mathematics

MDPI

Article

Deep Multi-Task Learning for an Autoencoder-Regularized Semantic Segmentation of Fundus Retina Images

Ge Jin [1,†], Xu Chen [1] and Long Ying [2,*,†]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2] School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China
* Correspondence: lying@nuist.edu.cn or lorin_ying@hotmail.com
† The authors contributed equally to this work.

Abstract: Automated segmentation of retinal blood vessels is necessary for the diagnosis, monitoring, and treatment planning of the disease. Although current U-shaped structure models have achieved outstanding performance, some challenges still emerge due to the nature of this problem and mainstream models. (1) There does not exist an effective framework to obtain and incorporate features with different spatial and semantic information at multiple levels. (2) The fundus retina images coupled with high-quality blood vessel segmentation are relatively rare. (3) The information on edge regions, which are the most difficult parts to segment, has not received adequate attention. In this work, we propose a novel encoder–decoder architecture based on the multi-task learning paradigm to tackle these challenges. The shared image encoder is regularized by conducting the reconstruction task in the VQ-VAE (Vector Quantized Variational AutoEncoder) module branch to improve the generalization ability. Meanwhile, hierarchical representations are generated and integrated to complement the input image. The edge attention module is designed to make the model capture edge-focused feature representations via deep supervision, focusing on the target edge regions that are most difficult to recognize. Extensive evaluations of three publicly accessible datasets demonstrate that the proposed model outperforms the current state-of-the-art methods.

Keywords: retinal vessel segmentation; VQ-VAE; edge attention

MSC: 68U10

Citation: Jin, G.; Chen, X.; Ying, L. Deep Multi-Task Learning for an Autoencoder-Regularized Semantic Segmentation of Fundus Retina Images. *Mathematics* **2022**, *10*, 4798. https://doi.org/10.3390/math10244798

Academic Editor: Teng Li

Received: 14 November 2022
Accepted: 9 December 2022
Published: 16 December 2022

1. Introduction

Disruptions to the structure of the retinal blood vessels are common side effects of several disorders. Alterations in the anatomy of the blood vessels in the retina can lead to vascular diseases, such as vascular stenosis, capillary sclerosis, and micro-adenomas, if they go untreated [1]. To improve the prognosis of patients, it is necessary to intervene in the early stage of the disease, and retinal vessel segmentation plays a significant role in the diagnosis of eye-related diseases [2]. Usually, retinal blood vessels are manually segmented by doctors. These morphological data were formerly obtained through eye inspections, which were not only time-consuming and prone to human error but also subjective. Thus, to solve the aforementioned problems, it is necessary to introduce an automatic retinal vessel segmentation method, and this task has attracted numerous research interest [3,4].

In actual application scenarios, automatic retinal vessel segmentation is very challenging for the following reasons. First, the scale and shape of retinal vessels vary greatly. The retinal vessel area occupies 1–20 pixels in the images. As a second point, the semantics of the retinal arteries' anatomy are complex. The optic disc, diseased regions, hemorrhage, and exudates are all sources of potential confusion in retinal fundus imaging. Third, the low contrast between retinal vessels and surrounding tissue makes it difficult for the model to segment targets correctly.

Abundant efforts have been dedicated to tackle the aforementioned challenges. Early research focused on the use of various hand-crafted features to segment retinal blood vessels. Huang et al. [5] created a vessel detection approach and incorporated the measurement findings using Bayesian decision making to provide a confidence value for each blood vessel segment in order to characterize the key properties of retinal vessels. However, the strategy's restricted representation capabilities of the hand-crafted features [6–8] make it ineffective when dealing with very big datasets comprising multiple complex situations.

The dramatic advancements in hardware over the past few years have made deep learning an immediate necessity [9–11]. Ronneberger et al.'s [12] U-Net model has had the largest influence in the area of medical image segmentation. Subsequently, a plethora of ideas was made to enhance performance by incorporating new types of deep supervision [13,14]. The symmetrical U-shaped structure that these models are based on allows for the gradual extraction of contextual information via the continuous link between the convolutional and down-sampling layers. For retinal blood vessel segmentation, U-shape structure-based methods have become the mainstream at present. Researchers tried diverse models [15–19] to preserve local and global semantic information, prevent spatial information loss, tackle varying scales, and segment tiny parts.

Although former U-shaped structure models have achieved outstanding performance in many different scenes, some challenges still need to be handled for the specific retinal blood vessel segmentation task. Challenge (1): Extraction of features from deep stages provides high-level features with rich semantic information but insufficient resolution, while extraction of features from shallow stages yields low-level features with rich spatial details but insufficient global semantic information. Challenge (2): The fundus retina images coupled with precise blood vessel segmentation are relatively rare and hard to access. Because the manual segmentation of the blood vessels is time-consuming and labor-intensive. Moreover, manual annotation is a subjective task, with accuracy impacted by the physician's clinical knowledge and personal bias. Challenge (3): Previous methods did not pay enough attention to the edge information of the target, resulting in low segmentation accuracy of the target edge region. Due to the low contrast and high geometric complexity of the vessel edge, this area is the most difficult part to segment for the retinal blood vessel segmentation task.

In this work, we propose a novel encoder–decoder architecture based on the multi-task learning paradigm to address these challenges and further improve the performance of the vessel segmentation task. To cope with Challenge (1) and Challenge (2), the VQ-VAE (Vector Quantized Variational AutoEncoder) module is incorporated into the U-shaped structure, which can regularize the shared encoding process by reconstructing the input image, elevating the generalization ability of the model when precisely annotated data are relatively rare. Meanwhile, latent feature maps encoding multiple-level semantic information are also generated and fused with the features obtained by the image encoder to further enhance the feature representation. To tackle Challenge (3), an edge attention module is proposed to make the model capture edge-focused feature representations via deep supervision, concerning the target edge regions that are most difficult to recognize. The perception of edge information is enhanced. We have performed comprehensive experiments on DRIVE [20], CHASE-DB1 [21] and STARE [22]. The experimental outcomes demonstrate the proposed method is useful in improving the performance of the model.

To summarize, the contributions and novelty of the present study are highlighted as follows:

1. A novel encoder–decoder architecture based on the multi-task learning paradigm is proposed. The VQ-VAE module branch reconstructs input images to regularize the shared image encoder while generating and integrating hierarchical representations of the input image. This module not only alleviates the challenge caused by limited annotated data but also improves the representation ability of the model.

2. An edge attention module is proposed to learn edge-focused feature representations by deep supervision, which can induce the model to focus on the target edge regions that are most difficult to segment and improve the perception of edge information.
3. Comprehensive experiments are conducted on three public datasets, and experimental results show that our methods can achieve state-of-the-art performance.

