EFL-Net: An Efficient Lightweight Neural Network Architecture for Retinal Vessel Segmentation

Nasrin Akbari and Amirali Baniasadi
Department of Electrical and Computer Engineering, University of Victoria, Victoria, Canada

Keywords: Blood Vessel Segmentation, Deep Learning, Image Processing.

Abstract: Accurate segmentation of retinal vessels is crucial for the timely diagnosis and treatment of conditions like diabetes and hypertension, which can prevent blindness. Deep learning algorithms have been successful in segmenting retinal vessels, but they often require a large number of parameters and computations. To address this, we propose an efficient and fast lightweight network (EFL-Net) for retinal blood vessel segmentation. EFL-Net includes the ResNet branches shuffle block (RBS block) and the Dilated Separable Down block (DSD block) to extract features at various granularities and enhance the network receptive field, respectively. These blocks are lightweight and can be easily integrated into existing CNN models. The model also uses PixelShuffle as an upsampling layer in the decoder, which has a higher capacity for learning features than deconvolution and interpolation approaches. The model was tested on the Drive and CHASEDB1 datasets and achieved excellent results with fewer parameters compared to other networks such as ladder net and DCU-Net. EFL-Net achieved F1 measures of 0.8351 and 0.8242 on the CHASEDB1 and DRIVE datasets, respectively, with 0.340 million parameters, compared to 1.5 million for ladder net and 1 million for DCU-Net.

1 INTRODUCTION

The retina is a layer of light-sensitive nerve tissue located at the back of the eye that receives images and transmits them to the brain as electric signals through the optic nerve (Kolb, 2012). Changes in the retina and optic nerve may indicate certain diseases such as glaucoma (Salmon, 2019) or hypertensive retinopathy (HR) (Irshad and Akram, 2014), which can cause blurring of vision. As we age, the oxidative load increases, leading to higher levels of oxidative stress which can cause pathologies such as age-related macular degeneration or neuropathic complications of diabetes in the eye (Payne et al., 2014). Diabetic retinopathy (DR) is a condition that affects individuals with diabetes, causing gradual damage to the retina and potentially leading to vision loss. It is a major complication of diabetes that can threaten vision.

Primary eye care (PEC) can use a funduscopy examination to give an early screening for drug-induced retinal toxicity (Alberta et al., 2022). In the procedure of funduscopy examination, an ophthalmologist looks at the structures of the retina, retinal blood vessels, and optic nerve head (disk) of the eye (Walker et al., 1990). There are several ways to analyze retinal images and find diseases, one of which is retinal image segmentation, which can be divided into manual and automatic methods. Manual segmentation takes time and expertise, while automated algorithms are useful for early detection and treatment of eye diseases due to their increased accuracy, reduced cost, and faster speed compared to manual segmentation.

U-Net (Ronneberger et al., 2015) is an automatic model used to segment vessels in retina images and is one of the successful medical and biomedical image segmentation methods based on deep neural networks. Humans often struggle to distinguish blood vessel images from their distorted backgrounds, making it more difficult to detect diseases. As a result, developing practical algorithms to identify vessel images and their surroundings would be useful (Yang et al., 2022).

Deep neural networks (DNNs) have been shown to be effective in automatically learning reliable and complex features from raw data without the need for manual feature engineering (Ronneberger et al., 2015; Zhou et al., 2021; Gu et al., 2019; Li et al., 2020). These techniques have achieved significant success in the fields of computer vision and medical health. Research on retinal vessel segmentation using DNNs has proposed various architectures for this task, however,

Akbari, N. and Baniasadi, A.
EFL-Net: An Efficient Lightweight Neural Network Architecture for Retinal Vessel Segmentation.
DOI: 10.5220/0011754700003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 920-927
ISBN: 978-989-758-634-7; ISSN: 2184-4321

