IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 7, 2019, accepted May 24, 2019, date of publication June 3, 2019, date of current version June 12, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2920616*

# Deep Learning Models for Retinal Blood Vessels Segmentation: A Review

**TOUFIQUE AHMED SOOMRO[1,5], (Member, IEEE), AHMED J. AFIFI[2], LIHONG ZHENG[1], SHAFIULLAH SOOMRO[3], JUNBIN GAO[4], OLAF HELLWICH[2], AND MANORANJAN PAUL[1], (Senior Member, IEEE)**

[1]School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW 2795, Australia
[2]Computer Vision and Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany
[3]Department of Basic Science and Related Studies, Quaid-e-Awam University of Engineering, and Science Technology, Nawabshah 67480, Pakistan
[4]The University of Sydney Business School, The University of Sydney, Camperdown, NSW 2006, Australia
[5]Electronic Engineering Department, QUEST, Larkana Campus, Pakistan

Corresponding author: Toufique Ahmed Soomro (etoufique@yahoo.com; toufique_soomro@quest.edu.pk)

This work was supported by CSU Compact Fund.

**ABSTRACT** This paper presents a comprehensive review of the principle and application of deep learning in retinal image analysis. Many eye diseases often lead to blindness in the absence of proper clinical diagnosis and medical treatment. For example, diabetic retinopathy (DR) is one such disease in which the retinal blood vessels of human eyes are damaged. The ophthalmologists diagnose DR based on their professional knowledge, that is labor intensive. With the advances in image processing and artificial intelligence, computer vision-based techniques have been applied rapidly and widely in the field of medical images analysis and are becoming a better way to advance ophthalmology in practice. Such approaches utilize accurate visual analysis to identify the abnormality of blood vessels with improved performance over manual procedures. More recently, machine learning, in particular, deep learning, has been successfully implemented in this area. In this paper, we focus on recent advances in deep learning methods for retinal image analysis. We review the related publications since 1982, which include more than 80 papers for retinal vessels detections in the research scope spanning from segmentation to classification. Although deep learning has been successfully implemented in other areas, we found only 17 papers so far focus on retinal blood vessel segmentation. This paper characterizes each deep learning based segmentation method as described in the literature. Analyzing along with the limitations and advantages of each method. In the end, we offer some recommendations for future improvement for retinal image analysis.

**INDEX TERMS** Retinal colour fundus images, convolutional neural networks, retinal vessels segmentation.

## I. INTRODUCTION

Eye diseases, such as diabetic retinopathy (DR) and Diabetic maculopathy (MD), are the main causes of global blindness. MD has a long preclinical phase where visual acuity is not affected and once the patient presents to the eye clinic the vision loss can often not be recovered. Early detection can prevent disease progression and protect vision. Therefore the study and the analysis of retinal vessel geometric characteristics such as vessel diameter, branch angles, and branch lengths have become the basis of medical applications related to early diagnosis and effective monitoring of retinal pathology. Retinal image assessment by ophthalmologists is a vital step for identification of retinal pathology.

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Sarwar Bajwa.

The computerized image analysis has been successfully applied in various medical applications since 1982 [1], [2]. Aiming at increasing both efficiency and the accuracy, various automatic computerized medical image analysis approaches have been developed since then [3]. Image segmentation has been seen as a wide application at that stage before the 1990s. Recently, more artificial intelligence-based approaches such as supervised methods have become popular in this area. Pattern recognition and machine learning methods are very productive due to their success in medical image analysis system [1], [4].

However, implementing a robust computerized method for medical image analysis is still a challenging task nowadays due to the complexity and heterogeneity of the retinal images [5]. Often it requires user interaction to diagnose the diseases by searching the most distinguished features from

the images [6]. There has been a consistent research effort made in the area of retinal image processing. Some data mining approaches have targeted extracting the most efficient features. Low-level features such as morphological features, colour features, vessel width, texture, sharpness of vessel boundary as well as the coarseness inside a blood vessel have been incorporated, although some require advanced technology and are not suitable for remote clinics without trained specialists. Middle-level features such as a saliency map extracted from fundus images are applied to train classifiers with the purpose of extracting retinal blood vessels [7]. Others treat such tasks as solving an optimal function in terms of minimizing the energy cost or distances. Examples include graph-based approach, K-nearest neighbours (KNN), Gaussian mixture model and entropy thresholding based [8], [9] model.

Different machine learning based approaches have shown better performance in this area. Support Vector Machine (SVM), decision tree, AdaBoost, Naive Bayes and Random Forest [10], [11] are examples of these and have been investigated by researchers. Moreover, ensemble of machine learning classifying algorithms were proposed to improve the final accuracy while combining the outputs of different retinal image processing algorithms [3].

Deep learning based approaches are playing an important role in the segmentation of digital images, and these approaches outperform existing methods based on image processing techniques. The main advantage of using deep learning is the suitability of training of large database for classification or segmentation which makes deep learning based approaches different and effective from other previous implemented methods. In 2006, a deep convolutional neural network (CNN) structure was designed for different imaging tasks of natural images [12] that intended to classify different features of colour images. AlexNet [12], VGGNet [13] and the inception of the architecture of GoogLeNet [14] have been applied successfully to a large variety of objects identification and semantic segmentation. In the area of retinal image analysis, the CNN is used to detect retinal vessels and to classify patch features into different regions of vessels and in different connecting cases of the vessels. Several researchers have worked on CNN based models for detection of vessels but the accuracy diminishes as the blood vessels get smaller in diameter.

Krizhevsky *et al.* [12] contributed to the watershed transformation of the image by using the ImageNet challenge in 2012, and they proposed a CNN model called AlexNet, winning the challenge with a significant margin. Due to the good performance of CNN model, the medical image analysis community too has considered implementation of CNN.

In this paper, we review the deep learning based articles published within the recent five years in medical image analysis. More specifically, this paper focuses on papers on retinal color fundus images for segmentation as well as abnormality analysis in the retinal fundus image. Three main purposes of doing such a survey of deep learning methods on retinal image analysis include

1) Demonstrating that the deep learning based methods have influence in the retinal image analysis as compared to other techniques.
2) Highlighting the specific contribution which solves the accurate segmentation of retinal blood vessels problems.
3) Studying different proposed deep learning models on retinal images and observing their performance as well as the limitation, and proposing the suitable methodology to overcome such limitations in order to improve performance.

We also list different types of medical image analysis applications using deep learning in Figure 1. Figure 2 shows the number of papers that discuss different modalities such as MRI, CT, X-ray and fundus images. The first deep learning based paper on retinal fundus images was published (Kaggle Diabetic Retinopathy challenge 2015, image from van Grinsven *et al.* [15]). There have been only 17 deep learning based papers published since then (as shown in Figure 2), 12 of which are related to retinal vessels segmentation and that is the focal area of our interest.

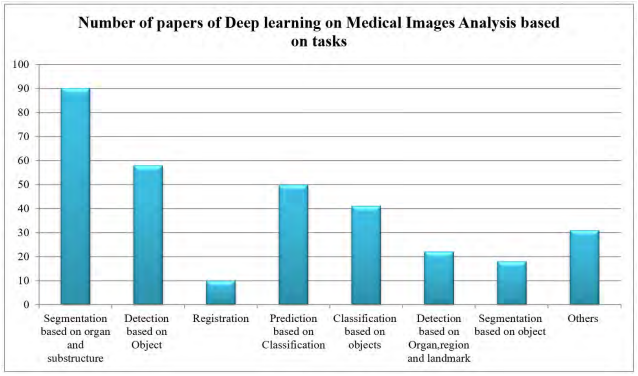



**FIGURE 1.** Number of papers describing deep learning application on medical images based on tasks.

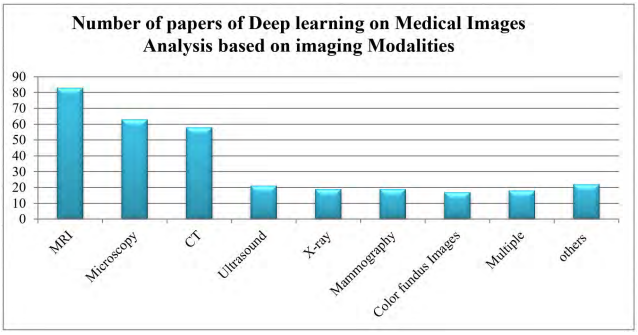



**FIGURE 2.** Number of papers describing deep learning applications on medical images based on imaging modalities.

Here, we present a detailed analysis of these 12 methods by investigating and comparing their advantages as well as

limitations. Using this approach we identify the best method for fast and accurate retinal image analysis that will hopefully ease the issues faced by Australia's regional community of unmet demand for professional clinical services. The finding of this paper will assist the early diagnose of retinal diseases by providing an efficient and reliable approach based on the most recent machine learning analysis. This paper foresees the future direction of how the new technology will help limit vision loss, improving the early diagnosis accuracy for effective treatments and thereby significantly improve the life quality of patients with eye diseases. This paper explores and identifies a flexible and reliable computerized system that can automatically identify the change in the blood vessels and predict the potential disease as early as possible.

The paper is organized as follows. Section II contains an overview of deep learning methods. Section III explains the CNN in details. Section IV contains detailed explanation of importance of segmentation by using CNN. Section V describes the details of measuring parameters and the used database in the existing methods. Section VI reviews the state of art of retinal vessels segmentation methods based on deep learning, and analyze the methodologies of each method. Section VII analyze the main issues in retinal vessels methods based deep learning. Section VIII outlines the concluding remarks and future research directions.

## II. OVERVIEW OF DEEP LEARNING METHODS

Deep learning is one of machine learning methods. It consists of hierarchical structured layers that can translate input data to meaningful outputs in a black box model. Deep learning methods have wide applications within different research analysis such as graphical modeling of data, neural networks, parameters optimization, image analysis, pattern recognition and signal processing. Many deep learning models in various applications are based on the model, proposed by Yann LeCun for hand written recognition by using deep supervised backpropagation convolutional network [16].

In recent six years, deep learning has become a hot research area in the field of computer vision and image processing owing to the achievement in different applications. Krizhevsky *et al.* [12] was the first to propose the deep CNN for image classification through ImageNet classification challenge.

Along with image classification, object detection and segmentation tasks were also achieved by using CNN models as reported in many research works [17].

A rich family of deep learning methods have been developed since 1990 [18], such as Deep Neural Networks (DNN) [18], Auto-encoders(AEs) and Stacked Auto-encoders (SAEs) Neural Network [19], Restricted Boltzmann Machines (RBM) [20], Deep Belief Network (DBN) [21] and much more. Among all these studies, deep convolutional neural network demonstrate with better performance on a variety of tasks in images/signal processing and computer vision.

In this section, we elaborate the basic concepts of deep from its initial implementation to advance level.

### A. DEEP NEURAL NETWORK

Deep neural network is a simple normal neural network with more hidden layers so that it becomes deeper. The deep neural architecture can be considered as the generalization of a linear or logistic regression neural network architectures. Each neuron is activated in a linear combination of input and some learning parameters, which are followed by an element-wise nonlinear formation. Mathematically, it is defined as.

$$a = \sigma(w^T x + b). \tag{1}$$

where $a$ represents activation neuron, $\sigma$ represents the element wise non-linearity, $x$ represents the input and $w$ and $b$ represent a set of learning parameters.

A neural network architecture consists of a number of layers $L$ of a weighted neuron through which activation is performed. Multi-Layer Perceptron (MLP) is a class of feedforward neural network with two or more layers between input and output layers. The feedforward means that data is flowing in one direction from input to output layers. The backpropagation learning algorithm is used to train the MLP. MLP is used in many applications such as pattern classification, recognition, approximation, and prediction. MLP mostly solves the problems which are not linearly separable as shown in Figure 3.

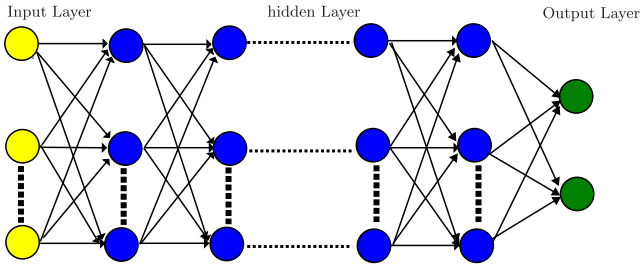



**FIGURE 3.** Deep neural network.

Many researchers use the DNN to train a model from data and it received popularity in 2007 and 2008 [22]–[24]. However, there are alternative tactics to train a whole deep neural network end-to-end in a supervised way. These alternative tactics can be better implemented by specified type of neural network, i.e., the convolutional neural network (CNN).

Nowadays, CNN gets more popular in medical image processing and becomes a good choice for the researchers in medical image analysis. The following subsections explain the basic types of neural networks and step by step implementation of the method with their limitations and advantages.