The remainder of this study is structured as follows: Section 2 presents the results of the literature review. Section 3 discusses the methodology. The experimental results and analyses are presented in Section 4. Finally, the conclusion is presented in the last section.

2. Related Work

Experts in the field of image processing have devoted more and more attention to the difficulty of segmenting retinal vascular pictures in recent years. Retinal imaging of vascular segmentation is a field that has seen numerous techniques developed. In this article, we provide a taxonomy and comparison of various methods for segmenting the retinal vasculature.

With the successful application of CNN in classification tasks, some researchers subsequently explored its potential for retinal blood vessel segmentation and achieved good performance. Liskowski and Krawiec [23] introduced a six-layer convolutional neural network (CNN) for vascular segmentation in the retina. Before training the model, the training samples were preprocessed with global contrast normalization (GCN) and zero-phase component analysis (ZCA whitening). Samuel et al. [24] proposed a novel network for segmenting retinal blood vessels from retinal fundus images. The method is based on transfer learning with a pre-trained VGG-16 model as its backbone network. Soomro et al. [25] implemented a model for the retinal blood vessel segmentation task. To alleviate visual complexity such as low contrast, uneven lighting, and noise, they applied a morphological operation and principal component analysis (PCA) to preprocess the image. Additionally, a novel post-processing technique was used to eliminate unwanted noise. Wu et al. [26] also applied PCA at the preprocessing stage to reduce the dimension of the input images. The framework was effective overall, but it lacked sophistication due to its lack of small veins and poorly connected blood vessels.

To address the limitations of CNN architectures, Long et al. [27] proposed the fully convolutional networks (FCN) for semantic image segmentation. This model and its variants are widely applied in the field of medical image segmentation. Atli et al. [28] designed an FCN model that up-samples before down-sampling to accommodate both thick and thin blood vessels. To avoid losing contextual information in the training phase, the technique additionally incorporated residual modules. Li et al. [29] built an FCN with skip connections and included active learning. In this case, the proposed model's performance was enhanced through iterative training. In order to avoid the problem of spatial loss of information, Luo et al. [30] developed a size-invariant FCN for retinal blood vessel extraction from retinal images. FCN models have achieved outstanding success in the segmentation of retinal blood vessels. FCN-based retinal vascular segmentation has achieved promising results; however, its predictions often lack crisp boundaries and ignore spatial coherence.

In the field of medical image segmentation, the most commonly used framework is the U-shape based network. It can produce fine segmentation results on small datasets, and further, the local and global semantic information is preserved. Sathananthavathi et al. [31] swapped out the convolutions for Atrous convolutions to broaden the receptive field and, hence, reduce spatial loss of information. The attention gate technique was developed by Li et al. [32] to protect the segmented retinal blood vessels from being masked by irrelevant foreground elements. Similarly, the authors of [33] also used the weighted attention gate approach to filter out irrelevant information. To enhance thin vessel segmentation, Mishra et al. [34] built a basic U-net and used data-aware deep supervision. The average input retinal vascular width was calculated and compared to the effective receptive fields of the different layers to identify the ones that extract vessel features most strongly.

Retinal artery segmentation in real-time using high-resolution images was proposed by Laibacher et al. [35] using a network that makes use of bottleneck modules and bilinear up-sampling to reduce the number of parameters. Adjusting the U-Net architecture with the help of variable receptive fields, Jin et al. [36] were able to decrease vascular missingness due to differences in vessel size, scale, characteristics, and shape in the retina. In [15], Fu et al. introduced a brand new form of network they called Deep Vessel. In order to fully comprehend these multi-level pictorial representations, a multi-layer, deep neural network with a side-output layer was developed. Alom et al. [16] proposed an extension of the U-shape architecture utilizing the power of U-Net, Residual Network, as well as a recurrent convolutional neural network. The authors in [17] combined attention mechanisms and a U-net framework to achieve better results in pancreas image segmentation. CE-Net [18] is able to increase the receptive field and segment smaller blood vessels because it incorporates the residual multi- kernel pooling module and the dense atrous convolution module. For the purpose of segmenting retinal images, Li et al. [37] presented a new approach that is based on a topological vascular tree. To deduce the topological vascular tree from retinal images, the method uses a global graph-based decision with pixel-wide separation and a complete set of node connections. Liu et al. [38] proposed a light-weight network, dubbed as FR-UNet. The model generates the full-resolution representation of features by extending the parallel convolution layer horizontally, and a novel feature fusion module is proposed to aggregate all-scale representations, finally generating the full-resolution representation of features. In the same period, there were two more studies that achieved state-of-the-art results, namely SGL [39] and RV-GAN [40], respectively. The authors of [39] assume that the group truth of the training examples given by clinicians are incomplete and noisy, resulting in a lack of annotations of some vascular fragments. A Study Group Learning (SGL) method is proposed, which includes k-fold cross validation and knowledge distillation, to improve the robustness of the model on noise data, and a learned enhancement map can provide better visualization. A novel Volumetric Memory Network (VMN) is proposed in [41]. The model can automatically segment 3D medical images interactively. First, a user hints at the 2D slice and automatically generates the initial 2D segmentation, then the VMN propagates it to all slices bidirectionally and refines the segmentation. The authors in [42] provided a prototype view of semantic segmentation. A non-parameter scheme based on a non-learning prototype is proposed. The model represents each class as a set of non-learnable prototypes, which only depends on the average characteristics of several training pixels in this class, rather than the previous method to learn the single weight/query vector of each class in a fully parameterized way.

3. Method

The proposed model is described in detail below. First, the overall structure of the whole network in Section 3.1 will be described. The specifics of our proposed methods are then discussed in Sections 3.2 and 3.3.

3.1. Network Architectures Overview

The proposed model is designed based on an asymmetric U-shaped architecture, which consists of the image encoding network, the segmentation decoder and two branches. The backbone network is a modified Resnet-based network, and the workflow of the proposed approach is displayed in Figure 1.

