

Article

Retinal OCTA Image Segmentation Based on Global Contrastive Learning

Ziping Ma ¹, Dongxiu Feng ^{2,*}, Jingyu Wang ¹ and Hu Ma ¹¹ College of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China² College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China

* Correspondence: 20207141@stu.nun.edu.cn

Abstract: The automatic segmentation of retinal vessels is of great significance for the analysis and diagnosis of retinal related diseases. However, the imbalanced data in retinal vascular images remain a great challenge. Current image segmentation methods based on deep learning almost always focus on local information in a single image while ignoring the global information of the entire dataset. To solve the problem of data imbalance in optical coherence tomography angiography (OCTA) datasets, this paper proposes a medical image segmentation method (contrastive OCTA segmentation net, COSNet) based on global contrastive learning. First, the feature extraction module extracts the features of OCTA image input and maps them to the segment head and the multilayer perceptron (MLP) head, respectively. Second, a contrastive learning module saves the pixel queue and pixel embedding of each category in the feature map into the memory bank, generates sample pairs through a mixed sampling strategy to construct a new contrastive loss function, and forces the network to learn local information and global information simultaneously. Finally, the segmented image is fine tuned to restore positional information of deep vessels. The experimental results show the proposed method can improve the accuracy (ACC), the area under the curve (AUC), and other evaluation indexes of image segmentation compared with the existing methods. This method could accomplish segmentation tasks in imbalanced data and extend to other segmentation tasks.

Keywords: medical image processing; image segmentation; optical coherence tomography angiography; retinal vascular plexus; imbalanced data; contrastive learning; convolutional neural network



Citation: Ma, Z.; Feng, D.; Wang, J.; Ma, H. Retinal OCTA Image Segmentation Based on Global Contrastive Learning. *Sensors* **2022**, *22*, 9847. <https://doi.org/10.3390/s22249847>

Academic Editors: Sidong Liu, Roneel V. Sharan and Hao Xiong

Received: 18 November 2022

Accepted: 12 December 2022

Published: 14 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Optical coherence tomography angiography (OCTA), a non-invasive imaging technique, has been increasingly utilized for imaging the retinal and choroidal vascular system at capillary-level resolution. Compared with previous angiographic methods, OCTA has the merits of easier handling, faster imaging, and greater imaging depth, thus facilitating quantitative assessment of the morphological structure of the retinal and choroidal vascular system. Abnormalities on OCTA images usually indicate the presence of a number of diseases, such as early glaucomatous optic neuropathy [1], diabetic retinopathy [2], and age-related macular degeneration [3]. Some other recent studies have shown that deep microvascular morphological changes revealed by OCTA images are associated with Alzheimer's disease [4] and mild cognitive impairment [5]. This opens up new ideas to investigate the relationship between abnormalities in retinal vessels and various neurodegenerative diseases. This shows that the precise segmentation of retinal vessels, especially deep capillaries, is of great importance to clinicians.

Before the popularity of deep learning, most of the automatic vessel segmentation methods were based on filters, and such methods are generally referred to as traditional methods. Azzopardi et al. [6] designed a rod structure selection filter based on the COSFIRE filter, which can effectively extract information about branching structure compared to other filters. A detailed review of the methods for conventional vessel detection and

segmentation can be found in the literature [7]. A common limitation of all these traditional vessel segmentation methods is that they usually require manual tuning of parameters for a specific dataset, which makes the algorithms less scalable.

Compared to traditional methods, deep learning performs better generalization. However, deep learning methods applied to medical images still face challenges, which Zhou et al. [8] attribute to the characteristics of medical images itself: the lack of standardized big data; the multimodality of medical data; and the long-tailed distribution of morbidity. Such problems also exist in the field of ophthalmic images, and more detail can be found in earlier work [9–13]. In previous retinal vascular datasets, the number of samples is generally low, such as in DRIVE [14], CHASE-DB1 [15], etc. To alleviate the shortage of such datasets, Li et al. [16] released the largest multimodal OCTA image dataset, OCTA-500, and also proposed the IPN image projection network. Taking advantage of the vertical consistency of 3D OCTA images, the IPN method compresses information in a 3D image to a 2D image and finally outputs segmentation maps. Ma et al. [17] constructed a dataset called ROSE with retinal superficial and deep vascular annotations, respectively, to make up for the lack of datasets in this domain. The segmentation network OCTA-Net was also proposed to first generate an initial confidence map of blood vessels and then optimize the vessel details using a finetune module.

Among many deep learning-based medical image segmentation networks, the core ideology of U-Net [18] is the most prevalent, and our COSNet also takes advantage of it. The U-shaped structure facilitates the extraction of semantic information, while the skip connection contributes to restore positional information. In addition, the U-Net structure can be applied to all types of medical images virtually. In recent years, many variants based on the U-Net architecture have also emerged. Huang et al. [19] proposed a UNet3+ segmentation network based on full-scale jump connections and deep supervision. Full-scale skip connections combine low-level details with high-level semantics of feature maps at different scales, whereas deep supervision learns hierarchical representations from full-size aggregated feature maps. Isensee et al. [20] addressed the selection of pre-processing, architecture, training methods, and post-processing methods to propose a nn-UNet segmentation network that can automatically configure parameters, outperforming most methods on 23 international biomedical competition datasets. Tomar et al. [21] proposed the feedback attention network FANet for the network's inability to efficiently learn information from different iteration batches, which combines feature maps from the current and previous batches together so that the masks of the previous batches constitute hard attention to learn the feature maps of different convolutional layers. Although the above methods achieve good results in image segmentation tasks, they focus only on intra- and inter-pixel semantic information of a single image, which prevents them from forming a structured feature space (large inter-class distance and small intra-class distance), thus limiting the segmentation performance.