### B. AUTO-ENCODERS (AEs) AND STACKED AUTO-ENCODERS (SAEs) NEURAL NETWORKS

Auto-encoders are neural networks that is used for unsupervised learning where the target values are set to be equal to the inputs. They are arranged by wright matrice and biases

as shown in the Equation 2. If the hidden layer has the same size as the input, and there is no further addition of non-linearities then model is used for learning the identity function. The main vital parameter in the AEs network is the use of a non-linear activation function to calculate the latent representation. In most case, the $|x|$ is larger than $|h|$ ( where $h$ represents hidden layer) because it makes data be projected onto a lower dimensional subspace for representing the latent structure of the input. Later, the sparsity and regularization can be utilized for the formulation of relevant structure.

$$h = \sigma(W_{x,h}x + b_{x,h}). \tag{2}$$

In 2010. Vincent *et al.* [19] proposed a denoising auto-encoder. The proposed model is trained to reconstruct the input from data where a noise is added to the input data. There is another representation of AEs in which AEs layers are placed on top of each other. Such an arrangement of AEs is known as Stacked Auto-encoders (SAEs) Neural Networks shown in Figure 4. A staked autoencoder is a neural network consisting of multiple autoencoders where the output of each layer is wired to the inputs of the successive layer.

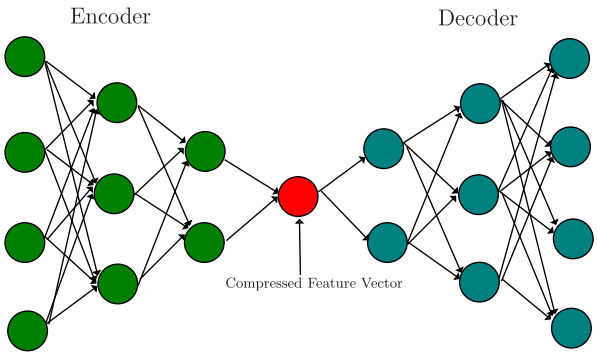



**FIGURE 4.** Deep auto-encoder (AE).

### C. RESTRICTED BOLTZMANN MACHINES (RBMs) AND DEEP BELIEF NETWORKS (DBNs)

Restricted Boltzmann Machines (RBMs) are a the class of machine learning model based on the Markov Random Field (MRF) that contains an input layer and hidden layer through the latent feature representation shown in Figure 5.

The layers connections are based on the bidirectional way, it means that input vector $x$ can be achieved from the latent feature representation $h$. This RBM model is known as a generative model, it meant that researchers can get a sample of RBM data, and generate new data points impending from the distribution of the latent feature on which data are trained. This whole system can be explained by energy function. The energy function is achieved for specified state $(x, h)$ of input and hidden connected units of the network. It is represented mathematically as Equation (3) along with $c$ and $b$ are bias terms.

$$E(x, h) = h^T Wx - c^T x - b^T h. \tag{3}$$

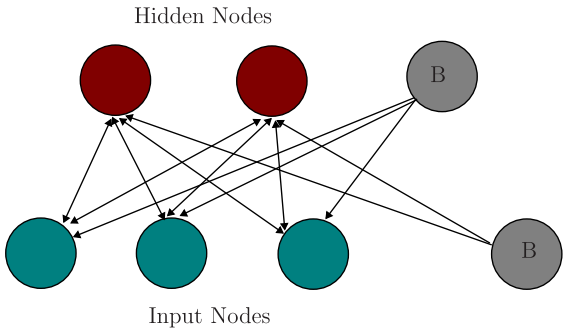



**FIGURE 5.** Restricted boltzmann machines (RBMs) model.

The energy function system has different states, and the probability state of the system is well explained by simply tossing the energy into an exponential function, and normalised each possible state as shown in Equation (4) below

$$P(x, h) = \frac{1}{z}\exp\{-E(x, h)\}. \tag{4}$$

where $Z$ is the partition function and it is generally in an intractable form. This intractable function is a conditional interference in the form of computing $h$ condition on input data and vice versa, and it gives result in the form of Equation (5). The similar expression is held for $P(h_j|x)$ due to symmetric network.

$$P(h_j|x) = \frac{1}{1 + \exp\{-b_j - W_j x\}}. \tag{5}$$

Later, researchers [23], [24] analyzed the RBMs and SAEs network, and gave the formation new technique named as Deep Belief Networks (DBNs). DBNs are basically SAEs but the AEs layers are replaced by RBMs. To train DBNs, the first step is to learn a layer of features from visible units. Then, the the learned layers are treated as visible units to train the next layer. At the end, the whole DBN is trained when the learning of the final hidden layer is achieved.

## III. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Convolutional Neural Networks are biologically-inspired of multi-layer perceptrons (MLPs). They are well-known deep learning architectures inspired by the natural visual perception mechanism of the living creatures. The history of CNNs started with the experiments conducted by Hubel and Wiesel in 1959 [25]. They found that the cells in animals' visual cortex are responsible for detecting the light in the receptive fields. In 1990, LeCun *et al.* [26] published a paper where they described the modern framework of CNN. They introduced a neural network called LeNet-5 which is used to classify the handwriting digits. They used backpropagation algorithm to train the neural network. However, due to the lack of data and the computation power at that time, the proposed networks couldn't perform well in large-scale problems. After that, many researchers have developed methods to overcome some problems encountered in training deep CNNs. In 2012, Krizhevsky *et al.* [12] proposed a deep CNN

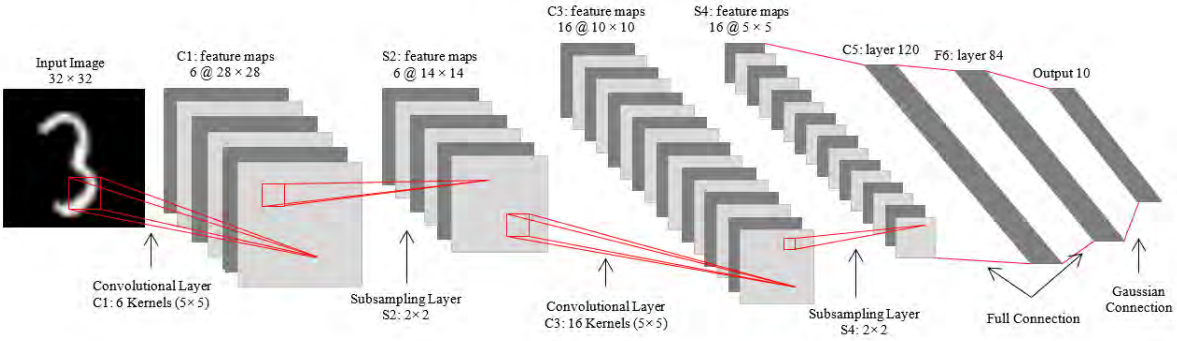



**FIGURE 6.** LeNet-5 architecture [26].

architecture to solve the problem of image classification. They showed significant improvements and outperformed results upon previous methods. They won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [27]. The proposed architecture is similar to LeNet-5 but with a deeper structure, and after that it was known as AlexNet. With the success of AlexNet, many CNNs architecture have been proposed to improve the performance and get more accurate results such as ZFNet [28], VGGNet [13], GoogleNet [14], and ResNet [29]. ResNet is the deepest architecture and it is 20 times deeper than AlexNet and 8 times deeper than VGGNet. It is observed that deeper architectures are better for the classification tasks. The reason behind this is that the network can approximate the output because of the increasing of the nonlinearity and extracting more representative features.

### A. CNN ARCHITECTURE COMPONENTS

Many CNN architectures have been proposed for solving different tasks. The general computation blocks are similar, and they are different in some layers depending on the tasks they want to solve. For example, LeNet-5 [26] and AlexNet [12] are almost similar in the general building blocks but AlexNet is deeper than LeNet-5. The main blocks of the CNNs are the convolutional layers, the pooling layers, and the fully-connected layers (FC). Each block has its own specifications and features, and it can vary from one architecture to another according to the problem to be solved. Below, we will explain clearly each building block of the CNNs. Figure 6 shows LeNet-5 architecture [26].

#### 1) CONVOLUTIONAL LAYER

Convolutional layers are the main building blocks of the CNNs. They learn the feature representation of the input images by performing convolutions over the inputs. The convolutional layer consists of several kernels which are used to compute different features from the input images. They ensure that the local connectivity-neurons are connected to a small region of the input which is known as the receptive field. The extracted feature maps are calculated by convolving the input with the kernels and then add the bias parameters to the feature. The convolutional layer has many kernels, and they are applied to the input image to calculate the output feature map. Each kernel is shared by all special locations of the input. The advantage of weight sharing is to reduce the complexity of the model and the training process of the network becomes easier. Mathematically, consider $x$ as the input image, $W$ is the kernel, and $b$ is the bias for the convolutional layer. The feature map $z$ generated from this layer is calculated as:

$$z = Wx + b. \tag{6}$$

Many researchers proposed different types of convolutional layers to improve the feature representation and to learn some kind of invariance. Tiled CNN [30] is one kind of the enhanced convolutional layer that tiles and multiplied different feature maps learn rotation and scale invariant features. Dilated CNN [31] is another recent development of the standard CNN, where more hyper-parameters are introduced to the convolutional layer. This strategy enhances the performance for tasks which need a large receptive field when they make the prediction, such as scene segmentation and speech synthesis and recognition. Also, there are other improved convolutional layers such as Network in Network (NIN) [32] for classification task which replaces the normal convolutional layer with multilayer perceptron convolution layer (mlpconv). It is like a micro neural network with more complex structures rather than conventional convolutional layer that uses linear filters followed by a nonlinear activation function. NIN enhances the model discriminability for local patches within the receptive field for classification task. Transposed Convolution [28] is another type of convolutional layer that is used to reconstruct the input again in regression tasks such as image segmentation, depth estimation, visualization, and image super-resolution. In the literature, this layer is called deconvolutional layer. The deconvolutional layer first upsamples the input to a specified factor and then performs a normal convolution operation on the upsampled result. Figure 7 illustrates the convolution operation.

#### 2) ACTIVATION FUNCTION

CNNs have some linear components and nonlinear components. The activation functions are the nonlinear com-

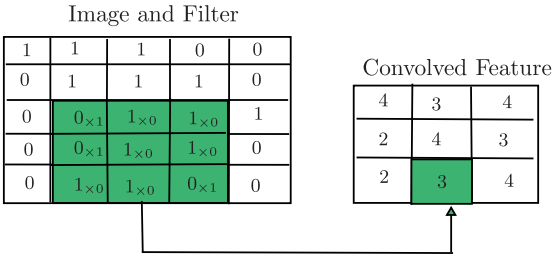



**FIGURE 7.** Convolution operation [27].

ponents which follow the convolutional layers to introduce the nonlinearities to the CNN to detect the nonlinear features and to improve the CNN performance. Rectified Linear Unit (ReLU) is one of the most popular activation function used in CNNs. It has been shown that CNNs can be trained efficiently using ReLU [33]. ReLU is defined as:

$$a = \max(z, 0). \tag{7}$$

where $z$ is the input to the activation function and $a$ is the output. ReLU keeps the positive part of the input and prunes the negative part to zero. Another version of ReLU is leaky ReLU (LReLU) [34] that defines a parameter $\lambda$ in range $(0, 1)$ to compress the negative part rather than mapping it to zero. Mathematically, LReLU is defined as:

$$a = \max(z, 0) + \lambda \min(z, 0). \tag{8}$$

This makes a small and non-zero gradient when the unit is not active (negative value).

Exponential Linear Unit (ELU) is another activation function that enables faster learning of the CNN and improves the accuracy of the classification task. Like ReLU and LReLU, ELU [35] sets the positive values to identity and the negative part is used for fast learning and it is robust to noise. Mathematically, ELU is defined as:

$$a = \max(z, 0) + \min(\lambda(e^z - 1), 0). \tag{9}$$

where $\lambda$ is a controlling parameter to saturate ELU for negative inputs.

The last activation function we want to talk about is the sigmoid function ($\sigma$) [36]. Sigmoid function is used often in artificial neural networks to introduce the nonlinearity in the model. It takes real numbers and squashes them into range $(0, 1)$. Mathematically, the sigmoid function is defined as:

$$a = \sigma(z) = \frac{1}{1 + e^{-z}}. \tag{10}$$

In general, the activation functions are applied to the output of convolutional layer in the CNN to add the nonlinearity to the output to project the values from some range to a desired range. The "Activation Function" term is biologically inspired. In real brain, neurons get signals from other neurons, and decided whether or not to fire by taking the cumulative input into account. This decision based on the cumulative input is the output from the activation function. Figure 8 shows the activation function we have discussed.

#### 3) POOLING LAYER

The purpose of the pooling layer is to ensure the shift-invariance and lowers the computational burden by reducing the resolution of the feature maps. It is usually placed after the convolutional layer. It takes the feature map which is generated from the convolutional layer and outputs a single value for each receptive field (pooling window) according to the pooling operation. The pooling layer performs max pooling [37], sum, and mean pooling [38]. Figure 9 shows different pooling operations.

