sensors

MDPI

Article

Unsupervised Domain Adaptation with Shape Constraint and Triple Attention for Joint Optic Disc and Cup Segmentation

Fengming Zhang †, Shuiwang Li † and Jianzhi Deng *

Guangxi Key Laboratory of Embedded Technology and Intelligent Information Processing, College of Information Science and Engineering, Guilin University of Technology, Guilin 541006, China
* Correspondence: dengjzh@glut.edu.cn
† These authors contributed equally to this work.

Abstract: Currently, glaucoma has become an important cause of blindness. At present, although glaucoma cannot be cured, early treatment can prevent it from getting worse. A reliable way to detect glaucoma is to segment the optic disc and cup and then measure the cup-to-disc ratio (CDR). Many deep neural network models have been developed to autonomously segment the optic disc and the optic cup to help in diagnosis. However, their performance degrades when subjected to domain shift. While many domain-adaptation methods have been exploited to address this problem, they are apt to produce malformed segmentation results. In this study, it is suggested that the segmentation network be adjusted using a constrained formulation that embeds prior knowledge about the shape of the segmentation areas that is domain-invariant. Based on IOSUDA (i.e., Input and Output Space Unsupervised Domain Adaptation), a novel unsupervised joint optic cup-to-disc segmentation framework with shape constraints is proposed, called SCUDA (short for Shape-Constrained Unsupervised Domain Adaptation). A shape constrained loss function is novelly proposed in this paper which utilizes domain-invariant prior knowledge concerning the segmentation region of the joint optic cup–optical disc of fundus images to constrain the segmentation result during network training. In addition, a convolutional triple attention module is designed to improve the segmentation network, which captures cross-dimensional interactions and provides a rich feature representation to improve the segmentation accuracy. Experiments on the RIM-ONE_r3 and Drishti-GS datasets demonstrate that the algorithm outperforms existing approaches for segmenting optic discs and cups.

Keywords: unsupervised; shape constraint; attention

check for updates

Citation: Zhang, F.; Li, S.; Deng, J. Unsupervised Domain Adaptation with Shape Constraint and Triple Attention for Joint Optic Disc and Cup Segmentation. *Sensors* **2022**, *22*, 8748. https://doi.org/10.3390/s22228748

Academic Editors: Sang-Woong Lee, O-Joun Lee, Muhammad Adnan Khan and Ngoc Dung Bui

Received: 3 October 2022
Accepted: 9 November 2022
Published: 12 November 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

Glaucoma is the second most common blinding disease after cataracts [1]. Using fundus pictures, the ratio of the vertical height of the optic cup to the optical disc can be used to determine an early diagnosis of glaucoma. Therefore, it has become a hot topic of research to accurately delineate the optic cup from the optic disc in fundus images and to accurately perform the CDR calculation. At present, deep learning-based techniques for segmenting the optic cup-optical disc have been proven to be effective and have attracted increasing attention in the field. Sevastopolsky et al. [2], for example, proposed a U-Net deep learning network-based method for segmenting the optic cup-optical disc by minimizing the number of convolutional kernels and network complexity. Fu et al. [3] proposed converting the Cartesian coordinates of fundus images into polar coordinate form, and used a U-Net neural network with multi-scale inputs and multi-scale outputs to achieve better performance in optic cup–optical disc segmentation. Most optic cup–optical disc segmentation models work best when the distribution of the test set and training set are the same. Nevertheless, these models tend to perform worse when applied to target domains other than the one they were trained on. This problem is known as a

domain shift or distributional shift. Domain adaptation is usually utilized to cope with this problem. According to the information considered for the target task, domain adaptation can be divided into three types, namely, unsupervised, semi-supervised, and supervised domain adaptation. Among them, unsupervised domain adaptation is the one we are most concerned with here. A number of unsupervised domain adaptive algorithms have been proposed for the mitigation of domain shifts in biomedical image segmentation [4–6]. For instance, studies of source and target domain domains based on common invariance properties [4,5] concentrate on partitioning the input space of the network. In order to ensure that the segmentation network's output space is invariant and that the segmentation maps of the source and target regions have the same spatial and geometric shape, [6] employed adversarial learning. Chen et al. [7] proposed an unsupervised framework called IOSUDA for the joint segmentation of the optic cup and optical disc. This framework focuses on separating shared features and stylized features for feature alignment, achieving input and output space alignment, and reducing performance degradation. Although these methods have achieved remarkable performance, they are apt to produce malformed segmentation regions, as demonstrated in Figure 1, that are very far from the real shapes of the optic cup and optical disc. Here, we propose to overcome this issue using a formulation with constraints that, based on the shape of the segmentation region, contain domain-invariant prior information for segmentation networks. The intuition behind our work is that shape information is a strong and valuable prior for optic cup and disc segmentation, as geometrically the optic cup or disc is very close to a round shape. The effectiveness of shape constraints has been proven very recently in 3D pancreas segmentation [8], motivating us to make use of it for the task at issue here. As seen in Figure 1, our method is capable of providing more realistic segmentation results with the proposed shape constraint.

On the other hand, the U-Net [9], a very effective but highly underutilized network introduced by Ronneberger et al. in 2015 for medical image segmentation, serves as the segmentation sub-network in IOSUDA. In order to locate and extract invariant features from the dataset, Zhang et al. [10] suggested a transferable attention U-Net model that used two discriminators and an attention module. Zhao et al. [11] added an attention gate between the encoder-decoder of U-Net in order to concentrate more on the target region, resulting in an attention U-Net architecture. These works suggest that attention mechanisms are effective in boosting the performance of U-Net, which inspires us to attempt a more advanced attention approach for further improvement. Recently, the use of channel attention, spatial attention, or both has been suggested in several studies on computer vision problems as a way to enhance the feature representation ability of by convolutional layers in order to enhance the performance of neural networks. For instance, the Squeeze-and-Excitation (SE) module [12] calculates channel attention and improves performance at a fraction of the cost. Moreover, the Convolutional Block Attention Module (CBAM) [13] and the Bottle-neck Attention Module (BAM) [14] both emphasise the combination of spatial attention and channel attention. Both the BAM (i.e., Bottle-neck Attention Module) and CBAM (i.e., Convolutional Block Attention Module) place emphasis on the union of spatial and channel attention. The Convolutional Triple Attention Module [15] is a lightweight yet effective attention mechanism that calculates attention weights by way of capturing interactions of cross dimensions using a three-branch structure. The segmentation performance of the segmentation sub-network U-Net is improved in this paper using a Convolutional Triple Attention Module (CTAM).

The following may be said about this paper's contributions:

- We propose a novel unsupervised adaptive framework with shape constraint, called SCUDA, for joint segmentation of the optic cup–optical disc in order to address the problem that existing methods are very likely to produce malformed segmentation regions.
- We exploit a convolutional triple attention module to improve the segmentation network, which is able to capture cross-dimensional interactions and provides rich feature representation in order to boost segmentation accuracy.

- We conducted a number of extensive experiments on the RIM-ONE_r3 dataset and the Drishti-GS dataset to demonstrate the performance of our performed SCUDA framework. The experimental findings verify that SCUDA outperforms the other tested model in terms of performance.

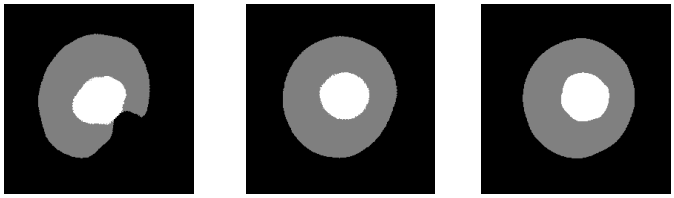


Figure 1. Comparison of the segmentation results between the state-of-the-art method IOSUDA [16] and our SCUDA method on a fundus image. The abbreviation GT refers to ground truth.

The remainder of the paper is structured as follows: we review related work and describe our methodology in Sections 2 and 3, respectively; experimental findings are discussed in Section 4; and the work is concluded in Section 5.