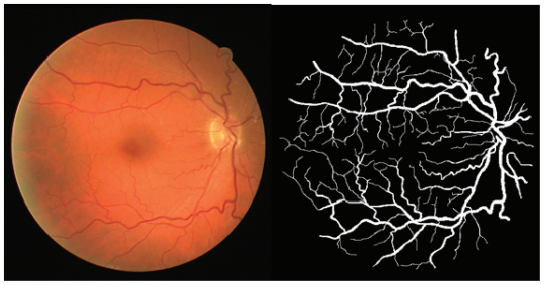

Figure 1: A retinal image from DRIVE dataset (left) and retinal vessel segmentation (right) (Staal et al., 2004).

our observations indicate that many of these models are not optimal in terms of architecture and number of parameters. Figure 2 illustrates the F1 measure and complexity of well-known DNN-based models (Azad et al., 2019; Zhou et al., 2021; Gu et al., 2019; Mou et al., 2021; Li et al., 2020; Zhuang, 2018; Zahangir Alom et al., 2018; Guo et al., 2021; Yang et al., 2022). More research is needed to improve the number of parameters and accuracy of current models. Our research aims to design a lightweight CNN architecture with fewer parameters that can achieve similar or better results in retinal vessel segmentation compared to state-of-the-art networks.

The remainder of the paper is structured as follows: Section 2 describes the proposed network architecture, Section 3 compares experimental results to state-of-the-art neural networks using the CHASEDB1 and DRIVE datasets, and Section 4 provides concluding remarks and discussion of future research directions.

2 METHODOLOGY

U-Net is a convolutional neural network that was designed for image segmentation in the field of biomedicine (Ronneberger et al., 2015). It is an improvement on the previously developed FCN - "Fully convolutional networks for semantic segmentation" (Long et al., 2015). Its ability to perform well with small training datasets has made it the most reliable architecture for the semantic segmentation of biological images.

The U-Net architecture consists of four encoder blocks on the left side, known as the contracting path, and four decoder blocks on the right side referred to as the expansive path. The encoder captures features from the input image and reduces its resolution through pooling layers, while the decoder part reconstructs the image and restores object details through skip connections between the encoder and decoder layers. While the U-Net model has been successful in various tasks, it has several drawbacks including a large number of parameters (31.031 million) and poor performance on retinal vessel segmentation (F1 score of 0.7783 on the CHASEDB1 dataset). In order to address these issues, we analyzed additional papers and their blocks and developed a solution that introduces two new blocks for enhanced feature extraction: the Resnet Branches Shuffle Block (RBSB) and the Dilated Separated Down Block (DSDB). We also employed efficient layers such as pixel shuffle deconvolution and interpolation techniques in the decoder path of the U-Net model (Shi and Caballero, 1874). Our modified U-Net model aims to improve performance on retinal vessel segmentation tasks

2.1 Our Proposed Architecture

Inspired by the U-Net (Ronneberger et al., 2015) and ShuffleNetV2 (Ma et al., 2018) models, we propose the Efficient and Fast Lightweight Neural Network (EFL-Net) for retinal vessel segmentation. Our goal is to create a lightweight and accurate deep learning model for this task. To increase the receptive field of the U-Net model, we introduce the Resnet Branches Shuffle Block (RBS block) and the Dilated Separable Down Block (DSD block) to our architecture. The encoder path of the U-Net model consists of four stages, each comprising EFL-Net, RBS, and DSD blocks for feature extraction and downsampling. In the decoder path, we use the PixelShuffle layer (Shi and Caballero, 1874) for upsampling instead of the deconvolution layer and add the encoder and decoder features rather than concatenating them to reduce computation. A diagram of our architecture is shown in Figure 3. A Dropout Block layer with a batch normalization layer after each convolution layer is included after each RBS block. In the following section, we will provide a more in-depth discussion of the core concepts that form the basis of our architecture and how they contribute to the overall design.

2.2 Resnet Branches Shuffle Block

In this paper, we present an improved feature extraction method for image classification tasks. Our approach is based on the Res2Net (Gao et al., 2019) and ShuffleNetV2 (Ma et al., 2018) architectures, and aims to enhance the feature extraction capabilities of the ShuffleNetV2 basic unit.

