IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 September 2023, accepted 8 November 2023, date of publication 16 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3333364

RESEARCH ARTICLE

Dual-Branch U-Net Architecture for Retinal Lesions Segmentation on Fundus Image

MING YIN[1], (Member, IEEE), TOUFIQUE AHMED SOOMRO[2], (Senior Member, IEEE), FAYYAZ ALI JANDAN[3], (Graduate Student Member, IEEE), AYOUB FATIHI[4], FAISAL BIN UBAID[5], MUHAMMAD IRFAN[6], AHMED J. AFIFI[7], SAIFUR RAHMAN[6], SERGII TELENYK[8], AND GRZEGORZ NOWAKOWSKI[8]

[1] School of Semiconductor Science and Technology, South China Normal University, Foshan 528225, China
[2] Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science & Technology, Larkana Campus, Larkana 77111, Pakistan
[3] Electrical Engineering Department, Quaid-e-Awam University of Engineering, Science & Technology, Larkana Campus, Larkana 77111, Pakistan
[4] Faculty of Geosciences and Environment, Institute of Earth Sciences, University of Lausanne (UNIL), 1015 Lausanne, Switzerland
[5] Computer Science Department, Sukkur IBA University, Sukkur 65200, Pakistan
[6] Electrical Engineering Department, College of Engineering, Najran University, Najran 61441, Saudi Arabia
[7] Institute of Industrial Information Technology (IIIT), Karlsruhe Institute of Technology (KIT), 76187 Karlsruhe, Germany
[8] Faculty of Electrical and Computer Engineering, Cracow University of Technology, 31155 Cracow, Poland

Corresponding author: Toufique Ahmed Soomro (toufique_soomro@quest.edu.pk)

The Article Processing Charge (APC) was funded by the Faculty of Electrical and Computer Engineering, Cracow University of Technology and the Ministry ofScience and Higher Education, Republic of Poland (grant no. E-1/2024).

ABSTRACT Deep learning has found widespread application in diabetic retinopathy (DR) screening, primarily for lesion detection. However, this approach encounters challenges such as information loss due to convolutional operations, shape uncertainty, and the high similarity between different lesions types. These factors collectively hinder the accurate segmentation of lesions. In this research paper, we introduce a novel dual-branch U-Net architecture, referred to as Dual-Branch (DB)-U-Net, tailored to address the intricacies of small-scale lesion segmentation. Our approach involves two branches: one employs a U-Net to capture the shared characteristics of lesions, while the other utilizes a modified U-Net, known as U2Net, equipped with two decoders that share a common encoder. U2Net is responsible for generating probability maps for lesion segmentation as well as corresponding boundary segmentation. DB U-Net combines the outputs of U2Net and U-Net as a dual branch, concatenating their segmentation maps to produce the final result. To mitigate the challenge of imbalanced data, we employ the Dice loss as a loss function. We evaluate the effectiveness of our approach on publicly available datasets, including DDR, IDRiD, and E-Ophtha. Our results demonstrate that DB U-Net achieves AUPR values of 0.5254 and 0.7297 for Microaneurysms and soft exudates segmentation, respectively, on the IDRiD dataset. These results outperform other models, highlighting the potential clinical utility of our method in identifying retinal lesions from retinal fundus images.

INDEX TERMS Deep learning, neural network, U-net, computer-aided diagnostic, retinal lesions segmentation.

I. INTRODUCTION

Diabetes federations predict that the number of people with diabetes will rise from 463 million to 700 million over the next 25 years if sufficient measures are not taken to combat the spread of diabetes [1]. As a common chronic complication of diabetes, DR remains one of the top five causes of irreversible blindness in adults [2]. In clinical practice, a fundus image is a projection of the fundus captured by a monocular camera onto a 2D plane. There are parameters in the fundus image, such as optic disc (OD), macula, fovea, blood vessels or some lesions related to DR:

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

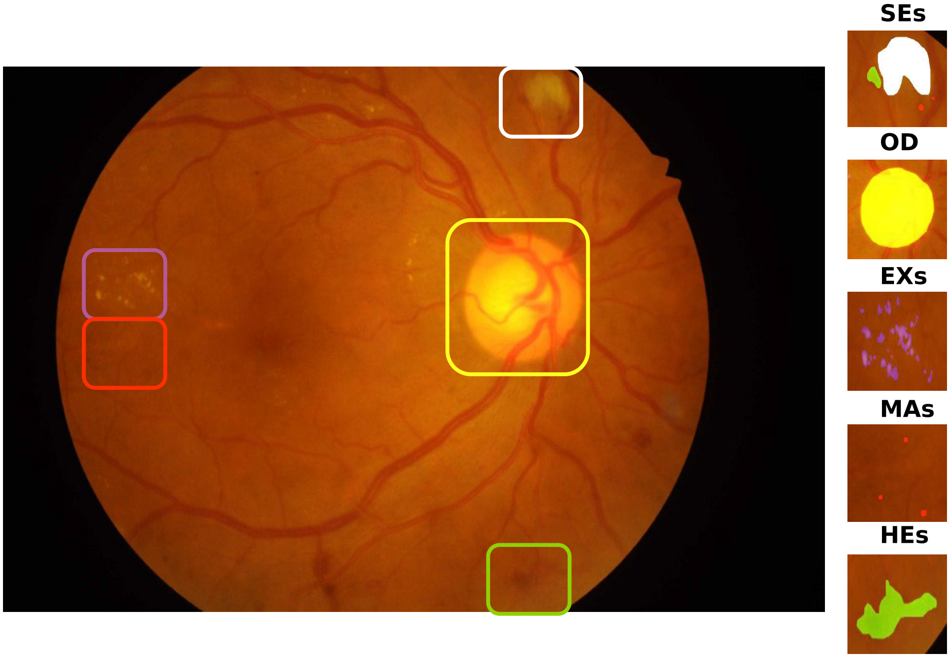


FIGURE 1. Illustration of the retinal image by highlighting normal structures (optic disc) and abnormalities associated with DR in different color: MAs, HEs, SEs, and EXs.

microaneurysms (MA), hemorrhages (HE), hard exudates (EX) and soft exudates (SE). Microaneurysms (MAs) are the first visible signs of DR, which are small swellings in the tiny blood vessels of the retina that appear as tiny, round, red spots. HEs form due to blood leakage and appear as a small white dot or spot. EXs and SEs are bright object [3], as listed in Figure 1. Fundus images can be acquired non-invasively and economically, making them more suitable for large-scale screening. Retinal lesions can be visualized on fundus images [4]. Thus, ocular screening by fundus images is important in the diagnosis of DR. It is therefore essential that treatment can be administered for the prevention of vision loss.

The analysis of the lesions from the retinal fundus image is represented with their shape, their texture and their location of appearance which are the main indicators for evaluating the evolution of the disease [5]. Quantitative analysis of fundus images is important, but analysis of the visualization base has played an important role for disease diagnosis in the screening process. But Performing a manual analysis is a laborious task, and the diagnosis of anterior diabetic retinopathy relies on automated lesion segmentation. However, there are certain constraints in the automated lesion detection procedure.

1) The retinal fundus images suffered from various issues such as noise, uneven illumination and variable low contrast, as well as abnormalities in their parameters such as retinal vessels, drusen and optic disc. Another major issue is that the lesion size of the retinal fundus image is smaller than the background and it is challenging to detect the lesion.
2) The shape, texture, and color of the lesion in retinal fundus images make it challenging to detect the lesion, since the shape and color of MAs and HEs are almost identical and appear as red dots in the fundus and other side images, the SEs and EXs appeared as bright spots. Many researchers [6], [7], [8] have misclassified lesion detection, and it is difficult to establish appropriate classes for lesion detection from retinal fundus images.

Deep learning-based methods especially fully convolutional neural networks have achieved great success in medical image segmentation. The fully convolutional networks (FCNs) [9] and U-Net [10] played an important role in segmentation especially segmentation on medical images. After successfully use of U-Net, we implemented a novel approach to improve the performance of retinal lesion segmentation from color retinal fundus images. U-Net contains an encoder and symmetric decoder to perform the segmentation. The encoder is employed to extract features, while the feature extraction processes are connected with the decoder. The stack of convolutional layer, batch normalization, rectified linear unit layer (ReLU), and the following max-pooling layer are the basic architecture for extracting features for segmentation of retinal lesion. With the increasing layers, all information related to the lesion is extracted as high-level semantic information while a part of low-level semantics includes information on color, texture, and shape, and such information is not the main purpose for segmentation else it plays an important role for segmentation tiny lesion. Inspiring by [11], Takikawa et al. propose a new two-stream CNN architecture for semantic segmentation that explicitly shape information as a separate processing branch, that processes information in parallel to the classical stream. Furthermore, Zhou et al. propose a Bilateral branch network (BBN) [12], each branch of which performs its duty separately. The feature maps from the base network are extracted to perform different segmentation simultaneously and this method may work well on other similar types of tasks.