Also, there are other versions of pooling layers proposed for some tasks, such as Spatial Pyramid Pooling (SPP) [38] that can generates a fixed length of features regardless of the input size.

#### 4) FULLY CONNECTED LAYERS (FC)

In classification tasks, fully connected layers (FC) [12] are used at the end of the CNN after the convolutional layers and the pooling layers. Fully connected layers aim to generate specific semantic information. The neurons in the fully connected layers have full connection to all neurons in the previous layer. It can be considered as a special case of a convolutional layer with the receptive field size is equal to one. Usually, dropout is used after the fully connected layers to avoid the CNN from overfitting.

### B. REGULARIZATION

One of the most problematic issues regarding CNN training is overfitting. Overfitting happens when the model fits too well to the training dataset, and it cannot generalize to new examples that were not in the training dataset. So, overfitting is an unneglectable problem in deep CNNs [39]. There are many proposed solutions to reduce the overfitting effectively, such as Dropout [40], $L1$ regularization, and $L2$ regularization [39]. In deep learning, Dropout is widely used as regularization after the fully connected layers. It deletes or deactivates some neurons so that not all connections between the layers are activated at that time during training. It can also be applied after the convolutional layers. However, it is not preferable to add them in the first layers because dropout causes information to get lost. And if the information is lost in the first layers, it will be also lost for the whole network and this will affect the performance of the network. During testing time, dropout layers are bypassed and they are not active.

### C. OPTIMIZATION AND LOSS FUNCTIONS

Training a CNN is a problem of global optimization. To find the best values of the weights for each layer, a loss function should be selected and minimized. To optimize the CNN parameters, Stochastic Gradient Descent (SGD) [41] is commonly used.

To optimize a deep CNN, many steps should be done such as preparing the training dataset, designing the CNN and initializing the weights, choosing the loss function, and

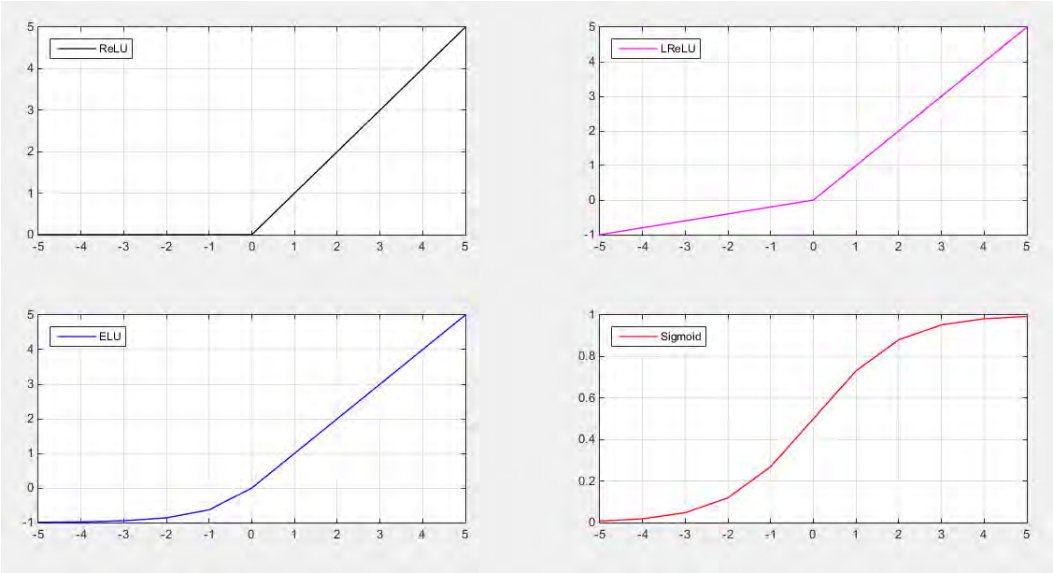



**FIGURE 8.** Different activation functions.

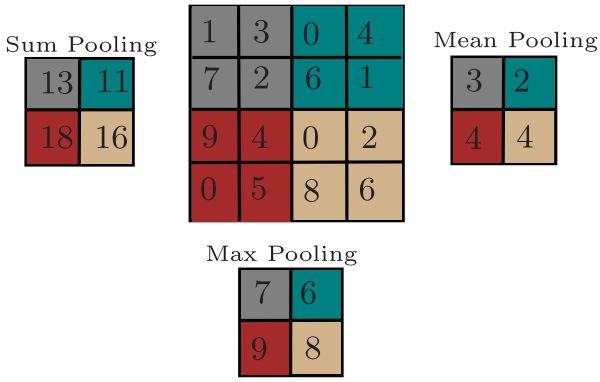



**FIGURE 9.** Different pooling operations with 2 × 2 filters and stride 2.

training the CNN using backpropagation to update the network weights. Below, we discuss each step briefly.

It is well known that deep CNN almost depends on the availability of the training data. The more training data we have, the better trained model we can get. However, in some applications such as medical problems, the available datasets are small and we cannot rely on them to train a deep CNN. To overcome this issue, data augmentation [42] is applied to the training dataset. The purpose of the data augmentation is to increases the training data by transforming the images into new data without altering the nature of the content. The commonly used data augmentation methods include simple geometric transformations such as mirroring, shifting, scaling, rotating, and spatial cropping [12]. This method not only helps to overcome the problem of training data scarcity but also trains the network and updates the network weights more accurately.

After preparing the training data, global data normalization is usually applied on the training dataset to transform the data to zero-mean and unit variance. However, during training and as the data flow deeper through the network, the distribution of the input data changes, which affects the training process and the network accuracy. Batch Normalization (BN) is applied on the input data in some layers to avoid this problem [43]. It fixes the means and the variances of the layer inputs by computing them after each mini-batch rather than the entire training set.

Deep CNN has numerous of parameters to be optimized. A proper initializing of these parameters is important for fast convergence and to avoid the vanishing gradient problem. Many methods have been proposed to initialize the weights. In [12], Krizhevsky *et al.* initialize the weights of the network from a zero-mean Gaussian distribution with standard deviation 0.01. Another weights initialization method is Xavier [30]. The idea in Xavier method is to initialize the weights from a Gaussian distribution with zero mean and a variance of $\frac{2}{(n_{in}+n_{out})}$, where $n_{in}$ is the number of neurons feeding into the layer and $n_{out}$ is the number of the neurons that the result is fed to. Xavier method and its improved version allow deep networks to be trained and they converge fast.

To optimize the CNN, backpropagation algorithm is used to train the CNN which uses gradient descent to update the CNN parameters. Among the optimization methods, Stochastic Gradient Descent (SGD) is commonly used to estimate the gradient on the basis of a single randomly picked example $(x^i, y^j)$ from the training dataset:

$$\theta_{i+1} = \theta_i - \eta_i \nabla_\theta \mathcal{L}(\theta_i; x^i, y^j). \tag{11}$$

where $\theta$ is the network parameters, $x^i$ is the input and $y^j$ is the output. $\mathcal{L}$ is the objective function that is used to train the network, and $\eta$ is the learning rate. In practice, the network parameters are updated with respect to mini-batch.

To guarantee the convergence and speedup the learning process, momentum [44] is proposed to make the current gradient depends on historical batches. The classical momentum update accumulates a velocity vector in the relevant direction, and it is given by:

$$v_{i+1} = \gamma v_i - \eta_i \nabla_\theta \mathcal{L}(\theta_i; x^i, y^j);$$
$$\theta_{i+1} = \theta_i + v_{i+1}. \tag{12}$$

where $v_{i+1}$ is the current velocity vector and $\gamma$ is the momentum term which is usually set to 0.9. Nesterov momentum [45] is another way of using momentum in gradient descent optimization that moves in the direction of the previous accumulated gradient, calculates the gradient and then updates the parameters.

To update the network weights and set them to accurate values, we have to compare the predicted values with the ground-truth. To do this, we have to define a loss function that measures how close the prediction from the ground-truth is. Choosing the loss function is an important step because it is used to guide the training process of the CNN and measure the error between the predicted values and the ground-truth to correct and update the network weights. Also, it reflects the nature of the problem that the CNN has to solve (either a classification problem or a regression problem). There are many loss functions that are commonly used for the classification problems.

Multiclass Support Vector Machine (SVM) [46] is set up so that the SVM wants the correct class of the input image to have the highest score within the incorrect classes' scores by some fixed margin $\Delta$. Suppose that the input image is $x_i$ and its label is $y_i$ and the output from the network activations is $s = f(x_i; W)$. Consider the score of the $j$-th class is the $j$-th elements, $s_j = f(x_i; W)_j$, then the Multiclass SVM for the $i$-th image is given by:

$$\mathcal{L}_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta). \tag{13}$$

Another commonly used loss function for classification is Softmax loss [47]. It is a combination of multinomial logistic loss and softmax. Consider that $x_i$ is the input and $y_i$ is the ground-truth (the class, $k = 1, \ldots, K$), the output from the activations is $s = f(x_i; W)$, and the prediction given from the softmax is:

$$P(Y = y_i | X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_k}}.$$
$$\mathcal{L}_i = -\log P(Y = y_i | X = x_i).$$
$$\mathcal{L}_i = -\log \frac{e^{s_k}}{\sum_j e^{s_k}}. \tag{14}$$

In regression problems, it is important to predict real-valued quantities such as depth values in predicting depth images. $L2$ norm [48] is used commonly in regression problem, that computes the loss between the predicted values the ground-truth and then measures the $L2$ squared norm. It is defined as:

$$\mathcal{L}_i = \|f - y_i\|_2^2. \tag{15}$$

where $f$ is the output from the network activations $y_i$ is the corresponding ground-truth. Usually $L2$ is squared to make the gradient simpler and not to change the optimal parameters. $L1$ norm [48] is also used as a loss function in regression problem, and it is the sum of absolute values along each dimension. It is given by:

$$\mathcal{L}_i = \|f - y_i\|_1 = \sum |f_j - (y_i)_j|. \tag{16}$$

It is important to note that choosing the loss functions affects the training process and the performance of the CNN. For example, it is known that $L2$ loss is less robust to the outliers because the outliers can introduce huge gradients. However, Tukey's biweight loss [49] is more robust against the outliers in regression problem.

## IV. CNN-BASED IMAGES SEGMENTATION

Image segmentation is the process of assigning a label to each pixel in an image. So, it is similar to dividing the image into multiple regions so that all pixels in one region have common characteristics. At the image level, the regions have meaningful shapes and they can help in analyzing and understanding the image. Image segmentation is a computationally expensive process because it is done on the pixel level of the image. It has many practical applications in different fields, such as object detection [17], face and pedestrian detection and localization [50]. In the medical field, image segmentation is used to locate tumors [51], measure tissue volumes [52], and retinal vessels segmentation [53].

Before introducing deep learning techniques, conventional approaches have been proposed to solve image segmentation task such as thresholding, clustering methods, and histogram-based methods. Nowadays, deep learning represents a leading-edge technology that is validated by the research community, where neural network existed for decades [54]. In particular, deep learning refers to solve the imaging problem through training a neural network model. The word deep contains the idea of having a number of layers of processing image/signal units, which is similar to the structure of the neurons in the brain.

The main advantage of deep learning is their capability to automatically learn the best features according to their training dataset. This allows avoiding the time-consuming process of manual selection of the features of a particular task either it is classification or regression. In retinal vessels segmentation, segmenting tiny vessels is a very difficult task as most convention methods missed them during the segmentation process, while the automated detection of retinal blood vessels has better performance by using the deep learning methods. Carefully designed CNN can detect these tiny vessels and segment them, even it is usually difficult for an expert to analyze these blood vessels as it is reported in [55]. CNN model is proposed to overcome this issue (as shown in Figure 10). Various methods are developed based on deep

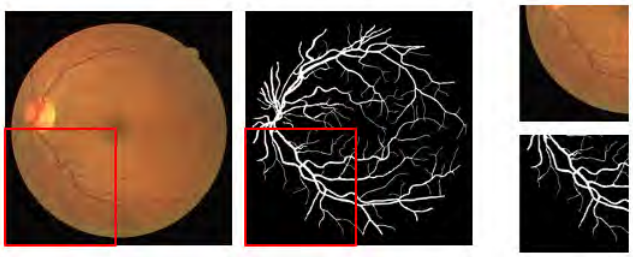


**FIGURE 10.** Issue of segmentation of retinal vessels particularly small vessels.

learning for retinal image analysis and we elaborate these methods in the next sections.

