DAB-UNET: Dual Attention Block UNET Segmentation for Diabetic Retinopathy Utilizing an Encoder-Decoder Residual

Haithem Kareem Abass[1], Mohammed Al-Mukhtar[2], and Ammar S. Al-Zubaidi[2,*]

[1] Computer Engineering Department, Al-Mansour University College, Baghdad, Iraq
[2] Computer Centre, University of Baghdad, Baghdad, Iraq
Email: Haithem.kareem@muc.edu.iq (H.K.A.); Mohammed.abdul@cc.uobaghdad.edu.iq (M.A.-M.); Ammar.sabah@ccc.uobaghdad.edu.iq (A.S.A.-Z.)
*Corresponding author

***Abstract*—Fundus images play an essential role in ophthalmic diagnostics for the detection of many eye illnesses. The experiment begins with a thorough image pre-processing technique, which includes clipping the circular borders, scaling the image, enhancing the contrast, removing noise, and augmenting the data. The new combined block applies to extracting distinctive deep feature representations, which help to detect the first shape of the edges of each lesion. It is namely the Attention Block and the Conv-Deconv UNET model. Attention Block is subsequently implemented in order to augment the robustness and quality of feature depictions derived from a pair of DR images. The Dual Attention Block for the backbone, which is supplemented with hierarchical bottleneck attention, is what we propose here referred to as Dual Attention Block UNET (DAB-UNET). Bottleneck Attention Blocks and Dual Attention Blocks greatly improve a model's ability to concentrate on essential features, boosting its performance in complex tasks such as image segmentation. When these attention mechanisms are built into architectures like DAB-UNET, they make the network faster and more accurate, letting it pick up on small, specific details. This is particularly beneficial in areas like medical imaging, where high precision is essential. In order to emphasize retinal anomalies that are significant for fovea macula and Diabetic Retinopathy (DR) semantic segmentation in the deteriorated retina, the network is made up of a unique bottleneck attention block. We trained Mask-Region based Convoluting Neural Network (RCNN) model that comprises of a backbone for eliminating Oculus Dexter (OD) regions. Moreover, the proposed block combines self-attention with channel attention in order to highlight these abnormalities. Our results indicate that DAB-UNET is potentially very effective for identifying landmarks even when dealing with different types of retinal degenerative disorders.**

***Keywords*—Dual Attention Block UNET (DAB-UNET), UNET network, ResNet model, diabetic retinopathy, attention block**

I. Introduction

The condition known as diabetic retinopathy is one of the consequences that may arise as a result of diabetes. This condition is caused by damage to the blood vessels of the retina, which is the light-sensitive tissue that is situated in the lower back of the eye. The third greatest cause of blindness in the United States is glaucoma, which is defined by progressive damage to the optic nerve and the subsequent loss of visual field because of this damage [1]. Diabetic Retinopathy is a condition that is associated with diabetes that specifically impacts the blood vessels in the retina. On the other hand, Glaucoma is largely caused by elevated pressure inside the eye, which in turn damages the optic nerve.

Glaucoma is a chronic optic neuropathy that occurs with age and is the primary cause of permanent blindness on a global scale. According to other forecasts, it is projected that by 2040 [2], the global population of individuals with glaucoma would approach around 112 million, with a disproportionate impact on Asian and African nations. On the other hand, the growing prevalence of glaucoma, especially in developing countries, highlights the crucial need of effective glaucoma diagnosis and treatment.

Several techniques have shown that image segmentation holds great promise for computer vision. In particular, U-Net architecture which is based on fully convolutional networks is rather promising for image segmentation tasks. Using these two encoding-decoding structures, it captures both the local and global properties. Accordingly, a combination of low-level and high-level convolutional features is achieved by the use of skip connections in the U-Net architecture, which restricts its capacity to efficiently combine valuable features and use contextual information. Utilizing image segmentation methods offers creative solutions to intricate segmentation problems. The use of Convolutional Neural Networks (CNNs) has shown substantial advancements and confirmed efficacy in detecting intricate image patterns [3]. Image segmentation

Manuscript received October 15, 2024; revised December 4, 2024; accepted January 21, 2025; published May 20, 2025.

technology has been widely used in Medical Imaging Analysis (MIA) in the medical profession, resulting in improved efficiency and accuracy of physicians' diagnosis [4]. Fundus Fluorescein Angiography (FFA) is a crucial method for evaluating retinal disorders. To accomplish automated and uniform labeling of FFA images, it is advantageous to train on annotated FFA images using different Convolutional Neural Networks (CNN) [5]. There are a number of phases involved in the BT-Net Convolutional Neural Network (CNN) model that is used for the identification and quantification of a brain tumor. These steps include preprocessing, skull stripping, and tumor segmentation. [6]. Segmenting medical images, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans, enable more precise detection and measurement of different abnormalities. for instance, lung segmentation [7] and liver segmentation [8]. In addition, the technology of image segmentation is essential for the planning of surgical procedures and the guidance of intraoperative procedures. In spite of the significant progress that has been achieved in the field of image segmentation for medical applications, there are still challenges that need to be solved. Included in these challenges are the capacity to deal with difficult backgrounds, the enhancement of segmentation accuracy, and the further improvement of algorithm stability. These challenges should be taken into consideration. By using spatial attention processes, the network is able to selectively concentrate on significant regions in space. This enables the network to prioritize key characteristics in the skip connections while minimizing the impact of irrelevant data. Chen *et al.* [9] implemented spatial attention techniques in the decoding process due to the significant role of spatial information in image segmentation tasks.

Multi scale technique [10], matched filter [11], and mathematical morphology [12] are applied for vessel segmentation of fundus images. The purpose of these strategies was to arrive at conclusive forecasts by use of feature extractors that were created manually. For the purpose of distinguishing vasculature and lesions from one another, the MCA algorithm together with the relevant transformations, is used. Following that, the Morlet Wavelet Transform is used in order to improve the look of the retinal vessels. For the final vessel map, adaptive thresholding is the method that is used. It presents a modified deep Convolutional Neural Network (CNN) model that can accurately predict the stage of glaucoma (mild, moderate, or severe) [13]. Our model utilizes characteristics commonly used by ophthalmologists and validates its predictions by comparing them to variables derived via visualization by the CNN layers.

The segmentation of retinal vessels has shown substantial advancements because of the application of deep learning techniques [14, 15]. Especially, Deep Neural Networks (DNNs) are used in order to accomplish the necessary outcomes by using their remarkable capabilities in terms of automated feature learning and end-to-end learning. However Segmentation in retinal imaging is challenging due to the complexity of retinal lesions, which can manifest in a variety of ways, including different sizes, shapes, and colors. Various forms of these lesions exist, ranging from microscopic microaneurysms to bigger hemorrhages, exudates, and neovascular formations. Each possesses distinct characteristics that complicate accurate identification. It's also possible for the quality of the images to vary, which can make it challenging to see lesions clearly because of things like low resolution, low contrast, or artifacts.

A notable problem emerges from data annotation, essential for training segmentation algorithms. Annotating retinal pictures is labor-intensive, necessitates specialized expertise, and may be subjective, resulting in possible discrepancies among various annotators. The inconsistency of labeled data might adversely impact the training of segmentation algorithms, thereby diminishing their accuracy and generalizability.

These problems can have a big effect on how well segmentation algorithms work at finding and separating complications of Diabetic Retinopathy (DR). If the segmentation isn't done right, some lesions might be missed or put in the wrong category. This makes automated methods less useful for finding and diagnosing Diabetic Retinopathy (DR) early on. Accurate and consistent segmentation is crucial for enhancing therapeutic results and enabling timely assistance for those at risk of vision loss.

Both the inability to regain the information that was lost during the decoding phase and the loss of uninterrupted resolution that occurred during the encoding phase are primarily responsible for this. This paper introduces a retinal vascular segmentation network that is named FRD-Net. It is designed to be particularly effective in addressing this issue. Backbone network and Multi-scale Feature Fusion Module (MFFM) are the two primary components that apply FRD-Net at its core [16]. The primary contributions of this work may be stated as follows:

1. Combining with a median filter, we present an efficient image enhancement approach based on the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm to increase the contrast and decrease noise in fundus images. We contend that training deep learning models using pre-processed DR-Fundus image data instead of raw data directly will significantly improve their capability to learn more meaningful feature information. Moreover, this method may help to lower the computational complexity required to generate an optimally trained model.
2. novel Dual Attention Bottleneck (DAB) combines the block used self-attention [17] and channel attention [18], the purpose of highlighting retinal abnormalities that are critical for landmark identification in the retina that has deteriorated.
3. In order to accomplish the formation of UNET, we include the DAB block into bottleneck skip connections throughout all tiers of a U-NET backbone network.