Furthermore, as previously mentioned, John et al.'s proposal [13], [14] underscores the advantages of auxiliary information and the dual-branch architecture in augmenting the acquisition of contextual features. This enhancement contributes to improved lesion segmentation performance. We propose an end-to-end dual-branch U-Net architecture lesion segmentation framework (DB U-Net) that contains two branches which are composed of U-Net and a modified U-Net respectively. The output of U-Net is supervised by red/bright lesions compared with ground truth (GT) to learn common features among red/bright lesions. The modified U-Net namely U2Net consists of two decoders: lesions segmentation biased learning decoder and boundary of lesion biased learning decoder. To preserve low-level information that might be lost during the down-sampling process, U2Net incorporates additional boundary information through an extra decoder. This involves combining two branches with an image patch to create feature maps, which are then inputted into the fusion module as the final step in the DB U-Net process. The research paper's novelty lies in its comprehensive approach to addressing the challenges of retinal lesion segmentation in diabetic retinopathy. It combines deep learning techniques with innovative architectural modifications to improve automatic lesion detection's accuracy and diagnostic capabilities in retinal fundus images. The following are the main contributions of this research work. Our main contributions and novelty are as follows

1) We address the challenges inherent in the automatic lesion detection process within retinal fundus images, including image quality, small lesion sizes, and the similarity in appearance of different lesions. We introduce deep learning techniques to overcome these challenges, specifically focusing on fully convolutional neural networks (FCNs) and U-Net. This proposed network aims to enhance accuracy in retinal lesion segmentation, leveraging the potential of these advanced neural networks.
2) We present innovative architectural modifications to improve segmentation performance, including a two-stream CNN architecture that explicitly considers shape information and a dual-branch U-Net architecture (DB U-Net) featuring two distinct branches. Additionally, we propose a supervised learning approach where the model's output is guided by specific types of lesions (red/bright), facilitating improved learning of common features among these lesions compared to ground truth (GT).
3) We introduce innovative enhancements to segmentation models, including incorporating boundary information through an extra decoder in U2Net to preserve low-level details during the down-sampling process, potentially enhancing segmentation accuracy. Additionally, in our DB U-Net architecture, we implement a fusion module at the end, combining feature maps from two branches and image patches. This integration aims to improve further the model's ability to capture essential features for segmentation.
4) Our evaluation encompasses three publicly accessible datasets: IDRiD [15], E-Ophtha [16], and DDR [17]. Through ablation studies, we analyze the impact of various design choices on lesion segmentation performance. Comparative assessments reveal that DB U-Net achieves remarkable and competitive results in comparison to state-of-the-art segmentation models like U-Net, DeepLab v3+ [18], and other segmentation models. This stands as a significant contribution, as our method consistently outperforms alternative approaches across diverse databases.

This paper is organized as follows. Section II discuss related works. Section III provides a comprehensive architecture of proposed model. Section IV explains the datasets and configure for perform an experiments. Section V analyses the quantitative results and experimental performance. Finally, we present conclusions in Section VI.

II. RELATED WORK

Table 1 below provides a thorough review of previous research, presenting a comparative analysis of existing work. Additionally, a detailed discussion of these earlier contributions is presented.

Table 1 showcases a comparative analysis that emphasizes the variety of datasets, lesion types, neural network architectures, and distinctive features employed across various studies focused on detecting and segmenting retinal lesions. Each study employs distinct techniques and innovations to enhance model accuracy, illustrating the continuous research and progress in this domain. The subsequent explanations delve into the specifics of each method, outlining their contributions and limitations as explained below.

Haloi et al. [19] present a novel method for early diabetic retinopathy screening, focusing on MA detection in color fundus images. They employ deep neural networks with dropout training and max-out activation functions, eliminating the need for preprocessing and manual feature extraction. While claiming substantial improvements, quantitative evidence is lacking. The method achieves state-of-the-art accuracy on benchmark datasets. Still, it faces limitations, including data diversity, interpretability, computational demands, false positives/negatives, dataset-dependent performance, lack of detailed comparisons, data balance, and robustness to noise-highlighting the need for further research and refinement.

Chudzik et al. [20] present an innovative approach to automated MA detection in fundus images, a crucial component of diabetic retinopathy screening. They employ a patch-based fully convolutional neural network with batch normalization layers and use the Dice loss function, simplifying the

TABLE 1. Summary of several researches for lesion detection/segmentation.

Techniques	Dataset	Leison	Model	Characteristics
Haloi et al [19]	Messidor	MAs	CNN	–
Chudzik et al [20]	E-ophtha	MAs	FCN	Dice Loss
Kou et al [21]	E-ophtha	MAs	U-Net	Deep Residual Model, Recurrent Convolutional
Sarhan et al [22]	IDRiD	MAs	CNN	Multi Scale ,Selective Sampling, Embedding Triplet Loss
Theelen et al [23]	Messidor	HEs	CNN	Selective Sampling
Zheng et al [24]	Messidor	EXs	U-Net	Ensemble, cGAN
Yan et al [25]	IDRiD	All	U-Net	Global-level and Local-level Information
Guo et al [26]	E-Ophtha,DDR	All	CNN	Multi-Channel Bin Loss and Multi Scale
Wang et al [27]	IDRiD,DDR	All	CNN	Scale-aware Attention, Multi Scale
Liu et al [28]	IDRiD,DDR	All	CNN	Dual-branch Network , Dual-sampling Modulated and Dice Loss

process with just three processing stages, contrasting with methods requiring up to five. Notably, the paper demonstrates successful knowledge transfer between datasets within the MA detection domain, potentially enhancing adaptability across different data sources. The method's evaluation on popular datasets, including E-Ophtha, DIARETDB1, and ROC, showcases robust performance, surpassing state-of-the-art methods based on the FROC metric. It excels in achieving high sensitivities for low false positive rates, enhancing its promise for diabetic retinopathy screening. However, the study does not explicitly address potential limitations such as further diversity evaluation, model interpretability, handling class imbalance, and comprehensive comparisons with existing methods, factors essential for assessing its real-world relevance.

Kou et al. [21] present the deep recurrent U-Net (DRU-Net), a deep learning approach for MA segmentation in diabetic retinopathy diagnosis. MAs are critical indicators, but manual annotation is cumbersome, prompting the need for automation. The DRU-Net combines U-Net, deep residual models, and recurrent convolutions to enhance feature accumulation, addressing low contrast and small MA challenges. It achieves impressive results on E-Ophtha and IDRiD datasets, notably an accuracy of 0.9999, an AUC of 0.9943 on E-Ophtha, and 0.987 AUC on IDRiD. It outperforms U-Net, FCNN, and ResU-Net, establishing itself as a state-of-the-art MA segmentation method. However, potential limitations include computational demands, generalizability, interpretability, data augmentation, which should be considered for practical use beyond specific datasets.

Sarhan et al. [22] present a two-stage deep learning approach for MA segmentation in diabetic retinopathy detection, underscoring the importance of deep learning in fundus image analysis. MAs are vital markers of diabetic retinopathy progression. Their method leverages multiple input scales, allowing for consideration of features at various resolutions, crucial for accommodating MA size variations. Additionally, selective sampling enhances computational efficiency by focusing the model's attention on key image regions. Embedding triplet loss is introduced to enhance classification model discriminative power, resulting in a significant 30.29% relative improvement over fully convolutional neural networks (FCNs) in MA segmentation. However, the study lacks discussion of potential limitations, including computational complexity, generalizability, interpretability and dataset diversity. These considerations are vital for assessing the method's practical applicability beyond its reported performance.

Theelen et al. [23] present a method aimed at improving CNN training for medical image analysis, with a focus on hemorrhage detection in colored fundus images. They address the challenge of time-consuming CNN training by introducing selective sampling, dynamically choosing misclassified negative samples to prioritize informative data during the learning process. Their results show a substantial reduction in training time, maintaining or improving performance, achieving AUC values of 0.894 and 0.972 on two datasets, and demonstrating potential for model generalization. However, limitations include the method's application specificity, questions about generalizability to diverse data sources, lack of detail on heuristic sampling criteria, and the need for clinical validation in real healthcare settings. These considerations are crucial when assessing the method's practical applicability beyond its promising yet specialized performance.

Zheng et al. [24] present a deep learning approach for detecting retinal exudates, an early indicator of diabetic retinopathy (DR). They tackle challenges in deep convolutional neural network (DCNN) application by introducing an ensemble convolutional neural network (MU-net) to cope with limited labeled data and adopting conditional generative adversarial networks (cGANs) for mitigating severe class imbalance. This strategy enhances model robustness and generalization across diverse datasets and clinical scenarios. The method demonstrates significant performance improvements, reflected in higher F1 scores at the lesion level and increased accuracy at the image level compared to non-cGAN approaches. However, limitations such as the need for computational resource requirements, dataset diversity, and model interpretability must be considered when assessing its practical applicability beyond benchmark datasets.

Yan et al. [25] propose an innovative method for segmenting small lesions in high-resolution retinal images. They acknowledge downsampling and patch-based methods' limitations and introduce mutually local-global U-nets to balance local and global context. While their method shows promise, quantitative comparisons with existing techniques are lacking, and they plan to collect more data for future research. They also suggest the model's potential for broader applications beyond retinal lesion segmentation, although concrete evidence is missing. Further validation and exploration are needed. Guo et al. [26] introduce a significant contribution to DR and diabetic macular edema diagnosis by developing the L-Seg multi-lesion segmentation model. This model addresses challenges related to the diagnosis of these conditions by simultaneously segmenting four types of lesions in fundus images. L-Seg is notable for being the first small object segmentation network capable of concurrently handling soft exudates, hard exudates, microaneurysms, and hemorrhages. The method incorporates a multi-scale feature fusion technique to enhance its performance. It introduces a multi-channel bin loss to address the class imbalance and loss-imbalance issues during training. Extensive evaluations on various datasets showcase L-Seg's superiority over other deep learning models and traditional methods, particularly excelling in small lesion segmentation. The limitations highlight the need for further research to ensure the model's applicability and robustness beyond the evaluated datasets and challenges.

Wang et al. [27] contribute to the field of diabetic retinopathy diagnosis by addressing the challenging task of multiple lesion segmentation. Their work introduces a scale-aware attention (SAA) block designed to effectively handle variations in lesion scales. Through extensive experimentation, they establish the superiority of the SAA block over existing attention mechanisms, achieving state-of-the-art results in the domain. However, the study falls short in terms of clinical validation and comprehensive comparisons with existing methods. Additionally, it does not delve into considerations related to computational resource requirements and scalability. These limitations highlight the necessity for further research and real-world validation to ascertain the practical applicability of their approach.