In general image segmentation, one of the leading architecture proposed for image segmentation and dense prediction was Fully Convolutional Network (FCN) [56] in 2014. From its name, one can see that this architecture is composed of convolutional layers without using fully-connected layers. This allows the segmentation image to be generated from any input image with arbitrary size. After that, different architectures have been proposed for image segmentation. U-Net [57] is an encoder-decoder architecture which was proposed to segment medical images. The encoder part extracts the features from the input image and reduces the resolution of the input image through the pooling layers, and the decoder part reconstruct the image and recovers the object details with the help of the skip connections between the encoder and the decoder layers. SegNet [58] is another architecture proposed for image segmentation. The authors introduce the transferring of max pooling indices instead of the whole values from the encoder to the decoder using skip connections to improve the segmentation resolution for faster training. In general, CNN architectures show impressive results in image segmentation, and there is still a large room for improvement that can be done to enhance the segmentation accuracy.

### A. CONVOLUTIONAL NEURAL NETWORKS FOR RETINAL VESSELS SEGMENTATION

Basically, CNNs are a special type of artificial neural networks (ANNs) where the CNNs are designed to process data with grid-like structure. They receive a stack of input images with grid-like structure and output a group of feature maps that can be used for solving different tasks. To design a CNN model for segmenting the retinal vessels, the model should be able to generate the segmented image accurately and the available databases which will be used for training should be considered. Also, the loss function that will be used for training the model should be selected carefully. In classification problems for example, the CNN models consist of a consecutive convolutional layers followed by activation functions for the non-linearity and then pooling layers to decrease the feature maps size and pool the features. At the end of the models, fully-connected layers are used as task-specific layers to generate the desired output.

In semantic segmentation and labeling, the fully-connected layers can be removed or replaced by convolutional layers (often with $1 \times 1$ kernels). Also, early-stage feature maps could help in semantic problems because they still contain the finer details of the scenes and objects in of the input image. The proposed model in [56] was perhaps the first to investigate the idea of merging feature maps with different levels of abstraction. In medical image segmentation, this trend led to the U-net [57]. With this idea, the encoder-decoder models became more popular for the medical image segmentation task. In the encoder-decoder model, the encoder network is a fully convolutional network where each convolutional layer is followed by an activation function. The decoder part is mirrored from the encoder, and it upsamples the feature maps generated from the encoder and try to segment the input image into different regions. For retinal vessels segmentation, the model should be able to classify each pixel as a vessel pixel or a background pixel. The output should be with the same resolution as the input image. The critical issue with the retinal vessels segmentation is to accurately detect and extract the tiny vessels. In the encoder-decoder model, skip connections are used to transfer the feature maps from the encoder to the decoder which help in the segmentation process and detect vessels with sharper edges.

With respect to the loss functions, it is important to select a suitable loss function that can represent a useful meaning when comparing the output and the ground truth. The loss function is used to train the model and optimize the model weights. In the literature, different loss functions were used for retinal vessels segmentation. It depends on how the ground truth is presented. Looking to the ground truth images, the images are almost black (90% of the pixels are background), and the vessels' pixels form 10% of the whole image. If we train the network without considering this, the trained model will be biased towards the background. In the literature, [59] proposed to use softmax with log-likelihood as a loss function. They classify the pixels into 4 categories; the background, the optic disc, the fovea, and the blood vessels. In [60], the output with two channels depth is generated from a softmax, where each pixel has two probability values that the maximum value indicates the class of the pixel (either a vessel or a background). Dice Loss is used as a loss function to optimize the model. Reference [53] uses the class-balancing cross entropy loss function [61] to train the proposed model and generate the segmented image. This loss function takes care of the heavily biased class distributions and balance the training process.

## V. MEASURING PARAMETERS AND RETINAL IMAGES DATABASES

### A. MEASURING PARAMETERS

The performance of the existing methods for retinal vessels segmentation was compared with the ground truth image of the corresponding image. To measure the performance of a method in segmenting retinal vessels, three metrics

are calculated: accuracy, sensitivity and specificity. These parameters are explaining as follows,

Accuracy is defined as the ratio of sum of correctly identified vessels and non-vessels to the sum of total number of pixels, is calculated as

$$\text{Accuracy}(AC) = (tp + tn)/(tp + fp + fn + tn). \quad (17)$$

Sensitivity, defined as the ratio of correctly identified vessels to the total number of vessels, is computed as

$$\text{Sensitivity}(Se) = tp/(tp + fn). \quad (18)$$

Specificity, defined as the ratio of correctly detected non-vessels to the total number of non-vessels, is measured as

$$\text{Specificity}(Sp) = tn/(tn + fp). \quad (19)$$

where *tp*,*tn*,*fp* and *fn* represent the identification of the vessels and non-vessels pixels, respectively abbreviating as true positive, true negative, false positive and false negative. Accuracy is the dominant parameter that gives the information of overall classification performance of vessels pixels.

### B. RETINAL IMAGES DATABASES

There are several publicly available databases for analyzing retinal images. The two most commonly used databases are the Digital Retinal Images for Vessel Extraction (DRIVE) [62] and Structured Analysis of the Retina (STARE) [63] databases.

Digital Retinal Images for Vessel Extraction (DRIVE) database is used in existing methods, and it contains forty images. These images were taken during an eye screening program in Netherland. DRIVE image has a resolution of $768 \times 584$ pixels. Forty images are divided into two groups, and each group contains 20 images. One group of 20 images is named as test images, and other 20 images group is named as training images. Each image contains with its own mask images that delineate the field of view. For training images group, a single manual segmentation of the blood vessels is available as the gold standard image. For the test, group images contain two manual segmentation images or ground truth images.

Structured Analysis of the Retina (STARE) database is also used for assessing existing methods, and it contains twenty images. Ten images of STARE database among twenty images contain pathologies, and thus these images provided a good chance to know the capability of segmentation methods on different DR stages. The STARE image database contains a resolution of $605 \times 700$. The STARE database contains two manually segmented images generated by two expert retinal vessel observers.

## VI. THE STATE-OF-ART

The researchers are implementing methods for analyzing the retinal fundus images based on deep learning algorithms. A wide variety of applications are addressing such as segmentation of retinal blood vessels and analysis of abnormalities in the retinal fundus images. There are around 12 deep learning publications focusing on segmentation, and other seven publications addressing the issues of abnormalities analysis. In this work, we analyze the performance of these methods to identify better indication of disease progression for early treatment. The retinal blood vessels segmentation leads to a robust and accurate technique for detection of retinal blood vessels. The analysis of each existing method of retinal blood vessels segmentation along with their advantages and limitations are explained in details and listed in Table 1.

### A. ANALYSIS METHODOLOGY OF EXISTING DEEP LEARNING METHODS FOR RETINAL BLOOD VESSEL EXTRACTION

Zhang *et al.* [64] was one of the first researchers who worked on detection of retinal blood vessels based on neural network techniques. They used some self organizing map (SOM) as pre-processing techniques to train the network to get retinal blood vessels, and their proposed unsupervised method gave a good retinal vessels segmentation. Their method consists of three steps. First, a multidimensional feature vector from the green channel of the RGB retinal image intensity is constructed, and then vessels enhanced intensities feature is achieved by using morphological operation. Next, a self-organizing map (SOM) with the concept of pixel clustering is used as a classifier, which is a type of unsupervised neural network. Finally, each neuron of the output layer of SOM is classified as retinal vessel neuron or retinal non-vessel neuron by applying Otsu's threshold method to achieve final vessel segmentation.

Zillya *et al.* [65] proposed a method to identify of retinal blood vessels by using ensemble learning late fusion on convolutional neural network (CNN). Their method is based on the following three steps,

1) An entropy sampling method is used to select the informative points to reduce the computational complexity for performing the uniform sampling.
2) A CNN model based on convolutional filters was used.
3) A softmax logistic classifier is used to fuse the output of all learned filters, and to test the trained model on DRIVE database.

Their observation was based on the graph cut algorithm because the output of the classifier is subjected to an unsupervised graph followed by a convex hull transformation to achieve the final segmentation of retinal vessels.

Maji *et al.* [66] proposed a method by using ConvNet-ensemble based CNN architecture for processing the fundus color image in order to get retinal blood vessel. They used ensemble learning based on multiple models which seek to promote diversity among models with a combination, and it helps to reduce the problem of overfitting of training data. The main contribution of their method was an integration of ensemble learning with ConvNet to improve the generalization and this approach was a heuristics independent. In general, it provides a useful solution to solve complex medical data like retinal images.

**TABLE 1. Overview of papers using deep learning techniques for retinal image segmentation. All works use CNNs.**

| Method | Year | Key Technique |
|---|---|---|
| Zhang et al [64] | 2015 | Morphological operation,Self Organizing Map (SOM). |
| Zilly et al [65] | 2016 | Ensemble learning based Convolutional Neural Network (CNN) architectures. |
| Maji et al [66] | 2016 | Ensemble learning based Convolutional Neural Network (CNN) architectures. |
| Tan et al [59] | 2016 | Normalisation technique, Convolutional Neural Network (CNN). |
| Liskowski et al [55] | 2016 | Global contrast Normalisation, Deep Neural Network (DNN). |
| Fu et al [67] | 2016 | Convolutional Neural Network (CNN). |
| Wu et al [68] | 2016 | Convolutional Neural Network (CNN),PCA. |
| Yao et al [69] | 2016 | Convolutional Neural Network (CNN). |
| Fu et al [70] | 2016 | Multiscale and multilevel CNN. |
| Mahapatra et al [71] | 2016 | Convolutional Neural Network (CNN). |
| Maninis et al [72] | 2016 | Convolutional Neural Network (CNN). |
| Soomro et al [53] | 2017 | Image enhancement, Convolutional Neural Network (CNN). |
| Li et al [73] | 2018 | Multi-scale convolutional neural network. |
| Guo et al [74] | 2018 | Convolutional Neural Network (CNN). |
| Chudzik et al [75] | 2018 | Convolutional Neural Network (CNN). |
| Hajabdollahi et al [76] | 2018 | Convolutional Neural Network (CNN). |
| Yan et al [77] | 2018 | Convolutional Neural Network (CNN). |
| Soomro et al [60] | 2018 | Image enhancement, Convolutional Neural Network (CNN). |

Tan *et al.* [59] proposed a seven-layer convolutional neural network (CNN) method. Their method not only detects the retinal blood vessels but also the optic disc and fovea of the retinal image. Due to detection of more than one feature at the same time, the accuracy of segmenting the retinal blood vessels was negatively affected. They used the normalization technique before segmentation to remove non-uniform background and noise in retinal images in such way it makes consistency in background lighting and contrast. The contribution of their work contains the selection of every effective vessels pixel in fundus image of three color channels before feeding into following CNN. Output layer contains four neurons which represents background, optic disc, fovea and blood vessels. Their method did not show satisfactory performance but it is the first method can do multi-task at the same time with a single CNN.

Liskowski and Krawiec [55] proposed an unsupervised deep neural network (DNN) for analyzing retinal blood vessels. They used pre-processing steps such as global contrast normalization, zero-phase whitening, and augmentation using geometric transformations and gamma corrections for enhancing the intensities of blood vessels against their background. They used various classifiers such as Bayesian classifier and a nearest-neighbor classifier to classify the retinal image features. The features were learned with different ways to label vessels and non-vessels pixels for achieving vessels segmented image. Their method was performed much better than other existing methods. They suggested that learned features can be more efficient if using the other strong classifier in terms of detecting more vessel pixels correctly. However, retinal images suffer from background noise, uneven illumination and varying low contrast. The varying low contrast of retinal fundus images introduces difficulty to detect the vessels. Image enhancement technique can be included rather than improving variety of the features. Their approach did not show a good performance in case of tiny vessels detection.

Fu *et al.* [67] developed a novel method for detecting retinal blood vessels based on probability map using CNN. They formulated the vessel extraction with consideration of boundary detection problem. The CNN is used to generate retinal vessels probability map. The vessel probability map differentiates vessels and background pixels in low contrast region. Their method performs better in the pathological images due to robustness of differentiating vessels and background pixels in low contrast region. Afterwards, they utilized fully-connected Conditional Random Fields (CRFs) with combination of the discriminative vessel probability

map and long-range interactions between pixels to achieve 95% accuracy of segmentation.

Wu *et al.* [68] developed a method for retinal blood vessels extraction by using CNN. They used PCA based nearest neighbor to local vessel pixels structure distribution. Afterwards, a generalized probabilistic tracking framework was used to segment retinal vessels. The main step of their method is to perform CNN for all image pixels either vessels pixels or background pixels. It loses the efficiency in final results. The combination of vessels tracking based on CNN and their probability distribution of vessels intensities improves the efficiency of their method, and provides more information of pixels of retinal blood vessels network. Their proposed method improves the efficiency of full image CNN results but loses minimal unconnected vessels pixels. This method can be improved if fewer pixels are segmented in the tracking approach as compared to the full CNN method.

