frontiers in Medicine

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/fmed.2021.761050

Pyramid-Net: Intra-layer Pyramid-Scale Feature Aggregation Network for Retinal Vessel Segmentation

Jiawei Zhang[1,2,3,4†], *Yanchun Zhang*[4,5,6*], *Hailong Qiu*[1†], *Wen Xie*[1†], *Zeyang Yao*[1†], *Haiyun Yuan*[1], *Qianjun Jia*[1], *Tianchen Wang*[3], *Yiyu Shi*[3], *Meiping Huang*[1*], *Jian Zhuang*[1*] and *Xiaowei Xu*[1*]

[1] Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital, Guangdong Cardiovascular Institute, Guangdong Academy of Medical Sciences, Guangzhou, China, [2] Shanghai key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai, China, [3] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States, [4] Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Wenzhou, China, [5] Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China, [6] College of Engineering and Science, Victoria University, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Jun Feng,
Northwest University, China

Reviewed by:
Juanying Xie,
Shaanxi Normal University, China
Márton Szemenyei,
Budapest University of Technology and Economics, Hungary
Erlei Zhang,
Northwest A&F University, China

***Correspondence:**
Yanchun Zhang
yanchun.zhang@vu.edu.au
Meiping Huang
huangmeiping@126.com
Jian Zhuang
Zhuangjian5413@163.com
Xiaowei Xu
xiao.wei.xu@foxmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: 19 August 2021
Accepted: 05 November 2021
Published: 07 December 2021

Citation:
Zhang J, Zhang Y, Qiu H, Xie W, Yao Z, Yuan H, Jia Q, Wang T, Shi Y, Huang M, Zhuang J and Xu X (2021) Pyramid-Net: Intra-layer Pyramid-Scale Feature Aggregation Network for Retinal Vessel Segmentation. Front. Med. 8:761050. doi: 10.3389/fmed.2021.761050

Retinal vessel segmentation plays an important role in the diagnosis of eye-related diseases and biomarkers discovery. Existing works perform multi-scale feature aggregation in an inter-layer manner, namely **inter-layer feature aggregation**. However, such an approach only fuses features at either a lower scale or a higher scale, which may result in a limited segmentation performance, especially on thin vessels. This discovery motivates us to fuse multi-scale features in each layer, **intra-layer feature aggregation**, to mitigate the problem. Therefore, in this paper, we propose Pyramid-Net for accurate retinal vessel segmentation, which features intra-layer pyramid-scale aggregation blocks (IPABs). At each layer, IPABs generate two associated branches at a higher scale and a lower scale, respectively, and the two with the main branch at the current scale operate in a **pyramid-scale** manner. Three further enhancements including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections are proposed to boost the performance. We have evaluated Pyramid-Net on three public retinal fundus photography datasets (DRIVE, STARE, and CHASE-DB1). The experimental results show that Pyramid-Net can effectively improve the segmentation performance especially on thin vessels, and outperforms the current state-of-the-art methods on all the adopted three datasets. In addition, our method is more efficient than existing methods with a large reduction in computational cost. We have released the source code at https://github.com/JerRuy/Pyramid-Net.

Keywords: deep learning, neural network, feature aggregation, pyramid scale, retinal vessel segmentation

1. INTRODUCTION

The subtle changes in the retinal vascular, including vessel width, tortuosity, and branching features, indicate mass eye-related diseases, such as diabetic retinopathy (1), glaucoma (2), and macular degeneration (3). Meanwhile, those characteristics are important biomarkers for numerous systemic diseases, including hypertension (4) and cardiovascular diseases (5). Retinal

vessel segmentation is one of the cornerstones to access those characteristics, particularly for automatic retinal image analysis (6, 7). For example, hypertensive retinopathy is a retinal disease, which is caused by hypertension. Increased vascular curvature or stenosis can be found in patients with hypertension (8). Conventionally, manual segmentation is laborious and time-consuming, and suffers subjectivity among experts. To improve efficiency and reliability and reduce the workload of doctors, the clinical practice puts forward high requirements for automatic segmentation (9).

Recently, deep neural networks have boosted the segmentation performance of retinal vessel segmentation (10, 12) by a large margin compared with traditional methods (13, 14). However, thin vessels cannot be segmented accurately. For example, **Figure 1** demonstrates a commonly-seen fundus image containing numerous thin vessels and thick vessels, and corresponding segmentation (11) and ground truth. We can easily notice that the thick vessels enjoy a promising performance, but the thin vessels suffer a big miss. A potential reason is that the continuous pooling operations in most neural networks are used to encode the features, which leads to a mass loss of appearance information and harms the segmentation accuracy, especially on thin vessels. Note that in practice, it is also difficult to segment these thin vessels for experts due to low contrast and ambiguousness. Currently, some works have been proposed to tackle the above problems, e.g., a particular processing branch for thin vessels (12), a new loss function to emphasize thin vessels (10). However, the segmentation performance is still limited considering the clinical requirement of retinal image analysis.

Meanwhile, **multi-scale feature aggregation** to fuse coarse-to-fine context information has been popular to segment thin/small objects (15–19). There are mainly two approaches: input-output level category and intra-network level category. In the input-output level category, connections exist between inputs at various scales and corresponding intermediate layers (15), or between the intermediate layers and the final predictions with corresponding scales (18). In the intra-network level category, features from previous layers are adjusted in channel numbers and spatial dimension and then aggregated with the ones in the later layer (16). However, current multi-scale feature aggregation works in an inter-layer manner, **inter-layer feature aggregation**, which can only fuse features at either a lower scale or a higher scale. For example, in the encoder, feature maps at the lower scale cannot be fused by that at the current scale because of the processing order of the layers. A possible solution is to fuse multi-scale features in each layer, **intra-layer feature aggregation**, to consider features at both the high scale and the low scale.

Motivated by the above discoveries, in this paper, we propose Pyramid-Net for accurate retinal vessel segmentation. In each layer of Pyramid-Net, intra-layer pyramid-scale aggregation blocks (IPABs) are employed in both the encoder and the decoder to aggregate features at pyramid scales (the higher scale, the lower scale, and the current scale). In this way, two associated branches at the higher scale and the lower scale are generated to assist the main branch at the current scale. Therefore, coarse-to-fine context information is shared and aggregated in each layer, thus improving the segmentation accuracy of capillaries. To further improve the performance, three optimizations, including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections, are applied to IPABs. We have conducted comprehensive experiments on three retinal vessel image segmentation datasets, including DRIVE (20), STARE (21), and CHASE-DB1 (22) with various segmentation networks. The experimental results show that our method can significantly improve the segmentation performance, especially on thin vessels, and achieves state-of-the-art performance on the three public datasets. In addition, our method is more efficient than the existing method with a large reduction in computational cost.

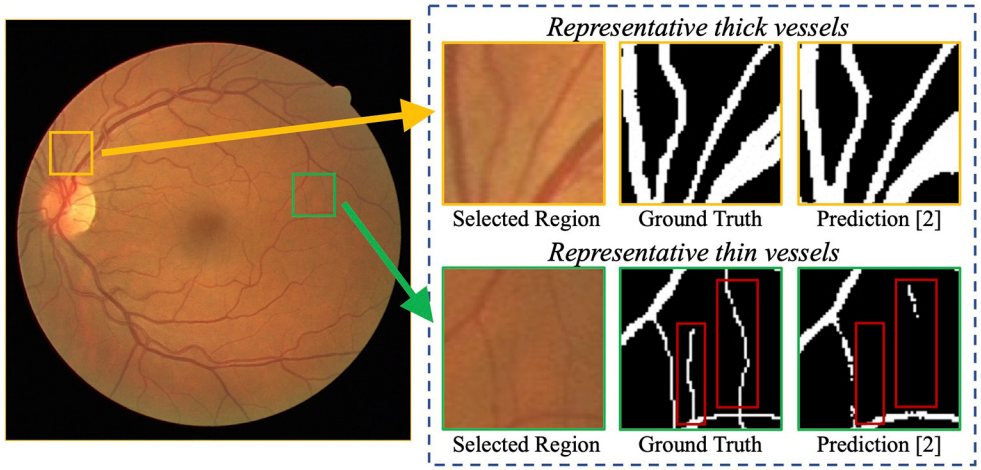

FIGURE 1 | Examples of challenging thin vessels in retinal vessel segmentation. The retinal fundus image (left) contains numerous thin vessels (1–2 pixels wide) and thick vessels (3 pixels wide or more) (10). Regions of representative thin and thick vessels, and their corresponding ground truth and predictions (11) are shown in the right. It can be noticed that the thick vessels obtain a better segmentation performance, while the thin vessels suffer a big miss (indicated by red rectangles).