2. Related Work

2.1. Unsupervised Domain Adaptation

A fairly common type of transfer learning is domain adaptation, which generally refers to using a model from one domain and apply it another domain that is only subtly different [17]. Unsupervised domain adaptation in classification is generally built on image and feature alignment [18–21] between source and target domains. For instance, Long et al. [22] proposed a new network architecture, Deep Adaptation Network (DAN), that used an optimal multi-core selection method for average embedding matching and was able to reduce domain differences. Bousmalis et al. [18] considered shared and private representations of each domain. Unsupervised domain adaptive segmentation has been used for many scenarios, including across various medical images. For example, according to Chen et al. [23], the network can be trained using images from the source domain, with the target domain's image style being the same as that of the source domain. Huo et al. [24] proposed a Synthetic Segmentation Network (SynSegNet) in order to stylize images from the source domain into those from the target domain. Song et al. [25] introduced several assumptions for feature space extraction; based on this, each loss function was derived and optimized. In addition, to compare the feature spaces of the source, target, and output domains with one another, Chen et al. [26] proposed Synergistic Image and Feature Alignment (SIFA).

2.2. Optic Cup–Optical Disc Segmentation

Early work in optic cup–optical disc segmentation focused on hand-crafted features [27–29], usually implemented first for target region detection [30,31]. Convolutional neural network-based approaches [3,32,33] have significantly improved accuracy and generalizability. A convolutional neural network for segmentation based on lifting trees was designed by Zilly et al. [32]. Fu et al. [33] proposed the Disc-aware Ensemble Network (DENet) for automated glaucoma screening, which integrates data from local optic disc regions with features from global fundus images. A U-Net based M-Net was proposed by Fu et al. [3] to segment the optic disc-cup, with the segmentation issue converted to a multi-label issue. In addition, a number of semi-supervised methods [34,35] have been proposed to alleviate the problem of insufficient truth labels of the original data. However, these models lack generalization in the face of domain shifting. Recently, unsupervised domain adaptation has made a splash in segmentation of optic cup–optical disc cross data sets [7,36,37]. In order to solve instability in adversarial learning, Liu et al. [36] pro-

posed Collaborative Feature Ensembling Adaptive (CFEA), which makes use of adversarial learning for both the network's output and intermediate representations. Wang et al. [37] proposed Boundary- and Entropy-driven Adversarial Learning (BEAL), which employs two boundary and entropy discriminators to effectively solve the problem of a target domain's high entropy and fuzzy boundary. For joint segmentation of the optic disc and cup, Chen et al. [7] proposed an IOSUDA framework including feature and output space alignment while simultaneously introducing adversarial learning into the learning process of segmentation networks; shared features of multiple domains are introduced in the input space. In this paper, we propose an unsupervised domain adaptation with a shape constraint for joint optic disc and cup segmentation. The comparison of our method with previous approaches in terms of used dataset, learning method, supervision method, and use of U-Net, GAN, attention mechanism, and prior geometric constraint (or not) is summarized in Table 1.

Table 1. Comparison of advantages and disadvantages of our method (SCUDA) with previous methods.

Method	Year	Dataset	Learning Method	Supervisio Method	U-Net	GAN	Attention	Geometric Constraint
[28]	2011	Non-public	Traditional learning	Supervised	×	×	×	×
[27]	2013	SiMES SCES			×	×	×	×
[29]	2013	Non-public			×	×	×	Disc contains cup
[32]	2015	Drishti-GS	Deep learning	Supervised	×	×	×	×
[3]	2018	ORIGA			√	×	×	×
[33]	2018	SCES SINDI			√	×	×	×
[11]	2021	DRIONS-DB Drishti-GS			√	×	√	×
[35]	2019	ORIGA REFUGE	Deep learning	Semi-supervised	×	√	×	×
[34]	2022	RIGA			√	×	×	×
[36]	2019	REFUGE	Deep learning	Unsupervised	√	√	×	×
[37]	2019	Drishti-GS RIM-ONE-r3 REFUGE			√	√	×	×
[7]	2021	Drishti-GS RIM-ONE-r3 REFUGE			√	√	×	×
Ours	2022	Drishti-GS RIM-ONE-r3 REFUGE			√	√	√	Circular-like region

2.3. Attentional Mechanism

In recent years, many researchers have proposed combining attention mechanisms with deep Convolutional Neural Networks (CNNs) to improve large-scale visualization. Double Attention Networks (A^2-Nets) [38] use a "double attention block" method that counts and propagates information-rich global features from the input image/video over the entire time and space. Global Second Order Pooling Network (GSoP-Net) [39] uses second-order pooling to collect important features from the entire input space and then distributes them to make further layers easier to verify and disseminate. In addition, an innovative NL block combined with an SE block has been proposed by Global-Context Networks (GC-Net) [40] in an effort to more effectively combine contextual representation with channel weighting. Images can be be segmented and classified using attention processes as well. Criss-Cross networks (CCNet) [41] and SPNet [42] have proposed a new cross-attention module that captures rich contextual information on its cross-paths. A

pipeline based on two top-down and two bottom-up attention modules has been presented by Xiao et al. [43] for classifying images.

3. Our Approach

3.1. SCUDA Framework

The proposed SCUDA model inherits the IOSUDA pipeline, and is formed from two parts: the image translation model and the segmentation model. Figure 2 shows the overview of SCUDA. The images from the source domain (X_s), the truth labels from the source domain (Y_s), and the images from the target domain (X_t) are the data utilized in training. The picture translation model applies unsupervised transformation between the source and target domains with the goal of learning the shared content features and the corresponding style features. Here, X_{s-t} denotes the transformed image dataset; conversely, X_{s-s} denotes the reconstructed image dataset, while the combination of content and style features is represented by the symbol $\oplus$. In addition, a shape-constrained loss function L_{shape} for segmentation is designed to incorporate the prior shape information of the segmentation region of the optic cup–optical disc, with the purpose of constraining the segmentation region predicted by the network to ensure that it lies within a feasible configuration space. Moreover, a convolutional triple attention module (CTAM) is adopted for purpose of improving the codec of the segmentation network, which can establish interdependencies between channels or spatial locations to achieve cross-dimensional interactions and boosts segmentation performance. The datasets generated by target-domain and source-domain segmentation are denoted by the variables Y_t and Y_s, respectively. The segmentation network may be optimized via adversarial learning of the segmentation maps of the source and target domains. Additionally, the segmentation maps produced by the target domain are superior.

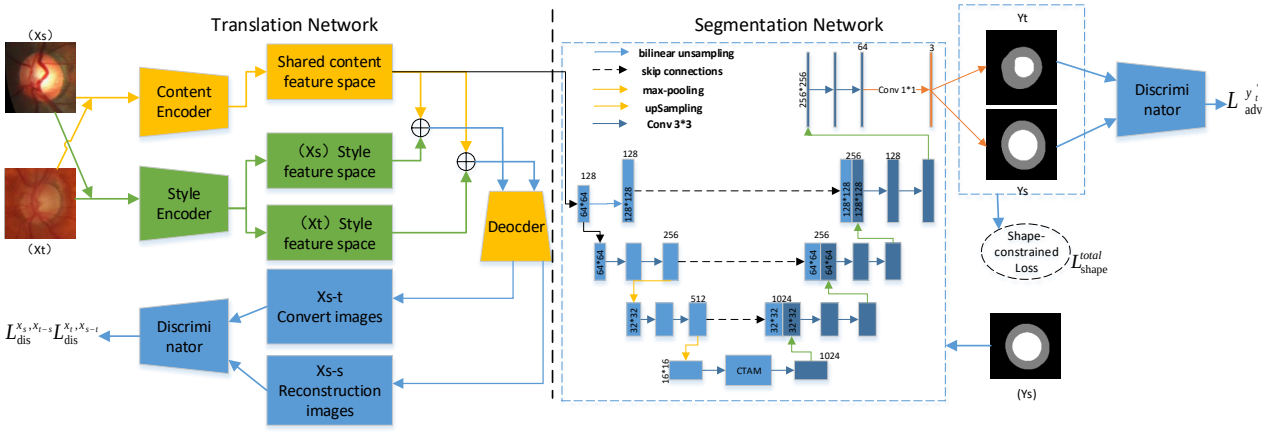


Figure 2. Overview of our proposed SCUDA framework. The left side of the split line is the image translation sub-model and the right side is the image segmentation sub-model; X_{s-s} denotes the reconstructed image dataset and X_{s-t} denotes the converted image dataset, while Y_t and Y_s denote the segmentation map datasets produced by the splitter network. Additionally, the symbol $\oplus$ shows the combination of content and style features.