Yao *et al.* [69] proposed a method based on the CNN and post-processing steps to get the retinal vessels segmentation. The CNN checks each pixel with its neighbors of the retinal fundus image for making better contrast of retinal vessels against their background. Segmentation of retinal fundus image is achieved by using binarization as post-processing. The binarization contains two steps: First, they used local multi-scale and global normalization imaging technique to achieve the initial binarization results. They named this step as local binarization. Secondly, they used the morphological operation to improve segmented vessels image. The multiscale binarization is a good idea to get a well segment image but their proposed CNN model needs to improve to learn better-enhanced feature image. Their method can be improved if they use multiple scales with multiple CNN and ensemble each output. On the other hand, they should consider the morphological characteristics of retinal blood vessels as prior information during their post-processing step.

In [70], a deep CNN model combined with conditional random field (CRF) was proposed for retinal vessels segmentation. It consists of the following components,

1) They consider the retinal blood vessel segmentation issue as a boundary detection task.
2) The CNN is designed to learn a multi-scale discriminative representation with side output layers. It acts as a classifier that produces a companion local output for early layers [61].
3) Conditional Random Field (CRF) is used to model for long-range interaction between the pixels.
4) CNN and CRF layers are combined into an integrated deep network and named as Deep Vessel output.

Mahapatra *et al.* [71] implemented a method for retinal vessels detection which is based on the local saliency map and CNN model to get the vessels image. They used the unsupervised information from local saliency maps and the supervised information from the trained convolutional neural network. The final result is achieved by a combination of both saliency maps from unsupervised information and they train CNN from supervised information. The novelty of their method is the calculation of saliency values for every image pixel at multiple scales to provide the global and local image information. Such vessel pixel information provides an additional information for CNN. The combination of two-information improves the performance of the method as compared to other methods. The main purpose of their method is an assessment of image quality in order to analyze the retinal vessels, and the low computation of their method makes it possible for quick assessment of image quality and recommends the patient for timely treatment.

Maninis *et al.* [72] proposed a framework for detection of retinal blood vessels and optic disc segmentation. The method is based on deep Convolutional Neural Networks (CNN), and they named their approach as Deep Retinal Image Understanding (DRIU). Their CNN architecture is based on the VGG network which is mostly used for image classification. In their method, the fully connected layers at the end of the network are removed. Four max-pooling layers are used between the convolutional layers, which separate the architecture into five stages. Between the pooling layers, feature maps are created by convolutions with different filters of the same size. The deeper network gives more information of features. They achieved two different feature maps for segmentation of retinal vessels and optic disc because they observed that the coarse feature maps at the last stage of CNN did not help with vessel detection, and thin vessels are not detected properly due to the fact that they down scaled the image layers by layers. The final vessel image of their method is based on probability map in which pixels detected as vessel or optic disc, and the method gives a good performance. However, their method does not show better results with images of optic disc shadow.

Soomro *et al.* [53] proposed a method based on convolution neural network along with pre- and post-processing steps. They observed the three main issues impacting the final segmentation without preprocessing images. These issues are uneven illumination, noise, and low varying contrast. They select grey well-contrast image by using RGB into grey conversion. Many researchers just process green band image of retinal fundus image for vessels segmentation,else red and blue band also well-contrast. They use RGB to grey conversion to give better contrast images than green band images. First, they include pre-processing like morphological and PCA techniques. In the second step, the authors proposed a fully Convolutional Neural Network (CNN) and train it to get fine vessels features. But when there are noise pixels, tiny vessels are not detected properly. To address this, they add a post-processing step for getting well binarization images that removes noisy pixels. The output of their method is much better as compared to all other existing methods.

Li *et al.* [73] proposed the method of analyzing fundus images by segmentation of retinal vessels, based on supervised vessel segmentation by the deep learning method. They found the problem of the imbalance of the retinal vessels which limits the improvement of the accuracy of

the segmentation. They used the multi-scale CNN-based deep learning model with a modified loss function, and they called it the focal loss function. The improved loss function for deep learning to better handle the imbalance problem of vessel pixel segmentation. They achieved better results due to the use of the CNN multi-scale structure and the label processing method.

Guo *et al.* [74] had proposed a method of vessel segmentation as well as pre-processing techniques to improve low contrast. They proposed CNN multi-scale and multi-level with deeply supervised CNN and short connection for the extraction of retinal vessels. They used short connections to transfer the semantic information of the vessels between the side-output layers of the network. There are two types of short connections: forward short connections and backward short connections. A forward short connection could transmit low-level semantic information at a high level, and backward short connection could transmit retinal vessel network information at a low level, resulting in a well-segmented image of the vessels. For further validation, they use a structural similarity measure to evaluate the segmented image of the vessels.

Chudzik *et al.* [75] proposed the method based on two-stage vessel segmentation from retinal fundus images. In the first stage, they used CNN to correlate an image patch with a corresponding groundtruth; they used totally random trees embedding for this task. In a second step, training patches are given to CNN for the creation of the visual codebook. This codebook was used to form the nearest neighbour vectors created by the propagation of CNN's invisible patches. The main contribution of their method to the generation of segmentation patches that were not detected during the training process. The performance of the proposed framework demonstrated the validation of their method for segmentation of retinal blood vessels.

Hajabdollahi *et al.* [76] proposed the CNN method based on a combination of CNN quantization and pruning for retinal vessel segmentation. They used the quantized technique on fully connected layers and a pruning tactic on the convolutional layers to form a very efficient and sample network structure. Their method worked well on the most used database and gave better results with reduced complexity.

Yan *et al.* [77] proposed both imbalances the detection of thin and thick retinal vessels. Because thick and thin vessels are important for proper detection. Numerous deep learning methods as well as image-based retinal vein detection method attempt to simultaneously segment thin and thick vessels using pixel-level unit loss that processes all pixels (vessels. or non-vessel pixels) with the same consideration. But there is a high imbalance ratio between thick and thin vessels, and most thick vessels dominate thin vessels. This is one of the reasons that most segmented retinal methods have not used thin vessels, which leads to poor performance. They proposed the method for separately segmenting thick and thin vessels, and their method is based on three-steps in-depth learning tactics. The first step was based on the detection of thick vessels and the second step on the detection of thin vessels, and the last step was based on the stage of fusion to refine the results by recognizing other thin and thick vessels and to give the image of the final well-segmented vessel.

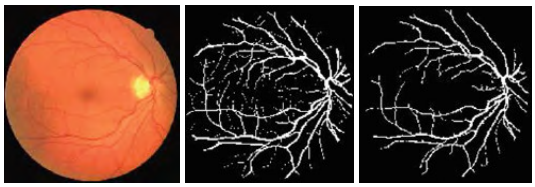



**FIGURE 11.** Results of Zhang et al [64]. (a) Retinal color image. (b) Ground truth image. (c) Output image.

Soomro *et al.* [60] proposed the method based on the CNN model with the dice loss function. The first time dice loss function is used for segmentation of retinal vessels. The method involved two steps. The first step was based on the pre-processing steps to eliminate uneven illumination. The main purpose of preprocessing steps is to make the training process more efficient. The second step was based on the CNN model that contained a variational auto-encoder that is a modified version of U-Net. The main contribution of their method to the implementation of the CNN model is to replace all layers of pooling by progressive convolution and deeper layers. It maintained the resolution and gave exactly the same resolution segmented image as the original image resolution. The proposed method gave better performances and made it possible to overcome the problem of imbalance related to the segmentation of the retinal vessels.

### B. ANALYSIS PERFORMANCE OF EXISTING DEEP LEARNING METHODS FOR RETINAL BLOOD VESSEL EXTRACTION

To give a clear picture of performance of each single method, we summarize their results in Table 2 that can show capability of detection of retinal blood vessels. We elaborate the strength and weakness of each method.

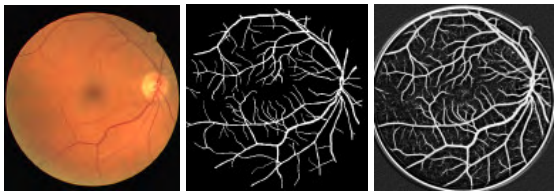
Zhang *et al.* [64] validated their method only on DRIVE database, and their proposed method achieved 0.940.. As shown in Figure 11, the tiny vessels are not detected properly. No validation was performed on the pathological images.

Maji *et al.* [66] validated their method on DRIVE database, and their method shows good performance in terms of vessels detection. As shown in Figure 12, tiny vessels are dropped. Also, there is no validation conducted on pathological images. Proper vessels length detection is not observed due to the unresolved issue of circle boundary around retinal vessels, and proper vessels were not analysed due to background noise pixels.

Liskowski's [55] method achieved the detection accuracy of retinal vessels (shown in Figure 13) of 0.949 on STARE and DRIVE database. Again, the tiny vessels are not observed properly, and there is non-uniformity between blood vessels

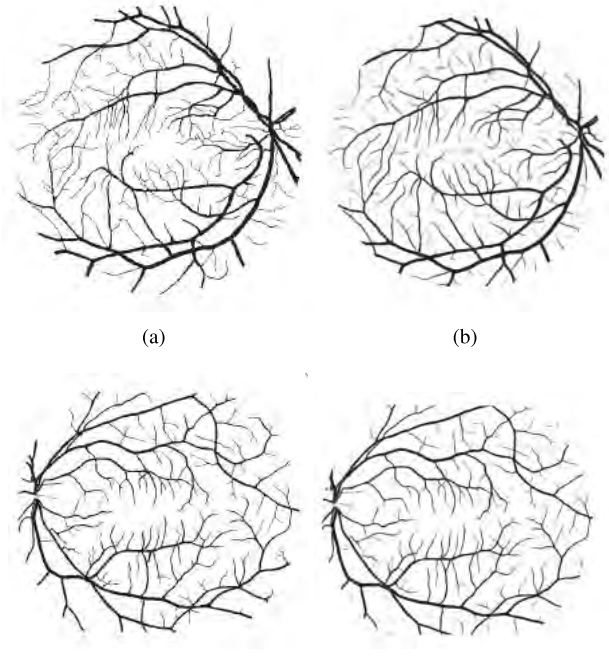
**TABLE 2.** Performance analysis of segmentation model.

| | | DRIVE | | | | *STARE* | | | |
|---|---|---|---|---|---|---|---|---|---|
| Methods | Year | *Se* | *Sp* | *AC* | *AUC* | *Se* | *Sp* | *AC* | *AUC* |
| Zhang et al [64] | 2015 | - | - | 0.940 | - | - | - | - | - |
| Maji et al [66] | 2016 | - | - | 0.947 | - | - | - | - | - |
| Liskowski et al [55] | 2016 | - | - | 0.949 | 0.973 | - | - | 0.949 | 0.982 |
| Fu et al [67] | 2016 | 0.760 | - | 0.952 | - | 0.741 | - | 0.958 | - |
| Wu et al [68] | 2016 | - | - | - | 0.97 | - | - | - | - |
| Yao et al [69] | 2016 | 0.773 | 0.960 | 0.936 | - | - | - | - | - |
| Maninis et al [72] | 2016 | - | - | - | 0.822 | - | - | - | 0.831 |
| Fu et al [70] | 2016 | 0.729 | - | 0.947 | - | 0.714 | - | 0.954 | - |
| Tan et al [59] | 2017 | 0.753 | 0.969 | 0.926 | - | - | - | - | - |
| Soomro et al [53] | 2017 | 0.746 | 0.917 | 0.948 | 0.831 | 0.748 | 0.922 | 0.947 | 0.835 |
| Guo et al [74] | 2018 | 0.789 | 0.978 | 0.954 | 0.979 | - | - | - | -. |
| Chudzik et al [75] | 2018 | 0.788 | 0.974 | - | 0.964 | 0.826 | 0.980 | - | 0.983. |
| Hajabdollahi et al [76] | 2018 | - | - | - | - | 0.782 | 0.977 | 0.961 | 0.97. |
| Yan et al [77] | 2018 | 0.763 | 0.982 | 0.954 | 0.975 | 0.774 | 0.986 | 0.964 | 0.983. |
| Soomro et al [60] | 2018 | 0.739 | 0.956 | 0.948 | 0.844 | 0.748 | 0.962 | 0.947 | 0.855. |



(a) (b) (c)

**FIGURE 12.** Results of Maji et al [66]. (a) Retinal color image. (b) Ground truth image. (c) Output image.



(a) (b) (c) (d)

**FIGURE 13.** Results of Paweł Liskowski's method [55]. First column ((a) and (c)) Ground truth image. Second column ((b) and (d))Output image.

and background. Normalization technique could be applied to improve the proposed method.

Fu *et al.* [67] improved their method based on CNN for retinal segmentation of blood vessels. As shown in Figure 14, the tiny vessels are not detected, and the performance of their method is not validated on the challenging images (Images contains abnormalities or pathologies). Their method gives much better sensitivity of 0.760 on DRIVE database and 0.741 on STARE database along with an average accuracy of 0.95.