Overall, this work makes the following contributions:

1) We discovered that thin vessels suffer a big miss in the segmentation results of existing methods;
2) We proposed Pyramid-Net for retinal vessel segmentation in which intra-layer pyramid-scale aggregation blocks (IPABs) aggregate features at the higher, current, and lower scales to fuse coarse-to-fine context information in each layer;
3) We further propose three enhancements: pyramid input enhancement, deep pyramid supervision, and pyramid skip connections to boost the performance;
4) We conducted comprehensive experiments on three public vessel image datasets (DRIVE, STARE, and CHASE-DB1), and our method achieves the state-of-the-art performance on three datasets.

The remainder of this paper is organized as follows. Section 2 introduces related works and the motivation of the proposed method. Section 3 details the overall framework of the proposed Pyramid-Net, including IPABs and three optimizations (pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections). Section 4 first introduces datasets, implementation, and evaluation. Second, quantitative evaluations on three vessel image datasets, comparisons with the state-of-the-art algorithms, and several visual retinal segmentation results are presented. Third, several ablation studies that included evaluating the thin vessel, ablation analysis, and cross-training evaluation are discussed. Section 5 concludes the paper.

2. RELATED WORK AND MOTIVATION

2.1. Vessel Image Segmentation

With the emergence of numerous public-available retinal image datasets (20–22), the supervised vessel segmentation methods became popular in the community. Commonly-seen supervised methods consist of two steps: feature extraction and classification. Some methods extracted the color intensity (24) and principle components (25) from the images, while some methods utilized wavelet (26) and edge responses (27). In terms of classification, various classic classifiers, including Support Vector Machine (SVM) (28), perceptron (29), random decision forests (30), and Gaussian model (26) are commonly seen and widely used in traditional supervised vessel image segmentation. Recently, in the light of fully convolutional networks (FCNs) (31) and U-Net (23), data-driven deep learning-based methods have demonstrated promising results and dominated the area of vessel image segmentation. Yan et al. (10) pointed out that the training loss tends to ignore the loss of thin vessels and is dominated by the thick vessels, which may be caused by the imbalance between thin vessels and thick vessels. Furthermore, Yan et al. (12) explored a three-stage network separating the segmentation of thick vessels, thin vessels, and the vessel fusion into different stages to make full use of the difference between thick and thin vessels to improve the overall segmentation performance. Considering that the consecutive pooling may lead to accuracy loss, CE-Net (32) encodes the high-dimension information and preserves spatial information to improve the overall segmentation. HA-Net (33) dynamically assigns the regions in the image hard regions or simple regions, and then introduces attention modules to help the network concentrate on the hard region for accurate vessel image segmentation. Meanwhile, some works introduce the spatial attention (34) and the channel attention (34) to the vessel segmentation domain and achieve promising results. The proposed method extends considerably to our previous work (35), which only supply some simplified evaluation on two public available vessel segmentation datasets. In this work, we have added a new module named "pyramid skip connections," which furthers boost the performance. Meanwhile, we have added another widely-used dataset (STARE) to demonstrate the generalization of our proposed Pyramid-Net. Moreover, in terms of the analysis, we have supplied in-depth analyses of our method including evaluation on thin vessel segmentation, ablation analysis, and cross-training evaluation.

2.2. Motivation

Multi-scale feature aggregation is widely used in medical image segmentation, which fuses the previous feature maps with different scales to improve the network performance. As shown in **Figure 2**, recent works (36–39) introduced multi-scale feature aggregation to strengthen feature propagation, alleviate the vanishing gradient problem, and improve the overall segmentation. We divide those methods into two major categories: input-output level and intra-network level.

Input-output level category: The connections exist between inputs at various scales and corresponding intermediate layers, or between the intermediate layers and the final predictions with corresponding scales. For example, Wu et al. (40) generated multi-scale feature maps by max-pooling and up-sampling layer and employed two sub-models to extract and aggregate features at multiple scales. MIMO-Net (41) fused scaled input images with multiple resolutions into the intermediate layers of the network in the encoder, and optimized the features in the decoder to improve the overall segmentation performance. MILD-Net (42) fused scaled original images with multiple resolutions to alleviate the potential accuracy decline caused by max-pooling.

Intra-network level category: In this approach, features from previous layers are adjusted in channel numbers and spatial dimension and then aggregated with the ones in the later layer. For ease of discussion, we discuss the network structures of related works based on the U-Net as shown in **Figure 2**. Note that U-Net is the most widely-used network in medical image segmentation. These works contain three main approaches: dense connections in the encoder (encoder sub-level), dense connections in the decoder (decoder sub-level) and dense connections in the cross of the encoder and the decoder (cross sub-level): (1) Encoder sub-level: (15) aggregated the scale inputs into the intermediate layers in the encoder to alleviate the accuracy decline caused by pooling; (2) Decoder sub-level: Dense decoder short connections (18) made full use of the feature maps in the decoder by fusing them with the feature maps in later layers; (3) Cross sub-level: Complete bipartite networks (16) inspired by the structure of complete bipartite graphs connected every layer in the encoder and the decoder.

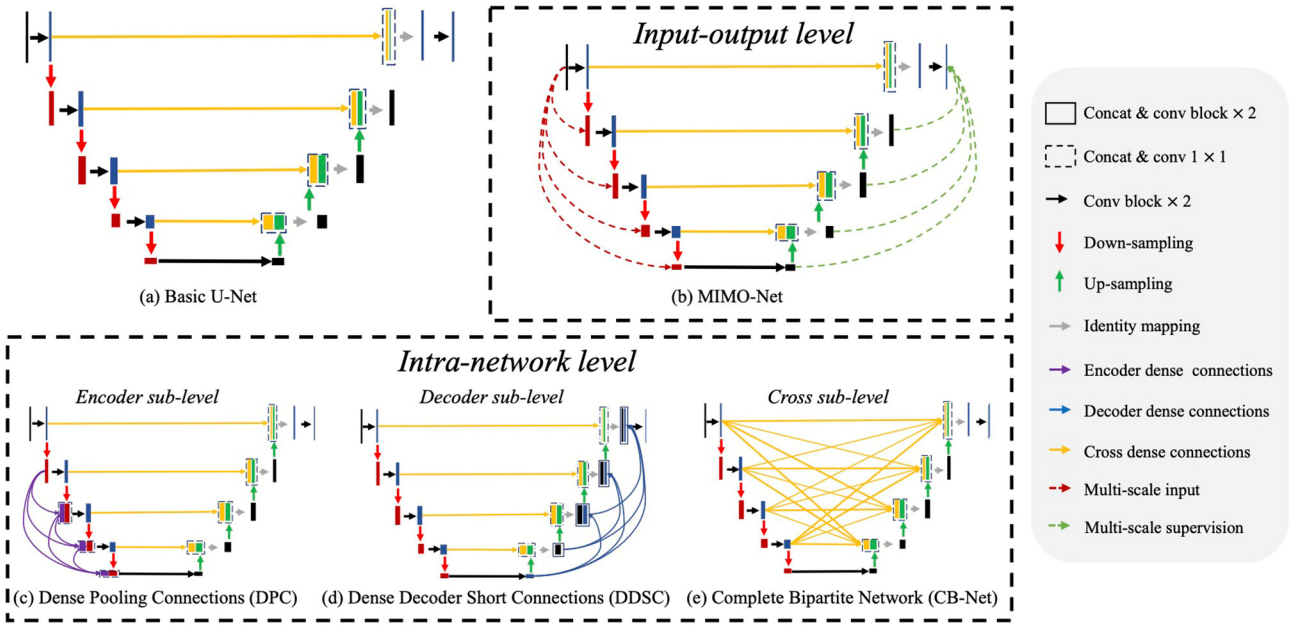

FIGURE 2 | Illustrations of network structures of **(a)** basic U-Net (23) and **(b–e)** existing multi-scale feature aggregation methods, which mainly consist of two major categories: input-output level and intra-network level. The input-output level category means that the network employs multiple scaled inputs, and the scaled ground truth supervises the inter feature maps. In the intra-network level category, the encoder level, the decoder level, and the cross-level indicate implemented multi-scale feature aggregation in the encoder, the decoder, and their cross, respectively.