3.2. The Proposed Shape-Constrained Loss Function

Shape information is an important and meaningful a priori indicator for organ segmentation in medical images. Although different datasets of fundus images may appear quite different due to scanning machines, procedures, stages, etc., they should have the same representation of anatomical structures, i.e., contours, of the optic cup and optical disc, which are all circular-like, or at least not very far from a circle. This prior shape information can be used as an indicator to constrain the segmentation results. Specifically, the result of segmentation of a fundus image corresponds to the two contour boundaries of the optic cup and optical disc, respectively, as shown in Figure 3. In the GT diagram in Figure 3, the

green contour line indicates the optic disc segmentation and the blue contour line indicates the optic cup segmentation. We denote the set of coordinates of the contour boundaries of the optic cup–optical disc as I_{cup} and I_{disc}, respectively. Accordingly, the equation for calculating the center of mass of the optic disc is expressed as

$$(C_X, C_Y) = C = \frac{1}{k}\sum_{i=1}^{k} m_i, m_i \in I_{disc}, \tag{1}$$

where (C_X, C_Y) denotes the centroid of the optic disc, C_X, C_Y are the x-coordinate and y-coordinate component, respectively. Similarly, the equation for the center of mass of the optic cup is expressed as

$$(D_X, D_Y) = D = \frac{1}{k}\sum_{i=1}^{k} n_i, n_i \in I_{cup}, \tag{2}$$

where (D_X, D_Y) is the centroid of the apparent cup, D_X is the x-coordinate of the center of mass of the apparent cup, and D_Y is the y-coordinate of the center of mass of the apparent cup. An illustration of computed centroids is shown in Figure 4, marked by dots.

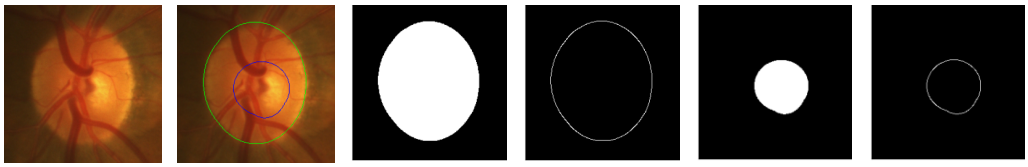


Figure 3. Illustration of the region and boundary of the optic cup and optical disc.

Figure 4. Illustration of the estimated centroids of optic disc (**left**) and optic cup (**right**), which have been marked in color.

If the contour boundary of a region is a circle, the distances of each point on the contour to its centroid are equal, and consequently are the same as their mean. In view of this, we use the mean deviation of distances from their mean as a measure of the deviation of a circular-like contour, which is normalized by dividing the mean distance in order to eliminate the scale variations. The proposed shape-constrained loss function for segmenting the optic cup is formulated as follows:

$$L_{cup} = \sqrt{\sum_{i=1}^{k}\left(E_{cup}^{i} - m_{cup}\right)^2}/m_{cup}, \tag{3}$$

where $E_{cup}^{i} = \|C - m_i\|^2$, $m_i \in I_{cup}$ is the distance of the ith point $(i \in [1, k])$ on the contour of optic cup region to its centroid, k denotes the number of points on the discrete contour, and m_{cup} represents the mean distance, defined by

$$m_{cup} = \frac{\sum_{i=1}^{k} E_{cup}^{i}}{k}. \tag{4}$$

By the same token, we can define the shape-constrained loss function for segmenting the optic disc, which is denoted by L_{disc}. Taken together, we obtain the shape constrained loss function L_{shape} for segmenting fundus images as follows:

$$L_{shape} = L_{disc} + L_{cup}. \tag{5}$$

3.3. Total Loss Function

The loss function of the SCUDA framework includes the loss of the image translation module and the loss of the image segmentation module in addition to the shape-constrained loss. For the image translation module, let E_C denote the content encoder, E_S the style encoder, C the shared content feature space, S_S and S_T the style feature space in the source domain and the target domain, respectively, G the shared decoder, and $L1$ the $L1$ distance. For a source domain image $x_s \in X_S$, $c_s, c_t \in C$, $s_s \in S_S$, $s_t \in S_T$, the source domain image loss $L_{rec}^{x_s}$, the source domain image content feature loss $L_{rec}^{c_s}$, and the source domain image style feature loss $L_{rec}^{s_s}$ are defined as follows:

$$L_{rec}^{x_s} = \mathbf{E}_{x_s}[G(E_C(x_s), E_S(x_s)) - x_s]_{L1}, \tag{6}$$

$$L_{rec}^{c_s} = \mathbf{E}_{c_s,s_t}[E_C(G(c_s, s_t)) - c_s]_{L1}, \tag{7}$$

$$L_{rec}^{s_s} = \mathbf{E}_{c_t,s_s}[E_S(G(c_t, s_s)) - s_s]_{L1}, \tag{8}$$

where $\mathbf{E}_z$ indicates computing the expectation of a function of z. The target domain image loss $L_{rec}^{x_t}$, its content feature loss $L_{rec}^{c_t}$, and its style feature loss $L_{rec}^{s_t}$ are defined analogously to the loss of the source domain image. For source domain to target domain image translation, the discriminator D_1 aims to distinguish the target domain image x_t from the transformed image x_{s-t}, while the discriminator D_2 aims to distinguish the source domain image x_s from the transformed image x_{t-s}, with the former loss function being denoted by $L_{dis}^{x_t,x_{s-t}}$ and the latter by $L_{dis}^{x_s,x_{t-s}}$. The $L_{dis}^{x_t,x_{s-t}}$ loss is formulated as

$$L_{dis}^{x_t,x_{s-t}} = \mathbf{E}_{c_s,s_t}[\log(1 - D_1(G(c_s, s_t)))] + \mathbf{E}_{x_t}[\log D_1(x_t)]. \tag{9}$$

The $L_{dis}^{x_s,x_{t-s}}$ loss function is defined similarly to $L_{dis}^{x_t,x_{s-t}}$. The total loss of the image translation model is defined as follows:

$$L_{tra}^{total} = \mu_1(L_{rec}^{x_s} + L_{rec}^{x_t}) + \mu_2(L_{rec}^{c_s} + L_{rec}^{c_t}) + \mu_3(L_{rec}^{s_s} + L_{rec}^{s_t}) + \mu_4(L_{dis}^{x_t,x_{s-t}} + L_{dis}^{x_s,x_{t-s}}), \tag{10}$$

where μ_1, μ_2, μ_3, μ_4 denote the weights of each component. In the image segmentation module, the segmentation of the optic cup–optical disc is converted to a multi-classification assignment with the segmentation label map $y_s \in R^{H\times W\times C}$, where $H \times W$ is the image height and width and C is the number of categories. The segmentation network takes c_s as input to obtain a predicted segmentation map $y_s^{'}$; similarly, c_t is taken as input to obtain a predicted segmentation map $y_t^{'}$. In addition, the role of the discriminator D is to determine that $y_s^{'}$ is true and $y_t^{'}$ is spurious. The output size of the discriminator is $m \times n$, and its loss function is defined by

$$L_{dis}^{y_s^{'},y_t^{'}} = \sum_{m,n} \log((D(y_s^{'}))^{(m,n)}) + \log(1 - (D(y_t^{'}))^{(m,n)}). \tag{11}$$

The split loss function of $y_s^{'}$ is as follows:

$$L_{seg}^{y_s,y_s^{'}} = -\sum_{h\in H,w\in W}\sum_{c\in C} y_s^{(h,w,c)} \log {y_s^{'}}^{(h,w,c)}. \tag{12}$$

In order to make $y_t^{'}$ and $y_s^{'}$ have similar definitions, the discriminator is confused in order to judge the patches of $y_t^{'}$ as true. The adverse loss is defined by

$$L_{adv}^{y_t^{'}} = \sum_{m,n} \log\left(\left(D(y_t^{'})\right)^{(m,n)}\right) \tag{13}$$