To this end, we propose the RBS block (Figure 4), which modifies the number of split channels in the input feature maps of the ShuffleNetV2 basic unit. The input, which has n channels, is divided into n/4 groups with the same number of channels each, denoted as

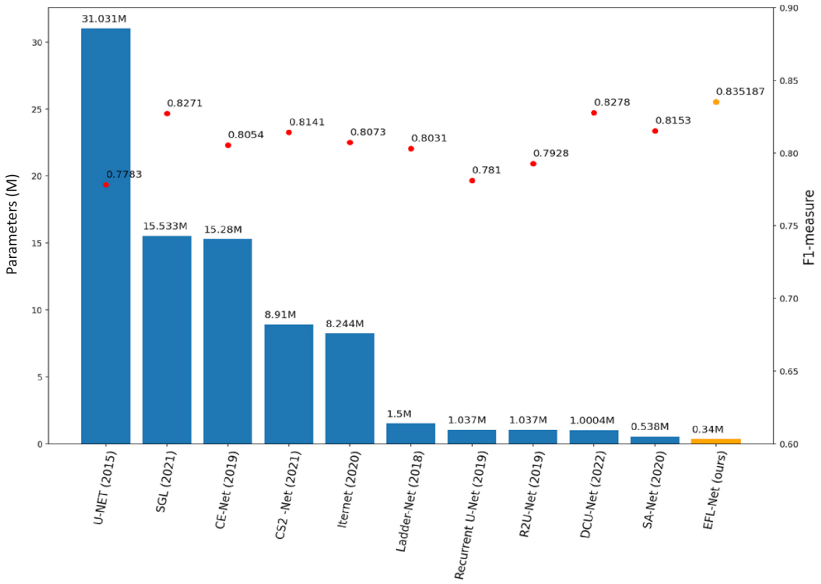

Figure 2: Accuracy and number of parameters of several retinal vessel segmentation papers in the past five years (CHASEDB1 dataset (Owen et al., 2009)).

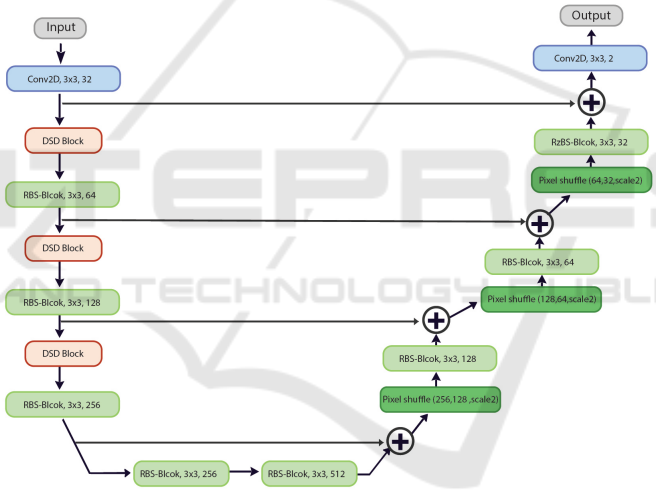

Figure 3: EFL-Net Architecture.

xi, where $i \in \{1,2,3,4\}$. Each xi has the same spatial size and is passed through a bottleneck unit Ci(), resulting in the output yi. The output of the previous bottleneck unit, C$i-1$(), is then added to the current group xi and passed through the bottleneck unit Ci(). This results in an output with a larger receptive field than xi. In addition, the channels xi for $i > 1$ are aggregated with x1 to reuse features. The remainder of the RBS block is identical to the ShuffleNetV2 convolution block.

RBS block can be easily incorporated into any network as a lightweight feature extractor. Our experimental results demonstrate the superiority of the RBS block over the original ShuffleNetV2 basic unit in terms of multi-scale feature extraction and the number of parameters. As shown in Figure 4, the architecture consists of several blocks, which we will describe in detail in the following section.

