






# DR-VNet: Retinal Vessel Segmentation via Dense Residual UNet

Ali Karaali<sup>1,2</sup>(✉) , Rozenn Dahyot<sup>2,3</sup> , and Donal J. Sexton<sup>1,2,4</sup> 

<sup>1</sup> The Irish Longitudinal Study on Ageing (TILDA), School of Medicine, Trinity College Dublin, Dublin, Ireland  
`{karaalia,dosexton}@tcd.ie`

<sup>2</sup> ADAPT: Science Foundation Ireland (SFI) Research Centre for Digital Media Technology, Dublin, Ireland  
`rozenn.dahyot@mu.ie`

<sup>3</sup> Department of Computer Science, Maynooth University, Maynooth, Ireland

<sup>4</sup> Department of Nephrology, St. James's Hospital, Dublin, Ireland

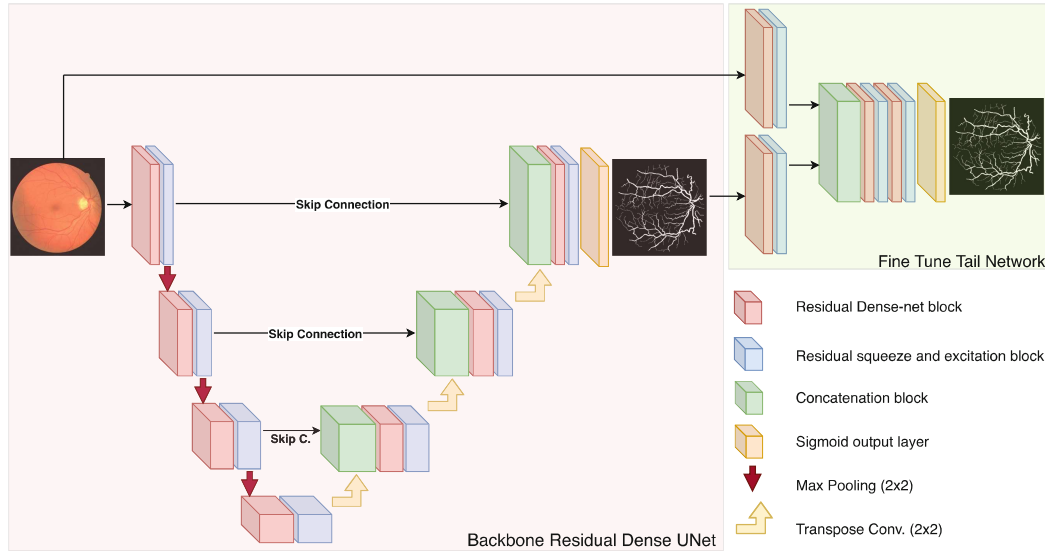
**Abstract.** Accurate retinal vessel segmentation is an important task for many computer-aided diagnosis systems. Yet, it is still a challenging problem due to the complex vessel structures of an eye. Numerous vessel segmentation methods have been proposed recently, however more research is needed to deal with poor segmentation of thin and tiny vessels. To address this, we propose a new deep learning pipeline combining the efficiency of residual dense net blocks and, residual squeeze and excitation blocks. We validate experimentally our approach on three datasets and show that our pipeline outperforms current state of the art techniques on the sensitivity metric relevant to assess capture of small vessels.

**Keywords:** Retinal image · Vessel segmentation · Eye

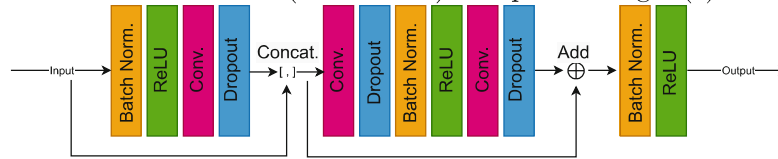
## 1 Introduction

Retinal fundus images have been widely used as a supportive tool to screen, diagnose and treat various systemic diseases, such as cardiovascular disorders [20], kidney diseases [3], and eye-related pathologies, as the retina is the only tissue in the human body where vascular structures can be visualized in a non-invasive manner for clinical examination. Most of the aforementioned diseases might manifest as changes in the morphological structure of retinal blood vessels, consequently various salient and dangerous diseases (e.g. cardiovascular diseases, high blood pressure) can be detected before they create more dangerous and irreversible conditions. Segmentation of the retinal blood vessels is one of the crucial steps for retinal fundus image analysis. However, segmenting the retinal blood vessels manually is a very time consuming and intensive task, and requires not only specific medical training but also technical expertise. In order to mitigate the workload of the health workers, computerized segmentation strategies have garnered great interest in recent years, and many segmentation methods have been proposed [2, 6, 14]. However, one of the main drawbacks of existing retinal