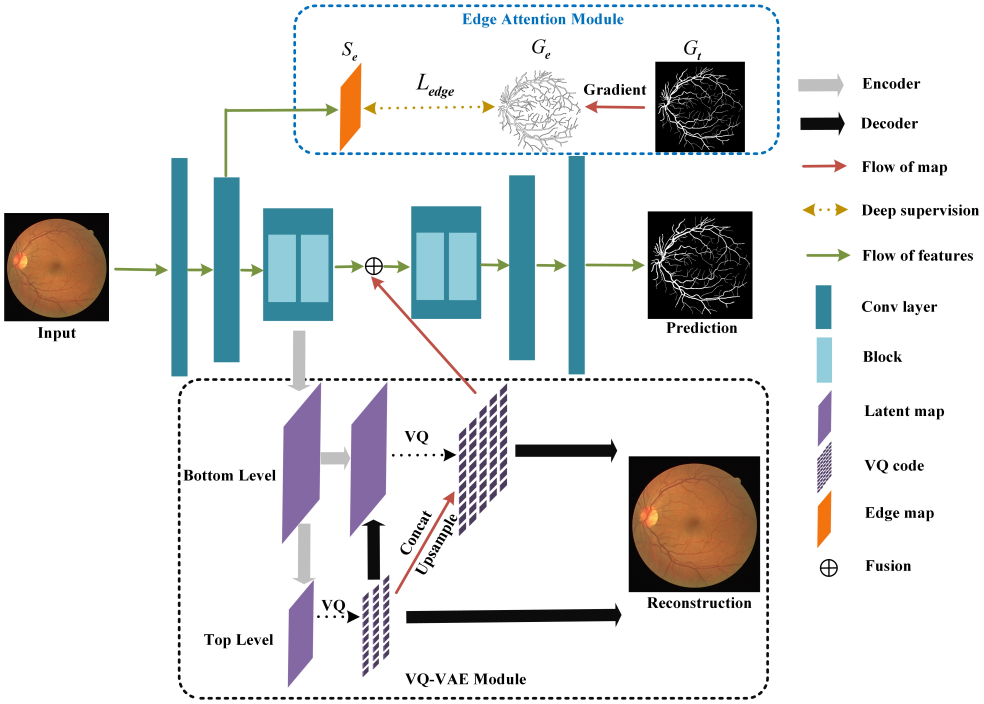

Figure 1. Overview of the proposed model, S_e denotes edge map, G_e denotes edge ground-truth map and G_t denotes ground truth map.

There are two separate branches in the segmentation network, the edge attention module and the VQ-VAE module. Combined with the segmentation decoder, all the branches are trained in a multi-tasking paradigm. First, the retinal blood vessel images are fed into the backbone network. Considering that low-level features retain abundant edge information, the low-level features are fed into the edge attention module. Guided by the gradient of annotated segmentation results, the model can be encouraged to focus on the edge information of the target. Subsequently, the feature maps from the intermediate layer are fed into the VQ-VAE branch to reconstruct the input images and regularize the shared encoding procedure. To further enhance the semantic feature representation, VQ-VAE combines the intermediate feature maps produced by the backbone network with the top-level and bottom-level latent maps produced by VQ-VAE module. The model is trained on three public datasets, i.e., DRIVE, CHASE-DB1 and STARE. Substantial experiments have shown that the proposed approach works.

3.2. Edge Attention Module

Many studies [43–45] have proved that edge information can provide effective constraints in the process of guiding the segmentation of targets. Due to the particularity of retinal fundus images, the edge of vascular tissue is extremely complex and has low contrast with the surrounding background, so the edge of blood vessels is the most difficult to segment. Moreover, the training images of the retinal fundus datasets are very limited, so complex modules may cause overfitting. Unlike [45], we only use the latent features generated from shallow networks and edge ground truth maps, and supervised training is conducted between them to regularize the shared network parameters. This module can improve the sensitivity of the model to edge information by only increasing a very small number of parameters.

It is universally acknowledged that the shallow features in the model contain a lot of low-level information, such as color, texture, and edge information. Therefore, we regularize the shallow shared parameters of the model to induce the model to pay more attention to edge information. The low-level features with the appropriate resolution are fed into the edge attention module to explicitly learn the edge-focused feature representations. Specifically, the low-level features get through a convolution layer and generate edge

maps (S_e). Then, we utilize the standard binary cross entropy loss function to measure the difference between the generated edge maps and the edge ground truth maps (G_e) derived from the ground truth images (G_t). The whole workflow is detailed in Figure 1.

$$L_{edge} = -\sum_{x=1}^{w}\sum_{y=1}^{h}\left[G_e \log(S_e) + (1 - G_e)\log(1 - S_e)\right] \quad (1)$$

where (x, y) are the coordinates of each pixel in the predicted edge map S_e and edge ground-truth map G_e. The S_e is calculated using the gradient of the ground-truth map G_t. Additionally, w and h denote the width and height of the corresponding map, respectively. The edge attention module can improve the sensitivity of the model to edge information and effectively improve the performance of the model.

3.3. VQ-VAE Module

Training a trustworthy and effective segmentation model requires access to a large number of high-quality annotations. Hand delineation for vascular tissue is often unable to obtain appropriate high-quality annotations for medical image segmentation because it is label-intensive and time-consuming. In order to reduce reliance on annotations and improve generalizability, a modified VQ-VAE module is developed based on the input image reconstruction task. Our module's layout was inspired by the winning entry to the BraTS 2018 challenge, submitted by Andriy et al. [46], who developed a Variational Auto-Encoder (VAE) model that was able to overcome a deficiency in training data. The proposed VQ-VAE module provides input reconstruction guidance to the training process, imposing regularization on the shared encoder to produce representations containing more image-intrinsic spatial and semantic structure patterns. Reconstructing the original images requires a sequence of vector quantized codes. The VQ-VAE produces two layers of latent codes, the first of which reflects broad features such as an object's shape and geometry, while the second represents finer details such as its texture. In addition, the two levels of representation are combined into a single, more accurate one using up-sampling and concatenation methods. To further improve the model's performance, the fused characteristics are then utilized in a subsequent network.

The VQ-VAE can be classified as the VAE family, which has an encoder, code, and decoder. The difference is that the code is not directly generated by the encoder, but is obtained through vector quantization. Using fewer resources, the module can recreate images with higher coherence and quality. While the model's foundation is built on probability, it should nevertheless be able to represent the full range of the true distribution and accommodate various forms of input data. Moreover, the VQ-VAE module can generate high-level and low-level latent code, which, respectively, contain the low-level characters (edge information, colors, textures) and semantic knowledge of the input images. Making full use of this automatically extracted knowledge can effectively improve the feature representation ability of the model. Thus, the aforementioned concerns prompted us to implement the VAE-based branch to enhance the model's feature expression and robustness.

An encoder, decoder, and a shared codebook are all housed in the VQ-VAE branch. To do this, the encoder transforms the data into a set of latent variables, which are then used by the decoder to recover the original data. The distance between the input vector x and the prototype vectors $e_k, k \in 1 \ldots K$ is quantized by the encoder's nonlinear mapping, yielding the quantized output vector $E(x)$. Specifically, K possible vectors in the codebook are generated by replacing each vector $E(x)$ with the index of the nearest prototype vector in the codebook. Finally, the indices are transmitted to the decoder, which still uses another nonlinear function to map them back to the codebook vectors to which they originally corresponded, thus reconstructing the data.

$$Quantize(E(x)) = e_k \; where \; k = \arg\min_j \left\| E(x) - e_j \right\| \quad (2)$$

For the VQ-VAE, we use a three-term objective function, as proposed by [47]:

$$L(x, D(e)) = \|x - D(e)\|_2^2 + \|sg[E(x)] - e\|_2^2 + \beta\|sg[e] - E(x)\|_2^2 \quad (3)$$

Quantized code for the training example x is shown by e. The initial component, which indicates the data fidelity term, guarantees accurate reconstructions with small errors. The final two terms are an ingenious addition that brings the encoder's output into line with the codebook's vector space. The second term is utilized in the codebook, where $sg[\cdot]$ means stop gradient. It adjusts the output of the encoder, $E(x)$, to be near to the selected codebook, e. The final term keeps $E(x)$ relatively close to the chosen codebook vector and limits the parameter's variance. In this paper, the β is set to 0.25.

