IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 30, 2021, accepted July 20, 2021, date of publication August 3, 2021, date of current version August 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3102176

Retinal Vessel Segmentation Using Deep Learning: A Review

CHUNHUI CHEN[1], JOON HUANG CHUAH[1], (Senior Member, IEEE), RAZA ALI[1,2], (Member, IEEE), AND YIZHOU WANG[3], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Faculty of Information and Communication Technology, BUITEMS, Quetta 87300, Pakistan
[3]Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China

Corresponding author: Joon Huang Chuah (jhchuah@um.edu.my)

This work was supported by the University of Malaya Faculty Research Grant under Grant GPF009A-2018.

ABSTRACT This paper presents a comprehensive review of retinal blood vessel segmentation based on deep learning. The geometric characteristics of retinal vessels reflect the health status of patients and help to diagnose some diseases such as diabetes and hypertension. The accurate diagnosis and timing treatment of these diseases can prevent global blindness of patients. Recently, deep learning algorithms have been rapidly applied to retinal vessel segmentation due to their higher efficiency and accuracy, when compared with manual segmentation and other computer-aided diagnosis techniques. In this work, we reviewed recent publications for retinal vessel segmentation based on deep learning. We surveyed these proposed methods especially the network architectures and figured out the trend of models. We summarized obstacles and key aspects for applying deep learning to retinal vessel segmentation and indicated future research directions. This article will help researchers to construct more advanced and robust models.

INDEX TERMS Retinal vessel segmentation, fundus images, deep learning, convolutional neural network.

I. INTRODUCTION

The fundus retina image is the only deeper microvascular system that can be observed non-invasively. The retinal vessel map contains abundant geometric characteristics, such as vessel diameter, branch angles, and branch lengths. These geometric characteristics reflect clinical and pathological features, which are used to diagnose hypertension, diabetes, and atherosclerosis [1]–[4]. The ophthalmologist uses retinal blood vessels to diagnose vascular and vascular system lesions related diseases, which interprets diabetic retinopathy (DR) and diabetic maculopathy (MD). These are the leading causes of global blindness. Retinal image assessment has been an indispensable step for the identification of retinal pathology.

Retinal fundus image illustrates retina structure, such as retinal blood vessel tree, optic disk (OD), fovea, macula, and abnormal structures, as shown in Figure 1. The retinal blood vessel tree is composed of the central retinal artery, vein, and branches. An abnormality may include microaneurysms (MAs), haemorrhages, exudates and cotton wool spots [5].

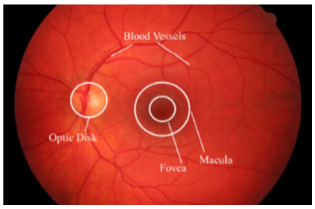

FIGURE 1. Annotated structure of retina [6].

Precise identification and diagnosis of eye abnormalities and their timely medication are vital in preventing blindness. Initially, trained experts would manually segment the retinal blood vessels, but that was an expensive, tedious and time-consuming process [7]. Moreover, the complexities of the image cause inconsistency of vessel map segmented by different experts [8], due to the lower contrast between vessels and backgrounds, uneven illumination, various abnormalities and variation in vessel width/shape. These facts inspire the development of automatic retinal vessel segmentation with minimal human interference.

Several supervised and unsupervised methods are developed and used to automate the segmentation of retinal

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojie Ju.

vessels. Earlier, unsupervised methods are the most common approach for automatically segmenting the retinal vessels, which do not rely on any annotation for segmentation [9], [10]. These methods are roughly divided into matching filter [11]–[13], vascular tracing based segmentation [14]–[16] and model-based segmentation methods [17]. Unsupervised methods show some defects in their performance because they cannot benefit from the hand-labelled ground truth.

Unlike unsupervised methods, supervised models are trained by using annotations and can benefit from the ground truth. Supervised models conduct retinal vessel segmentation in two stages: feature extraction and pixels classification. Features can be further divided into handcrafted features or automatically learned features. In machine learning, the process of feature extraction from fundus images is manual, and some typical classifiers are adopted, such as k-nearest neighbour classifier (KNN) [18] and support vector machine (SVM) [19]. Manual feature selection can leverage domain-knowledge well, but it also lacks generalization ability since it is application-specific and cannot learn new features automatically [20].

Deep learning, especially convolutional neural networks (CNNs), has gained much attention for image analysis [21], [22]. Deep learning methods learn features automatically by using massive data with less human inference. They have better generalization ability and recognition capability because they can learn different level patterns automatically and will not be limited by a specific application. In 2012, Krizhevsky, *et al.* [23] proposed AlexNet for image recognition. For image segmentation and identification, VGGNet [24] and GoogleNet [25] were introduced. Long, *et al.* [26] proposed fully convolutional networks (FCN) for image semantic segmentation, which made dense predictions in a sliding window fashion and thus speeded up the segmentation.

Several review articles on retinal blood vessel segmentation have been published [10], [27]–[30]. However, Mookiah, *et al.* [10], Khan, *et al.* [28], Badar, *et al.* [29], Li, *et al.* [30] did not focus on deep learning methods for vessel segmentation, whereas the techniques discussed in Soomro, *et al.* [27] are published several years ago. Therefore, in this review article, we discussed publications of recent six years for retinal blood vessel segmentation based on deep learning.

All the papers were retrieved by conducting iterative and exhaustive searches in IEEE Xplore, Springer Link and ScienceDirect databases.

We applied the following search command to index terms of both journal papers and conference papers: ''fundus image'' AND (''retinal vessel'' OR ''retinal blood vessel'') AND (''segmentation'' OR ''extraction'' OR ''detection'' OR ''identification'') AND (''deep learning'' OR ''convolutional neural network'' OR ''CNN'' OR ''fully convolutional network'' OR ''FCN'' OR ''generative adversarial network'' OR ''GAN''). We only selected original studies from 2016 that formulated retinal vessel segmentation as the main task instead of the intermediate task.

This article is organized as follows. In section II, we discuss deep learning and convolutional neural networks. In section III, we introduce the datasets used for retinal vessel segmentation and performance evaluation metrics for proposed models. In section IV, we analyze the existing models for retinal segmentation based on deep learning. In section V, we discuss retinal vessel segmentation according to the analysis of existing models. Section VI concludes the article and points out future research directions.

II. OVERVIEW OF DEEP LEARNING

Deep learning models are composed of hierarchically structured layers which translate input information into a meaningful output. Deep learning has been developed a rich family since 1990 [31], such as deep neural networks (DNNs) [31], auto-encoders (AEs) [32] and stacked auto-encoders (SAEs) [33] neural networks, deep belief network (DBNs) [34], [35], restricted Boltzmann machines (RBMs) [36], convolution neural networks (CNNs) [37], recurrent neural networks (RNNs) [38], [39], generative adversarial networks (GANs) [40] and graphic neural networks (GNNs) [41].

In this section, we have discussed the most widely used CNNs architectures for image computer vision tasks.

A. CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Convolutional Neural Networks (CNNs) are inspired by multi-layered perceptrons (MLPs) and are widely used for image processing such as classification, segmentation, and localization. Hubel and Wiesel [42] conducted a first experiment based on CNN, indicated that cells in the cat's visual cortex were responsible for detecting light in corresponding receptive fields. LeCun, *et al.* [37] proposed another CNN based network that recognized the handwritten digits. The network was composed of convolution operation and pooling operation, trained by the back-propagation algorithm. Later, LeCun, *et al.* [43] proposed the LeNet-5 for document recognition. However, these architectures were not widely used due to the lack of training data and computation power at that time. Krizhevsky, *et al.* [23] proposed a powerful deep CNN for image classification, called AlexNet. The model had significant improvement and outperformed all existing methods, also won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [44]. AlexNet architecture is deeper than LeNet-5 and utilizes the ReLU activation function. Figure 2 and Figure 3 show the architecture of LeNet-5 and AlexNet, respectively.

Encouraged by AlexNet, a lot of research has been done based on CNN architectures. Several applications based on deeper architectures were proposed to improve the performance. VGG Net [24] was the first to explore much deeper networks, which stacked small, fixed sized kernels in each convolution layer. Simonyan and Zisserman [24] proposed deeper CNNs with different numbers of convolu-

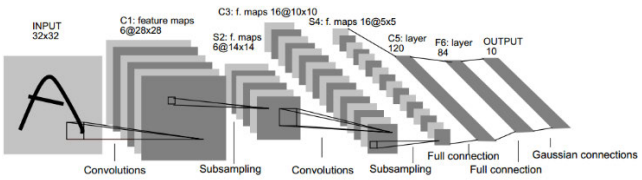

FIGURE 2. Architecture of the LeNet-5 [43]. C: convolution layer, S: sub-sampling layer, F: fully connected layer.

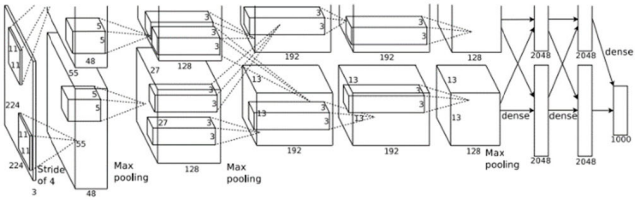

FIGURE 3. Architecture of AlexNet [23].

tion layers, such as 13, 16 and 19. Finally, VGG19 with 19 convolutional layers won the ImageNet challenge of 2014. Szegedy, *et al.* [25] introduced GoogleNet which contains 22 layers and adopts the Inception module [45].

1) CNN ARCHITECTURE COMPONENTS

CNN architectures are composed of hierarchically structured layers with optimized parameters. This section will interpret the main components of a CNN.

a: CONVOLUTIONAL LAYER

Convolutional layer is the main layer in CNNs that extracts features from input data or feature maps. The convolutional layer contains several stacked convolution kernels to conduct convolution operations. In the convolution operation (see Figure 4), a convolution kernel slides from left to right and from up to down, and it multiplies with a specific region of input or feature map elementwise to produce a value, which is known as feature extraction. The specific region is called the receptive field. These special regions share kernels, known as weight sharing which reduces the complexity of the model and makes the training process easier. Mathematically, the feature map z generated by convolution kernel can be expressed as:

$$z = W * x + b \tag{1}$$

where x is the input image, W is the convolution kernel, while b is the bias for the convolution layer.

b: BATCH NORMALIZATION

The input or feature maps generated by convolutional layers may vary greatly, so for large or small values sent to the activation function they face a problem of vanishing/exploding gradients, which hamper the training process [38], [46]. To address this problem, batch normalization [47] is proposed to accelerate the training process, which scales the input of

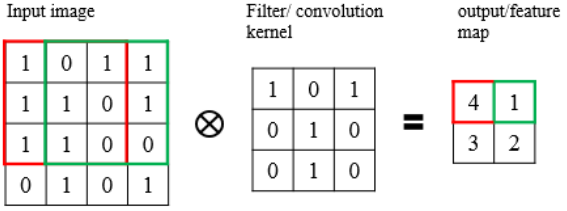

FIGURE 4. Convolution operation. Stride=1 and assume bias=0.

activation function and reduces internal covariate shift by applying normalization operation to each mini-batch. Generally, batch normalization is performed before the activation function, but the function can also be used after the activation function based on application.

c: ACTIVATION FUNCTION

An activation function is a type of mathematical function that maps input non-linearly, which is applied to improve the feature representation ability of networks. It often follows convolutional layers and uses feature maps as input in neural networks. Sigmoid function [48] is a prevalent option for activation function, which is defined as:

$$y = sigmoid\ (x) = \frac{1}{1 + e^{-x}} \tag{2}$$

where x is the input and y represents the output. The sigmoid function experiences the vanishing gradient problem for very large or very small input.

ReLU [49] is another frequently used activation function. It is expressed as:

$$\mathrm{y} = \mathrm{ReLU}\ (\mathrm{x}) = \max(\mathrm{x}, 0) \tag{3}$$

where x is the input of ReLU function and y represents its output. It preserves the positive part in feature maps and prunes the negative part to 0. ReLU can alleviate the problem of vanishing gradient since its gradient is 1 when the input is positive, no matter how large it is.