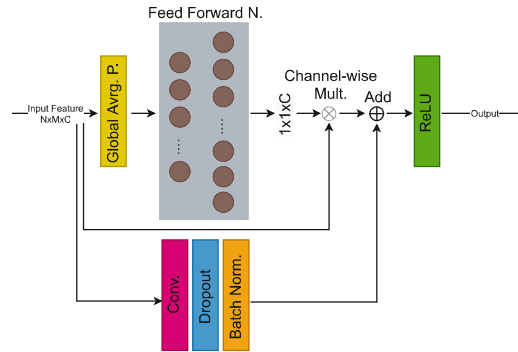
vessel segmentation methods is that they present a poor sensitivity rate where the thin and tiny vessel branches are located, and they tend to be miss-classified by most of the existing methods.



(a) Overview: Residual Dense-Net (pink blocks) is explained Fig. 1(b), Residual squeeze and excitation block architecture (blue blocks) is explained Fig. 1(c).



(b) Residual dense-net (RDN) block architecture.



(c) Residual squeeze and excitation block architecture.

**Fig. 1.** Our proposed architecture uses a modified Backbone Residual Dense UNet that is extended with a Fine Tune Tail Network.

In this paper we propose a supervised method called DR-VNet for retinal vessel segmentation based on Convolutional Neural Network (CNN), aiming to overcome the aforementioned sensitivity rate problem while keeping the other accuracy metrics (e.g. specificity, accuracy, Area Under the ROC Curve (AUC))

at a high level. The technical details of each part of our pipeline are presented in Sect. 3 and validated on three datasets (cf. Sect. 4) against the current leading state of the art methods (reviewed Sect. 2).

## 2 Related Works

Existing retinal vessel segmentation methods can be broadly classified into two categories: supervised and unsupervised methods.

**Unsupervised methods** are mostly rule-based methods, and segmentation is carried out by utilizing the visual or geometric information such as contrast levels, vessel structure and other manually designed features [2].

Bankhead et al. [1] proposed a method based on a modified wavelet transform. More precisely, their method extract blood vessels by filtering the image with the isotropic undecimated wavelet transform (IUWT) and binarizing the filtered image with a percentile-computed threshold. Nguyen et al. [18]’s work, on the other hand, is based on the line operators [21]. Their method amplifies blood vessel pixels by filtering the fundus image with a kernel that enhances the pixels that belong to the lines at different orientations. Recently, Li et al. [14] proposed a method that is based on a deep tube marked point process (MPP) model [13], which has been originally proposed to detect short and/or long tubes in a given hyper-spectral image. As the blood vessel in retinal fundus images are visually similar to tubes, the authors take advantage of the already trained MPP model bypassing the need for labeled images.

**Supervised methods** utilize a group of samples to train a classifier that discriminates the vessel pixels from the background eye tissue, which can be referred as a binary classification problem. These methods are machine learning and/or deep neural network (DNN) based methods.

Soares et al. [23] proposed a supervised segmentation algorithm that utilizes the two dimensional - multi scale Gabor wavelet filter responses as features. Then, three types of classifiers is used to segment vessel pixels: Gaussian mixture model (GMM), K-nearest neighbor (KNN) and least mean square error (LMSE). More recently, Liskowski et al. [15] proposed a CNN based method that consists of consecutive convolutional & max pooling and fully connected layers. The proposed network inputs  $27 \times 27$  image patches, and classifies them whether centered on vessel pixels or not. Moreover, Guo et al. proposed a UNet [22] based method called Structured Dropout UNet [7]. The method is inspired by a recently developed neural block called DropBlock [4], which does not exploit the traditional dropout at convolutional layers, but instead, utilizes a structured dropout block. Their work is extended in [5] by employing a modified residual block structure and a spatial attention block [25]. Later on, Guo et al. [6] introduced a new block called Modified Efficient Channel Attention (MECA). Their module enhances the discriminative capacity of the modified UNet shape architecture by weighting the feature map channels independently.

Following the vast majority of recent methods relying on deep learning, we present next a new deep learning pipeline aiming at improving sensitivity for detecting thin blood vessels (see Fig. 1).