2.3 ShuffleNetV2 Basic Unit

The basic unit in ShuffleNetV2 (Ma et al., 2018) is a block of layers that includes a depthwise separable convolution (Chollet, 2017), a pointwise convolution (Chollet, 2017), and a shuffle operation (Zhang et al., 2018). This unit is used to construct larger network architectures in a way that reduces computational complexity while maintaining representational capacity. The use of depthwise separable convolutions and the shuffle operation also make the network

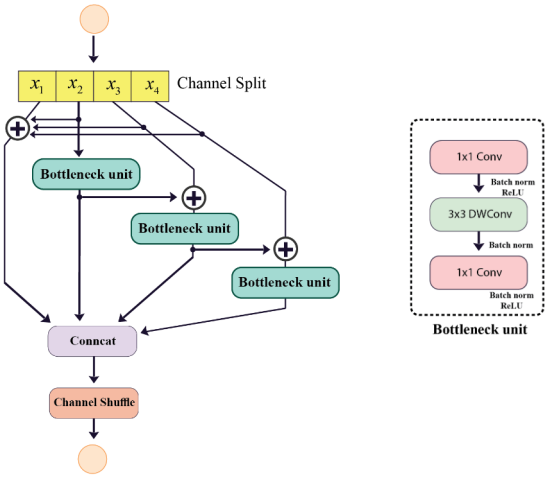

Figure 4: a) ShuffleNetV2 basic unit (Ma et al., 2018), b) Resnet Branches Shuffle Block (RBS block), DWConv stands for depth-wise convolution.

more efficient by reducing the number of parameters and computations required.

2.4 Depthwise Separable Convolution

A depthwise separable convolution (Chollet, 2017) is a way of decomposing a standard convolution in a convolutional neural network into two smaller convolutions: a depthwise convolution that applies a separate kernel to each input channel, and a pointwise convolution that mixes the output channels as shown in Figure 5) . This can reduce the number of parameters and computations in a model and make it more efficient, particularly for mobile and embedded applications.

2.5 Channel Shuffle

A channel shuffle (Zhang et al., 2018) operation is a way of rearranging the channels of a feature map by interleaving them into groups. It is used to allow a convolutional neural network to mix and combine the information in different channels more flexibly and is often used with depthwise separable convolutions to increase representational capacity while maintaining efficiency.

2.6 Dilated Separable down Block (DSD)

The receptive field (RF) is a crucial concept in the design of convolutional neural networks (CNNs). As described in (Luo et al., 2016), the RF at each layer is the size of the region in the input that contributes to generating a particular feature in the output. In order to accurately predict the boundaries of objects in the input image, such as organs, tumors, or vessels, it is necessary to provide the model with access to all relevant parts of the image. In a CNN, each neuron controls a specific part of the data and is exposed to different parts of the input data during the convolution process, filling a segmented area known as the local receptive field. In this paper, we propose the use of the Dilated Separable Down block (DSD block) as a method for increasing the RF of the network and improving its ability to predict object boundaries.

The Dilated Separable Down block (DSD) block, as shown in Figure 6, consists of two branches of 3x3 group convolutional layers with a stride size of 2 and a pointwise convolutional layer. Note that different dilation rates are applied to the different groups to extract multi-scale features. The results of the two branches are concatenated to improve the ability of the network to represent features. This block can be used in place of pooling layers to increase the expressive power of the model. The use of dilated convolution allows the model to increase its field of view without increasing the number of parameters.