However, when the input is negative, the output of ReLU and its gradient is always equal to 0. It can reduce overfitting, but it also obstacles CNN architecture to learn in some cases because of zero i.e. disconnection of neurons. LeakyReLU (LReLU) was proposed to address the problem of zero gradients when the input is negative for ReLU function [50]. It preserves the positive part fully, but it also preserves the negative part and scales them in a ratio λ (range 0 to 1). It is expressed as:

$$\mathrm{y} = \mathrm{LReLU}\ (\mathrm{x}) = \max\ (\mathrm{x}, 0) + \lambda \min(0, \mathrm{x}) \tag{4}$$

when its input is negative, both output and gradient are non-zero values.

Generally, Softmax function is applied to the final layer as activation function for the multi-class classification task. It is expressed as:

$$y(x)_i = \frac{e^{x_i}}{\sum_{i=1}^{K} e^{x_i}} \tag{5}$$

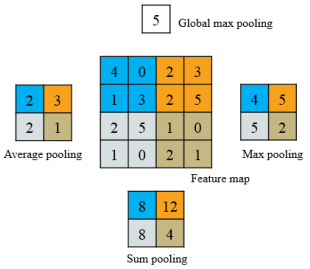

FIGURE 5. Pooling operations, with 2 × 2 filter and stride=2.

where x is input vector and x_i is its component. K-dimension output means K-class classification. $y(x)_i$ is the output which means the probability of the input vector is classified into the *ith* class.

d: POOLING LAYER

The feature map out of the convolutional layer records the position of pixels precisely, so it is very sensitive to the location of features. The high sensitivity means a small movement of the position of features, such as rotation and shift, will lead to a different map, which will decrease the robustness of CNNs. Usually, a pooling layer with pooling operation is applied after the convolutional layer to reduce specific feature positioning reliance and ensure the shift-invariance of CNNs. At the same time, pooling operations can also reduce the resolution of feature maps, and then reduce the burden of computation.

Pooling operations can be categorized as max-pooling [51], average pooling [52], and sum pooling [53], [54]. Figure 5 shows how pooling operations work: a sliding window is placed upon feature maps and max value, average value, or sum of the value in this window is calculated as output. Especially, if the size of the pooling window equals the size of the feature map, it is referred to as global pooling, otherwise, it is local/regional pooling.

e: FULLY CONNECTED LAYER

Fully connected layers (FCs) are flattened layers that generate specific semantic information. Each neuron in the fully connected layer has a connection with all the neurons in the previous layer, then all activations can be computed with matrix multiplication followed by biases.

2) LOSS FUNCTION

The loss function is used to evaluate the difference between the predicted result and desired result. An appropriate loss function can measure the difference between the result and label properly and guide a fast and correct training process. Following are some popular loss functions used in CNN architectures.

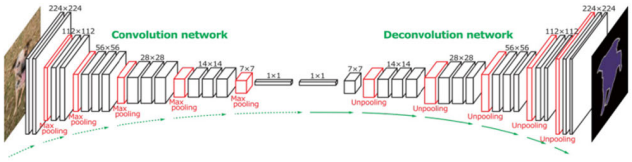

FIGURE 6. Architecture of an FCN [26].

For multi-class classification tasks, the most used loss function is cross-entropy function loss. It is given as:

$$\mathcal{L} = -\sum_{i=1}^{M} c_i \log(p_i) \tag{6}$$

where M is the number of classes, c_i is the practical label of an input belongs to *ith* class so it is 0 or 1, p_i is the probability of the input predicted by networks.

Cross-entropy can also be applied to binary classification since the sigmoid function is a special case of the Softmax function. Here, cross-entropy is known as binary cross-entropy loss function, it is expressed as:

$$\mathcal{L} = -(\mathrm{y}\log(\mathrm{p}) + (1-\mathrm{y})\log(1-\mathrm{p})) \tag{7}$$

where y is ground truth and p is predicted value.

Dice coefficient (DSC) is a statistical indicator that can be used to evaluate the similarity between two images. It is represented as:

$$\mathrm{DSC} = \frac{2|GT \cap SR|}{|GT| + |SR|} \tag{8}$$

where $|GT|$ represents the ground truth magnitude while $|SR|$ represents the segmentation result magnitude, $|GT \cap SR|$ represents the common elements between GT and SR. Based on the Dice coefficient, Dice loss (DSL) is another loss function widely applied to image segmentation tasks. Dice loss is expressed as:

$$\mathrm{DSL} = 1 - \mathrm{DSC} = 1 - \frac{2|GT \cap SR|}{|GT| + |SR|} \tag{9}$$

B. FULLY CONVOLUTIONAL NETWORKS (FCNS)

Long, *et al.* [26] proposed fully convolutional networks (FCNs) which replaced fully connected layers with up-sampling layers. The feature maps were up-sampled to the same size as the input images, and thus dense predictions were made by the network. The proposed FCN architecture is shown in Figure 6. Compared with traditional CNNs, FCNs can predict each pixel in an image or image patch, so it is more suitable and fast for image segmentation tasks.

C. U-NET

Ronneberger, *et al.* [55] proposed U-net which has symmetrical encoder-decoder structure and skip connections from encoding path to decoding path. Features were extracted in the encoder and images were reconstructed in the decoder. Skip connections sent low-level feature maps generated in the

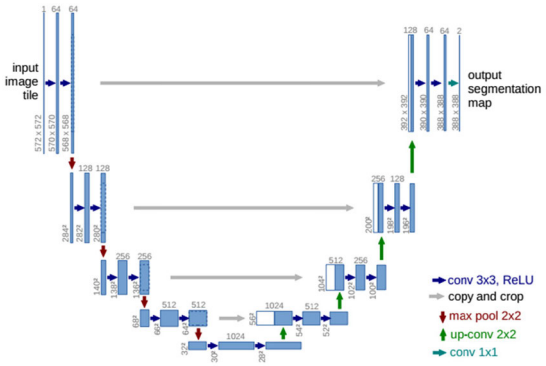

FIGURE 7. Architecture of U-net [55].

encoder to the decoder directly. Since low-level feature maps contained local information while high-level feature maps contained global information, then the proposed U-net integrated low-level and high-level feature maps and thus made the better prediction. Figure 7 shows the U-net architecture.

III. DATABASE AND EVALUATION METRICS FOR RETINAL VESSEL SEGMENTATION

Retina locates in the inner layer of the eyewall. A digital fundus camera attached with a low-power microscope is used to acquire retinal fundus images. The pupil of the human eye is the entry/exit point for fundus camera illumination and imaging light beams on the retina. The retinal fundus images can also be obtained through EasyScan camera based on Scanning Laser Ophthalmoscopy (SLO) [56]. SLO has the advantage of lower light exposure and has a better contrast between vessels and background due to the confocal design [57].

There are many publicly available databases for retinal vessel segmentation [10]. Here we just introduce several main databases. DRIVE [18], STARE [58], CHASE_DB1 [59] and HRF [6] are the most used publicly available databases. DRiDB [60] and ARIA [61], [62] are also available for retinal vessel segmentation but less used in recent years. Images in these six databases were obtained by the colorful fundus photography technique. In addition, two other databases, IOSTAR [63] and RC-SLO [64], whose samples were obtained by SLO, can also be used for retinal vessel segmentation. Table 1 indicates the brief information of these databases.

Generally, pixels in FOV of fundus images are classified as vessel pixel (positive) or non-vessel pixel (negative). To measure the identification of pixels, ground truth labels are compared with pixel identifications. On this basis, there are four basic pixel measures i.e., TP (true positives), FP (false positives), FN (false negatives), and TN (true negatives). Table 2 shows the measures of these elements through pixels.

Several evaluation metrics are defined to evaluate the performance of segmentation networks. Some of the prevalent metrics are listed in Table 3.

TABLE 1. Summary of 2-D fundus Image Datasets used for retinal vessel segmentation.

Dataset	Year	Resolution	Total Images	FOV	Format
STARE	2000	605x700	20	35°	PPM
DRIVE	2004	768x584	40	45°	JPEG
ARIA	2006	768x576	143	50°	TIFF
HRF	2009	3504×2336	45	45°	JPEG
CHASE_DB1	2011	999x960	28	30°	TIFF
DRiDB	2013	720x576	50	45°	BMP
IOSTAR	2015	1024x1024	30	45°	JPEG
RC-SLO	2015	360x320	40	/	TIFF

TABLE 2. Pixel measures in vessel segmentation.

Classification result		Ground truth	
		Vessel	Non-vessel
Segmentation result	Vessel	TP	FP
	Non-vessel	FN	TN

In addition, the Receiver operating characteristic curve (ROC curve) is a plot that summarizes the trade-off between TPR and FPR of a model under different thresholds. Therefore, the ROC curve can be utilized to compare different models under the identical threshold or a specific model under different thresholds. Similar to the ROC curve, the Precision-Recall curve (PR curve) illustrates the trade-off between Precision and Recall. The area under the curves, AUC-ROC (the area under the ROC curve), and AUC-PR (the area under the PR curve), are available to evaluate the overall performance of the networks.

IV. EXISTING MODELS FOR RETINAL VESSEL SEGMENTATION

In this section, we category and analyze various methods for retinal vessel segmentation according to their network architecture.

A. CNN FOR RETINAL VESSEL SEGMENTATION

Earlier, some researchers adopted CNNs with only several layers to segment vessels. We review 7 CNNs and summarize their performance evaluations in Table 4.

Fan and Mo [65] applied a 5-layer CNN to vessel segmentation and extracted image patches in the green channel as input. According to the comparison between R, G and B channels, the green channel provides the best vessel-background contrast than red and blue channels. They used L2-norm

TABLE 3. Evaluation metrics for image segmentation.

Matric	Expression
Sensitivity	$Sen = \frac{TP}{TP + FN}$
Specificity	$Spe = \frac{TN}{TN + FP}$
Precision	$Pre = \frac{TP}{TP + FP}$
Accuracy	$Acc = \frac{TP + TN}{TP + FP + TN + FN}$
F1-score	$F1 = \frac{2TP}{2TP + FP + FN}$
Jaccard Similarity	$JS = \frac{\lvert GT \cap SR \rvert}{\lvert GT \cup SR \rvert}$
Matthews Correlation Coefficient	$MCC = \frac{\frac{TP}{N} - S * P}{\sqrt{S * P * (1 - S) * (1 - P)}}$
G-mean	$G = \sqrt{Spe * Sen}$
False Positive Rate (FPR)	$FPR = \frac{FP}{TN + FP}$

as the loss function and adopted an optimized threshold to generate the binary vessel map.

Liskowski and Krawiec [66] proposed a CNN with 6 layers for retinal vessel segmentation. They applied global contrast normalization (GCN) and zero-phase component analysis (ZCA whitening) to training images in the pre-processing phase. GCN reduced the uneven illumination in images and ZCA abstracted features from universal characteristics and thus focused on the higher-order correlations.

Khalaf, *et al.* [67] constructed a CNN with 7 layers. They divided pixels in an image into 3-class: background, large vessel and small vessel to reduce the intra-classes variance. They extracted the green channel of images and applied adaptive histogram equalization (AHE) and top-hat filtering to the green channel in the pre-processing phase. The green channel and AHE increased image contrast and suppressed noise, and top-hat filtering enhanced vessels in training images.

Vengalil, *et al.* [68] proposed to fine-tune an existing network DEEPLAB-COCO-LARGEFOV using retinal fundus image patches. They replaced the last layer by a convolutional layer and applied a threshold to obtain final vessel maps. They did not adopt any image processing technique because they thought it may lead to undesired outcomes or harm vessel structures.

Tan, *et al.* [69] constructed a 7-layer CNN to make predictions for multiple objects in fundus images, including optic disc, fovea and retinal vessels. They extracted image patches in different channels with different sizes and resized them. Utilizing multiple channels can provide more information which is helpful for multi-object classification.

Guo, *et al.* [70] proposed a CNN with 6 layers and introduced a reinforcement sample learning scheme that trained the network on samples with poor performance. The proposed scheme allows researchers to train networks with fewer iterations of epochs and less training time as well as increased network performance.

Uysal and Güraksin [71] proposed a CNN model with several convolutional layers, and they also introduced transposed convolution to up-sample feature maps. Their proposed model made pixel-wise identification and did not perform well.

