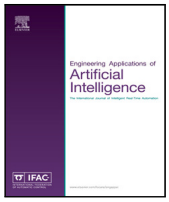
Engineering Applications of Artificial Intelligence 128 (2024) 107454

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Survey paper

A review of retinal vessel segmentation for fundus image analysis

Qing Qin, Yuanyuan Chen *

College of Computer Science, Sichuan University, No. 24 South Section 1, Yihuan Road, Chengdu, 610065, China

ARTICLE INFO

Keywords:
Medical imaging
Fundus images
Retinal vessels
Segmentation

ABSTRACT

The fundus is the only region where arteries, veins and capillaries can be directly observed. Morphological changes of retinal vessels in the fundus are signals for the appearance of many fundus and cardiovascular diseases. Consequently, the segmentation of retinal vessels is crucial for diagnosing and screening various diseases. In recent years, a large amount of research publications has been published on retinal vessel segmentation. This paper offers a comprehensive review of retinal vessel automatic segmentation research, covering both traditional methods and deep learning methods, including unsupervised and supervised learning methods. Especially, the supervised learning methods are summarized and analyzed from three aspects: traditional CNN-based, GAN-based, and UNet-based methods. This paper also presents an overview of the development of retinal vessel automatic segmentation and analyzes the advantages and disadvantages of existing segmentation methods. The results are shown in two at-a-glance tables. Finally, our work provides a faster and better look to recognize and understand the field of retinal vessel segmentation.

1. Introduction

In recent years, the prevalence of visual impairment due to eye diseases and related causes has been on the rise, making eye health a prominent area of concern. According to surveys, at least 2.2 billion individuals globally suffer from visual impairment, at least 1 billion of whom are impacted by preventable diseases or treatable conditions (Organization et al., 2019). Consequently, early screening of fundus diseases, timely detection, and early treatment are essential for addressing these issues.

The diagnosis and screening of fundus disease are achieved by examining fundus images, extracting information from these images, and conducting analyses. The fundus, located at the back of the eye, comprises the retina, optic disk, macula, central retinal artery, and vein. It is the only region where arteries, veins and capillaries can be directly observed (Chen et al., 2021), which can reflect the dynamics and health of blood circulation throughout the body.

The morphological changes of retinal blood vessels are often the basis for diagnosing certain diseases. Retinal vessel segmentation is the basis for studying and analyzing the retinal vessel morphology, such as vessel thickness and curvature. This has become a critical technology to implement the intelligent disease screening and computer-aided diagnosis. For example, the degree of narrowing in retinal blood vessel arterioles reflects the condition of the peripheral blood vessels and can also indicate the presence of cerebrovascular and coronary artery vascular diseases. Once the vessels are separated, the caliber of the retinal blood vessels can be measured by calculation, which can assist in the diagnosis of cardiovascular diseases, including atherosclerosis, coronary heart disease, stroke, etc. In addition, screening for many fundus diseases can be conducted based on changes in retinal vessel structure. Diabetic retinopathy (DR) patients with high blood sugar content and high viscosity may experience retinal hypoxia, along with fundus and vitreous hemorrhage, which can lead to retinopathy. Clinical manifestations of DR include capillary dilation, venous twist and dilation, and tube diameter inequality (Anon, 2021). Hypertensive retinopathy exhibits clinical manifestations such as retinal artery vascular stenosis and a change from an acute to a blunt angle at large branches of arteries (Anon, 2021). Retinopathy of prematurity (ROP) is a proliferative retinopathy occurring in immature or low-weight infants. Patients with ROP present roundabout and dilated arteries and veins, with brush-like capillaries visible at the peripheral vessel endings of the retina (Anon, 2019). Rapid development of glaucoma can also cause a shift in the position of the retinal blood vessels (Anon, 2021). Therefore, the analysis of diameter, branch angle, branch length, and tortuosity of retinal blood vessels has become crucial in the diagnosis and screening of cardiovascular diseases, such as atherosclerosis and coronary heart disease, along with DR, hypertensive retinopathy, ROP, and other fundus diseases. Retinal vessel segmentation is the basis for analyzing and studying retinal blood vessels structures.

However, retinal vessel segmentation is different from other medical image segmentation tasks. Firstly, blood vessels are smaller than other

* Corresponding author.
E-mail address: chenyuanyuan@scu.edu.cn (Y. Chen).

https://doi.org/10.1016/j.engappai.2023.107454
Received 25 October 2022; Received in revised form 13 July 2023; Accepted 4 November 2023
Available online 20 November 2023

segmentation targets, making imaging quality crucial for segmentation performance. And retinal vessel segmentation requires specialized fundus cameras to capture the fundus image. Therefore, the specific camera used and external shooting conditions may impact the quality of the fundus image (noise, contrast, illumination, blur, etc.), directly affect the retinal vessel segmentation performance. Additionally, retinal blood vessels are numerous, small, curved, and variably shaped, and can be parallel or intersecting. This often results in issues such as missed detection of poorly contrasted small vessels, broken vessels at bifurcations or crossover points, and segmentation of parallel vessels into a large one or merging of two close vessels into a single one. These factors make retinal vessel segmentation more difficult than other medical image segmentation tasks. Moreover, retinal lesions, such as exudates, cotton wool spots, hemorrhages, and neovascularization, are common interference information in retinal vessel segmentation. All these factors make retinal vessel segmentation a challenging task. The key to successful retinal vessel segmentation lies in a method that can segment small blood vessels under uneven illumination, low contrast and various complexities within fundus images, resolve the issue of broken blood vessel, and exhibit strong generalization ability. Due to the difficulty and urgency of retinal vessel segmentation, methods for performing this task are continuously optimized and developed. The development timeline is shown in Fig. 1.

In the early days, ophthalmologists mainly segmented retinal vessels manually, a process that was laborious, time-consuming, and costly. Manual segmentation not only demands significant human resources, but also requires staff to possess professional knowledge and clinical experience, resulting in a severe waste of experts' time and energy. Furthermore, manual segmentation is inefficient, making it impossible to carry out large-scale fundus disease screening. In response to these issues, artificial intelligence-assisted fundus disease screening methods have garnered increasing attention from scholars, leading to the gradual development of automatic retinal vessel segmentation methods.

The initial automatic retinal vessel segmentation methods were predominantly unsupervised, meaning that they did not rely on any labeled data. These methods operated by extracting vessels and background features in fundus images, discovering their interrelationships, and subsequently segmenting the vessels. The most common approaches included matched filters, vessel tracking, and morphological reconstruction-based methods. However, from the perspective of segmentation performance, unsupervised methods have certain limitations. Firstly, due to the absence of guidance provided by labeled data, they are entirely reliant on the intrinsic image characteristics and the feature information extracted from these images, resulting in limited segmentation accuracy. Moreover, unsupervised methods still require some degree of human intervention (such as selecting features, setting filters, etc.), resulting in weak generalizability.

Compared with unsupervised methods, supervised methods add labeled data (usually provided by experts) to train the model. Supervised methods involve two stages: feature extraction and classification. Feature extraction can be divided into manually selected features and automatically learned features. In traditional methods, features are manually selected from fundus images, followed by pixel-by-pixel classification by the classifier based on the extracted features to achieve retinal vessel segmentation. However, these methods are relatively subjective in feature selection, require professional expertise, and are unsuitable for large-scale fundus retinal vessel segmentation. As deep learning techniques have developed, they have been increasingly applied in image processing and analysis. Deep learning methods can automatically learn features from a large amount of data, reducing the need for human intervention, and improving the generalization and accuracy of automatic segmentation methods.

In 2012, AlexNet (Krizhevsky et al., 2012), proposed by Krizhevsky et al. won the ImageNet competition and marked a significant milestone in the resurgence of interest in neural networks. This event is widely considered as the catalyst that triggered the rapid development and widespread adoption of deep learning in various fields. Building on the success of AlexNet, subsequent breakthroughs were achieved in deep learning with the introduction of VGG network (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2015), and ResNet (He et al., 2016), gaining renewed attention in the research community. In 2014, Long et al. proposed the fully convolutional network (FCN) (Long et al., 2015), which abandoned fully connected layers and instead used deconvolution layers to upsample the feature maps of the last convolution layer. This enabled the feature maps to be gradually restored to the same size as the input image, facilitating pixel-level classification of the image. FCN solves the problem of semantic-level image segmentation, which has enabled semantic segmentation to enter a new field and has served as the basis for a large number of retinal vessel automatic segmentation models (Dasgupta and Singh, 2017; Feng et al., 2017; Hu et al., 2018; Oliveira et al., 2018; Yan et al., 2018). In the same year, Goodfellow et al. proposed the Generative Adversarial Network (GAN) (Goodfellow et al., 2014), which laid the foundation for the proposal of a semi-supervised model for retinal vessel segmentation, and also provided a new solution to address the issue of lower data availability for retinal vessel segmentation. In 2015, Ronneberger et al. proposed UNet (Ronneberger et al., 2015), a U-shaped network with skip connections based on a fully convolutional network, referred to as . Due to its accurate segmentation and good performance, a large number of variant models based on UNet (Alom et al., 2019; Oktay et al., 2018; Xiao et al., 2018; Gu et al., 2019; Guo et al., 2019; Jin et al., 2019; Laibacher et al., 2019; Li et al., 2019; Zhuang, 2018; Li et al., 2020; Wang et al., 2020; Du et al., 2021; Guo et al., 2021; Hu and Liu, 2021; Wu et al., 2021; Dong et al., 2022; Huang et al., 2022; Li et al., 2022a; Shen et al., 2022) have been proposed, leading to another new stage for retinal vessel segmentation. In 2016, Cambridge proposed SegNet (Badrinarayanan et al., 2017), which is based on the FCN and VGG networks. Unlike UNet, SegNet uses a pooling index to perform upsampling through the inverse process of pooling in the decoder part, meaning that the upsampling process does not require learning. In the same year, the idea of GAN was applied to semantic segmentation (Luc et al., 2016), and the concept of structured prediction was proposed (Liskowski and Krawiec, 2016). In 2017, Lahiri et al. used GAN for retinal vessel segmentation (Lahiri et al., 2017), in an attempt to alleviate the work of manually labeling retinal vessels in a semi-supervised manner. With the development of neural networks, especially the rise of DNN, network depth has increased and various new networks have been proposed, which has enabled the accuracy of vessel segmentation to improve; at present the accuracy of vessel segmentation on the DRIVE dataset has reached 97.9% (Kamran et al., 2021).

The remainder of this paper is structured as follows. In Section 2, we present the databases for retinal vessel segmentation. Section 3 discusses and analyzes existing retinal vessel segmentation models. In Section 4, we introduce the evaluation metrics of retinal vessel segmentation models and compare the performance of the state-of-the-art models. Finally, Section 5 present the discussion before concluding this paper.

2. Datasets

The fundus, which includes the optic disk, blood vessels, retinal tissue, and choroid, offers a unique opportunity for direct observation of arteries, veins, and capillaries with the naked eye. To facilitate scientific research, specialized fundus cameras are used to capture fundus images for retinal vessel segmentation. These images are processed in a straightforward manner to create datasets that can be applied to further analysis and investigation.

In recent years, numerous retinal vessel segmentation datasets have been developed. The most commonly publicly available datasets for retinal vessel segmentation are DRIVE, STARE, and CHASEDB1. While MESSIDOR, ARIR, REVIEW, HRF, and others are also suitable for retinal

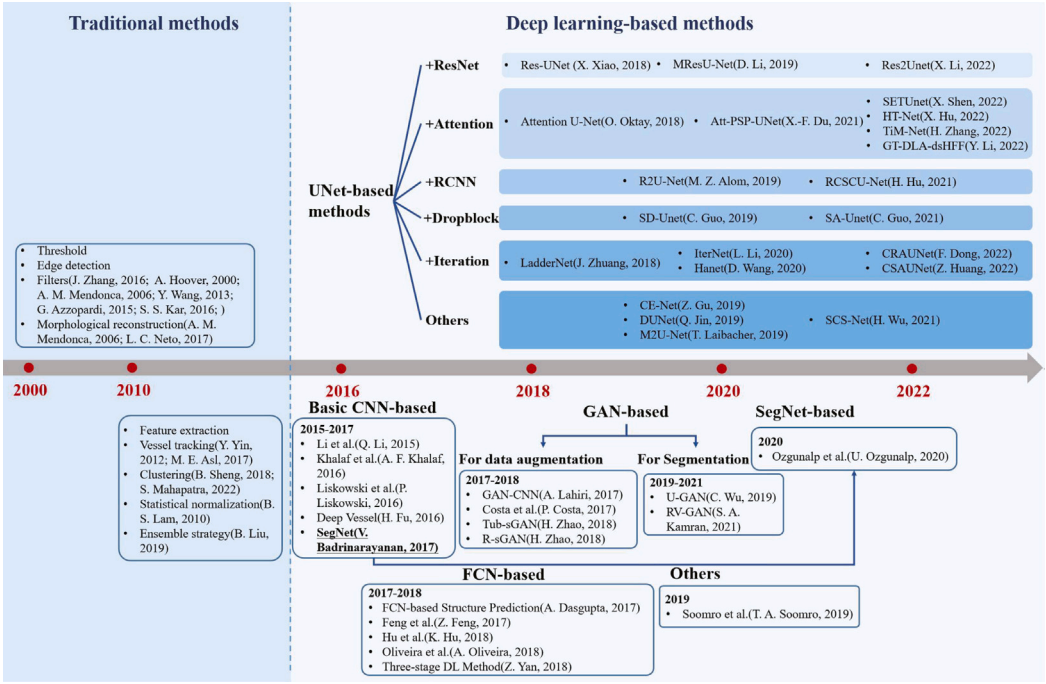


Fig. 1. History of the automatic segmentation of retinal vessels.