2.7 PixelShuffle

PixelShuffle (Shi and Caballero, 1874) is a type of upsampling layer used in convolutional neural networks (CNNs) to increase the resolution of the output feature maps. It has the advantage of being able to achieve a higher resolution output than other upsampling methods and being more efficient, as it does not require the use of additional convolutional kernels or the insertion of zeros into the feature maps. These

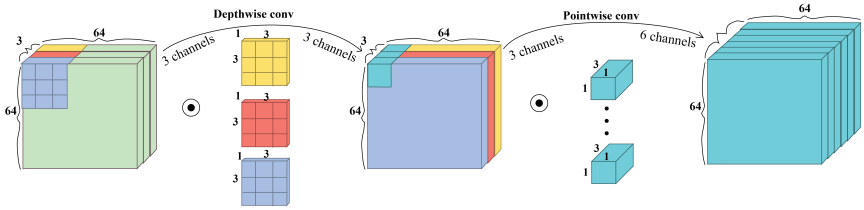

Figure 5: Depthwise separable convolution (Pandey, 2018).

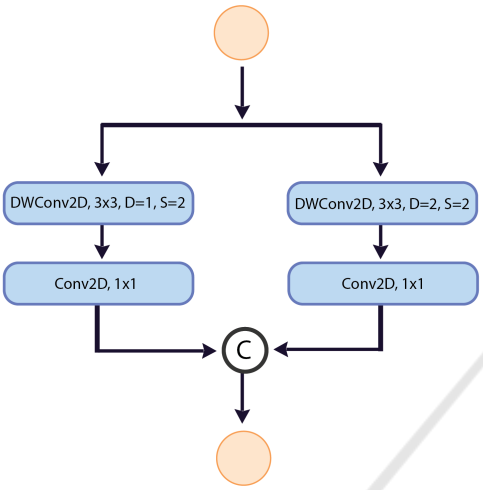

Figure 6: Dilated Separable Down Block.

properties make PixelShuffle a useful tool for tasks that require a high degree of spatial resolution and for use in mobile and embedded applications where computational resources are limited.

3 EXPERIMENTAL ENVIRONMENT AND RESULT

3.1 DATASETS and Data Preprocessing

The DRIVE (Owen et al., 2009), and CHASEDB1 (Owen et al., 2009) are both publicly available datasets for retinal segmentation. The DRIVE dataset consists of 40 2D RGB images with a resolution of 565 x 584 pixels, with 20 images in both the training and test sets. The CHASEDB1 dataset includes 28 images, and has a resolution of 999 x 960 pixels. There are 20 training images and 8 test images in the CHASEDB1 dataset. The model's performance was evaluated on both datasets using the ground truth labels provided by the first expert.

We enhanced the size of the dataset by implementing data augmentation techniques. To focus on the relevant information and eliminate unnecessary processing, we employed a mask or field of view (FOV) to extract patches from the input image that only contained vessels. While the DRIVE dataset includes a binary mask, the CHASEDB1 dataset does not. Therefore we manually created a mask for the CHASEDB1 dataset.

Before training, the data was pre-processed to remove noise and uneven lighting in fundus images. The green channel of the RGB image was chosen as it allows for better visualization of blood vessels. The data was then normalized and scaled, and contrast limited adaptive histogram equalization (Zuiderveld, 1994) and gamma adjustment were applied to improve the contrast between the foreground and background

3.2 Evaluation Approaches

The performance of a segmentation model can be evaluated by comparing its results to the ground truth (GT) and considering four scenarios: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). TP is the number of correctly classified blood vessel pixels, FP is the number of incorrectly classified background pixels as vessels, FN is the number of incorrectly classified vessel pixels as background, and TN is the number of correctly classified background pixels. In addition to these four indicators, the model's performance can also be evaluated using the following criteria: sensitivity (SE), specificity (SP), accuracy (ACC), precision (Pr), and F-Measure (F1).

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Table 1: Performance comparison between the EFL-Net and some state-of-the-art models on DRIVE.