Liu et al. [28] introduce a dual-branch network designed to segment hard exudates in color fundus images. These exudates vary significantly in size, and class imbalance issues complicate their segmentation. The dual-branch network employs two branches with partially shared weights, allowing it to effectively learn features and classifiers for hard exudates of different sizes. During training, they utilize a novel dual-sampling modulated Dice loss, prioritizing the segmentation of large exudates before addressing smaller ones. Their experimental evaluations, conducted on publicly available datasets for hard exudate segmentation, demonstrate the superior performance of the dual-branch network compared to existing methods that use both CBCE (Class Balanced Cross-Entropy) loss and Dice loss. This suggests their novel network architecture and loss function significantly enhance segmentation accuracy. However, it's important to acknowledge certain limitations. The study primarily focuses on showcasing the effectiveness of their dual-branch network but does not thoroughly explore potential limitations related to its clinical applicability. Additionally, the comparison with existing methods is somewhat limited, warranting further research for a more comprehensive evaluation of the network's real-world potential and potential drawbacks.

The comparative analysis of these methods highlights their innovative approaches to detecting and segmenting retinal lesions. Nevertheless, each method comes with its own set of limitations. Although these approaches show promise, they frequently lack thorough evaluations and do not fully account for real-world challenges like diverse datasets, computational requirements, generalization, and interpretability. In light of these observations, we will now introduce our proposed approach, which aims to address the limitations identified in the existing methods.

III. PROPOSED METHOD

The proposed methods contain different tasks and the model of proposed methods is shown in Figure 2. Each part of these methods is elaborated below.

A. PROPOSED ARCHITECTURE

In the architecture section, we explained the DB-U-Net model of our proposed method, as shown in Figure 2. The network model contained the dual branches of the network with the fusion module. The first branch of the proposed network is U-Net and the second branch of the proposed network is U2Net. The U2Net model is based on a multi-task learning network that explores the boundaries of information that can extract lesions and give information about appropriate boundaries. Next, we used the fusion features based on the branch feature maps to produce the image patches for precise segmentation. In the training process, U2Net on the branch is performed by predicting the target lesion and the corresponding boundary while the other branch was supervised by a red and bright lesion label to learn the common characteristic of a similar lesion and their prediction process is explained below and in the next section we elaborate the subnet architecture of our proposed method.

B. U-NET ARCHITECTURE

The main objective is to detect the lesion as much as possible, and the U-Net branch is used to train the common features based on the red and bright lesion to detect the required lesion. The U-Net is composed of the encoder and the decoder with a convolutional layer instead of fully connected layers. This process is used to convert input images into binary image maps. It takes the image patches with $X \in \Re^{H\times W\times 3}$ and it is a process of entering and exiting segmentation of the red and bright lesion and it is denoted by $\widehat{Y_2}$. In this proposed method, the residual architecture according to LinkNet [29] can optimize the final performance and we modify the residual block [30] replaced by a convolution block in the typical U-Net. Additionally, we have reduced the

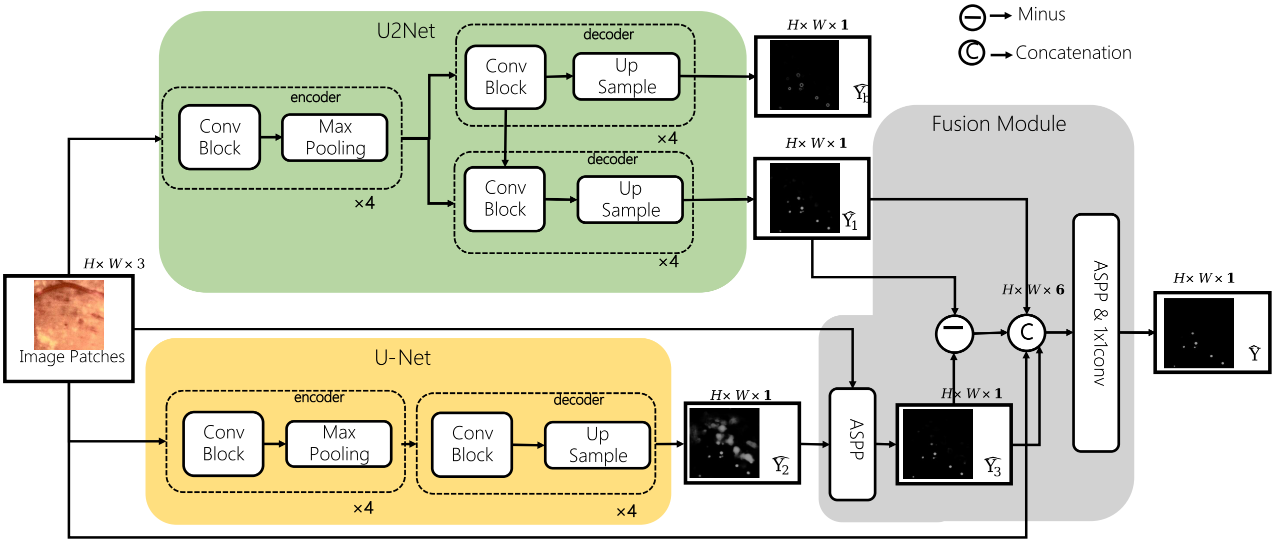


FIGURE 2. **An overview of the proposed segmentation framework by highlighting the three modules in different colors. The two branches of DB U-Net are composed of U-Net and U2Net, respectively. The fusion module takes segmentation maps from two branches as input and outputs the final segmentation result, denoted as $\hat{Y}$.**

number of convolution cores to half of the typical architecture by proposing to increase the trade-off between speed and accuracy.

$$F_{conv}(X) = ReLU(BN(Conv(X)). \quad (1)$$

$$X_{out} = Pooling(F_{conv}(F_{conv}(X)) + Conv_{1\times1}(X)). \quad (2)$$

where X, X_{out} are the input and output the residual block in downsampling path. $BN(\cdot)$ denotes the batch normalization layer. $ReLU(\cdot)$ denotes rectified linear unit layer. $Conv(\cdot)$ denotes convolutional layer and $Conv_{1\times1}(\cdot)$ is an identity mapping function. $Pooling(\cdot)$ is a max-pooling function.

C. U2NET ARCHITECTURE

The proposed U2Net architecture is shown in the Figure 3.

The deep neural network acts as a black box to extract the feature from the input. Downward sampling path depends on increasing layers to extract low-level information from exudates such as color, shape. High-level functionality such as the border is phased out as the [31] pattern is implemented. In this research work, we designed the modified U-Net model named U2Net model. The U2Net model is used to overcome the lack of information, as shown in Figure 3. The decoder in U2Net is to share the characteristics of its encoder to predict the lesion and its boundary respectively as shown in Figure 2. The boundary information in U2Net is introduced as an auxiliary by an additional decoder to avoid the loss of low-level information in order to obtain a good segmentation of the exudates.

The architecture of U2Net is implemented based on U-Net because it is based on a convolutional block replaced by the residual block with a number of convolution cores per convolutional layer reduced by half. The green block in the Figure 2 represents the input $X \in R^{H\times W\times 3}$ and the output of U2Net: lesion segmentation map $\hat{Y}_1$, and the corresponding boundary segmentation map $\hat{Y}_b$ respectively.

D. ARCHITECTURE OF FUSION MODULE

The segmentation map of red/bright lesion and target lesion are obtained from U-Net and U2Net respectively. The fusion module takes the image patches and the segmentation map as the input to get the final segmented lesion. Firstly, the segmentation of red/bright lesion $\hat{Y}_2$ is concatenated with image patch X along the channel dimension in order to forming a new feature map. In this step, the image X and segmentation of red/bright lesion $\hat{Y}_2$ are concatenated within the fusion module to stack the channels together. Then we merge the feature map using an Atrous Spatial Pyramid Pooling (ASPP) [32] which is containing multiple atrous convolutions with different sampling rates to capturing the context of the image at multi-scales. It combines segmentation map from U-Net branch with image patch X and the output of ASPP $\hat{Y}_3$ supervised by the target lesion according to groundtruth.

Finally, we obtain two segmentation $\hat{Y}_1$ and $\hat{Y}_3$. Moreover, we assume the difference between$\hat{Y}_1$ and $\hat{Y}_3$ can be viewed as different regions. So, to highlight these regions, we subtract $\hat{Y}_3$ and $\hat{Y}_1$ element-wise and take its absolute value, which is contributed to encourage fusion module focus on the difference between the two feature maps. We concatenate the aforementioned segmentation map and denote it as $X_{concat} \in R^{H\times W\times 6}$ $X_{concat} \in R^{H\times W\times 5}$. The following ASPP and 1×1 convolution layer transform X_{concat} to the final segmentation $Y \in R^{H\times W\times 1}$. The algorithm for the fusion module is presented below (see Algorithm 1).