### 3 Proposed Method

Our proposed CNN architecture DR-VNet consists of two cascaded sub-networks: a *Backbone Residual Dense* network and a *Fine-tune Tail* network.

The *Backbone Residual Dense* sub-network is inspired by the well known UNet [22], where three down-sampling and up-sampling blocks are used instead of four, and the original convolutional layers are replaced by two novel blocks: (i) Residual dense-net block (Fig. 1(b)); (ii) Residual squeeze and excitation block (Fig. 1(c)). The *Fine-tune Tail sub-network*, on the other hand, is a shallow network that consists of three consecutive RDN and RSE blocks, which combine the output of the *Backbone Residual Dense* sub-network and the input retinal image. The aim is to fine-tune the output of the *Backbone Residual Dense* sub-network to produce the final retinal vessel segmentation image.

**Residual dense-net (RDN) block** is a novel neural architecture that is based on DenseNet [11] and ResNet [8] types of neural structures (see Fig. 1(b)). The proposed neural block consists of two sequentially connected sub-blocks that utilize dense and residual connectivity patterns.

The first sub-block, which employs the dense connectivity pattern, is based on the idea of connecting a layer’s output to the following layer’s input in a feed-forward fashion, as in [11]. In contrast to summation as in a residual block, the connection is performed by concatenation. To do that, we define a composite function of  $\mathcal{H}$  that consists of consecutive Batch Normalization, ReLu, Convolution and Dropout operations, and applied to the input features  $X$ . More formally, the process can be defined as  $y_d = [\mathcal{H}(X; \{W_i\}), X]$ , where  $[\cdot]$  represents the concatenation operation, and  $y_d$  is the output of the first sub-block. Although this sub-block might have more than one composite function  $\mathcal{H}$  (i.e. a deeper architecture), we opt to use a single unit to keep the network light-weight.

The second sub-block, on the other hand, employs the residual connectivity pattern, which consists of a mapping function  $\mathcal{F}$  with 2-layers of Convolution, Dropout, Batch Normalization, and ReLu operations (see Fig. 1(b)). More precisely, the RDN block is defined as  $y_{RDN} = \delta(\mathcal{BN}(\mathcal{F}(y_d; \{W_i\}) + y_d))$ , where  $y_{RDN}$  represent the output of the RDN block,  $\mathcal{BN}$  represents the Batch Normalization operation, and  $\delta$  represents the second non-linearity function ReLu. This block is designed as a feature extractor inspired by DenseNet and ResNet neural architectures, as they present a good leverage for many CNN tasks [8, 11].

**Residual squeeze and excitation (RSE) block** is inspired by ResNet [8] and a recently developed Squeeze and Excitation (SE) types of neural architectures [10]. In general, SE neural blocks aim at exploring channel interdependencies so as to selectively emphasize informative channels and suppress less useful ones. Residual blocks are designed to smooth the information flow across the layers so as to facilitate the optimization process. Given the success of ResNet blocks [8] and SE blocks [10], we propose to combine ResNet onto SE block to design a novel neural block (Fig. 1(c)) to enhance the extracted features and to facilitate the optimization process.

The proposed RSE block consists of two parallel branches:

1. a SE computational branch that transforms the given input  $X \in \mathbb{R}^{N' \times M' \times C'}$  to a calibrated feature map  $\hat{X} \in \mathbb{R}^{N \times M \times C}$ ,
2. and a standard convolutional block with dropout and batch normalization layers that represents the residual mapping function for the RSE block.

More formally, we define the RSE block as  $y_{RSE} = \delta(\mathcal{F}(X; \{W_i\}) + \hat{X})$ , where  $X$  and  $y_{RSE}$  represent the input and output feature maps of the RSE block,  $\hat{X}$  represents the output of SE computational branch,  $\mathcal{F}(X; \{W_i\})$  illustrates the residual mapping function with weights  $W_i$ , and  $\delta$  is the final non-linearity function ReLu. In this work, we use a single  $\mathcal{F}$  composite function of Convolution, Dropout and Batch Normalization respectively.