The total loss of the image segmentation network is defined as follows:

$$L_{seg}^{total} = \delta_1(L_{seg}^{y_s,y_s^{'}}) + \delta_2(L_{adv}^{y_t^{'}}). \tag{14}$$

where δ_1 and δ_2 denote the weights of each component. Because of the source and target domains, there are four terms of shape constrained losses during gradient backpropagation. The total shape-constrained loss function L_{shape}^{total} is therefore

$$L_{shape}^{total} = L_{shape}^{x_s} + L_{shape}^{x_t} + L_{shape}^{x_{s-t}} + L_{shape}^{x_{t-s}}. \tag{15}$$

Taken together, the total loss of the proposed model is

$$L_{total}^{'} = L_{tra}^{total} + \rho_1 L_{seg}^{total} + \rho_2 L_{shape}^{total}. \tag{16}$$

3.4. Convolutional Triple Attention Module (CTAM)

The shared feature content obtained by the image translation model is later fed to the segmentation network for segmentation. Concretely, The segmentation network makes use of an adjusted U-Net, and as the shared content features are downsampled from the original image to be used in the segmentation, the first two downsampling layers of the network are eliminated to satisfy the dimensionality requirement. Convolutional Triple Attention Module (CTAM) [15], a compact yet powerful attention module, is designed and deployed to the interface between the innermost encoder and decoder of U-Net in this paper to further enhance the segmentation network. CTAM captures cross-dimensional interactions of a tensor input by establishing inter-dimensional correlations through a rotation operation and subsequent residual transformations. By computing the attention weights, it generates a large number of feature representations and produces a refined tensor with the same form as the input. The detailed structure of the CTAM is shown in Figure 5. CTAM contains three parallel branches, two each to capture the interaction between channel dimension C and a spatial dimension, i.e., H or W. The output of all three lines is determined using a straightforward averaging method, with one line being utilized to develop spatial attention. More specifically, CTAM accepts an input tensor $x \in R^{C\times H\times W}$, where C denotes the number of channels and H and W denote the height and width of the spatial feature mapping, respectively, which is first passed to each of the three branches. The height and the channel dimension create an interaction in the first branch. Then, x is rotated 90° counterclockwise along with the H axis, recorded as x_1 with the shape $(W \times H \times C)$, which is minimized to $x_1^{'}$ with the shape $(2 \times H \times C)$ after Z-pool; $x_1^{'}$ later goes through the convolution layer, followed by a batch normalization layer. Moreover, attention weights are obtained by sending them to the sigmoid activation layer. To retain the basic input form of x, the created weights are employed in x_1 and the result is rotated 90° clockwise along with the H axis. The tensor of the first branch that is generated at the conclusion is defined as x_1^*.

Likewise, in the second branch, a 90° counterclockwise rotation along the W axis is applied to x with the same principle as in the first branch to obtain the refined x_2^*. The last branch, where the z-pool reduces the channels of the input tensor x to two, produces x_3, which has the shape $(2 \times H \times W)$, and is then processed by a convolution layer. Then, it proceeds successively through a batch normalization layer. Through the sigmoid activation layer, the output generates an attention weight with the shape $(1 \times H \times W)$; the tensor of the final branch generated at the end is defined as x_3^*. The refined tensor of shape $(C \times H \times W)$ generated by a simple averaging pool of data generated by three branches.

To sum up, for an input tensor $x \in R^{C\times H\times W}$, the following equation illustrates how the refinement tensor y is obtained from the three branches:

$$y = \frac{x_1^*\omega_1 + x_2^*\omega_2 + x_3^*\omega_3}{3}. \tag{17}$$

where ω_1, ω_2, and ω_3 are the three cross-dimensional attention weights calculated in the triple attention.

It is worth noting that the incorporation of CTAM into U-Net is based on the following considerations. Despite being widely used, U-Net can be further improved for various segmentation tasks, especially through attention mechanisms, with the motivation of ensuring that the network devotes more focus to the important parts of the data. Remarkably, the parameters related to attention mechanisms can be learned without introducing additional losses. Many works have used attention mechanisms to improve U-Net for medical segmentation [44–47], including segmentation of the optic disc and cup [10,11,48]. However, the attention methods used in these works require quite a number of learnable parameters, which can easily lead them to suffer from overfiting problems in view of the limited training data in many medical segmentation tasks. Fortunately, a cheap and very effective attention method, namely, CTAM, was proposed in [15] with the aim of capturing cross-dimension interaction while computing attention weights to provide rich feature representations. It has demonstrated the ability to provide similar or better performance to the alternatives. In light of these advantages, in this paper we apply this triplet attention method to boost the performance of U-Net. Because the attention triplet receives an input tensor and outputs a refined tensor of the same shape, it can be applied to any layer to enhance the feature representation there. To avoid increasing too many parameters, we only apply it to the deepest layer of U-Net, as this is the layer with the most abstract representation, which we believe should have the greatest effect on the final result. Trivially, the CTAM becomes an identity map when, say, the convolutional layers in the CTAM have zero kernels and the cross-dimensional attention weights sum to 1. Hopefully, a CTAM can be learned that performs better than this trivial case.

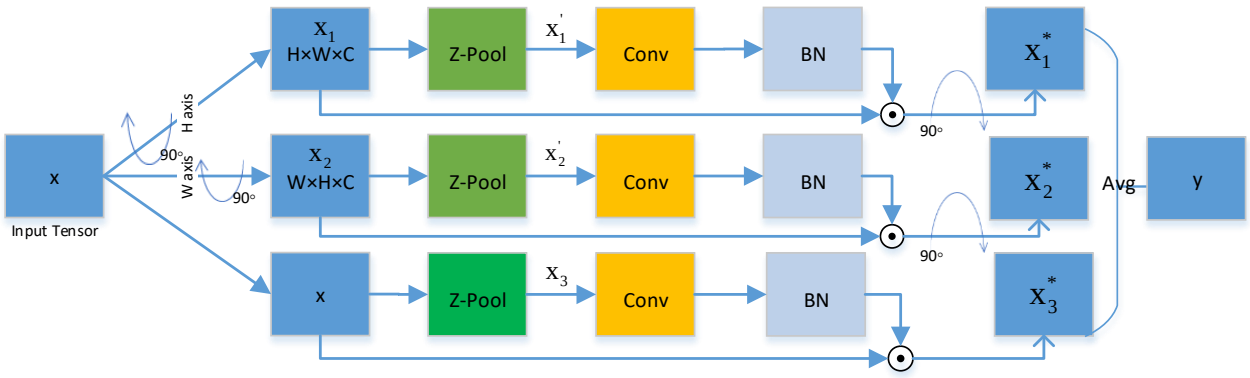


Figure 5. Illustration of the architecture of the convolutional triple attention module (CTAM).

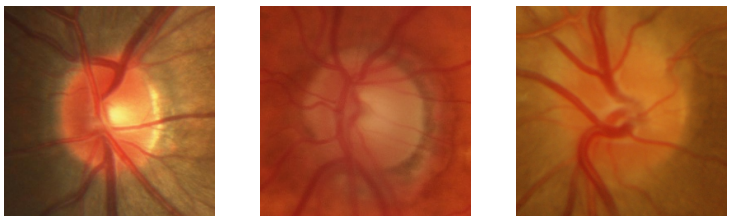
3.5. Implementation Details

The network model for this experiment used the Pytorch framework, and training/testing was performed on an RTX3090 with 24 GB of memory. A pre-trained model [3] was used to locate the optic cup and optical disc region of the fundus images in the experimental dataset. Training images were then obtained by cropping and scaling, and the training images were normalized, randomly inverted, and cropped for input. In addition, random seeds were fixed in the experiment. The size of the input training image was 256 × 256 pixels. The whole model framework was optimized using the Adam method with a batch size of 2 and a training period of 400 epochs, and the initial learning rate was set to 10^{-4}.

4. Experiments

4.1. Datasets

The RIM-ONE_r3 [49] dataset, the Drishti-GS [28] dataset, and the REFUGE [50] dataset were the three publicly available fundus imaging datasets used in this experiment. They have different appearances, as shown in Figure 6. Following [7], the datasets from the source and target domains were split into a training set and a test set for this experiment. The RIM-ONE_r3 dataset with Drishti-GS was employed as the target domain, while the REFUGE training set served as the source domain. Table 2 shows the statistical distribution of the data.