Methods	Year	F1	SE	SP	Acc	AUC	Parameters (M)
U-NET (Ronneberger et al., 2015)	2015	0.7783	0.8288	0.9701	0.9578	0.9772	31.031
SGL (Zhou et al., 2021)	2021	0.8271	0.869	0.9843	0.9771	0.992	15.533
CE-Net (Gu et al., 2019)	2019	0.8054	0.8093	0.9797	0.9641	0.9834	15.28
CS2-Net (Mou et al., 2021)	2021	0.8141	0.8329	0.9784	0.9651	0.9851	8.91
Iternet (Li et al., 2020)	2020	0.8073	0.797	0.9823	0.9655	0.9851	8.244
Ladder-Net (Zhuang, 2018)	2018	0.8031	0.7978	0.9818	0.9656	0.9839	1.5
Recurrent U-Net (Zahangir Alom et al., 2018)	2019	0.781	0.7459	0.9836	0.9622	0.798	1.037
R2U-Net (Zahangir Alom et al., 2018)	2019	0.7928	0.7756	0.982	0.9634	0.9815	1.037
DCU-Net (Yang et al., 2022)	2022	**0.8278**	0.8075	0.9841	0.9664	0.9872	1.0004
SA-Net (Hu et al., 2021)	2020	0.8153	0.8573	0.9835	0.9755	0.9905	0.538707
EFL-Net (ours)	2022	**0.8242**	0.7957	**0.9802**	0.9567	0.9803	0.340

Table 2: Performance comparison between our EFL-Net and some state-of-the-art models on CHASEDB1.

Methods	Year	F1	SE	SP	Acc	AUC	Parameters (M)
U-NET (Ronneberger et al., 2015)	2015	0.8174	0.7537	0.982	0.9531	0.9755	31.031
BCDU-Net (d=3) (Azad et al., 2019)	2019	0.8224	0.8007	0.9786	0.956	0.9789	20.659
SGL (Zhou et al., 2021)	2021	0.8316	0.838	0.9834	0.9705	0.9886	15.533
CE-Net (Gu et al., 2019)	2019	0.8243	0.8276	0.9735	0.9545	0.9794	15.28
CS2-Net (Mou et al., 2021)	2021	0.8228	0.8154	0.9757	0.9553	0.9784	8.91
Iternet (Li et al., 2020)	2020	0.8205	0.7735	0.9838	0.9573	0.9816	8.244
Ladder-Net (Zhuang, 2018)	2018	0.8202	0.7856	0.981	0.9561	0.9793	1.5
Recurrent U-Net (Zahangir Alom et al., 2018)	2019	0.8155	0.7751	0.9816	0.9556	0.9782	1.037
R2U-Net (Zahangir Alom et al., 2018)	2019	0.8171	0.7792	0.9813	0.9556	0.9784	1.037
DCU-Net (Yang et al., 2022)	2022	0.8272	0.8115	0.978	0.9568	0.981	1.0004
SA-Net (Hu et al., 2021)	2020	0.8263	0.8212	0.984	0.9698	0.9864	0.538707
EFL-Net (ours)	2022	**0.8351**	0.7977	0.9860	0.9651	**0.9868**	0.340

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Specificity (SP) is the ratio of correctly segmented background pixels to the total number of actual background pixels, while sensitivity (SE) is the ratio of correctly segmented blood vessel pixels to the total number of actual blood vessel pixels. Accuracy (Acc) shows the percentage of total image pixels that were correctly segmented. Precision measures the quality of the model's positive predictions, while recall measures the quality of negative predictions. A higher precision value indicates that the model's architecture is better trained on the given data. F1 is the weighted harmonic mean of precision and recall.

3.3 Loss Function

The focal loss (Lin et al., 2017) is a method for addressing the class imbalance between foreground and background pixels in a dataset, and is defined as follows:

$$FL(\mathrm{p}_t) = -\alpha_t * (1-\mathrm{p}_\mathrm{t})^\gamma * log(\mathrm{p}_t)\, where \begin{cases} p & if\, y = 1 \\ 1-p & else \end{cases} \quad (1)$$

In the focal loss, the predicted probability of the network output is denoted by p and the focusing parameter, which can be adjusted, is denoted by γ. Samples that are easy to classify contribute less to the loss values, while samples that are difficult to classify contribute more, causing the model to focus more on the latter.