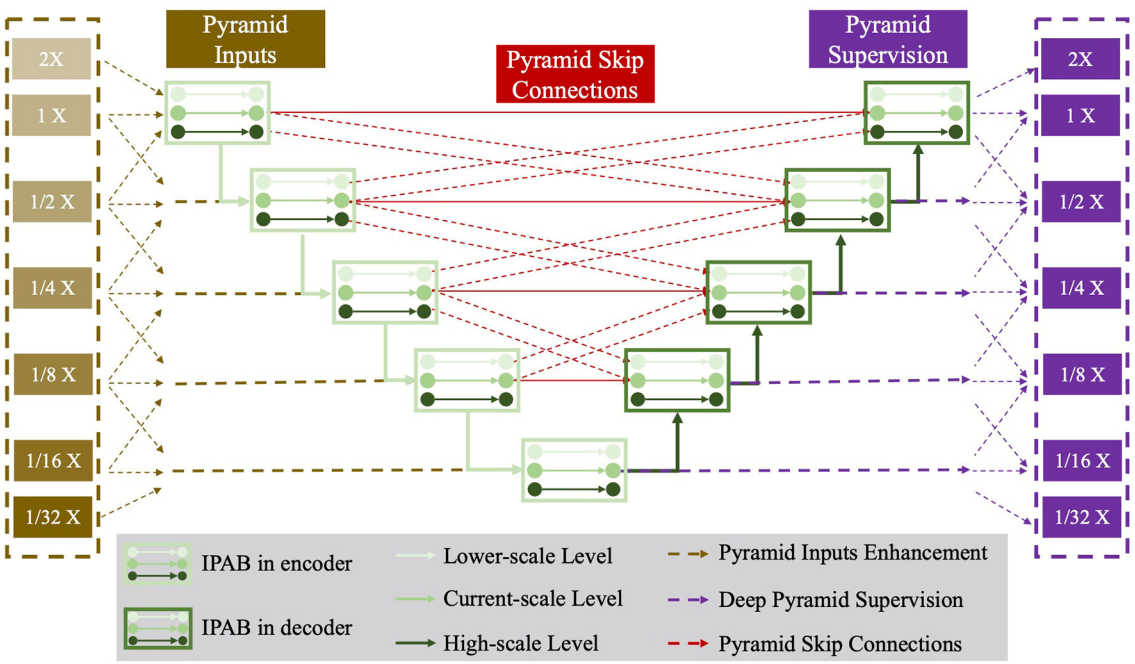

FIGURE 3 | The network structure of the proposed Pyramid-Net. IPABs (green rectangle) not only aggregate features at **pyramid scales** [the current scale (green line), the higher scale (dark green line) and the lower scale (bright green line)] containing coarse-to-fine context information. Meanwhile, pyramid input enhancement (yellow rectangle), deep pyramid supervision (purple rectangle), and pyramid skip connections (rad rectangle) are employed to further improve the overall segmentation. Best viewed in color.

Though multi-scale feature aggregation can significantly improve segmentation performance, we discover that they usually work in an inter-layer manner, **inter-layer feature aggregation**. In such a manner, features at either a lower scale or a higher scale are fused by the current layer. For example, in the encoder, feature maps at the lower scale cannot be fused by that at the current scale because of the processing order of the layers. The same phenomenon also exists in the decoder. Note that a successful segmentation needs to consider both feature maps at high scales for global localization information and low scales for detailed appearance information. Thus, we may mitigate the above problem by performing multi-scale feature aggregation in each layer of the network, **intra-layer feature aggregation**. How to obtain the multi-scale features in each layer becomes another problem. We may use pooling and upsampling to obtain two associated branches operating on a higher scale and a low scale, respectively. In this way, there exist three branches at three different scales (namely **pyramid scales**) in each layer, which is like a ResNet block (43). In this way, we may aggregate coarse-to-fine context information from pyramid-scale feature maps in each layer to further improve the segmentation performance.

3. METHODS

In this section, we first introduce IPABs and then describe three optimizations, including pyramid input enhancement, deep pyramid supervision, and pyramid skip connections. **Figure 3** presents the structure details of Pyramid-Net.

3.1. Intra-layer Pyramid-Scale Aggregation Block

Intra-layer pyramid-scale aggregation block are based on the ResNet block (43), which is widely adopted in deep learning. **Figure 4** illustrates the structure of the ResNet block (43), which is formulated as

$$X_{l+1} = f(X_l) + X_l, \tag{1}$$

where X_l and X_{l+1} are the input and the output of the current layer, while $f(\cdot)$ represents the main branch of the current layer. ResNet learns the additive residual function $f(\cdot)$ with respect to the unit input through a shortcut connection between them. Meanwhile, the multi-scale feature aggregation inspires us to propose associated branches to learn coarse-to-fine features in each residual branch. **Figure 4** illustrates the detailed structures of traditional ResNet blocks and our IPABs. Different from ResNet blocks, in each layer, IPABs generate two associated branches to aggregate coarse-to-fine feature maps to assist the main branch at the current scale. In each branch, the processing steps are almost the same as those in traditional ResNet blocks. Some extra steps such as up-sampling and down-sampling are adopted at the higher and the lower scales to adjust scales. In order to reduce the potential increase of computational cost, the number of channels of the inputs X_l in the main branch has been reduced to half, while the number of channels of resized inputs X_l^p and X_l^d in the associated branches is reduced to one-fourth. The feature maps with channel adjustment are fed to the processing steps at three scales and are processed in parallel. The three outputs at pyramid scales are then concatenated. The whole process is formulated as follows,

$$\widetilde{X}_{l+1} = H(f(\widehat{X}_l^p), f(\widehat{X}_l), f(\widehat{X}_l^d)) + X_l, \tag{2}$$

where X_l^p and X_l^d are the up-sampled and the down-sampled results of the current input X_l with channel adjustment, respectively. $\widehat{X}_l^p$, $\widehat{X}_l$ and $\widehat{X}_l^d$ are the enhanced results using pyramid input enhancement, which only exists in the encoder and is detailed in section 3.2. Meanwhile, $\widehat{X}_l^p$, $\widehat{X}_l$, and $\widehat{X}_l^d$ are replaced by $\hat{X}_l^p$, $\hat{X}_l$, and $\hat{X}_l^d$ in the decoder, which represents the enhancement results by pyramid skip connections and are detailed in section 3.4. $H(\cdot)$ represents the aggregation process, which performs re-scaling and feature concatenation. $\widetilde{X}_{l+1}$ is the strengthened results of X_{l+1} by IPAB.