REFUGE RIM-ONE_r3 Drishti-GS

Figure 6. Example fundus images from different datasets. From left to right: REFUGE [50], RIM-ONE_r3 [49], and Drishti-GS [28].

Table 2. Statistical distribution of the RIM-ONE_r3, the Drishti-GS, and REFUGE datasets.

Data	RIM-ONE_R3	Drishti-GS	REFUGE
Image size	1072 × 1424	2047 × 1760	2124 × 2056
Quantity of training images	99	50	400
Quantity of test images	60	51	0
Target domain	Target domain	Target domain	Source Domain

4.2. Evaluation Metrics

This experiment uses the IoU coefficient of the optic cup and optic disc along with their Dice coefficient as evaluation indicators. TP (True Positives), FP (False Positives), and TN (True Negatives) are the number of pixels in the segmentation which match the ground truth (for TN/TP) or do not (FP/FN):

$$Dice = \frac{2TP}{FP + 2TP + FN}, \tag{18}$$

$$IoU = \frac{TP}{FP + TP + FN}. \tag{19}$$

The higher the $Dice$ and IoU values, i.e., the closer they are to 1, the better the segmentation performance of the model. IoU_{OD} and $Dice_{OD}$ denote the IoU and Dice values of the optic disc, respectively, while IoU_{OC} and $Dice_{OC}$ denote the IoU and Dice values of the optic cup, respectively.

4.3. Quantitative and Qualitative Analysis

We compare our method with five state-of-the-art methods for segmenting the optic cup–optical disc on two datasets, namely, RIM-ONE_r3 [49] and Drishti-GS [28], to show the efficacy of the SCUDA framework proposed in this study. The methods for comparison are classified into two types. One kind is a model without domain adaptation, such as CycleGAN [51] and Pix2Pix [16]. CycleGAN is an image transformation model based on mismatch, which can transform fundus images into segmentation images to achieve target segmentation. Numerous studies have utilized Pix2Pix, a conditional adversarial

generative network (cGAN), for segmentation tasks [52,53]. Another type of unsupervised domain adaptive models include SynSeg-Net [24], SIFA [26], and IOSUDA [7]. In the input space, SynSeg-Net provides picture alignment. Feature alignment and output space alignment are combined by SIFA. Therefore, in our evaluation, CycleGAN and Pix2Pix are trained using the source domain dataset. On the other hand, SynSeg-Net, SIFA, IOSUDA, and the SCUDA model proposed in this paper are trained using data from the source domain and the unlabeled target domain's training portion, while the test data come from the target domain. Table 3 reports the experimental results. As can be seen, the RIM-ONE_r3 dataset is more difficult than the Drishti-GS dataset, as all the metrics of the tested methods are significantly lower on the former, reflecting the more severe domain drift of the RIM-ONE_r3 dataset compared to the Drishti-GS dataset. Remarkably, our SCUDA method achieves the best performance in terms of all metrics on both datasets. For example, on the RIM-ONE_r3 dataset, our method outperforms the second-best method, IOSUDA, by 1.83%, 2.02%, 1.66%, and 1.73% in IoU_{OD}, IoU_{OC}, $Dice_{OD}$, and $Dice_{Oc}$, respectively. On the Drishti-GS dataset, our method likewise outperforms the second-best method, again IOSUDA, by 1.26%, 1.70%, 1.41%, and 1.49%, respectively. Results such as those above demonstrate how well our proposed SCUDA model works.

On eight test samples from RIM-ONE_r3 and Drishti-GS, Figures 7 and 8 compare our method qualitatively to two state-of-the-art methods, including the baseline IOSUDA method and SIFA. Concretely, the first and second columns of Figures 7 and 8 show fundus images and the corresponding ground truths, and other columns show the fundus images with the boundary of the optic cup–optical disc marked by different methods. The green contour lines in the figure indicate the optic disc segmentation results and the blue ones indicate the optic cup segmentation results. It can be observed that our SCUDA approach demonstrates better segmentation results with relatively smooth and accurate segmentation contours in all these cases, regardless of whether the fundus images contain clear contours or blur contours, while the other methods produce malformed segmentations in most of these cases. We ascribe this to the effectiveness of the proposed shape constraint, which embeds domain-invariant prior knowledge concerning the circular shape of the optic cup and optical disc into our model.

Table 3. Comparative experimental results of CycleGAN, Pix2Pix, SynSeg-Net, SIFA, IOSUDA, and our proposed SCUDA on the RIM-ONE_r3 test set and the Drishti-GS test set.

Datasets	Model	IoU_{OD} (%)	IoU_{OC} (%)	$Dice_{OD}$ (%)	$Dice_{Oc}$ (%)
RIM-ONE_r3	CycleGAN [51]	70.41	49.76	82.08	64.27
	Pix2Pix [16]	69.57	52.12	81.77	67.81
	SynSeg-Net [24]	71.92	52.69	83.27	67.93
	SIFA [26]	74.67	52.84	84.17	68.03
	IOSUDA [7]	83.06	59.63	90.14	72.32
	SCUDA (Ours)	**84.89**	**61.65**	**91.80**	**74.05**
Drishti-GS	CycleGAN [51]	80.63	45.29	89.12	60.35
	Pix2Pix [16]	82.27	56.02	89.51	69.13
	SynSeg-Net [24]	79.70	49.45	88.36	64.62
	SIFA [26]	83.04	57.29	88.90	70.64
	IOSUDA [7]	89.08	64.91	93.77	77.49
	SCUDA (Ours)	**90.34**	**66.61**	**95.18**	**78.98**

4.4. Ablation Study on the Impact of the Weight of the Shape Constraint

We evaluated the proposed SCUDA on the RIM-ONE_r3 dataset with regard to the weight of the shape constraint loss in order to investigate the effects of the shape constraint weight on the effectiveness of segmentation. The weight ranges were from 0.2 to 2.0 with a step size of 0.2. The four metrics of SCUDA for the different weights are shown in Table 4.

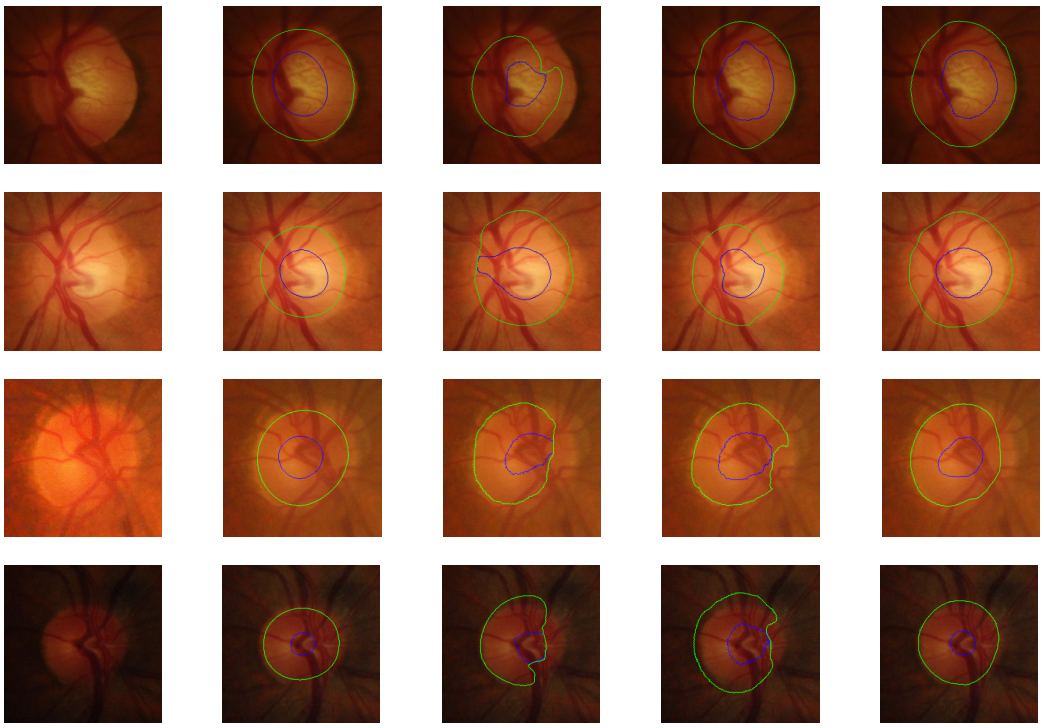


Figure 7. Representative visual examples of the RIM-ONE_r3 test set; the green contours indicate the boundary of the optical disc and the blue contours indicate the boundary of the optic cup. From left to right: original images, GT, and the results of SIFA, IOSUDA, and our proposed SCUDA.