Most current medical image segmentation methods rely entirely on accurate annotation, and medical image data are not as easily accessible as natural images. To address the data integration cost problem, researchers have introduced non-fully supervised methods to image segmentation tasks to reduce the reliance on annotation. In this paper, we focus on contrast learning, and the application of contrast learning to image segmentation is described below. Commonly used semantic segmentation methods for contrast learning are usually pre-trained on large unlabeled datasets first and then used for segmentation tasks, such as in the literature [22]. To solve the problem of training overfitting on a small amount of labeled data, Zhao et al. [23] proposed a new contrastive learning pre-training strategy, which is first pre-trained with label-based contrast loss and then fine-tuned with cross-entropy loss to improve intra-class compactness and inter-class separation. Zhong et al. [24] proposed a semi-supervised semantic segmentation method that both maintains the continuity of the prediction space for the input transform and also ensures the contrastive nature of the feature space. In contrast to the previous stepwise contrast learning segmentation, the contrast loss and semi-supervised segmentation training in

this work are performed jointly. Alonso et al. [25] proposed a contrast learning module to perform pixel-level feature representation of similar samples from labeled and unlabeled data, and this work obtained good results in scenarios where the proportion of labeled data is small. Although all these methods noted above have achieved good results in the corresponding fields, they have not completely solved the problems of category imbalance and small target segmentation in medical image segmentation. COSNet in this paper also refers to similar loss function forms and utilizes the memory bank approach.

To address the problems of poor detail and class imbalance in small target segmentation tasks, we propose a global contrastive learning method for retinal OCTA image segmentation. Compared with previous methods, the method in this paper focuses on both single image information and global information (the whole dataset), which can achieve better results in a small sample medical dataset containing two kinds of annotations. The main contributions of this paper are as follows:

1. A two-branch contrastive learning network for retinal OCTA image segmentation is proposed. The model is able to effectively extract features of vascular images by learning superficial and deep vascular annotations while avoiding the feature vanishing of deep vessels. A segmentation head and a projection head are added at the end of the decoder to obtain both segmentation mapping and pixel embedding;
2. In this paper, a new pixel contrastive loss function is proposed. By saving same-class pixel queues and pixel embeddings in memory bank, features within a single image can be learned as well as same-class features in the whole dataset. The network model is guaranteed to learn more hard-to-segment samples, thus alleviating the class imbalance problem and improving the segmentation performance;
3. A contrast-limited adaptive histogram equalization (CLANE) method with fixed area is used for retinal OCTA images to mitigate noise caused by imaging artifacts.

The subsequent sections of the text contain Methods and Theory, Experiments and Results, Discussion, and Conclusions.

2. Methods and Theory

We propose a fully-supervised contrastive learning model for retinal OCTA image segmentation called contrastive OCTA segmentation net (COSNet). This method is mainly divided into three parts: **feature extraction module**, **contrastive learning module**, and **fine-tune module**, as shown in Figure 1.

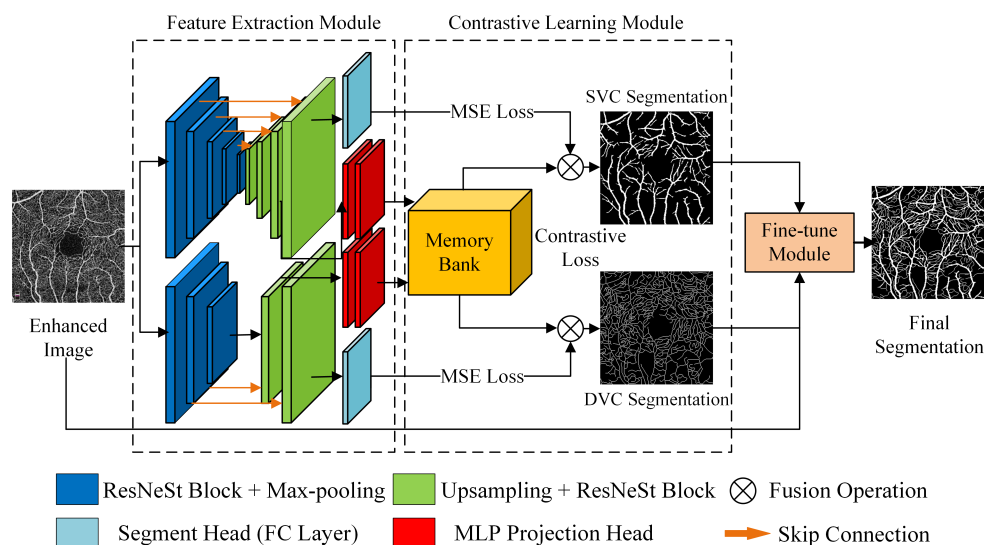


Figure 1. COSNet architecture.

2.1. Feature Extraction Module

In the feature extraction module, a two-branch U-shaped network is adopted to handle data with two levels of annotation. This architecture has a partially shared “encoder + decoder” structure. The upper branch deals with superficial vessels (pixel-level label), and the lower branch deals with deep vessels (centerline-level), which can balance the importance of the two levels of information. Unlike previous segmentation networks, our method adds both a segmentation head and an MLP mapping head after the “encoder + decoder” structure. As shown in Figure 1, the ResNeSt module is used as the basic module for feature extraction, and the shallow vessel branch has five encoder layers and corresponding decoder layers, whereas the deep vessel branch has three encoder layers and corresponding decoder layers. Eventually, the segmentation mapping is obtained via a 1×1 convolutional layer, and the pixel embedding of the feature map is obtained by the MLP projection head behind the penultimate layer of the decoder.

As shown in Figure 2, we employ the ResNeSt [26] module as the basic unit, which serves the purpose of treating a series of representations as a combination of several feature groups and then using channel attention on these feature groups.