Table 1
Details of retinal vessel segmentation datasets.

Database	Year	Image	Resolution	Abnormal	Total	Patients	Labels	FOV	Format	SOTA Method	F1-Score
STARE	2000	Fundus	605 × 700	10	400	–	20	35°	PPM	R2U-Net (Alom et al., 2019)	0.8475
DRIVE	2004	Fundus	768 × 584	7	40	–	40	45°	JPEG	RV-GAN (Kamran et al., 2021)	0.8690
CHASE_DB1	2011	Fundus	1280 × 960	0	28	14	28	35°	TIF	RV-GAN (Kamran et al., 2021)	0.8957
MESSIDOR	2004	Fundus	1440 × 960 2240 × 1488 2304 × 1536	660	1200	–	1200	45°	TIF	Ding et al. (2015)	–
ARIR	2006	Fundus	768 × 576	82	143	–	143	50°	TIF	–	–
DIARRTDB0	2007	Fundus	1500 × 1150	20	130	–	0	50°	PNG	–	–
DIARETDB1	2007	Fundus	1500 × 1150	84	89	–	0	50°	PNG	Kar and Maity (2016b)	–
REVIEW	2008	Fundus	3584 × 2438 1360 × 1024 2160 × 1440 3300 × 2600	–	16	–	16	–	–	Jebaseeli et al. (2019)	–
HRF	2009	Fundus	3504 × 2336	30	45	45	45	45°	JPEG	VGN (Shin et al., 2019)	0.8151
ROC	2009	Fundus	1058 × 1061 1389 × 1383 768 × 576	–	100	–	–	45°	JPEG	–	–
DRiDB	2013	Fundus	720 × 576	–	50	–	0	45°	BMP	–	–
LES-AV	2018	Fundus	1620 × 1444 1958 × 2196	1	22	–	22	30°/45°	PNG	–	–
VAMPIRE	2011	FA	3900 × 3072	4	8	–	8	200°	–	Zhao et al. (2015)	0.7370
IOSTAR	2015	SLO	1024 × 1024	0	30	–	30	45°	JPEG	Zhang et al. (2016)	–
RC-SLO	2015	SLO	360 × 320	0	40	–	40	–	TIF	Srinidhi et al. (2018)	0.72

vessel segmentation, they have rarely been used in recent years, partly because some of them are not available. Additionally, a few fundus laser image datasets such as IOSTAR and RC-SLO that can use for retinal vessels segmentation. With the development of ultra-wide field (UWF) fundus color imaging technology, some ultra-wide field fundus datasets like VAMPIRE have also emerged. Compared with traditional fundus color imaging, which provides 30–75° views, UWF imaging can offer a wider view up to 200°, providing more comprehensive information. Table 1 provides further details on these datasets.

The Structured Analysis of the Retina (STARE) (Hoover et al., 2000) database consists of 400 retinal images, including 20 labeled images that can be used for retinal blood vessel segmentation. Among them, 10 fundus images are slightly healthy with no signs of disease, and the others show retinal diseases of varying severity. The images are captured by a TopCon TRV-50 fundus camera with a 35-degree field of view (FOV).

Digital Retinal Images for Vessel Extraction (DRIVE) (Staal et al., 2004) is a database containing a random selection of 40 images from

a diabetic retinopathy screening program in a population aged 25–90 years of age in the Netherlands. The images are acquired with a Canon CR5 non-mydriatic 3CCD camera with a 45-degree FOV. Seven out of the included 40 retinal fundus images show signs of mild early diabetic retinopathy. The 40 images in the DRIVE dataset have been divided into a training set and a test set, each including 20 images.

The CHASE_DB1 (Owen et al., 2009) database contains data derived from the Child Health and Heart Studies in England. The database consists of left and right fundus photographs of 14 children, for a total of 28 fundus images. None of the 28 fundus images in this dataset showed signs of disease.

Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology within the Scope of Diabetic Retinopathy (MESSIDOR) (Decencière et al., 2014) is a project of the French Ministry of Research and Defense and the largest retinal dataset currently available. The images in MESSIDOR are captured by three ophthalmology departments using non-mydriatic Topcon TRC NW6 cameras.

The Automated Retinal Image Analysis (ARIR) (Farnell, 2007; Farnell et al., 2008) database was collected in the UK from 2004 to 2006, and the images are captured with a Zeiss fundus camera with a 50-degree FOV. The database consists of 143 fundus images: of these, 23 are abnormal fundus images with AMD, 59 are fundus images with DR, and the rest are a control group of healthy retinal fundus images.

ImageRet (Kälviäinen et al., 2007; Kauppi et al., 2006) is a project of the FinnWell Technology Program for diabetic retinopathy coordinated by the Finnish Funding Agency for Technology and Innovation, Tekes. The database was officially published in Tampere on January 14, 2008, and is divided into two datasets, DRARETDB0 and DRARETDB1; see Table 1 for details.

Retinal Vessel Image Set of Estimation of Widths (REVIEW) (Al-Diri et al., 2008) was established by the University of Lincoln in 2008 for retinal vessel width estimation. The database is divided into four sets, namely HRIS (high-resolution image set), VDIS (vascular disease image set), CLRIS (central light reflection image set) and KPIS (kick point image set); all images have a higher resolution than those in DRIVE.

The High Resolution Fundus (HRF) (Budai et al., 2013) database was constructed by a collaborative research group. The entire dataset has 45 fundus images, including 15 healthy images, 15 diabetic retinopathy patient images, and 15 glaucoma patient images. The images were taken with a Canon CR-1 fundus camera with a 45-degree FOV.

RetinoPathy Online Challenge (ROC) (Niemeijer et al., 2009) includes 100 images obtained when a microtumor detection contest was held at the University of Iowa in 2009. The database is divided into two sets, a test set and a training set, with 50 images per set. These images are acquired with a Topcon NW100 and a Canon CR5-45NM underside camera with a 45-degree FOV.

The Diabetic retinopathy image dataset (DRiDB) (Prentašić et al., 2013) was created to help scientists from around the world test and develop new image processing methods. The database consists of fifty retinal images acquired with a 45-degree FOV; see Table 1 for details.

LES-AV (Orlando et al., 2018) includes 22 images, including 1 pathological image with a resolution of 1958×2196 and 21 normal fundus images with a resolution of 1620×1444. It also includes manual annotations that distinguish the corresponding vessels of arteries and veins and specifies whether an image is of a healthy eye or one with glaucoma.

Vascular Assessment and Measurement Platform for Images of the Retina (VAMPIRE) (Perez-Rovira et al., 2011) includes eight ultra-wide-field retinal angiography images acquired using an OPTOS P200C camera. Four of the images are of pathological retinas, while the other four are of healthy retinas.

IOSTAR (Zhang et al., 2016) is a vessel segmentation database of Scanning Laser Ophthalmoscopy (SLO) images, captured by an EasyScan camera based on SLO technology with a 45-degree FOV. It consists of 30 images in total, which have been annotated by two different experts.

The RetinaCheck-Scanning Laser Ophthalmoscopy (RC-SLO) (Abbasi-Sureshjani et al., 2015) database contains 40 image patches acquired by the EasyScan camera, which utilizes an SLO technique. The images in RC-SLO include most of the difficult cases for vessel segmentation, such as high curvature changes, central vessel reflex, microvessels, crossings/bifurcations, and background artifacts.

As the field of retinal vessel segmentation continues to expand and technology develops, the datasets for retinal vessel segmentation are also constantly developing and expanding. Both in terms of quantity and image quality, these datasets show a gradual upward trend. Furthermore, due to the increase in the number of imaging methods, the diversity of image data is also expanding to include laser scanning imaging and fluorescein fundus angiography. These developments will significantly aid in the growth and advancement of the field of retinal vessel segmentation.

3. Methods

The automatic segmentation of retinal vessels has garnered significant interest among researchers, leading to the development of numerous distinct methods for performing this task. In order to improve the accuracy of the segmentation methods, scholars have conducted a series of explorations into both the segmentation model and the image preprocessing operations. In this section, we will introduce the data preprocessing methods commonly employed in retinal fundus vessel segmentation and discuss existing automatic segmentation methods from the perspectives of unsupervised learning and supervised learning, respectively.

3.1. Fundus image preprocessing

Retinal fundus image datasets are often characterized by small sample sizes, limited label information, and characterized by inconsistent image quality. Fundus images generally contain substantial noise and exhibit extremely low contrast, making it challenging to accurately segment vessels from the background. Therefore, in order to reduce the influence of these factors and achieve more precise vessel segmentation, a range of preprocessing operations (e.g., contrast enhancement, noise reduction) are typically performed on the original images. Preprocessing methods commonly used in retinal vessel segmentation include geometric transformation (Shen et al., 2022; Xiao et al., 2018; Hu and Liu, 2021; Wu et al., 2021; Guo et al., 2019), grayscale processing (Dong et al., 2022; Wang et al., 2020; Li et al., 2019; Shen et al., 2022; Xiao et al., 2018; Jin et al., 2019; Hu and Liu, 2021), gamma correction (Wang et al., 2020; Jin et al., 2019; Hu and Liu, 2021; Dong et al., 2022), limited contrast histogram equalization (Wang et al., 2020; Li et al., 2019; Shen et al., 2022; Xiao et al., 2018; Jin et al., 2019; Hu and Liu, 2021; Dong et al., 2022), among others.

Geometric Transformation. Given that most extant datasets suitable for retinal vessel segmentation suffer from an insufficient number of labels, data augmentation techniques such as cropping, rotating, horizontal flipping, and vertical flipping are frequently employed to bolster the dataset, thereby ensuring effective model learning and preventing overfitting issues.

Grayscale Processing. Owing to the exceedingly low contrast of retinal fundus images, vessel segmentation is a difficult task. During image preprocessing, fundus images are usually converted to grayscaled, which not only enhances contrast but also bolsters the model's operational speed. Grayscale processing entails transforming a color RGB image into a grayscale image. Widely used image grayscale processing methods include the component method (Hu and Liu, 2021; Jin et al., 2019) (one of the three components in the color image is selected as the image gray value, and after comparison, it is found that the retinal vessels and the background in the green channel have better contrast),

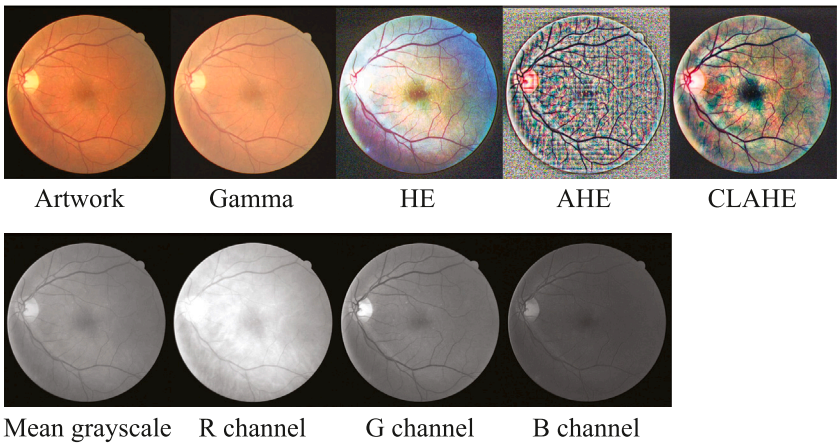


Fig. 2. Examples of preprocessing methods applied to a fundus image.

the maximum value method (the maximum value of the brightness of the three components in the color image is used as the gray value of the grayscale image), the averaging method (the brightness of the three components of the color image is averaged to obtain a gray value), and the weighted average method (according to the importance and indicators, the three components are weighted and averaged with different weights to obtain the final grayscale image).

Gamma Correction. Gamma correction is a method of editing the gamma curve of an image to perform non-linear tone editing on the image. This process increases the proportion of light and dark areas within the image signal, thereby enhancing image contrast. Gamma correction is also employed as a preprocessing technique for contrast enhancement.

Limited Contrast Histogram Equalization (CLAHE). CLAHE is a variant of histogram equalization(HE), which aims to expand the gray level range of the image or make the gray level evenly distributed, thereby increasing contrast and making the image details clearer so as to achieve the purpose of image enhancement. However, since HE cannot effectively enhance obvious bright and dark within the image, adaptive histogram equalization (AHE) was developed. AHE calculates the histogram for multiple local image region, redistributes brightness, and alters image contrast. While AHE is more suitable for improving local contrast image details, it can also excessively magnifies noise of relatively uniform areas of the image. The benefit of CLAHE is that it applies contrast limitation for each small area block, which overcomes the issue of over-amplified caused by AHE and now is widely adopted.