The problem statement revolves on the fundus images, which provide a complex dataset because to several problematic characteristics. One such aspect is the presence of optic disk and hard exudates, which have similar pixel values. The complicated network of connecting blood arteries in the eye hampers the identification of tiny anomalies, resulting in reduced accuracy. Through the use of real-time monitoring capabilities, it becomes feasible to swiftly act, hence optimizing the results for patients and improving the treatment of illnesses. It is possible for healthcare organizations to lower their burden and prioritize patients who need immediate attention if they maximize the use of medical resources and improve illness detection systems. They are able to adopt a comprehensive approach to the management of diabetic retinopathy as a result of this. Our proposed model segment optic disc and extract vessel blood for increase the accurate detection.

The aim of our proposed model is to develop a system that is capable of processing images and reliably diagnosing the amount of diabetic retinopathy. This is the goal of the model that has been developed.

Overall, the advancements that have been made in medical imaging and diagnostic processes have the potential to bring about substantial changes, especially in the case of diseases such as diabetic retinopathy. The ability to get very precise and comprehensive retinal images is what makes the fundus imaging system stand out among these technologies. Using a fundus camera to capture images of the retina, this diagnostic network would let doctors evaluate the illness and identify diabetic retinopathy.

Finally, the key objective of this study is to significantly enhance the diagnosis of diabetic retinopathy by creating innovative and effective methods for assessing the disease and providing optimal therapy to patients.

1. The aim is to enhance the accuracy of segmenting diabetic retinopathy stages by using pre-processing techniques such as transformation matrix, non-local mean denoising autoencoder, and image filtering.
2. in order to use data augmentation that is specific to each grade to correct the imbalance in the data.

The rest of the sections of this work are organized as follows: Section II provides an overview of the existing research in the field. Section III presents the planned network. The experimental setup is described in Section IV. Section V analyzes the empirical findings. Section VI addresses the experimental design, whereas Section VII provides the paper's conclusion.

II. Literature Review

Several automated vessel segmentation approaches based on deep learning have been suggested [19–21]. introduced a deep learning network named SegNet. This network utilizes an encoder-decoder architecture with max-pooling indices to efficiently learn features and reconstruct images. Additionally, to use transfer learning-based models to cope with limited annotated training resources, minimize training overhead, and automatically extract features.

In this part, we will provide a concise overview of the associated work that has been done in the past concerning the U-NET Network, the introduction of the attention mechanism, and vessels & optical disk segmentation.

A. U-NET Network

The U-NET architecture is a network based on Fully Convolutional Networks (FCNs). The topology of this system is similar to that of FCN, NUNET [22] and UNET++ [23], since it makes use of encoders and decoders, along with skip connections. The U-Net network is distinguished by its symmetrical network architecture, with an encoder on the left side to capture contextual information and a decoder on the right side to accurately locate and restore the feature map size. The output feature maps of the encoder-part are duplicated and cropped. These duplicated and cropped feature maps are then fused with the deconvolution feature maps of the decoder. The resulting fused feature maps are transmitted to the next layer for upsampling. During the process of upsampling in the U-Net network, a large number of feature channels are able to transmit contextual information to higher resolution levels. The multi-attention MA-UNET design utilizes a conventional encoder-decoder topology. This process involves the sequential application of convolutions and down-sampling at the encoder step [24]. As a consequence, the feature maps produced have reduced resolutions but still include compact, high-dimensional semantic representations. Afterwards, the decoder undertakes a continuous process of convolution and up-sampling to restore the segmentation result to its original size. In order to enhance the backbone's capacity to extract fine-grained features, it is suggested that a residual encoder be developed, which would be based on a simple attention module. The encoder-decoder architecture underpins the deep supervised network known as UNET++ [25]. A number of nested, dense skip connections are used by the encoder and decoder subnetworks to produce feature fusion. This reduction in semantic loss between the feature mappings is accomplished via the use of these connections. The combined model of UNet++ with FPN has achieved accurate results in [26].

B. Introduction of the Attention Mechanism

Introduce a network called Spatial Attention [27] U-Net (SA-UNet) that is designed to be lightweight and does not rely on a large number of annotated training examples. This network may also be used in a data augmentation approach to make better use of the existing annotated samples. Spatial attention makes it easier for the network to focus on important parts of an image or feature map. This lets the model put the most important information at the top of the list and lessen the impact of less important or irrelevant data. By learning to allocate more attention weights to essential spatial regions, the network can eliminate noise, background clutter, and other distractions that might impede the job at hand. This approach generates an attention map that highlights the most critical areas of the feature map relevant to the job, such as object detection, segmentation, or classification. We frequently employ a convolutional layer or other techniques to generate the attention map. We employ these strategies to discern spatial connections and highlight aspects essential for understanding the input. This enables the model to concentrate on the areas that significantly influence the decision-making process, hence enhancing overall performance.

Spatial attention enhances the network's ability to manage intricate spatial connections by concentrating on essential spatial regions. In medical imaging, spatial attention enables the model to concentrate on critical structures, such as tumors or lesions, which may be diminutive or irregularly shaped and easily overlooked without this concentration. In real pictures, spatial attention enables the model to focus on relevant items or areas while disregarding insignificant background components. Spatial attention not only diminishes noise but also facilitates the network's acquisition of more pertinent and distinctive characteristics; hence, it enhances the model's generalization capabilities. This makes the model more resistant to changes in input, such as changes in lighting, orientation, or scale, while still being able to focus on the most important parts.

Ultimately, spatial attention enhances the efficiency and precision of model input processing. This renders it an effective instrument for jobs requiring a comprehensive, context-aware comprehension of spatial relationships within an image or feature map. Spatial attention facilitates the model's ability to concentrate on pertinent regions and manage intricate spatial connections. This results in enhanced feature representation, improved performance, and more precise predictions across several domains, including computer vision and medical diagnostics

In Ref. [28], by paying attention to hierarchical bottlenecks, the U-Net architecture was revised, and its application for fundus analysis was shown. DAB-block is used for the segmentation of retinal landmarks. It is suggested that an end-to-end encoder-decoder network, known as DRNet [29], be used for the purpose of localizing OD and Fovea centers. In order to compensate for the spatial feature that is lost in the encoder as a result of pooling, we suggest the use of residual skip connection in our DRNet. The skip connection that has been suggested does not instantly concatenate low-level feature maps from the early layers of the encoder with the matching same scale decoder. Wang *et al.* [30] provided the Efficient Channel Attention (ECA) module as a potential solution. In order to improve the learning of effective channel attention, a unique local cross-channel interaction technique was presented. This strategy eliminated the dimensionality reduction. A dual attention network was presented in [31], which included both spatial attention and channel attention modules. The purpose of this network was to dynamically integrate local characteristics with their global dependencies.

C. Vessels & Optic Disk Segmentation

Retinopathy may result in significant fluctuations in the shape, color, and size of the optic disc. Muhammed [32] recommended an examination of local feature spectrums. During the training step, each candidate is recreated using a set of local characdteristics. The building of the dictionary employs a sparse selection method, while the classification utilizes the k-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms. An approach that is based on template matching is described. Gui *et al.* [33] contains this method. The use of adaptive template size design and the examination of blood vessel patterns that are present on the surface of the optic disc are components that are necessary for the identification of the optical disc. A phase that involves removing blood vessels through the use of Alternating Sequential Filtering (ASF) and removing bright areas through the use of morphological reconstruction is included into the procedure before the subsequent stage of segmenting the optic disc.

Converting the traditional two-dimensional searching space into a one-dimensional searching space, which would result in a more expedient localization process, was one of the suggestions made by Orlando *et al.* [34]. The encoding of the x and y coordinates, as well as the development of two projections of the image properties of the OD, are both components of the approach. A study was provided by Orlando *et al.* [35], in which they created a deep learning model for glaucoma, in which an unsupervised Convolutional Neural Network (CNN) retrieved the models' characteristics. Between normal and glaucomatous patients, a Cup-to-Disc Ratio (CDR) threshold of 0.5 was used to differentiate between the two. In Ref. [36], a system that is based on deep learning and is called RDCU-Net has been presented for the purpose of segmenting three sections of brain tumors. When compared to the traditional U-Net, this technique overcomes challenges such as direct feature fusion, poor accuracy of segmentation at area edges, and lower resolution. In this strategy, the Dilated Convolution (DC) block is responsible for fusing both high-level and low-level features together. This model achieves a better level of precision than the approaches that are now in use by using two different methodologies and a smaller number of parameters than those that are already in use. High-level and low-level characteristics are integrated using the first method, which involves the use of Dilated Convolution (DC) blocks that have proportionate conversion rates.