In the SE computational branch, firstly, the input features undergo a squeeze operation so as to shrink the feature maps across their spatial dimensions and produce a channel descriptor vector  $v \in \mathbb{R}^{C'}$  that defines a statistic for each channel of the input features. This is accomplished by using the global average pooling operation:

$$v_c = \frac{1}{N \times M} \sum_k \sum_l X_c(k, l), \quad (1)$$

where  $k$  and  $l$  represent the spatial locations,  $c$  is the channel of interest,  $N$  and  $M$  are the spatial dimensions. The channel-wise statistic vector  $v$  is then re-calibrated through two fully-connected (FC) layers of one hidden layer of size  $\frac{C}{r}$  with ReLU activations and one output layer of size  $C$  with Sigmoid activations. Formally, the recalibration is obtained via a simple gating mechanism  $u_c = \sigma(W_2 \delta(W_1 v))$ , where  $\sigma$  and  $\delta$  represent the Sigmoid and ReLu activation functions, and  $W_1$  &  $W_2$  are the weights of the fully connected layers respectively. Then, the output of the SE branch is obtained by rescaling the input features with the re-calibrated channel-wise statistic vector  $\hat{X} = X \otimes u_c$ , where  $\otimes$  refers to channel-wise multiplication. This neural block is proposed as a transition block in order to modify the weights of each channel of feature maps so that the informative channels can be emphasized further during the information flow in the network.

**The network architecture** is designed as a light-weight encoder-decoder type neural network (Fig. 1(a)). The network contains two consecutive cascaded sub-networks. First sub-network, as noted previously, is based on the famous UNet architecture, which is referred as *Backbone Residual Dense* sub-network; and the second sub-network is based on a shallow CNN, which is referred as *Fine-tune Tail* sub-network.

The *Backbone Residual Dense* sub-network consists of three down-sampling layers, a latent layer, three up-sampling layers and an output block. Each layer contains a RDN block and a transition RSE block. RDN block is carried out as a feature extractor by using the DenseNet and ResNet connectivity patterns [8, 11]. RSE block is utilized as a transition block that modifies the weights of each channel so as to emphasize the informative ones further. None of these blocks alter the spatial resolution of the input feature maps. Spatial resolution is changed via down-sampling layers by utilizing max-pooling layers of  $2 \times 2$  kernel size with 2 pixel

stride in the encoder side, and is conducted through transpose convolutions in the decoder side. The output block, which contains a single convolution operation of  $1 \times 1$  filters with Sigmoid activation, receives the output of the last up-sampling layer, and yields an initial estimate for the vessel map.

The *Fine-tune Tail* sub-network is a very shallow CNN block, which is designed to fine-tune the initial estimates of the *Backbone Residual Dense* sub-network. The fine-tuning operation is conducted as follows: first, a single-layer of RDN and RSE neural block is applied to the input and output of *Backbone Residual Dense* sub-network; then, outputs are concatenated and sent to a 2-layer of RDN and RSE neural block; and finally, the output block, which has similar architecture as the output block of the backbone sub-network, is fed-forward by the previous layer. The vessel map is then extracted by thresholding.

## 4 Experiments

**Datasets:** Three publicly available datasets of colour retinal images have been used to evaluate the proposed network. The DRIVE<sup>1</sup> dataset [24] consists of 40 retinal images (resolution  $565 \times 584$  pixels), which are divided into training and test sets, and each set contains 20 images. The CHASE DB<sup>2</sup> dataset [19] contains 28 retinal images with a resolution of  $999 \times 960$  pixels, where they are acquired from both the left and right eye of 14 children. The STARE<sup>3</sup> dataset [9] consists of 20 retinal images with a resolution of  $700 \times 605$  pixels. Each dataset provides expert annotations for the retinal vessels, and these annotations are used to train the network and to quantitatively evaluate the results.

**Implementation Details and Training Procedure:** Given a color retinal fundus image  $I$ , our algorithm starts with zero padding it in the four margins to a set size  $H \times W$ . Then, the zero padded images are fed-forward to the retinal vessel segmentation network in order to train the network in two phases:

1. In the first phase, we train the *Backbone Residual Dense* sub-network, using an initial learning rate of  $10^{-3}$  that is divided by 10 at every 50 epochs, converging after 150 epochs.
2. In the subsequent phase, we train the full network by cascading the *Fine-tune Tail* sub-network to the backbone sub-network, using a similar learning rate strategy. In this phase the weights from *Backbone Residual Dense* sub-network are frozen, and the remaining weights in the *Fine-tune Tail* sub-network are learned in 100 epochs.

Our pipeline is implemented using Tensorflow, with a composite loss function that consists of a weighted sum between the binary cross-entropy  $L_b$  and the Dice loss  $L_d$  [16]:

$$\mathcal{L} = \lambda_1 L_b + \lambda_2 L_d, \quad (2)$$

<sup>1</sup> <https://drive.grand-challenge.org/>.

<sup>2</sup> <https://www.kaggle.com/khoongweihao/chasedb1>.