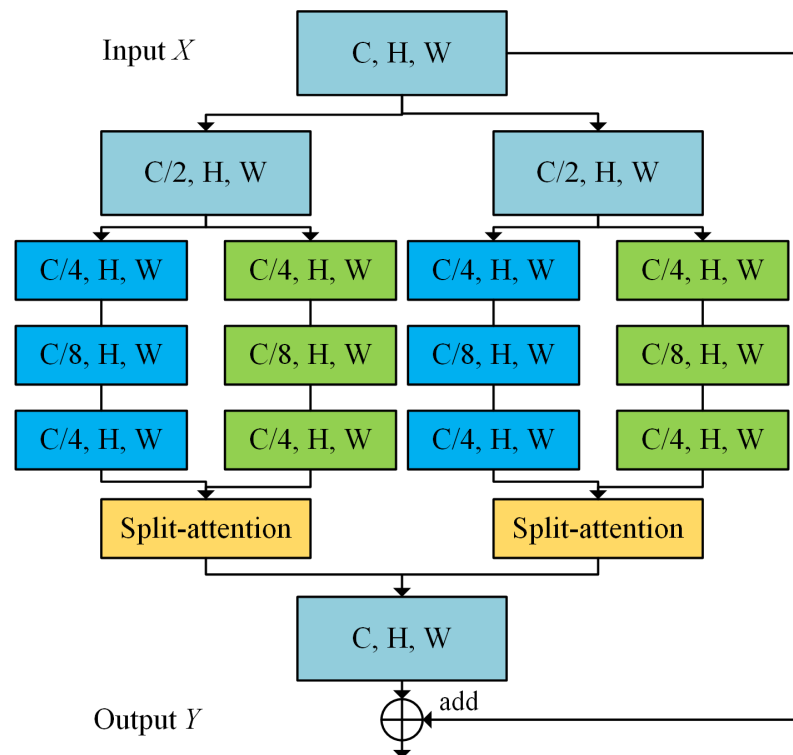


Figure 2. ResNeSt block (only two branches are shown for brevity).

The variables C , H , and W represent channels, height, and width, respectively, of the feature map. The feature input is divided equally into K cardinal groups, which are then fed into R identical cardinal blocks. Thus, the initial input feature map is divided into G ($G = K \times R$) groups along the channel dimension. Each branch contains the structure of “ 1×1 convolutional layer + BN + Relu” and “ 3×3 convolutional layer + BN + Relu”, resulting in a feature map with one-fourth of the original channels. In addition, each cardinal block is followed by a split attention to integrate feature map.

As shown in Figure 3, the feature map from two branches (U_1 and U_2) are fused, and then the channel-wise statistics S are generated via global pooling. It can be expressed as:

$$S_c^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c^k(i, j). \quad (1)$$

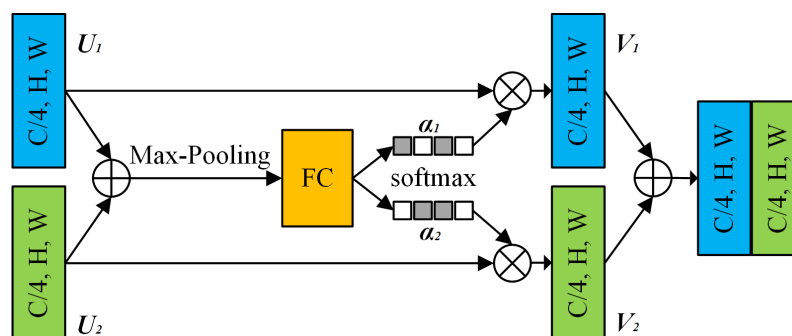


Figure 3. Split-attention block (only two branches are shown for brevity).

The weighted fusion of the cardinal group representations is implemented using the channel dimension soft attention, where the feature maps of each channel are generated using a weighted combination. This process can be represented as:

$$V_c^k = \sum_{i=1}^R \alpha_i^k(c) U_{R(k-1)}. \quad (2)$$

where

$$\alpha_i^k(c) = \begin{cases} \frac{\exp(G_i^c(s^k))}{\sum_{j=0}^R \exp(G_j^c(s^k))} & R > 1, \\ \frac{1}{1 + \exp(-G_i^c(s^k))} & R = 1. \end{cases} \quad (3)$$

Next, all the base arrays are concatenated to get the output, and by adding them up with shortcut $\tau(X)$, the total module output is $Y = V + \tau(X)$.

2.2. Contrastive Learning Module

Currently, the loss functions used in most semantic segmentation methods focus only on local information in a single image and ignore the global context, i.e., same-class samples in the entire dataset. Examples include the cross-entropy loss function and the Dice loss function. In medical image segmentation networks, cross-entropy and Dice loss are usually used in a weighted manner to solve the problem of excessive oscillation of the loss function during training. However, the performance of such loss functions in segmentation becomes poor when the segmented target occupies a very small proportion of the full image. Considering the requirement of the loss function for detail difference sensitivity in the vessel segmentation task, we select the mean-square error loss (MSE) as part of the loss function. The mean-square loss function is defined as follows:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_{pred} - y_{gt})^2. \quad (4)$$

To address the category imbalance problem, our approach borrows the idea of MoCo [27] to construct sample pairs, i.e., a memory bank. Unlike the MoCo method, the memory bank proposed in this paper stores a queue for each class. In each class, only a small number of pixels are randomly selected from the latest batch of images and are queued. At the same time, the same-class pixel embedding in each image is stored by mean-pooling. This approach achieves a better balance between learning speed and sample diversity than storing all pixel samples directly or updating only the last few batches. Specifically, suppose there are a total of N training images, C semantic classes, D is the dimension of the pixel embedding, and T is the length of the pixel queue, then the image pixel embedding size is $C \times N \times D$; the pixel queue size is $C \times T \times D$. The final size of the memory bank is $M = (N + T) \times C \times D$. The advantage of this approach is that more representative hard-to-segment samples could be stored, and memory consumption is lower.

The most vital step of contrast learning is how to reasonably select difficult samples. In OCTA images, difficult samples refer to those detailed parts of the deep vessels, where the width of the tiniest part is only one pixel. It can be further shown from Kalantidis et al. [28] that an increasing number of negative samples (background) over-simplifies the training, which leads to an ineffective penalty for contrastive loss. We use a hybrid strategy that includes hard sample mining and random sampling. The first few most difficult samples are sampled from the top of the memory bank using difficult sample mining, whereas the other half is sampled randomly. This hybrid sampling method allows pixel contrastive loss to focus more on the difficult pixels for segmentation.

The global pixel contrastive loss proposed in this paper is defined as follows:

$$L_i^{NCE} = \frac{1}{|P_i|} \sum_{i^+ \in P_i} -\log \frac{\exp(i \cdot i^+ / \tau)}{\exp(i \cdot i^+ / \tau) + \sum_{i^- \in N_i} \exp(i \cdot i^- / \tau)}. \quad (5)$$

where i^+ and i^- represent positive and negative samples, respectively; P_i and N_i represent the set of positive and negative pixel embeddings, respectively (positive and negative samples are from the entire dataset); “ \cdot ” represents the inner product; τ represents a temperature hyperparameter. When pixel i belonging to class \bar{C} is calculated, the pixel embedding of the same class \bar{C} is considered as a positive sample and the other classes are considered as negative samples.