The channel attention module selectively emphasizes interdependent channel maps by integrating associated features among all channel maps. To improve the efficiency of feature extraction, we also employ an attention mechanism (44, 45) in IPAB as follows,

$$\mathbf{\Phi}(\widetilde{\mathbf{X}}_{l+1}) = \mathbf{Q}(\Phi_{Avg}(\widetilde{\mathbf{X}}_{l+1})) + \mathbf{Q}(\Phi_{Max}(\widetilde{\mathbf{X}}_{l+1})). \tag{3}$$

$$\mathbf{\Psi}(\widetilde{\mathbf{X}}_{l+1}) = \boldsymbol{\sigma}(\mathbf{\Phi}(\widetilde{\mathbf{X}}_{l+1}) \otimes \widetilde{\mathbf{X}}_{l+1}. \tag{4}$$

where $\mathbf{\Psi}(\cdot)$ is the operation of attention process, $\mathbf{Q}$ is the conventional operation using 1×1 kernels for channel adjustment, and $\boldsymbol{\sigma}$ is the activation function. Average-pooling $\Phi_{Avg}(\cdot)$ and max-pooling $\Phi_{Max}(\cdot)$ are adopted to aggregate channel information. By utilizing IPAB, each layer of the network aggregates the feature with pyramid scales, which helps fuse coarse-to-fine context information to improve the overall segmentation performance.

3.2. Pyramid Input Enhancement

Pyramid input enhancement fuses the input image with multiple scales to IPABs to reduce the loss of information caused by re-scaling and enhance feature fusion. Pooling operations with various pooling sizes are used to guarantee spatial resolution consistency. Particularly, in each layer, the input image is scaled at higher, current, and lower scales, and fed to three parallel processing steps at multiple scales in the IPAB. Pooling operations over larger regions successively reinforce the scale and translation invariance while reducing noise sensitivity at the same time as more and more context information is added. The aggregation should facilitate discrimination between relevant features and local noises. The above three pyramid-scale images are concatenated with corresponding outputs of up-sampling, down-sampling, and channel adjustment, respectively. Suppose that X_l is denoted as the input of the current layer, and X_l^p, and X_l^d are results at the higher scale and the lower scale, respectively. Meanwhile, I_{l-1}, I_l and I_{l+1} are the scaled inputs of X_l^d, X_l, and X_l^p

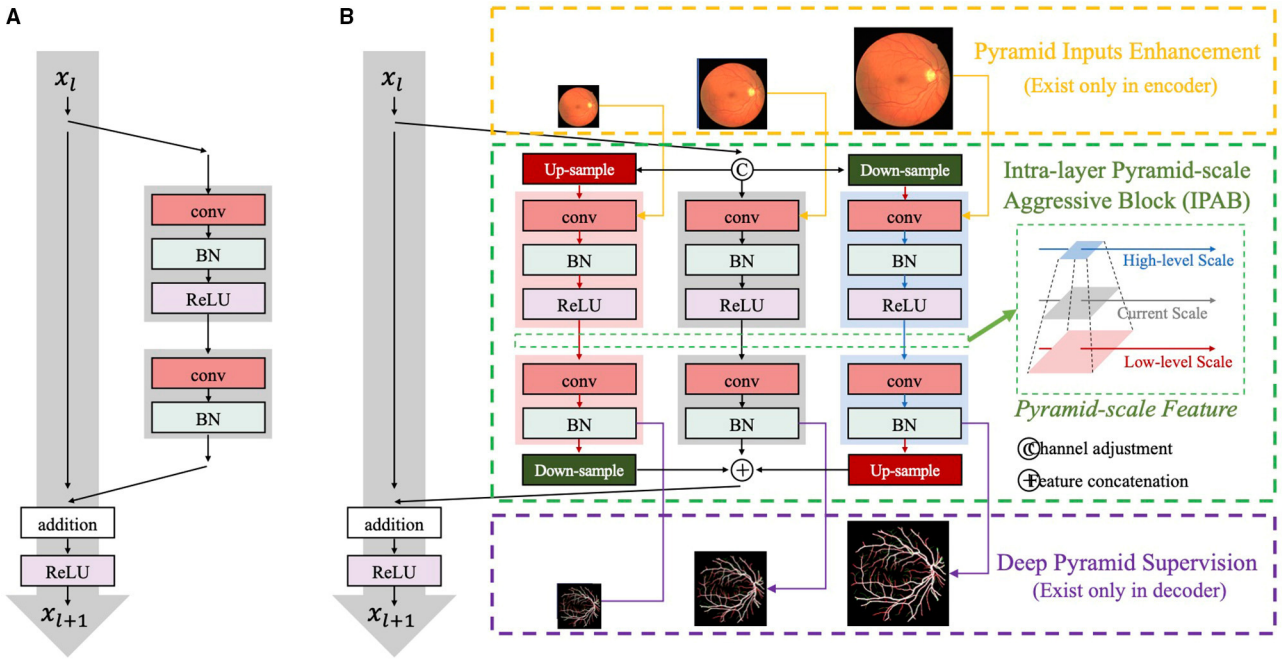

FIGURE 4 | The network structure of **(A)** ResNet blocks and **(B)** our intra-layer pyramid-scale aggregation blocks (IPABs). IPABs (marked by green rectangles) aggregate coarse-to-fine features at the current scale and both the higher scale and the lower scale (**pyramid scales**). Meanwhile, pyramid input enhancement (marked by yellow rectangles) and deep pyramid supervision (marked by purple rectangles) are employed to fuse the original images with corresponding scales, and supervise the intermediate results in each layer of the decoder, respectively.

with the same size, respectively. The fusion process of the current scale is formulated as follows,

$$\widehat{X}_{l-1} = H(X_l^d, \mathbf{W}^d(I_{l-1})), \tag{5}$$

$$\widehat{X}_l = H(X_l, \mathbf{W}(I_l)), \tag{6}$$

$$\widehat{X}_{l+1} = H(X_l^p, \mathbf{W}^p(I_{l+1})), \tag{7}$$

where $\mathbf{W}^p(\cdot), \mathbf{W}^d(\cdot)$, and $\mathbf{W}(\cdot)$ represents 3×3 convolutional operations and is applied before concatenating to the pyramid-scale features, and $H(\cdot)$ denotes channel adjustment.

3.3. Deep Pyramid Supervision

Deep pyramid supervision optimizes feature maps at multiple scales to improve the segmentation of multi-scale objects and fast the training process. Similar to pyramid input enhancement, deep pyramid supervision connects the intermediate layer to the final prediction thus fusing coarse-to-fine context information. Particularly, the feature maps at multiple scales from each IPAB in the decoder are fed into a plain 3×3 convolutional layer followed by Sigmoid function. Deep pyramid supervision at the lth scale of the decoder can be defined as,

$$L_l = L(Y_l^p, M_{l-1}) + L(Y_l, M_l) + L(Y_l^d, M_{l+1}). \tag{8}$$

The ground truths M are scaled to the same size as the pyramid-scale feature maps for deep supervision, e.g., Y_l^p, Y_l, and Y_l^d are supervised by the corresponding ground truth M_{l-1}, M_l, and M_{l+1}, respectively. Note that the feature maps in each layer can be directly fused with the final prediction and optimized without massive convolutional processing. Therefore, deep pyramid supervision can be adapted to different depths for different tasks in training, which supply adaptive model capacity, thereby facilitating the segmentation of objects with different scales.

3.4. Pyramid Skip Connections

Pyramid skip connections perform feature reuse among the three scaled feature maps (the higher scale, the current scale, and the lower scale) in each IPAB module. Suppose that X_l is the input of the current layer in the decoder, and X_l^p, and X_l^d are the results at the higher scale and the lower scale, respectively. Meanwhile, $(\tilde{X}_l^p, \tilde{X}_{l+1}, \tilde{X}_{l+2}^d)$, $(\tilde{X}_{l-1}^p, \tilde{X}_l, \tilde{X}_{l+1}^d)$, and $(\tilde{X}_{l-2}^p, \tilde{X}_{l-1}, \tilde{X}_l^d)$ are three groups of learned feature maps from the encoder, and feature maps in each group have the same spatial dimension with the corresponding scaled input $\hat{X}_{l-1}$, $\hat{X}_l$, and $\hat{X}_{l+1}$, respectively. The fusion process of the current scale is formulated as follows,

$$\hat{X}_{l-1} = H(X_l^d, H(\tilde{X}_l^p, \tilde{X}_{l+1}, \tilde{X}_{l+2}^d)), \tag{9}$$

$$\hat{X}_l = H(X_l, H(\tilde{X}^p_{l-1}, \tilde{X}_l, \tilde{X}^d_{l+1})), \quad (10)$$

$$\hat{X}_{l+1} = H(X^p_l, H(\tilde{X}^d_l, \tilde{X}_{l-1}, \tilde{X}^p_{l-2})), \quad (11)$$

where $H(\cdot)$ denotes channel adjustment. We can see that features at the current-scale l can reuse and aggregate feature maps at most five scales ($l-2, l-1, l, l+1,$ and$l+2$).