Aaron Wu's method [68] was validated on DRIVE database only. Even though their method achieved 0.97 area under curve (AUC), there is no disclosure of the parameters in their experimental design. The output from their method is shown in Figure 15, from which we can observe that the tiny vessels are completely dropped. Their method was not validated on the pathological images. An image enhancement technique can be applied on the images before training the CNN to improve the vessels detections results.

Zhenjie Yao's method [69] gave much better sensitivity 0.773 on DRIVE database as compared to above methods but their method gave low accuracy of 0.936. No validation was performed on STARE dataset nor on the pathological images. The output of their method is shown in Figure 16. As we observed, the tiny vessels are not detected even some large vessels were missing. Their method can be improved if image enhancement technique used as a pre-processing step.

Maninis *et al.* [72] proposed a CNN model for retinal vessels segmentation and optic disc segmentation. The proposed

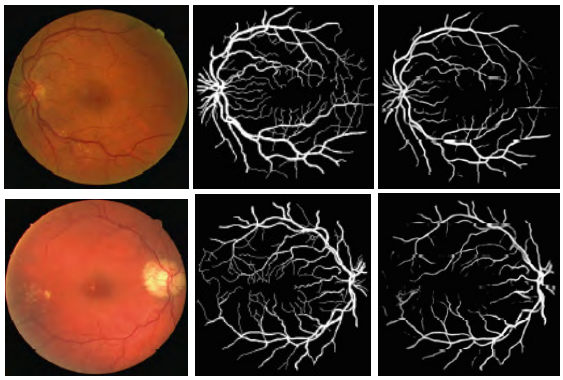
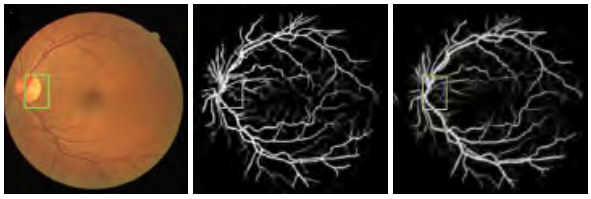
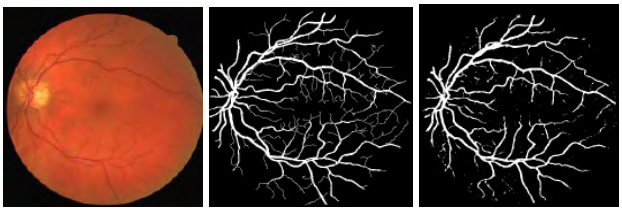


**FIGURE 14.** Results of Fu et al [67] From left to right: Retinal images, the corresponding ground truth and the output images.



(a) (b) (c)

**FIGURE 15.** Results of Aaron Wu's method [68]. (a) Retinal Images. (b) Ground images. (c) Aaron Wu's method [68] output Images.



(a) (b) (c)

**FIGURE 16.** Results of Zhenjie Yao's method [69]. (a) Retinal images. (b) Ground images. (c) output images.

method was validated on DRIVE and STARE databases but they reported only AUC of 0.822 on DRIVE database and 0.831 on STARE database. Figure 17 shows their method has the capability to segment the fine vessels but no validation was reported on pathological images. Their method can be one good tool for analysis retinal blood vessels after validation on challenging images.

Huazhu Fu's [70] method was based on CNN for detection of tiny vessels. Their methods gave a good accuracy around average 95% on DRIVE and STARE database, but sensitivity is lower around 72% as compared to another methods. From the results of Fu's method shown in Figure 18 we can observe that tiny vessels are not detected properly. The performance of the method on the pathological images was not addressed. Their method can be improved by using the post-processing steps for removing noisy pixels to achieve fine vessels image.

Tan *et al.* [59] validated their method on DRIVE database. The performance of their method was comparatively better

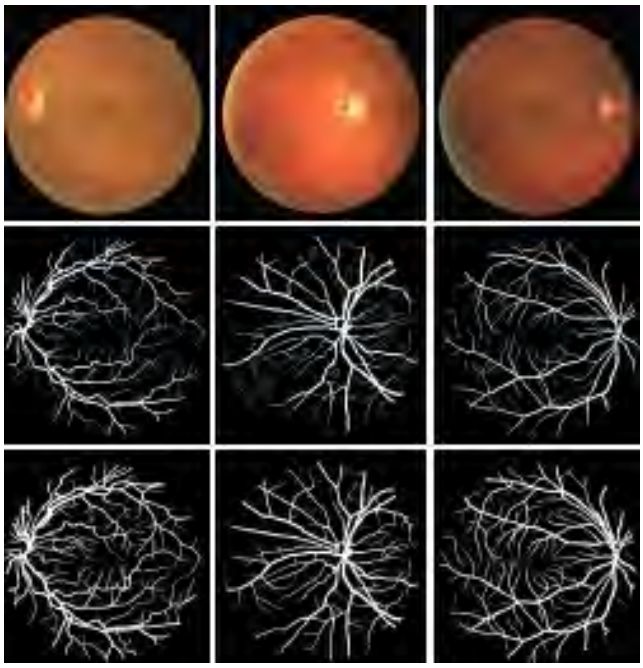


**FIGURE 17.** Result of Kokitsi Maninis's method [72]. TFrom top to botton: row input images, ground truth images and segmented images.

**FIGURE 18.** Results of Huazhu Fu's method [70]. From left to right: retinal images, ground truth and output images.

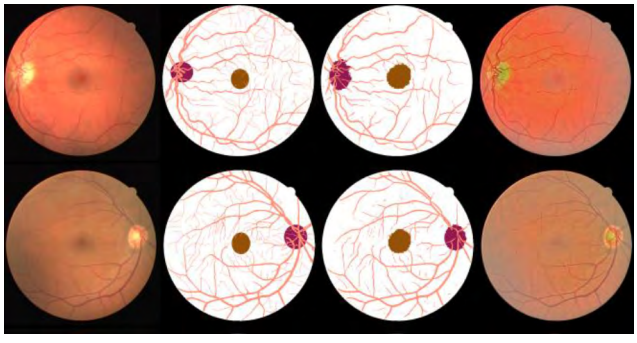


**FIGURE 19.** Result of Tan et al [59]. From left to right: input images, ground truth, segmentation images and the normalized images.

than the above method [72]. They also measured and reported sensitivity, specificity and accuracy. Their method gave a good sensitivity of 0.75 but the accuracy drops. The reason of

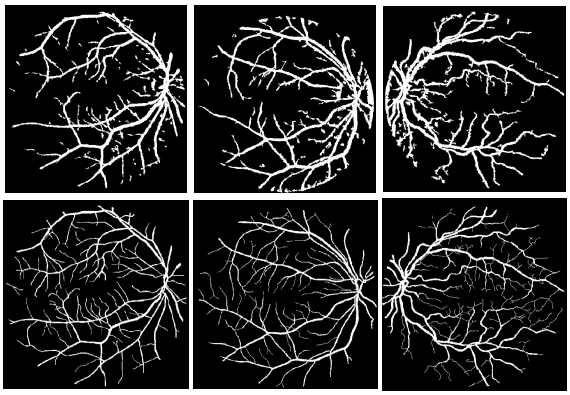


**FIGURE 20.** Comparison results of Soomro et al [53] on challenging images of DRIVE and STARE databases. From top to bottom: the segmented images and their corresponding ground truth.

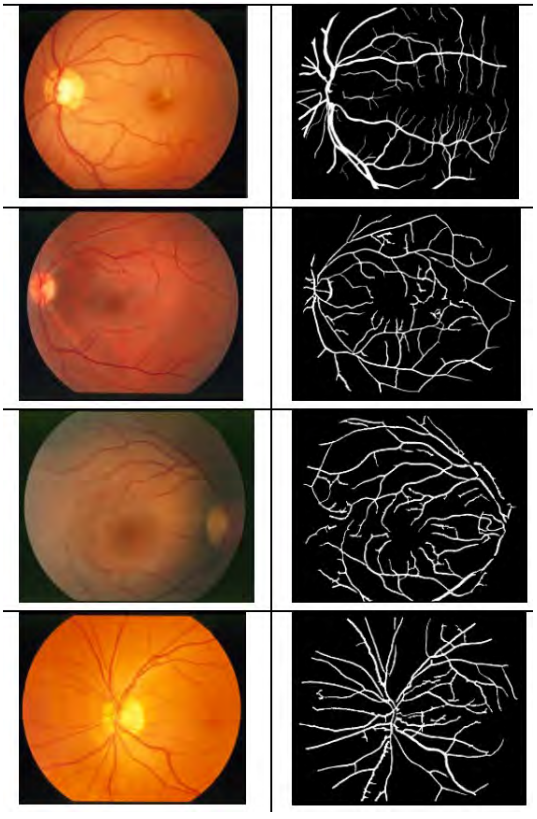


**FIGURE 21.** Result of Li et al. [73]. Input images (left) and the segmented images (right).

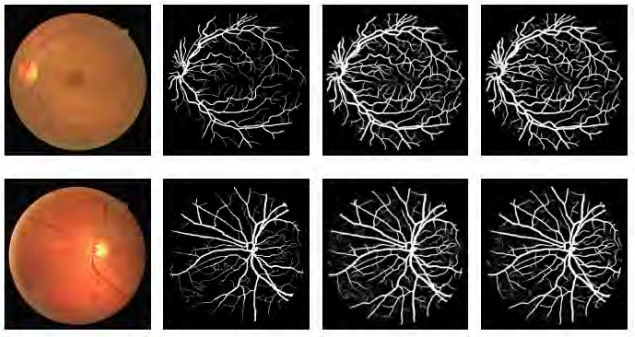


**FIGURE 22.** Result of Guo *et al.* [74].From left to right: the input images, the ground truth, vessel segmentation result of image-level input S-DSN and vessel segmentation result of patch-level input S-DSN.

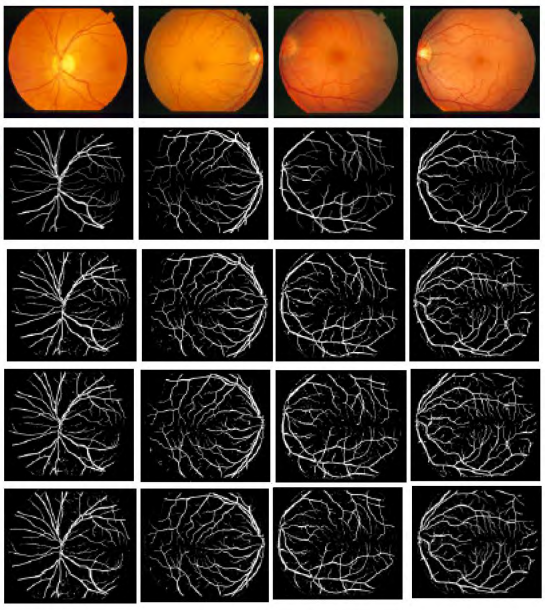


**FIGURE 23.** Result of Hajabdollahi and al [76]. From top to bottom: the input images, the ground truth, CNN with original parameters, CNN with FCLs quantization and CNN with FCLs quantization and FCLs pruning.

good sensitivity was the use of local contrast normalization technique but their method still suffers from losing the tiny vessels as shown in Figure 19. They did not validate their method on pathologies images.

Soomro *et al.* [53] proposed method based CNN along with pre-and-post processing. The performance on both DRIVE and STARE is sensible because they mostly focused on the detection of tiny vessels. They achieved the sensitivity of 75% on both DRIVE and STARE database along with average around 95% accuracy on both databases. Their method can be improved further for detecting tiny vessels as some of the vessels are dropped by using improved CNN model. They reported all measurement parameters that helps to analyze retinal images properly, especially as they validated their method on challenging images as shown in Figure 20.

Li *et al.* [73] used a database containing retinal fundus images of 5620 patients to validate their method for segmentation of the retinal blood vessels. The proposed method achieved the optimum accuracy of 0.949, but tiny vessels are missing, as shown in Figure 21, just as they have not compared their method to other existing methods.

Guo *et al.* [74] validated their proposed method on the DRIVE database and obtained the best performances compared to the other existing methods. Their method was able to segment the tiny vessels as shown in the Figure 22 and overall gave the best performance with a specificity of 0.9802, a sensitivity of 0.789, an accuracy of 0.956 and an AUC of 0.9802.