Therefore, combining the respective advantages of mean-square loss and global pixel contrastive loss, the objective function proposed is as follows:

$$L_{obj} = \sum_i (L_{MSE} + \lambda_{NCE} L_i^{NCE}). \quad (6)$$

Ultimately, both the superficial vascular complexes (SVC) branch and the deep vascular complexes (DVC) branch of COSNet utilize Equation (6) as the objective function, so the segmentation loss function equation of this module is defined as follows:

$$L_{total} = \lambda_{SVC} L_{SVC} + \lambda_{DVC} L_{DVC}. \quad (7)$$

2.3. Fine-Tune Module

To address the problem of detailed breakpoints in vessel segmentation, we directly follow the fine-tune module in OCTA-Net [17], as shown in Figure 4.

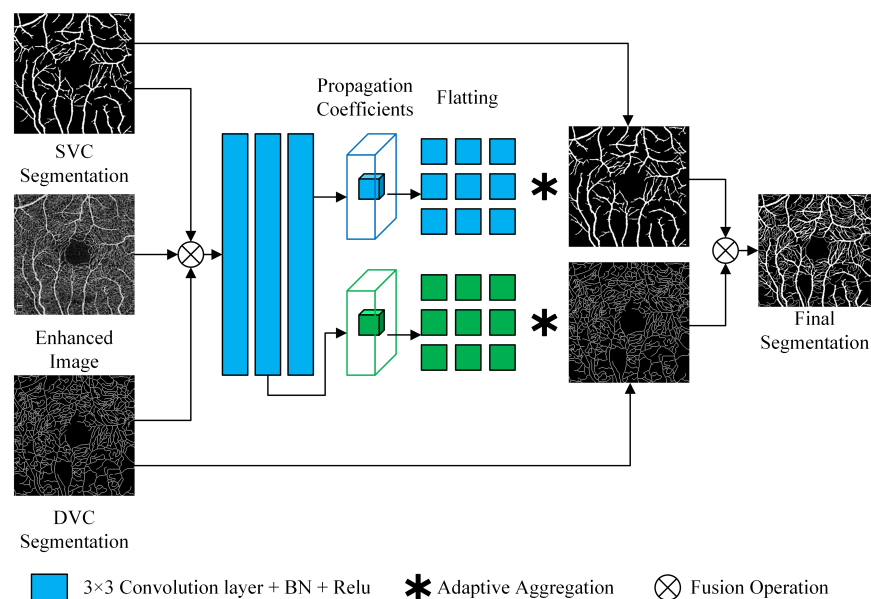


Figure 4. Fine-tune block.

First, the SVC prediction map, DVC prediction map, and original image (all single channel) obtained from the previous module are channel-fused into a 3-channel input. The input passes through two and three convolutional layers (DVC and SVC), respectively, and then this module generates a normalized local propagation coefficient mapping of size $m \times m$, for all positions of the output of the convolutional layers, denoted as Equation (8), where h_i^P denotes the confidence value of neighbor P at position i , and $m \times m$ is the size of the propagating neighbor.

$$\omega_i^P = \frac{\exp(h_i^P)}{\sum_{t=1}^{m \times m} \exp(h_i^t)}, P \in [1, m \times m]. \quad (8)$$

Local propagation coefficient ω_i^P of position i is multiplied by SVC and DVC confidence maps from the previous module, then aggregated to the centroid to yield fine-tuned result, denoted as Equation (9):

$$Pred_i = \sum_{P=1}^{m \times m} \omega_i^P \cdot f_i^P \quad (9)$$

where $Pred_i$ is the final prediction vector for position i , and f_i^P is the confidence vector for neighbor P at position i . Finally, the fine-tuned SVC segmentation image and the DVC segmentation image are combined into a binary image.

In the fine-tuning module, we set the size of the aggregated neighbor coefficient m to 3 while using Dice loss as the fine-tune loss function. Dice loss function is denoted as Equation (10), where P_i , G_i represent the prediction and ground truth, respectively. In addition, a small positive constant σ is used to avoid the numerical problem and accelerate convergence of training.

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N P_i G_i + \sigma}{\sum_{i=1}^N P_i^2 + \sum_{i=1}^N G_i^2 + \sigma}. \quad (10)$$

3. Experiments and Results

In this section, we perform a comparison test and an ablation study in order to verify the effectiveness of COSNet on retinal vessel segmentation tasks (class-imbalanced and complex-structured vessel).

3.1. Experimental Configuration

3.1.1. Dataset and Augmentation

The ROSE dataset is the first open-source OCTA dataset for segmentation, constructed by the Ningbo Institute of Materials affiliated with the Chinese Academy of Sciences, Southern University of Science and Technology, University of Liverpool, UK, IIAI, University of Southern California, and several research and clinical institutions in China and abroad. The ROSE dataset consists of two subsets named ROSE-1 and ROSE-2, which, in total, contain 229 OCTA images. The OCTA-500 dataset [29], built by Qiang Chen's group from Nanjing University of Science and Technology, contains multimodal image data and multiple labels and different depth OCTA projection. All datasets are divided into a training set and a validation set in the ratio of 7:3. The details of the datasets are shown in Tables 1 and 2.

In addition to using common augmentation methods (e.g., random cropping, rotation, etc.), we employ a contrast-constrained adaptive histogram equalization (CLANE) method for preprocessing images in a fixed region. This method could enhance the detail features in DVC while reducing the noise generated by artifact problem. As can be seen in Figure 5, the DVC details in the green box are enhanced.

Table 1. Details of the ROSE dataset.