3.4 Experimental Environment and Parameters

In this work, we trained our network from scratch for 200 epochs using a batch size of 256. We initialized the weights with random values and used the Adam optimizer with default parameters and an initial learning rate of 0.001. The learning rate was updated at each epoch using a cosine function attenuation strategy. Our experiments were conducted on a server with a Linux operating system, 2.30 GHz processor, 128 GB RAM, an NVIDIA TESLA P100 GPU, and the Pytorch 1.7.0 framework.

3.5 Experimental Result

Tables 1 and 2 present a comparison of the performance of our proposed architecture with existing methods on the DRIVE and CHASEDB1 datasets. The results show that our model outperforms other methods on the CHASEDB1 dataset and produces

Table 3: Ablation experiment on RBS and DSD blocks. The EFL-Net is trained and evaluated on CHASEDB1.

Method	F1	SE	SP	Acc	AUC	Parameters (M)
EFL-Net (Standard convolution + maxpooling2D + cross entropy loss)	0.747399	0.6839	0.9815	0.9489	0.82892	1.936738
EFL-Net (Standard convolution + DSD block + cross entropy loss)	0.770642	0.7031103	0.986	0.957	0.979272	1.788498
EFL-Net (ShuffleNetV2 basic unit + DSD block + cross entropy loss)	0.815429	0.739147	0.990829	0.962996	0.985666	0.410922
EFL-Net (RBS block + DSD block + cross entropy loss)	**0.801086**	**0.950963**	0.947362	**0.9648**	**0.988168**	0.340950
EFL-Net (RBS block + DSD block+ Focal loss)	**0.8351**	**0.7693**	0.9891	**0.9648**	**0.9871**	0.340950

satisfactory results on the DRIVE dataset. In particular, our model achieves the highest F1 score of 0.8351 and the highest specificity of 0.9860 on the CHASEDB1 dataset, demonstrating its superiority in retinal vessel segmentation compared to previous works. When comparing the results of our model (EFL-Net) with other networks, it is clear that our network achieves equivalent or better results compared to the best performing networks with a smaller number of parameters.

We conducted ablation experiments on the CHASEDB1 dataset to study the impact of different components of our proposed architecture. In the first experiment, we compared the performance of networks with and without DSD blocks (using MaxPooling instead) in the encoder. The results, shown in rows 1 and 2 of Table 3, indicate that the DSD block is superior to MaxPooling in terms of F1 score.

In the second experiment, we replaced the RBS block with the ShuffleNetV2 basic unit to evaluate the performance of the RBS block in the proposed EFL-Net architecture. By comparing the performance of EFL-Net with the RBS block and DSD block (EFL-Net (RBS block + DSD block)) to EFL-Net with the ShuffleNetV2 basic unit and DSD block (EFL-Net (ShuffleNetV2 basic unit + DSD block)), we found that the RBS block significantly improved the model's F1 score by approximately 1.47%.

In the third experiment, we compared the performance of our model using cross entropy loss and focal loss. The results showed that the model using focal loss achieved better performance.

The results of our model on the DRIVE and CHASEDB1 datasets are shown in Figure 7. The original images of retinal vessels are shown in the first column, followed by the output of our network for vessel segmentation in the second column. The third column shows the binary result obtained by applying a threshold to the network output, and the ground truth for each input image is displayed in the last column.

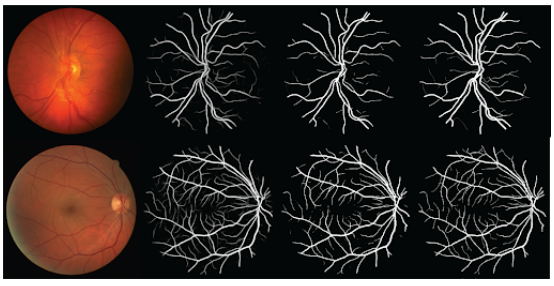

Figure 7: Retinal vessel images from the CHASEDB1 (first row) and DRIVE (second row) datasets.