E. LOSS FUNCTION

During the training process, we simultaneously train the U-Net and U2Net sub-networks, incorporating the fusion

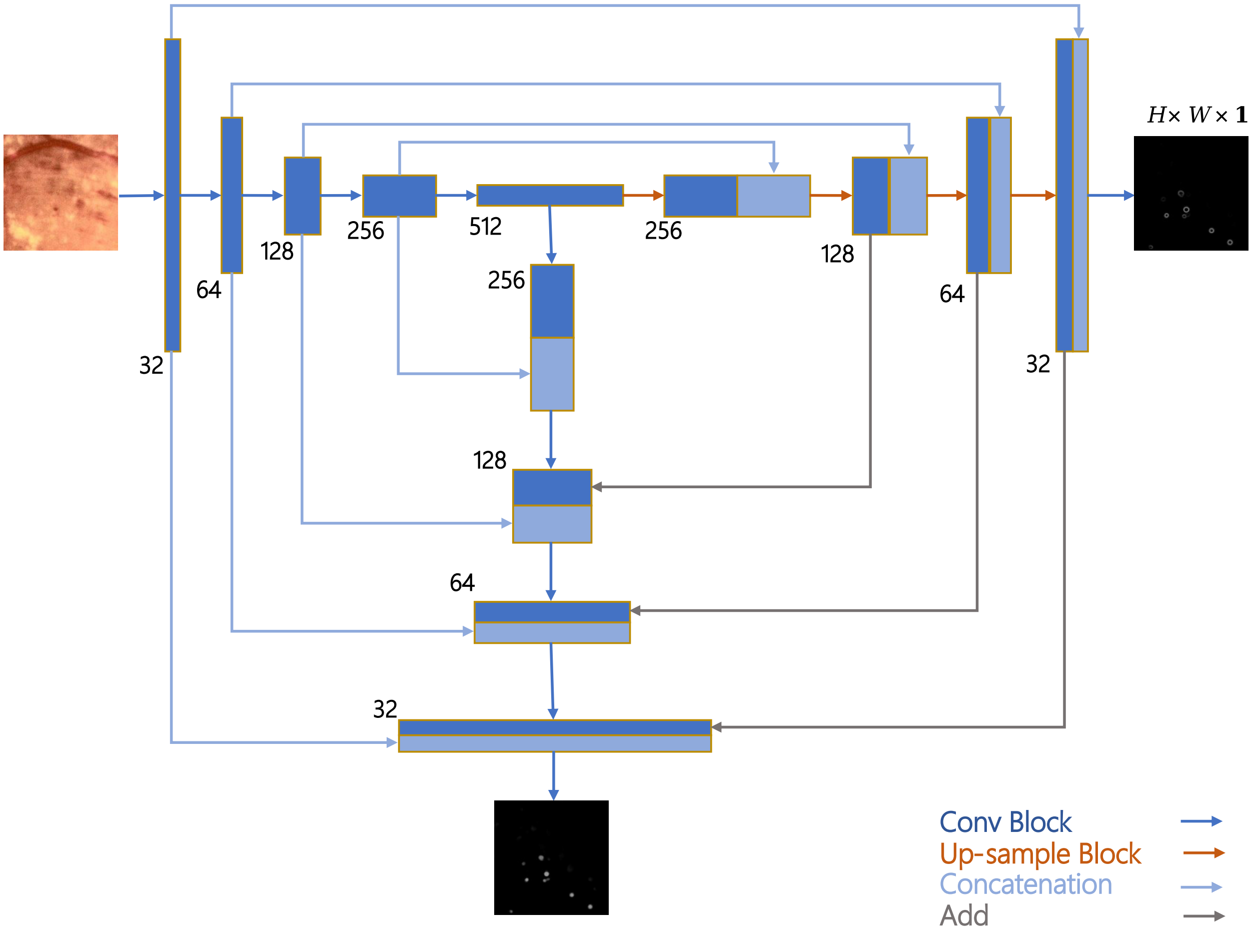


FIGURE 3. An overview of U2Net. Dual decoders share the same latent representation from the encoder. Two upsampling paths take care of lesion and boundary segmentation.

```
Algorithm 1 The Fusion Module Algorithm
Input: Image patch $X$, Lesion segmentation map $\hat{Y}_1$
       from U2Net, Segmentation map of red/bright
       lesion $\hat{Y}_2$ from U-Net
Output: Final segmented lesion map $\hat{Y}$
$\hat{Y}_2' = Concat(\hat{Y}_2, X)$;
$\hat{Y}_3 = ASPP(\hat{Y}_2')$;
$\hat{Y}_{1|3} = |\hat{Y}_3 - \hat{Y}_1|$;
$X_{concat} = Concat(\hat{Y}_{1|3}, \hat{Y}_3, \hat{Y}_1, X)$;
$\hat{Y} = Conv_{1\times 1}(ASPP(X_{concat}))$;
```

module to supervise both segmentation and boundary map predictions. The total loss function is expressed as follows:

$$L = L_{U-Net}\left(Y_b, \hat{Y}_b\right) + L_{U2Net}\left(Y_2, \hat{Y}_2\right) + L_{Fusion}\left(Y, \hat{Y}\right). \tag{3}$$

where L represents the total loss function.

$L_{U-Net}\left(Y_b, \hat{Y}_b\right)$ is the loss function specific to the U-Net sub-network, which measures the error between the ground truth segmentation data Y_b and the predicted segmentation $\hat{Y}_b$.

$L_{U2Net}\left(Y_2, \hat{Y}_2\right)$ is the loss function specific to the U2Net sub-network, which measures the error between the ground truth Y_2 and predicted $\hat{Y}_2$ values of segmentation.

$L_{Fusion}\left(Y, \hat{Y}\right)$ is the loss function related to the fusion module, which evaluates the discrepancy between the ground truth Y and the predicted $\hat{Y}$ fusion outcomes.

Equation 3 combines these three loss components to create a comprehensive loss function that guides the training of the U-Net and U2Net sub-networks with the fusion module. The objective during training is to minimize this total loss L which helps improve the accuracy of segmentation and boundary map predictions.

The modules employ the Dice loss as their loss function. “The Dice loss, as referenced in [33], serves as a valuable metric for gauging the extent of overlap between the Ground Truth (GT) and the segmented output. It relies on the Dice coefficient for its calculation. This approach obviates the necessity to meticulously fine-tune the balance between foreground and background elements within the data.

Given the inherent imbalance in the distribution of lesion pixels and background pixels, the adoption of the Dice loss as the primary loss function is a strategic choice. Mathematically, the Dice loss is represented as follows:

$$L_{dice} = 1 - \frac{2\sum_{x\in\Omega} p_l(x)g_l(x)}{\sum_{x\in\Omega} p_l^2(x) + \sum_{x\in\Omega} g_l^2(x)}. \quad (4)$$

In this context, $p_l(x)$ represents the probability assigned to pixel x for belonging to class l and $g_l(x)$ corresponds to a vector indicating the ground truth label, where it assumes a value of one for the correct class and zero for all other classes. This formulation of the Dice loss effectively addresses the challenge posed by unbalanced training data. Consequently, there is no need to introduce weighting parameters between various classes, such as the background and the vessel tree, during training. This makes the loss function particularly suitable for binary segmentation tasks.

IV. EXPERIMENT

A. DATASET

The Indian Diabetic Retinopathy Picture (IDRiD) [15] is provided by the 2018 ISBI Grand Challenge on Segmentation and Classification of Diabetic Retinopathy. It is used in this research work. Three problems can be solved using this database: lesion segmentation, disease classification, and OD detection and segmentation. For lesion segmentation, the dataset consists of 81 challenging fundus images with 54 images for training and the remaining 27 images for testing, each having a resolution of 4288 × 2848 and pixel-level lesion annotations. Specifically, there are 81 MA frames, 81 EX frames, 80 HE frames, 40 SE frames. For training data, each image was selected with an ROI of $2,560 \times 3,840$ and then scaled down by t times. For AM segmentation, the size of the ROI was halved ($t = 2$). Each image was divided into 24 patches with 320 pixels for training. For the other segmentation of the lesion, the cropped image was reduced by 4 times ($t = 4$) and was divided into 6 patches with 320 × 320.

DDR [17] is the largest dataset proposed in 2019 for DR screening, containing 13,673 images obtained from 147 hospitals, covering 23 provinces in China. These images were captured using 42 types of fundus cameras with a 45 degree field of view and range in resolution from 1088 × 1920 to 3456 × 5184. For lesion segmentation, DDR provides 757 fundus images with pixel-level annotation. There are 383 images for training, 149 for validation, and 225 for testing. We scaled all fundus images by 4 times ($t = 4$), while 320 × 320 patches were uniformly cropped from these images for training.

The E-Ophtha [16] is a publicly available dataset that consists of two parts: E-Ophtha EX and E-Ophtha MA. In our experience, E-Ophtha MA is adopted only, which consists of 148 images with MAs or small HEs and 233 healthy images with resolution ranging from 1440 × 960 to 2544 × 1696 pixels and provides annotations at the pixel level for AM segmentation. In our experiments, we used 100 images for training and the remaining 48 images for testing. For lesion segmentation, we scaled these images to 1360 × 2048 and then cropped the images to 1280 × 1920. Each cropped image was divided into 24 patches with 320 × 320.

B. PREPROCESSING

There are serious challenges such as uneven illumination, high variability in contrast, and background noise from data acquisition devices on the original fundus images. To track these issues, we developed an image enhancement method inspired by [5] and [34] to mitigate the influence of the aforementioned challenges. We apply histogram equalization (HE) and contrast-limited adaptive histogram equalization (CLAHE) [35] on the brightness channel of the LAB fundus image. Concretely, a Gaussian filter is used to remove the noise caused by the device and zoom in at first. Next, we transform the color space of the images from RGB to LAB. HE distributes the pixel intensities of the image according to all the information of the image to improve the contrast. But it also amplifies background noise. CLAHE can remove noise and retain detail by limiting contrast. After pre-processing, the original fundus images are split into multiple image patches in uniform resolution and used for data augmentation.