4. EXPERIMENTS

4.1. Datasets

We used three public available retinal vessel datasets, DRIVE (20), STARE (21), and CHASE-DB1 (22) for evaluation. The images in the three datasets are collected using digital retinal imaging, a standard method of documenting the appearance of the retina. More details of the datasets are as follows.

DRIVE: The DRIVE dataset (20) consists of 40 images with a resolution of 565 × 584 pixels, which were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45-degree field of view (FOV). Two trained human observers labeled the vessels in all images, and the ones from the first observer were used for network training. The dataset has been divided into a training and a test set (20), both of which contain 20 images.

CHASE-DB1: The CHASE-DB1 dataset (22) contains vascular patch images with a resolution of 999 × 960, which were acquired from 28 eyes of 14 ten-year-old children. Since images were captured in subdued lighting and the operators adjusted illumination settings, the images contain more illumination variation in CHASE-DB1 compared with the DRIVE datasets. Following the configuration in Li et al. (46), the first 20 images and the remaining 8 images are employed as the training set and the test set, respectively.

STARE: The STARE dataset (21) consists of 20 equal-sized images with a resolution of 700 × 605 pixels. Each image is with a 35° FOV, and half of the images of eyes are with ocular pathology. As the training set and the test set are not explicitly specified, the same leave-one-out cross-validation is adopted (33) for performance evaluation, where models are iteratively trained on 19 images and tested on the rest images. Liking other methods (10), manual annotations generated by the first observer are used for both training and test.

4.2. Implementations

All experiments were conducted on an Nvidia GeForce Titan X (pascal) containing 12 GB memory. Meanwhile, we employed CE-Net (32), one of the state-of-the-art methods in retinal vessel segmentation, as the backbone models to implement IPABs, pyramid input enhancement, deep pyramid supervision, and pyramid skip connections. Normalization of the training data has been implemented. In order to express the details of multi-scale feature fusion more clearly, we use U-Net as the basic network to explain, which is widely used in the medical image segmentation domain. In practice, we use the state-of-the-art method CE-Net to replace U-Net to obtain better performance. During training, we adopted Adaptive Moment Estimation (Adam) as the learning optimizer with a batch size of 4. Data augmentation operations including horizontal flip, vertical flip, and diagonal flip are used to enlarge the train samples. We use a threshold to obtain the final segmentation from pixel probability vectors. Particularly, the pixels with values smaller than the threshold are assigned to the background class, and the remaining pixels with values equal to or greater than the threshold are categorized as the vessel class. The final prediction is the ensemble of the segmentation output of the vessel images, its rotation (90°), and its flip (horizontal and vertical).

TABLE 1 | Performance comparison of Pyramid-Net and the state-of-the-art methods on the DRIVE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
FCN (31)	74.89	96.21	94.13	95.67
U-Net (23)	75.31	96.45	94.45	96.01
DeepVessel (11)	76.12	97.68	95.23	97.52
(10)	76.53	98.18	95.42	97.52
(47)	77.92	98.13	95.56	97.84
(40)	78.44	98.07	95.67	98.19
CE-Net (32)	83.09	97.47	95.45	97.79
BTS-DSN (48)	78.91	98.04	95.61	98.06
(49)	79.16	98.11	95.70	98.10
(50)	79.40	98.16	95.67	97.72
Vessel-Net (51)	80.38	98.02	95.78	98.21
MResU-Net (52)	79.69	97.99	-	97.99
CTF-Net (53)	78.49	98.13	95.67	97.88
Hybrid-Net (6)	**83.53**	97.51	95.79	-
HA-Net (33)	79.91	98.13	95.81	98.23
Pyramid-Net	82.38	**98.19**	**96.26**	**98.32**

Bold values mean the state-of-the-art performance.

4.3. Evaluation Metrics

We introduce four evaluation metrics including Sensitivity (Sens), Specificity (Spec), Accuracy (Acc), and Area Under the ROC Curve (AUC) to validate our proposed Pyramid-Net. The metrics are calculated as follows:

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \quad (12)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}), \quad (13)$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}). \quad (14)$$

True positive (TP) and true negative (TN) present that pixels are correctly classified to objects or backgrounds, respectively. Meanwhile, pixels will be labeled as false positive (FP) or false negative (FN), if they are misclassified to objects or backgrounds, respectively.

4.4. Quantitative Results

We compared our Pyramid-Net with existing state-of-the-art works on three vessel image segmentation datasets (DRIVE, CHASE-DB1, and STARE). **Tables 1–3** illustrate the comparison results of Pyramid-Net and the current state-of-the-art methods.

TABLE 2 | Performance comparison of Pyramid-Net and the state-of-the-art methods on the CHASE-DB1 dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
(54)	76.15	95.75	94.67	96.23
(46)	75.07	97.93	95.81	97.16
(55)	81.94	97.39	96.30	-
(10)	76.33	98.09	96.10	97.81
(47)	77.56	98.20	96.34	98.15
FCN (31)	76.41	98.06	96.07	97.76
(56)	81.55	97.52	96.10	98.04
(48)	78.88	98.01	96.27	98.40
(50)	80.74	98.21	96.61	98.12
(51)	81.32	98.14	96.61	98.60
Three-stage (12)	76.41	98.06	96.07	97.76
CTF-Net (52)	79.48	**98.42**	96.48	98.47
Hybrid-Net (6)	81.76	97.76	96.32	-
HA-Net (33)	**82.39**	98.13	96.70	98.70
Pyramid-Net	81.17	98.26	**96.89**	**98.92**

Bold values mean the state-of-the-art performance.

TABLE 3 | Performance comparison of Pyramid-Net and the state-of-the-art methods on the STARE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
(54)	73.20	98.40	95.60	96.70
(57)	77.91	97.58	95.54	97.48
(58)	76.80	97.38	-	-
(10)	75.81	98.46	96.12	98.01
(56)	75.95	98.78	96.41	98.32
Three-stage (12)	77.35	98.57	96.38	98.33
MResU-Net (52)	81.01	97.95	-	98.16
Hybrid-Net (6)	79.46	98.21	96.26	-
HA-Net (33)	81.86	98.44	96.73	98.32
Pyramid-Net	**82.35**	**98.87**	**97.19**	**98.62**

Bold values mean the state-of-the-art performance.

For the DRIVE dataset, Pyramid-Net achieves a high score of 82.38, 98.19, 96.26, and 98.32% on Sens, Spec, Acc, and AUC, respectively, and outperforms state-of-the-art methods in three metrics including Spec, Acc, and AUC. In terms of Sens, CE-Net achieves the best performance of 83.09%, while our method achieves a comparable result, which is 0.71% lower. Overall, Pyramid-Net achieves higher overall performance than CE-Net. For the CHASE-DB1 dataset, compared with the state-of-the-art results, the proposed Pyramid-Net achieves high score of 81.17, 98.26, 96.89, and 98.92% for Sens, Spec, Acc, and AUC, respectively, which consistently enjoys a better performance than all the current state-of-the-art methods. For the STARE dataset, Pyramid-Net achieves a promising score of 82.35, 98.87, 97.19, and 98.62% for Sens, Spec, Acc, and AUC, respectively, which is also consistently better than all the current state-of-the-art methods. The consistent improvements in **Tables 1**–**3** indicate the effectiveness and robustness of our Pyramid-Net.