Li *et al.* [37] presented a unique approach for the segmentation of retinal vessels that makes use of the BCOSFIRE filter for the diagnosis of diabetic retinopathy and glaucomam. Contour enhancement, area of interest extraction, and morphological filtering are all integrated into the procedure, which allows it to get accurate results while minimizing the amount of time spent processing. DRIVE, Stare, and CHASEDB1 datasets have all been evaluated, and the results reveal that it has competitive performance metrics, which makes it appropriate for real-time diagnosis. In the initial pipeline that Faria *et al.* [38] developed, we coupled the capabilities of eXplainable AI (XAI) with eight well-known CNN models that had already been educated. In addition to enabling rapid image classification, the second pipeline employs Attention U-Net, Trans U-NET, and Swin-UNET designs to tackle the difficult task of retinal blood vessel segmentation. This is done in order to get the desired results. using meticulous segmentation as a method.

Due to the distinctive needs and difficulties associated with retinal image registration, deep learning techniques in this field have been approached with caution. One of these issues is the preservation of sparsely defined structures, such as vasculature, across very uniform backdrops. Other challenges include the evolution of illnesses and the anticipated massive displacement alterations. Despite the fact that deep learning has been used for image classification and segmentation, its use in retinal image registration is still in need of improvement, and there have only been a few research that have tested its potential. Nevertheless, these algorithms are able to acquire the ability to map intricate connections between images and precisely align them, even when applied to massive datasets [39].

This study uses a DeepDR network that has been trained from beginning to finish, making use of features from the lesionaware sub-network as well as the original images. This is in contrast to prior research that utilized numerous CNNs in order to detect and categorize lesions.

As a consequence of this all-encompassing strategy, the grading results were enhanced, and the performance of diagnosing different stages of DR in real-world datasets performed much better [40]. Jabbar *et al.* [41] develop a computer-based solution that utilizes diagnostic and analysis techniques to accurately detect and classify different stages of diabetic retinopathy. On the other hand, weakly supervised learning [42] and conv model given good results. Upon conducting an extensive examination of the existing literature, it was found that there was few research that used CNN, however yielded significant outcomes. Even among the group of people who saw positive results, the method was found to be computationally intensive, therefore necessitating the use of sophisticated computer equipment.

III. METHODOLOGY

The primary function of a network is to acquire feature maps at various spatial resolutions. To achieve this objective, we use ResNet 50 Model [43]. In order to provide support for the evidential classifier that is based on the Dempster-Shafer theory, CNN gathers features from the input data. In particular, deep convolutional neural networks, often known as CNNs, have emerged as one of the most widely used architectures for deep learning applications in the field of image categorization [44]. The feature maps, in addition to the original image, were then processed into a modified U-NET architecture. Within the framework of the U-Net design [45], a contracting sub-network is connected to a symmetric upsampling sub-network in such a way that the representation that is produced by the last layer of the upsampling route matches to the dimensions of the layer that comes immediately after the final layer. Our model evaluated on three public different datasets which is described in the datasets section. The purpose of this research is to offer a unique model that was developed for the purpose of OC and OD segmentation by MaskRCNN. This model was especially targeted to enhance glaucoma screening efforts. Through the use of the UNET model, Dual Attention Block, and pretrained modules, we were able to optimize the structure of the model. Further, we offer a technique for the processing of fundus photographs, with the goal of standardizing and improving fundus illustrations.

A. Datasets

Our experiment used three different datasets. Firstly, published in 2011, the first Retinal IMage Database version for Optic Nerve Evaluation (RIM-ONE) [46]. The dataset consists of three versions. The RIM-ONE, STARE, and IDRiD databases, frequently utilized for retinal image analysis, possess several limitations: RIM-ONE possesses a limited sample size, inconsistent picture quality, and insufficient comprehensive annotations, hindering the ability to generalize and construct stable models. STARE can only hold 40 pictures, and its main purpose is to separate retinal vessels, which makes it less useful for other tasks. It also experiences class imbalance and annotation difficulties. IDRiD doesn't have a smooth distribution of conditions, picture quality varies, and there aren't many annotations for some lesions, which makes it less useful for training deep models. Common issues across datasets include inconsistent annotations, noisy or low-quality pictures, and insufficient diversity in demographics and circumstances, all of which hinder models' ability to generalize and perform well. Notwithstanding these problems, these datasets retain their value; integrating them with additional datasets or employing augmentation techniques may mitigate some of these drawbacks.

Along with the particular uncertainty and even incorrect usage noted in certain situations of the three versions released, prompted us to propose updating and integrating them into a new, publicly accessible version named RIM-ONE DL (Rim-ONE for Deep Learning). Fig. 1 shows some samples of RIM-ONE DL dataset.

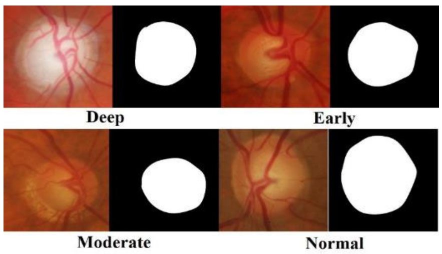

Fig. 1. Samples of RIM-ONE DL datasets for Glaucoma and glaucoma suspicious and Normal phase.

Secondly, STARE [47]. The dataset has a total of 400 images. The camera used was the Topcon TRV-50 fundus camera, which has a Field of View (FOV) of 35 degrees. Fig. 2 illustrates some samples of the STARE dataset.

Fig. 2. Samples of STARE dataset at different stages of Diabetic retinopathy.

Third, datasets are used for the localization of OD centers, whereas the IDRiD dataset [48] is employed specifically for the localization of Fovea centers. On the other hand, the lack of a validation set in some datasets is absolutely necessary for the effective operation of supervised learning systems. The selection of the training, validation, and testing sets was accomplished by the use of a cross-validation approach. We used 5-fold cross-validation to test model performance. We randomly mixed the dataset and divided it into five equal folds to prevent data leakage. Each fold was used as a validation set once, while the other four folds were utilized for training. We chose this method to strike a balance between computational efficiency and performance stability.

Annotations were verified by two retinal specialists after the images were taken in India using a digital fundus camera (Kowa-VX-10a). The images were manually labeled after they were collected. Several examples of the IDRiD dataset are shown in Fig. 3.

Fig. 3. Samples of IDRiD dataset including retinal vessel segmentation mask.

B. Optic Disk Removal Model

Due to the optic disc's presence in the fundus image of the retina, it is difficult to differentiate between light and dark lesions. Because the two lesions have similar intensity levels. the neural network has a hard time distinguishing between them. Therefore, in order to reduce its negative effect on the forecasts, it is necessary to remove the optic disc [49].

Given the varied appearances of Optic Disc (OD) in both normal and Glaucoma situations, we have developed a comprehensive feature descriptor by combining several characteristics to ensure its robustness. For the purpose of extracting glaucomatous qualities, such as texture, intensity, color moments, and histogram attributes, we presented a combination of many different forms of content-based features. The primary benefit of the optic disk removal technique is its ability to generate images without Areas of Interest (ROI) including Optic Disk (OD) pixels [50]. This enhances the performance of exudates identification or segmentation, particularly when the exudates have a similar yellow hue to the OD pixels. Following the training of a Mask-RCNN model that includes a backbone, we were able to eliminate OD. We train Mask R-CNN to eliminate all regions related to the optic disk before feeding the dataset to our model for predicting segmentation. This enhances the performance of exudate identification or segmentation, particularly when the exudates have a similar yellow hue to the OD pixels. Following the training of a Mask-RCNN model that includes a backbone, we were able to eliminate OD. Furthermore, it is a two-head branch detector in addition to a region proposal network. To extract features, the backbone is responsible for doing so. The Region Proposal Network (RPN) determines the region of interest that the various branches of the head will include. While the succeeding head is responsible for categorizing, the initial head is responsible for facilitating bounding box regression. Mask-RCNN, an improvement on Faster-RCNN, incorporates an additional fully connected network branch, such as segmentation, which utilizes the region of interest throughout the training process.

Furthermore, it is a two-head branch detector in addition to a region proposal network. In order to extract features, the backbone is responsible for doing so. The region Proposal Network (RPN) is responsible for determining the region of interest that will be included into the various branches of the head. While the succeeding head is responsible for categorizing, the initial head is responsible for facilitating bounding box regression. Mask-RCNN is an improvement on Faster-RCNN that incorporates an extra fully connected network branch, such as segmentation, that makes use of the region of interest throughout the training process.

An illustration of the process of optic disc segmentation, which is an essential part of the approach for detecting diabetic retinopathy, may be seen in Fig. 4. It displays the segmentation technique that removes the optical disc from retinal fundus pictures, which makes it easier to precisely identify lesions. This is accomplished by using visual representation. It is possible to have a better understanding of the impact that segmentation has on the visibility of lesions, such as microaneurysms and hemorrhages, by comparing the pictures taken before and after the procedure.