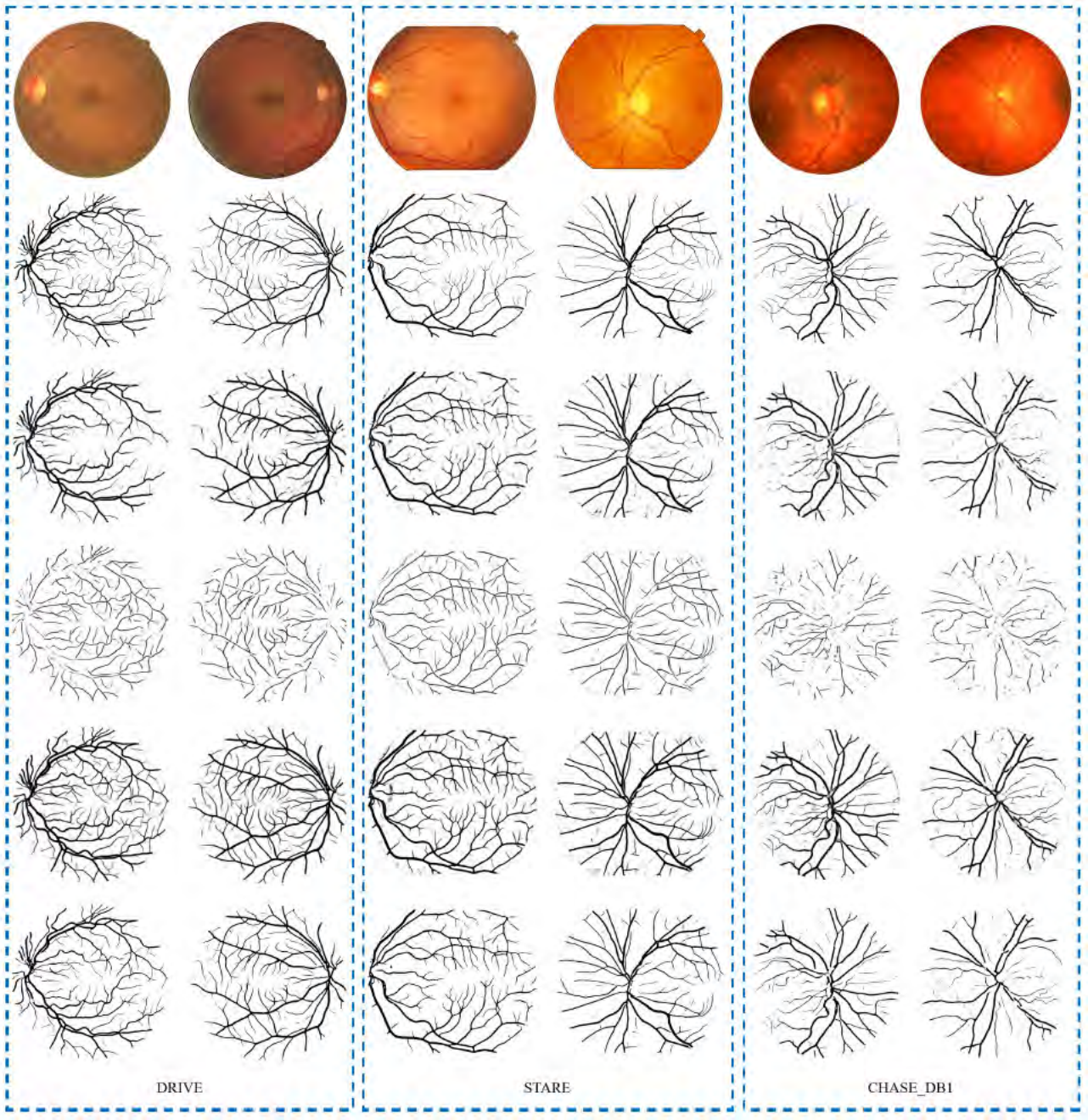



**FIGURE 24.** Result of Yan et al [77]. From top to bottom: the input images, manual annotations, probability maps of thick vessels, probability maps of thin vessels, final probability cards and corresponding hard segmentation images of Yan's method.

Chudzik *et al.* [75] validated their method on the DRIVE and STARE databases, which performed well with a sensitivity of 0.788 and an AUC of 0.964 on the DRIVE database, and a sensitivity of 0.826 and an AUC of 0.983 on the STARE database. However, the accuracy was not reported and no visualized segmented images were shown. It is very difficult to analyze the tiny vessels and predict the robustness of the method without observing the segmented images and analyze the accuracy.

Hajabdollahi *et al.* [76] validated their method on one of the most challenging databases called STARE database and obtained good performances with an accuracy of 0.961 and AUC of 0.97. The proposed method allows to segment small vessels as shown in the Figure 23. This method can be considered as one of the source codes for the detection of portable retinal vessels.

Yan *et al.* [77] have validated the results of their method on three databases DRIVE, STARE and CHASE-DB1, and

**TABLE 3.** Performance analysis of segmentation model.

| | Main three issues | | |
|---|---|---|---|
| Algorithms | Thin Vessel Detection | Workable on Pathological Images | Noise Issue |
| Zhang et al [64] | × | × | ✓ |
| Zilly et al [65] | × | × | ✓ |
| Maji et al [66] | × | × | ✓ |
| Tan et al [59] | × | × | × |
| Liskowskie et al [55] | × | ✓ | ✓ |
| Fu et al [67] | × | × | ✓ |
| Wu et al [68] | × | × | ✓ |
| Yao et al [69] | × | × | ✓ |
| Fu et al [70] | × | ✓ | ✓ |
| Mahapatra et al [71] | × | × | ✓ |
| Maninis et al [72] | × | × | ✓ |
| Toufique et al [53] | ✓ | ✓ | ✓ |
| Li et al [73] | × | × | ✓ |
| Guo et al [74] | ✓ | ✓ | ✓ |
| Hajabdollahi et al [76] | × | ✓ | ✓ |
| Yan et al [77] | ✓ | ✓ | ✓ |
| Toufique et al [60] | × | ✓ | ✓ |

their method can detect tiny small vessels and also works in the presence of pathologies, as shown in Figure 24. The performance of their method on DRIVE and STARE is shown in Table 2 against other existing methods. The performance of their method on the CHASE-DB1 database was a sensitivity of 0.7641, a specificity of 0.9806, a precision of 0.9607 and an AUC of 0.9776.

Soomro *et al.* [60] also validated its method on the DRIVE and STARE databases, and their method gave a performance comparable to that of other existing methods. But it requires a lot of improvements because their method does not detect the thin and appropriate vessels and gives a shadow to the optic disc, as shown in Figure 25, and because of the optic disc shadow, the overall performance of their method is affected.

From the analysis of the above methods, we compare the capability of each method in terms of detection of accurate blood vessels as well as the performance of detection of tiny vessels. As per our analysis, each method makes a good contribution for detecting retinal vessels only over certain aspects. We would like to recommend that a good methodology needs three steps to detect proper vessels. First, a preprocessing step is proposed to handle the issue of varying low contrast, such as image contrast and normalization technique. Second, a careful design of CNN should be considered because of the nature of training images. The third step is optional. If there are noisy pixels then post-processing is required otherwise this step can be omitted. We recommend these as on basis of analysis of methodology and

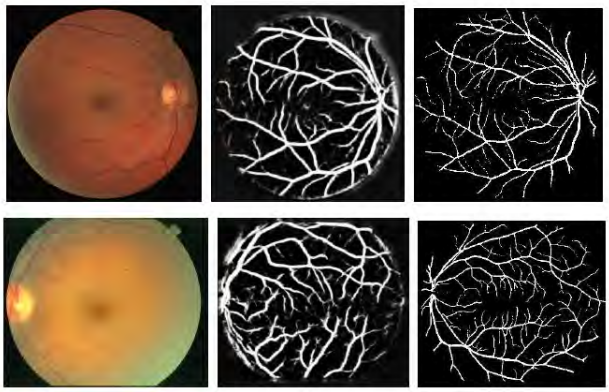


**FIGURE 25.** Toufique et al [60]. From left to right: the input images, the output images, the corresponding ground truth images. Top: DRIVE image. Bottom: STARE image.

performance of methods. The methods such Fu *et al.* [70], Liskowski*et al.* [55], Soomro *et al.* [53], Guo*et al.* [74] and Yan*et al.* [77] achieve good performance based on CNN models along with pre-and-post processing steps. In next section, we highlight the main issues and provide more comparison of existing methods for future work contribution in this area.

## VII. MAIN ISSUES IN RETINAL VESSELS METHODS

After analysing limitation and capabilities of each retinal vessels segmentation method based on deep learning, we identify

that three main issues should be considered to implement a robust retinal blood vessels segmentation process. These issues are

1) In order to improve the sensitivity, the retinal blood vessels segmentation method based on deep learning technique should be capable to detect the tiny vessels. It means that the sensitivity of method should be improved.
2) The retinal blood vessel segmentation method should be capable to work on challenging images such as images contains abnormalities, central of light reflex and uneven illumination along with noise. The best way to analyze the performance of the method is to validate it on STARE database because it contains almost 50% challenging images.
3) Most segmented vessels images contain noise pixels in the background. Therefore, a robust design and trained CNN model or novel post-processing is required to handle the noisy pixels. The removal of the noisy pixels will give better visualization of retinal blood vessels, and it will be easier for the ophthalmologist to diagnose a disease and recommend proper treatment.

We compared the previous methods according to the capability of solving these three core issues. Table 3 contains all CNN based retinal blood vessels segmentation methods. It represents the capabilities of these methods regarding detection of tiny vessels, working on pathological images and resolving the issue of noise. We rummage-sale check marks (✓) and X-mark (×) symbol. The ✓shows that problem was successfully considered. The X-mark (×) symbol indicates that problem was not considered. We clearly observed that all methods solved the issue of noise, but they did not handle main issues of observation of tiny vessels and capability of working on pathologies. No method properly handles the issue of tiny vessels detection except some tiny vessels are detected by Toufique's method. It is this reason why we recommend proposing a novel contrast enhancement technique to enhance vessel network in order to detect tiny vessels properly in later stages. The methods such Fu *et al.* [70], Liskowskie *et al.* [55],Toufique *et al.* [53], Guo *et al* [74] and Yan*et al* [77] proposed models for segmentation of retinal vessels based on normalization techniques as pre-processing. This enable their methods to work on pathological images. The main reason of detection of tiny vessels in Toufique's method [53] is the use of post-processing step to remove the noise pixels from the background. However, there are a few tiny vessels that are still not detected.

## VIII. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

A large diversity of deep learning techniques has been tested on the retinal fundus images. The deep learning methods give improved results for detection of retinal blood vessels. Lack of observing the progress of diabetes leads to develop a various abnormality in the retina damage retinal vessels, and finally lose the vision. Therefore, timely treatment is required to save the vision. Timely detection of eye diseases especially DR is of significant importance to protect vision. This can be done by an accurate detection of retinal blood vessels of DR effect image.

There are many proposed algorithms for segmentation of retinal blood vessels in past 34 years, but the retinal blood vessels segmentation methods based on deep learning are only proposed after 2015. The main purpose of detection of retinal blood vessels is analyzing of the intensity of retinal vessels, removing noise pixels, varying low contrast issues in order to detect proper vessels network. The good training of the model can lead to achieve accurate detection of vessels through using Convolutional Neural Network (CNN).

This research work provides a complete survey on the existing retinal fundus images using deep learning techniques. We have analyzed all existing segmentation based deep learning tactics for diagnostic of DR or eye disease. The main contribution of this review paper is to identify the strengths and weaknesses of difference methods, so that one can further develop robust algorithm for the screening of eye diseases based on deep learning method. The early stage detection for eye diseases will provide directions for quick treatment thereby protecting the vision of patient.

After the study of retinal blood vessels segmentation methods, three steps in using color retinal fundus images have been identified. We conclude that a robust computerized system of DR screening based on deep learning method has the capability to give accurate results for detection of retinal image features such as pathologies and retinal blood vessels. It clearly shows that the performance of the existing methods drops due to low contrast of tiny vessels which are not extracted easily. There exist a few other issues in retinal images such as non-uniform background (background pixels are much higher than blood vessels pixels), noise, varying and low contrast in different regions of the image, central light reflex, presence of abnormalities, etc. These issues make it tough to detect tiny vessels, which still remains challenge in retinal image processing. Our proposed methodology has potential to improve the results. Overall, this computerized system for eye disease analysis inevitably reduces the workload of the experts in the processing of the retinal images, rapidly analyzes the disease progression and identifies candidates for early treatment.

## ACKNOWLEDGMENT

*(Toufique A. Soomro and Ahmed J. Afifi contributed equally to this work.)*

## REFERENCES

[1] T. A. Soomro, J. Gao, T. M. Khan, A. F. M. Hani, M. A. U. Khan, and M. Paul, "Computerised approaches for the detection of diabetic retinopathy using retinal fundus images: A survey," *Pattern Anal. Appl.*, vol. 20, no. 4, pp. 927–961, 2017.
[2] S. Soomro, F. Akram, J. H. Kim, T. A. Soomro, and K. N. Choi, "Active contours using additive local and global intensity fitting models for intensity inhomogeneous image segmentation," *Comput. Math. Methods Med.*, vol. 2016, Sep. 2016, Art. no. 9675249.
[3] M. M. Fraza, P. Remagninoa, A. Hoppea, B. Uyyanonvarab, A. R. Rudnickac, C. G. Owenc, and S. A. Barmana, "Blood vessel segmentation methodologies in retinal images—A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, 2012.