From Table 4 we can see that most CNNs just produced about 94% segmentation accuracy. We suppose that it is because CNNs have only several convolutional layers and do not have strong feature representation capacity, then they can only segment the basic structure and misclassified most of the vessel boundaries and thin vessels, so they are less used in recent years.

B. FCN FOR RETINAL VESSEL SEGMENTATION

FCNs can make dense and excellent predictions for each pixel in an image patch [26]. In this survey, we review 7 FCNs and list their performance evaluations in Table 4.

Oliveira, *et al.* [75] proposed an FCN and added skip connections to propagate features from shallow layer to deeper layer. They also explored the multiscale nature of the vascular system by using stationary wavelet transform (SWT) which added extra channels to input. Their result illustrated that the deep learning method can benefit from domain knowledge.

Jiang, *et al.* [74] used a network based on the fully convolutional version of AlexNet. They applied Gaussian smooth to reduce the discontinuity between FOV and the replaced region. The segmented vessels were thicker than ground truth, so Jiang, *et al.* [74] applied a 9×9 filter to refine the result and reduce noise in the post-processing phase.

Dasgupta and Singh [73], Soomro, *et al.* [76] also proposed FCNs for retinal vessel segmentation. Soomro, *et al.* [76] formulated a 2-classes classification task while Dasgupta and Singh [73] regarded the task as a multi-label inference task. Soomro, *et al.* [76] introduced principal component analysis (PCA) to convert RGB images into well contrast grayscale images.

Li, *et al.* [77] constructed an FCN with skip connections and introduced active learning to retinal vessel segmentation. Active learning used fewer manually labelled samples to improve the segmentation accuracy of blood vessels. The performance of the proposed model was increased in the iterative training process.

Since the consecutive down-sampling operations in the encoder lead to loss of information, which is critical to determine vessel boundaries and thin vessels. Luo, *et al.* [72] proposed a size-invariant fully convolutional

TABLE 4. Performance evaluations of CNNs for retinal vessel segmentation.

Reference	Database	Sensitivity	Specificity	Accuracy	AU-ROC	Kappa
Fan and Mo [65]	DRIVE	0.7814	0.9788	0.9612	/	/
	STARE	0.7234	0.9799	0.9614		
	CHASE_DB1	0.9702	0.9702	0.6761		
Liskowski and Krawiec [66]	DRIVE	0.7763	0.9768	0.9495	0.972	0.7781
	STARE	0.7867	0.9754	0.9566	0.9785	0.7622
Khalaf, et al. [67]	DRIVE	0.8397	0.9562	0.9456	/	/
Vengalil, et al. [68]	HRF	/	/	0.9394	0.894	/
Tan, et al. [69]	DRIVE	0.7537	0.9694	/	/	/
Guo, et al. [70]	DRIVE	/	/	0.9199	0.9652	/
	STARE			0.9220	0.9444	
Uysal and Güraksin [71]	DRIVE	0.7548	0.9682	0.9419	/	/
	STARE	0.7377	0.9735	0.9471		

TABLE 5. Performance evaluations of FCNs for retinal vessel segmentation.

Reference	Dataset	Precision	Sensitivity	Specificity	Accuracy	AU_ROC
Luo, et al. [72]	DRIVE	/	/	0.9741	0.9628	/
Dasgupta and Singh [73]	DRIVE	0.8498	0.7691	0.9801	0.9533	/
Jiang, et al. [74]	DRIVE	/	0.754	0.9825	0.9624	0.981
	STARE		0.8352	0.9846	0.9734	0.9900
	CHASE_DB1		0.8640	0.9745	0.9668	0.9810
	HRF		0.8010	0.8010	0.9650	0.9777
Oliveira, et al. [75]	DRIVE	/	0.8039	0.9804	0.9576	0.9821
	STARE		0.8315	0.9858	0.9694	0.9905
	CHASE_DB1		0.7779	0.9864	0.9653	0.9855
Soomro, et al. [76]	DRIVE	/	0.87	0.985	0.956	0.986
	STARE		0.848	0.986	0.968	0.988
	CHASE_DB1		0.886	0.982	0.976	0.985
	HRF		0.829	0.962	0.962	0.978
Li, et al. [77]	DRIVE	/	0.7752	0.9883	0.9697	0.9844
Atli and Gedik [78]	DRIVE	/	0.7987	0.9854	0.9689	0.9851
	STARE		0.6574	0.9933	0.9682	0.9748
	CHASE_DB1		0.7876	0.9845	0.9676	0.9892

neural network (SIFCN) to reduce its effect. They hold the size of feature maps in each layer by padding and assigning strides and thus reduces loss of information.

Atli and Gedik [78] proposed a fully convolutional network and they were the first to use up-sampling and down-sampling to capture thin and thick vessels, respectively. Their proposed model made some over segmentation and did not produce a very good performance, especially on STARE database.

FCNs have more convolutional layers that can extract high-level features, therefore, FCNs have performed better than other architectures as shown in Table 4 and Table 5. However, the segmentation results produced by FCNs are not enough fine, and the edges of segments are too blurry and smooth. FCNs also ignore the spatial consistency of pixels in pixel-wise segmentation. Conditional random field (CRF) [79] can be introduced to improve the segmentation of FCNs [80], [81].

C. U-NET FOR RETINAL VESSEL SEGMENTATION

U-net has a symmetric architecture and skip connection is applied to send feature maps from encoder to decoder directly [55]. Low-level feature maps contain rich detailed information while high-level have better global information, therefore, U-net can capture local and global information to make better decisions. In this survey, we review 32 U-shaped networks and list their performance evaluations in Table 6.

Guo, *et al.* [79] proposed a U-net and introduced structured dropout to regularize it. The proposed structured dropout is inspired by DropBlock [80] and discards continuous regions of feature maps in a ratio. Sule and Viriri [81] proposed a U-net and applied transpose convolution to the expanding path to recover the lost information.

Zhang and Chung [82] regarded retinal vessel segmentation as a multi-class classification task and introduced an edge-aware mechanism. They divided pixels into 5 classes: background, thick vessels, thin vessels, background near

thick vessels and background near thin vessels. The network can pay more attention to the boundary areas of vessels in this way. They leveraged deep supervision to ease optimization.

Mishra, *et al.* [83] proposed a simple U-net and introduced data-aware deep supervision to improve thin vessel segmentation. They computed the average input retinal vessel width and matched it with the layer-wise effective receptive fields to find layers that extract vessel features preeminently, and then add auxiliary layers there.

Laibacher, *et al.* [84] proposed a U-shaped network for retinal vessel segmentation which was built on pre-trained components of MobileNetV2. It was the first network to run in real-time on high resolution images. It utilized bottleneck modules and bilinear up-sampling to reduce the number of parameters so that the model could be employed on mobile and embedded systems. The network was trained by using a hybrid loss that combined binary cross-entropy and Jaccard index.

Jin, *et al.* [85] introduced deformable convolution to retinal vessel segmentation. The deformable convolution block adjusted the receptive fields adaptively by learning offsets and therefore captured the retinal vessels at various shapes and scales. The proposed deformable U-net produced better performance than U-net and deformable convolution network [86] on DRIVE, STARE and CHASE_DB1and two other datasets: WIDE [87] and SYNTHE [88].

Similar to Luo, *et al.* [72], Wang, *et al.* [89] also wanted to reduce information loss caused by consecutive down-sampling layers. They introduced a feature refinement path to U-net which sent low-level feature maps to high-level layers in encoder and decoder, respectively. The proposed feature refinement path can improve the detailed representation ability of the encoder and the discriminative ability of the decoder. While Yin, *et al.* [90] proposed to add multi-scale grayscale images to each stage of the encoder and decoder to reduce information loss and help information recovery.

Dharmawan, *et al.* [91] proposed a new directionally sensitive blood vessel enhancement method that combined CLAHE with a new match filter to detect micro vessels. The new matched filter was based on multi-scale and orientation modified Dolph-Chebyshev type I function. Their method detected more micro vessels than common CLAHE but still produced many mistakes.

Residual learning [92] was also introduced to increase the depth of networks as well as alleviate vanishing/exploding gradients. It was applied to building blocks [93]–[97] or skip connections [97], [98].

Dilated convolution [99] was also introduced to retinal vessel segmentation to enlarge the receptive fields [100]–[103]. Lopes, *et al.* [100] also tested the effect of different down-sampling techniques, that is, max-pooling, convolution with 2×2 kernel and convolution with 3×3 kernel. They obtained better results when using convolution as down-sampling operations, which is consistent with Soomro, *et al.* [76].

Jiang, *et al.* [101] arranged dilated rates deliberately to obtain a dense sampling of input and thus avoid the chessboard effect. They also introduced depthwise separable convolution [104] to reduce the computation cost and the number of parameters.

Soomro, *et al.* [103] introduced morphological transform and fuzzy C-means to the pre-processing of images. In the post-processing phase, they applied morphological reconstruction to remove small objects in segmented results. Mou, *et al.* [97] introduced probability regularized walk (PRW) algorithm to reconnect fractured vessels. PRW is an extension of the random walk algorithm [105] on probability maps.

There is a black ring around the field of view (FOV) in fundus images. Networks should pay more attention to the FOV since the black ring does not contain any information. Attention mechanism [106] has been applied to locate the region of interest (ROI) and strengthen feature representations in retinal vessel segmentation. Luo, *et al.* [107], Lian, *et al.* [108], Lv, *et al.* [109] made attention masks manually with the same size as original images to locate ROI. Wang, *et al.* [110], Li, *et al.* [111], Li, *et al.* [112], Fu, *et al.* [113] Tang, *et al.* [114] designed attention modules to strengthen feature representations, and their attention maps were learned by networks instead of assigned by experts.

Yan, *et al.* [115] introduced a novel joint loss to alleviate the highly unbalanced pixel ratio between thick and thin vessels in fundus images. They divided vessels into thin vessels and thick vessels to alleviate the unbalance problem. The joint loss includes pixel-wise and segment-level loss which emphasizes more on the thickness consistency of thin vessels.

Nasery, *et al.* [116] proposed a new data augmentation approach. They leveraged vignetting masks to create more annotated fundus images. Their method just adjusted the illumination condition of images but did not change the geometric and morphologic characteristics.

Galdran, *et al.* [117] proposed a new metric for retinal vessel segmentation and tested it using U-net. They introduced normalized mutual information to evaluate the segmentation quality. The new metric was applied to raw vessel probability map and can instruct the selection of threshold to binarize the vessel probability map.

Alvarado-Carrillo, *et al.* [118] focused on the curvilinear structures in vessels, so they proposed Distorted Gaussian Matched Filters (D-GMFs) with adaptive parameters and added them to the beginning and end of a U-net. They did not conduct an ablation study so we cannot know the effect of their proposed D-GMF Adaptive Unit.

Considering the large-scale variants of vessels and semantic variants existing in fundus images, Wu, *et al.* [119] proposed to adjust the receptive field adaptively to capture multi-scale features, they also adaptively fused features to extract more semantic information. They obtained a good result but still need to pay more attention to thin vessels.

From Table 6 we can see that U-net can produce about 96% segmentation accuracy, which is higher than FCNs'. U-net can reuse low-level information by concatenating feature maps, which also increases the computation burden, so the

input is a small image patch cropped from whole images. The network cannot identify pixels well because an image patch contains less information than a whole image, and it is also constrained by a limited receptive field, although dilated convolution can enlarge the receptive field.

D. MULTI-MODEL NETWORK FOR RETINAL VESSEL SEGMENTATION

Lots of researchers had found the limited prediction capability of a single model, so they proposed multi-model networks for stronger prediction ability. Most of them followed the spirit of U-net and FCN and employed encoder-decoder structure to form sub-models. We review 19 multi-model networks and summary their performance evaluations in Table 7.

Some research segmented thin/thick vessels or vessel boundaries/centers separately, then fused the segments to complete a whole segmentation [122]–[125]. These methods can be regarded as coarse-and-fine segmentation because thick/thin vessels or boundary/center vessels were segmented concurrently and separately.

Yan, *et al.* [122] proposed a three-stage segmentation network for retinal vessels using three sub-networks. The segmentation of the whole vessel tree was divided into three sub-tasks: thick vessel segmentation using FCN, thin vessel segmentation using U-net and fusion of segmentations.