The VQ-VAE module can not only improve the performance of the model but also reduce the dependence of the model on annotations. The reason is that in the process of reconstruction, the encoder implicitly obtains more patterns, while promoting the model to generate more representative features. This module helps the model to consistently achieve good training accuracy for any random initialization and maintains the characteristics generated by the model invariant.

4. Experimental Results

In this section, we detail the experiments of the proposed methodology on three freely available datasets. The available datasets, metrics for measuring progress, and technological issues are briefly outlined before going into the details of the implementation. In addition, studies of ablation are conducted to guarantee that the proposed technique is effective. The experimental findings support the idea that our proposed method is comparable to state-of-the-art models.

4.1. Evaluation Datasets

The proposed retinal vascular segmentation model is evaluated by using data from three publically available databases (DRIVE, STARE, and CHASEDB1). Since FoV masks are not provided by CHASEDB1 and STARE, we built them by ourselves to ensure uniformity across all experiments [48].

- DRIVE: 40 fundus retinal images are included in this dataset. All images were collected by a Canon CR5 nonmydriatic 3CCD camera with a 45-degree field of view (FOV) and cropped from 565 × 584 pixels to 448 × 448 pixels. The dataset contains seven retinal fundus images of diabetes patients. Moreover, the dataset is split into two subsets, that is, training set (20 images) and testing set (20 images).
- CHASEDB1: Each image in this dataset of 28 pictures depicts a vascular patch and has a resolution of 999 × 960 pixels. Fourteen kids' left and right eye retinal fundus images are stored in the database. All images were taken from a 30 degree FOV. The first 20 photos are used as a training set, while the last 8 are utilized as a test set, as stated in [49].
- STARE: In this dataset containing 20 retinal fundus images, half of them have pathological signs. The resolution of the images is 700 × 605. We used 20% of the images as the validation and test sets.

4.2. Evaluation Metric

To evaluate our model more comprehensively, five metrics are introduced for evaluation, including sensitivity (SE), specificity (SP), accuracy (ACC), and the area under the ROC curve (AUC), which are calculated by the following equations:

$$SE = \frac{|TP|}{|TP + FN|} \quad (4)$$

$$SP = \frac{|TN|}{|FP + TN|} \quad (5)$$

$$ACC = \frac{|TP + TN|}{|TP + TN + FN + FP|} \tag{6}$$

True positive (TP) means that the positive sample is correctly classified; false negative (FN) means that the positive sample is wrongly classified as a negative sample; false positive (FP) indicates that the negative samples are wrongly classified as positive samples; true negative (TN) indicates that negative samples are correctly classified as negative samples. In addition, the area under the curve (AUC) of the receiver operating characteristic curve (ROC) is used to evaluate segmentation accuracy based on recall and precision. When a method's area under the curve (AUC) becomes closer to 1, it performs better in segmenting blood vessels.

4.3. Implementation Details

We implemented our network in the Pytorch framework and trained models on an NVIDIA Quadro P5000 GPU with 12 GB memory. Moreover, we utilized the binary cross-entropy loss as the objective function. The Adam optimizer with an initialization learning rate of 1×10^{-3} was applied in the training phase, and the weight decay was set to 0.001. The batch size was set to 2 in the experiments. In order to prevent the gradient from exploding or disappearing during training, we adopted the method proposed by He et al. [50] to initialize the whole network. Moreover, the network was trained for 50 epochs.

Due to the particularity and scarcity of medical image data, some sampling and data augmentation strategies were introduced to prevent overfitting and induce a more generalized network. All three datasets' images were converted into grayscale images and resized to 512×512. Moreover, we conducted uniform normalization on all images, and a 48×48 sliding window with a stride of 6 was introduced to generate patches from vessel images. Then, we conducted horizontal flipping, vertical flipping, and random rotation in the training phase to prevent the model from overfitting and increase the diversity of training samples. In particular, we also used random erase to induce the model to be more sensitive to boundary information. Note that there is no patch extraction operation during the test phase, but full-size images were used as input.

4.4. Segmentation Results

In this work, our model was experimented on three public datasets, including DRIVE, CHASEDB1, and STARE. Results from U-Net [12], DeepVessel [15], R2U-Net [16], AttU-Net [17], CE-Net [18], IterNet [37], SGL [39], and RV-GAN [40] are compared with the proposed method to verify the effectiveness of our method. The outcomes are shown in Table 1.

To quantitatively analyze the experimental results, we perform a statistical comparison based on several important metrics, including SE, SP, ACC, and AUC to evaluate the proposed method and compare it with eight state-of-the-art methods on all three datasets. In addition, we also display the number of parameters of each model in Table 2. As can be seen from Table 1, SGL, RV-GAN, and our method achieved the highest scores on some evaluation metrics, respectively. It shows that our method can achieve comparable performance to state-of-the-art methods. As can be seen in Table 1, the proposed method outperforms most models in four metrics. It achieves a SE of 83.86% in the DRIVE dataset. Some studies [3,39] have verified that higher SE indicates the model is more sensitive to edge information, which further demonstrates that the proposed method has a great ability to focus on microvascular structures. Among the models, RV-GAN as a generative adversarial framework achieves the highest SP, ACC, and AUC on DRIVE and highest SE, and AUC on STARE. The architecture introduces two generators and two multi-scale autoencoding discriminators for better microvessel localization and segmentation. However, due to the complexity of the architecture, it takes a longer time and more computing resources for training to converge the model. The parameters of the network are about 14.8 M, which is about 1.7 times that of our method. Furthermore, the experiments were

conducted on the CHASEDB1 dataset, and the comparison results of different methods are shown in the middle of Table 1. The proposed method obtained the highest value on the SP and ACC with 98.65%/97.80% and a comparable AUC of 98.98%, which is good proof of the validity of the proposed method. It is worth mentioning that SGL achieves the highest SE and AUC on CHASEDB1. The authors proposed a Study Group Learning (SGL) framework to improve the generalization ability of the learned model and better address the missing annotation problems in the training set. It applies the cascade method and cross-validation-based pseudo-label generation strategies, which greatly increases model complexity and training time. Like RV-GAN, the number of parameters of SGL is also large, which is 1.8 times of our model. However, the performance of the proposed method is extremely close to SGL and RV-GAN. Last, the performance comparisons on the STARE dataset are summarized at the bottom of Table 1. The proposed method surpasses other state-of-the-art models in terms of SP and ACC. As can be seen in Table 2, although the number of the proposed method parameters is similar to U-Net, R2U-Net, and AttU-Net, it can obtain much better results. In conclusion, compared with state-of-the-art methods, our model can not always obtain the best result, but it can provide a good trade-off between model complexity and segmentation performance.