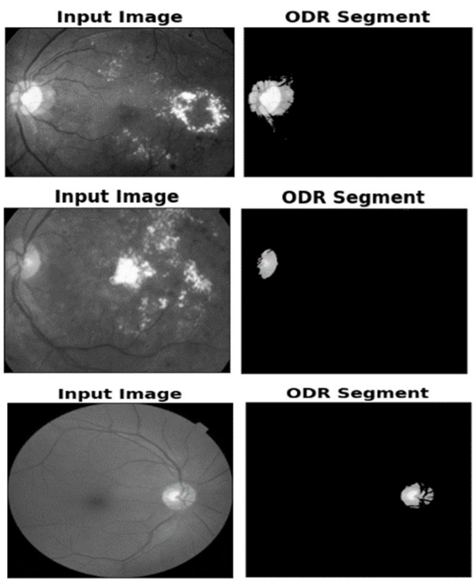

Fig. 4. Optic disk removal segmentation.

C. Preprocessing Techniques

The process starts with the enhancement of contrast. The contrast of fundus images decreases as the pixel distance from the image's center increases. To alleviate this condition, adaptive contrast equalization improves the contrast while simultaneously resolving the problems of noise and oversaturation in associated regions.

CLAHE [33] is mostly used for the purpose of improving the contrast with digital medical photographs. It has been shown that this strategy is more effective in this particular domain compared to both the conventional histogram equalization and the adaptive histogram equalization. It provides localized contrast enhancement, enabling the improvement of selected picture regions without affecting others. The use of a clip limit mitigates excessive contrast enhancement in regions characterized by uniform intensity or substantial noise. Its adaptability renders it appropriate for uses including medical imaging, satellite images, and low-light photography. Furthermore, CLAHE proficiently preserves and enhances nuanced information that conventional histogram equalization may overlook. This approach is very successful in enhancing visual contrast in low-light or foggy circumstances.

Algorithm 1. Data pre-processing Pipeline for enhancement and augmentation of images.

Input Layer: Images of fundus datasets.
Output Layer: Augmented Images.
1- function CLAHE (Image, clip limit) *// Contrast Enhancement.*
Grayscale image = convert to grayscale(image)
clahe = create CLAHE (clip Limit = cliplimit)
enhanced image = clahe apply (grayscale image)
return Enhanced images.
2-function illumination correction (image, desired intensity) // Illumination Correction
Local average intensity = calculate local average intensity (image)
Correction factor = desired intensity / local average intensity.
Corrected image = image ∗ correction factor
return corrected image
set: augmented_images = [ ]
For i in images do *// Cropping*
Cropped image = crop_image(image)
Augmented images_append (cropped_image)
For angle in [90, 120, 180, 270] do *// Rotation*
Rotated image = rotate image (image, angle)
Augmented images append (rotated image)
End For
    Horizontally flipped image = flip image horizontally (image)
    Vertically flipped image = flip image vertically(image)
    Augmented images extend ([horizontally flipped image,
Vertically flipped image])
return augmented images

The CLAHE algorithm is used to the luminosity channel in order to keep shading distances constant. The image is converted into the l×a×b space in order to accomplish this particular objective. As a result, the brightness is improved with the use of CLAHE [23]. The preprocessing steps are described in Algorithm 1. For enhancement and augmentations process.

Edges in an image are defined as sudden shifts or discontinuities that may effectively store almost the same amount of information as pixels. The canny edge detection approach involves reducing noise by smoothing the image and then detecting sudden changes in intensity. The procedure involves many processes, including noise reduction, intensity gradient detection, non-maximum suppression, and hysteresis thresholding.

This method utilizes the pixel-area correlation to resample the data, which leads to an output that is devoid of noise.

The normalizing method is used to modify the boundaries of pixel values. It might also be referred to as contrast stretching or histogram stretching. The purpose of this process is to decrease the amount of noise in the image and then return the resulting values to the intensity level.

In order to do this, it is necessary to divide each pixel by 255, so scaling them to a range between zero and one. The contrast enhancement that was conducted on the image reveals the intricacies of the optic disc, which is an essential stage in the process of doing research on the area of interest.

In order to address the issue of overfitting caused by the scarcity of fundus datasets, the photos were enhanced. Additionally, one of the goals of the enhancement was to reduce the gap that existed between the sample groups. Image rotation, scaling, cropping, horizontal reflection (mirror), and flipping (vertical, diagonal, and orthogonal) were some of the augmentation alterations that were performed. The OD detection system may be able to recognize a higher number of issues in fundus images and variations in image acquisitions as a result of these advancements. The rationale for using the augmentation technique with the transfer learning approach is because pre-trained models mostly utilize genuine images.

D. Model Structure

A pertained ResNet-50 model applies for getting feature extraction at the initial step of feed data. The model using weights trained on a large dataset e.g., ImageNet. This enables the model to provide decent initial feature representations even in cases where uses different datasets. The process of the network starts with loading a pretrained ResNet-50 model after that modify the ResNet-50 to output intermediate feature maps. Next, Extract layer for instantiate the encoder and pass an input image through it. Finally, save all shapes of the feature maps. After that, these feature maps, in addition to the initial images, were included into a modified version of the U-Net framework.

Dual Attention Block DAB add to skin connection. For the purpose of producing the exudates, OD, and vessels segmentation mask of expected, the outputs of the DAB blocks were sampled and aggregated. Our objective is to add DAB layers to the process of reducing computational complexity while simultaneously avoiding a significant increase in the complexity of the computations. We redesigned the U-Net to accommodate multiple DAB blocks by establishing local bottleneck structures in skip-connection pairs. To emphasize key contributions of DAB is made up of channel and content attention modules, and combining these two types of attention models increases the number of robots in the model while reducing complexity.

It was observed that the performance of the network was significantly improved by incorporation of a self-attention mechanism into the bottleneck layers, which included a decreasing route, the attention module, and an expanding path.

Fig. 5 represents the proposed model of DAB block architecture; the framework includes channel attention. As a result, we redesigned U-Net in order to accommodate numerous DAB blocks into the network. This was accomplished by establishing local bottleneck structures in each skip-connection pair. function at different spatial resolutions.

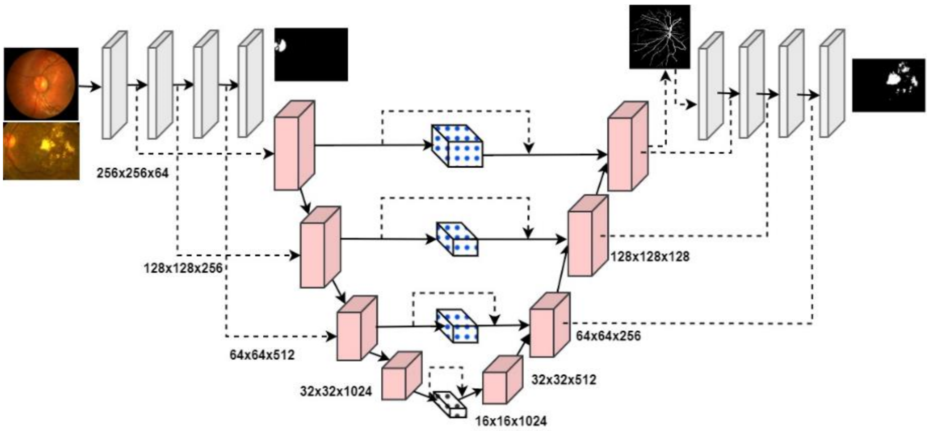

Fig. 5. DAB block architecture, the framework includes channel attention, which produces (AC) as output of red box, and content attention employing multi-head self-attention, which produces (AS) as output blue box.

DAB is formed up of channel and content attention modules; the combination of these two types of attention models will result in an increase in the number of robots in the model while simultaneously reducing the complexity. However, in each spatial feature map, content attention was devoted to the pixels that were represented individually.

After each down-convolution block was finished, the features were pooled and then down-sampled before being transferred to the block. This process was repeated until that block was finished. This was then followed by up-sampling to the size that they had been at when they were first established. If they are processed in this fashion, the pairs of down-sampling and up-sampling convolution blocks may be considered to be local bottleneck structures that work at distinct spatial resolutions. This is because they are able to illustrate Fig. 6 using the layers *q*, *k*, and *v*, we were able to determine the query from Dense-CNN layers.

The Attention Score (AC) that was produced as a result of the content attention was a combination of the key (*k*) and the query vectors *q*.

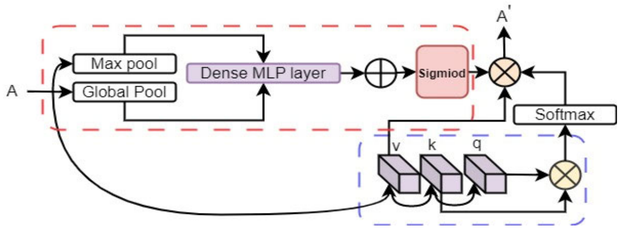

Fig. 6. The architecture of DAB-block. A U-Net model, equipped with a novel attention block and a redesigned skip-connection pathway, is used to simultaneously locate the fovea and segment the optic disc.