As shown in Fig. 2, we randomly selected a fundus image from the datasets and performed a series of image preprocessing operations. As Fig. 2 clearly shows, CLAHE is the operation that makes the feature clearest in the fundus image, while in the grayscale operation, the G channel grayscale most effectively distinguishes vessels and background within the fundus image. Consequently, these preprocessing operations will be employed in our experiments. In addition to the above preprocessing operations, generative networks are also occasionally used to generate training data for data augmentation during the preprocessing stage. These methods will be discussed in detail in Section 3.3.2.

3.2. Unsupervised method

Due to the scarcity of available data, especially labeled data, initial automatic retinal vessel segmentation methods predominantly employed unsupervised learning methods. These methods do not rely on label information; instead, they achieve vessel segmentation by extracting vessel and background features within fundus images and discerning their relationships. Early methods used edge detection algorithm to detect vessels outlines and subsequently attempted segmentation. However, features extracted through edge detection algorithms sometimes failed to represent vessel characteristics or were not suitable for distinguishing vessels from the background, resulting in poor segmentation results. Subsequently, matched filter-based segmentation methods emerged.

The matched filter attempts to maximize the power ratio between the positive and negative classes at the output, thereby enabling image segmentation to be realized after filtering. In 1989, Chaudhuri et al. (1989) proposed a method for retinal vessel detection using a two-dimensional matched filter. Their method extracted features based on the optical and spatial characteristics of retinal fundus images approximating the grayscale profile of blood vessel cross-sections with Gaussian-shaped curves. The matched filter of the signal was employed to detect pixel-level linear segments of blood vessels, followed by a search for vessel segments in all possible directions to accomplish vessel detection. Later, Hoover et al. (2000) and Al-Rawi et al. Al-Rawi et al. (2007) successively proposed the use of matched filters for vessel segmentation. Hoover et al. combined the local and global features of vessels during vessel segmentation, while Al-Rawi et al. improved the segmentation effect by using better parameters. In 2013, Wang et al. (2013) used the matched filter of the multi-wave tube kernel to achieve automatic blood vessel segmentation, specifically removing non-vessels boundary portions while segmenting vessels, which further improved segmentation performance. Subsequently, in 2015, Azzopardi et al. (2015) introduced a new filter method that selectively responds to blood vessels based on the COSFIRE (Combination Of Shifted Filter Responses) method. Due to the satisfactory results achieved by filters in vessel segmentation, this approach has become widely adopted in automatic retinal vessel segmentation. Even after the development of supervised learning, a large number of researchers continue to explore filter-based segmentation methods (Kar and Maity, 2016a; Zhang et al., 2016; Rezaee et al., 2017; Zhao et al., 2017). In 2016, Zhang et al. (2016) proposed a novel filter based on a 3D rotation framework, achieving a comprehensive evaluation/benchmark for vessel segmentation through left-turn processing of orientation scores. Their methods represented a huge improvement in handling low-resolution data. Subsequently, Zhao et al. (2017) proposed an unsupervised segmentation method applicable to both 2D and 3D images.

Filter-based methods are a large branch of unsupervised methods for automatic retinal vessel segmentation. However, these methods derive vessel and background segmentation from pixel value differences without considering inherent vessel characteristics such as curvature, complex shapes, and the challenges associated with identifying tiny vessels. Therefore, these methods fail to satisfy the accuracy requirements of segmentation. In addition to filter-based methods, mainstream automatic retinal vessel segmentation methods focusing on the characteristics of the blood vessels themselves include methods based on

vessel tracking (Yin et al., 2012; Asl et al., 2017) and morphological reconstruction (Mendonca and Campilho, 2006; Neto et al., 2017).

In 2006, Mendonca et al. (Mendonca and Campilho, 2006) employed the vessel centerline algorithm to achieve vessel segmentation based on vessel morphology enhancement and morphology reconstruction. In 2012, Yin et al. (2012) proposed a method using probabilistic tracking based on the blood vessel tracking algorithm. Vessel tracking algorithms involve manual selection of initial points followed by iterative vessel edge detection. In the paper published by Yin et al. local vessel sectional intensity profiles were estimated using Gaussian-shaped curves, after which a Bayesian method with the maximum a posteriori (MAP) probability criterion was applied to identify local vessel structures and extract edge points from these candidate templates. However, vessel tracking methods often encounter difficulties with blood vessel gaps and discontinuities. To address the issue of thin vessels, the vessel tracking method proposed by Asl et al. (Asl et al., 2017) made great progress in processing such vessels and detecting vessel branches. In the same year, Neto et al. (2017) created a curvature map by combining rough segmentation contours obtained in the previous stage during morphological reconstruction, resulting in obtaining finer segmentation results and implementing a coarse-to-fine vessel segmentation method.

Besides the above mainstream methods, numerous researchers have attempted to solve vascular issues through traditional methods. Considering the influence of lesions on vessel segmentation, Lam et al. proposed a statistical normalization method (Lam et al., 2010) for retinal vessel segmentation that utilizes the concave and convex features of lesions, vessels, and the background to achieve segmentation. Sheng et al. (2018) employed superpixels as the fundamental unit for vessel segmentation, implementing clustering to achieve vessel and background segmentation. Taking advantage of different unsupervised learning methods, Liu et al. proposed a novel unsupervised ensemble policy network, the no-reference evaluation network (Liu et al., 2019), which works by fusing the segmentation results of multiple networks. Although unsupervised methods cannot benefit from label information guidance, the scarcity of annotated medical imaging data remains a prevalent issue in medical imaging. Therefore, continued research in unsupervised learning methods is still of great importance and warrants scholarly attention. In recent years, Mahapatra et al. presented a novel clustering technique for segmentation (Mahapatra et al., 2022) that utilized vessel connectivity and considered pixel spatial information to realize automatic retinal vessel segmentation.

The unsupervised methods mentioned above are mainly traditional methods, including signal processing methods (e.g., filters, edge detection) and other traditional methods. Although unsupervised methods have made great progress on the retinal vessel segmentation task, they still have limitations that are difficult to overcome. Unsupervised methods extract useful features from the image data without guidance from label information, leading to increased training difficulty and complexity. Moreover, most unsupervised methods require human involvement in the training process (such as filter setting, mode selection, etc.), which leads to poor model generalizability. Therefore, supervised methods generally outperform unsupervised methods.

3.3. Supervised methods

Supervised learning methods involve labeling image data based on specific criteria and classifying the image data according to these labels, essentially utilizing label data to guide model training. Widely employed in image classification tasks, supervised learning is also applicable to segmentation, which is pixel-level image classification. Retinal vessel segmentation is achieved through binary classification of vessels and background for each fundus image pixel. In supervised learning methods, there are also many traditional methods for vessel segmentation, such as Ghaderi et al. (2007). However, the supervised learning method introduced in this paper is mainly based on the neural network method, which is a mathematical model or computational model that imitates the structure and function of biological neural networks, and is used to estimate or approximate functions.

Significant developments in neural network research include the proposal of CNNs (Convolutional Neural Networks) and DNNs (Deep Neural Networks), both capable of establishing high nonlinear mapping relationship between data and labels through training, thereby achieving remarkable results in classification tasks. Accordingly, DNNs have been applied in various mainstream methods in medical image segmentation. The evolution of supervised learning-based automatic retinal vessel segmentation methods has progressed through several stages, from traditional CNNs to fully convolutional networks, GAN-based networks, and UNet-based networks. This section will introduce existing automatic retinal vessel segmentation methods from the perspectives of traditional CNN-based, GAN-based, and UNet-based. Additionally, due to the large size of retinal fundus images and the limited ability of equipment, current retinal vessel segmentation methods typically divide the entire image into several patches during training. These patches are segmented and predicted separately before being reassembled to reconstruct the original image.

Formally, let X represent the vector form of all pixels in a color fundus image, and Y denote the vector form of the associated label, which corresponds to the segmentation results. Y^+ and Y^- signify vessel pixels and background pixels in the label, respectively. Finally, let W symbolize the network parameters and $\sigma(\cdot)$ represent the activation function.

3.3.1. Traditional CNN-based methods

Sometimes, a fundus image is cut into multiple small patches, which are segmented individually before being stitched together to reconstruct the original fundus image. This strategy of cutting the fundus image into smaller patches allows the model to read the image more quickly and can also enhance segmentation performance to a certain extent. In this paper, assuming that the length and width of the fundus image are equal and defined as a, and the length and width of small patches are also equal and defined as s, we let the ratio of a and s be denoted as n; $n = a/s$.

CNNs are characterized by local connections, weight sharing and hierarchical structures. Numerous neurons are present in a CNN, with each neuron receiving local field information to extract feature through a layer-by-layer convolution operation. Low-level features are then combined to obtain high-level feature information for the final task. Although this process involves many parameters, weight sharing across all neurons at each level substantially reduces the number of parameters and the training difficulty. As CNNs automatically learn and extract features, which reduces human intervention, resulting in improved generalizability. Thus, CNNs have been extensively applied across various domains, including retinal vessel segmentation.

In 2016, Li et al. (2015) treated retinal vessel segmentation as a cross-modality data transformation problem and employed a DNN to establish a function that maps retinal images to segmentation maps. The network consisted five layers and used a denoising autoencoder (DAE) to pretrain the first layer. Auto-encoding represents a feature mapping from input to the hidden layer. In DAE the input X is concatenated with label-data Y and a zero vector, while the output comprises only the pair of X and Y. Thus, the network is forced to learn the relationship between X and Y in order to achieve the purpose of denoising, which is superior to directly inputting X into the auto-encoder. Pre-training with DAE also minimizes interference from irrelevant information. During network training, a mean squared error loss function is employed for model optimization. Khalaf et al. (2016) separately identified large and small blood vessels, with their method's architecture diagram shown in Fig. 3(f). The last fully connected layer of the network consists of three neurons, each representing the probability that the central pixel belongs to a large blood vessel, a small blood vessel, or the background. The method uses the category of the central pixel as the final prediction result, which is a classic pixel-level prediction.

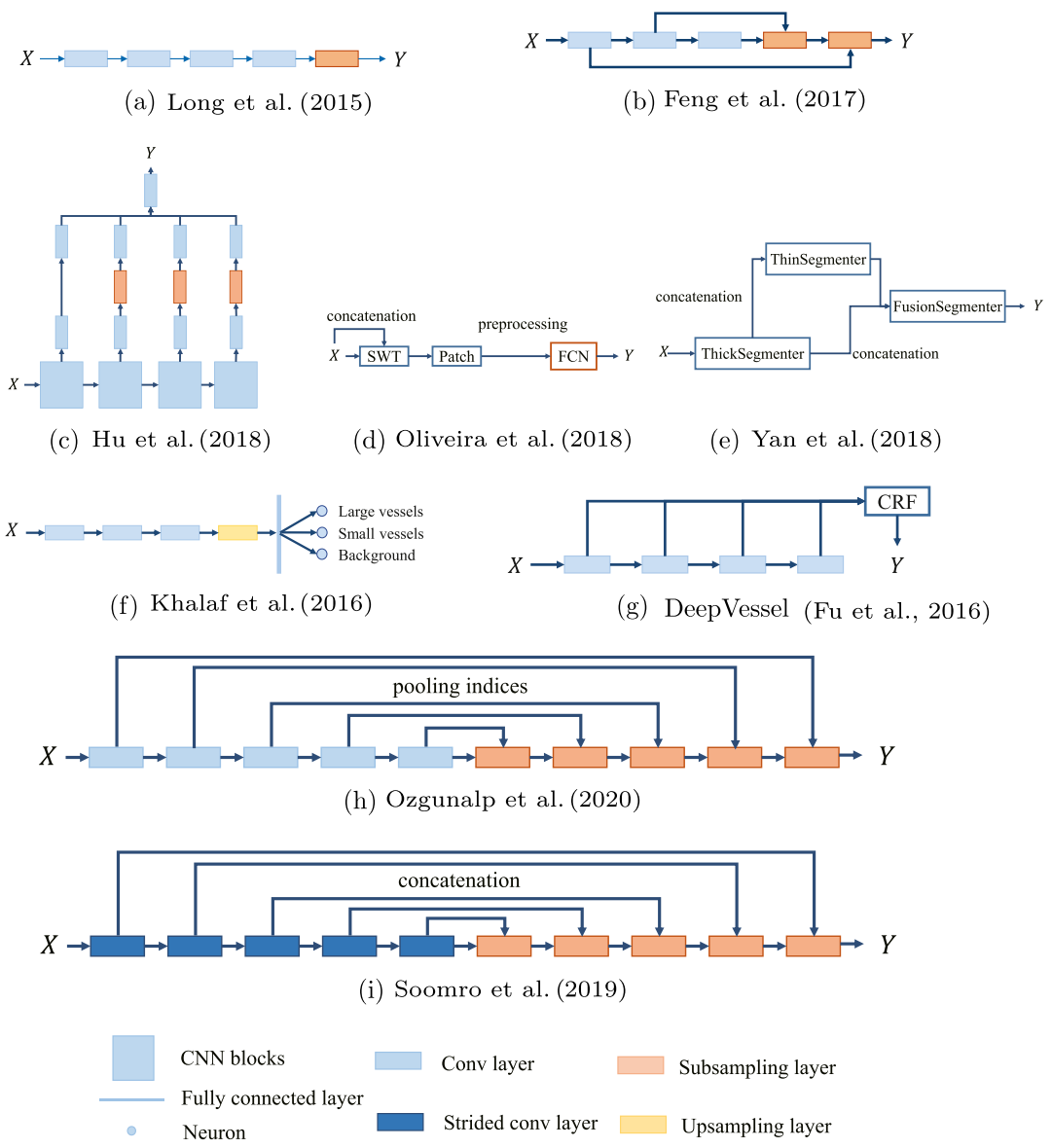


Fig. 3. An overview of CNN-based network architectures.