[4] S. Soomro, F. Akram, A. Munir, C. Ha Lee, and K. N. Choi, "Segmentation of left and right ventricles in cardiac MRI using active contours," *Comput. Math. Methods Med.*, vol. 2017, Aug. 2017, Art. no. 8350680.

[5] S. Soomro, A. Munir, and K. N. Choi, "Hybrid two-stage active contour method with region and edge information for intensity inhomogeneous image segmentation," *PLoS ONE*, vol. 13, no. 1, 2018, Art. no. e0191827.

[6] U. T. V. Nguyen, A. Bhuiyan, L. A. F. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognit.*, vol. 46, no. 3, pp. 703–715, 2013.

[7] X. Xu, W. Ding, M. D. Abrámoff, and R. Cao, "An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image," *Comput. Methods Programs Biomed.*, vol. 141, pp. 3–9, Apr. 2017.

[8] K. Narasimhan and K. Vijayarekha, "Automatic grading of images based on retinal vessel tortuosity analysis," *Indian J. Sci. Technol.*, vol. 8, no. 29, pp. 1–5, 2015.

[9] T. A. Soomro, T. M. Khan, M. A. U. Khan, J. Gao, M. Paul, and L. Zheng, "Impact of ICA-based image enhancement technique on retinal blood vessels segmentation," *IEEE Access*, vol. 6, pp. 3524–3538, 2018.

[10] K. Bhatia, S. Arora, and R. Tomar, "Diagnosis of diabetic retinopathy using machine learning classification algorithm," in *Proc. IEEE 2nd Int. Conf. Next Gener. Comput. Technol. (NGCT)*, Oct. 2016, pp. 347–351.

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

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

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

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

[15] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1273–1284, May 2016.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

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

[19] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[20] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 791–798.

[21] N. Lopes and B. Ribeiro, "Deep belief networks (DBNS)," *Mach. Learn. Adapt. Many Core Mach. Practical Approach. Stud. Big Data*, vol. 7, pp. 155–186, 2015.

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

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

[24] B. Yoshua, L. Pascal, P. Dan, and L. Hugo, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 153–160.

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

[26] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 1989, pp. 396–404.

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

[28] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[30] J. Ngiam, Z. Chen, D. Chia, W. P. Koh, V. Q. Le, and Y. A. Ng, "Tiled convolutional neural networks," *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 1, pp. 1279–1287, 2010.

[31] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–13.

[32] M. Lin and Q. Chen, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–10.

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[34] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 30, 2013, pp. 1–6.

[35] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–14.

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

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

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

[39] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," 2015, pp. 1–38, *arXiv:1512.07108*. [Online]. Available: https://arxiv.org/abs/1512.07108

[40] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, pp. 1–18, *arXiv:1207.0580*. [Online]. Available: https://arxiv.org/abs/1207.0580

[41] R. G. J. Wijnhoven and P. H. N. de With, "Fast training of object detection using stochastic gradient descent," in *Proc. IEEE 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 424–427.

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

[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[44] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[45] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1139–1147.

[46] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[47] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. ICML*, 2016, pp. 507–516.

[48] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," 2017, *arXiv:1702.05659*. [Online]. Available: https://arxiv.org/abs/1702.05659

[49] A. J. Afifi and O. Hellwich, "Object depth estimation from a single image using fully convolutional neural network," in *Proc. IEEE Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov./Dec. 2016, pp. 1–7.

[50] A. Mateus, D. Ribeiro, P. Miraldo, and J. C. Nascimento, "Efficient and robust pedestrian detection using deep learning for human-aware navigation," 2016, pp. 1–16, *arXiv:1607.04441*. [Online]. Available: https://arxiv.org/abs/1607.04441

[51] S. Trebeschi, J. M. Joost van Griethuysen, D. M. J. Lambregts, M. J. Lahaye, C. Parmer, F. C. H. Bakers, N. H. G. M. Peters, R. G. H. Beets-Tan, and H. J. W. L. Aerts, "Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR," *Sci. Rep.*, vol. 7, p. 5301, Jul. 2017.
[52] Y. Wang, Y. Qiu, T. Thai, K. Moore, H. Liu, and B. Zheng, "A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images," *Comput. Methods Programs Biomed.*, vol. 144, pp. 97–104, Jun. 2017.
[53] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. A. U. Khan, M. Paul, and L. Zheng, "Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov./Dec. 2017, pp. 1–8.
[54] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[55] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.
[56] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, vol. 1. Cham, Switzerland: Springer, 2015, pp. 234–241.
[58] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," 2015, pp. 1–10, *arXiv:1505.07293*. [Online]. Available: https://arxiv.org/abs/1505.07293
[59] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *J. Comput. Sci.*, vol. 20, pp. 70–79, May 2017.
[60] T. A. Soomro, O. Hellwich, A. J. Afifi, M. Paul, J. Gao, and L. Zheng, "Strided U-Net model: Retinal vessels segmentation using dice loss," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, vol. 1, Dec. 2018, pp. 1–8.
[61] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403.
[62] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[63] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, Jr., H. F. Jelinek, and J. Michael Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.
[64] J. Zhang, Y. Cui, W. Jiang, and L. Wang, "Blood vessel segmentation of retinal images based on neural network," in *Image and Graphics. ICIG* (Lecture Notes in Computer Science), vol. 9218, Y. J. Zhang, Eds. Cham, Switzerland: Springer, 2015.
[65] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Comput. Med. Imag. Graph.*, vol. 55, pp. 28–41, Jan. 2017.
[66] D. Maji, A. Santara, P. Mitra, and D. Sheet, "Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images," 2016, pp. 1–4, *arXiv:1603.04833*. [Online]. Available: https://arxiv.org/abs/1603.04833
[67] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Sep. 2016, pp. 698–701.
[68] A. Wu, Z. Xu, M. Gao, M. Buty, and D. J. Mollura, "Deep vessel tracking: A generalized probabilistic approach via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, vol. 1, Apr. 2016, pp. 1363–1367.
[69] Z. Yao, Z. Zhang, and L.-Q. Xu, "Convolutional neural network for retinal blood vessel segmentation," in *Proc. 9th Int. Symp. Comput. Intell. Design (ISCID)*, vol. 1, Dec. 2016, pp. 406–409.
[70] H. Fu, Y. Xu, S. Lin, D. Wing, K. Wong, and J. Liu, "DeepVessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 132–139.
[71] D. Mahapatra, P. K. Roy, S. Sedai, and R. Garnavi, "Retinal image quality classification using saliency maps and CNNs," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2016, pp. 172–179.
[72] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 140–148.
[73] M. Li, Q. Yin, and M. Lu, "Retinal blood vessel segmentation based on multi-scale deep learning," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Sep. 2018, vol. 15, nos. 117–123, pp. 1–7.
[74] S. Guo, Y. Gao, K. Wang, and T. Li, "Deeply supervised neural network with short connections for retinal vessel segmentation," 2018, pp. 1–14, *arXiv:1803.03963*. [Online]. Available: https://arxiv.org/abs/1803.03963
[75] P. Chudzik, B. Al-Dir, F. Caliv, and A. Hunter, "DISCERN: Generative framework for vessel segmentation using convolutional neural network and visual codebook," in *Proc. IEEE 40th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, vol. 1, Jul. 2018, pp. 5934–5937.
[76] M. Hajabdollahi, R. Esfandiarpoor, K. Najarian, N. Karimi, S. Samavi, and S. M. Reza-Soroushmeh, "Low complexity convolutional neural network for vessel segmentation in portable retinal diagnostic devices," in *Proc. ICIP*, Oct. 2018, pp. 2785–2789.
[77] Z. Yan, X. Yang, and K.-T. T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, to be published.

**TOUFIQUE AHMED SOOMRO** received the B.E. degree in electronic engineering from the Mehran University of Engineering and Technology, Pakistan, in 2008, the M.Sc. degree in electrical and electronic engineering (research) from University Technologi PETRONAS, Malaysia, in 2014, and the Ph.D. degree in AI and image processing from the School of Computing and Mathematics, Charles Sturt University, Australia. He was a Research Assistant for 6 months with the School of Business Analytic in Cluster of Big Data Analysis, University of Sydney Australia. He is currently an Assistant Professor with the Electronic Engineering Department, QUEST, Larkana Campus, Pakistan. His research interests include most aspects of image enhancement methods, segmentation methods, classifications methods, and image analysis for medical images.

**AHMED J. AFIFI** was born in 1985. He received the bachelor's and M.Sc. degrees in computer engineering from Islamic University-Gaza, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Computer Vision and Remote Sensing Research Group, Technische Universität Berlin. During his master's degree, he was interested in digital image processing and patter recognition. His research interests include computer vision, deep learning, 3D object reconstruction from a single image, and medical image analysis.

**LIHONG ZHENG** received the Ph.D. degree in computer science from the University of Technology, Sydney, Australia, in 2008. She is currently a Senior Lecturer with the School of Computing and Mathematics, Charles Sturt University, Australia. She is leading Imaging and Sensing Research Group to conduct high quality research in machine learning, image processing, and information and communications technology (ICT) area. She has published more than 80 high-quality journal and conference papers. She led a team who received the 2nd place of the IoT Spartans Challenge, in 2017. She is a member of the Australian Computer Society (ACS) and the Australian Computer Society—Artificial Intelligence Committee. In 2019, she received the academia award of Women in IT by Cisco. As a technical referee, she has been serving many top-ranked IEEE and Elsevier journals and IEEE flagship conferences, and sitting on the Organizing Committee of many international IEEE conferences and workshops.

**SHAFIULLAH SOOMRO** received the B.E. degree from QUEST, Nawabshah, Pakistan, in 2008, the M.E. degree from MUET, Jamshoro, Pakistan, in 2014, and the Ph.D. degree in computer science from Chung-Ang University, Seoul, South Korea, in 2018. He is currently an Assistant Professor in computer science with the Quaid-e-Awam University of Engineering, Science and Technology, Larkana, Pakistan. His research interests include motion tracking, object segmentation, and 3D image recognition.

**JUNBIN GAO** received the B.Sc. degree in computational mathematics from the Huazhong University of Science and Technology (HUST), China, in 1982, and the Ph.D. degree from the Dalian University of Technology, China, in 1991. He was a Professor in computer science with the School of Computing and Mathematics, Charles Sturt University, Australia. From 1982 to 2001, he was an Associate Lecturer, a Lecturer, an Associate Professor, and a Professor with the Department of Mathematics, HUST. From 2001 to 2005, he was a Senior Lecturer and a Lecturer in computer science with the University of New England, Australia. He is currently a Professor of big data analytics with The University of Sydney Business School, The University of Sydney. His current research interests include machine learning, data analytics, Bayesian learning and inference, and image analysis.

**OLAF HELLWICH** was born in 1962. He received the B.S. degree in surveying engineering from the University of New Brunswick, Fredericton, NB, Canada, in 1986, and the Ph.D. degree in linienextraktion aus SAR-Daten mit einem Markoff-Zufallsfeld-Modell from the Technische Universität München, München, Germany, in 1997. He headed the Remote Sensing Group, Department of Photogrammetry and Remote Sensing, Technische Universität München. Since 2001, he has been a Professor with the Technische Universität Berlin (TUB), Berlin, Germany, initially for photogrammetry and cartography, and since 2004 for Computer Vision and Remote Sensing. From 2006 to 2009, he was the Dean of the Faculty of Electrical Engineering and Computer Science, TUB. His research interests include 3-D object reconstruction, object recognition, synthetic aperture radar remote sensing, and discovery and use of object shape priors in 3-D reconstruction. He was a recipient of the Hansa Luftbild Prize of the German Society for Photogrammetry and Remote Sensing, in 2000.

**MANORANJAN PAUL** received the Ph.D. degree from Monash University, Australia, in 2005. He was a Postdoctoral Research Fellow with the University of New South Wales, Monash University, and Nanyang Technological University. He is currently an Associate Professor and the Director of the Computer Vision Lab, and a Leader of the E-Health Research Group, Charles Sturt University (CSU), Australia. He has supervised 15 Ph.D. students to completion. He has published more than 160 refereed publications including 50 journals. His current research interests include video analytics, E-health, wine technology, and imaging/signal processing. He received the ICT Researcher of the Year 2017 awarded by the Australian Computer Society. He obtained more than AUD 2.5M competitive external grant money including two Australian Research Council (ARC) Discovery Project Grants. He is a General Co-Chair of PSIVT 2019, and a Program Co-Chair of PSIVT 2017 and DICTA 2018. He is currently an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and *EURASIP Journal in Advances on Signal Processing*. He has conducted invited keynote speeches in IEEE DICTA 2017 and 2013, WoWMoM 2014, and ICCIT 2010.