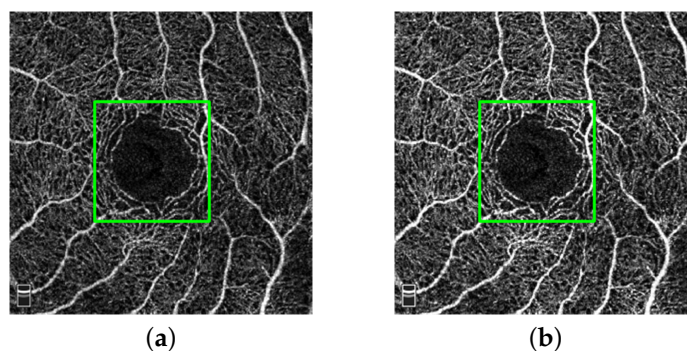
	ROSE-1	ROSE-2
Acquisition device	Optovue, USA	Heidelberg OCT2, Germany
Number	117	112
Resolution	304 × 304	512 × 512
Image type	SVC, DVC, SVC + DVC ¹	SVC
Annotation type	pixel and centerline level	centerline level
Disease type	Alzheimer's disease, macular degeneration, glaucoma, etc.	macular degeneration

¹ SVC stands for superficial vascular complex angiography; DVC stands for deep vascular complex angiography.

Table 2. Details of the OCTA-500 * dataset.

	OCTA-3M	OCTA-6M
Number	200	300
Resolution	304 × 304	400 × 400
Image type	SVC	SVC
Annotation type	pixel level	pixel level

* OCTA-500 contains images of different depths (FULL, ILM-OPL, and OPL-BM); we use only the ILM-OPL data.

**Figure 5.** Original image (a) and enhanced image (b).

3.1.2. Development Environment, Parameter Configuration, and Evaluation Metrics

The experimental hardware environment is an NVIDIA A40 graphics card, 80 GB running memory, and an AMD EPYC 7543 processor on AutoDI-GPU server; the software environment is the Ubuntu18 system, Python 3.6.5, Pytorch 1.8.2 framework. Our COSNet method trains 200 epochs in both the segmentation and fine-tune stages, with the following settings: batch size of 16, temperature τ of 0.07, memory bank size of 400, Adam optimization with the initial learning rate of 0.0005, and weight decay of 0.0001. To speed up training, COSNet makes use of the ResNeSt50 pre-trained weight.

The following six metrics were tested: area under the curve (AUC), accuracy (ACC), G-mean, Kappa, Dice, and false discovery rate (FDR). They are defined in Equations (11)–(16), where P stands for positive and N for negative.

$$\text{AUC} = \frac{\sum (p_i, n_j)_{p_i > n_j}}{P \times N}. \quad (11)$$

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (12)$$

$$\text{G-mean} = \sqrt{\frac{TP \times TN}{(TP + FN)(TN + FP)}}. \quad (13)$$

$$\text{Kappa} = \frac{\text{ACC} - p_e}{1 - p_e}, p_e = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}. \quad (14)$$

$$\text{Dice} = \frac{2 \times TP}{FP + FN + 2 \times TP}. \quad (15)$$

$$\text{FDR} = \frac{FP}{TP + FP}. \quad (16)$$

3.2. Comparison Test

To validate the superiority of COSNet on OCTA image segmentation tasks, COSNet is compared with existing methods on ROSE and OCTA-500. To ensure fairness, batch size is set to 16 and epoch of 400 (except COSFIRE), as listed in Tables 3–6.

Table 3. Performance comparison on ROSE-1.

Method	AUC	ACC	G-Mean	Kappa	Dice	FDR
COSFIRE [6]	0.8764	0.8978	0.7253	0.6125	0.6673	0.0985
U-Net [18]	0.9028	0.8859	0.8038	0.6310	0.7015	0.2889
nnU-Net [20]	0.9109	0.8996	0.8185	0.6687	0.7311	0.2253
FANet [21]	0.9285	0.9057	0.8223	0.6815	0.7406	0.2230
OCTA-Net [17]	0.9371	0.9098	0.8335	0.7022	0.7570	0.2045
COSNet (Ours)	0.9452	0.9133	0.8402	0.7097	0.7645	0.2013

Table 4. Performance comparison on ROSE-2*.

Method	AUC	ACC	G-Mean	Kappa	Dice	FDR
COSFIRE [6]	0.7783	0.9210	0.7745	0.5698	0.6143	0.3890
U-Net [18]	0.8365	0.9317	0.8000	0.6174	0.6559	0.3546
nnU-Net [20]	0.8459	0.9342	0.8077	0.6346	0.6696	0.3328
FANet [21]	0.8536	0.9373	0.8214	0.6578	0.6935	0.3236
OCTA-Net [17]	0.8603	0.9386	0.8313	0.6721	0.7078	0.3018
COSNet (Ours)	0.8674	0.9398	0.8386	0.6738	0.7104	0.3002

* Because ROSE-2 contains only centerline-level annotations, COSNet uses only the DVC branch.

Table 5. Performance comparison on OCTA-3M (ILM-OPL).

Method	AUC	ACC	G-Mean	Kappa	Dice	FDR
COSFIRE [6]	0.8542	0.8645	0.7544	0.6583	0.7402	0.2213
U-Net [18]	0.9156	0.9001	0.8021	0.6847	0.8054	0.2394
nnU-Net [20]	0.9358	0.9089	0.8087	0.7065	0.8365	0.2175
FANet [21]	0.9520	0.9275	0.8322	0.7344	0.8556	0.2160
OCTA-Net [17]	0.9524	0.9246	0.8457	0.7857	0.9085	0.2037
COSNet (Ours)	0.9676	0.9345	0.8628	0.7992	0.9168	0.1857

Table 6. Performance comparison on OCTA-6M (ILM-OPL).

Method	AUC	ACC	G-Mean	Kappa	Dice	FDR
COSFIRE [6]	0.8248	0.8392	0.7406	0.6084	0.7078	0.2859
U-Net [18]	0.8876	0.8802	0.8085	0.7158	0.8045	0.3189
nnU-Net [20]	0.8965	0.8854	0.8196	0.7295	0.8158	0.3064
FANet [21]	0.9057	0.8935	0.8172	0.7354	0.8365	0.2930
OCTA-Net [17]	0.9196	0.9107	0.8349	0.7536	0.8754	0.2778
COSNet (Ours)	0.9388	0.9253	0.8457	0.7768	0.8869	0.2874

Except for FDR in Table 3, COSNet obtained the best performance. Although COSFIRE achieved the best score on the FDR metric, it was significantly lower than the deep learning method on other metrics. The filter-based method worked better on the simple structure of vessels. However, in the ROSE dataset, the superficial and deep microvascular structures are highly dense and intricate, which makes the segmentation effect of the traditional method poorer. In addition, the artifacts, low resolution, and low contrast of OCTA imaging also increase segmentation difficulty. In contrast, deep learning methods can extract higher-level distinguishing representations from local and global features, resulting in better performance on segmentation tasks. To visualize segmentation performance of different methods, we show the ROC curves for ROSE-1 and ROSE-2 in Figure 6. The comparison of segmentation results is also shown in Figures 7 and 8. (We mainly focus on tasks containing two kinds of annotations, so only the results of ROSE-1 are shown).