Sathananthavathi and Indumathi [123] also proposed a coarse-and-fine strategy. They constructed 2 parallel FCNs, the first FCN was larger and trained with ground truth to extract thick and moderate vessels, while the second FCN was smaller and trained with skeletonized ground truth to extract thin vessels legibly. Outputs of both FCNs were integrated to generate the overall vessel segmentation.

Yang, *et al.* [124] proposed an improved U-net, whose encoder was used as a backbone to extract features, they arranged 2 decoders to segment thin and thick vessels, respectively. Finally, they added a fusion network to fuse the output of two decoders.

Wang, *et al.* [121] constructed a U-net which is composed of one encoder and three decoders. They used one decoder to generate a coarse probability map and divided an image into 'hard' or 'easy' regions according to the probability map. They used two other decoders to segment vessels in 'easy' and 'hard' regions independently. Finally, they fused all feature maps produced by 3 decoders to generate the final vessel map. They also introduced an attention gate to give more weight to vessel feature responses in decoders.

Tian, *et al.* [125] proposed a multi-model network to learn high- and low-frequency information, respectively. They applied Gaussian high-pass filter and Gaussian low-pass filter to original fundus images to obtain high-frequency or low-frequency information. They sent obtained high- and low-frequency information to two sub-networks with encoder-decoder structures, respectively. They fused the output of two sub-networks to get the final vessel map.

More researchers proposed coarse-to-fine segmentation by cascading several sub-networks. The following sub-network can inherit the learning experiences of previous sub-models [126]–[132]. Generally, they added intra- and inter- skip connections to send low-level feature maps and learned knowledge to deeper layers and sub-networks. The followed sub-network segmented vessels coarsely and the following sub-network refined vessel maps. The following sub-model used segmented results of previous sub-models and original images as input. Wu, *et al.* [128] added an auxiliary layer to the followed network to get an auxiliary loss, so their model was trained by main supervision and auxiliary supervision.

Guo, *et al.* [133] introduced an incremental learning strategy by cascading five CNNs. They trained the next CNN using the same samples as previous ones and enhanced it by feeding samples that were not performed well in the previous CNN. Finally, the final decision of each pixel was made using a voting scheme on the multiple CNNs results.

Lian, *et al.* [108] observed existing models always applied global pre-processing operations to images that will lose local information. They applied global and local operations simultaneously to enhance the contrast of images.

Tang, *et al.* [134] proposed a network with five identical and parallel sub-networks for ensemble learning. The input of each sub-network was grayscale images by extracting R-G channel image data with different proportions. Probability maps produced by five sub-models were averaged to generate the final segmentation result.

Zou, *et al.* [135] also formulated the task as a multi-class classification task to detect thin vessels with a width less than 2 pixels. They constructed 2 networks, one is for generating labels, one is for retinal vessel segmentation while the last one is for label simplification.

Cherukuri, *et al.* [136] proposed a domain-enriched network that was composed of two parts: a representation network to geometric features from fundus images and a residual task network to make a pixel-level prediction using the obtained features. Their method obtained a good performance but there are still non-vessel pixels identified as vessel pixels.

To consider the graphical structure of vessel shape, Shin, *et al.* [137] proposed a vessel graphic network that combined a graph neural network (GNN) [41] with a CNN to jointly utilize both local appearance and global vessel structure. They did not obtain a good result because their model misclassified many non-vessel pixels as vessel pixels.

Tajbakhsh, *et al.* [138] proposed an error correction mechanism that can learn from segmentation mistakes. The proposed network is divided into three sub-networks: a U-net to produce an initial segmentation, a network to produce diverse but representative error patterns and another U-net to make up the mistake of the initial segmentation map.

From Table 7 we can see that multi-model networks can produce about 96.3% segmentation accuracy, which have slight improvement compared with single networks. However, multi-model networks are also more difficult to train and have a higher computation burden.

TABLE 6. Performance evaluations of U-nets for retinal vessel segmentation.

Reference	Database	Precision	Sensitivity	Specificity	Accuracy	AU_ROC	F1-score	Others
Guo, et al. [79]	DRIVE	0.8335	0.7891	0.9848	0.9674	0.9836	/	0.9764(JS)
	STARE	0.8605	0.7548	0.9899	0.9725	0.9850		0.9763(JS)
	CHASE_DB1	0.8486	0.7559	0.9900	0.9738	0.9872		0.9738(JS)
Sule and Viriri [81]	DRIVE	/	0.7092	0.982	0.9447	0.9721	/	/
Zhang and Chung [82]	DRIVE	/	0.8723	0.9618	0.9504	0.9799	/	/
	STARE		0.7673	0.9901	0.9712	0.9882		
	CHASE_DB1		0.7670	0.9909	0.9770	0.9900		
Mishra, et al. [83]	DRIVE	/	0.8916	0.9601	0.9540	0.9724	/	/
	STARE		0.8805	0.9651	0.9601	0.9763		
	CHASE_DB1		0.8771	0.9634	0.9571	0.9742		
Laibacher, et al. [84]	DRIVE	/	/	/	0.9630	0.9714	0.8091	/
	CHASE_DB1				0.9703	0.9666	0.8006	
	HRF				0.9635	/	0.7814	
Jin, et al. [85]	DRIVE	0.8529	0.7963	0.9800	0.9566	0.9802	0.8237	0.9566(JS)
	STARE	0.8777	0.7595	0.9878	0.9641	0.9832	0.8143	0.9642(JS)
	CHAS_DB1	0.7630	0.8155	0.9752	0.9610	0.9804	0.7883	0.9610(JS)
	HRF	0.8593	0.7464	0.9874	0.9651	0.9831	--	--
Wang, et al. [89]	DRIVE	/	0.83	0.984	0.968	0.978	/	/
Yin, et al. [90]	DRIVE	/	0.7614	0.9837	0.9604	0.9846	/	/
	CHASE_DB1		0.7993	0.9868	0.9783	0.9869		
Dharmawan, et al. [91]	DRIVE	0.8157	0.8314	0.9726	/	/	0.8235	0.8992(Gmean) 0.7991(MCC)
	STARE	0.8412	0.7924	0.9827			0.8161	0.8124(Gmean) 0.7959(MCC)
	HRF	0.7838	0.8136	0.9770			0.7984	0.8916(Gmean) 0.7775(MCC)
Xiuqin, et al. [93]	DRIVE	/	0.931	0.9863	0.965	0.9811	/	/
Li, et al. [94]	DRIVE	/	0.7969	0.9799	/	0.9799	0.8237	0.8837(Gmean)
	STARE		0.8101	0.9795		0.9816	0.8147	0.8905(Gmean)
Khan, et al. [95]	DRIVE	/	0.8252	0.9787	0.9649	0.9780	/	//
	STARE		0.8397	0.9792	0.9659	0.9810		
	CHASE_DB1		0.844	0.9810	0.9722	0.9830		
Guo, et al. [96]	IOSTAR	/	0.8082	0.9854	0.9713	0.9873	/	0.8017(MCC)
	RC-SLO		0.8151	0.9879	0.9744	0.9848		0.8190(MCC)
Mou, et al. [97]	DRIVE	0.8132	0.9783	0.9607	/	/	/	/
	STARE	0.8398	0.9761	0.9698				
	CHASE_DB1	0.8275	0.9768	0.9648				
Adarsh, et al. [98]	DRIVE	/	0.7979	0.9794	0.9563	0.9795	0.8227	/
Lopes, et al. [100]	DRIVE	/	0.7903	0.9813	0.9567	/	/	/
Jiang, et al. [101]	DRIVE	/	0.7839	0.989	0.9709	0.9864	0.8246	
	STARE		0.8249	0.9904	0.9781	0.9927	0.8492	
	CHASE_DB1		0.7839	0.9894	0.9721	0.9866	0.8062	
Biswas, et al. [102]	DRIVE	/	0.7823	0.9814	0.9561	0.9794	/	/
Soomro, et al. [103]	DRIVE	/	0.802	0.974	0.959	0.948	/	/
	STARE		0.801	0.969	0.961	0.945		
Luo, et al. [107]	DRIVE	/	0.8075	0.9814	0.9663	0.9846	0.8203	
	STARE		0.8437	0.9762	0.9684	0.9765	0.8419	
Lv, et al. [109]	DRIVE	/	0.7941	0.9798	0.9558	0.9847	0.8216	0.9568(JS)
	STARE		0.7598	0.9878	0.9640	0.9824	0.8142	0.9638(JS)
	CHASE_DB1		0.8167	0.9704	0.9608	0.9865	0.7892	0.9603(JS)
Wang, et al. [110]	DRIVE	/	0.8071	0.9782	0.9565	0.9801	0.8251	/
	STARE		0.8432	0.9845	0.9702	0.9825	0.8516	
	CHASE_DB1		0.8427	0.9836	0.9706	0.9824	0.8105	
Li, et al. [111]	DRIVE	/	0.7921	0.9810	0.9568	0.9806	/	--
	STARE		0.8352	0.9823	0.9678	0.9875		--
	CHASE_DB1		0.7818	0.9819	0.9635	0.9810		--
	IOSTAR		0.7322	0.9802	0.9544	0.9623		0.7308(MCC)
	RC-SLO		0.8452	0.9807	0.9696	0.9842		0.8119(MCC)
Li, et al. [112]	DRIVE	/	0.8145	0.9883	0.9769	0.9895	/	/
	STARE		0.8505	0.9889	0.9797	0.9924		
	CHASE_DB1		0.9334	0.9862	0.9803	0.9912		

TABLE 6. ***(Continued.)*** **Performance evaluations of U-nets for retinal vessel segmentation.**

Fu, et al. [113]	DRIVE	/	0.8342	0.9732	0.9555	0.9795	0.8267	0.9125(AU_PR)
	STARE		0.8412	0.9807	0.9658	0.9863	0.8401	0.9250(AU_PR)
	CHASE_DB1		0.8312	0.9816	0.9644	0.9839	0.8237	0.9074(AU_PR)
Tang, et al. [114]	DRIVE	0.9682	0.9682	/	0.9551	0.9769	0.8155	/
	STARE	0.9932	0.9745		0.9687	0.9883	0.8312	
Yan, et al. [115]	DRIVE	/	0.7653	0.9818	0.9542	0.9752	/	/
	STARE		0.7581	0.9846	0.9612	0.9801		
	CHASE_DB1		0.7633	0.9809	0.9610	0.9781		
	HRF		0.7881	0.9592	0.6647	0.9437		
Nasery, et al. [116]	DRIVE	/	/	/	/	0.9787	/	0.9562(IoU)
Galdran, et al. [117]	DRIVE	/	/	/	0.9567	0.9750	/	/
Alvarado-Carrillo, et al. [118]	DRIVE	/	0.7960	0.9799	0.9772	/	0.8233	/
	STARE		0.7904	0.9843	0.9837		0.8141	
	CHASE_DB1		0.7530	0.9863	0.9798		0.8077	
Wu, et al. [119]	DRIVE	/	0.8289	0.9838	0.9697	0.9837	/	/
	STARE		0.8207	0.9839	0.9736	0.9877		
	CHASE_DB1		0.8365	0.9839	0.9744	0.9867		
	HRF		0.8114	0.9823	0.9687	0.9842		
	IOSTAR		0.8255	0.9830	0.9706	0.9865		
	LES[120]		0.8267	0.9881	0.9751	0.9827		
Wang, et al. [121]	DRIVE	/	0.7991	0.9813	0.9581	0.9823	0.8293	/
	STARE		0.8186	0.9844	0.9673	0.9881	0.8379	
	CHASE_DB1		0.8239	0.9813	0.9670	0.9871	0.8191	
	HRF		0.7803	0.9843	0.9654	0.9837	0.8074	
	ISOTAR		0.7538	0.9893	0.9652	0.9859	0.8161	
	RC-SLO.		0.8681	0.9797	0.9699	0.9911	0.8350	

E. GENERATIVE ADVERSARIAL NETWORK (GAN) FOR RETINAL VESSEL SEGMENTATION

GAN [40] is a type of deep unsupervised learning model, which is composed of a generator and a discriminator. In this survey, we review 13 GANs and list their performance evaluations in Table 8.