Table 1. Comparison with state-of-the-art methods on three classical datasets: DRIVE, CHASEDB1, and STARE.

DRIVE Dataset				
Methods	SE (%)	SP (%)	ACC (%)	AUC (%)
U-Net [12]	79.15 ± 0.23	98.08 ± 0.31	96.40 ± 0.13	97.64 ± 0.08
DeepVessel [15]	78.83 ± 0.18	98.13 ± 0.23	96.09 ± 0.12	97.83 ± 0.05
R2U-Net [16]	79.23 ± 0.56	98.03 ± 0.35	96.54 ± 0.07	98.02 ± 0.08
AttU-Net [17]	78.82 ± 0.17	98.48 ± 0.35	96.49 ± 0.19	98.03 ± 0.07
CE-Net [18]	80.15 ± 0.22	98.16 ± 0.19	96.59 ± 0.16	98.11 ± 0.09
IterNet [37]	79.95 ± 0.26	98.26 ± 0.08	96.57 ± 0.17	98.13 ± 0.06
SGL [39]	83.80 ± 0.09	98.34 ± 0.08	97.05 ± 0.13	98.86 ± 0.04
RV-GAN [40]	79.27 ± 0.10	**99.69 ± 0.11**	**97.90 ± 0.15**	**98.87 ± 0.05**
Ours	**83.86 ± 0.13**	98.37 ± 0.28	97.35 ± 0.08	98.82 ± 0.05
CHASEDB1 Dataset				
Method	SE (%)	SP (%)	ACC (%)	AUC (%)
U-Net [12]	76.17 ± 0.86	98.61 ± 0.69	97.16 ± 0.25	97.92 ± 0.15
DeepVessel [15]	75.84 ± 0.54	98.34 ± 0.54	97.18 ± 0.14	97.85 ± 0.13
R2U-Net [16]	81.45 ± 0.71	98.40 ± 0.71	97.21 ± 0.13	98.01 ± 0.07
AttU-Net [17]	77.21 ± 1.01	98.50 ± 0.98	97.26 ± 0.18	98.07 ± 0.06
CE-Net [18]	80.42 ± 0.39	98.39 ± 0.33	97.23 ± 0.36	98.06 ± 0.09
IterNet [37]	79.97 ± 1.55	98.47 ± 1.05	97.31 ± 0.24	98.26 ± 0.12
SGL [39]	**86.90 ± 0.24**	98.43 ± 0.23	97.71 ± 0.19	**99.20 ± 0.08**
RV-GAN [40]	81.99 ± 0.07	98.06 ± 0.13	96.97 ± 0.24	99.14 ± 0.03
Ours	83.29 ± 0.64	**98.65 ± 0.37**	**97.80 ± 0.07**	98.98 ± 0.10
STARE Dataset				
Method	SE (%)	SP (%)	ACC (%)	AUC (%)
U-Net [12]	78.39 ± 1.36	98.71 ± 0.96	96.88 ± 0.48	97.93 ± 0.15
DeepVessel [15]	78.83 ± 0.94	98.14 ± 0.81	97.13 ± 0.41	98.14 ± 0.11
R2U-Net [16]	78.69 ± 0.99	98.62 ± 0.56	96.97 ± 0.33	98.09 ± 0.09
AttU-Net [17]	79.03 ± 1.06	98.56 ± 0.74	97.22 ± 0.45	98.22 ± 0.10
CE-Net [18]	79.16 ± 0.86	98.53 ± 1.11	97.15 ± 0.25	98.17 ± 0.07
IterNet [37]	80.86 ± 0.53	98.46 ± 0.68	97.23 ± 0.36	98.29 ± 0.07
RV-GAN [40]	**83.26 ± 0.27**	98.64 ± 0.37	97.54 ± 0.15	**98.87 ± 0.06**
Ours	81.35 ± 0.85	**98.74 ± 0.48**	**97.54 ± 0.19**	98.84 ± 0.05

Table 2. Parameter comparison with state-of-the-art methods.

Model	U-Net [12]	R2U-Net [16]	AttU-Net [17]	CE-Net [18]	IterNet [37]	SGL [39]	RV-GAN [40]	Ours
Parameters (M)	7.8	8.3	7.2	14.4	8.6	15.5	14.8	8.8

The example segmentation results on three databases are also shown in Figure 2. It can be observed that the proposed method is sensitive to the edge region of the blood vessels with low contrast and has achieved good segmentation results. In particular, we also display the bad segmentation examples in Figure 3. The whole training phase is a multi-task learning process; one task is semantic segmentation, and the other is the reconstruction task. In the experiment, we found that if the number of epochs is set too large, the performance of segmentation will decline significantly, and the entire model is more suited to reconstruction task as it even fails to converge. Thus, the number of epochs is set to 50. To obtain a better view of our model, we also drew the loss curve and ROC curve in Figure 4. As can be observed, the model converges well within 50 epochs.