4.5. Qualitative Results

The visual comparisons between Pyramid-Net and the state-of-the-art methods, including DeepVessel and CE-Net on the DRIVE dataset and the CHASE-DB1 dataset are shown in **Figure 5**. White (TP) and black (TN) pixels are correct predictions of vessels and the background, respectively, while red (FP) and green (FN) pixels are incorrect predictions. In **Figure 5**, dark yellow rectangles contain the selected areas used for detail comparison, and the bright yellow rectangles contain the zoomed area in the dark yellow rectangle. We can notice that current methods enjoy a good performance on the segmentation of main retinal vessels, but the effect on some capillaries is poor. For example, Row 1 of **Figure 5** shows that the result of DeepVessel misses a large number of thin vessels on the DRIVE dataset, and that of CE-Net obtains a much better accuracy on thin vessels. However, in Row 2, there is no significant difference between the results of the two methods. In both Rows 1 and 2 of **Figure 5**, our method can achieve much higher accuracy, but we can still notice that our method cannot segment them correctly if the vessels are too thin. We can further observe that our method has much fewer false-negative pixels (indicated by green) than the other two. This may due to the fact that our proposed IPABs can consider more scales thus improving the segmentation accuracy. Overall, our proposed Pyramid-Net evidently improves the segmentation performance, especially for those narrow, low-contrast, and ambiguous retinal vessels.

4.6. Evaluation on Thin Vessels

In the previous subsection, the results in **Figure 5** indicate that though the main vessels enjoy a promising segmentation performance, the segmentation of thin vessels always suffers a big miss in the prediction. In practice, it is challenging to segment the thin vessels from the complex retina background, which are always low-contrast and extremely narrow (1–2 pixels). Thus, in this subsection, to evaluate the effectiveness of Pyramid-Net on thin vessels, we compared Pyramid-Net with the state-of-the-art methods on an additional dataset only containing thin vessel labels. Vessels with a width of 1 or 2 pixels are commonly regarded as the thin vessels in the DRIVE dataset. To avoid potential unfair in the evaluation on the manual addition label of the thin vessel, we distinguish thick vessels from thin vessels by an opening operation (10). The evaluation results are summarized in **Table 4**. It can be noticed that Pyramid-Net achieves a high ACC score of 96.26, 96.51, and 91.64% on all vessels, thick vessels, and thin vessels, respectively. Overall, our method outperforms the state-of-the-art methods on all metrics. As for the thin vessel segmentation, our methods achieve an improvement of 4.73% over backbone model CE-Net and outperforms the state-of-the-art method by about 3.86%. The experiment results indicate that our Pyramid-Net is particularly effective on thin vessels.

4.7. Ablation Analysis

To justify the effectiveness of IPABs, pyramid input enhancement, deep pyramid supervision, and pyramid skip connections in the proposed Pyramid-Net, we conduct ablation analysis using the DRIVE dataset as a vehicle. The ablation experimental results are summarized in **Table 5**. We use CE-Net

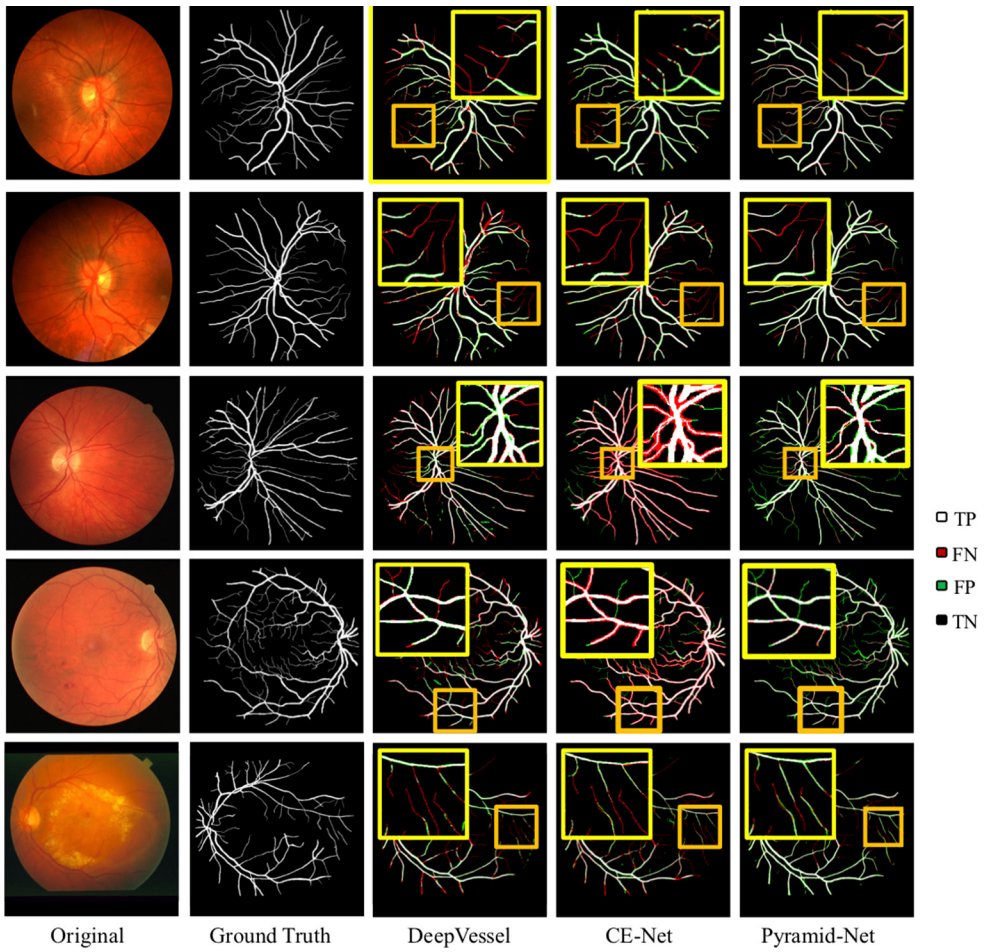

FIGURE 5 | Visual comparison of Pyramid-Net and the state-of-the-art methods including DeepVessel (11) and CE-Net (32) on DRIVE (Row 1–2), CHASE-DB1 (Row 3–4), and STARE (Row 5) datasets. White (TP) and black (TN) pixels indicate correct predictions of object and background, respectively, while red (FP) and green (FN) pixels indicate incorrect predictions. The dark yellow rectangle contains the area used to compare segmentation details, and the bright yellow rectangle contains the zoomed area in the dark yellow rectangle. Best viewed in color.

(32) as our backbone, which achieves a good score of 95.45 and 97.79% on Acc and on AUC, respectively. Firstly, we evaluate the effectiveness of IPABs on the backbone. Benefiting from aggregating coarse-to-fine context information from pyramid scale in each layer, the backbone model with IPABs achieves improvements of 0.62% on Acc and 0.30% on AUC. Second, we evaluate pyramid input enhancement and deep pyramid supervision to feed the original image at multiple scales into the network and supervise the immediate layers contains features at various scales. In **Table 5**, we can notice that the above two optimizations achieve improvements of more than 0.10 and 0.07% in AUC, respectively. Third, pyramid skip connections connect the encoder and the decoder and make full use of the features from multiple layers and scales in the encoder, which achieves an improvement of about 0.15% on AUC. Overall, integrating the pyramid-scale concept into the design of the basic unit and skip connections can obviously improve the network segmentation, and the other two optimizations also bring some improvement.

TABLE 4 | Performance comparison on thick and thin vessels of Pyramid-Net on the DRIVE dataset.

Method	All vessel (%)	Thick vessel (%)	Thin vessel (%)
(10)	95.42	95.78	87.78
CE-Net (32)	95.45	95.96	86.91
Pyramid-Net	**96.26**	**96.51**	**91.64**

Bold values mean the state-of-the-art performance.