However, the accuracy of this method is limited. In response, structured prediction patterns were explored by Liskowski and Krawiec (2016). Their structured prediction includes $s \times s$ neurons, and each neuron represents a pixel. The output is a set of two-dimensional vectors, enabling simultaneous prediction of all pixels within $n \times n$ patches of size $s \times s$. In order to consider the influence of local noise and lesions, and to learn vessel features comprehensively at multiple scales, Fu et al. (2016) regard vessel segmentation as a boundary detection task. They employed side output layers to learn rich hierarchical representations and used conditional random field (CRF) layers to consider non-local information comprehensively. This method led to more accurate vessel segmentation as illustrated in the architecture diagram (Fig. 3(g)). To address the class imbalance between vessels and background pixels, Fu et al. follow the HED (Xie and Tu, 2015) to train the model using a class-balanced cross-entropy loss function:

$$L_s^{(m)}\left(W, w^{(m)}\right) = -\frac{\left|Y^-\right|}{|Y|}\sum_{j\in Y^+}\log\sigma\left(y_j^{(m)}\right) - \frac{\left|Y^+\right|}{|Y|}\sum_{j\in Y^-}\log\left(1-\sigma\left(y_j^{(m)}\right)\right), \tag{1}$$

where $L_s^{(m)}$ denotes the loss value of the mth side-output layer, $w^{(m)}$ represents the parameters of the mth layer, and $y_j^{(m)}$ signifies the output value of sample i on the mth side-output layer.

In 2015, Long et al. (2015) proposed the fully convolutional network, the architecture diagram of which is shown in Fig. 3(a). The original network ends with a fully connected layer, which converts the two-dimensional feature map output from the convolution into a one-dimensional vector. This fully connected layer serves as a highly purified feature. FCN replaces the fully connected layer with convolution and employs a 1×1 convolution for dimensionality reduction. The final convolution kernel has a number of channels equal to the predicted data's dimension, ensuring that the data is not flattened and that the image retains its spatial information. Moreover, compared to the large parameters set of the fully connected layer, the fully convolutional layer is not only compatible with different sizes of input, but also reduces the number of network parameters.

FCN (Long et al., 2015) uses deconvolution to upsample the feature map, restoring it to the original image size and performing pixel-level segmentation. This method solves the semantic-level image segmentation problem and provides a novel idea for the structured prediction in retinal vessel segmentation. Dasgupta and Singh (2017) were the first

to apply FCN for structured prediction in retinal vessel segmentation. Notably, the FCN loses some information during the downsampling process. To ensure that important information is preserved, Feng et al. (2017) combined high-level features with low-level features obtained through upsampling using skip connections, improving the issue of information losing during upsampling (the architecture diagram is shown in Fig. 3(b)). To capture multi-scale information and further improve the segmentation performance, Oliveira et al. (2018) employed stationary wavelet transform (SWT) to add a new channel to the input before feeding the segmentation map to the FCN, which enriched input information and helped to realize multi-scale information analysis for vessel segmentation (the architecture diagram is shown in Fig. 3(d)). Furthermore, Hu et al. (2018) fused the mid-layer feature map and output of the FCN structure and incorporated a CRF at the end of the network to utilize more spatial context information. This method considered the influence of all pixels to achieve a multi-scale segmentation network, as illustrated in Fig. 3(c). During model training, Hu et al. proposed an improved class-balanced cross-entropy loss function, as shown in Eq. (2):

$$L_s^{(m)}\left(W, w^{(m)}\right) = -\beta \sum_{j\in Y^+} \log\left(\sigma\left(y_j^{(m)}\right)\right) - \alpha(1-\beta) \sum_{j\in Y^-} \log\left(1-\sigma\left(y_j^{(m)}\right)\right), \tag{2}$$

where $L_s^{(m)}\left(W, w^{(m)}\right)$ represents the loss of the mth side-output, and $w^{(m)}$ denotes the parameters of the mth side-output. The weights α and β serves to balance the classes; here, α is a hyperparameter, and β is the ratio of background pixels to all pixels. $\sigma\left(y_j^{(m)}\right)$ signifies the probability map obtained by applying a sigmoid function to activation value $y_j^{(m)}$ at pixel j.

The improvement of the loss function mentioned above solves the imbalance between the vessel and background to a certain degree. Nevertheless, when vessels are extremely thin, segmentation remains challenging. To ensure accurate segmentation of these vessels, Yan et al. proposed a three-stage segmentation model (Yan et al., 2018) that explores the segmentation of thick and thin blood vessels respectively. The results of these two stages are then combined to achieve precise retinal vessel segmentation. The architecture diagram of this model is shown in Fig. 3(e). In addition to FCN, Badrinarayanan et al. (2017) proposed SegNet, which retains the index in the downsampling pooling layer and reuses the pooling index during upsampling. Compared with the FCN, SegNet does not need to learn during upsampling and requires fewer learning parameters. Subsequently, Ozgunalp et al. (2020) employed SegNet for retinal vessel segmentation, with the architecture shown in Fig. 3(h). In addition, Soomro et al. (2019) proposed a novel network framework based on the FCN, replacing the pooling layer with a strided convolutional layer during downsampling and addressing the imbalance between vessels and backgrounds during model training using the class-balanced cross-entropy loss function shown in Eq. (1) (see Fig. 3(i) for the model architecture).

3.3.2. GAN-based methods

In 2014, Goodfellow et al. proposed generative adversarial networks (GAN) (Goodfellow et al., 2014). GAN is a generative model that involves a competition between a generator and a discriminator during the training process. On the one hand, GAN learns data distributions, and imitates them to generate samples for purpose data augmentation. In comparison to other models, GAN can generate clearer and more realistic samples. On the other hand, GAN directly employ the task model as a generator, and through adversarial training, the prediction results obtained by the task model get closer and closer to the ground truth, thereby achieving the training purpose of improving the task model's performance. This concept of generator and discriminator adversarial training has provided researchers with innovative ideas for problem-solving, leading to a new era in the automatic segmentation of retinal vessels. The frameworks of the mentioned GAN-based architectures are shown in Fig. 4.

Generator implements data augmentation. Lahiri et al. (2017) employed GAN for retinal image patch synthesis to alleviate the manual labeling efforts required in retinal vessel segmentation. In their method, retinal image patches are utilized: the generator maps the noise space to actual retinal images, synthesizing retinal image patches. Subsequently, both labeled retinal image patches and synthetic image patches from the dataset are input into the discriminator. Through an adversarial game between the discriminator and the generator, the synthesized images become increasingly similar to real images, ultimately achieving data augmentation and enabling the segmentation model to learn more feature information. However, these authors only considered the original image information and overlooked the guidance information contained in the ground truth. Costa et al. (2017) initially learned the distribution of retinal vessel images via an auto-encoding model to obtain the latent space distribution, and then employed this learned distribution to synthesize blood vessel maps. A GAN-like structure is subsequently used, leveraging the synthetic vessel map obtained by the auto-encoding model to generate retinal fundus images, which form a GAN generator. The pairs (of synthetic vessel maps and corresponding generated retinal fundus images, and of the fundus images in the dataset and the label maps) are fed into the discriminator. Both the generator and discriminator are optimized simultaneously through adversarial training, while the auto-encoding model is optimized using backpropagation. In order to further enhance the generated image's resemblance to real images and strengthen generalization, Zhao et al. introduced a style transfer framework to GAN and proposed Tub-sGAN (Zhao et al., 2018a). These authors incorporated style transfer into the GAN framework for the first time and proposed the nonlinear variant network R-sGAN (Zhao et al., 2018b) of GRU (Chung et al., 2014). They integrated the R-sGAN unit as a generator into the GAN generator and include the GAN discriminator as part of the loss function to ensure the diversity of the synthetic images.

Generator as task model. Besides using the generator of GAN for retinal fundus images synthesis, the generator can be directly employed for retinal vessel segmentation. The U-GAN (Wu et al., 2019) proposed by Wu et al. which uses the retinal vessel automatic segmentation model as the GAN's generator, then sends the segmentation results and ground truth to the discrimination network at the same time. The segmentation results are continuously optimized through adversarial training. To improve the segmentation performance, Kamran et al. added the SFA module to the generator's automatic segmentation model to fuse the feature information of different scales, thus proposing RV-GAN (Kamran et al., 2021). The proposed architecture employs two generators and two discriminators: the feature maps obtained by the first generator are fed into the second generator, and the loss values of the two discriminators are combined, enabling better localization and segmentation of retinal vessels. Furthermore, the feature matching loss (Wang et al., 2018) performs semantic segmentation by analyzing the features merged and extracted from the discriminator. The feature matching loss is formulated in Eq. (3). In their model training, the authors adopted a newly proposed weighted feature matching loss (see Eq. (4)), which is calculated by aggregating the features from each downsampling and upsampling block of the discriminator's encoder and decoder. To train the model, both real and synthetic segmentation maps are fed into the network consecutively.

$$L_{fm}\left(G, D_{enc}\right) = E_{x,y}\frac{1}{N}\sum_{i=1}^{k}\left\|D_{enc}^{i}(x,y) - D_{enc}^{i}(x, G(x))\right\|, \tag{3}$$

$$L_{wfm}\left(G, D_n\right) = E_{x,y}\frac{1}{N}\sum_{i=1}^{k}\lambda_{enc}^{i}\left\|D_{enc}^{i}(x,y) - D_{enc}^{i}(x, G(x))\right\| + \lambda_{dec}^{i}\left\|D_{dec}^{i}(x,y) - D_{dec}^{i}(x, G(x))\right\|, \tag{4}$$

where N represents the number of features, λ_{enc} and λ_{dec} denote the internal weight multipliers assigned to each extracted feature map. The

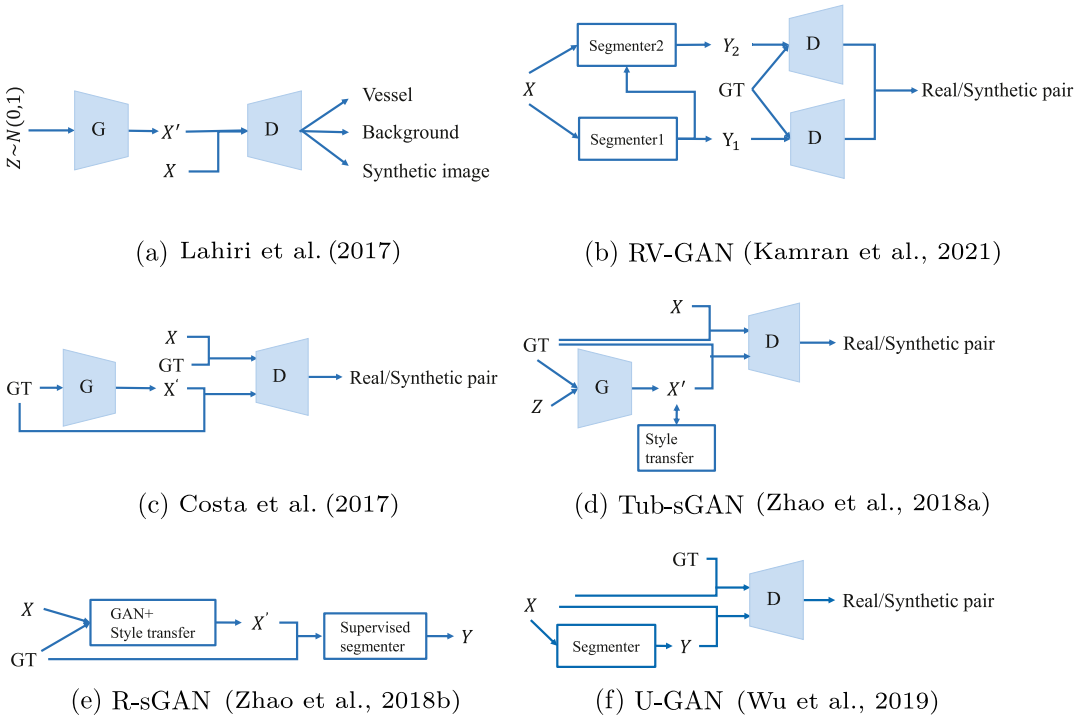


Fig. 4. An overview of some GAN-based network architectures.

weight values are within the range of $[0, 1]$, and their sum equals to 1. Typically, a higher weight is assigned to the decoder feature maps than the encoder ones.