C. DATA AUGMENTATION

Data augmentation is an essential method to improve model robustness and accuracy by artificially augmenting the training data available in deep learning. In this article, there are two methods of data augmentation: geometric and lightweight. The first includes vertical flip, horizontal flip, and rotation, which are processed on both the input image and its corresponding ground truth segmentation. While the latter adjusts the input brightness based on gamma correction and only applies to the input image. Before training, several transformations randomly combine and apply to each input image patch. In this whole process, We employed publicly available databases for validation of our proposed, and each of these databases includes images with ground truth or pixel-level annotations. These resources were utilized to validate and compare the effectiveness of the method we have proposed.

D. EXPERIMENT DETAIL

The experimental environment is single Inter CPU Intel (R) Xeon (R) CPU e5-2620 V3 @2.40GHz and NVIDIA GeForce GTX 1080Ti, with 16G video memory. The model proposed in this paper is implemented in Python 3.5 and PyTorch 1.12. For training, each model has trained 1,000 epochs. We performed the Adam algorithm with a batch size of 8, a max iteration of 1,000, a momentum of 0.9, a weight decay of 5e-4, and an initial learning rate of 1e-3. In the comparative experiments, we employed ResNet-101 as the backbone architectures of DeepLab v3+ and adapted COCO pretrained weights as the initial weights. The learning rate was initialized to 1e-4 to fine tune DeepLab v3+.

E. EVALUATION METRICS

The performance of model for lesions segmentation was evaluated by Area under Precision-Recall curves (AUPR) [36] as AUPR was used for evaluation by ISBI IDRiD challenge [15] in 2018. Precision-Recall curves (PR curve)is used and PR curve is a plot of the *PPV* (y-axis) and the *Sen* (x-axis) for different probability thresholds which is set as 33 equally spaces instance from 0 to 1 in probabilities in our experiment. AUPR are recommended for imbalanced binary classification task where Area under ROC curves (AUC) may provide an excessively optimistic view of theperformance [36], [37]. PR curves are recommended for tasks with imbalanced binary classification models where ROC curves may provide an excessively optimistic view of the performance [36], [37]. Other evaluation metrics are defined in Table 2.

TABLE 2. Definitions of the evaluation metrics [24].

Evaluation Metrics	Mathematical Formula
Accuracy	$\frac{(TP+TN)}{(TP+TN+FN+FP)}$
Precision	$\frac{TP}{(TP+FP)}$
Sensitivity	$\frac{TP}{(TP+FN)}$
Specificity	$\frac{TN}{(TN+FP)}$
F1	$\frac{(2*TP)}{(2*(TP+FP+FN))}$

Where *TP* indicates True Positive; *FP*: False Positive; *TN*: True Negative; *FN*: False Negative

V. EXPERIMENT RESULT

In this section, we introduce the results of our proposals on three public datasets. Initially, we provide an ablation study to show the effectiveness of each component. Comparing with other extensive models and show qualitative results of our method are also performed. AUPR was used as the main evaluation metric as same as the 2018 ISBI grand challenge. And the Illustration of lesion segmentation results on a fundus image from IDRiD is shown in Figure 4.

A. ABLATION STUDIES

From Table 3 to Table 5, we tested the impact of different modules on the results on IDRiD, DDR, E-Ophtha MA datasets respectively.

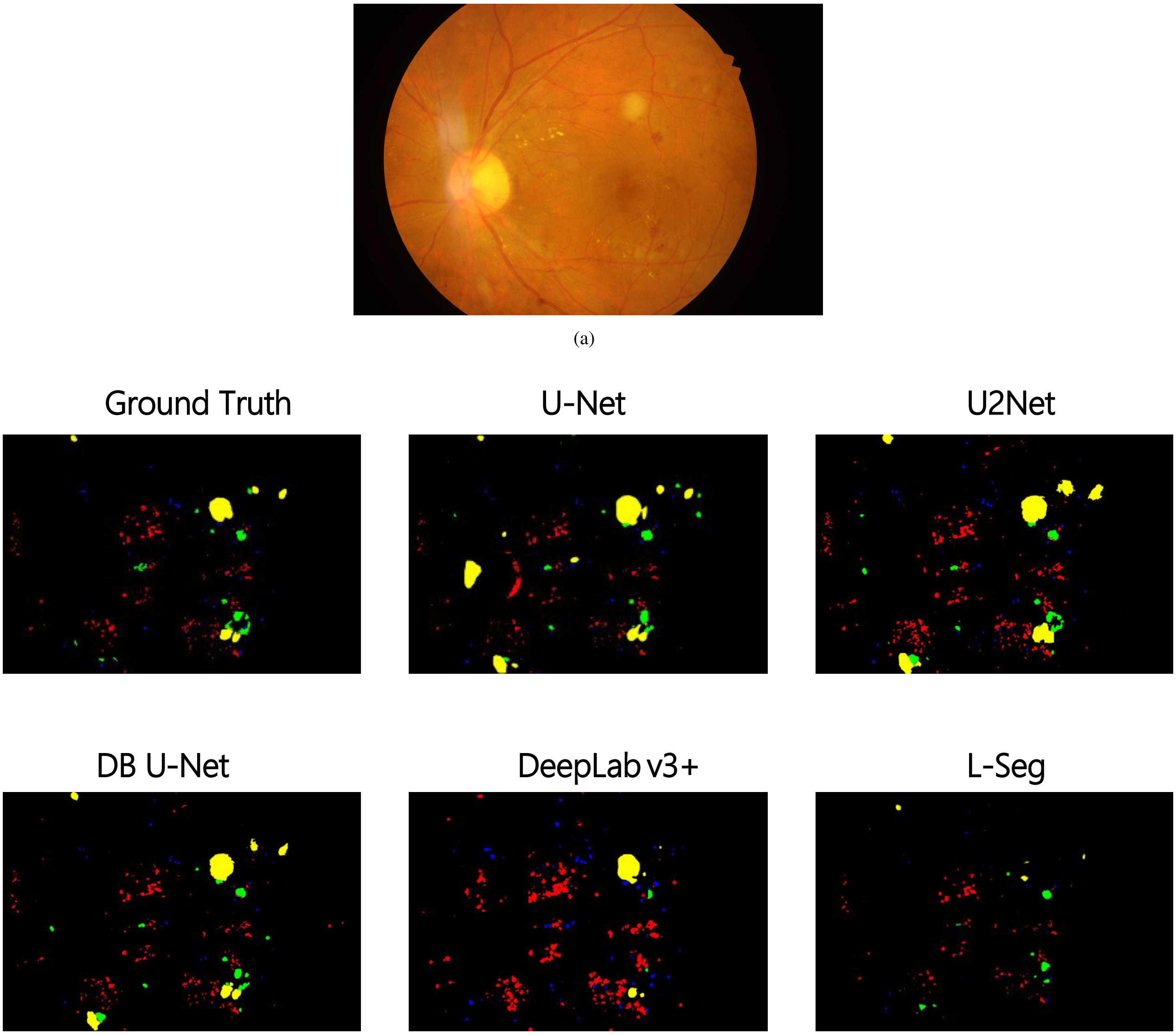
1) The external decoder was employed in our ablation studies. We used the conventional U-Net as a baseline model and utilized the U2Net's output as a segmentation result to assess lesion segmentation performance. When compared to the U-Net, the external decoder, specifically designed to learn boundary features, exhibited a significant enhancement in AUPR for lesions, except for SEs. This observation underscores the valuable contribution of auxiliary information in enhancing the network's performance, particularly in the context of small-scaled lesion areas.

TABLE 3. Performance of various models on IDRiD dataset. (best AUPR value are shown in bold).

Model	Lesion	AUC	AUPR	Acc	F1
U-Net	MA	0.8920	0.2844	0.9984	0.3466
	HE	0.9379	0.5433	0.9895	0.5311
	EX	0.9573	0.7973	0.9929	0.7356
	SE	0.9467	0.594	0.9985	0.5986
U2Net	MA	0.9234	0.4773	0.9987	0.4882
	HE	0.9538	0.5986	0.9905	0.5789
	EX	0.9951	0.8734	0.9941	0.7931
	SE	0.9073	0.5443	0.9986	0.4457
U-Net+U2Net	MA	0.9834	0.5153	0.9987	0.5172
	HE	0.9438	0.621	0.9905	0.5926
	EX	0.952	0.8545	0.9938	0.7872
	SE	0.972	0.5119	0.9985	0.5786
Proposed Method	MA	0.9870	**0.5254**	0.9988	0.5243
	HE	0.9642	**0.6609**	0.9913	0.6356
	EX	0.9944	**0.8878**	0.9946	0.8114
	SE	0.9540	**0.7297**	0.9989	0.6824

2) Incorporating the dual-branch architecture into our research framework, we labeled the final segmentation output as DB U-Net to evaluate its performance. In Table 3, it is evident that DB U-Net yielded the highest AUPR value compared to all the other methods under consideration. The segmentation results of DB U-Net, as illustrated in Table 4, effectively distinguished similar abnormal regions while maintaining a high sensitivity to lesions. This architectural approach, featuring dual branches and a fusion module, improves performance in balancing detection and classification tasks. However, when examining the results in 4, we noticed that U2Net achieved the highest AUPR, whereas DB U-Net outperformed others in the F1 score. This indicates that our model excelled at a specific threshold value for achieving the best results. Nevertheless, it's important to note that DB U-Net exhibited some instability when subjected to various factors, such as changes in illumination, resolution, or lesion size. This instability may arise due to the branch responsible for red/bright lesion segmentation, which introduces information about similar lesions. Furthermore, in cases of low resolution, the fusion module may struggle to differentiate the target lesion from the fused feature map.