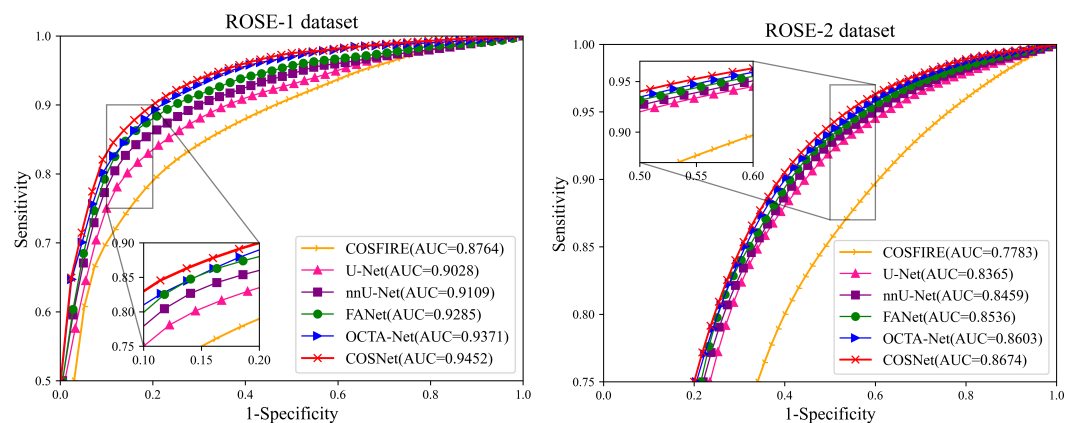


Figure 6. ROC curves for ROSE-1 (left) and ROSE-2 (right).

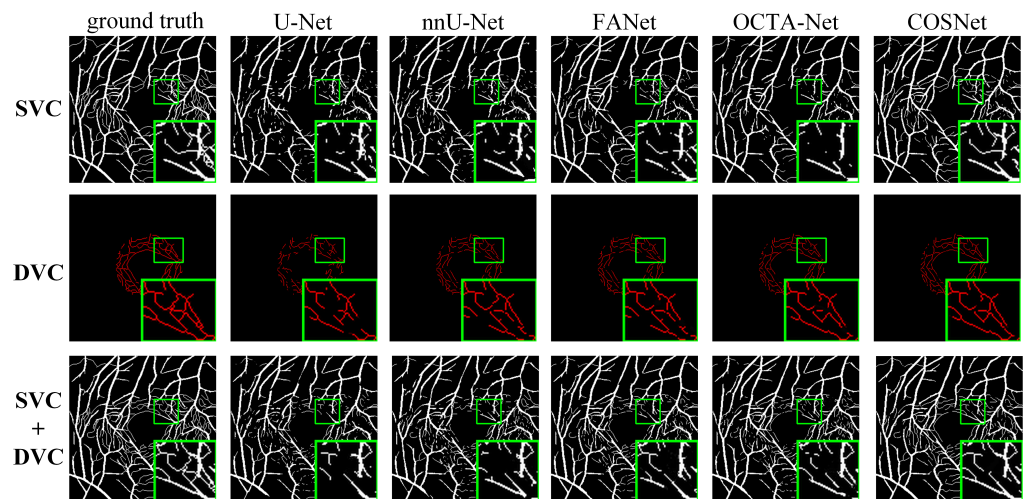


Figure 7. Segmentation results for ROSE-1.

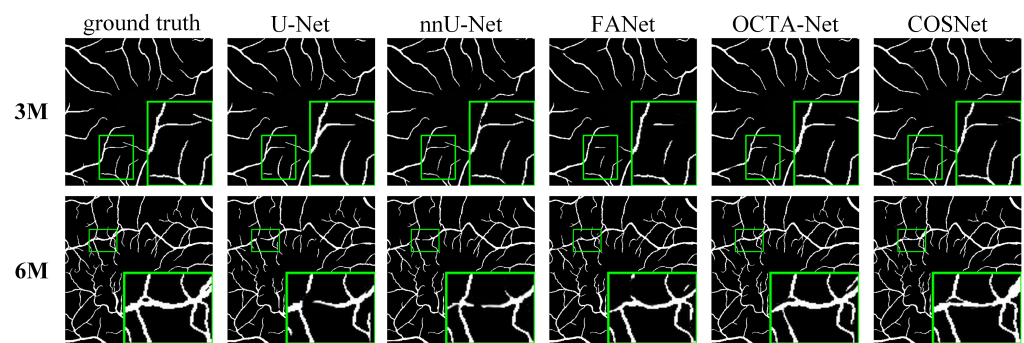


Figure 8. Segmentation results for OCTA-500 (ILM-OPL).

Deep learning-based segmentation methods are superior to traditional methods such as COSFIRE for the OCTA image segmentation task. The basic structure of these segmentation networks is “encoder + decoder”, which is able to extract the deep semantic information. In addition, the use of various attention blocks can also help to improve the AUC score, such as the feedback attention mechanism of FANet. The COSNet proposed in this paper outperforms other approaches not only by using the ResNeSt block as the backbone of the “encoder + decoder” structure, but also by adopting a new pixel contrast loss function, which improves the information extraction ability of the network under the scenario of class imbalance. In addition, the fine-tune module can refine position information of DVC. From the green box in Figures 7 and 8, it is easy to see that the previous methods have more breakpoints at deep vascular locations. In contrast, COSNet outperforms the other methods in the scenario of complex structure and class imbalance of microvessels.

3.3. Ablation Study

The ablation experiments in this subsection mainly explore the performance of different combinations of methods on ROSE-1. The role of skip connection has been verified many times in previous studies, so no separate experiments are done. Each set of experiments uses pre-training weight of ResNest50 and trains 400 epochs separately to record the highest score.