Most generative models adopted encoder-decoder structure with some improvement modules, such as dense block [139], Wu, *et al.* [140], and dilated convolution [139], [141], deep supervision [142], attention mechanism [140], [143], skip connections [144], Inception module [145] and others [146]. CNNs were widely used as generators [139]–[141], [144], [146]–[148], but U-net can also be adopted [143], [145].

Son, *et al.* [148] explored several models for the discriminators: pixel-GAN, patch-GAN and image-GAN. The results indicate that patch-GAN performs better than others including the one with a single generator.

Park, *et al.* [149] chained two U-nets in the generator and used residual convolution blocks as building blocks in both generator and discriminator. They utilized automatic color equalization (ACE) to enhance images in the pre-processing phase with the Lanczos resampling method to smooth the vessel branches and reduce false negatives in the post-processing phase.

GANs based on semi-supervised learning were also explored to address the problem of lacking annotated data. Huo, *et al.* [150] proposed a semi-supervised framework that combined GAN and self-training scheme, and they adopted particle swarm optimization (PSO) [151] algorithm to choose the hyperparameters in semi-supervised learning since self-training is sensitive to hyperparameters. They obtained 0.9550/0.8419 of AUC_ROC and AUC_PR on the DRIVE database when using 0.1 labelled and 0.9 unlabeled data. Lahiri, *et al.* [152] also trained a GAN based on semi-supervised learning to learn from both labeled and unlabeled data. They only used 3K annotated image patches to make patch-wise predictions and obtained 0.95/0.96 accuracy and 0.96/0.94 AUC on DRIVE and STARE databases, respectively. Their models outperformed simple U-net and used less annotated data, but they also did not segment vessels as accurately as other improved models based on supervised learning.

From Table 8 we can see that GANs produced about 96% segmentation accuracy, which is similar to U-net. Park, *et al.* [149] obtained the best performance. Compared with CNNs, we need to train generators and discriminators alternatively in GANs, which is more troublesome.

F. OTHER NETWORK FOR RETINAL VESSEL SEGMENTATION

Researchers also proposed networks that cannot be categorized into the forgoing classes due to their unique architectures. The performance of these methods is listed in Table 9. Ngo and Han [153], Guo, *et al.* [154], Li, *et al.* [155] adopted multiple input branches to capture multi-scale spatial information. Further, all the feature maps generated by each branch are combined to make predictions. In addition, Guo, *et al.* [154] applied the K-dimensions tree integrated with the hessian matrix to reconnect the broken segments in the post-processing stage. Some broken vessels were reconnected and the vessel map

TABLE 7. Performance evaluations of multi-model networks for retinal vessel segmentation.

Reference	Database	Sensitivity	Specificity	Accuracy	AU_ROC	F1-score	Others
Lian, et al. [108]	DRIVE	0.8278	0.9861	0.9692	/	/	0.8637(Pre)
	STARE	0.8342	0.9916	0.9740			0.8823(Pre)
Wang, et al. [121]	DRIVE	0.7991	0.9813	0.9581	0.9823	0.8293	/
	STARE	0.8186	0.9844	0.9673	0.9881	0.8379	
	CHASE_DB1	0.8239	0.9813	0.9670	0.9871	0.8191	
	HRF	0.7803	0.9843	0.9654	0.9837	0.8074	
	ISOTAR	0.7538	0.9893	0.9652	0.9859	0.8161	
	RC-SLO.	0.8681	0.9797	0.9699	0.9911	0.8350	
Yan, et al. [122]	DRIVE	0.7631	0.9820	0.9538	0.9750	/	/
	STARE	0.7735	0.9857	0.9638	0.9833		
	CHASE_DB1	0.7641	0.9806	0.9607	0.9776		
Sathananthavathi and Indumathi [123]	DRIVE	0.7287	0.9818	0.9594	/	/	/
	STARE	0.8172	0.9624	0.9531			
Yang, et al. [124]	DRIVE	0.8353	0.9751	0.9579	/	0.8297	/
	STARE	0.7946	0.9821	0.9626		0.8155	
	CHASE_DB1	0.8176	0.9776	0.9632		0.7997	
Tian, et al. [125]	DRIVE	0.8639	0.969	0.958	0.956	/	/
Xia, et al. [126]	DRIVE	0.7979	0.9857	0.9685	/	/	/
Wang, et al. [127]	DRIVE	0.7849	0.9813	0.9567	0.9788	0.8241	/
	STARE	0.9024	0.9934	0.9849	0.9960	0.9184	
	CHASE_DB1	0.7948	0.9842	0.9648	0.9847	0.8220	
Wu, et al. [128]	DRIVE	0.7996	0.9813	0.9582	0.9830	/	/
	STARE	0.7963	0.9863	0.9672	0.9875		
	CHASE_DB1	0.8003	0.9880	0.9688	0.9894		
Hu, et al. [129]	DRIVE	0.8312	0.9751	0.9567	0.9821	0.8303	0.8055(MCC)
	CHASE_DB1	0.8044	0.9861	0.9658	0.9867	0.8242	0.8065(MCC)
Budak, et al. [130]	DRIVE	0.7439	0.9900	0.9685	0.9822	0.9829	/
	STARE	0.8196	0.9871	0.9735	0.9868	0.9852	
Francia, et al. [131]	DRIVE	0.8160	/	0.9696	/	0.8250	0.9341(Pre)
	CHASE_DB1	0.8258		0.9766		0.9312	0.8366(Pre)
Li, et al. [132]	DRIVE	0.7735	0.9838	0.9573	0.9816	0.8205	/
	STARE	0.7715	0.9886	0.9701	0.9881	0.8146	
	CHASE_DB1	0.7970	0.9823	0.9655	0.9851	0.8073	
Guo, et al. [133]	DRIVE	0.9859	0.7046	0.9613	0.9737	0.7613	/
	STARE	0.9861	0.5629	0.9539	0.9539	0.6502	
Tang, et al. [134]	DRIVE	0.8564	0.9710	0.9574	0.9822	/	0.7984(MCC)
	STARE	0.8162	0.9869	0.9695	0.9898		0.8066(MCC)
	CHASE_DB1	0.8106	0.9807	0.9654	0.9850		0.7700(MCC)
	HRF	0.7782	0.9843	0.9631	0.9843		0.7969(MCC)
Zou, et al. [135]	DRIVE	0.7761	0.9792	0.9519	/	0.8129	/
	STARE	0.8120	0.9895	0.9704		0.8553	
Cherukuri, et al. [136]	DRIVE	0.8426	0.9823	0.9603	0.9844	0.8220	
	STARE	0.8667	0.9871	0.9734	0.9930	0.8364	
	CHASE_DB1	0.8025	0.9874	0.9693	0.9858	0.8211	
	HRF	0.8144	0.9733	0.9588	-	0.7832	
Shin, et al. [137]	DRIVE	0.9382	0.9255	0.9271	0.9802	/	/
	STARE	0.9598	0.9352	0.9378	0.9877		
	CHASE_DB1	0.9463	0.9364	0.9373	0.9830		
	HRF	0.9546	0.9329	0.9349	0.9838		
Tajbakhsh, et al. [138]	CHASE_DB1	/	/	/	/	0.8150	/
	ARIA					0.7200	

became cleaner after post-processing. Accuracy and sensitivity were increased while specificity was decreased after post-processing. Li, *et al.* [155] introduced sparse variables into the label design and improved the cross-entropy loss function to address the unbalance of samples. Li, *et al.* [155] got better sensitivities than [153] and [154] because they solved the class imbalanced issue.

Holistically-nested edge detection (HED) network made a significant advancement on edge detection in an image [162]. It is a single-stream network with multiple side outputs and final predictions are made by fusing multi-scale side outputs. Inspired by the HED network, Mo and Zhang [20], Guo, *et al.* [158] integrated feature maps generated in different stages of networks to generate the final probability

TABLE 8. Performance evaluations of GANs for retinal vessel segmentation.

Reference	Database	Sensitivity	Specificity	Accuracy	AU_ROC	AU_PR	F1-score
Tu, et al. [139]	DRIVE	0.784	0.985	0.9571	0.985	/	/
Wu, et al. [140]	DRIVE	0.7798	0.982	0.9615	/	/	/
Ma, et al. [141]	DRIVE STARE	/	/	/	0.9817 0.9847	0.9249 0.9265	0.8310 0.8492
Dong, et al. [142]	DRIVE STARE	0.8574 0.8126	0.9828 0.9837	0.9729 0.9716	/	/	0.8342 0.8074
Zhou, et al. [143]	DRIVE STARE CHASE_DB1 HRF	0.8294 0.8812 0.8435 0.8310	0.9812 0.9781 0.9782 0.9730	0.9563 0.9671 0.9630 0.9559	0.9830 0.9863 0.9872 0.9693	0.8397(Pre) 0.7952(Pre) 0.8013(Pre) 0.8115(Pre) 0.9021(Gmean) 0.9283(Gmean) 0.9083(Gmean) 0.8992(Gmean)	0.8345 0.9359 0.8218 0.8211
Yang, et al. [144]	DRIVE STARE	0.834 0.8334	0.982 0.9897	0.956 0.9663	0.9786 0.9734	0.8821 0.8718	/
Guo, et al. [145]	DRIVE	0.8283	0.9726	0.9542	0.9772	0.9058	0.8215
Rammy, et al. [146]	DRIVE STARE	/	/	/	0.9817 0.9845	0.9154 0.917	0.832 0.842
He and Jiang [147]	DRIVE STARE	0.832 0.855	0.982 0.990	0.963 0.973	0.980 0.991	/	/
Son, et al. [148]	DRIVE STARE	/	/	/	0.9810 0.9873	0.9145 0.9226	0.8275 0.8378
Park, et al. [149]	DRIVE STARE CHASE_DB1 HRF	0.8346 0.8324 -- --	0.9836 0.9938 -- --	0.9706 0.9876 0.9736 0.9761	0.9868 0.9873 0.9859 0.9852	0.8163 0.8306 0.7979 0.7845 (MCC)	0.8324 0.8370 0.8110 0.7972
Huo, et al. [150]	DRIVE	/	/	/	0.9550	0.9419	/
Lahiri, et al. [152]	DRIVE STARE	/	/	0.95 0.96	0.96 0.94	/	/

map. Lin, *et al.* [156], Hu, *et al.* [157] also constructed single-stream networks based on VGG Net and applied fully-connected CRFs [163] to get the final binary segmentation result. CRFs utilized multiscale feature maps in different stages to make full use of spatial contextual information. CRFs can mitigate noise and edge blurring acting as a global smoothness regularizer.

Feng, *et al.* [159] proposed a cross-connected network (CcNet) that has two parallel paths. They used two CRM (convolution-ReLU-Max pooling) modules as building blocks and formed cross-connections between these two paths. They sent feature maps produced by each module in the upper path to each module in the lower path. They concatenated all feature maps generated in the lower path to generate final vessel maps. Since these cross-connections, CcNet can learn multi-scale features efficiently.

Zhuo, *et al.* [160] used three dense blocks to form a straight network and added two bottleneck blocks between them, which aimed to reduce the model complexity and computation cost. Similar to Luo, *et al.* [72], they also maintained the size of feature maps by cancelling down-sampling layers to reduce information loss of tiny vessels. In addition, considering the existing evaluation metrics should not be equally important since great unbalance exists between vessels and non-vessels, Zhuo, *et al.* [160] proposed a new evaluation index named fusion score, which converts multiple evaluation metrics into a single target. It is expressed in Equation 9:

$$\mathrm{FS} = \frac{3 * F1 * MCC * Gmean}{F1 * MCC + F1 * Gmean + MCC * Gmean} \quad (10)$$

They got a fusion score of 0.8339, 0.8449 on DRIVE and STARE databases, respectively.

To reduce the high-frequency information loss caused by consecutive down-samplings in the encoder, Noh, *et al.* [161] introduced a scale-space approximated CNN (SSA-Net). It is a single-stream network with residual connection and skip connections. They inserted up-sampling layers in the feature generation phase to generate size-invariant feature maps and thus reduce spatial scale-space distortions.