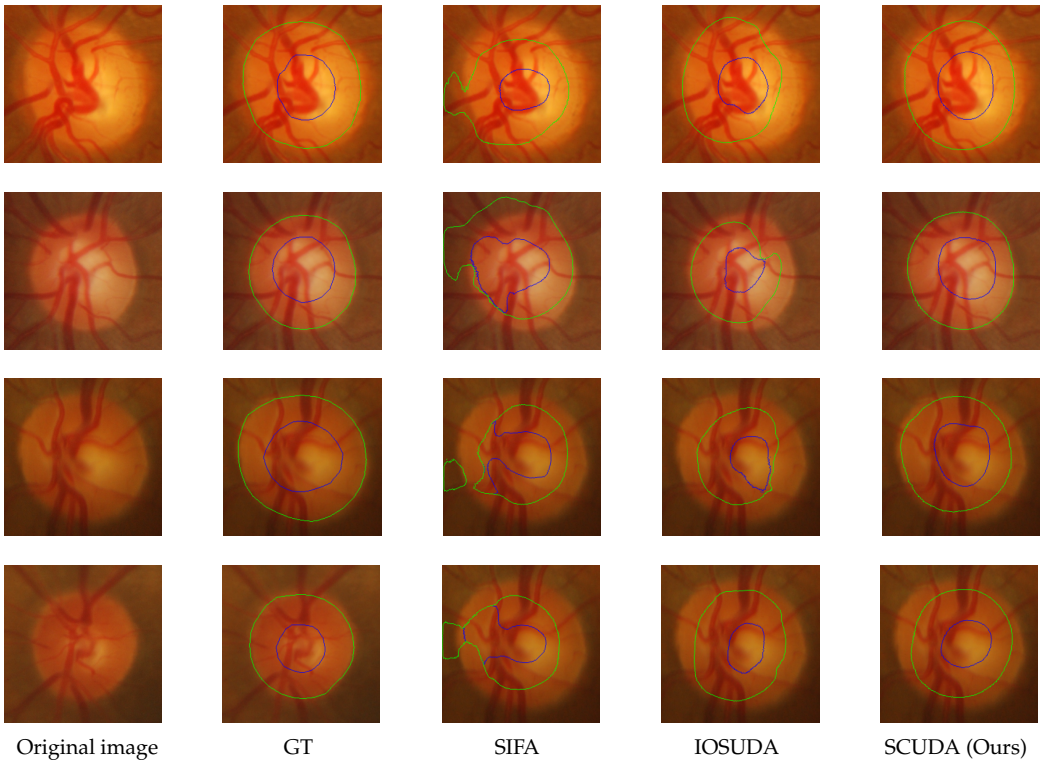


Figure 8. Representative visual examples of the Drishti-GS test set; the green contours indicate the boundary of the optical disc and the blue contours indicate the boundary of the optic cup. From the left to right: original images, GT, and the results of SIFA, IOSUDA, and our proposed SCUDA.

Table 4. Illustration of the varied performance of the proposed method on the RIM-ONE_r3 test set with different weights of the shape constraint loss function.

	0.2	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**
IoU_{OD} (%)	84.74	84.57	84.75	84.49	84.50	**84.84**	83.82	84.82	84.73	84.72
IoU_{OC} (%)	59.80	60.19	60.63	60.42	**61.19**	60.12	60.21	60.12	59.35	60.76
$Dice_{OD}$ (%)	91.50	91.44	91.55	91.41	91.42	**91.65**	90.94	91.59	91.58	91.50
$Dice_{OC}$ (%)	72.55	72.96	73.07	73.19	**73.65**	72.97	72.86	72.87	72.17	72.99

It can be seen that when the weight of the loss function is 1.2, SCUDA achieves the best IoU_{OD} and $Dice_{OD}$, which are 84.84% and 94.65%, respectively; when the weight is 1.0, SCUDA again achieves the best IoU_{OC} and $Dice_{OC}$, which are 61.19% and 73.65%, respectively. Overall, the best performance is obtained when the weight is 1.0, which is the default setting of the weight in our proposed SCUDA. The trend of the average of the four metrics as the weight changes from 0.2 to 1.8 is plotted in Figure 9 to provide a more intuitive grasp of the influence of this weight. Note that the average of the four metrics is plotted by the gray dotted line. As can be seen, when the weight changes from 0.2 to 1.0, the IoU_{OC} and $Dice_{OC}$ show an increasing trend overall, except for a drop at 0.8. Although the increasing trend is not obvious for IoU_OD and Dice_OD, apparent drops can be observed at 1.4 for both metrics. On average, when the weight increases from 1.0 to 1.8, a decreasing trend is observed on the whole, except for a rise at 1.6. These results suggest that the segmentation performance can be improved if the shape constraint is imposed moderately. In order to understand the impact of the weight of the shape constraint more intuitively, we show five examples of segmentation with different weights in Figure 10. It can be seen that, the segmentation results become visually better and better as the weight goes from 0.4 to 1.0. This justifies the effectiveness of the shape constraint for optic cup–optical disc segmentation and conforms to the fact that the shape-constrained loss function is based on an approximately (though not strictly) correct assumption, namely, that a constraint that is too strong leads to false prior information being imposed on the trained model.

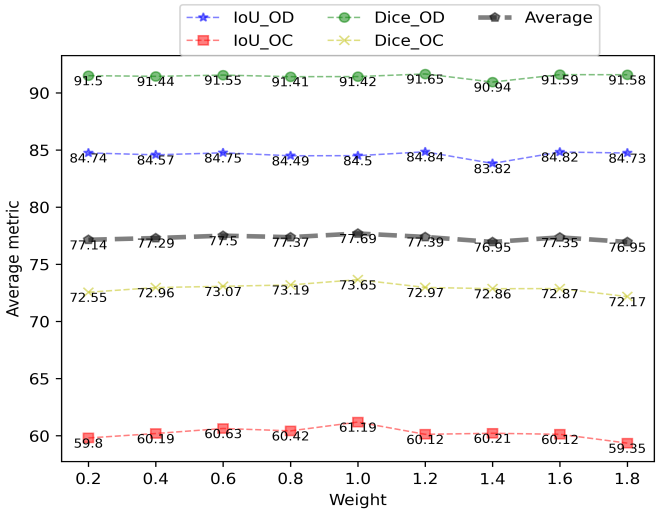


Figure 9. Illustration of trends in the values and means of the four indicators relative to the weights of the shape constraint loss.