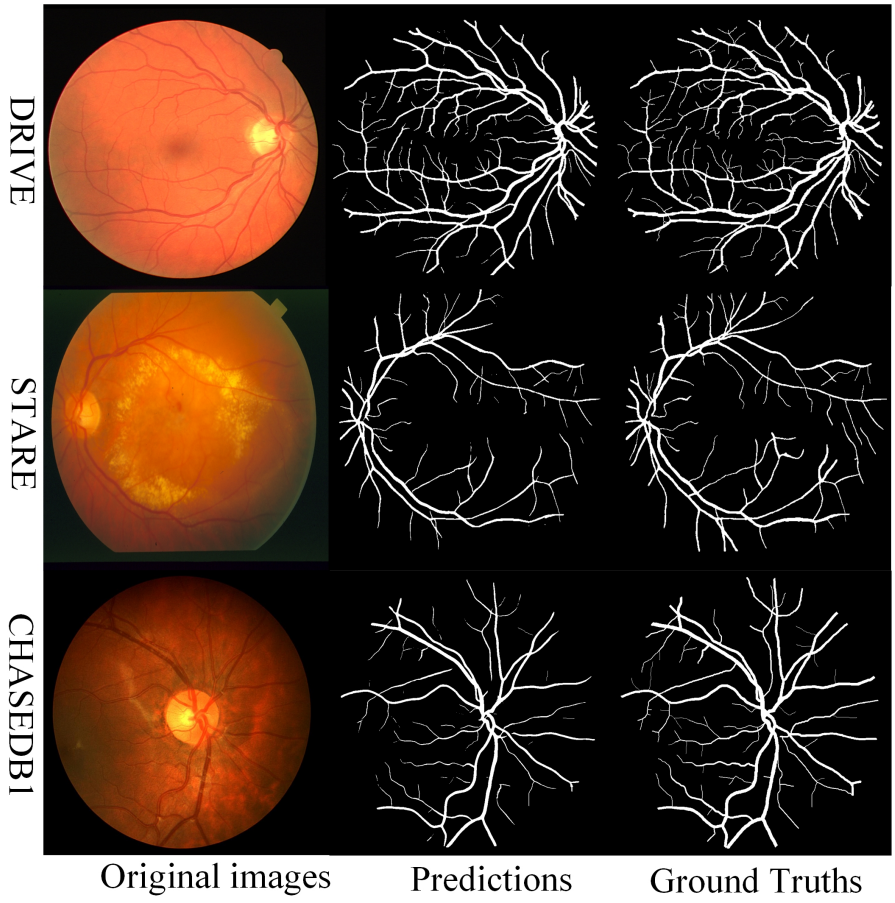

Figure 2. Example segmentation results on three databases.

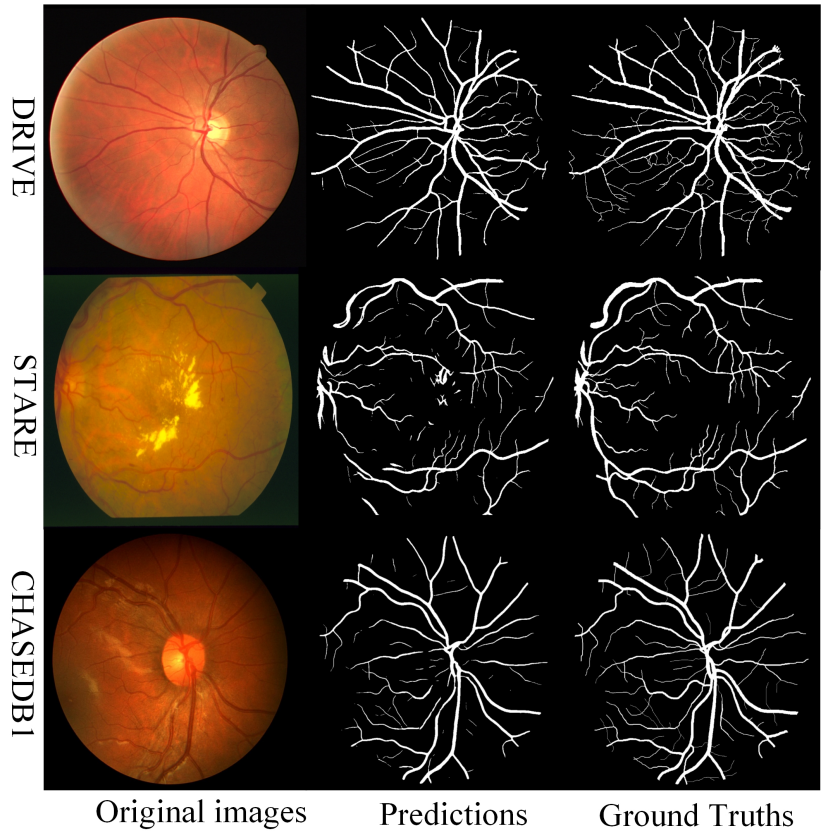

Figure 3. Bad segmentation examples on three databases.

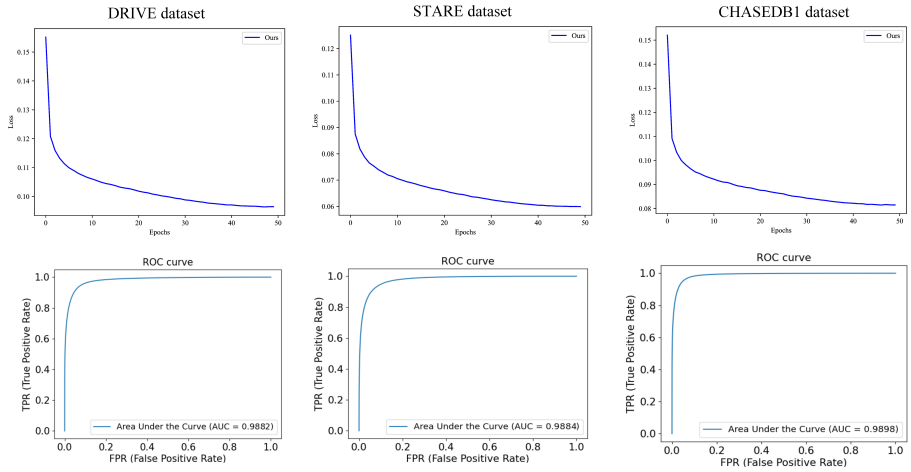

Figure 4. Loss curve and ROC curve on three datasets.

4.5. Ablation Study

In this section, the proposed VQ-VAE module and edge attention module are analyzed in detail. The VQ-VAE module regularizes the encoder parameters in the network while reconstructing the input images. It makes the encoder more robust and can generate features that are more representative. Furthermore, the high and low-level latent features are fused with the intermediate features of the backbone network to enhance the feature representation. Since the low-level features contain rich edge information, we utilize it to generate edge maps and feed it into the edge attention module to explicitly induce the model to learn features that focus on edges. The ablation study validates the proposed methods, and the results are presented in Table 3. The baseline is the backbone network, which removes the VQ-VAE module and the edge attention module. As shown in Table 3,

compared with 'Baseline', 'Baseline + VVM' improves the performance from 79.35%/96.16% to 82.31%/97.24% in terms of SE/ACC. Compared with the 'Baseline', the proposed EAM module (referred to as 'Baseline + EAM') increases the SE/ACC by 1.79%/0.71% (from 79.35%/96.16% to 81.14%/96.87%). Finally, we add the VQ-VAE module and edge attention module to the Baseline. The results show that the proposed methods obtain the best results on the three metrics. Many statistical comparisons and component analyses have visual representations in Figure 5. The effectiveness of the proposed modules in segmenting vessels of varying scales, including some miniature vessels that the baseline network cannot handle efficiently, is illustrated with an example from retinal vascular segmentation. In conclusion, the proposed method has the potential to significantly enhance the model's efficiency.

Table 3. Comparison of ablation studies on DRIVE dataset. VVM stands for VQ-VAE module, EAM denotes the eage attention module.