<sup>3</sup> <https://cecas.clemson.edu/~ahoover/stare/>.



where  $\lambda_1$  and  $\lambda_2$  are the weighting parameters for the loss functions. We used  $\lambda_1 = 1$  and  $\lambda_2 = 0.5$  for both training phases. The scaling ratio for all RSE blocks has been set to  $r = 2$ . As for training time, the convergence is accomplished less than 5 h on average for each dataset on a Tesla K40m NVidia GPU.

Data augmentation with random rotations, horizontal, vertical and diagonal flips are applied to all the training images of the three datasets. Adam optimizer is utilized with a batch size of 2 for DRIVE and STARE datasets and a batch size of 1 for CHASE DB. For the size adjustment, we use  $H = W = 592$ ,  $H = W = 1008$  and  $H = W = 704$  for DRIVE, CHASE DB and STARE datasets respectively as in [6]. Finally, the datasets are partitioned as follows:

1. DRIVE, 90-10 % train-validation data separation is used from the training set of 20 images, and a testing set of 20 images are used for testing;
2. CHASE DB, first 20 images are utilized for training-validating (90-10) %, and the last 8 images are used for testing;
3. STARE, following the previous models [6,17], we adopt a 4-fold cross-validation strategy for training and testing.

The segmented vessel images (output of our pipeline) are cropped back to the original size, then a thresholding operation is applied, and finally the results are compared with the expert annotations provided with the datasets. More precisely, a vessel pixel at location  $x$  in the output image  $I_o$  is validated if the corresponding pixel location has a higher confidence value than a certain threshold  $T$ , i.e., if  $T < I_o(x)$  (we have set  $T = 0.5$ ). The code and results are shared online<sup>4</sup>.

**Quantitative Evaluation:** Table 1(a), 1(b) and 1(c) summarize the average results of a multiple run of our proposed approach along with the standard error and competitive state of the art methods [6, 17, 26] on the aforementioned three datasets. The following metrics are reported: Specificity (Sp), Sensitivity (Se), Accuracy (Acc), Area Under the ROC Curve (AUC), and the G-mean ( $G = \sqrt{Se \times Sp}$ ) to quantitatively compare our retinal vessel segmentation approach with other competitive methods.

We first illustrate the quantitative results for DRIVE dataset in Table 1(a). As it can be seen from the table, our approach achieves the highest sensitivity rate among the all other approaches (3.7 % better than the second best [6]), while keeping the other evaluation metrics at a promising level such as specificity, accuracy and AUC. Furthermore, our approach presents the highest G-mean score, which indicates that the proposed approach has a promising trade-off between specificity and sensitivity rates. Likewise the quantitative results for CHASE dataset in Table 1(b) shows that our proposed approach presents the best performance in terms of Sensitivity rate by presenting 6.8 % higher value than the second best [6] with a significant rise in G-mean score, while keeping the other metrics at a desired level among all competing methods. Notably, standard

<sup>4</sup> <https://github.com/alikaraali/DR-VNet>.

**Table 1.** Comparison with state of the art approaches (Vessel-Net [26], 2019; DDNet [17], 2020; CAR-UNet [6], 2021) on the three datasets: DRIVE (Table 1(a)), CHASE DB (Table 1(b)) and STARE (Table 1(c)). Each result is reported as an average over 5 runs for DRIVE and CHASE DB datasets, and 4-Fold Cross Validation for STARE dataset along with the standard errors. Our pipeline outperforms significantly other approaches systematically for Sensitivity (Se) and G-mean scores, while maintaining excellent performances for Specificity (Sp), Accuracy (Acc) and Area Under the ROC Curve (AUC).

(a) Quantitative evaluations on the DRIVE dataset.

Method	Sp $\uparrow$	Se $\uparrow$	Acc $\uparrow$	AUC $\uparrow$	G-mean $\uparrow$
Vessel-Net [26]	0.9802	0.8038	0.9578	0.9821	0.8876
DDNet [17]	0.9788	0.8126	0.9594	0.9796	0.8918
CAR-UNet [6]	<b>0.9849</b>	0.8135	<b>0.9699</b>	<b>0.9852</b>	0.8951
DR-VNet	0.9795	<b>0.8512</b>	0.9682	0.9848	<b>0.9127</b>
stderr $\times 10^{-3}$	0.5	3.2	0.2	0.5	1.4

(b) Quantitative evaluations on the CHASE DB dataset.

