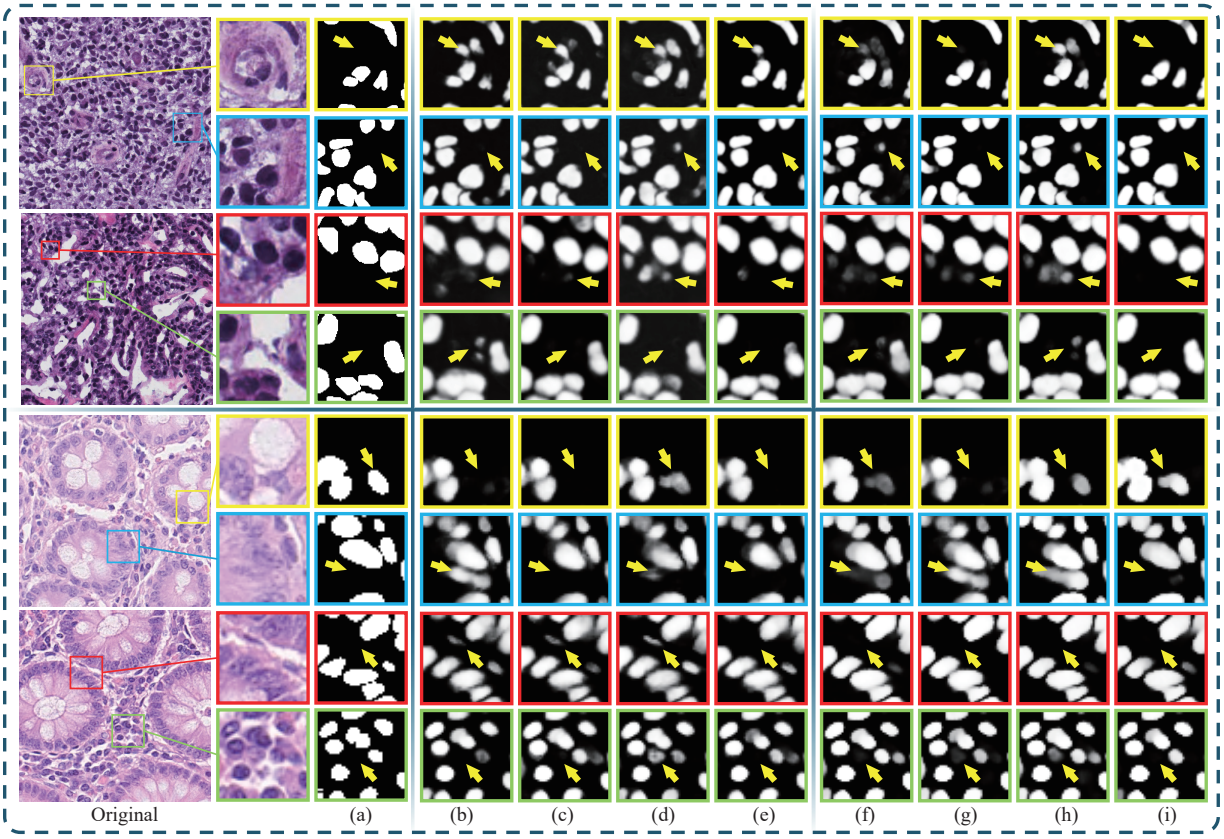


Fig. S.XII: Visualization results of structure exploration experiments. (a) Ground truth of the original images. (b) and (f) Results obtained using SerialUnet and PCNet with traditional concatenation connections. (c) and (g) Results obtained using single hard connections in the two baseline networks.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

[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.