IV. Result and Discussion

We implement the suggested DAB-UNet model using the Keras framework, with TensorFlow serving as the backend. We run all experiments on an NVIDIA TITAN XP GPU with 12 GB of RAM. We selected this GPU for deep learning projects due to its optimal combination of processing capacity and resource availability. In preparation for further tests, we modified the training technique such that it would continue until 60 epochs passed. To fine-tune the model's hyperparameters, we used a grid search to determine which parameters provide optimal performance on the test data provided. It is critical to choose the best hyperparameters to guarantee effective model training. Here, we've set the learning rate at 0.0001 and fixed the number of training epochs at 50. We further enhanced the image's patches by rotating them by 90 degrees. It entails delineating a range of potential values for each hyperparameter and assessing all conceivable combinations within this established grid.

Starting with a learning rate of 10^3, the training regimen was implemented, and during the training process, a learning rate decay method was used. We used the Adam optimizer with a momentum value of 0.9 in order to get the model parameters to their optimal state.

However, the number of n^{th} convolutional layer of our proposed DAB-UNET determined for accurate segment. the f-map Xn input is used to generate a final output f-map Xn output by computing the proposed residual-skip connection of Resnet50 model as: Xn output = S (Xn input) + Xn input. Where *S* is the stack of convolution layers in the shortcut-path of the algorithm.

Accordingly, the output of a residual skip connection where Xnout $\in \mathcal{R}^{B\times H\times W\times E}$. This function is then concatenated with the scaled n^{th} decoder output.

Firstly, where Y^n out $\in \mathcal{R}^{B\times H\times W\times D}$ is computed to regain the lost spatial information of model. all features aggregate to represent the final prediction.

Second, where X^n $C \in \mathcal{R}^{B\times H\times W\times (G+E)}$ and (B, H, W, G, and *E) respectively* denote the channel concatenation, batch size [B], height [H], width [W], and depth of X^n_{out}.

On the other hand, Batch normalization layers are applied after each convolution in both the encoder and decoder. This is done to standardize the inputs to the layers inside each mini batch, in order to address the issue of internal covariate shift. Furthermore, the DAB-UNET utilizes the sigmoid activation function to generate a 2D feature map, which may be used for segmentation or regression purposes. In the process of segmentation, the output maps are processed by thresholding in order to get binary masks of the OD. The output heatmaps that approximate the intended localization are subjected to a two-dimensional argmax operation in order to ascertain the spatial coordinates of the OD and Fovea centers. However, when compared to the backdrop, the OD of ROIs are much tiny. In the event if a typical measure is used, which considers both the background and the foreground pixels in equal measure, it would include the introduction of bias towards the background pixels.

$$IoU_s(y,\hat{y}) = \frac{\sum_{i=1}^{N} y_i \times \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y} - \sum_{i=1}^{N} y_i \times \hat{y}_i} \quad (1)$$

The actual label, the projected label, and the total pixel count are denoted by and N, respectively. In Eq. (1), the measure of similarity predicted between the real label and the annotation is equal to the product of y and y. On the other hand, the architecture that was provided results in the model obtaining binary cross-entropy and mean squared errors, both of which are used as loss functions for segmentation and regression, respectively. The loss functions for both tasks are optimized without modification by using a momentum optimizer with a starting learning-rate of 0.001 and exponential decay values (*β1-first layer, β2-second layer*) as the beginning values. Over the course of ten epochs, the LR will decrease by ten percent in the event that validation loss does not improve. Setting the epoch to 100 at the beginning of the training process, we terminate the training phase with a callback when the test loss stops getting better.

A grid search is used to determine which parameters provide the optimal performance on the test data provided in order to fine-tune the model's hyperparameters. It is critical to choose the best hyperparameters to guarantee effective model training. Here, we've set the learning rate at 0.0001 and fixed the number of training epochs at 50. The image's patches were further augmented by rotating them by 90°. A total of 1156 MA image patches were utilized to train the DCNN. Furthermore, the training process included the use of 5000 images patches that were not from Massachusetts.

Down-sampling and up-sampling are two methods that are used in a U-Net to encode spatial information to various channels. We are of the opinion that channel-wise attention is an excellent method for making use of this information at the bottleneck layers, which often include a great number of channels.

The input feature map A was sent through the average pooling and max pooling layers simultaneously, which resulted in each channel being compressed to a single value. After being sent through a single, shared multi-layer perceptron with one hidden layer inside the model, these two feature maps were then merged together in order to calculate the final channel attention score (A_C).

$$A' = \text{ReLU}(A_S)\sigma(A_C)\nu \quad (2)$$

The value vector was scaled not just according to the content Attention Score (AS), but also according to the channel Attention Score (AC). This was done in order to ensure that the value vector was accurate. Eq. (2) represents the output of the DAB block, which is denoted by the A'. The SoftMax algorithm was applied to all of the attention heads, and the symbol σ refers to the sigmoid function. The exact location of the OD border will be of great assistance in the investigation of the progression of a variety of eye illnesses and the outcomes of therapy. There is a possibility that failures might be attributed to the fact that pictures are often very inhomogeneous. Additionally, lesions have the potential to generate false targets or alter the predicted OD characteristics, particularly in the vicinity of its boundaries. A unique technique is presented to enhance the identification of the Optic Disc (OD) border. This method involves localizing the OD area, extracting the segmentation of blood vessels, and using an active contour model with a variational level set formulation.

It has been shown that skip connections are successful in recovering fine-grained information of the target objects; they are also capable of constructing segmentation masks that include fine details even when applied to complicated backgrounds. Skip connections are also essential to the success of instance-level segmentation models like Mask-RCNN, which allows the segmentation of OD and eliminates the false-positive results.

This is because skip connections allow for the success of these models. When it comes to the retinal images, the most luminous region is the Optic Disc (OD), and the blood vessels start from the center section of the OD. Accordingly, there are a great number of ships that are crossing their limit, which makes its division more difficult. The use of the vesselness filter, which is a method that is often utilized for the purpose of improving the visibility of blood vessels, is what allows for the identification of blood vessels.

However, the suggested approach was used to segment the Optic Disc (OD) of the IDRiD and RIMONE data sets. The retinal Optic Disk border detection mechanism yields a result that classifies each pixel. Each pixel is categorized as either OD (optical density) or non-OD. Next, we evaluate each binary image obtained by comparing it with the matching ground truth. This evaluation is done by calculating four performance metrics. Fig. 7 illustrates the final output results of STARE images as predicted by our proposed model, while Fig. 8 showed Results segmentation of RIM-ONE DL datasets.

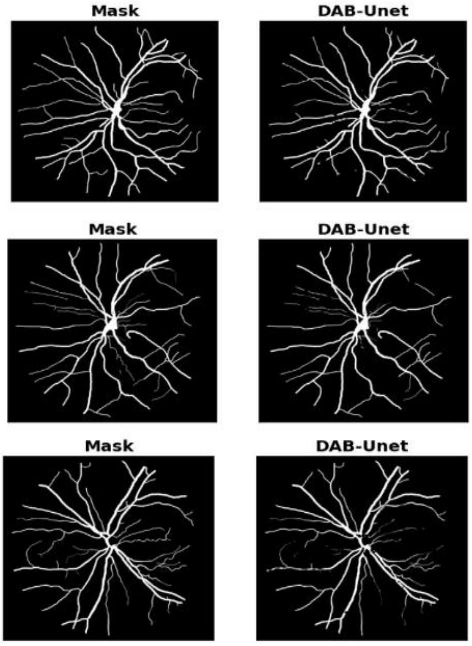

Fig. 7. Results segmentation of STARE datasets.

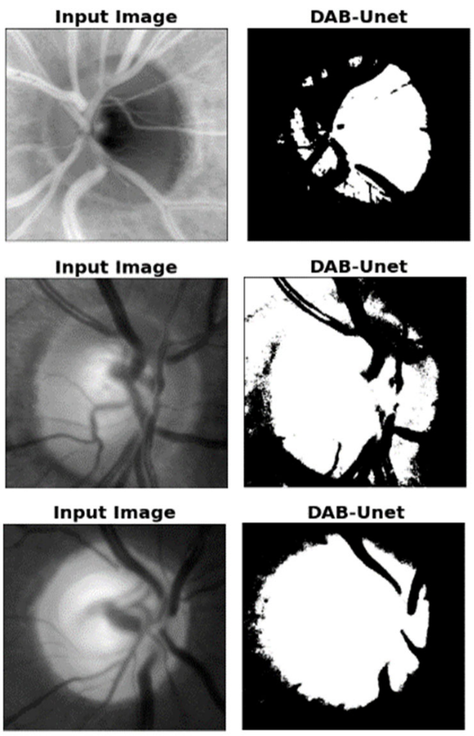

Fig. 8. Results segmentation of RIM-ONE DL datasets.