Method	Sp $\uparrow$	Se $\uparrow$	Acc $\uparrow$	AUC $\uparrow$	G-mean $\uparrow$
Vessel-Net [26]	0.9814	0.8186	0.9661	0.986	0.8963
DDNet [17]	0.9773	0.8268	0.9637	0.9812	0.8989
CAR-UNet [6]	<b>0.9839</b>	0.8439	<b>0.9751</b>	<b>0.9898</b>	0.9112
DR-VNet	0.9733	<b>0.9120</b>	0.9694	0.9884	<b>0.9421</b>
stderr $\times 10^{-3}$	0.1	6.3	1.0	0.3	2.5

(c) Quantitative evaluations on the STARE dataset.

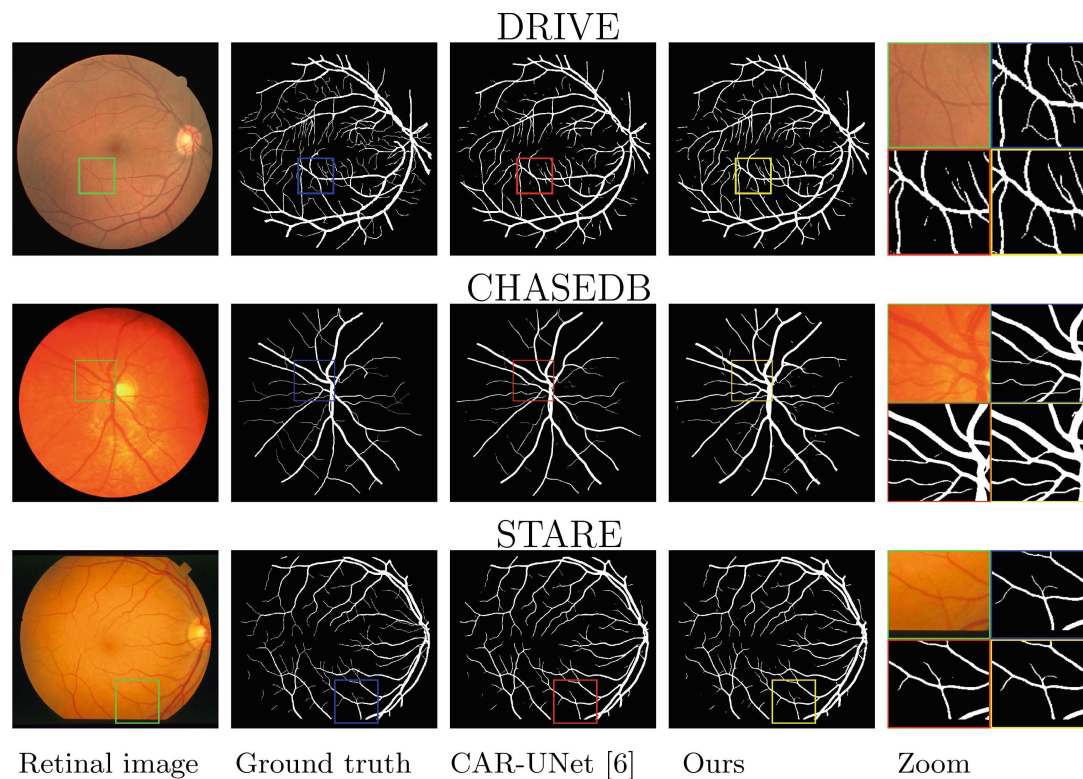
Method	Sp $\uparrow$	Se $\uparrow$	Acc $\uparrow$	AUC $\uparrow$	G-mean $\uparrow$
DDNet [17]	0.9769	0.8391	0.9685	0.9858	0.9053
CAR-UNet [6]	<b>0.9850</b>	0.8445	<b>0.9743</b>	<b>0.9911</b>	0.9097
DR-VNet	0.9841	<b>0.8572</b>	0.9744	0.9847	<b>0.9183</b>
stderr $\times 10^{-2}$	0.1	3.4	0.2	1.0	1.8

errors computed with training our approach 5 times highlight that our results are significant in showing that our approach provides improvements for these metrics for both datasets.

Finally, the quantitative results for the STARE dataset are summarized in Table 1(c). As it can be seen from the table, the SOTA methods present similar scores on this dataset in terms of used evaluation metrics. However, the proposed approach yields again a significantly higher sensitivity rate and the G-mean score on average when compared to the state of the art methods.