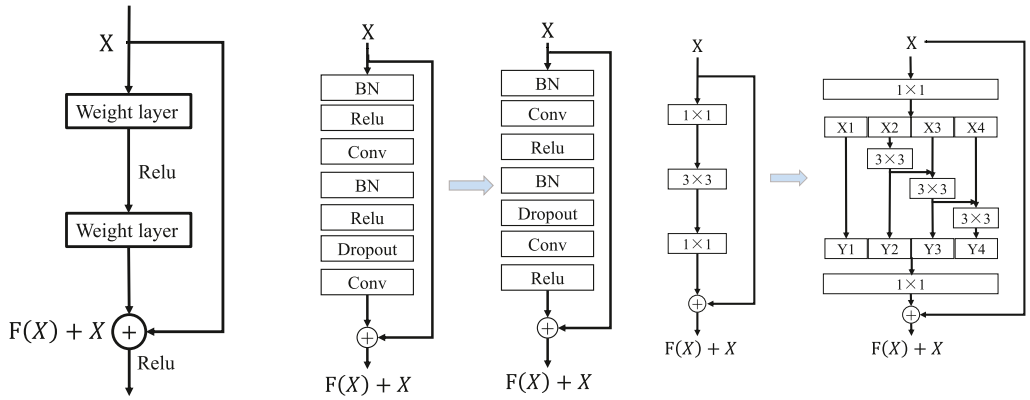
3.3.3. *Unet-based methods*

In addition to the basic framework of FCN (Long et al., 2015), GAN (Goodfellow et al., 2014), and SegNet (Badrinarayanan et al., 2017), the UNet (Ronneberger et al., 2015) proposed by Ronneberger et al. in 2015 for biomedical image segmentation quickly attracted research attention and pushed medical image segmentation to new heights. GAN can be used not only for data augmentation but also for achieving good performance by directly using the segmentation network as a generator. Typically, when the generator is a segmentation network, most of them are based on UNet (Ronneberger et al., 2015). The difference is that during the training of UNet-based networks, the update of network parameters is based on the learning of data samples, whereas in GAN, the parameter update of the segmentation network used as a generator comes from the backpropagation of the discriminator.

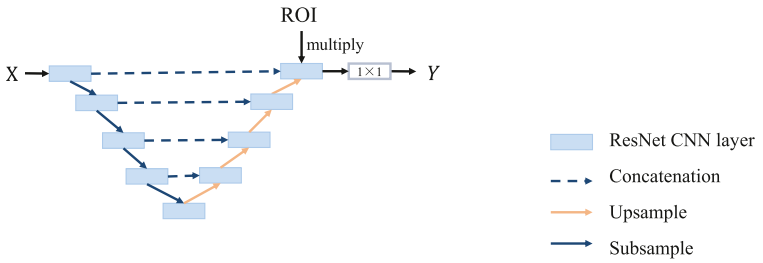
The UNet (Ronneberger et al., 2015) utilizes a U-shaped network structure to extract and integrate both contextual and location information, and employs cross-layer connection to fuse low-level location information with deep-level semantic information through, achieving excellent segmentation results. Unlike FCN, which fuses the features by using the pixel value of the feature map response position, UNet concatenates features across channels. By combining UNet with other classic network structures, many new automatic retinal vessel segmentation models have been proposed. In the following sections, we will introduce this series of models, which are classified based on the model structure referenced by the various variant models or the network concepts employed.

ResNet. Inspired by the success of ResNet (He et al., 2016), researchers have incorporated it into UNet, leading to the proposal of novel models. Xiao et al. (2018) presented Res-UNet by organically combining the concepts of ResNet and Attention. Res-UNet introduces skip connections and employs an attention mechanism by extracting circular regions of interest (ROI) and dark backgrounds, enabling the model to concentrate on target ROI regions while disregarding irrelevant noise backgrounds. Li et al. (2019) proposed an innovative residual structure based on the traditional residual structure, wherein the batch normalization layer places before the activation unit, and the new residual network is applied to UNet's convolution. This results in the creation of MResU-Net. Inspired by Res2Net, Li et al. replaced the standard convolution in UNet with Res2Net (Gao et al., 2019) and proposed a novel network structure, Res2Unet (Li et al., 2022a), which integrates a channel attention mechanism to facilitate inter-channel communication. The architectures of these UNet-based networks, which incorporate ResNet ideas, are depicted in Fig. 5.

Attention. The ResNet concept is widely employed in various UNet variant frameworks. On this basis, researchers have incorporated the popular attention mechanism (Vaswani et al., 2017) to further enable the network to acquire more information about important regions and improve segmentation accuracy. As early as 2018, Xiao et al. incorporated the attention mechanism in Res-UNet (Xiao et al., 2018). Rather than emphasizing the ROI region, Oktay et al. integrated the attention mechanism into the skip connection part, striving to capture more important information within the model's internal structure. They fused the features of the encoder part with the upsampled feature map, performed convolution and sigmoid operations, and obtained an importance score ranging between 0 and 1. This attention score is then assigned to the feature map input by the encoder part, so that the attention mechanism is integrated into the UNet structure. Accordingly, the Attention U-Net (Oktay et al., 2018) is proposed, and the architecture of it is presented in Fig. 6(a). Additionally, gathering information at multiple scales is crucial to avoid losing vital information. Du et al. proposed Att-PSP-UNet (Du et al., 2021), which employs the attention mechanism from Attention U-Net in the skip connection part, and replaces the traditional pooling layer in the encoder with the Pyramid Pool layer. The Att-PSP-UNet framework is illustrated in Fig. 6(b). The transformer network, derived from the attention mechanism, has also gained widespread recognized for its effectiveness. The core component of transformer is self-attention, and its structure is fundamentally an encoder–decoder architecture.



(a) ResNet (He et al., 2016) Block (b) Residual block in MResU-Net (Li et al., 2019) (c) Residual block in Res2Unet (Li et al., 2022a)



(d) Res-UNet (Xiao et al., 2018)

Fig. 5. An overview of UNet-based network architectures that add the idea of ResNet.

In recent years, applying transformers to computer vision has become increasingly common, and retinal vessel segmentation is no exception. In 2022, Shen et al. (2022) introduced a self-attention mechanism into vessel segmentation by focusing on effective weights through a transformer network layer based on a global attention map. The architecture diagram is presented in Fig. 6(c). Concurrently, they replaced standard convolution blocks with residual blocks, added spatial attention and channel attention blocks to these residual blocks to improve feature extraction performance. Hu et al. incorporated transformer to CNN at the bottom of U-shaped network, proposing HT-Net (Hu et al., 2022). HT-Net (Hu et al., 2022) considers shallow features and long-range information interactions for optimizing feature extraction and embeds feature fusion block in shallow levels to enrich the feature space. At the same time, a feature refinement block is added to address blood vessel scale changes by fusing multi-scale feature information, resulting in excellent performance from HT-Net. TiM-Net (Zhang et al., 2022) combines the advantages of transformer and M-Net (Fu et al., 2018), designing a dual-attention mechanism based on channels and spaces. It introduces the self-attention mechanism from the transformer into the skip connection in UNet to re-encode the features, achieving good segmentation performance. In the same year, in order to address the challenge of simultaneously capturing global and local features in fundus images and enhancing fine vessel segmentation, Li et al. (Li et al., 2022b) proposed a global transformer and dual local attention network through deep-shallow hierarchical feature fusion. The detailed network structure is depicted in Fig. 6(f).

Dropout. As network depth increases, overfitting becomes a significant factor affecting network performance. However, traditional dropout in the convolutional layer is unsuitable due to the spatial correlation between activated neurons in the convolutional layer. It randomly deactivates feature neurons during the training process. Influenced by DropBlock (Ghiasi et al., 2018) proposed by Ghiasi et al. Guo et al. added a structured dropout to each convolutional layer and proposed a structured dropout UNet, named SD-Unet (Guo et al., 2019). SD-Unet deletes continuous regions from the feature map of the layer, prompting the network to focus on learning features from other parts by deactivating section of adjacent region in the feature map. These methods help to reduce the issue of overfitting in the network. Subsequently, Guo et al. incorporated a Spatial Attention Module (SAM) to SD-Unet and proposed SA-UNet (Guo et al., 2021). The SAM employs the spatial relationship between features to generate a spatial attention map, which enhances the network's ability to capture spatial information and improve the segmentation performance. The architectures of these frameworks are illustrated in Fig. 7.

RCNN. RCNN (Girshick et al., 2014), which is commonly used in the field of target detection, trains a deep network for feature extraction and uses the selective search method to pre-extract a series of candidate regions that are more likely to be targets. It then extracts features from only these candidate regions for subsequent judgment, which greatly improves the target detection effect. Inspired by the idea of RCNN, Alom et al. (2019) proposed two recurrent convolutional neural network architectures, RU-Net and R2U-Net, based on the UNet framework. RU-Net replaces some traditional convolutional layers in the original UNet with recurrent convolutional layers and accumulates features at different time steps, resulting in better and stronger feature representation. R2U-Net incorporates the same circular convolution as RU-Net and adds the idea of ResNet to improve the training of deep networks. Influenced by the structure of RCNN, Hu et al. directly connected the left part of the original skip connection in UNet to the input layer and replaced the original direct connection with a circular convolution operation to obtain more information, an approach named RCSCU-Net (Hu and Liu, 2021). RCSCU-Net resolves the issue of the excessive amount of coupling between skip connections and convolutional layers in traditional UNet. These approaches all adopt the ideas of ResNet and RCNN and integrate them into the UNet framework in different ways to obtain different variant architectures that improve segmentation performance. The architectures are presented in Fig. 8.

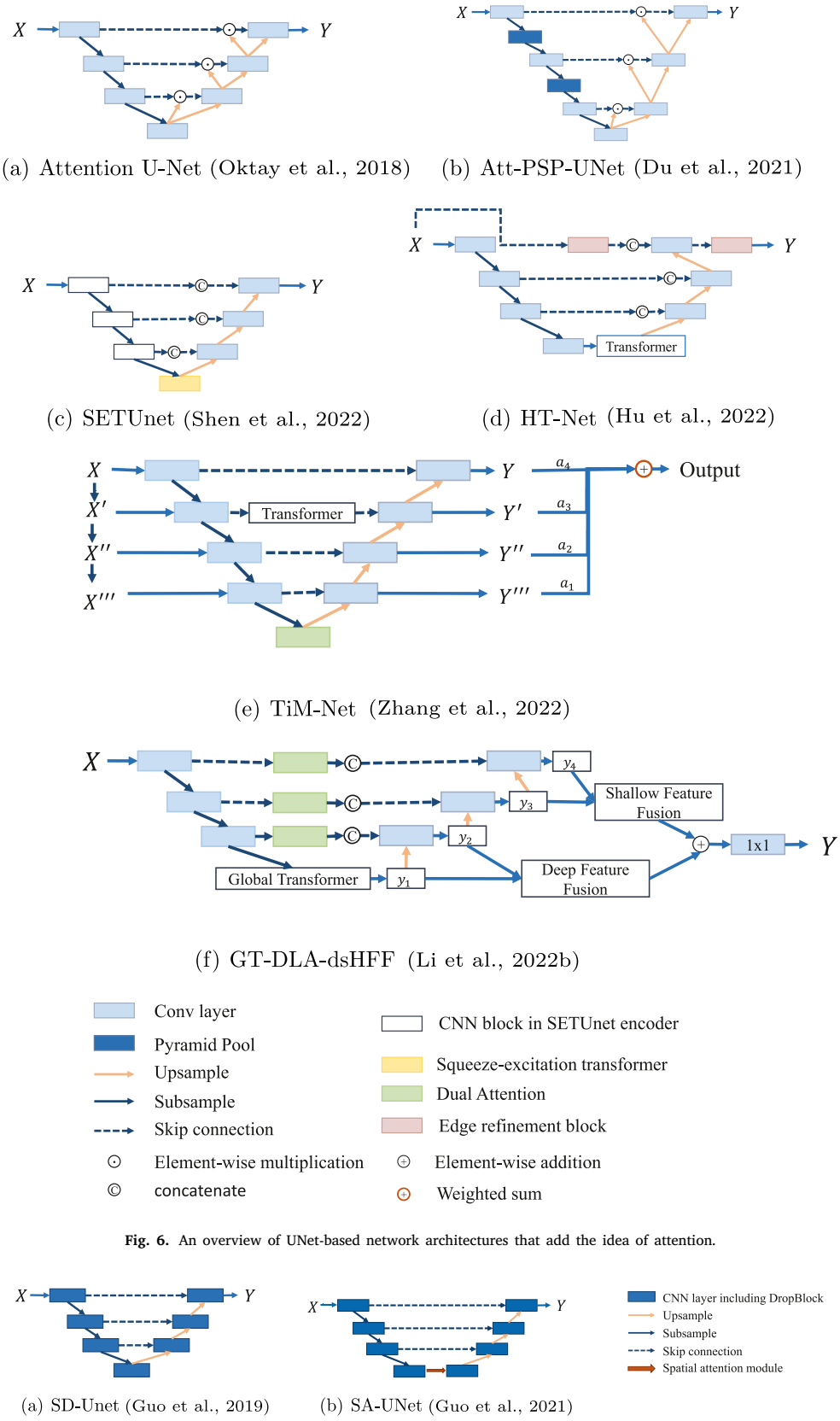
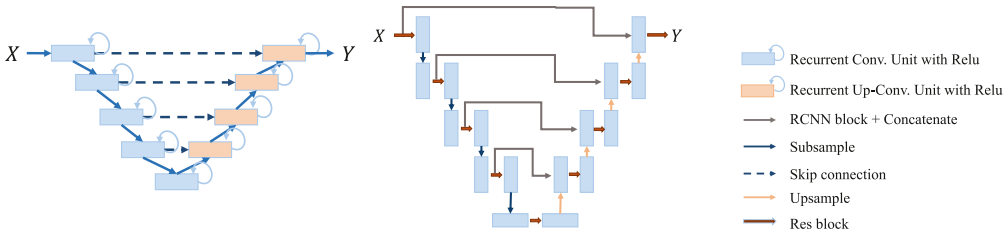


Fig. 6. An overview of UNet-based network architectures that add the idea of attention.