CLAHE is widely applied to fundus images to enhance image contrast, which has two parameters: size of the contextual region and clip limit. Most researchers just used default values for CLAHE, but Aurangzeb, *et al.* [164] introduced PSO to CLAHE to find optimal parameter values. They did not propose a new network but just applied their method to existing models.

From Table 9 it can be observed that Noh, *et al.* [161] obtained the best performance in these methods, and these

TABLE 9. Performance evaluations of other networks for retinal vessel segmentation.

Reference	Dataset	Sen	Spe	Acc	AUC_ROC	F1-score	Others
Mo and Zhang [20]	DRIVE	0.7779	0.978	0.9521	0.9782	/	/
	STARE	0.8147	0.9844	0.9674	0.9885		
	CHASE_DB1	0.7661	0.9816	0.9599	0.9812		
Ngo and Han [153]	DRIVE	0.7464	0.9836	0.9533	0.9752	/	/
Guo, et al. [154]	DRIVE	0.7560	0.9839	0.9625	0.9782	/	/
Li, et al. [155]	DRIVE	0.8347	0.9796	0.9510	0.9792	/	/
	STARE	0.8231	0.9782	0.9560	0.9743		
Lin, et al. [156]	DRIVE	0.7632	/	0.9536	/	/	/
	STARE	0.7423		0.9603			
	CHASE_DB1	0.7815		0.9587			
Hu, et al. [157]	DRIVE	0.7772	0.9793	0.9533	0.9759	/	/
	STARE	0.7543	0.9814	0.9632	0.9751		
Guo, et al. [158]	DRIVE	0.7800	0.9806	0.9551	0.9796	0.8208	0.7923(MCC)
	STARE	0.8201	0.9828	0.9660	0.9872	0.9362	0.8142(MCC)
	CHASE_DB1	0.7888	0.9801	0.9627	0.9840	0.7983	0.7733 (MCC)
Feng, et al. [159]	DRIVE	0.7625	0.9809	0.9528	0.9678	/	/
	STARE	0.7709	0.9848	0.9633	0.97		
Zhuo, et al. [160]	DRIVE	0.8432	0.9681	0.9520	0.9754	0.8163	0.7905(MCC) 0.8044(MCC)
	STARE	0.8630	0.973	0.9620	0.9824	0.8233	0.9031(Gmean) 0.9158(Gmean)
Noh, et al. [161]	DRIVE	0.8354	0.9746	0.9569	0.9820	/	/
	STARE	0.8537	0.9864	0.9764	0.9921		
	CHASE_DB1	0.8523	0.9871	0.9778	0.9916		

methods produced about 95.5% segmentation accuracy. These models did not have a strong learning capacity because most of them have only several convolutional layers, and their architectures may not be very suitable for this task.

V. DISCUSSION

In this survey, we have reviewed 89 deep learning models for retinal vessel segmentation, which indicate deep learning has been widely applied to segment retinal vessels. Earlier, researchers applied simple CNNs for vessel segmentation on DRIVE and STARE databases [69], [165]. Moreover, lots of researchers have proposed various improved models for retinal vessel segmentation. FCNs and U-nets were the most leveraged to make dense predictions because of the excellent performance [73], [79], [81]. Later, different improvement modules such as residual block, dilated convolution and attention mechanism, were introduced with U-net to improve the performance of proposed models [101], [102], [107]. On the other hand, researchers also proposed multi-model networks to get a stronger identification capability [122], [128], and others introduced GANs to vessel segmentation [128]. Multi-branch networks and HED-shaped networks were also used for vessel segmentation [153], [166].

A. CHALLENGES IN RETINAL VESSEL SEGMENTATION USING DEEP LEARNING

According to the existing research, the following challenges are encountered while using deep learning for retinal vessel segmentation:

1) Lack of well labelled training samples. Although there is a large number of fundus images, acquiring annotated data is very difficult to obtain since it requires professional doctors and takes a significant amount of time and cost.
2) Low quality of existing image samples. It hinders deep learning models to learn better feature representations. Image noise, uneven illumination, low contrast especially for thin vessels, centerline reflection and various structures (pathological region, fovea, macula, optic disc) decrease the performance of proposed models.
3) Class imbalance problem of training samples. The different number of positive and negative examples available for training degrades the performance of networks. Class imbalance problem not only exists between foreground and background, but also in thick vessels and thin vessels. Deep learning models tend to classify pixels in boundaries as non-vessels pixels because the number of non-vessels pixels are in large quantities as compared to vessel pixels. The network performs worse on thin vessels than thick vessels since the misclassification of pixels in thin vessels has less influence on the total loss.

B. KEY ASPECTS FOR SUCCESSFUL RETINAL VESSEL SEGMENTATION

From the analysis of existing methods, a successful model should be able to detect vessels under uneven illumination, low contrast and various regions in fundus images. At the

same time, it should be robust enough for images and have strong generalization ability. We identify some key aspects for successful and robust retinal vessel segmentation, which are as follow:

1) Raw image enhancement. Using image enhancement techniques pre-processing phase, such as the conversion of RGB images to grayscale images, normalization, contrast limited adaptive histogram equalization (CLAHE) [167] and gamma correction increase the image quality [73], [93], [107]. We can also adopt morphological operations [168] to increase the quality of images [76], [123].
2) Data augmentation. The publicly available databases are too small to train a network, so we can utilize regular data augmentation techniques [169] to enlarge the training dataset, such as rotating, flipping, shifting, mirroring and cropping images into image patches [23], [107], [108], [122], [156]. Also, we can leverage transfer learning for this task, such as VGGNet [20], [156]–[158], ResNet [158] or a fully convolutional version of AlexNet [74].
3) A Well-designed model. A well-designed model could capture more spatial information, reduce loss of local information and reuse low-level feature maps for accurate segmentation. From the segmentation result, U-net and multi-model networks have a better performance than CNNs and FCNs, that is because they have more convolutional layers then they can extract features better. In addition, skip connections also help the reuse of low-level information, which is very important in identification. Some proposed GANs also obtained high accuracy. In addition, dilated convolution is a good option to enlarge the receptive field and capture more spatial information but still keep the same number of parameters [99]. Residual learning can increase network depth and alleviate network degradation at the same time [92]. A dense connection can make full use of feature maps generated by all previous layers and thus decrease model complexity and mitigate vanishing gradient [170].
4) Proper loss function. A proper loss function could lead models to pay more attention to vessels, especially thin vessels. Researchers can adopt improved loss functions, such as weighted cross-entropy loss function, to solve the imbalance problem [82], [125], [158].
5) Vessel map enhancement. The segmentation result contains noise and isolated small vessels, so we can use a matched filter or morphological transform to illuminate them in the post-processing phase [74], [103]. The vessel segments are broken in some cases, and we can reconnect fractured vessels by some techniques, such as PRW and K-dimensions tree [97], [154]. Better visualization of the vessel map helps ophthalmologists diagnose disease easier.
6) Abundant validation: we cannot only verify our models using a single database, but also cross-validate networks to evaluate their generalization ability. In cross-validation, a network is trained using samples from one dataset but tested using another dataset [20], [75], [123], [128]. Even we can conduct mixed validation for further check. We can train a network using mixed samples from several databases and test it using the rest samples from these databases [74].

From the analysis of the reviewed articles, several proposed research in terms of models and strategies to improve the performance of networks, such as incremental learning strategy [133], various improvement modules [93], [101], [121], coarse-to-fine segmentation [127], there is still no model can segment vessels perfectly, including segmentation of vessel boundaries and thin vessels, segmentation of background between two closed vessels, segmentation of vessel under the presence of abnormalities and various structures, segmentation of vessel in cross-connections and robust segmentation between different databases. In addition, the segmented vessels are still fractured and broken in most results, which invites researchers to investigate further to reconnect fractured vessels.

Although deep learning has been widely applied to retinal vessel segmentation, there are still some limitations. Compared with human beings, deep learning has less generalization capacity. Compared with conventional methods, such as matched filtering methods and vessel tracing methods, deep learning is more uninterpretable, and it needs massive data and GPUs in training processes, which are not available and expensive for users in some cases.

VI. CONCLUSION

Geometric characteristics of retinal vessels reflect clinical and pathological features. The ophthalmologist uses vessel maps to diagnose diseases, such as DR and MD. Precise diagnoses of eye abnormalities and their timely treatment are important in preventing global blindness.

Computerized automatic segmentation for retinal blood vessels is inspired since manual segmentation of retinal blood vessels is expensive and time-consuming. In the past, researchers proposed different methods for automatic retinal vessel segmentation. Unsupervised models are limited by their accuracy. Machine learning algorithms require handcrafted features and thus are limited by their generalization ability. Currently, deep learning models have been greatly used to image segmentation including retinal images since they do not need any handcrafted features and outperform existing unsupervised methods.

This article reviews publications of recent six years for retinal vessel segmentation based on deep learning. The main contribution of our works is to analyze recent models and find out new trends for retinal vessel segmentation. It will be helpful for researchers and industrialists to develop a robust model for retinal vessel segmentation.

REFERENCES

[1] Early Treatment Diabetic Retinopathy Study Research Group, "Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12," *Ophthalmology*, vol. 98, pp. 823–33, 1991.

[2] J. J. Kanski and B. Bowling, *Clinical Ophthalmology: A Systematic Approach*. Amsterdam, The Netherlands: Elsevier, 2011.

[3] A. Fathi and A. R. Naghsh-Nilchi, "Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation," *Biomed. Signal Process. Control*, vol. 8, no. 1, pp. 71–80, Jan. 2013.

[4] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S. J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, and S. Haffner, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes Care*, vol. 35, no. 3, pp. 556–564, 2012.

[5] S. W. Franklin and S. E. Rajan, "Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images," *Biocybern. Biomed. Eng.*, vol. 34, no. 2, pp. 117–124, 2014.

[6] T. Kohler, A. Budai, M. F. Kraus, J. Odstrcilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proc. 26th IEEE Int. Symp. Comput. Based Med. Syst.*, Jun. 2013, pp. 95–100.

[7] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *Proc. SPIE*, vol. 5370, pp. 648–656, May 2004.

[8] P. Dai, H. Luo, H. Sheng, Y. Zhao, L. Li, J. Wu, Y. Zhao, and K. Suzuki, "A new approach to segment both main and peripheral retinal vessels based on gray-voting and Gaussian mixture model," *PLoS ONE*, vol. 10, no. 6, Jun. 2015, Art. no. e0127748.

[9] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images—A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, Oct. 2012, doi: 10.1016/j.cmpb.2012.03.009.

[10] M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, R. M. Anjana, A. S. Doney, C. N. A. Palmer, and E. Trucco, "A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101905.

[11] M. Al-Rawi, M. Qutaishat, and M. Arrar, "An improved matched filter for blood vessel detection of digital retinal images," *Comput. Biol. Med.*, vol. 37, no. 2, pp. 262–267, 2007.

[12] S. S. Kar and S. P. Maity, "Blood vessel extraction and optic disc removal using curvelet transform and kernel fuzzy c-means," *Comput. Biol. Med.*, vol. 70, pp. 174–189, Mar. 2016.

[13] K. Rezaee, J. Haddadnia, and A. Tashk, "Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization," *Appl. Soft Comput.*, vol. 52, pp. 937–951, Mar. 2017.

[14] Y. Yin, M. Adel, and S. Bourennane, "Retinal vessel segmentation using a probabilistic tracking method," *Pattern Recognit.*, vol. 45, no. 4, pp. 1235–1244, Apr. 2012.

[15] M. Nergiz and M. Akın, "Retinal vessel segmentation via structure tensor coloring and anisotropy enhancement," *Symmetry*, vol. 9, no. 11, p. 276, Nov. 2017.

[16] J. Kaur and D. Mittal, "A generalized method for the detection of vascular structure in pathological retinal images," *Biocybern. Biomed. Eng.*, vol. 37, no. 1, pp. 184–200, 2017.

[17] D. Kaba, A. G. Salazar-Gonzalez, Y. Li, X. Liu, and A. Serag, "Segmentation of retinal blood vessels using Gaussian mixture models and expectation maximisation," in *Proc. Int. Conf. Health Inf. Sci.* London, U.K.: Springer, 2013, pp. 105–112.

[18] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[19] X. You, Q. Peng, Y. Yuan, Y.-M. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2314–2324, 2011.