3) The highlight of our model output incorporates the Fusion module, which plays a pivotal role in improving the performance of DB U-Net, our proposed method. The Fusion module demonstrates its importance by exhibiting improved network performance over alternative methods. Essentially, this module helps in effectively detecting diseases in input data. The Fusion module is an essential component that merges information from different branches or sources within the network, enabling a holistic understanding of the data. By integrating this module into the architecture, DB U-Net gains the ability to extract and utilize valuable information from multiple sources, leading to more accurate and robust disease detection capabilities.


(a)

(b)

FIGURE 4. Figure (a) displays the original color fundus image. Figure (b) provides a visual representation of segmentation maps, presenting the outcomes of lesion segmentation on the IDRiD dataset. In this representation, microaneurysms (MAs) are denoted in blue, hard exudates (HEs) in green, hemorrhages (EXs) in red, and soft exudates (SEs) in yellow. This illustration highlights the results of segmenting four distinct types of retinal lesions using our proposed deep learning model, organized from left to right: MAs, HEs, SEs, and EXs within fundus images. The uppermost row of images represents the ground truth (GT) reference. Similarly, the second, third, and fourth rows correspond to lesion segmentation achieved by the first stage (U-Net), the second stage (U2Net), and the third stage (referred to as Fusion, employing the DB U-Net model), respectively. The grayscale value assigned to each pixel reflects the probability of the presence of a lesion.

Essentially, the Fusion module serves as a hub in our model, allowing it to connect the collective power of its components and deliver superior performance in disease detection tasks. Its role in information fusion is instrumental in the overall success of the DB U-Net model.

B. COMPARATIVE ANALYSIS:PERFORMANCE ON THE PUBLIC E_OPHTHA_EX DATASET

Our evaluation on the publicly available e_ophtha_EX dataset reveals that our proposed method outperforms other state-of-the-art techniques in various aspects as shown in Table 6. Specifically, our method shows significant improvements in sensitivity (3.06% to 9.72%), precision (1.06% to 5.61%), and F1-score (2.02% to 8.08%) compared to recent studies. While competitive with a leading method by Zheng et al. regarding specificity and accuracy, there is a slight gap in sensitivity, precision, and F1-score. Our method demonstrates superior performance in critical metrics, making it a strong contender in medical image analysis applications, albeit with some nuanced differences compared to the top-performing alternative. We conducted a computation time analysis, revealing that our proposed method achieved a swift 0.141second computation time, whereas no other method in the study reported their timing.

TABLE 4. Performance of various models on DDR dataset.

Model	Lesion	AUC	AUPR	Acc	F1
U-Net	MA	0.9725	0.2211	0.9995	0.2821
	HE	0.8268	0.4722	0.9934	0.4586
	EX	0.9452	0.5775	0.9962	0.5673
	SE	0.9257	0.1482	0.9990	0.2102
U2Net	MA	0.9449	**0.2252**	0.9995	0.2702
	HE	0.8555	0.4590	0.9936	0.4619
	EX	0.9175	0.5629	0.9962	0.5557
	SE	0.9139	**0.2410**	0.9982	0.1397
U-Net+U2Net	MA	0.9568	0.1647	0.9995	0.2785
	HE	0.945	0.328	0.9921	0.3742
	EX	0.9434	0.5878	0.996	0.5665
	SE	0.9782	0.102	0.9991	0.1751
Proposed Method	MA	0.9678	0.1918	0.9996	0.2828
	HE	0.9489	**0.4952**	0.9937	0.4991
	EX	0.9586	**0.5844**	0.9956	0.5647
	SE	0.9665	0.2345	0.999	0.2122

TABLE 5. Performance of various models on E-Ophtha(MA) dataset.

Model	Lesion	AUC	AUPR	Acc	F1
U-Net	MA	0.9573	0.4188	0.9998	0.4524
U2Net	MA	0.9778	0.4252	0.9998	0.4204
U-Net+U2Net	MA	0.9567	0.4222	0.9998	0.4526
Proposed Method	MA	0.9408	**0.4382**	0.9998	**0.4626**

C. COMPARATIVE ANALYSIS WITH TOP 10 IDRID LESION SEGMENTATION TEAMS

In this section, we employed AUPR (Area under Precision-Recall curve) as the evaluation metric, aligning with the criteria used in the IDRiD challenge. The IDRiD challenge, hosted by the IEEE International Symposium on Biomedical Imaging (ISBI) conference, focuses on analyzing fundus images. To gauge the effectiveness of our method, we conducted a comparative analysis against the top 10 teams participating in the lesion segmentation competition of the IDRiD challenge. As illustrated in Table 7, our proposed approach secured the top position in microaneurysms (MA) segmentation, ranked second in hemorrhage (HE) segmentation, and achieved first place in both hard exudate and soft exudate segmentation. Worth noting is that the top-performing teams in the challenge adopted different network architectures for each specific segmentation task. Additionally, they encountered the complexity of fine-tuning numerous hyper parameters during the training phase. Consequently, these high-performing teams were obligated to test four distinct models for each corresponding segmentation task during the evaluation phase. In contrast, our study adopted a single unified network architecture, requiring only minor adjustments to the hyperparameter settings. Despite this streamlined approach, our proposed method is able to achieve results that are on par with the performance of the top-performing teams.

D. OVERALL COMPARATIVE ANALYSIS

In Table 8 and Table 9, we compared against published state-of-the-art methods IDRiD, DDR, E-Ophtha (MA) datasets. In IDRiD, we compare our framework with other published deep learning methods: L-seg [26], Local-Global U-Net [25], Multi-scale Net [22], Deeplab v3+, and their AUPR score are summarized in Table 8. We can observe that Local-Global U-Net performed well on HEs, EXs segmentation. The Local-Global U-Net is an efficient network that combines local details and the global context by integrating the decoder parts of a global level U-Net and a patch-level one. It resulted in an AUPR value of 0.711 and 0.889 higher than that of the other model. Similarly, Multi-scale Net also introduces multi-scales information which uses multi-scale input with embedding triplet loss. The triple loss minimizes the distance between the lesion patches while increasing the distance between the lesion patch and the healthy one. Multi-scale Net report an AUPR value of 0.4196.

DeepLab v3+ [18] extends DeepLabv3 [44] by adding an effective decoder module and utiles the depthwise separable convolution to both ASPP and decoder modules. In our experiments, it takes the ResNet-101 model as a backbone network. DeepLab v3+ shows poor performance in HEs and SEs segmentation on IDRiD. From Table 10, we observed that there is no obvious gap compared with other models on DDR. Its performance on different datasets may be limited by the scale of training data. As mentioned before, DDR is the largest dataset and has enough data for training while IDRiD provided less data on HEs and SEs segmentation. Table 10 offers a comparative analysis of various segmentation methods using AUPR (Area Under the Precision-Recall Curve) values within the DDR dataset. Overall, U-Net and ResUNet deliver moderate performance, with ResUNet excelling in HE and SE segmentation. DenseUNet stands out with strong MA, HE, and SE segmentation results, while UNet++ showcases superior performance in EX and SE segmentation. Att-UNet exhibits consistent but not outstanding results across all metrics, while PSPNet falls behind with lower AUPR values. DeepLab v3+ also registers relatively low AUPR values, indicating suboptimal performance within the DDR dataset. L-Seg is referenced but lacks specific AUPR values. In contrast, EfficientNet-B0+SAA displays impressive results, especially in EX and SE segmentation. Dual PSPNet+DSM lacks a specific AUPR value in the Table 10, making its performance unclear. The proposed method emerges as a strong contender, particularly excelling in HE segmentation, albeit with the highest parameter count among all methods.

IFLYTEK-MIG and VRT are the teams in IDRiD competition. They resulted in a AUPR value of 0.5017,0.4951 and ranked No.1 and 2 on the MAs segmentation task of the competition, respectively. IFLYTEK-MIG proposed a cascaded CNN-based approach with U-Net containing three stages: a coarse segmentation model, a cascade classifier, and a fine segmentation model. VRT modified U-Net, the upsampling layers of which have the same number of feature maps with layers concatenated, and they adjusted the number of downsampling layers according to the type of lesion.

L-Seg is an end-to-end multi-lesion segmentation model with a multi-scale feature fusion method and proposes a novel multi-channel bin loss to handle the cases of both class-imbalance and loss-imbalance problems. In Table 8,

TABLE 6. Evaluation of Exudate detection on e_ophtha_EX dataset.

Model	Lesion based results					
	SE	SP	PR	ACC	F1	Time
U-NET [38]	79.86	99.97	78.77	99.95	79.31	–
∗Playout et al. [5]	80.02	—	78.50	—	79.25	–
∗Zheng et al. [24]	94.12	99.98	91.25	99.96	92.66	–
Fraz et al. [39]	81.20	94.60	90.91	89.25	—	–
Zhang et al. [40]	74	—	72	—	—	–
Imani and Pourreza [41]	80.32	99.83	77.28	—	—	–
Javidi et al. [42]	80.51	99.84	77.30	—	—	–
Guo et al. [26]	84.17	—	83.45	—	83.81	–
∗EAD-Net [43]	92.77	99.98	89.06	99.97	90.87	–
∗proposed Method	94.89	99.99	92.12	99.98	92.89	0.141s

SE: sensitivity; SP: specificity; PR: precision; ACC: accuracy; F1: F1 score.
∗ are methods based on U-net.

TABLE 7. Comparative analysis with top 10 IDRiD lesion segmentation teams.