Methods	SE (%)	SP (%)	ACC (%)	AUC (%)
Baseline	79.35 ± 0.09	97.95 ± 0.11	96.16 ± 0.04	97.84 ± 0.03
Baseline + VVM	82.31 ± 0.18	**98.43 ± 0.25**	97.24 ± 0.09	98.55 ± 0.06
Baseline + EAM	81.14 ± 0.15	98.11 ± 0.15	96.87 ± 0.09	98.19 ± 0.04
Baseline + VVM + EAM	**83.86 ± 0.13**	98.37 ± 0.28	**97.35 ± 0.08**	**98.82 ± 0.05**

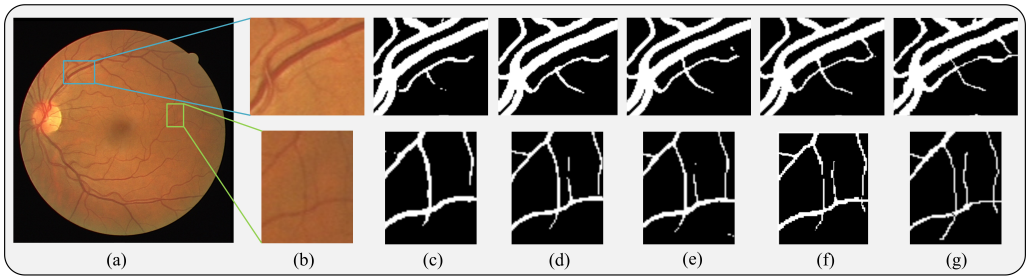

Figure 5. Some typical visual results for different methods in our ablation study on the DRIVE dataset. (**a**) Original image, (**b**) detailed view, (**c**) Baseline, (**d**) Baseline + VVM, (**e**) Baseline + EAM, (**f**) Baseline + VVM + EAM, (**g**) ground truth.

4.6. Effectiveness of the VQ-VAE Module

The VQ-VAE module reconstructs the input images while regularizing the shared parameters of the encoder and generating the high-level features and the low-level features. The reconstruction process induces the model to generate more representative features and enhances the robustness of the model. The two level features and the features generated by the encoder are fused to further improve the feature representation ability of the model. In this section, we mainly conduct experiments to verify that the VQ-VAE module can reduce the dependence of the model on annotations. First, we randomly select half of the training data in the DRIVE dataset as the training set, and the test set remains unchanged. Then, the 'Baseline' and 'Baseline + VVM' are trained on the modified training set and validated on the standard test set, respectively. The results are shown in Table 4. As can be observed, after reducing the training data, all metrics of the baseline network declined significantly. However, the baseline network with the VQ-VAE module still achieves competitive results without significant degradation. It demonstrates that the proposed method can reduce the dependence of the model on the training data and improve its robustness.

Table 4. Statistical comparison of the effectiveness of the VQ-VAE module.

Data amount = Full				
Methods	SE (%)	SP (%)	ACC (%)	AUC (%)
Baseline	79.35 ± 0.09	97.95 ± 0.11	96.16 ± 0.04	97.84 ± 0.03
Data amount = 1/2				
Methods	SE (%)	SP (%)	ACC (%)	AUC (%)
Baseline	74.42 ± 0.39	95.74 ± 0.31	93.78 ± 0.19	95.92 ± 0.08
Baseline + VVM	78.62 ± 0.18	96.75 ± 0.23	95.28 ± 0.12	97.11 ± 0.07

4.7. Limitations

There is a serious class imbalance problem in medical images, especially when the TP class (vessels) is significantly smaller than the TN class (the rest of the image) in the retinal images. The targets we focus on only account for a small part of the whole image, resulting in large regions dominating small regions. Thus, we should pay more attention to the small targets (thin vessel). In addition, the whole framework performs two sub-tasks; one is reconstruction, and the other is semantic segmentation. In the experiment, we found that when the number of epochs increases to a high level, the image segmentation performance will decline rapidly, and the model will pay more attention to the reconstruction sub-task. For this, we display some failure cases in Figure 3. In future work, we should solve the class imbalance problem and make the model training process more stable.

5. Conclusions

In this paper, we present a novel multi-task learning-based network to comprehensively address three challenges that remain in retinal vessel segmentation. The whole network consists of four parts: the image encoding network, the segmentation decoder, the VQ-VAE module, and the edge attention module. The edge attention module is capable of effectively inducing the encoder to capture the edge information of the target and explicitly concern the edge area of the target by deep supervision, which is important for retinal blood vessel segmentation. The VQ-VAE module conducts the input image reconstruction task to regularize the parameters of the encoding network, generating and fusing multi-level spatial and semantic features to incorporate local and global information. It not only improves the model's performance but also helps consistently achieve good training accuracy for any random initialization. Furthermore, the process of regularization can improve the generalization ability and reduce the dependence on sufficient accurate annotated training data. By using three publicly available retinal fundus datasets (DRIVE, CHASEDB1 and START) for in-depth comparative analysis, it is shown that the proposed method can be compared with the state-of-the-art method. We believe the proposed approaches are readily transferable to other medical image segmentation scenarios where few training data and complex anatomical semantics pose significant challenges.

Author Contributions: G.J.: conceptualization, methodology, software, investigation, writing—original draft preparation X.C.: formal analysis, writing—reviewing and editing, validation L.Y.: software, resources, writing—reviewing and editing, formal analysis, conceptualization, investigation, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61902193 and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Data Availability Statement: Not applicable.

Conflicts of Interest: We declare that we have no conflict of interest with each other and with other people or organizations that can inappropriately influence our work. The funders provides software, resources, writing—reviewing and editing, formal analysis, conceptualization, investigation, and the decision to publish the results in this work.