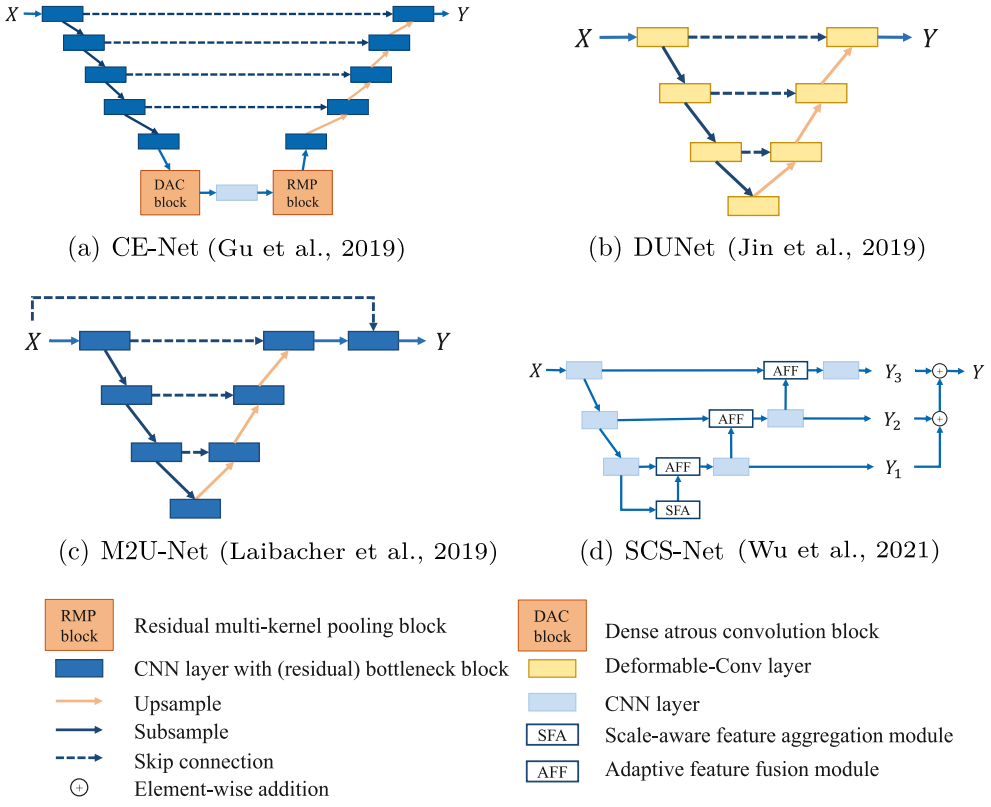
Fig. 7. An overview of UNet-based network architectures that introduce the idea of dropout used for vessel segmentation.

Other Module. In addition to the UNet variant networks described above, there are other UNet-based variant structures. Laibacher et al. proposed M2U-Net (Laibacher et al., 2019) by replacing the convolution blocks in UNet with bottleneck blocks with stride 1 and stride 2, as shown in Fig. 9(c). The CE-Net (Gu et al., 2019) proposed by Gu et al. employs ResNet for feature extraction, followed by the addition of a dense atrous convolution (DAC) module and a residual multi-kernel pooling (RMP) module at the bottom of the U-shaped network to



(a) R2U-Net (Alom et al., 2019) (b) RCSCU-Net (Hu and Liu, 2021)

Fig. 8. An overview of UNet-based network architectures that reference the structure of RCNN.



(a) CE-Net (Gu et al., 2019) (b) DUNet (Jin et al., 2019)

(c) M2U-Net (Laibacher et al., 2019) (d) SCS-Net (Wu et al., 2021)

Fig. 9. An overview of UNet-based network architectures that add other modules.

retain more spatial information while obtaining more high-level feature information, as illustrated in Fig. 9(a). The deformable convolution has a larger sampling field of view than standard convolution and a flexible receptive field that is conducive to matching targets of different scales. Jin et al. replaced the traditional convolution with deformable convolution and proposed DUNet (Jin et al., 2019), as presented in Fig. 9(b). In order to solve the impact of different retinal vessel sizes on segmentation, Wu et al. inserted three efficient modules into UNet, namely a scale-aware feature aggregation module (SFA), an adaptive feature fusion module (AFF), and a multi-level semantic supervision (MSS), and proposed SCS-Net (Wu et al., 2021), as shown in Fig. 9(d). The SFA module extracts hidden multi-scale contextual information and aggregate multi-scale features effectively, while the AFF module replaces the ordinary skip connections in the traditional UNet model and guides the fusion of adjacent hierarchical features by adaptively fusing semantic information and spatial information, thus capturing more distinguishable semantic information. Finally, the MSS module learns more global semantic representations from the side output layers, thereby improving segmentation accuracy by assigning auxiliary supervision to the decoder.

Iteration. In addition to modifying the U-shaped structure with various modules, some researchers have repeated the U-shaped structure to improve the final segmentation results. Zhuang employed two U-shaped structures and replaced the original channel number splicing with pixel value addition inside the U-shaped network. The decoder of the first U-shaped network and the encoder of the second U-shaped network were connected by adding pixel values, resulting in LadderNet (Zhuang, 2018), as shown in Fig. 10(a). LadderNet also adopted a parameter-sharing residual block to simplify training. Li et al. proposed IterNet (Li et al., 2020), a deeper iterative network composed of a UNet model and a mini-UNet that iterate multiple times, and its framework is illustrated in Fig. 10(c). IterNet splices the results of all previous UNets for use as the current new input, achieving finer segmentation. Weight sharing is still used in the training process. Even so, given that vessels differ in terms of their segmentation difficulty, some vessels are still difficult to segment accurately. Wang et al. proposed HAnet (Wang et al., 2020), which consists of one encoder and three decoders, as illustrated in Fig. 10(d). The encoder is employed to extract features, while one decoder acquired a coarse segmentation result and divided the features into easy-to-segment or difficult-to-segment, then the other two decoders dedicated to further segmenting retinal vessels. This method of segmenting the easy area and the difficult area in a targeted manner achieves a certain amount of improvement. Dong et al. proposed CRAUNet (Dong et al., 2022) by combining the multi-scale fusion

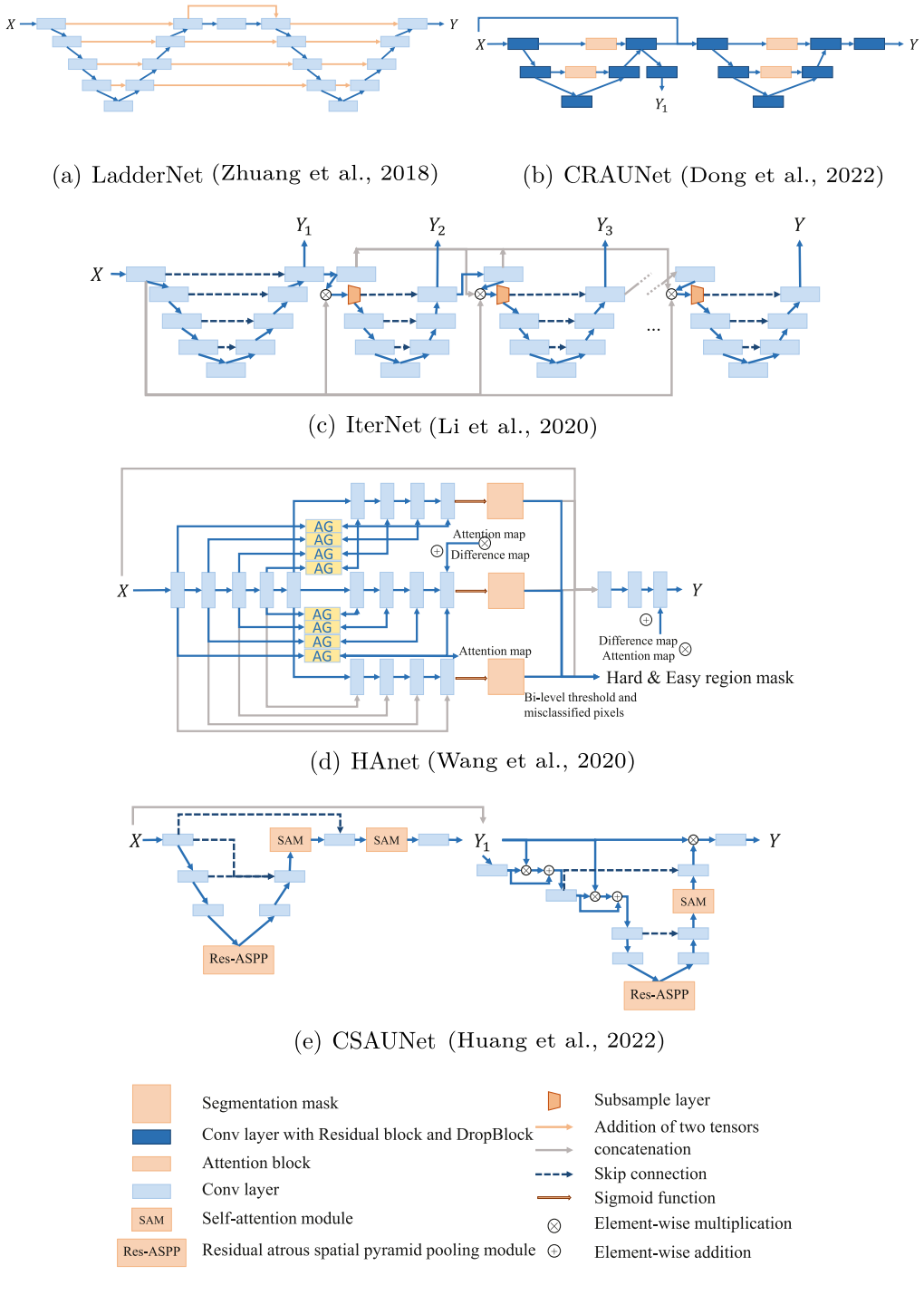


Fig. 10. An overview of UNet-based network architectures that add the idea of iteration.

channel attention (MFCA) module and the ResNet structure with DropBlock, and its network architecture is shown in Fig. 10(b). CRAUNet cascades two U-shaped networks, and incorporates the feature maps of the first U-shaped network with the original images before feeding them into the second U-shaped network for retraining, leading to more accurate segmentation results. In the same year, Huang et al. adopted the coarse-to-fine approach and proposed cascaded self-supervised network, named CSAUNet (Huang et al., 2022) (the framework is shown in Fig. 10(e)). CSAUNet first pre-trains a self-supervised U-shaped network to roughly segment retinal fundus images. The coarse segmentation results are then treated as spatial attention maps and fed into the second subdivided U-shaped network, which jointly trains the two models to generate the final segmentation result.

During the training of various UNet-based network architectures, cross-entropy loss, dice coefficient, binary cross-entropy, and some improved loss functions based on cross-entropy loss are among the most commonly used loss functions. Cross-entropy loss describes the distance between two probability distributions, and minimizing the cross-entropy loss can make the predicted distribution closer to the true distribution. The formula for cross-entropy loss function is shown in

Table 2
Performance comparison of unsupervised methods and CNN-based supervised methods.