4 CONCLUSION

EFL-Net is a lightweight network designed to improve the accuracy and speed of blood vessel segmentation. It uses two custom modules, the ResNet Branches Shuffle Block (RBS) and the Dilated Separable Down block (DSD), which have a high capacity for feature extraction. The RBS block is based on the shuffle Net block and the DSD block expands the network's receptive field while reducing feature size without losing important information. In the upsampling path, the network uses PixelShuffle instead of deconvolution or interpolation. The network has 0.34 million parameters and demonstrated good performance on two datasets, achieving F1 scores of 0.8242 on the DRIVE dataset and 0.835187 on the CHASEDB1 dataset.

REFERENCES

Alberta, I. B., Rahmani, S. A., et al. (2022). Retinal impairment associated with long-term use of ritonavir among hiv patients: A systematic review for primary eye care practice. *International Journal of Retina*, 5(1):48–48.

Azad, R., Asadi-Aghbolaghi, M., Fathy, M., and Escalera, S. (2019). Bi-directional convlstm u-net with densley connected convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. (2019). Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662.

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., and Liu, J. (2019). Ce-net: Context encoder network for 2d medical image segmentation.

IEEE transactions on medical imaging, 38(10):2281–2292.

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., and Fan, C. (2021). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In *2020 25th international conference on pattern recognition (ICPR)*, pages 1236–1242. IEEE.

Hu, J., Wang, H., Wang, J., Wang, Y., He, F., and Zhang, J. (2021). Sa-net: A scale-attention network for medical image segmentation. *PloS one*, 16(4):e0247388.

Irshad, S. and Akram, M. U. (2014). Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. In *2014 Cairo International Biomedical Engineering Conference (CIBEC)*, pages 133–136. IEEE.

Kolb, H. (2012). Simple anatomy of the retina.

Li, L., Verma, M., Nakashima, Y., Nagahara, H., and Kawasaki, R. (2020). Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3656–3665.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Luo, W., Li, Y., Urtasun, R., and Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems*, 29.

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131.

Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A. F., et al. (2021). Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874.

Owen, C. G., Rudnicka, A. R., Mullen, R., Barman, S. A., Monekosso, D., Whincup, P. H., Ng, J., and Paterson, C. (2009). Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. *Investigative ophthalmology & visual science*, 50(5):2004–2010.

Pandey, A. (2018). Depth-wise convolution and depth-wise separable convolution.

Payne, A. J., Kaja, S., Naumchuk, Y., Kunjukunju, N., and Koulen, P. (2014). Antioxidant drug therapy approaches for neuroprotection in chronic diseases of the retina. *International journal of molecular sciences*, 15(2):1865–1886.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Salmon, J. F. (2019). *Kanski's Clinical Ophthalmology E-Book: A Systematic Approach*. Elsevier Health Sciences.

Shi, W. and Caballero, J. (1874). Ferenc husz á r, johannes totz, andrew p aitken, rob bishop, daniel rueckert, and zehan wang. 2016. real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Conf. on computer vision and pattern recognition (CVPR)*, volume 1883.

Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., and Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509.

Walker, H. K., Hall, W. D., and Hurst, J. W. (1990). Clinical methods: the history, physical, and laboratory examinations.

Yang, X., Li, Z., Guo, Y., and Zhou, D. (2022). Dcu-net: a deformable convolutional neural network based on cascade u-net for retinal vessel segmentation. *Multimedia Tools and Applications*, 81(11):15593–15607.

Zahangir Alom, M., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv e-prints*, pages arXiv–1802.

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856.

Zhou, Y., Yu, H., and Shi, H. (2021). Study group learning: Improving retinal vessel segmentation trained with noisy labels. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 57–67. Springer.

Zhuang, J. (2018). Laddernet: Multi-path networks based on u-net for medical image segmentation. *arXiv preprint arXiv:1810.07810*.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics gems*, pages 474–485.