Model (team)	MAs	HEs	Hard exudates	Soft exudates
VRT (1st)	0.4951	**0.6804**	0.7127	0.6995
PATech (2nd)	0.4740	0.6490	0.8850	—
iFLYTEK-MIG (3rd)	0.5017	0.5588	0.8741	0.6588
SOONER (4th)	0.4003	0.5395	0.7390	0.5369
SHAIST (5th)	—	—	0.8582	—
lzyuncc_fusion (6th)	—	—	0.8202	0.6259
SDNU (7th)	0.4111	0.4572	0.5018	0.5374
CIL (8th)	0.3920	0.4886	0.7554	0.5024
MedLabs (9th)	0.3397	0.3705	0.7863	0.2637
AIMIA (10th)	0.3792	0.3283	0.7662	0.2733
Proposed Method	**0.5254**	0.6609	**0.8878**	**0.7297**

The results are based on AUPR (Area under Precision-Recall curve).

TABLE 8. AUPR value of other published methods on IDRiD dataset.The result of iFLYTEK-MIG* and VRT* are borrowed from the Leaderboard of the IDRiD Challenge.

Model	MA	HE	EX	SE
DeepLab v3+ [18]	0.4544	0.4913	0.8442	0.2826
Multi-Scale Net [22]	0.4196	-	-	-
Local-Global U-Net [25]	0.525	**0.711**	**0.889**	0.720
iFLYTEK-MIG* [15]	0.5017	0.5588	0.8741	0.6588
VRT* [15]	0.4951	0.6804	0.7127	0.6995
L-Seg [26]	0.4627	0.6374	0.7945	0.7113
EfficientNet-B0+SAA [27]	0.4152	0.6704	0.8812	0.7281
Dual PSPNet+DSM [28]	-	-	0.7767	-
Proposed Method	**0.5254**	0.6609	0.8878	**0.7297**

TABLE 9. Performance of other published methods on E-Ophtha-MA dataset.

Model	AUC	AUPR	Acc	F1
L-Seg [26]	-	0.1687	-	-
DeepLab v3+ [18]	0.8630	0.3492	0.9998	0.4204
Proposed Method	0.9408	**0.4382**	0.9998	0.4626

L-Seg ranked No. 3 on SE segmentation, No. 4 on HE segmentation. In Table 8 and Table 10, L-Seg is the only end-to-end unified framework that generates multi-lesion segmentation results and shows competitive performance compared with DeepLab v3+, U2Net, and DB U-Net on DDR dataset.

There are various evaluation metrics with the current state-of-the-art methods (e.g., AUC, AUPR, F1). The IDRiD grand challenge provided a great opportunity to compare our performance standardized metrics. From Table 8, we observe that DB U-Net achieved the best performance of 0.5254 and 0.7297 on MAs and SE segmentation and ranked No.3 and No.2 on HE and EX segmentation respectively. As shown in Table 10 and Table 9, DB U-Net achieved the highest AUPR value on MAs, HEs, and EXs segmentation on DDR and E-ophtha MA. The improvement in fundus lesions segmentation shows the capability of effectively handling both data imbalance problems and lesion segmentation under the complex background.

VI. DISCUSSION

For lesions segmentation, several modified architectures, as well as effective methods, have been employed. In our work, we attempt to explore the dual-branch architecture and auxiliary information improves the performance of our proposal in terms of precision and sensitivity. There are still some defects.

Our experiments show that the aforementioned methods lead to a highly effective architecture that significantly boosts performance on lesion segmentation, especially scatter and smaller objects. However, our proposal fails to distinguish target lesions from the background or the other lesions that belong to the same group (see Figure 4). This indicates that our work is short of capturing the global context. Introducing multi-scale information may optimize this issue. Just as [25], a segmentation framework integrates the decoder parts of a global-level U-net and a patch-level.

As mentioned before, MAs are the earliest clinical signs of DR but the ratio of object pixels to background pixels is approximately 0.10% on IDRiD. In the DDR and E-ophtha datasets, the ratio is 0.02%, 0.01%, respectively. Due to PR curve is deployed to evaluate models instead of ROC curves, the misclassified pixels have an enormous impact on AUPR value while these are of no consequence to DR screening in medical practice. A question that the lack of standardized evaluation metrics naturally arises.

VII. CONCLUSION

As a chronic eye disease, the timely treatment is of great significance and prospect in terms of the patients with diabetic retinopathy. Based on deep learning, computer-aided diagnostic technology plays an important role in disease screening. In this paper, we propose a network with dual-

TABLE 10. AUPR value of other published methods on DDR dataset.

Model	MA	HE	EX	SE	Param
U-NET [38]	0.4586	0.0970	0.0955	0.1094	7.86
ResUNet [48]	0.4042	0.1532	0.0904	0.869	7.25
DenseUNet [49]	0.5459	0.3732	0.1804	0.2434	6.58
UNet++ [50]	0.4527	0.1205	0.1154	0.794	7.95
Att-UNet [51]	0.4626	0.1130	0.0894	0.1273	7.96
PSPNet [52]	0.3497	0.2118	0.557	0.0909	7.89
L-Seg [26]	0.1052	0.3586	0.5546	0.2648	–
DeepLab v3+ [18]	0.0843	0.3529	0.5668	0.2302	7.67
EfficientNet-B0+SAA [27]	**0.1933**	0.4456	**0.6269**	**0.3747**	
Dual PSPNet+DSM [28]	-	-	0.5424	-	-
Proposed Method	0.1918	**0.4952**	0.5825	0.2345	8.45

branch architecture to improve the segmentation of scattered and small lesions in fundus images. We introduce edge information and parallel architecture to address the issue of segmenting various size lesions. We evaluated our work on public datasets and obtained competitive performance, which demonstrates that the efficiency of our proposal network. However, the fusion module is constricted by the quality of the segmentation map from branches. We found that optimizing the feature space through auxiliary information helps the model focus on small region. Furthermore, the structure of double branches can compare the prediction of the network on the two branches, and regard the areas with differences as easily confused areas. The purpose of double branches is to mine the complementary information of features and obtain better feature representation to improve the final segmentation performance.

There are still many problems of lesion segmentation being to be solved. More researches are necessary to further explore the practical application of the automatic diabetic retinopathy diagnosis system. For example, medical related tasks are more difficult to obtain labels. Patient privacy, professional labeling and other factors limit the scale of the dataset. Under the constraints, weak-supervised learning, semi-supervised learning, few-shot learning and even zero-shot learning can reduce the dependence of the model on data. How to use less data to get a better and more robust model is an expected research.

ACKNOWLEDGMENT

The authors acknowledge the support from the Deanship of Scientific Research, Najran University. Kingdom of Saudi Arabia, for funding this work under the Research Groups funding program grant code number (NU/RG/SERC/12/15). This research was also supported by the National Natural Science Foundation of China (No. 62376101) to manage lab resources.

REFERENCES

[1] *International Diabetes Federation (IDF) Diabetes Atlas*, 10th ed. Brussels, Belgium, 2021. [Online]. Available: https://www.diabetesatlas.org

[2] G. L. Ong, L. G. Ripley, R. S. Newsom, M. Cooper, and A. G. Casswell, "Screening for sight-threatening diabetic retinopathy: Comparison of fundus photography with automated color contrast threshold test," *Amer. J. Ophthalmol.*, vol. 137, no. 3, pp. 445–452, Mar. 2004.

[3] N. Salamat, M. M. S. Missen, and A. Rashid, "Diabetic retinopathy techniques in retinal images: A review," *Artif. Intell. Med.*, vol. 97, pp. 168–188, Jun. 2019.

[4] V. G. Edupuganti, A. Chawla, and A. Kale, "Automatic optic disk and cup segmentation of fundus images using deep learning," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2227–2231.

[5] C. Playout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2434–2444, Oct. 2019.

[6] M. Monemian and H. Rabbani, "Detecting red-lesions from retinal fundus images using unique morphological features," *Sci. Rep.*, vol. 13, no. 1, pp. 1–11, Mar. 2023.

[7] C. Santos, M. Aguiar, D. Welfer, and B. Belloni, "A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model," *Sensors*, vol. 22, no. 17, p. 6441, Aug. 2022.

[8] A. Sebastian, O. Elharrouss, S. Al-Maadeed, and N. Almaadeed, "A survey on diabetic retinopathy lesion detection and segmentation," *Appl. Sci.*, vol. 13, no. 8, p. 5111, Apr. 2023.

[9] 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.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5228–5237.

[12] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9716–9725.

[13] V. John, M. K. Nithilan, S. Mita, H. Tehrani, R. S. Sudheesh, and P. P. Lalu, "SO-Net: Joint semantic segmentation and obstacle detection using deep fusion of monocular camera and radar," in *Image and Video Technology*, J. J. Dabrowski, A. Rahman, and M. Paul, Eds. Cham, Switzerland: Springer, 2020, pp. 138–148.

[14] V. John, A. Boyali, S. Thompson, and S. Mita, "BVTNet: Multi-label multi-class fusion of visible and thermal camera for free space and pedestrian segmentation," in *Proc. Pattern Recognit., ICPR Int. Workshops Challenges*. Germany: Springer-Verlag, 2021, pp. 277–288.

[15] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, L. Xinhui, L. Gao, T. Wu, J. Xiao, F. Wang, B. Yin, Y. Wang, G. Danala, L. He, Y. Choi, Y. C. Lee, and F. Meriaudeau, "IDRiD: Diabetic retinopathy—Segmentation and grading challenge," *Med. Image Anal.*, vol. 59, Oct. 2019, Art. no. 101561.

[16] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, and A. Chabouis, "TeleOphta: Machine learning and image processing methods for teleophthalmology," *IRBM*, vol. 34, no. 2, pp. 196–203, Apr. 2013.

[17] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Inf. Sci.*, vol. 501, pp. 511–522, Oct. 2019.

[18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Computer Vision—ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 833–851.

[19] M. Haloi, "Improved microaneurysm detection using deep neural networks," 2015, *arXiv:1505.04424*.