References

1. Zana, F.; Klein, J.C. A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform. *IEEE Trans. Med. Imaging* **1999**, *18*, 419–428. [CrossRef] [PubMed]
2. Sinthanayothin, C. Automated localization of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br. J. Ophthalmal* **1999**, *83*, 231–238.
3. Wu, H.; Wang, W.; Zhong, J.; Lei, B.; Wen, Z.; Qin, J. Scs-net: A scale and context sensitive network for retinal vessel segmentation. *Med. Image Anal.* **2021**, *70*, 102025. [CrossRef] [PubMed]
4. Li, D.; Rahardja, S. BSEResU-Net: An attention-based before-activation residual U-Net for retinal vessel segmentation. *Comput. Methods Programs Biomed.* **2021**, *205*, 106070. [CrossRef]
5. Mo, J.; Zhang, L. Multi-level deep supervised networks for retinal vessel segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 2181–2193. [CrossRef]
6. Nian, F.; Li, T.; Wu, X.; Gao, Q.; Li, F. Efficient near-duplicate image detection with a local-based binary representation. *Multimed. Tools Appl.* **2016**, *75*, 2435–2452. [CrossRef]
7. Li, T.; Yan, S.; Mei, T.; Hua, X.S.; Kweon, I.S. Image decomposition with multilabel context: Algorithms and applications. *IEEE Trans. Image Process.* **2010**, *20*, 2301–2314.
8. Li, T.; Mei, T.; Yan, S.; Kweon, I.S.; Lee, C. Contextual decomposition of multi-label images. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2270–2277.
9. Nian, F.; Bao, B.K.; Li, T.; Xu, C. Multi-modal knowledge representation learning via webly-supervised relationships mining. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 411–419.
10. Zhang, J.; Liu, S.; Yan, H.; Li, T.; Mao, R.; Liu, J. Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions. *Phys. Med. Biol.* **2020**, *65*, 205013. [CrossRef]
11. Jiang, D.; Yan, H.; Chang, N.; Li, T.; Mao, R.; Du, C.; Guo, B.; Liu, J. Convolutional neural network-based dosimetry evaluation of esophageal radiation treatment planning. *Med. Phys.* **2020**, *47*, 4735–4742. [CrossRef]
12. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
13. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2016; pp. 424–432.
14. Khened, M.; Kollerathu, V.A.; Krishnamurthi, G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med. Image Anal.* **2019**, *51*, 21–45. [CrossRef]
15. Fu, H.; Xu, Y.; Lin, S.; Kee Wong, D.W.; Liu, J. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2016; pp. 132–139.
16. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv* **2018**, arXiv:1802.06955.
17. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
18. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [CrossRef] [PubMed]
19. Zhang, J.; Zhang, Y.; Xu, X. Pyramid u-net for retinal vessel segmentation. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1125–1129.
20. Staal, J.; Abràmoff, M.D.; Niemeijer, M.; Viergever, M.A.; Van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [CrossRef] [PubMed]
21. Owen, C.G.; Rudnicka, A.R.; Mullen, R.; Barman, S.A.; Monekosso, D.; Whincup, P.H.; Ng, J.; Paterson, C. Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program. *Investig. Ophthalmol. Vis. Sci.* **2009**, *50*, 2004–2010. [CrossRef]
22. 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* **2000**, *19*, 203–210. [CrossRef]
23. Liskowski, P.; Krawiec, K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 2369–2380. [CrossRef]
24. Samuel, P.M.; Veeramalai, T. VSSC Net: Vessel specific skip chain convolutional network for blood vessel segmentation. *Comput. Methods Programs Biomed.* **2021**, *198*, 105769. [CrossRef]

25. Soomro, T.A.; Afifi, A.J.; Gao, J.; Hellwich, O.; Khan, M.A.; Paul, M.; Zheng, L. Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network. In Proceedings of the 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, Australia, 29 November–1 December 2017; pp. 1–8.
26. Wu, A.; Xu, Z.; Gao, M.; Buty, M.; Mollura, D.J. Deep vessel tracking: A generalized probabilistic approach via deep learning. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1363–1367.
27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
28. Atli, I.; Gedik, O.S. Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 271–283. [CrossRef]
29. Li, W.; Zhang, M.; Chen, D. Fundus retinal blood vessel segmentation based on active learning. In Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 17–19 April 2020; pp. 264–268.
30. Luo, Y.; Cheng, H.; Yang, L. Size-invariant fully convolutional neural network for vessel segmentation of digital retinal images. In Proceedings of the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Republic of Korea, 13–15 December 2016; pp. 1–7.
31. Sathananthavathi, V.; Indumathi, G. Encoder enhanced atrous (EEA) unet architecture for retinal blood vessel segmentation. *Cogn. Syst. Res.* **2021**, *67*, 84–95.
32. Li, D.; Dharmawan, D.A.; Ng, B.P.; Rahardja, S. Residual u-net for retinal vessel segmentation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 1425–1429.
33. Lian, S.; Li, L.; Lian, G.; Xiao, X.; Luo, Z.; Li, S. A global and local enhanced residual u-net for accurate retinal vessel segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *18*, 852–862. [CrossRef] [PubMed]
34. Mishra, S.; Chen, D.Z.; Hu, X.S. A data-aware deep supervised method for retinal vessel segmentation. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1254–1257.
35. Laibacher, T.; Weyde, T.; Jalali, S. M2u-net: Effective and efficient retinal vessel segmentation for real-world applications. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
36. Jin, Q.; Meng, Z.; Pham, T.D.; Chen, Q.; Wei, L.; Su, R. DUNet: A deformable network for retinal vessel segmentation. *Knowl.-Based Syst.* **2019**, *178*, 149–162. [CrossRef]
37. Li, L.; Verma, M.; Nakashima, Y.; Nagahara, H.; Kawasaki, R. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 3656–3665.
38. Liu, W.; Yang, H.; Tian, T.; Cao, Z.; Pan, X.; Xu, W.; Jin, Y.; Gao, F. Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiograph Segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4623–4634. [CrossRef] [PubMed]
39. Zhou, Y.; Yu, H.; Shi, H. Study group learning: Improving retinal vessel segmentation trained with noisy labels. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2021; pp. 57–67.
40. Kamran, S.A.; Hossain, K.F.; Tavakkoli, A.; Zuckerbrod, S.L.; Sanders, K.M.; Baker, S.A. RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2021; pp. 34–44.
41. Zhou, T.; Li, L.; Bredell, G.; Li, J.; Unkelbach, J.; Konukoglu, E. Volumetric memory network for interactive medical image segmentation. *Med. Image Anal.* **2023**, *83*, 102599. [CrossRef] [PubMed]
42. Zhou, T.; Wang, W.; Konukoglu, E.; Van Gool, L. Rethinking Semantic Segmentation: A Prototype View. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2582–2593.
43. Yang, B.; Zhang, X.; Chen, L.; Yang, H.; Gao, Z. Edge guided salient object detection. *Neurocomputing* **2017**, *221*, 60–71. [CrossRef]
44. Wu, Z.; Su, L.; Huang, Q. Stacked cross refinement network for edge-aware salient object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7264–7273.
45. Zhou, T.; Li, J.; Wang, S.; Tao, R.; Shen, J. Matnet: Motion-attentive transition network for zero-shot video object segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 8326–8338. [CrossRef]
46. Myronenko, A. 3D MRI brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2018; pp. 311–320.
47. Razavi, A.; Van den Oord, A.; Vinyals, O. Generating diverse high-fidelity images with vq-vae-2. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 14866–14876.
48. Wu, Y.; Xia, Y.; Song, Y.; Zhang, Y.; Cai, W. NFN+: A novel network followed network for retinal vessel segmentation. *Neural Netw.* **2020**, *126*, 153–162. [CrossRef]
49. Li, Q.; Feng, B.; Xie, L.; Liang, P.; Zhang, H.; Wang, T. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans. Med. Imaging* **2015**, *35*, 109–118. [CrossRef]
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.