As shown in Table 7, we conducted ablation experiments on the COSNet architecture. Due to the form of the mean square loss, the difference between the predicted and true values are amplified, forming a strong supervision for the training, which is superior to the combination of cross entropy and Dice loss. In particular, our global contrastive loss function improves segmentation AUC by nearly 2 percentage points; other methods have less impact. This shows that our proposed combination of methods (Method 6) performs better on the ROSE dataset.

Table 7. Ablation experiment on ROSE-1.

Method	AUC	ACC	G-Mean	Kappa	Dice	FDR
1 ¹	0.8874	0.8645	0.7937	0.6673	0.7284	0.2524
2 ²	0.9199	0.9064	0.8285	0.6889	0.7463	0.2114
3 ³	0.9307	0.9092	0.8344	0.7002	0.7561	0.2135
4 ⁴	0.9321	0.9110	0.8375	0.7053	0.7604	0.2057
5 ⁵	0.9304	0.9002	0.8285	0.6974	0.7552	0.2103
6 ⁶	0.9452	0.9133	0.8402	0.7097	0.7645	0.2013

¹ Cross-entropy loss + Dice loss. ² Mse loss. ³ Mse loss + contrastive loss. ⁴ Mse loss + contrastive loss + image augmentation (CLANE). ⁵ Cross-entropy loss + Dice loss + contrastive loss + image augmentation (CLANE) + fine-tune.

⁶ Mse loss + contrastive loss + image augmentation (CLANE) + fine-tune.

4. Discussion

Experimental results show that our COSNet is generally superior to existing methods. In the case of small segmentation targets and class imbalance, methods that employ only strongly supervised loss functions (e.g., U-Net) usually perform poorly. The reason is that the strongly supervised loss pays attention only to information in single images, but the classes in a single image are inherently in an unbalanced state, which leads to the model inadequately learning information of minority class (vessels).

We believe that the essence of image segmentation lies in constructing a structured feature space (smaller intra-class distance and larger inter-class distance). Therefore, in the network architecture, our method extracts both local information (single images) and global information (same-class samples), which effectively alleviates class imbalance. Due to the characteristics of contrast learning, categorization by cosine similarity, distinctive target (vessels) and background are easily segmented. The segmentation results could be utilized for clinical diagnosis by the ratio of superficial and deep vessels width (e.g., Alzheimer's disease). The previous method was only designed for processing images containing superficial vascular annotations, so the clinical application is less extensive.

The experiments also find that using different sampling strategies has a great impact on the results, which also validates previous studies on contrast learning. The selection of sampling strategies needs to be further investigated with regard to the features of data. In the fine-tune stage, the breakpoint problem has not been completely solved, so there is still potential for improvement.

5. Conclusions

In summary, we propose a segmentation method based on global contrast learning for segmenting complex structured OCTA images, which includes a two-branch network architecture and a new pixel contrastive learning function. The core idea is to enable neural networks to learn similar features from a single image and the whole dataset simultaneously. In the scene of class imbalance, our method results in better segmentation accuracy. At the same time, COSNet is also transferable to other class-imbalanced classification or segmentation tasks. The segmentation results of retinal vascular plexus are available for the analysis and prediction of ocular related diseases, and some recent studies [30–32] have been published.

However, the resultant breakpoint of complex-structured vessel segmentation is still a problem that needs to be solved. We hope that more researchers will choose to open source their results in the future to promote the development of medical images.

Author Contributions: Methodology, Z.M. and D.F.; software, D.F.; experiment, D.F.; data processing, J.W. and H.M.; writing—original draft preparation, D.F.; writing—review and editing, D.F.; supervision, Z.M. and D.F.; project administration, D.F.; funding acquisition, Z.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is jointly supported by the National Natural Science Foundation of China (No. 61462002), Natural Science Foundation of Ningxia (No. 2022AAC03268, 2020AAC03215) and the Special Project of North Minzu University (FWNX21, 2021KJCX09).

Institutional Review Board Statement: The datasets we use has been open sourced, which have appropriate approvals from the institutional ethics committees, and written informed consent was obtained from each participant in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: ROSE dataset link: <https://imed.nimte.ac.cn/dataofrose.html>; OCTA-500 dataset link: <https://ieee-dataport.org/open-access/octa-500>; Download project code at <https://github.com/Pantherk1/COSNet>.

Acknowledgments: Thanks to the Intelligent Medical Imaging Research Group from CNITECH and Qiang Chen's team at Nanjing University of Science and Technology for providing the dataset.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OCTA	Optical coherence tomography angiograph
MLP	Multilayer perceptron
CLANE	Contrast-limited adaptive histogram equalization
COSNet	Contrastive OCTA segmentation net
MSE	Mean-squared error
MoCo	Momentum contrast
SVC	Superficial vascular complexes
DVC	Deep vascular complexes
AUC	Area under curve
ACC	Accuracy
TP	True positive
TN	True negative
FP	False positive
FN	False negative
ILM	Internal limiting membrane
OPL	Outer plexiform layer