[20] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 12, pp. 2181–2193, Dec. 2017, doi: 10.1007/s11548-017-1619-0.

[21] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[22] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M. L. Shyu, S. C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Sep. 2018.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[27] T. A. Soomro, A. J. Afifi, L. Zheng, S. Soomro, J. Gao, O. Hellwich, and M. Paul, "Deep learning models for retinal blood vessels segmentation: A review," *IEEE Access*, vol. 7, pp. 71696–71717, 2019, doi: 10.1109/access.2019.2920616.

[28] K. B. Khan, A. A. Khaliq, A. Jalil, M. A. Iftikhar, N. Ullah, M. W. Aziz, K. Ullah, and M. Shahid, "A review of retinal blood vessels extraction techniques: Challenges, taxonomy, and future trends," *Pattern Anal. Appl.*, vol. 22, no. 3, pp. 767–802, 2019.

[29] M. Badar, M. Haris, and A. Fatima, "Application of deep learning for retinal image analysis: A review," *Comput. Sci. Rev.*, vol. 35, Feb. 2020, Art. no. 100203.

[30] T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101971.

[31] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[32] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[33] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 1–38, 2010.

[34] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 153–160.

[35] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[36] G. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade* (Lecture Notes in Computer Science), vol. 7700, 2nd ed. 2012, pp. 599–619.

[37] Y. LeCun, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1990, pp. 396–404.

[38] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[41] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[42] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.

[43] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[45] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400

[46] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[48] H. N. Mhaskar and C. A. Micchelli, "How to choose an activation function," in *Proc. Adv. Neural Inf. Process. Syst.*, 1994, pp. 319–326.

[49] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.

[50] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013, p. 3.

[51] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 111–118.

[52] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 3304–3308.

[53] F. H. Hamker, "Predictions of a model of spatial attention using sum- and max-pooling functions," *Neurocomputing*, vol. 56, pp. 329–343, Jan. 2004.

[54] S. Aich and I. Stavness, "Global sum pooling: A generalization trick for object counting with small datasets of large images," 2018, *arXiv:1805.11123*. [Online]. Available: http://arxiv.org/abs/1805.11123

[55] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[56] R. H. Webb and G. W. Hughes, "Scanning laser ophthalmoscope," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 7, pp. 488–492, Jul. 1981.

[57] F. LaRocca, D. Nankivil, S. Farsiu, and J. A. Izatt, "True color scanning laser ophthalmoscopy and optical coherence tomography handheld probe," *Biomed. Opt. Exp.*, vol. 5, no. 9, pp. 3204–3216, 2014.

[58] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[59] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.

[60] P. Prentasic, S. Loncaric, Z. Vatavuk, G. Bencic, M. Subasic, T. Petkovic, L. Dujmovic, M. Malenica-Ravlic, N. Budimlija, and R. Tadic, "Diabetic retinopathy image database(DRiDB): A new database for diabetic retinopathy screening programs research," in *Proc. 8th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2013, pp. 711–716.

[61] D. J. Farnell. *Automated Retinal Image Analysis (ARIA) Data Set*. Accessed: Jun. 3, 2021. [Online]. Available: http://www.damianjjfarnell.com/?page_id=276

[62] D. J. J. Farnell, F. N. Hatfield, P. Knox, M. Reakes, S. Spencer, D. Parry, and S. P. Harding, "Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators," *J. Franklin Inst.*, vol. 345, no. 7, pp. 748–765, Oct. 2008.

[63] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. W. Pluim, R. Duits, and B. M. T. H. Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2631–2644, Dec. 2016.

[64] S. Abbasi-Sureshjani, I. Smit-Ockeloen, J. Zhang, and B. T. H. Romeny, "Biologically-inspired supervised vasculature segmentation in SLO retinal fundus images," in *Proc. Int. Conf. Image Anal. Recognit.* Niagara Falls, ON, Canada: Springer, 2015, pp. 325–334.

[65] Z. Fan and J.-J. Mo, "Automated blood vessel segmentation based on de-noising auto-encoder and neural network," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Jul. 2016, pp. 849–856.

[66] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016, doi: 10.1109/TMI.2016.2546227.

[67] A. F. Khalaf, I. A. Yassine, and A. S. Fahmy, "Convolutional neural networks for deep feature learning in retinal vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 385–388.

[68] S. K. Vengalil, N. Sinha, S. S. S. Kruthiventi, and R. V. Babu, "Customizing CNNs for blood vessel segmentation from fundus images," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jun. 2016, pp. 1–4.

[69] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *J. Comput. Sci.*, vol. 20, pp. 70–79, May 2017.

[70] Y. Guo, Ü. Budak, L. J. Vespa, E. Khorasani, and A. Şengür, "A retinal vessel detection approach using convolution neural network with reinforcement sample learning strategy," *Measurement*, vol. 125, pp. 586–591, Sep. 2018.

[71] E. Uysal and G. E. Güraksin, "Computer-aided retinal vessel segmentation in retinal images: Convolutional neural networks," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3505–3528, Jan. 2021.

[72] Y. Luo, H. Cheng, and L. Yang, "Size-invariant fully convolutional neural network for vessel segmentation of digital retinal images," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–7.

[73] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 248–251.

[74] Z. Jiang, H. Zhang, Y. Wang, and S.-B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Med. Imag. Graph.*, vol. 68, pp. 1–15, Sep. 2018, doi: 10.1016/j.compmedimag.2018.04.005.

[75] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Syst. Appl.*, vol. 112, pp. 229–242, Dec. 2018, doi: 10.1016/j.eswa.2018.06.034.

[76] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, and M. Paul, "Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation," *Expert Syst. Appl.*, vol. 134, pp. 36–52, Nov. 2019, doi: 10.1016/j.eswa.2019.05.029.

[77] W. Li, M. Zhang, and D. Chen, "Fundus retinal blood vessel segmentation based on active learning," in *Proc. Int. Conf. Comput. Inf. Big Data Appl. (CIBDA)*, Apr. 2020, pp. 264–268.

[78] I. Atli and O. S. Gedik, "Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation," *Eng. Sci. Technol., Int. J.*, vol. 24, no. 2, pp. 271–283, Apr. 2021.

[79] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, and W. Zhou, "SD-UNet: A structured dropout U-Net for retinal vessel segmentation," in *Proc. IEEE 19th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2019, pp. 439–444.

[80] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10727–10737.

[81] O. Sule and S. Viriri, "Enhanced convolutional neural networks for segmentation of retinal blood vessel image," in *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, Mar. 2020, pp. 1–6.

[82] Y. Zhang and A. C. Chung, "Deep supervision with additional labels for retinal vessel segmentation task," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Piscataway, NJ, USA: Springer, 2018, pp. 83–91.

[83] S. Mishra, D. Z. Chen, and X. S. Hu, "A data-aware deep supervised method for retinal vessel segmentation," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1254–1257.

[84] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 115–124.

[85] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019, doi: 10.1016/j.knosys.2019.04.025.

[86] 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)*, Oct. 2017, pp. 764–773.

[87] R. Estrada, C. Tomasi, S. C. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1688–1701, Aug. 2015.

[88] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Med. Image Anal.*, vol. 49, pp. 14–26, Jul. 2018.

[89] D. Wang, G. Hu, and C. Lyu, "FRNet: An end-to-end feature refinement neural network for medical image segmentation," *Vis. Comput.*, vol. 37, pp. 1101–1112, May 2021.

[90] P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep guidance network for biomedical image segmentation," *IEEE Access*, vol. 8, pp. 116106–116116, 2020.

[91] D. A. Dharmawan, D. Li, B. P. Ng, and S. Rahardja, "A new hybrid algorithm for retinal vessels segmentation on fundus images," *IEEE Access*, vol. 7, pp. 41885–41896, Mar. 2019.
[92] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[93] P. Xiuqin, Q. Zhang, H. Zhang, and S. Li, "A fundus retinal vessels segmentation scheme based on the improved deep learning U-Net model," *IEEE Access*, vol. 7, pp. 122634–122643, 2019, doi: 10.1109/ACCESS.2019.2935138.
[94] D. Li, D. A. Dharmawan, B. P. Ng, and S. Rahardja, "Residual U-Net for retinal vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1425–1429.
[95] T. M. Khan, M. Alhussein, K. Aurangzeb, M. Arsalan, S. S. Naqvi, and S. J. Nawaz, "Residual connection-based encoder decoder network (RCED-Net) for retinal vessel segmentation," *IEEE Access*, vol. 8, pp. 131257–131272, 2020.
[96] C. Guo, M. Szemenyei, Y. Yi, Y. Xue, W. Zhou, and Y. Li, "Dense residual network for retinal vessel segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1374–1378.
[97] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1392–1403, May 2020.
[98] R. Adarsh, G. Amarnageswarao, R. Pandeeswari, and S. Deivalakshmi, "Dense residual convolutional auto encoder for retinal blood vessels segmentation," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 280–284.
[99] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[100] A. P. Lopes, A. Ribeiro, and C. A. Silva, "Dilated convolutions in retinal blood vessels segmentation," in *Proc. IEEE 6th Portuguese Meeting Bioeng. (ENBENG)*, Feb. 2019, pp. 1–4.
[101] Y. Jiang, N. Tan, T. Peng, and H. Zhang, "Retinal vessels segmentation based on dilated multi-scale convolutional neural network," *IEEE Access*, vol. 7, pp. 76342–76352, 2019, doi: 10.1109/access.2019.2922365.
[102] R. Biswas, A. Vasan, and S. S. Roy, "Dilated deep neural network for segmentation of retinal blood vessels in fundus images," *Iranian J. Sci. Technol., Trans. Electr. Eng.*, vol. 44, no. 1, pp. 505–518, Mar. 2020, doi: 10.1007/s40998-019-00213-7.
[103] T. A. Soomro, A. J. Afifi, A. A. Shah, S. Soomro, G. A. Baloch, L. Zheng, M. Yin, and J. Gao, "Impact of image enhancement technique on CNN model for retinal blood vessels segmentation," *IEEE Access*, vol. 7, pp. 158183–158197, 2019.
[104] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861
[105] L. Lovász, "Random walks on graphs: A survey," *Combinat., Paul Erdos Eight*, vol. 2, no. 1, pp. 1–46, 1993.
[106] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[107] Z. Luo, Y. Zhang, L. Zhou, B. Zhang, J. Luo, and H. Wu, "Micro-vessel image segmentation based on the AD-UNet model," *IEEE Access*, vol. 7, pp. 143402–143411, 2019, doi: 10.1109/access.2019.2945556.
[108] S. Lian, L. Li, G. Lian, X. Xiao, Z. Luo, and S. Li, "A global and local enhanced residual U-Net for accurate retinal vessel segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 852–862, May 2021, doi: 10.1109/TCBB.2019.2917188.
[109] Y. Lv, H. Ma, J. Li, and S. Liu, "Attention guided U-Net with atrous convolution for accurate retinal vessels segmentation," *IEEE Access*, vol. 8, pp. 32826–32839, 2020, doi: 10.1109/access.2020.2974027.
[110] B. Wang, S. Wang, S. Qiu, W. Wei, H. Wang, and H. He, "CSU-Net: A context spatial U-Net for accurate blood vessel segmentation in fundus images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1128–1138, Apr. 2021, doi: 10.1109/JBHI.2020.3011178.
[111] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, Mar. 2021.
[112] K. Li, X. Qi, Y. Luo, Z. Yao, X. Zhou, and M. Sun, "Accurate retinal vessel segmentation in color fundus images via fully attention-based networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2071–2081, Jun. 2021.
[113] Q. Fu, S. Li, and X. Wang, "MSCNN-AM: A multi-scale convolutional neural network with attention mechanisms for retinal vessel segmentation," *IEEE Access*, vol. 8, pp. 163926–163936, 2020.
[114] X. Tang, B. Zhong, J. Peng, B. Hao, and J. Li, "Multi-scale channel importance sorting and spatial attention mechanism for retinal vessels segmentation," *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106353.
[115] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018.
[116] V. Nasery, K. B. Soundararajan, and J. Galeotti, "Learning to segment vessels from poorly illuminated fundus images," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1232–1236.
[117] A. Galdran, P. Costa, A. Bria, T. Araújo, A. M. Mendonça, and A. Campilho, "A no-reference quality metric for retinal vessel tree segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Piscataway, NJ, USA: Springer, 2018, pp. 82–90.
[118] D. E. Alvarado-Carrillo, E. Ovalle-Magallanes, and O. S. Dalmau-Cedeño, "D-GaussianNet: Adaptive distorted Gaussian matched filter with convolutional neural network for retinal vessel segmentation," *Geometry Vis.*, vol. 1386, p. 378, 2021.
[119] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "SCS-Net: A scale and context sensitive network for retinal vessel segmentation," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102025.
[120] J. I. Orlando, J. B. Breda, K. Van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant, "Towards a glaucoma risk index based on simulated hemodynamics from fundus images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Iowa City, IA, USA: Springer, 2018, pp. 65–73.
[121] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020, doi: 10.1109/JBHI.2020.3002985.
[122] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019, doi: 10.1109/JBHI.2018.2872813.
[123] V. Sathananthavathi and G. Indumathi, "Parallel architecture of fully convolved neural network for retinal vessel segmentation," *J. Digit. Imag.*, vol. 33, no. 1, pp. 168–180, Feb. 2020.
[124] L. Yang, H. Wang, Q. Zeng, Y. Liu, and G. Bian, "A hybrid deep segmentation network for fundus vessels via deep-learning framework," *Neurocomputing*, vol. 448, pp. 168–178, Aug. 2021.
[125] C. Tian, T. Fang, Y. Fan, and W. Wu, "Multi-path convolutional neural network in fundus segmentation of blood vessels," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 583–595, Apr. 2020.
[126] H. Xia, R. Zhuge, and H. Li, "Retinal vessel segmentation via a coarse-to-fine convolutional neural network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 1036–1039.
[127] K. Wang, X. Zhang, S. Huang, Q. Wang, and F. Chen, "CTF-Net: Retinal vessel segmentation via deep coarse-to-fine supervision network," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1237–1241.
[128] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "NFN +: A novel network followed network for retinal vessel segmentation," *Neural Netw.*, vol. 126, pp. 153–162, Jun. 2020.
[129] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, and J. Zhang, "S-UNet: A bridge-style U-Net framework with a saliency mechanism for retinal vessel segmentation," *IEEE Access*, vol. 7, pp. 174167–174177, 2019, doi: 10.1109/access.2019.2940476.
[130] Ü. Budak, Z. Cömert, M. Çıbuk, and A. Şengür, "DCCMED-net: Densely connected and concatenated multi encoder-decoder CNNs for retinal vessel extraction from fundus images," *Med. Hypotheses*, vol. 134, Jan. 2020, Art. no. 109426.
[131] G. A. Francia, C. Pedraza, M. Aceves, and S. Tovar-Arriaga, "Chaining a U-Net with a residual U-Net for retinal blood vessels segmentation," *IEEE Access*, vol. 8, pp. 38493–38500, 2020.
[132] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3656–3665.
[133] Y. Guo, Ü. Budak, and A. Sengür, "A novel retinal vessel detection approach based on multiple deep convolution neural networks," *Comput. Methods Programs Biomed.*, vol. 167, pp. 43–48, Dec. 2018.