We assess the segmented OD masks using sensitivity (Sn), accuracy (Acc), Intersection over Union (IoU). Whereas the mIoU measures the overlap between the real and projected OD masks, the mSn and mAcc are used to assess the false-negative area and the proportion of properly categorized pixels.

The effectiveness of our proposed model is showing the feature extraction with ResNet model 50 encoders before feed to scale UNET model of encoder-decoder. The second stage of our proposed model is used ResNet-50 model for the decoder that is responsible for eliminating all false positive errors. On the other hand, we preprocess stage with train Mask-RCNN model for extracting optic disk region, that helpful to increase the accurate model.

Fig. 9 shows overlay visualizations with randomly applied preprocessing to enhance the prediction and clarify small regions and tiny lesions. The red line predicts the hard extrudates while blue line detects the small regions of infected and soft exudates.

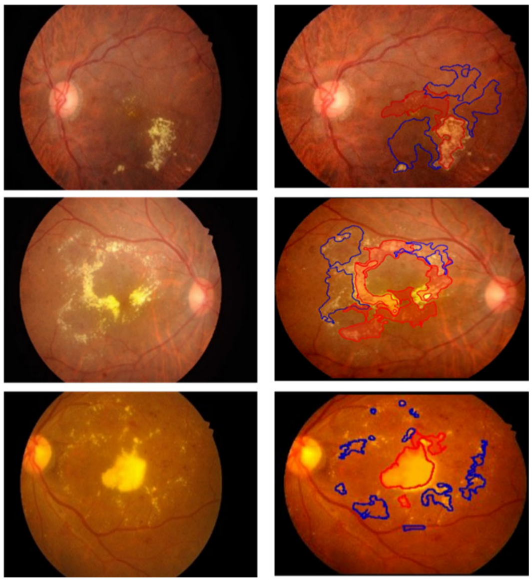

Fig. 9. Comparison of hard extrudates and soft extrudates with ground truth.

Table I shows results of OD segmentation using the proposed DAB-UNET with transfer learning (TR) and Data Augmentation (DA). Table II presents the results of OD segmentation using the proposed DAB-UNET on three publicly accessible datasets without data augmentation transfer learning.

TABLE I. THE RESULTS OF OD SEGMENTATION USING DAB-UNET WITH TRANSFER LEARNING AND DATA AUGMENTATION

Datasets	Sn%		Acc%		IoU%	
	Mean	Median	Mean	Median	Mean	Median
IDRiD	97.8	98.5	99.4	86.8	84.7	94.5
RIMONE	96.7	96.2	98.7	93.0	88.3	95.5
STARE	97.5	98.4	98.8	94.6	84.2	94.6

TABLE II. THE RESULTS OF OD SEGMENTATION DAB-UNET ON DIFFERENT DATASETS

Datasets	Sn%		Acc%		IoU%	
	Mean	Median	Mean	Median	Mean	Median
IDRiD	88.5	95.0	97.7	96.8	85.5	90.5
RIMONE	95.9	97.7	96.2	97.4	90.6	92.9
STARE	96.2	98.0	99.9	99.9	92.0	93.7

Moreover, we added a Dual Attention Block for UNET model that works effectively for extracting lesions with attention map. Fig. 10 illustrates the results of final semantic segmentation of our proposed DAB-UNET with gray scale and generated mask of OD and Hard Exudates lesions.

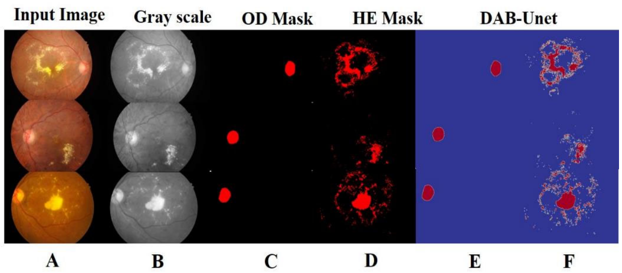

Fig. 10. A) illustrates input image dataset B) represent gray scale C) optic disk mask D) Hard exudates mask label E) and F) Final segmentation results of our proposed model DAB-Unet.

Fig. 11 illustrates the comparative ROC curve of three models. This performance improvement shows the effectiveness of the network architecture we suggested, like dual attention blocks, feature extraction, and skip connections. Furthermore, the UNET model incorporates a dual attention block that efficiently extracts lesions using an attention map.

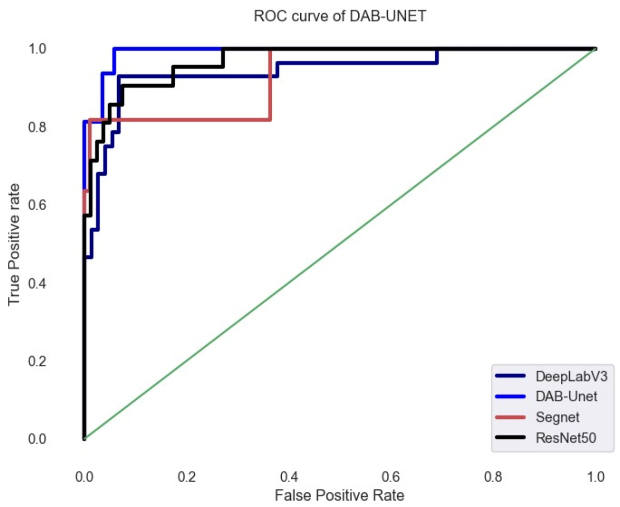

Fig. 11. DAB-UNET comparative ROC curve with three models.

Table III shows the State-of-Art methods used in some segmentation with different conv models with STARE dataset. According to Accuracy (ACC) and Intersection over Union (IoU), this comparison illustrates how various approaches fare in terms of their respective performance. Based on the results, it seems that the DAB-UNET technique with Resnet50 has obtained the maximum accuracy and IoU.

Table IV describes some experimental results of the suggested methodologies for retina blood vessel segmentation, which were then compared to other classical and deep learning-based algorithms for the IDRiD dataset.

DAB-UNET demonstrates competitive performance with elevated AUC and accuracy, signifying efficient segmentation of retinal blood vessels. The U-Net versions often exhibit commendable performance, with R2U-Net and Residual U-Net demonstrating robust outcomes. Conventional techniques, however successful, generally exhibit marginally worse performance metrics relative to deep learning-based alternatives.

TABLE III. EXPERIMENTAL RESULTS OF THE PROPOSED APPROACHES FOR RETINA BLOOD VESSEL SEGMENTATION WERE OBTAINED AND COMPARED AGAINST OTHER TRADITIONAL AND DEEP LEARNING-BASED APPROACHES (STARE)

Study	Methodology	Segmentation	Results
[51]	Adaptive Active Morphological Operation	Otsu Thresholding	ACC 60%
[52]	Deep CNN, Seg-net	DeepLabV3, Segnet	ACC 88%
[53]	Generative Adversarial Network (cGAN)	U-Net, HEDNet	84.05%
[54]	DR-Net, CNN	Regression	84.50% ACC
Segnet (Our)	Multilayer-CNN	Semantic segmentation	**57% ACC 54% IoU**
DeepLab V3 (Our)	Deep classify model	Extract features for semantic region	**88% ACC 80% IoU**
ResNet 50	Residual model	segmentation	**0.9509**
DAB-Unet	Resnet50, DAB-Unet	Segment	**98.86%ACC 94.6% IoU**

TABLE IV. EXPERIMENTAL RESULTS OF THE PROPOSED APPROACHES FOR RETINA BLOOD VESSEL SEGMENTATION WERE OBTAINED AND COMPARED AGAINST OTHER TRADITIONAL AND DEEP LEARNING-BASED APPROACHES FOR IDRID DATASET

Study	SE %	SP %	AC %	AUC %
Marin *et al.* [55]	69.40	97.70	95.20	**98.20**
Fraz [56]	75.48	97.63	95.34	**97.68**
Roychowdhury [57]	77.20	97.30	95.10	**96.90**
Liskowsk [58]	78.67	97.54	95.66	**97.85**
Qiaoliang Li [59]	77.26	98.44	96.28	**98.79**
WSF [60]	78.80	97.60	95.70	**95.90**
U-Net	82.70	98.42	96.90	**98.98**
CE-Net [61]	78.41	97.25	95.83	**97.87**
R2U-Net [62]	82.98	98.62	97.12	**99.14**
Residual U-Net (our)	**82.03**	**98.56**	**97.00**	**99.04**
Recurrent U-Net (our)	**0.8108**	**0.9871**	**0.9706**	**0.9809**
DAB-UNET (our)	**0.856**	**0.9857**	**0.9889**	**0.9909**

Table V illustrates the severity of Diabetic Retinopathy (DR), a complication of diabetes that affects the eyes, using the. The table provides a detailed comparison of different neural network architectures for classifying various severity levels of Diabetic Retinopathy (DR) on the IDRiD dataset. DAB-UNET significantly enhances the classification accuracy for different DR severity levels.