[20] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri, and A. Hunter, ''Microaneurysm detection using fully convolutional neural networks,'' *Comput. Methods Programs Biomed.*, vol. 158, pp. 185–192, May 2018.

[21] C. Kou, W. Li, W. Liang, Z. Yu, and J. Hao, ''Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network,'' *J. Med. Imag.*, vol. 6, no. 2, p. 1, Jun. 2019.

[22] M. H. Sarhan, S. Albarqouni, M. Yigitsoy, and N. Navab, ''Multi-scale microaneurysms segmentation using embedding triplet loss,'' in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Cham, Switzerland: Springer, 2019, pp. 174–182.

[23] 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.

[24] R. Zheng, L. Liu, S. Zhang, C. Zheng, F. Bunyak, R. Xu, B. Li, and M. Sun, ''Detection of exudates in fundus photographs with imbalanced learning using conditional generative adversarial network,'' *Biomed. Opt. Exp.*, vol. 9, no. 10, p. 4863, 2018.

[25] Z. Yan, X. Han, C. Wang, Y. Qiu, Z. Xiong, and S. Cui, ''Learning mutually local–global U-Nets for high-resolution retinal lesion segmentation in fundus images,'' in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 597–600.

[26] S. Guo, T. Li, H. Kang, N. Li, Y. Zhang, and K. Wang, ''L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images,'' *Neurocomputing*, vol. 349, pp. 52–63, Jul. 2019.

[27] W. Bo, T. Li, X. Liu, and K. Wang, ''SAA: Scale-aware attention block for multi-lesion segmentation of fundus images,'' in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5.

[28] Q. Liu, H. Liu, Y. Zhao, and Y. Liang, ''Dual-branch network with dual-sampling modulated dice loss for hard exudate segmentation in color fundus images,'' *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 1091–1102, Mar. 2022.

[29] A. Chaurasia and E. Culurciello, ''LinkNet: Exploiting encoder representations for efficient semantic segmentation,'' in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

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

[31] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, ''Domain generalization: A survey,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4396–4415, Apr. 2023.

[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[33] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[34] A. Dyanaa and P. V. Smruthi, ''Detection using dynamic shape features red lesion for diabetic retinopathy screening,'' *Int. J. Comput. Trends Technol.*, vol. 42, no. 1, pp. 1–6, Dec. 2016.

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

[36] T. Saito and M. Rehmsmeier, ''The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,'' *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432.

[37] K. Boyd, K. H. Eng, and C. D. Page, ''Erratum: Area under the precision-recall curve: Point estimates and confidence intervals,'' in *Machine Learning and Knowledge Discovery in Databases*, H. Blockeel, K. Kersting, S. Nijssen, and F. Železný, Eds. Berlin, Germany: Springer, 2013.

[38] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.

[39] M. M. Fraz, W. Jahangir, S. Zahid, M. M. Hamayun, and S. A. Barman, ''Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification,'' *Biomed. Signal Process. Control*, vol. 35, pp. 50–62, May 2017.

[40] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, and A. Erginay, ''Exudate detection in color retinal images for mass screening of diabetic retinopathy,'' *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, Oct. 2014.

[41] E. Imani and H.-R. Pourreza, ''A novel method for retinal exudate segmentation using signal separation algorithm,'' *Comput. Methods Programs Biomed.*, vol. 133, pp. 195–205, Sep. 2016.

[42] M. Javidi, A. Harati, and H. Pourreza, ''Retinal image assessment using bi-level adaptive morphological component analysis,'' *Artif. Intell. Med.*, vol. 99, Aug. 2019, Art. no. 101702.

[43] C. Wan, Y. Chen, H. Li, B. Zheng, N. Chen, W. Yang, C. Wang, and Y. Li, ''EAD-Net: A novel lesion segmentation method in diabetic retinopathy using neural networks,'' *Disease Markers*, vol. 2021, pp. 1–13, Sep. 2021.

[44] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking Atrous convolution for semantic image segmentation,'' 2017, *arXiv:1706.05587*.

[45] Z. Zhang, Q. Liu, and Y. Wang, ''Road extraction by deep residual U-Net,'' *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[46] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, ''H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes,'' *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[47] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, ''UNet++: A nested U-Net architecture for medical image segmentation,'' in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science), vol. 11045, D. Stoyanov et al., Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-00889-5_1.

[48] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, ''Attention U-Net: Learning where to look for the pancreas,'' 2018, *arXiv:1804.03999*.

[49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ''Pyramid scene parsing network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

MING YIN (Member, IEEE) received the Ph.D. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006. From 2012 to 2012, he was a Visiting Scholar with the School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW, Australia. From 2006 to 2022, he was with the School of Automation, Guangdong University of Technology, Guangzhou, China. Currently, he is a Professor with the School of Semiconductor Science and Technology, South China Normal University, Guangzhou. His research interests include computer vision, pattern recognition, and machine learning.

TOUFIQUE AHMED SOOMRO (Senior Member, IEEE) received the Ph.D. degree in AI and image processing from Charles Sturt University, Australia, in 2018. He is currently an accomplished academic with expertise in engineering and artificial intelligence (AI). He is also an Associate Professor and the Head of the Department of Electronic Engineering, QUEST, Larkano Campus, Pakistan. His research interests include image enhancement, segmentation, classification, and analysis for medical images. He has made significant contributions to the field, with 61 research publications in AI for medical imaging. He has collaborated on projects related to AI for biomedical imaging with the Ministry of Education of Saudi Arabia and Najran University. Additionally, he was honored as a Young Research Professor with the University of Technology, Guangdong, China, in 2019 and 2020. With his extensive knowledge and research expertise, he continues to advance the field of AI and image processing. As an Educator, he inspires students and contributes to the growth of electronic engineering. His accomplishments demonstrate his commitment to research excellence and his contributions to the scientific community.

FAYYAZ ALI JANDAN (Graduate Student Member, IEEE) is currently with the Electrical Engineering Department, Quaid-e-Awam University of Engineering, Science & Technology, Larkana Campus, Pakistan.

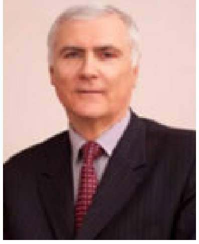
AYOUB FATIHI received the Dipl.-Ing. degree in geomatics and surveying engineering from IAVH2, Morocco. He is currently pursuing the Ph.D. degree with the Institute of Earth Sciences, Faculty of Geosciences and Environment, University of Lausanne (UNIL). His research interests include deep learning, computer vision, remote sensing, and medical imaging.

FAISAL BIN UBAID received the Ph.D. degree in computer science from Chongqing University, Chongqing, China. He is currently an Assistant Professor with the Computer Science Department, Sukkur IBA University, Sukkur, Pakistan. His research interests include artificial intelligence, machine learning, and software defined networks. Throughout his academic and professional career, he has actively contributed to the advancement of research in these areas and has published several research papers in reputable conferences and journals. He is committed to fostering innovative solutions and empowering the next generation of computer scientists through his teaching and research endeavors.

MUHAMMAD IRFAN received the Ph.D. degree in electrical and electronic engineering from Universiti Teknologi PETRONAS, Malaysia, in 2016. He has two years of industry experience (October 2009–October 2011) and six years of academic experience (January 2017) in teaching and research. Currently, he is an Associate Professor with the Electrical Engineering Department, Najran University, Saudi Arabia. He has authored more than 200 research articles in reputed journals, books, and conference proceedings (Google Scholar Citations 2900 and H-index 24). His main research interests include automation and process control, energy efficiency, condition monitoring, vibration analysis, artificial intelligence, the Internet of Things (IoT), big data analytics, smart cities, and smart healthcare.

AHMED J. AFIFI received the bachelor's and M.Sc. degrees in computer engineering from the Islamic University of Gaza (IUG), in 2008 and 2011, respectively, and the Ph.D. (Dr.-Ing.) degree from Technische Universität Berlin, Germany, in 2021. Currently, he is a Postdoctoral Researcher with the Helmholtz Institute Freiberg for Resource Technology (HIF) and the Karlsruhe Institute of Technology (KIT), focusing on 3D point cloud classification and segmentation. His research interests include computer vision, deep learning, 3D object reconstruction from a single image, and medical image analysis.

SAIFUR RAHMAN received the Ph.D. degree in electrical and electronic engineering. He has more than 12 years of academic experience in teaching and research. Currently, he is an Associate Professor with the College of Engineering, Najran University, Saudi Arabia. He has authored more than 70 research articles in reputed journals, books, and conference proceedings. His main research interests include artificial intelligence, the Internet of Things (IoT), big data analytics, smart cities, and smart healthcare.



SERGII TELENYK received the M.Sc. (Eng.), Ph.D., and D.Sc. (Higher Doctorate) degrees in information technologies from the Igor Sikorsky Kyiv Polytechnic Institute, Ukraine, in 1975, 1982, and 2000, respectively. He is currently a Professor with the Cracow University of Technology, where he has been involved in scientific and educational activities, since 2016. Additionally, he is also a Professor with the Igor Sikorsky Kyiv Polytechnic Institute. His research interests include IT infrastructure management, software engineering, mathematical logic, computational linguistics, artificial intelligence, control theory, and information systems design. He particularly focuses on methodologies related to the support of the life cycle of services in information systems, the design and implementation of services, the preparation and provision of services, and the development of services.

GRZEGORZ NOWAKOWSKI is currently an Assistant Professor in computer science with the Department of Automatics and Informatics, Cracow University of Technology, Poland. His scientific research interests include information systems, computational intelligence, databases, big data, cloud computing, and soft computing methods. He focuses on fuzzy database queries and their application in big data analysis.