[134] P. Tang, Q. Liang, X. Yan, D. Zhang, G. Coppola, and W. Sun, "Multi-proportion channel ensemble model for retinal vessel segmentation," *Comput. Biol. Med.*, vol. 111, Aug. 2019, Art. no. 103352.

[135] B. Zou, Y. Dai, Q. He, C. Zhu, G. Liu, Y. Su, and R. Tang, "Multi-label classification scheme based on local regression for retinal vessel segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Mar. 13, 2020, doi: 10.1109/TCBB.2020.2980233.

[136] V. Cherukuri, V. Kumar B. G., R. Bala, and V. Monga, "Deep retinal image segmentation with regularization under geometric priors," *IEEE Trans. Image Process.*, vol. 29, pp. 2552–2567, 2020.

[137] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101556.

[138] N. Tajbakhsh, B. Lai, S. P. Ananth, and X. Ding, "ErrorNet: Learning error representations from limited data to improve vascular segmentation," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1364–1368.

[139] W. Tu, W. Hu, X. Liu, and J. He, "DRPAN: A novel adversarial network approach for retinal vessel segmentation," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 228–232.

[140] C. Wu, Y. Zou, and Z. Yang, "U-GAN: Generative adversarial networks with U-Net for retinal vessel segmentation," in *Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2019, pp. 642–646.

[141] J. Ma, M. Wei, Z. Ma, L. Shi, and K. Zhu, "Retinal vessel segmentation based on generative adversarial network and dilated convolution," in *Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2019, pp. 282–287.

[142] Y. Dong, W. Ren, and K. Zhang, "Deep supervision adversarial learning network for retinal vessel segmentation," in *Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2019, pp. 1–6.

[143] Y. Zhou, Z. Chen, H. Shen, X. Zheng, R. Zhao, and X. Duan, "A refined equilibrium generative adversarial network for retinal vessel segmentation," *Neurocomputing*, vol. 437, pp. 118–130, May 2021.

[144] T. Yang, T. Wu, L. Li, and C. Zhu, "SUD-GAN: Deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation," *J. Digit. Imag.*, vol. 33, no. 4, pp. 946–957, 2020.

[145] X. Guo, C. Chen, Y. Lu, K. Meng, H. Chen, K. Zhou, Z. Wang, and R. Xiao, "Retinal vessel segmentation combined with generative adversarial networks and dense U-Net," *IEEE Access*, vol. 8, pp. 194551–194560, 2020.

[146] S. A. Rammy, S. J. Anwar, M. Abrar, and W. Zhang, "Conditional patch-based generative adversarial network for retinal vessel segmentation," in *Proc. 22nd Int. Multitopic Conf. (INMIC)*, Nov. 2019, pp. 1–6.

[147] J. He and D. Jiang, "Fundus image segmentation based on improved generative adversarial network for retinal vessel analysis," in *Proc. 3rd Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2020, pp. 231–236.

[148] J. Son, S. J. Park, and K.-H. Jung, "Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks," *J. Digit. Imag.*, vol. 32, no. 3, pp. 499–512, 2019.

[149] K.-B. Park, S. H. Choi, and J. Y. Lee, "M-GAN: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks," *IEEE Access*, vol. 8, pp. 146308–146322, 2020.

[150] Q. Huo, G. Tang, and F. Zhang, "Particle swarm optimization for great enhancement in semi-supervised retinal vessel segmentation with generative adversarial networks," in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*. Pakistan: Springer, 2019, pp. 112–120.

[151] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, Nov. 1995, pp. 1942–1948.

[152] A. Lahiri, V. Jain, A. Mondal, and P. K. Biswas, "Retinal vessel segmentation under extreme low annotation: A GAN based semi-supervised approach," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 418–422.

[153] L. Ngo and J.-H. Han, "Multi-level deep neural network for efficient segmentation of blood vessels in fundus images," *Electron. Lett.*, vol. 53, no. 16, pp. 1096–1098, 2017.

[154] J. Guo, S. Ren, Y. Shi, and H. Wang, "Automatic retinal blood vessel segmentation based on multi-level convolutional neural network," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.

[155] M. Li, Q. Yin, and M. Lu, "Retinal blood vessel segmentation based on multi-scale deep learning," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Sep. 2018, pp. 1–7.

[156] Y. Lin, H. Zhang, and G. Hu, "Automatic retinal vessel segmentation via deeply supervised and smoothly regularized network," *IEEE Access*, vol. 7, pp. 57717–57724, 2018, doi: 10.1109/access.2018.2844861.

[157] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, Oct. 2018.

[158] S. Guo, K. Wang, H. Kang, Y. Zhang, Y. Gao, and T. Li, "BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation," *Int. J. Med. Inform.*, vol. 126, pp. 105–113, Jun. 2019, doi: 10.1016/j.ijmedinf.2019.03.015.

[159] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," *Neurocomputing*, vol. 392, pp. 268–276, Jun. 2020, doi: 10.1016/j.neucom.2018.10.098.

[160] Z. Zhuo, J. Huang, K. Lu, D. Pan, and S. Feng, "A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105508.

[161] K. J. Noh, S. J. Park, and S. Lee, "Scale-space approximated convolutional neural networks for retinal vessel segmentation," *Comput. Methods Programs Biomed.*, vol. 178, pp. 237–246, Sep. 2019, doi: 10.1016/j.cmpb.2019.06.030.

[162] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403.

[163] H. M. Wallach, "Conditional random fields: An introduction," Dept. Comput. Inf. Sci., Univ. Pennsylvania, Philadelphia, PA, USA, Tech. Rep. MS-CIS-04-21, 2004, p. 22.

[164] K. Aurangzeb, S. Aslam, M. Alhussein, R. A. Naqvi, M. Arsalan, and S. I. Haider, "Contrast enhancement of fundus images by employing modified PSO for improving the performance of deep learning models," *IEEE Access*, vol. 9, pp. 47930–47945, 2021.

[165] A. Sengur, Y. Guo, U. Budak, and L. J. Vespa, "A retinal vessel detection approach using convolution neural network," in *Proc. Int. Artif. Intell. Data Process. Symp. (IDAP)*, Sep. 2017, pp. 1–4.

[166] J. Song and B. Lee, "Development of automatic retinal vessel segmentation method in fundus images via convolutional neural networks," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 681–684.

[167] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, P. S. Heckbert Ed. New York, NY, USA: Academic, 1994, pp. 474–485.

[168] H. Hassanpour, N. Samadiani, and S. M. M. Salehi, "Using morphological transforms to enhance the contrast of medical images," *Egyptian J. Radiol. Nucl. Med.*, vol. 46, no. 2, pp. 481–489, 2015, doi: 10.1016/j.ejrnm.2015.01.004.

[169] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A. C. Berg, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[170] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

CHUNHUI CHEN received the B.S. degree in energy and power engineering from Jiangsu University (UJS), Zhenjiang, China, in 2018. He is currently pursuing the Master of Industrial Electronics and Control Engineering degree with the University of Malaya (UM), Malaysia. His research interests include deep learning, image processing, and computer vision.

JOON HUANG CHUAH (Senior Member, IEEE) received the B.Eng. degree (Hons.) from the Universiti Teknologi Malaysia, the M.Eng. degree from the National University of Singapore, and the M.Phil. and Ph.D. degrees from the University of Cambridge. He is currently the Head of the VIP Research Group, Department of Electrical Engineering, Faculty of Engineering, University of Malaya. His research interests include image processing, computational intelligence, IC design, and scanning electron microscopy. He is a Chartered Engineer registered under the Engineering Council, U.K., and also a Professional Engineer registered under the Board of Engineers, Malaysia. He was the Honorary Treasurer of the IEEE Computational Intelligence Society (CIS) Malaysia Chapter and the Honorary Secretary of the IEEE Council on RFID Malaysia Chapter. He is also the Vice Chairman of the Institution of Engineering and Technology (IET) Malaysia Network. He is a fellow and the Honorary Secretary of the Institution of Engineers, Malaysia (IEM).

RAZA ALI (Member, IEEE) received the B.S. degree in telecommunication engineering from Balochistan University of IT, Engineering and Management Sciences (BUITEMS), Quetta, Pakistan, in 2009, and the master's degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2015. Since January 2013, he has been working as a Faculty Member with the Faculty of Information and Communication Technology, BUITEMS. He is currently a Ph.D. Researcher with the VIP Laboratory, University of Malaya, Malaysia. He is a Registered Engineer under the Pakistan Engineering Council. He is a Coordinator for awards and grants in IEEE Malaysia SAC (2020). His research interests include signal processing, computer vision, and deep learning for semantic segmentation.

YIZHOU WANG (Senior Member, IEEE) received the bachelor's degree in electrical engineering from Tsinghua University, in 1996, and the Ph.D. degree in computer science from the University of California at Los Angeles (UCLA), in 2005. From 2005 to 2007, he was a Research Staff with Xerox Palo Alto Research Center (Xerox PARC). He is a BoYa Professor with the Computer Science Department and the Vice Director of the Center on Frontiers of Computing Studies, Peking University. His research interests include computational vision, statistical modeling and learning, and medical image analysis.