4.8. Cross-Training Evaluation

To evaluate the generalization of Pyramid-Net, we performed a cross-training evaluation on the DRIVE dataset and the STARE dataset. We directly implemented our models trained on the source dataset and tested on the target dataset for fair comparisons. The experimental results are summarized in **Table 6**. Overall, our method achieves the state-of-the-art transfer performance on both configurations. Particularly, for the configuration that models are trained on the STARE dataset

TABLE 5 | Ablation analysis of Pyramid-Net on the DRIVE dataset.

Method	Acc (%)	AUC (%)
Baseline	95.45	97.79
Baseline + IPABs	96.07	98.09
Baseline + IPABs + pyramid input	96.10	98.15
Baseline + IPABs + Pyramid supervision	96.15	98.12
Baseline + IPABs + pyramid skip connection	96.21	98.24
Pyramid-Net	**96.26**	**98.32**

Bold values mean the state-of-the-art performance.

TABLE 6 | Cross-training evaluation on the DRIVE dataset and the STARE dataset.

Method	Sens (%)	Spec (%)	Acc (%)	AUC (%)
DRIVE (train) -> STARE (test)				
(12)	70.14	98.02	94.44	95.68
(56)	65.05	99.14	94.81	97.18
HA-Net (33)	71.40	98.79	95.30	97.58
Pyramid-Net	**75.71**	**98.86**	**95.57**	**97.78**
STARE (train) -> DRIVE (test)				
(12)	73.19	98.40	95.80	96.78
(56)	70.00	97.59	94.74	97.18
HA-Net (33)	81.87	**98.79**	95.30	97.58
Pyramid-Net	**82.67**	98.76	**95.36**	**97.72**

Bold values mean the state-of-the-art performance.

and tested on the DRIVE dataset, it can be noticed that the transfer model can achieve competitive results on Spec and suffer a big loss of accuracy on Sens. The potential reason is the imbalance between thick vessels and thin vessels in the STARE dataset. Manual annotations of the STARE dataset contain more thick vessels than thin vessels, which led that the pre-trained model on the STARE dataset obtains a bad segmentation performance of thin vessels on the DRIVE dataset. When the conditions are reversed, the above situation is alleviated, and the corresponding scores on Sens, Spec, Acc, and AUC on the STARE dataset are comparable with the model trained on the STARE dataset.

4.9. Comparison With Multi-Scale Aggregation Methods

To evaluate the effectiveness of the multi-scale information aggregated in the proposed Pyramid-Net, we compare existing multi-scale aggregation methods, including Dense Pooling Connections (15), Complete Bipartite Network (CB-Net) (16), Dense Decoder Short Connections (DDSC) (18), and U-Net++ (17) on the DRIVE dataset. For fair comparisons, we directly implement those different connection styles and our Pyramid-Net on U-Net (23). The comparison results and the *p*-values for the paired *t*-test are summarized in **Table 7**. Compared with existing methods, our method outperforms them by 0.65–0.99% and 0.67–1.50% on Acc and AUC, respectively. On the other hand, we also compare the computational cost of the proposed Pyramid-Net with existing methods. Obviously, existing methods improve the network performance and increase the computational cost by 16.38–493.74G (104.9–247.4%) on FLOPs from the numerous feature reuse. Particularly, our proposed Pyramid-Net achieves state-of-the-art performance with a computational cost reduced by 216.8G (64.7%) on FLOPs. The reason for the above phenomenon is the channel reduction in each IPAB. The channels' main branch is reduced to half, while the number of channels at associated branches is half of that of the main branch. Overall, our method achieves the state-of-the-art performance of 96.26% on Acc and 98.32% on AUC with a 64.7% reduction on FLOPs.

TABLE 7 | Comparison with existing multi-scale aggregation methods on the DRIVE Dataset.

Method	Acc (%)	AUC (%)	FLOPs	*p*-values
U-Net (23)	94.45	96.01	334.95G	<0.01
DPC (15)	95.56	97.65	351.33G	<0.01
CB-Net (16)	95.61	97.52	441.62G	<0.01
DDSC (18)	95.42	97.48	381.07G	<0.01
U-Net ++ (17)	95.27	96.82	828.69G	<0.01
CE-Net (32)	95.45	97.79	-	<0.05
Pyramid-Net	**96.26**	**98.32**	188.15G	-

Bold values mean the state-of-the-art performance.

5. CONCLUSION

In this paper, we introduced Pyramid-Net for accurate retinal vessel segmentation. In Pyramid-Net, the proposed IPABs are utilized to generalize two associated branches to aggregate coarse-to-fine feature maps at pyramid scales to improve the segmentation performance. Meanwhile, three optimizations including pyramid inputs enhancement, deep pyramid supervision, and pyramid skip connections are implemented with IPABs in the encoder, the decoder, and the cross of the two to further improve performance, respectively. Comprehensive experiments have been conducted on three retinal vessel segmentation datasets, including DRIVE (20), STARE (21), and CHASE-DB1 (22). Experimental results demonstrate that our IPABs can efficiently improve the segmentation performance, especially for thin vessels. In addition, our method is also much more efficient than existing methods with a large reduction in computational cost.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors/s.

AUTHOR CONTRIBUTIONS

XX is the guarantor of the manuscript. JZh implemented the experiments and wrote the first draft of the manuscript. HQ, WX, and ZY managed the result analysis. All authors contributed to drawing up the manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (no. 2018YFC1002600), the Science and Technology Planning Project of Guangdong Province, China (nos. 2017B090904034, 2017B030314109, 2018B090944002, and 2019B020230003), Guangdong Peak Project (no. DFJH201802), and the National Natural Science Foundation of China (no. 62006050).