Method	Year	DRIVE					STARE					CHASEDB1				
		Acc	Auc	Sp	Se	F1	Acc	Auc	Sp	Se	F1	Acc	Auc	Sp	Se	F1
Unsupervised Method																
Mendonca and Campilho (2006)	2006	.9485	–	–	–	–	.9466	–	–	–	–	–	–	–	–	–
Al-Rawi et al. (2007)	2007	.9532	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Lam et al. (2010)	2010	.9472	–	–	–	–	.9567	–	–	–	–	–	–	–	–	–
Wang et al. (2013)	2013	.9461	–	–	–	–	.9521	–	–	–	–	–	–	–	–	–
Azzopardi et al. (2015)	2015	.9427	.9571	.9707	.7526	–	.9467	.9487	.9689	.7543	–	.9411	.9434	.9651	.7257	–
Kar and Maity (2016a)	2016	.961	–	.9792	.7548	–	.973	–	.9788	.7577	–	–	–	–	–	–
Zhang et al. (2016)	2016	.9476	.9636	.9725	.7743	–	.9554	.9748	.9758	.7791	–	.9452	.9606	.9661	.7626	
Rezaee et al. (2017)	2017	.9463	–	.9793	.7189	–	.9532	–	.9741	.7202	–	–	–	–	–	–
Asl et al. (2017)	2017	–	–	.9732	.7428	–	–	–	.9706	.7419	–	–	–	.9767	.7535	–
Neto et al. (2017)	2017	–	–	.9631	.7942	–	–	–	.9537	.7695	–	–	–	–	–	–
Zhao et al. (2017)	2018	.958	.975	.979	.774	–	.957	.959	.976	.788	–	–	–	–	–	–
no-reference (Liu et al., 2019)	2019	.9559	.9779	.978	.8027	.8225	.9623	.9793	.9843	.7771	.8036	.9742	.9905	.9843	.8769	.8598
Mahapatra et al. (2022)	2022	.9605	–	.9844	.702	–	.9601	–	.9802	.6846	–	–	–	–	–	–
Traditional CNN-Based Supervised Method																
Li et al. (2015)	2016	.9527	.9738	.9816	.7569	–	.9628	.9879	.9844	.7726	–	.9581	.9716	.9793	.7507	–
Liskowski and Krawiec (2016)	2016	.9515	.971	.9806	.752	–	.9696	.988	.9866	.8145	–	–	–	–	–	–
DeepVessel (Fu et al., 2016)	2016	.9523	–	–	.7603	–	.9585	–	–	.7412	–	.9489	–	–	.713	–
Khalaf et al. (2016)	2016	.9472	–	.9562	.8397	–	–	–	–	–	–	–	–	–	–	–
Feng et al. (2017)	2017	.956	.9792	.9839	.7811	–	–	–	–	–	–	–	–	–	–	–
Dasgupta and Singh (2017)	2017	.9533	.9744	.9801	.7691	–	–	–	–	–	–	–	–	–	–	–
Oliveira et al. (2018)	2018	.9576	.9821	.9804	.8039	–	.9694	.9905	.9858	.8315	–	.9653	.9855	.9864	.7779	–
Hu et al. (2018)	2018	.9533	.9759	.9793	.7772	–	.9632	.9751	.9814	.7543	–	–	–	–	–	–
Soomro et al. (2019)	2019	.956	–	.985	.87	–	.968	–	.986	.848	–	.976	–	.982	.886	–
Yan et al. (2018)	2019	.9538	.975	.982	.7631	–	.9638	.9833	.9857	.7735	–	.9607	.9776	.9806	.7641	–

Table 3
Performance comparison of GAN-based and UNet-based supervised methods.

Method	Year	DRIVE					STARE					CHASEDB1				
		Acc	Auc	Sp	Se	F1	Acc	Auc	Sp	Se	F1	Acc	Auc	Sp	Se	F1
GAN-Based Supervised Method																
Lahiri et al. (2017)	2017	–	.945	–	–	–	–	–	–	–	–	–	–	–	–	–
Tub-sGAN (Zhao et al., 2018a)	2018	–	–	.9815	.8038	.8033	–	–	.9841	.7896	.7902	–	–	–	–	–
R-sGAN (Zhao et al., 2018b)	2019	–	–	.9766	.7641	.7602	–	–	.981	.7631	.7655	–	–	–	–	–
U-GAN (Wu et al., 2019)	2019	.9615	–	.982	.7798	–	–	–	–	–	–	–	–	–	–	–
RV-GAN (Kamran et al., 2021)	2021	.979	.9887	.9969	.7927	.869	.9754	.9887	.9864	.8356	.8323	.9697	.9914	.9806	.8199	.8957
UNet-Based Supervised Method																
Res-Unet (Xiao et al., 2018)	2018	.9655	–	–	.7715	–	.9693	–	–	.7469	–	–	–	–	–	–
M2U-Net (Laibacher et al., 2019)	2019	.963	.9714	–	–	.8091	–	–	–	–	–	.9703	.9666	–	–	.8006
CE-Net (Gu et al., 2019)	2019	.9545	.9779	–	.8309	–	–	–	–	–	–	–	–	–	–	–
DUNet (Jin et al., 2019)	2019	.9566	.9802	.98	.7963	.8237	.9641	.9832	.9878	.7595	.8143	.961	.9804	.9752	.8155	.7883
R2Unet (Alom et al., 2019)	2019	.9613	.9793	.9807	.7661	–	.9712	.9914	.9862	.8398	–	.9634	.9815	.982	.7756	–
LadderNet (Zhuang, 2018)	2019	.9561	.9793	.981	.7856	.8202	–	–	–	–	–	.9656	.9839	.9818	.7978	.8031
SD-Unet (Guo et al., 2019)	2019	.9674	.9836	.9848	.7891	.8042	.9725	.985	.9899	.7548	.802	.9738	.9872	.99	.7559	.7978
MResU-Net (Li et al., 2019)	2019	–	.9799	.9799	.7969	.8237	–	.9816	.9795	.8101	.8147	–	–	–	–	–
IterNet (Li et al., 2020)	2020	.9573	.9816	.9838	.7735	.8205	.9701	.9881	.9886	.7715	.8146	.9655	.9851	.9823	.797	.8037
HAnet (Wang et al., 2020)	2020	.9581	.9823	.9813	.7991	.8293	.9673	.9881	.9844	.8186	.8379	.967	.9871	.9813	.8239	.8191
SCS-Net (Wu et al., 2021)	2021	.9697	.9837	.9838	.8289	.8189	.9736	.9877	.9839	.8207	–	.9744	.9867	.9839	.8365	
SA-UNet (Guo et al., 2021)	2021	.9698	.9864	.984	.8212	.8263	–	–	–	–	–	.9755	.9905	.9835	.8573	.8153
Att-PSP-Unet (Du et al., 2021)	2021	.9556	.978	.981	.7814	.8176	–	–	–	–	–	.959	.9784	.9727	.8195	.7813
RCSCU-Net (Hu and Liu, 2021)	2021	.9624	.9814	.9773	.8251	–	.9673	.9839	.9752	.8119	–	–	–	–	–	–
CRAUNet (Dong et al., 2022)	2022	.9586	.983	–	.7954	.8302	–	–	–	–	–	.9659	.9864	–	.8259	.8156
SETUnet (Shen et al., 2022)	2022	.968	.9875	.9854	.8056	.8357	.9686	.9832	.9817	.814	.8	.9734	.9906	.9875	.825	.835
CSAUNet (Huang et al., 2022)	2022	.9676	.9758	.981	.834	–	.9728	.9876	.9862	.8304	–	–	–	–	–	–
Res2Unet (Li et al., 2022a)	2022	–	.9762	.9734	.8162	.8168	–	.9802	.976	.8234	.8038	–	–	–	–	–

Eq. (5). For binary classification problems with only two classes, the loss function is referred to as binary-classification cross-entropy loss, which is formulated as shown in Eq. (6).

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M} q_{ic}\log p_{ic}, \tag{5}$$

$$BCELoss = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log p_i + \left(1-y_i\right)\cdot \log\left(1-p_i\right)\right], \tag{6}$$

where N represents the total number of samples, while M represents the total number of classes. The value of q_{ic} is 1 if the true category of sample i is the same as c, and 0 otherwise. Moreover, p_{ic} denotes the predicted probability that the observation sample i belongs to category c. y_i represents the label of sample i, and p_i represents the probability that sample i is predicted to belong to the positive class.

4. Evaluation indicators and performance comparison

4.1. Evaluation indicators

In vessel segmentation tasks, vessel pixels are considered the positive class, while background pixels are deemed the negative class: True Positive (TP) denotes the correctly identified positive class, True

Negative (TN) represents the correctly identified negative class, False Positive (FP) refers to the negative class incorrectly identified as the positive class, and False Negative (FN) indicates the positive class misclassified as the negative class. The evaluation metrics employed in this context include precision, recall, sensitivity, specificity, accuracy, F1 score, and Area Under the ROC Curve (AUC). The calculation of these metrics involves TP, TN, FP, and FN. Precision is the ratio of TP to the sum of TP and FP, indicating the probability that the prediction is correct among all the predicted positive samples. Recall is the ratio of TP to the sum of TP and FN, representing the probability of correct predictions among all positive classes. Sensitivity denotes the model's responsiveness to the positive class and shares the same value as recall. Specificity is the ratio of TN to the sum of TN and FP. Accuracy is the proportion of correctly predicted pixels to all pixels, i.e., the ratio of the sum of TP and TN to the sum of TP, TN, FP and FN. The F1 score is the harmonic mean of precision and recall, balancing both metrics. AUC represents the area under the ROC curve, with values ranging from 0.5 to 1. The advantage of the ROC curve is that it does not change with the distribution of positive and negative samples in the test set. The Dice coefficient, more commonly used in medical images, displays the similarity between the predicted and the real sets and is numerically equivalent to the F1 score. Additionally, (1-Dice) is often used within the field of medical images as a type of loss function during model training. The formulas for precision, recall, sensitivity, specificity, accuracy, F1 score, and Dice are provided in Eqs. (7)–(11).

$$\text{precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{recall} = \text{sensitivity} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{specificity} = \frac{TN}{TN + FP} \tag{9}$$

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{10}$$

$$\text{F1 score} = \text{dice} = 2 \cdot \left(\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \right) \tag{11}$$

4.2. SOTA model performance comparison

Following the development of automatic retinal vessel segmentation networks, numerous state-of-the-art models have emerged. In this section, the performance of these models will be compared on the DRIVE, STARE and CHASEDB1 datasets. Table 2 presents the performance comparison of unsupervised learning methods and supervised learning methods based on traditional CNN, while Table 3 presents a performance comparison of supervised learning methods based on GAN and UNet.

5. Conclusion

As retinal vessel segmentation has gradually developed and grown, the incorporation of artificial intelligence methods to achieve automatic segmentation of vessels is the results of the development of the times. Datasets serve as the foundation for automatic segmentation methods. This paper summarizes these datasets for retinal vessel segmentation. It can be seen that the number of data, the imaging quality and the image diversity of the data have all improved to a certain extent from the original DRIVE published in 2000.

In terms of methods, unsupervised learning methods starts from the image itself and rely on image features to achieve segmentation. The mainstream unsupervised methods include filter-based methods, vascular tracking, and morphological reconstruction-based methods. However, unsupervised methods lack the guidance of labeled data and require human intervention, so it is more popular to use supervised learning methods to segment retinal vessels. The vessels segmentation of supervised learning methods evolved from the traditional CNN to FCN-based, GAN-based, and UNet-based. At present, most mainstream retinal vessels segmentation networks are based on UNet. Moreover, even in the field of medical image segmentation, UNet plays a pivotal role. In a series of UNet-based supervised learning methods, this paper discusses and summarizes according to the model structure referenced by the various variant models or the network concepts employed (such as ResNet, Attention, Dropout, RCNN and iteration, etc.).

This paper summarizes the datasets and discusses the methods of retinal vessel segmentation from two aspects of unsupervised and supervised learning, which provides a comprehensive review and summary of retinal vessel segmentation. Given the current context of retinal vessel segmentation, it is evident that the number of images in commonly used datasets is limited. In the future, we could devote ourselves to building a more comprehensive and massive retinal vessel segmentation datasets to foster the development of retinal vessel segmentation. In terms of retinal vessel segmentation methods, most CNN-based methods do not take into account the actual morphology of blood vessels. Investigating the potential benefits of combining morphological methods with CNN-based approaches to enhance vessel segmentation capabilities is a worthwhile future direction. In addition, segmenting small blood vessels in fundus images is a challenging task, especially when images have uneven illumination, low contrast, and complex regions. Further research and exploration are necessary to overcome these challenges and improve vessel segmentation. This paper offers valuable insights and guidance for researchers developing new segmentation models and for new scholars entering the field of retinal blood vessel segmentation.

CRediT authorship contribution statement

Qing Qin: Conceptualization, Methodology, Writing – original draft. **Yuanyuan Chen:** Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Abbasi-Sureshjani, S., Smit-Ockeloen, I., Zhang, J., Ter Haar Romeny, B., 2015. Biologically-inspired supervised vasculature segmentation in SLO retinal fundus images. In: International Conference Image Analysis and Recognition. Springer, pp. 325–334.

Al-Diri, B., Hunter, A., Steel, D., Habib, M., Hudaib, T., Berry, S., 2008. A reference data set for retinal vessel profiles. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 2262–2265.

Al-Rawi, M., Qutaishat, M., Arrar, M., 2007. An improved matched filter for blood vessel detection of digital retinal images. Comput. Biol. Med. 37 (2), 262–267.

Alom, M.Z., Yakopcic, C., Hasan, M., Taha, T.M., Asari, V.K., 2019. Recurrent residual U-Net for medical image segmentation. J. Med. Imaging 6 (1), 014006.

Anon, 2019. Clinical manifestations of retinopathy of prematurity. https://www.bohe.cn/article/view/153139.html.

Anon, 2021. Retinal vessel therapy. https://www.yxj.org.cn/detailPage?articleId=286232.

Asl, M.E., Koohbanani, N.A., Frangi, A.F., Gooya, A., 2017. Tracking and diameter estimation of retinal vessels using Gaussian process and radon transform. J. Med. Imaging 4 (3), 034006.

Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N., 2015. Trainable COSFIRE filters for vessel delineation with application to retinal images. Med. Image Anal. 19 (1), 46–57.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495.

Budai, A., Bock, R., Maier, A., Hornegger, J., Michelson, G., 2013. Robust vessel segmentation in fundus images. Int. J. Biomed. Imag. 2013.

Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M., 1989. Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans. Med. Imag. 8 (3), 263–269.

Chen, C., Chuah, J.H., Ali, R., Wang, Y., 2021. Retinal vessel segmentation using deep learning: a review. IEEE Access 9, 111985–112004.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M., Campilho, A., 2017. End-to-end adversarial retinal image synthesis. IEEE Trans. Med. Imaging 37 (3), 781–791.