Table VI illustrates the severity of Diabetic Retinopathy (DR), a complication of diabetes that affects the eyes, using the STARE dataset. DAB-UNET outperforms others significantly with 88.30%, followed by Attention UNet at 85.94%.

TABLE V. THE NETWORKS FOR THE VARIOUS DR SEVERITY LEVELS ON IDRID

Network	No DR %	Mild%	Moderate %	Severe %	Proliferative %
Residual U-Net	95.27	82.46	80.63	76.98	74.70
Recurrent U-Net	97.48	85.59	86.53	81.48	80.32
Attention UNet	98.80	88.73	87.63	83.74	85.94
U-Net ++	97.59	87.97	85.73	81.83	80.57
DAB-UNET (our)	**99.67**	**89.35**	**86.13**	**84.76**	**88.30**

TABLE VI. THE NETWORKS FOR THE VARIOUS DR SEVERITY LEVELS ON STARE

Network	**No DR %**	**Mild %**	**Moderate %**	**Severe %**	**Proliferative %**
Residual U-Net	98.43	84.79	83.46	73.94	76.23
Recurrent U-Net	97.89	86.74	87.37	80.30	84.31
Attention UNet	99.65	91.03	88.76	85.90	87.50
U-Net ++	96.74	88.89	86.69	83.78	84.47
DAB-UNET (our)	**99.56**	**90.67**	**88.04**	**86.65**	**87.83**

Finally, the current experiments put the suggested DAB-UNet model to the test on certain datasets, showing that it works well with the given conditions. To fully show how robust and useful it is, however, it is important to test the model's ability to work in a variety of imaging settings, such as those with different image capture equipment. In the future, we will apply augmentation strategies to enhance data augmentation approaches, imitating varied imaging situations during training to increase generalization.

V. CONCLUSION

The study that we have conducted represents a significant step forward in the area of early health detection. This is accomplished by carefully evaluating retinal blood vessels in fundus and DR images. In this research, we proposed the network that is referred to as DAB-UNET which is proposed with the purpose of automatically segregating optic disk circles and determining the locations of Fovea-centers region. Foreground pushing is the primary emphasis of the region-guided attention network block, while the cascaded partial decoder is responsible for aligning the high-level and low-level features, which ultimately leads to an improvement in the model's performance.

The proposed residual skip connection has been shown to improve segmentation and localization final results in comparison to traditional skip connections, such as those found in UNet. This is despite the fact that the structure is more lightweight than other skip connections. Our further plan was to evaluate the impact of the number of parameters, number of layers, and depth of the DAB-UNET to obtain the most preferred outcomes that are possible in the future. And since the recommended skip connection will also have a better capability to reconstruct spatial information that got lost in the pooling process of the encoder, the same will be used in other parts of medical pictures inclusive of segmentation and localization. This is because; right spatial information is critical in these areas.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ammar S. Al-Zubaidi, Mohammed Al-Mukhtar study concepts study design and analysis with interpretation; Mohammed Al-Mukhtar, provide diagrams and literature results of SOTA. Haithem Kareem Abass experimental studies, statistical analysis and manuscript editing. All authors had approved the final version

REFERENCES

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, Dec. 2010. doi: 10.1109/RBME.2010.2084567

[2] I. Rodriguez-Una and A. Azuara-Blanco, "New technologies for glaucoma detection," *Asia-Pacific Journal of Ophthalmology*, vol. 7, pp. 394–404, Nov. 2018. doi: 10.22608/APO.2018349

[3] A. S. Al-Zubaidi, M. Al-Mukhtar, and A. A. A. Baghdadi, "ALL-FABNET: Acute lymphocytic leukemia segmentation using a flipping attention block decoder-encoder network," *Journal of Image and Graphics*, vol. 12, no. 3, pp. 239–249, 2024. doi: 10.18178/joig.12.3.239-249

[4] T. Fernando, S. Denman, and D. Ahmedt-Aristizabal *et al.*, "Neural memory plasticity for medical anomaly detection," *Neural Networks*, vol. 127, pp. 67–81, Jul. 2020. doi: 10.1016/j.neunet.2020.04.011

[5] A. S. Al-Zubaidi, M. Al-Mukhtar, M. H. Al-hashimi, and H. Ijaz. "Skin lesion segmentation for melanoma using dilated DenseUNet," *Journal of ICT Research & Applications*, vol. 18, no. 1, pp. 21–35, Jun 2024. doi: 10.5614/itbj.ict.res.appl.2023.18.1.2

[6] I. M. Hameed, M. Al-Mukhtar, and A. S. Al-Zubaidi, "MCNet: Mask cell of multi class deep network for blood cells detection and classification," *International Journal of Intelligent Engineering & Systems*, vol. 18, no. 1, 2025. doi: 10.22266/ijies2025.0229.23

[7] K. B. Chen, Y. Xuan, A. J. Lin, and S. H. Guo, "Lung computed tomography image segmentation based on U-Net network fused with dilated convolution," *Computer Methods and Programs in Biomedicine*, vol. 207, 106170, Aug. 2021. doi: 10.1016/j.cmpb.2021.106170

[8] M. Ahmad, S. F. Qadri, M. U. Ashraf, K. Subhi, S. Khan, S. S. Zareen, and S. Qadri, "Efficient liver segmentation from computed tomography images using deep learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, May 2022. doi: 10.1155/2022/2665283

[9] H. Chen and K. Kim, "Multi-convolutional channel residual spatial attention U-NET for industrial and medical image segmentation,"

IEEE Access, vol. 12, pp. 76089–76101, 2024. doi: 10.1109/ACCESS.2024.3405957

[10] U. T. Nguyen, A. Bhuiyan, and L. A. Park *et al.*, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognit*., vol. 46, no. 3, pp. 703–715, 2013. doi: 10.1016/j.patcog.2012.08.009

[11] N. P. Singh and R. Srivastava, "Retinal blood vessels segmentation by using gumbel probability distribution function based matched filter," *Computer Methods and Programs in Biomedicine*, vol. 129, pp. 40–50, Jun 2016. doi: 10.1016/j.cmpb.2016.03.001

[12] E. Imani, M. Javidi, and H. R. Pourreza, "Improvement of retinal blood vessel detection using morphological component analysis," *Computer Methods and Programs in Biomedicine*, vol. 118, no. 3, pp. 263–279, Mar. 2015. doi: 10.1016/j.cmpb.2015.01.004

[13] N. K. J. Ali, and M. H. Senthi *et al.*, "Early detection of glaucoma: Feature visualization with a deep convolutional network," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 12, 2350508, Dec. 2024. doi: 10.1080/21681163.2024.2350508

[14] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 12, pp. 2181–2193, 2017. doi: 10.1007/s11548-017-1619-0

[15] C. Chen, J. H. Chuah, and R. Ali *et al.*, "Retinal vessel segmentation using deep learning: A review," *IEEE Access*, vol. 9, pp. 111985–112004, Aug. 2021. doi: 10.1109/ACCESS. 2021.3102176

[16] H. Huang, Z. Shang, and C. Yu, "FRD-Net: A full-resolution dilated convolution network for retinal vessel segmentation," *Biomedical Optics Express*, vol. 15, no. 5, pp. 3344–3365, 2024. doi: 10.1364/BOE.522482

[17] A. Vaswani, N. Shazeer, and N. Parmar *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[18] S. Woo, J. Park, J. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[19] L. Peng, L. Lin, and P. Cheng *et al.*, "Fargo: A joint framework for faz and rv segmentation from octa images," in *Proc. Ophthalmic Medical Image Analysis: 8th International Workshop*, Sept. 27, 2021, vol. 8, pp. 42–51. doi:10.1007/978-3-030-87000-3_5

[20] M. Li, W. Zhang, and Q. Chen. "Image magnification network for vessel segmentation in octa images," in *Proc. Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Oct. 2022, pp. 426–435. doi: 10.1007/978-3-031-18916-6_35

[21] M. Li and K. Huang *et al.*, "OCTA-500: A retinal dataset for optical coherence tomography angiography study," *Medical Image Analysis*, vol. 93, 103092, Apr. 2024. doi: 10.1016/j.media.2024.103092

[22] M. Al-Mukhtar, A. A. Abbas, A. H. Hamad, and M. H. Al-Hashimi, "Normalized-UNET segmentation for COVID-19 utilizing an encoder-decoder connection layer block," *Journal of Image and Graphics*, vol. 12, no. 1, 2024.