REFERENCES

1. Winder RJ, Morrow PJ, McRitchie IN, Bailie J, Hart PM. Algorithms for digital image processing in diabetic retinopathy. *Comput Med Imaging Graph*. (2009) 33:608–22. doi: 10.1016/j.compmedimag.2009.06.003
2. Mitchell P, Leung H, Wang JJ, Rochtchina E, Lee AJ, Wong TY, et al. Retinal vessel diameter and open-angle glaucoma: the blue mountains eye study. *Ophthalmology*. (2005) 112:245–50. doi: 10.1016/j.ophtha.2004.08.015
3. Yannuzzi LA, Negr ao S, Tomohiro I, Carvalho C, Rodriguez-Coleman H, Slakter J, et al. Retinal angiomatous proliferation in age-related macular degeneration. *Retina*. (2012) 32:416–34. doi: 10.1097/IAE.0b013e31823f9b3b
4. Ikram MK, Witteman JC, Vingerling JR, Breteler MM, Hofman A, de Jong PT. Retinal vessel diameters and risk of hypertension: the Rotterdam Study. *Hypertension*. (2006) 47:189–94. doi: 10.1161/01.HYP.0000199104.61945.33
5. Gishti O, Jaddoe VW, Felix JF, Klaver CC, Hofman A, Wong TY, et al. Retinal microvasculature and cardiovascular health in childhood. *Pediatrics*. (2015) 135:678–85. doi: 10.1542/peds.2014-3341
6. Yang L, Wang H, Zeng Q, Liu Y, Bian G. A hybrid deep segmentation network for fundus vessels via deep-learning framework. *Neurocomputing*. (2021) 448:168–78. doi: 10.1016/j.neucom.2021.03.085
7. Guo S, Li T, Kang H, Li N, Zhang Y, Wang K. L-Seg: an end-to-end unified framework for multi-lesion segmentation of fundus images. *Neurocomputing*. (2019) 349:52–63. doi: 10.1016/j.neucom.2019.04.019
8. Cheung CYl, Zheng Y, Hsu W, Lee ML, Lau QP, Mitchell P, et al. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology*. (2011) 118:812–8. doi: 10.1016/j.ophtha.2010.08.045
9. Xu X, Lu Q, Yang L, Hu S, Chen D, Hu Y, et al. Quantization of fully convolutional networks for accurate biomedical image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake city, UT: IEEE (2018). p. 8300–8.
10. Yan Z, Yang X, Cheng KT. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng*. (2018) 65:1912–23. doi: 10.1109/TBME.2018.2828137
11. Fu H, Xu Y, Lin S, Wong DWK, Liu J. Deepvessel: retinal vessel segmentation via deep learning and conditional random field. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens: Springer (2016). p. 132–9.
12. Yan Z, Yang X, Cheng KT. A three-stage deep learning model for accurate retinal vessel segmentation. *Biomed Health Inf IEEE J*. (2019) 23:1427–36. doi: 10.1109/JBHI.2018.2872813
13. Katz N, Goldbaum M, Nelson M, Chaudhuri S. An image processing system for automatic retina diagnosis. In: *Three-Dimensional Imaging and Remote Sensing Imaging. Vol. 902*. Los Angeles, CA: International Society for Optics and Photonics. (1988). p. 131–7.
14. Spencer T, Olson JA, McHardy KC, Sharp PF, Forrester JV. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput Biomed Res*. (1996) 29:284–302. doi: 10.1006/cbmr.1996.0021
15. Playout C, Duval R, Cheriet F. A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 101–8.
16. Chen J, Banerjee S, Grama A, Scheirer WJ, Chen DZ. Neuron segmentation using deep complete bipartite networks. *In: International Conference on Medical Image Computing and Computer-Assisted Intervention*. Quebec, QC: Springer (2017). p. 21–9. doi: 10.1007/978-3-319-66185-8_3
17. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada: Springer (2018). p. 3–11.
18. Bilinski P, Prisacariu V. Dense decoder shortcut connections for single-pass semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Granada (2018) p. 6596–605.
19. Ding H, Pan Z, Cen Q, Li Y, Chen S. Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing*. (2020) 380:150–61. doi: 10.1016/j.neucom.2019.10.097
20. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging*. (2004) 23:501–9. doi: 10.1109/TMI.2004.825627
21. Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging*. (2000) 19:203–10. doi: 10.1109/42.845178
22. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng*. (2012) 59:2538–48. doi: 10.1109/TBME.2012.2205687
23. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Munich: Springer International Publishing (2015).
24. Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD. Comparative study of retinal vessel segmentation methods on a new publicly available database. In: *Medical imaging 2004: image processing. Vol. 5370. International Society for Optics and Photonics*. San Diego, CA (2004). p. 648–56.
25. Sinthanayothin C, Boyce JF, Cook HL, Williamson TH. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol*. (1999) 83:902–10. doi: 10.1136/bjo.83.8.902
26. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging*. (2006) 25:1214–22. doi: 10.1109/TMI.2006.879967
27. Rangayyan RM, Ayres FJ, Oloumi F, Oloumi F, Eshghzadeh-Zanjani P. Detection of blood vessels in the retina with multiscale Gabor filters. *J Electron Imaging*. (2008) 17:023018. doi: 10.1117/1.2907209
28. Ricci E, Perfetti R. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging*. (2007) 26:1357–1365. doi: 10.1109/TMI.2007.898551
29. Franklin SW, Rajan SE. Retinal vessel segmentation employing ANN technique by Gabor and moment invariants-based features. *Appl Soft Comput*. (2014) 22:94–100. doi: 10.1016/j.asoc.2014.04.024
30. Zhang J, Chen Y, Bekkers E, Wang M, Dashtbozorg B, ter Haar Romeny BM. Retinal vessel delineation using a brain-inspired wavelet transform and random forest. *Pattern Recognit*. (2017) 69:107–23. doi: 10.1016/j.patcog.2017.04.008
31. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA: IEEE (2015). p. 3431–40.
32. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging*. (2019) 38: 2281–2292. doi: 10.1109/TMI.2019.2903562
33. Wang D, Haytham A, Pottenburgh J, Saeedi OJ, Tao Y. Hard attention net for automatic retinal vessel segmentation. *IEEE J Biomed Health Inf*. (2020) 24:3384–96. doi: 10.1109/JBHI.2020.3002985
34. Guo C, Szemenyei M, Yi Y, Zhou W, Bian H. Residual spatial attention network for retinal vessel segmentation. In: *International Conference on Neural Information Processing*. San Diego, CA: Springer (2020). p. 509–19.

35. Zhang J, Zhang Y, Xu X. Pyramid U-net for retinal vessel segmentation. In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, ON: IEEE (2021). p. 1125–9.
36. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: *Deep Learning and Data Labeling for Medical Applications*. Athens: Springer (2016). p. 179–87.
37. Zhang J, Jin Y, Xu J, Xu X, Zhang Y. Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. *arXiv preprint* arXiv:181200352. (2018).
38. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017). p. 6230–9.
39. Chen LC, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: Scale-aware semantic image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Las Vegas, NV: IEEE (2016). p. 3640–9.
40. Wu Y, Xia Y, Song Y, Zhang Y, Cai W. Multiscale network followed network model for retinal vessel segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 119–26.
41. Raza SEA, Cheung L, Epstein D, Pelengaris S, Khan M, Rajpoot NM. MIMO-Net: a multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* Melbourne, VIC: IEEE (2017). p. 337–40.
42. Graham S, Chen H, Gamper J, Dou Q, Heng PA, Snead D, et al. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Med Image Anal*. (2019) 52:199–211. doi: 10.1016/j.media.2018.12.001
43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Las Vegas, NV: IEEE (2016). p. 770–8.
44. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Salt Lake City, UT: IEEE (2018). p. 7132–41.
45. Li S, Zhang J, Ruan C, Zhang Y. Multi-stage attention-unet for wireless capsule endoscopy image bleeding area segmentation. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. San Diego, CA: IEEE (2019). p. 818–25.
46. Li Q, Feng B, Xie L, Liang P, Zhang H, Wang T. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans Med Imaging*. (2015) 35:109–18. doi: 10.1109/TMI.2015.2457891
47. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv[Preprint]*.arXiv:180206955. (2018) doi: 10.1109/NAECON.2018.8556686
48. Guo S, Wang K, Kang H, Zhang Y, Gao Y, Li T. BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. *Int J Med Inform*. (2019) 126:105–13. doi: 10.1016/j.ijmedinf.2019.03.015
49. Ma W, Yu S, Ma K, Wang J, Ding X, Zheng Y. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 769–78.
50. Wang B, Qiu S, He H. Dual encoding u-net for retinal vessel segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 84–92.
51. Wu Y, Xia Y, Song Y, Zhang D, Liu D, Zhang C, et al. Vessel-Net: retinal vessel segmentation under multi-path supervision. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen: Springer (2019). p. 264–72.
52. Li D, Dharmawan DA, Ng BP, Rahardja S. Residual u-net for retinal vessel segmentation. In: *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei: IEEE (2019). p. 1425–9.
53. Wang K, Zhang X, Huang S, Wang Q, Chen F. Ctf-net: retinal vessel segmentation via deep coarse-to-fine supervision network. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Iowa City, IA: IEEE (2020). p. 1237–41.
54. Roychowdhury S, Koozekanani DD, Parhi KK. Iterative vessel segmentation of fundus images. *IEEE Trans Biomed Eng*. (2015) 62:1738–49. doi: 10.1109/TBME.2015.2403295
55. Kassim YM, Palaniappan K. Extracting retinal vascular networks using deep learning architecture. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Kansas City, MO: IEEE (2017). p. 1170–4.
56. Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R. DUNet: A deformable network for retinal vessel segmentation. *Knowl Based Syst*. (2019) 178:149–62. doi: 10.1016/j.knosys.2019.04.025
57. Zhang J, Dashtbozorg B, Bekkers E, Pluim JP, Duits R, ter Haar Romeny BM. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Trans Med Imaging*. (2016) 35:2631–2644. doi: 10.1109/TMI.2016.2587062
58. Orlando JI, Prokofyeva E, Blaschko MB. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans Biomed Eng*. (2016) 64:16–27. doi: 10.1109/TBME.2016.2535311

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.