Dasgupta, A., Singh, S., 2017. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: 2017 IEEE 14th International Symposium on Biomedical Imaging. ISBI 2017, IEEE, pp. 248–251.

Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., et al., 2014. Feedback on a publicly distributed image database: the Messidor database. Image Anal. Stereol. 33 (3), 231–234.

Ding, Y., Ward, W.O., Duan, J., Auer, D., Gowland, P., Bai, L., 2015. Retinal vasculature classification using novel multifractal features. Phys. Med. Biol. 60 (21), 8365.

Dong, F., Wu, D., Guo, C., Zhang, S., Yang, B., Gong, X., 2022. CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation. Comput. Biol. Med. 105651.

Du, X.-F., Wang, J.-S., Sun, W.-z., 2021. UNet retinal blood vessel segmentation algorithm based on improved pyramid pooling method and attention mechanism. Phys. Med. Biol. 66 (17), 175013.

Farnell, D., 2007. Automated retinal image analysis (ARIA) data set. http://www.damianjjfarnell.com/?page_id=276.

Farnell, D.J., Hatfield, F.N., Knox, P., Reakes, M., Spencer, S., Parry, D., Harding, S.P., 2008. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. J. Franklin Inst. 345 (7), 748–765.

Feng, Z., Yang, J., Yao, L., 2017. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. In: 2017 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1742–1746.

Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging 37 (7), 1597–1605.

Fu, H., Xu, Y., Lin, S., Kee Wong, D.W., Liu, J., 2016. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 132–139.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2019. Res2net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43 (2), 652–662.

Ghaderi, R., Hassanpour, H., Shahiri, M., 2007. Retinal vessel segmentation using the 2-D morlet wavelet and neural network. In: 2007 International Conference on Intelligent and Advanced Systems. IEEE, pp. 1251–1255.

Ghiasi, G., Lin, T.-Y., Le, Q.V., 2018. Dropblock: A regularization method for convolutional networks. In: Advances in Neural Information Processing Systems. Vol. 31.

Girshick, R., Donahue, J., Darrell, T., Berkeley, U., Malik, J., 2014. R-CNN: Region-based convolutional neural networks. In: Proc. Comput. Vis. Pattern Recognit.. pp. 2–9.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. Vol. 27. Curran Associates, Inc, URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J., 2019. Ce-Net: Context encoder network for 2d medical image segmentation. IEEE Trans. Med. Imaging 38 (10), 2281–2292.

Guo, C., Szemenyei, M., Pei, Y., Yi, Y., Zhou, W., 2019. SD-UNet: A structured dropout U-Net for retinal vessel segmentation. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering. BIBE, IEEE, pp. 439–444.

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., Fan, C., 2021. Sa-Unet: Spatial attention U-Net for retinal vessel segmentation. In: 2020 25th International Conference on Pattern Recognition. ICPR, IEEE, pp. 1236–1242.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hoover, A., Kouznetsova, V., Goldbaum, M., 2000. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imag. 19 (3), 203–210.

Hu, H., Liu, Z., 2021. Retinal vessel segmentation based on recurrent convolutional skip connection U-Net. In: 2021 4th International Conference on Intelligent Autonomous Systems. ICoIAS, IEEE, pp. 65–71.

Hu, X., Wang, L., Li, Y., 2022. HT-Net: A hybrid transformer network for fundus vessel segmentation. Sensors 22 (18), 6782.

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

Huang, Z., Sun, M., Liu, Y., Wu, J., 2022. CSAUNet: A cascade self-attention u-shaped network for precise fundus vessel segmentation. Biomed. Signal Process. Control 75, 103613.

Jebaseeli, T.J., Durai, C.A.D., Peter, J.D., 2019. Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM. Optik 199, 163328.

Jin, Q., Meng, Z., Pham, T.D., Chen, Q., Wei, L., Su, R., 2019. DUNet: A deformable network for retinal vessel segmentation. Knowl.-Based Syst. 178, 149–162.

Kälviäinen, H., Uusitalo, H., Parkkinen, J., 2007. Imageret. https://www.it.lut.fi/project/imageret/#DOWNLOAD.

Kamran, S.A., Hossain, K.F., Tavakkoli, A., Zuckerbrod, S.L., Sanders, K.M., Baker, S.A., 2021. RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 34–44.

Kar, S.S., Maity, S.P., 2016a. Blood vessel extraction and optic disc removal using curvelet transform and kernel fuzzy c-means. Comput. Biol. Med. 70, 174–189.

Kar, S.S., Maity, S.P., 2016b. Retinal blood vessel extraction using tunable bandpass filter and fuzzy conditional entropy. Comput. Methods Programs Biomed. 133, 111–132.

Kauppi, T., Kalesnykiene, V., Kamarainen, J.-K., Lensu, L., Sorri, I., Uusitalo, H., Kälviäinen, H., Pietilä, J., 2006. DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms. In: Machine Vision and Pattern Recognition Research Group. Vol. 73. Lappeenranta University of Technology, Finland, pp. 1–17.

Khalaf, A.F., Yassine, I.A., Fahmy, A.S., 2016. Convolutional neural networks for deep feature learning in retinal vessel segmentation. In: 2016 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 385–388.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. Vol. 25. Curran Associates, Inc, URL https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Lahiri, A., Ayush, K., Kumar Biswas, P., Mitra, P., 2017. Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: Automated vessel segmentation in retinal fundus image as test case. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 42–48.

Laibacher, T., Weyde, T., Jalali, S., 2019. M2U-Net: Effective and efficient retinal vessel segmentation for real-world applications. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. CVPRW, pp. 115–124. http://dx.doi.org/10.1109/CVPRW.2019.00020.

Lam, B.S., Gao, Y., Liew, A.W.-C., 2010. General retinal vessel segmentation using regularization-based multiconcavity modeling. IEEE Trans. Med. Imag. 29 (7), 1369–1381.

Li, D., Dharmawan, D.A., Ng, B.P., Rahardja, S., 2019. Residual U-Net for retinal vessel segmentation. In: 2019 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1425–1429.

Li, X., Ding, J., Tang, J., Guo, F., 2022a. Res2Unet: A multi-scale channel attention network for retinal vessel segmentation. Neural Comput. Appl. 1–15.

Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T., 2015. A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging 35 (1), 109–118.

Li, L., Verma, M., Nakashima, Y., Nagahara, H., Kawasaki, R., 2020. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3656–3665.

Li, Y., Zhang, Y., Liu, J.-Y., Wang, K., Zhang, K., Zhang, G.-S., Liao, X.-F., Yang, G., 2022b. Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation. IEEE Trans. Cybern. 1–14. http://dx.doi.org/10.1109/TCYB.2022.3194099.

Liskowski, P., Krawiec, K., 2016. Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging 35 (11), 2369–2380.

Liu, B., Gu, L., Lu, F., 2019. Unsupervised ensemble strategy for retinal vessel segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 111–119.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Luc, P., Couprie, C., Chintala, S., Verbeek, J., 2016. Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408.

Mahapatra, S., Agrawal, S., Mishro, P.K., Pachori, R.B., 2022. A novel framework for retinal vessel segmentation using optimal improved frangi filter and adaptive weighted spatial FCM. Comput. Biol. Med. 147, 105770.

Mendonca, A.M., Campilho, A., 2006. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans. Med. Imaging 25 (9), 1200–1213.

Neto, L.C., Ramalho, G.L., Neto, J.F.R., Veras, R.M., Medeiros, F.N., 2017. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. Expert Syst. Appl. 78, 182–192.

Niemeijer, M., Van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G., Sánchez, C.I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., et al., 2009. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans. Med. Imaging 29 (1), 185–195.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Oliveira, A., Pereira, S., Silva, C.A., 2018. Retinal vessel segmentation based on fully convolutional neural networks. Expert Syst. Appl. 112, 229–242.

Organization, W.H., et al., 2019. World report on vision. URL https://www.who.int/publicationsdetail/worldreportonvision.

Orlando, J.I., Barbosa Breda, J., Keer, K.v., Blaschko, M.B., Blanco, P.J., Bulant, C.A., 2018. Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 65–73.

Owen, C.G., Rudnicka, A.R., Mullen, R., Barman, S.A., Monekosso, D., Whincup, P.H., Ng, J., Paterson, C., 2009. Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program. Investigat. Ophthalmol. Vis. Sci. 50 (5), 2004–2010.

Ozgunalp, U., Fan, R., Serener, A., 2020. Semantic segmentation of retinal vessels using SegNet. In: 2020 28th Signal Processing and Communications Applications Conference. SIU, pp. 1–4. http://dx.doi.org/10.1109/SIU49456.2020.9302055.

Perez-Rovira, A., MacGillivray, T., Trucco, E., Chin, K., Zutis, K., Lupascu, C., Tegolo, D., Giachetti, A., Wilson, P.J., Doney, A., et al., 2011. VAMPIRE: Vessel assessment and measurement platform for images of the Retina. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 3391–3394.

Prentašić, P., Lončarić, S., Vatavuk, Z., Benčić, G., Subašić, M., Petković, T., Dujmović, L., Malenica-Ravlić, M., Budimlija, N., Tadić, R., 2013. Diabetic retinopathy image database (DRiDB): a new database for diabetic retinopathy screening programs research. In: 2013 8th International Symposium on Image and Signal Processing and Analysis. ISPA, IEEE, pp. 711–716.

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

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Shen, X., Xu, J., Jia, H., Fan, P., Dong, F., Yu, B., Ren, S., 2022. Self-attentional microvessel segmentation via squeeze-excitation transformer Unet. Comput. Med. Imaging Graph. 97, 102055.

Sheng, B., Li, P., Mo, S., Li, H., Hou, X., Wu, Q., Qin, J., Fang, R., Feng, D.D., 2018. Retinal vessel segmentation using minimum spanning superpixel tree detector. IEEE Trans. Cybern. 49 (7), 2707–2719.

Shin, S.Y., Lee, S., Yun, I.D., Lee, K.M., 2019. Deep vessel segmentation by learning graphical connectivity. Med. Image Anal. 58, 101556.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Soomro, T.A., Afifi, A.J., Gao, J., Hellwich, O., Zheng, L., Paul, M., 2019. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. Expert Syst. Appl. 134, 36–52.

Srinidhi, C.L., Aparna, P., Rajan, J., 2018. A visual attention guided unsupervised feature learning for robust vessel delineation in retinal images. Biomed. Signal Process. Control 44, 110–126.

Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B., 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging 23 (4), 501–509.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. Vol. 30.

Wang, D., Haytham, A., Pottenburgh, J., Saeedi, O., Tao, Y., 2020. Hard attention net for automatic retinal vessel segmentation. IEEE J. Biomed. Health Inf. 24 (12), 3384–3396.

Wang, Y., Ji, G., Lin, P., Trucco, E., 2013. Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. Pattern Recognit. 46 (8), 2117–2133.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8798–8807.

Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., Qin, J., 2021. Scs-Net: A scale and context sensitive network for retinal vessel segmentation. Med. Image Anal. 70, 102025.

Wu, C., Zou, Y., Yang, Z., 2019. U-GAN: generative adversarial networks with U-Net for retinal vessel segmentation. In: 2019 14th International Conference on Computer Science & Education. ICCSE, IEEE, pp. 642–646.

Xiao, X., Lian, S., Luo, Z., Li, S., 2018. Weighted res-unet for high-quality retina vessel segmentation. In: 2018 9th International Conference on Information Technology in Medicine and Education. ITME, IEEE, pp. 327–331.

Xie, S., Tu, Z., 2015. Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1395–1403.

Yan, Z., Yang, X., Cheng, K.-T., 2018. A three-stage deep learning model for accurate retinal vessel segmentation. IEEE J. Biomed. Health Inf. 23 (4), 1427–1436.

Yin, Y., Adel, M., Bourennane, S., 2012. Retinal vessel segmentation using a probabilistic tracking method. Pattern Recognit. 45 (4), 1235–1244.

Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J.P., Duits, R., ter Haar Romeny, B.M., 2016. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE Trans. Med. Imaging 35 (12), 2631–2644.

Zhang, H., Zhong, X., Li, Z., Chen, Y., Zhu, Z., Lv, J., Li, C., Zhou, Y., Li, G., 2022. TiM-Net: transformer in M-Net for retinal vessel segmentation. J. Healthc. Eng. 2022.

Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L., 2018a. Synthesizing retinal and neuronal images with generative adversarial nets. Med. Image Anal. 49, 14–26.

Zhao, H., Li, H., Maurer-Stroh, S., Guo, Y., Deng, Q., Cheng, L., 2018b. Supervised segmentation of un-annotated retinal fundus images by synthesis. IEEE Trans. Med. Imaging 38 (1), 46–56.

Zhao, Y., Rada, L., Chen, K., Harding, S.P., Zheng, Y., 2015. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. IEEE Trans. Med. Imaging 34 (9), 1797–1807.

Zhao, Y., Zheng, Y., Liu, Y., Zhao, Y., Luo, L., Yang, S., Na, T., Wang, Y., Liu, J., 2017. Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter. IEEE Trans. Med. Imaging 37 (2), 438–450.

Zhuang, J., 2018. LadderNet: Multi-path networks based on U-Net for medical image segmentation. arXiv preprint arXiv:1810.07810.