The average running times of our approach for a single retinal image at inference are 0.25 s on the DRIVE dataset, 0.25 s on the STARE dataset, and 0.32 s on the CHASE DB dataset. For comparison the average running time for Guo et al.’s





**Fig. 2.** Retinal vessel segmentation results on three publicly available datasets. From left to right: Retinal image, ground truth, results of CAR-UNet [6], results of the proposed method (Ours), and zooms of patches for the corresponding images (to read from left to right as retinal image (green), ground truth (blue), and in second line, CAR-UNet [6] (red) and our proposed method (yellow)). (Color figure online)

approach [6] is 0.35 s on the DRIVE dataset, 0.35 s on the STARE dataset, and 0.45 s on the CHASE DB dataset and in [17], the average running time reported is 0.14 s for all datasets using a similar hardware configuration to ours.

**Qualitative Evaluation:** Figure 2 shows the original retinal vessel images, expert annotations for vessels (ground truth), results of the competitive method CAR-UNet [6], and results of the proposed method for the three datasets. Although visual analysis is very subjective, it can be observed from the figure that the proposed approach present visually coherent results by performing a clear segmentation not only of the thick vessels but also of around thin/tiny blood vessels, which can be better seen in the last column of Fig. 2, where the zoomed patches are shown for the original retinal image (upper left), ground truth annotation (upper right), result of the competitive method CAR-UNet [6] (bottom left), and result of the proposed method (bottom right).

### Ablation Study

The binary cross-entropy is one of the most widely used loss functions, and works very well for many types of binary classification tasks. On the other hand,

**Table 2.** Effectiveness of the loss function and fine-tuning sub-block on DRIVE dataset. BC. and D. represent the binary cross-entropy and Dice losses respectively.

Method	Sp $\uparrow$	Se $\uparrow$	Acc $\uparrow$	AUC $\uparrow$	G-mean $\uparrow$
Backbone with BC	<b>0.9850</b>	0.8100	0.9696	0.9866	0.8932
Backbone with D	0.9802	0.8467	0.9685	0.9471	0.9110
Backbone with BC&D	0.9811	0.8417	<b>0.9689</b>	<b>0.9868</b>	0.9087
Full Net. with BC&D	0.9793	<b>0.8519</b>	0.9682	0.9855	<b>0.9134</b>

**Table 3.** Effectiveness fine-tuning sub-block on CHASE DB dataset. BC. and D. represent the binary cross-entropy and Dice losses respectively.

Method	Sp $\uparrow$	Se $\uparrow$	Acc $\uparrow$	AUC $\uparrow$	G-mean $\uparrow$
Backbone with BC&D	<b>0.9815</b>	0.8675	<b>0.9743</b>	<b>0.9899</b>	0.9227
Full Net. with BC&D	0.9723	<b>0.9160</b>	0.9688	0.9887	<b>0.9438</b>

the Dice loss [16] is a commonly used loss function for medical imaging related segmentation tasks, which has several advantages over binary cross-entropy loss function, such as that the loss information is evaluated both locally and globally. We have proposed a composite loss function (cf. Eq. 2) that combines the traditional binary cross-entropy loss function with the dice loss and we explore its effectiveness along with the contribution of the fine-tuning sub-network for the DRIVE (largest) dataset.

We conduct experiments by applying different configuration to the loss function to validate which one presents the best result. First, we train the backbone sub-network using only the binary cross-entropy (BC) loss function; second, using only the Dice (D) loss function; third, by combining the binary cross-entropy and Dice loss functions (BC&D); and finally, we trained the backbone and the fine-tuning sub-networks together (i.e. full network) using the proposed composite loss function. It worths noting that experiments are conducted over a single run.

Inference results on the test set are summarized in Table 2: the Dice loss provides a clear improvement for the sensitivity rate (compare the first and second row in Table 2), yet there is a trade-off with other metrics, especially with AUC. We also note that there is a significant improvement in performance when the composite loss function is used (see the sensitivity rate at the third row in Table 2), instead of using either one of them alone. Considering the provided trade-off between sensitivity rate and other evaluation metrics, we opt to use the composite loss function.

Finally, the results obtained by training the full network (freezing the weights of backbone network and training the entire network) are shown at the forth row in Table 2. It can be observed that there is a slight improvement in sensitivity rate when the fine-tuning sub-network is cascaded to the backbone sub-network. Although the other evaluation metrics remain at a similar level when the net-

work is trained without fine-tuning, the proposed network (i.e. the full network) remarkably yields not only the best balance between the G-mean and AUC scores but also conceivable results with respect to the remaining evaluation metrics. The results of this ablation study are likewise confirmed on the CHASE DB dataset (cf. Table 3).