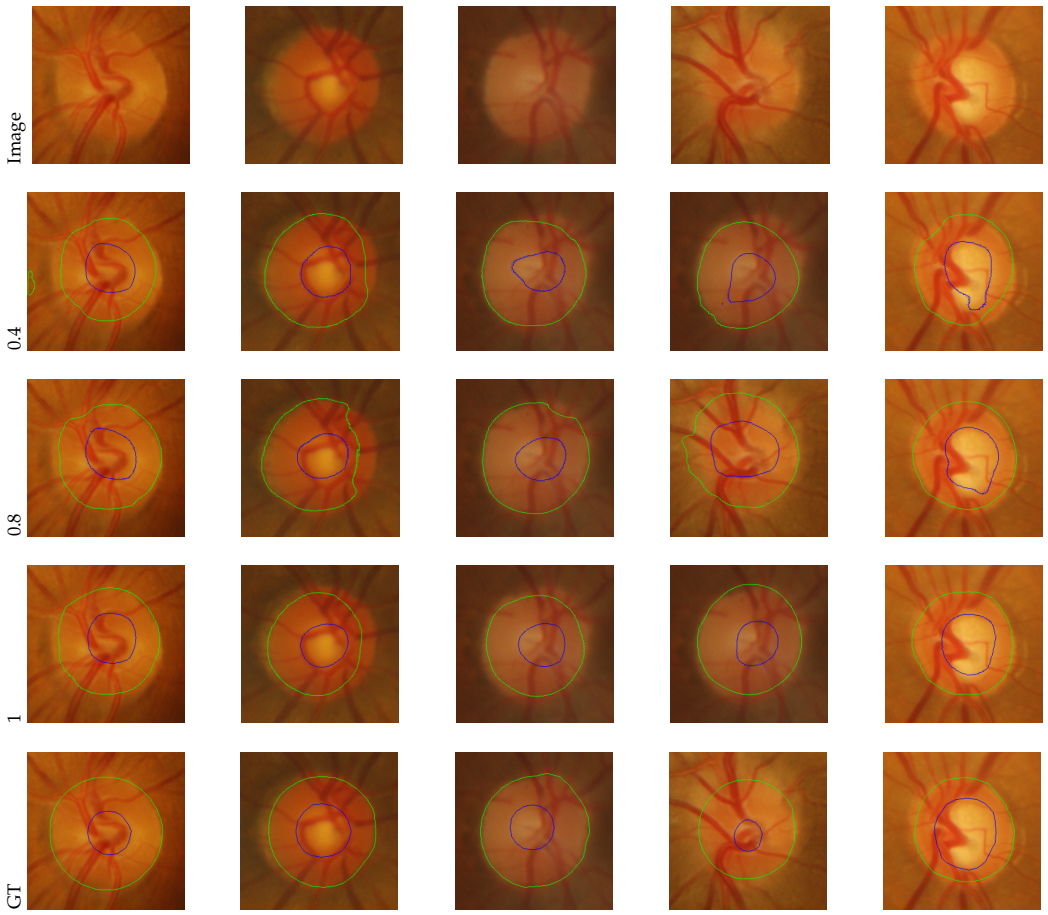


Figure 10. Optic cup–optical disc segmentation with different weights of the loss function.

4.5. Ablation Study on the Effect of the Proposed Components

To demonstrate the efficiency of the two components proposed in this paper, that is, the shape-constrained loss and the CTAM moduel, an ablation study was carried out. In this experiment, IOSUDA was the baseline model. Depending on whether or not each component was incorporated or not, there were four candidate models: (1) IOSUDA, (2) IOSUDA+ L_{shape}^{total}, (3) IOSUDA+CTAM, and (4) SCUDA. IOSUDA+L_{shape}^{total} denotes the addition of the shape-constrained loss function L_{shape}^{total} to the IOSUDA model, IOSUDA+CTAM indicates that the convolutional triple attention module CTAM was added to the IOSUDA model, and SCUDA indicates that the shape-constrained loss function L_{shape}^{total} and the convolutional triple attention module CTAM were both added to the IOSUDA model. These four models were evaluated on the RIM-ONE_r3 and Drishti-GS datasets. Table 5 reports the experimental results. To aid with more intuitive understanding, the results are plotted in the bar chart shown in Figure 11. It can be seen that, compared with IOSUDA, both IOSUDA+L_{shape}^{total} and IOSUDA+CTAM improve the IoU and Dice values of the optic cup and optical disc on the test dataset, which proves the effectiveness of the shape-constrained loss function and the CTAM module proposed in this paper. Specifically, taking $Dice_{OC}$ as an example, IOSUDA+L_{shape}^{total} and IOSUDA+CTAM show improvements of 1.33% and 0.34%, respectively, over the base IOSUDA model on the RIM-ONE_r3 dataset. As for IoU_{OD}, on the Drishti-GS dataset, IOSUDA+L_{shape}^{total} and IOSUDA+CTAM show improvements of 0.97% and 0.52%, respectively, over IOSUDA. Overall, the module result of IOSUDA+L_{shape}^{total} is better than the module of IOSUDA+CTAM, although the best outcome on both datasets is only reached when the two modules are combined, that is, in SCUDA. The effectiveness of the proposed components is therefore justified.

Table 5. Ablation study on the effect of the proposed components on the RIM-ONE_r3 and Drishti-GS datasets.

Datasets	Model	L_{shape}^{total}	CTAM	IoU_{OD} (%)	IoU_{OC} (%)	$Dice_{OD}$ (%)	$Dice_{OC}$ (%)
RIM-ONE_r3	IOSUDA [7]	×	×	83.06	59.63	90.14	72.32
	IOSUDA+L_{shape}^{total}	√	×	84.50	61.19	91.42	73.65
	IOSUDA+CTAM	×	√	83.60	60.21	90.45	72.66
	SCUDA (Ours)	√	√	**84.89**	**61.65**	**91.80**	**74.05**
Drishti-GS	IOSUDA [7]	×	×	89.08	64.91	93.77	77.49
	IOSUDA+L_{shape}^{total}	√	×	90.05	66.19	94.79	78.64
	IOSUDA+CTAM	×	√	89.60	65.23	94.25	77.87
	SCUDA (Ours)	√	√	**90.34**	**66.61**	**95.18**	**78.98**

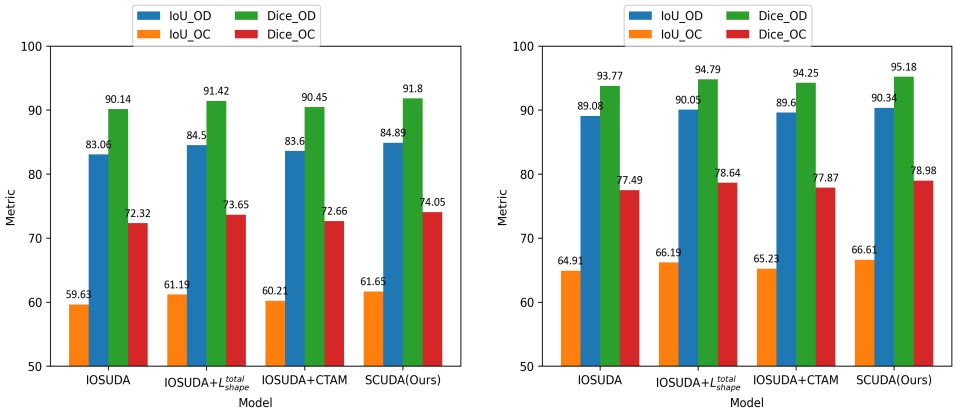


Figure 11. Illustration via bar chart of the effect of the proposed components evaluated on the RIM-ONE_r3 (**left**) and Drishti-GS (**right**) datasets.

5. Conclusions

In this paper, we propose an unsupervised domain adaptation with shape constraint for joint optic disc and cup segmentation, which we call SCUDA. A shape-constrained loss function is novelly proposed in this paper, which utilizes domain-invariant prior knowledge about the segmentation region of the optic cup–optical disc in fundus images to constrain the segmentation results during network training. Moreover, we design a convolutional triple attention module in the segmentation network that captures cross-dimensional interactions and provides rich feature representation in order to improve the segmentation performance of the network. Extensive experiments show that the proposed SCUDA framework outperforms state-of-the-art methods for segmentation of the optic cup and optical discs on both the RIM-ONE_r3 and Drishti-GS datasets.

Compared with existing method, we make the first attempt to use prior shape constraints to develop models for joint optic disc and cup segmentation, and use a cheaper yet more effective attention method to boost the performance of U-Net. It is worth noting that, in this work, the shape-constrained loss function is based on an approximate assumption, not a strictly correct one. Our future work will include investigating more realistic shape assumptions to construct constraints for training, along with a more effective and efficient attention mechanism for improving U-Net and novel frameworks of unsupervised domain adaptation for transfer learning.

Author Contributions: Conceptualization, J.D. and F.Z.; methodology, S.L. and F.Z.; software, F.Z.; validation, J.D., F.Z. and S.L.; formal analysis, F.Z.; writing—original draft preparation, F.Z.; writing—review and editing, S.L.; supervision, J.D. and S.L.; funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the National Natural Science Foundation of China (Grant no. 81660031) and the Guangxi Science and Technology Base and Talent Special Project (Grant no. Guike AD22035127).