References

- Akil, H.; Huang, A.S.; Francis, B.A.; Sadda, S.R.; Chopra, V. Retinal vessel density from optical coherence tomography angiography to differentiate early glaucoma, pre-perimetric glaucoma and normal eyes. *PLoS ONE* **2017**, *12*, e0170476. [[CrossRef](#)] [[PubMed](#)]
- Zhao, Y.; Zheng, Y.; Liu, Y.; Yang, J.; Zhao, Y.; Chen, D.; Wang, Y. Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy. *IEEE Trans. Med. Imaging* **2016**, *36*, 51–63. [[CrossRef](#)] [[PubMed](#)]
- Carnevali, A.; Mastropasqua, R.; Gatti, V.; Vaccaro, S.; Mancini, A.; D'aloisio, R.; Lupidi, M.; Cerquaglia, A.; Sacconi, R.; Borrelli, E.; et al. Optical coherence tomography angiography in intermediate and late age-related macular degeneration: Review of current technical aspects and applications. *Appl. Sci.* **2020**, *10*, 8865. [[CrossRef](#)]
- López-Cuenca, I.; Salobarra-García, E.; Gil-Salgado, I.; Sánchez-Puebla, L.; Elvira-Hurtado, L.; Fernández-Albarral, J.A.; Ramírez-Toraño, F.; Barabash, A.; de Frutos-Lucas, J.; Salazar, J.J.; et al. Characterization of Retinal Drusen in Subjects at High Genetic Risk of Developing Sporadic Alzheimer's Disease: An Exploratory Analysis. *J. Pers. Med.* **2022**, *12*, 847. [[CrossRef](#)] [[PubMed](#)]
- Li, Y.; Wang, X.; Zhang, Y.; Zhang, P.; He, C.; Li, R.; Wang, L.; Zhang, H.; Zhang, Y. Retinal microvascular impairment in Parkinson's disease with cognitive dysfunction. *Park. Relat. Disord.* **2022**, *98*, 27–31. [[CrossRef](#)] [[PubMed](#)]
- Azzopardi, G.; Strisciuglio, N.; Vento, M.; Petkov, N. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Med. Image Anal.* **2015**, *19*, 46–57. [[CrossRef](#)] [[PubMed](#)]
- Liu, X.; Deng, Z.; Yang, Y. Recent progress in semantic image segmentation. *Artif. Intell. Rev.* **2019**, *52*, 1089–1106. [[CrossRef](#)]
- Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc. IEEE* **2021**, *109*, 820–838. [[CrossRef](#)]
- Badar, M.; Haris, M.; Fatima, A. Application of deep learning for retinal image analysis: A review. *Comput. Sci. Rev.* **2020**, *35*, 100203. [[CrossRef](#)]
- Serte, S.; Serener, A.; Al-Turjman, F. Deep learning in medical imaging: A brief review. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4080. [[CrossRef](#)]
- Suganyadevi, S.; Seethalakshmi, V.; Balasamy, K. A review on deep learning in medical image analysis. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 19–38. [[CrossRef](#)]
- Zheng, W.; Liu, X.; Yin, L. Research on image classification method based on improved multi-scale relational network. *PeerJ Comput. Sci.* **2021**, *7*, e613. [[CrossRef](#)] [[PubMed](#)]
- Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; Santamaría, J.; Duan, Y. Robust application of new deep learning tools: An experimental study in medical imaging. *Multimed. Tools Appl.* **2022**, *81*, 13289–13317. [[CrossRef](#)]
- Staal, J.; Abramoff, 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](#)]
- Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S.A. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [[CrossRef](#)] [[PubMed](#)]

16. Li, M.; Chen, Y.; Ji, Z.; Xie, K.; Yuan, S.; Chen, Q.; Li, S. Image projection network: 3D to 2D image segmentation in OCTA images. *IEEE Trans. Med. Imaging* **2020**, *39*, 3343–3354. [[CrossRef](#)] [[PubMed](#)]
17. Ma, Y.; Hao, H.; Xie, J.; Fu, H.; Zhang, J.; Yang, J.; Wang, Z.; Liu, J.; Zheng, Y.; Zhao, Y. ROSE: A retinal OCT-angiography vessel segmentation dataset and new model. *IEEE Trans. Med. Imaging* **2020**, *40*, 928–939. [[CrossRef](#)]
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
19. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020), Barcelona, Spain, 4–8 May 2020*; pp. 1055–1059.
20. Isensee, F.; Jaeger, P.F.; Kohl, S.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [[CrossRef](#)]
21. Tomar, N.K.; Jha, D.; Riegler, M.A.; Johansen, H.D.; Johansen, D.; Rittscher, J.; Halvorsen, P.; Ali, S. Fanet: A feedback attention network for improved biomedical image segmentation. *arXiv* **2022**, arXiv:2103.17235.
22. Peng, J.; Wang, P.; Desrosiers, C.; Pedersoli, M. Self-paced contrastive learning for semi-supervised medical image segmentation with meta-labels. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 16686–16699.
23. Zhao, X.; Vemulapalli, R.; Mansfield, P.A.; Gong, B.; Green, B.; Shapira, L.; Wu, Y. Contrastive learning for label efficient semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021*; pp. 10623–10633.
24. Zhong, Y.; Yuan, B.; Wu, H.; Yuan, Z.; Peng, J.; Wang, Y.X. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021*; pp. 7273–7282.
25. Alonso, I.; Sabater, A.; Ferstl, D.; Montesano, L.; Murillo, A.C. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021*; pp. 8219–8228.
26. Zhang, H.; Wu, C.; Zhang, Z.; Zhu, Y.; Lin, H.; Zhang, Z.; Sun, Y.; He, T.; Mueller, J.; Manmatha, R.; et al. Resnest: Split-attention networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022*; pp. 2736–2746.
27. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020*; pp. 9729–9738.
28. Kalantidis, Y.; Sariyildiz, M.B.; Pion, N.; Weinzaepfel, P.; Larlus, D. Hard negative mixing for contrastive learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21798–21809.
29. Li, M.; Huang, K.; Xu, Q.; Yang, J.; Zhang, Y.; Ji, Z.; Xie, K.; Yuan, S.; Liu, Q.; Chen, Q. OCTA-500: A Retinal Dataset for Optical Coherence Tomography Angiography Study. *arXiv* **2020**, arXiv:2012.07261. <https://doi.org/10.48550/ARXIV.2012.07261>.
30. Li, Y.; Zeghlache, R.; Brahim, I.; Xu, H.; Tan, Y.; Conze, P.H.; Lamard, M.; Quellec, G.; Daho, M.E.H. Segmentation, Classification, and Quality Assessment of UW-OCTA Images for the Diagnosis of Diabetic Retinopathy. *arXiv* **2022**, arXiv:2211.11509. <https://doi.org/10.48550/ARXIV.2211.11509>.
31. Yuan, M.; Wang, W.; Kang, S.; Li, Y.; Li, W.; Gong, X.; Xiong, K.; Meng, J.; Zhong, P.; Guo, X.; et al. Peripapillary Microvasculature Predicts the Incidence and Development of Diabetic Retinopathy: An SS-OCTA Study. *Am. J. Ophthalmol.* **2022**, *243*, 19–27. [[CrossRef](#)] [[PubMed](#)]
32. Turker, I.C.; Dogan, C.U.; Dirim, A.B.; Guven, D.; Kutucu, O.K. Evaluation of early and late COVID-19-induced vascular changes with OCTA. *Can. J. Ophthalmol.* **2022**, *57*, 236–241. [[CrossRef](#)] [[PubMed](#)]