[23] C. Li, Y. Tan, and W. Chen *et al.*, "Attention UNET++: A nested attention-aware U-NET for liver CT image segmentation," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 345–349. doi: 10.1109/ICIP40778.2020.9190761

[24] Y. Sun, F. Bi, and Y. Gao *et al.*, "A multi-attention UNET for semantic segmentation in remote sensing images," *Symmetry*, vol. 14, no.5, p. 906, 2022. doi: 10.3390/sym14050906

[25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNET++: A nested U-NET architecture for medical image segmentation," in *Proc. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018*, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, vol. 4, pp. 3–11. doi: 10.1007/978-3-030-00889-5_1

[26] H. K. Abass, M. H. Al-hashimi, A. S. Al-Zubaidi, and M. Al-Mukhtar, "GastroFPN: Advanced deep segmentation model for gastrointestinal disease with enhanced feature pyramid network decoder," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 4, 2024. doi: 10.22266/ijies2024.0831.34

[27] C. Guo M. Szemenyei, and Y. Yi *et al.*, "Sa-UNET: Spatial attention U-NET for retinal vessel segmentation," in *Proc. 2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 1236–1242. doi: 10.1109/ICPR48806.2021.9413346

[28] S. Tang, Z. Qi, J. Granley, and M. Beyeler, "U-Net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina," in *Proc. Ophthalmic Medical Image Analysis: 8th International Workshop, OMIA 2021*, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, vol. 8, pp. 62–71.

[29] M. K. Hasan, M. A. Alam, and M. T. E. Elahi *et al.*, "DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image," *Artificial Intelligence in Medicine*, vol. 111, 102001, 2021. doi: 10.1016/j.artmed.2020.102001

[30] Q. Wang, B. Wu, and P. Zhu *et al.*, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020.

[31] J. Fu, J. Liu, and H. Tian *et al.*, "Dual attention network for scene segmentation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2019.

[32] L. A. Muhammed, "Localizing optic disc in retinal image automatically with entropy-based algorithm," *International Journal of Biomedical Imaging*, 2815163, 2018.

[33] B. Gui, R. J. Shuai, and P. Chen, "Optic disc localization algorithm based on improved corner detection," *Procedia Computer Science*, vol. 131, pp. 311–319, 2018. doi: 10.1016/j.procs.2018.04.169

[34] R. G. Ramani and J. J. Shanthamalar, "Improved image processing techniques for optic disc segmentation in retinal fundus images," *Biomed Signal Process Control*, vol. 58, 101832, 2020. doi: 10.1016/j.bspc.2019.101832

[35] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M.B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *Proc. 12th International Symposium on Medical Information Processing and Analysis*, 2017, vol. 10160, pp. 241–250. doi:10.1117/12.2255740

[36] M. Soltani-Gol, A. Asgharzadeh-Bonab, H. Soltanian-Zadehand, and J. Mazlum, "RDCU-Net: A multi-Scale residual dilated convolution U-Net with spatial pyramid pooling for brain tumor segmentation," *AUT Journal of Electrical Engineering*, vol. 56, no. 2, pp. 203–212. 2024. doi:10.22060/eej.2023.22395.5538

[37] W. Li, Y. Xiao, and H. Hu *et al.*, "Retinal vessel segmentation based on b-cosfire filters in fundus images," *Frontiers in Public Health*, vol. 10, 914973, 2022. doi: 10.3389/fpubh.2022.914973

[38] F. T. J. Faria, M. B. Moin, and P. Debnath *et al.*, "Explainable convolutional neural networks for retinal fundus classification and cutting-edge segmentation models for retinal blood Vessels from fundus images," arXiv preprint, arXiv:2405.07338, 2024.

[39] K. Wisaeng, "Retinal Image registration using partial intensity invariant feature descriptor and redundant keypoints elimination techniques," *IEEE Access*, vol. 12, pp. 83611–83628, 2024. doi: 10.1109/ACCESS.2024.3404242

[40] K. Ohri and M. Kumar, "Domain and label efficient approach for diabetic retinopathy severity detection," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35795–35824, 2024. doi: 10.1007/s11042-023-16908-3

[41] A. Jabbar, S. Naseem, J. Li, and T. Mahmood *et al.*, "Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, pp. 1–20. 2024. doi: 10.1007/s44196-024-00520-w

[42] Z. H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018. doi: 10.1093/nsr/nwx106

[43] M. I. Meyer, A. Galdran, A. M. Mendonça, and A. Campilho, "A pixel-wise distance regression approach for joint retinal optical disc and fovea detection," in *Proc. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference*, Granada, Spain, September 16–20, Part II, vol. 11, 2018, pp. 39–47.

[44] N. Salih, M. Ksantini, N. Hussein, D. B. Halima, A. A. Razzaq, and S. Ahmed. "An advanced approach for predicting ROP stages: Deep learning algorithms and belief function technique," *Iraqi Journal of Science*, vol. 65, no. 7, pp. 4047–4060, 2024. doi: 10.24996/ijs.2024.65.7.39

[45] A. H. Khudair and A. M. Radhi, "Diabetes diagnosis using deep learning," *Iraqi Journal of Science*, vol. 65, no.1, pp. 443–454, 2024. doi: 10.24996/ijs.2024.65.1.36

[46] F. J. F. Batista, T. Diaz-Aleman, and J. Sigut *et al.*, "Rim-one DL: A unified retinal image database for assessing glaucoma using deep learning," *Image Analysis and Stereology*, vol. 39, no. 3, pp. 161–167, 2020. doi:10.5566/ias.2346

[47] T. J. Jebaseeli, C. A. D. Durai, and J. D. Peter, "Segmentation of retinal blood vessels from ophthalmologic diabetic retinopathy images," *Computers & Electrical Engineering*, vol. 73, pp. 245–258, 2019. doi: 10.1016/j.compeleceng.2018.11.024

[48] P. Porwal, S. Pachade, and M. Kokare *et al.*, "Idrid: Diabetic retinopathy–segmentation and grading challenge," *Medical Image Analysis*, vol. 59, 101561, 2020. doi: 10.1016/j.media.2019.101561

[49] F. H. Mahmood and S. A. J. Aziz, "An improved segmentation technique for early detection of exudates of diabetic retinopathy disease," *Iraqi Journal of Science*, vol. 60, no. 11, pp. 2520–2530, 2019. doi: 10.24996/ijs.2019.60.11.25

[50] C. Wan and Y. Chen *et al.*, "EAD-Net: A novel lesion segmentation method in diabetic retinopathy using neural networks," *Disease Markers*, vol. 2021, 6482665, 2021. doi:10.1155/2021/6482665

[51] P. Khojasteh, L. A. P. Júnior, and T. Carvalho *et al.*, "Exudate detection in fundus images using deeply learnable features," *Computers in Biology and Medicine*, vol. 104, pp. 62–69, 2019. doi: 10.1016/j.compbiomed.2018.10.031

[52] D. Mohan, J.R.H. Kumar, and C. S. Seelamantula, "High-performance optic disc segmentation using convolutional neural networks," in *Proc. 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 4038–4042. doi: 10.1109/ICIP.2018.8451543

[53] M. K. Hasan, M. A. Alam, M. T. E. Elahi, S. Roy, and R. Martí, "DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image," *Artificial Intelligence in Medicine*, vol. 111, 102001, 2021. doi: 10.1016/j.artmed.2020.102001

[54] D. Marín, A. Aquino, ME. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2011. doi: 10.1109/TMI.2010.2064333

[55] M. M. Fraz, P. Remagnino, A. Hoppe, and B. Uyyanonvara *et al.*, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012. doi: 10.1109/TBME.2012.2205687

[56] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Blood vessel segmentation of fundus images by major vessel extraction and subimage classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1118–1128, 2015. doi: 10.1109/JBHI.2014.2335617

[57] P. Laskowski and K. Krawiec. "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 2369–2380, 2016. doi: 10.1109/TMI.2016.2546227

[58] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 109–118, 2016. doi: 10.1109/TMI.2015.2457891

[59] Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, and L. Luo *et al.*, "Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 438–450, 2018. doi: 10.1109/TMI.2017.2756073

[60] Z. Gu, J. Cheng, and H. Fu *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019. doi: 10.1109/TMI.2019.2903562

[61] M. Z. Alom, M. Hasan, and C. Yakopcic *et al.*, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," arXiv Preprint, arXiv:1802.06955, 2018.

[62] S. S. Honnahalli, H. Tiwari, and D. V. Chitragar, "Future fusion+ UNET (R2U-Net) deep learning architecture for breast mass segmentation," *Engineering Proceedings*, vol. 59, no. 1, p. 44, 2023. doi: 10.3390/engproc2023059044