## 5 Conclusion

We have proposed a supervised CNN based pipeline for retinal vessel segmentation that is on par with state of the art approaches on three benchmark datasets for metrics such as accuracy, specificity and AUC, but that outperforms significantly on metrics such G-mean and sensitivity. In practical terms our pipeline is performing better for not only segmenting the larger and thicker retinal blood vessel branches but also the more distal retinal blood vessel branches which are much smaller and thinner. Accurate image interpretation for both thick and thin vessels is important since subtle differences in vessel segmentation patterns may be linked with a specific systemic diseases or cardiovascular risk factors. Future work will look at combining information extracted from segmented images with other biological parameters using data from the Irish Longitudinal Study on Ageing (TILDA) [12] in an effort to develop novel prediction tools for cardiovascular disease and other important conditions.

**Acknowledgments.** This work was partly funded by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (13/RC/2106.P2) and is cofunded by the European Regional Development Fund, and also partly supported by Department of Nephrology, St. James’s Hospital, Dublin Ireland. Dr. Donal J. Sexton is funded by Health Research Board of Ireland: grant number ARPP-P-2018-011.

## References

1. Bankhead, P., Scholfield, C.N., McGeown, J.G., Curtis, T.M.: Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLOS ONE* **7**(3), 1–12 (03 2012)
2. Câmara Neto, L., Ramalho, G.L., Rocha Neto, J.F., Veras, R.M., Medeiros, F.N.: An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. *Expert Syst. Appl.* **78**, 182–192 (2017)
3. Farrah, T.E., Dhillon, B., Keane, P.A., Webb, D.J., Dhaun, N.: The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons. *Kidney Int.* **98**(2), 323–342 (2020)
4. Ghiasi, G., Lin, T.Y., Le, Q.V.: DropBlock: a regularization method for convolutional networks (2018)
5. Guo, C., Szemenyei, M., Yi, Y., Zhou, W., Bian, H.: Residual spatial attention network for retinal vessel segmentation. In: Yang, H., Pasupa, K., Leung, A.C.S., Kwok, J.T., Chan, J.H., King, I. (eds.) *Neural Information Processing*, pp. 509–519. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-63830-6\\_43](https://doi.org/10.1007/978-3-030-63830-6_43)

6. Guo, C., Szemenyei, M., Hu, Y., Wang, W., Zhou, W., Yi, Y.: Channel attention residual U-Net for retinal vessel segmentation. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1185–1189 (2021)
7. Guo, C., Szemenyei, M., Pei, Y., Yi, Y., Zhou, W.: SD-UNet: a structured dropout u-net for retinal vessel segmentation. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 439–444 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
9. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**(3), 203–210 (2000)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
11. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
12. Kearney, P.M., et al.: Cohort profile: the Irish longitudinal study on ageing. *Int. J. Epidemiol.* **40**(4), 877–884 (2011)
13. Li, T., Comer, M., Zerubia, J.: A connected-tube MPP model for object detection with application to materials and remotely-sensed images. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 1323–1327 (2018)
14. Li, T., Comer, M., Zerubia, J.: An unsupervised retinal vessel extraction and segmentation method based on a tube marked point process model. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1394–1398 (2020)
15. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **35**(11), 2369–2380 (2016)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation (2016)
17. Mou, L., Chen, L., Cheng, J., Gu, Z., Zhao, Y., Liu, J.: Dense dilated network with probability regularized walk for vessel detection. *IEEE Trans. Med. Imaging* **39**(5), 1392–1403 (2020)
18. Nguyen, U.T., Bhuiyan, A., Park, L.A., Ramamohanarao, K.: An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recogn.* **46**(3), 703–715 (2013)
19. Owen, C.G., et al.: Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. *Invest Ophthalmol. Vis. Sci.* **50**, 2004–10 (2009)
20. Poplin, R., et al.: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**(3), 158–164 (2018)
21. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans. Med. Imaging* **26**(10), 1357–1365 (2007)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MIC-CAI 2015*, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

23. Soares, J., Leandro, J., Cesar, R., Jelinek, H., Cree, M.: Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* **25**(9), 1214–1222 (2006)
24. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
25. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018*. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
26. Wu, Y., et al.: Vessel-Net: retinal vessel segmentation under multi-path supervision. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11764, pp. 264–272. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32239-7\\_30](https://doi.org/10.1007/978-3-030-32239-7_30)