Data Availability Statement: Not applicable.

Acknowledgments: Thanks are due to Yuehan Zhou from Guilin Medical University for funding assistance with the experiments and valuable discussions.

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

References

1. Tham, Y.C.; Li, X.; Wong, T.Y.; Quigley, H.A.; Aung, T.; Cheng, C.Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* **2014**, *121*, 2081–2090. [CrossRef] [PubMed]
2. Sevastopolsky, A. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognit. Image Anal.* **2017**, *27*, 618–624. [CrossRef]
3. Fu, H.; Cheng, J.; Xu, Y.; Wong, D.W.K.; Liu, J.; Cao, X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* **2018**, *37*, 1597–1605. [CrossRef]
4. Dou, Q.; Ouyang, C.; Chen, C.; Chen, H.; Heng, P.A. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. *arXiv* **2018**. arXiv:1804.10916.
5. Kamnitsas, K.; Baumgartner, C.; Ledig, C.; Newcombe, V.; Simpson, J.; Kane, A.; Menon, D.; Nori, A.; Criminisi, A.; Rueckert, D.; et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International Conference on Information Processing in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 597–609.
6. Tsai, Y.H.; Hung, W.C.; Schulter, S.; Sohn, K.; Yang, M.H.; Chandraker, M. Learning to adapt structured output space for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7472–7481.
7. Chen, C.; Wang, G. IOSUDA: An unsupervised domain adaptation with input and output space alignment for joint optic disc and cup segmentation. *Appl. Intell.* **2021**, *51*, 3880–3898. [CrossRef]
8. Yao, Y.; Liu, F.; Zhou, Z.; Wang, Y.; Shen, W.; Yuille, A.; Lu, Y. Unsupervised Domain Adaptation through Shape Modeling for Medical Image Segmentation. *arXiv* **2022**. arXiv:2207.02529.
9. 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: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
10. Zhang, Y.; Cai, X.; Zhang, Y.; Kang, H.; Ji, X.; Yuan, X. TAU: Transferable Attention U-Net for optic disc and cup segmentation. *Knowl.-Based Syst.* **2021**, *213*, 106668. [CrossRef]
11. Zhao, X.; Wang, S.; Zhao, J.; Wei, H.; Xiao, M.; Ta, N. Application of an attention u-net incorporating transfer learning for optic disc and cup segmentation. *Signal Image Video Process.* **2021**, *15*, 913–921. [CrossRef]
12. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
13. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
14. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**. arXiv:1807.06514.
15. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3139–3148.
16. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
17. Zhang, J.; Li, W.; Ogunbona, P.; Xu, D. Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–38. [CrossRef]
18. Bousmalis, K.; Trigeorgis, G.; Silberman, N.; Krishnan, D.; Erhan, D. Domain separation networks. *arXiv* **2016**, arXiv:1608.06019.
19. French, G.; Mackiewicz, M.; Fisher, M. Self-ensembling for visual domain adaptation. *arXiv* **2017**, arXiv:1706.05208.
20. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3722–3731.
21. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, Ju.; Isola, P.; Saenko, K.; Efros, A.A.; Darrell, T. Cycada: Cycle-Consistent Adversarial Domain Adaptation. 2017. Available online: https://proceedings.mlr.press/v80/hoffman18a.html (accessed on 8 Novermber 2022).
22. Long, M.; Wang, J. Learning Transferable Features with Deep Adaptation Networks. 2015. Available online: https://proceedings.mlr.press/v37/long15 (accessed on 8 November 2022).
23. Chen, C.; Dou, Q.; Chen, H.; Heng, P.A. Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 143–151.

24. Huo, Y.; Xu, Z.; Moon, H.; Bao, S.; Assad, A.; Moyo, T.K.; Savona, M.R.; Abramson, R.G.; Landman, B.A. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE Trans. Med. Imaging* **2018**, *38*, 1016–1025. [CrossRef]
25. Song, L.; Wang, C.; Zhang, L.; Du, B.; Zhang, Q.; Huang, C.; Wang, X. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognit.* **2020**, *102*, 107173. [CrossRef]
26. Chen, C.; Dou, Q.; Chen, H.; Qin, J.; Heng, P.A. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 2494–2505. [CrossRef]
27. Cheng, J.; Liu, J.; Xu, Y.; Yin, F.; Wong, D.W.K.; Tan, N.M.; Tao, D.; Cheng, C.Y.; Aung, T.; Wong, T.Y. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans. Med. Imaging* **2013**, *32*, 1019–1032. [CrossRef]
28. Joshi, G.D.; Sivaswamy, J.; Krishnadas, S.R. Optic Disk and Cup Segmentation From Monocular Color Retinal Images for Glaucoma Assessment. *IEEE Trans. Med. Imaging* **2011**, *30*, 1192–1205. [CrossRef]
29. Zheng, Y.; Stambolian, D.; O'Brien, J.; Gee, J.C. Optic Disc and Cup Segmentation from Color Fundus Photograph Using Graph Cut with Priors. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*; Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 75–82.
30. Lupascu, C.A.; Tegolo, D.; Rosa, L.D. Automated detection of optic disc location in retinal images. In Proceedings of the 2008 21st IEEE International Symposium on Computer-Based Medical Systems, Jyvaskyla, Finland, 17–19 June 2008; pp. 17–22.
31. Youssif, A.A.H.A.R.; Ghalwash, A.Z.; Ghoneim, A.A.S.A.R. Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *IEEE Trans. Med. Imaging* **2007**, *27*, 11–18. [CrossRef]
32. Zilly, J.G.; Buhmann, J.M.; Mahapatra, D. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 136–143.
33. Fu, H.; Cheng, J.; Xu, Y.; Zhang, C.; Wong, D.W.K.; Liu, J.; Cao, X. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans. Med. Imaging* **2018**, *37*, 2493–2501. [CrossRef]
34. Deng, J.; Zhang, F.; Li, S.; Bao, J. Towards Semi-Supervised Segmentation of Retinal Fundus Images via Self-Training. In Proceedings of the 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 22–24 July 2022; pp. 167–172. [CrossRef]
35. Liu, S.; Hong, J.; Lu, X.; Jia, X.; Lin, Z.; Zhou, Y.; Liu, Y.; Zhang, H. Joint optic disc and cup segmentation using semi-supervised conditional GANs. *Comput. Biol. Med.* **2019**, *115*, 103485. [CrossRef]
36. Liu, P.; Kong, B.; Li, Z.; Zhang, S.; Fang, R. CFEA: Collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 521–529.
37. Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.W.; Heng, P.A. Boundary and entropy-driven adversarial learning for fundus image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 102–110.
38. Chen, Y.; Kalantidis, Y.; Li, J.; Yan, S.; Feng, J. A^2-nets: Double attention networks. *arXiv* **2018**, arXiv:1810.11579
39. Gao, Z.; Xie, J.; Wang, Q.; Li, P. Global second-order pooling convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3024–3033.
40. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 27–28 October 2019, Seoul, Korea.
41. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 27–28 October 2019, Seoul, Korea; pp. 603–612.
42. Hou, Q.; Zhang, L.; Cheng, M.M.; Feng, J. Strip pooling: Rethinking spatial pooling for scene parsing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4003–4012.
43. Xiao, T.; Xu, Y.; Yang, K.; Zhang, J.; Peng, Y.; Zhang, Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 842–850.
44. Tong, X.; Wei, J.; Sun, B.; Su, S.; Zuo, Z.; Wu, P. ASCU-Net: Attention Gate, Spatial and Channel Attention U-Net for Skin Lesion Segmentation. *Diagnostics* **2021**, *11*, 501. [CrossRef] [PubMed]
45. Li, C.; Tan, Y.; Chen, W.; Luo, X.; He, Y.; Gao, Y.; Li, F. ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. *Comput. Graph.* **2020**, *90*, 11–20. [CrossRef]
46. Zhang, J.; Lv, X.; Zhang, H.; Liu, B. AResU-Net: Attention Residual U-Net for Brain Tumor Segmentation. *Symmetry* **2020**, *12*, 721. [CrossRef]
47. Petit, O.; Thome, N.; Rambour, C.; Themyr, L.; Collins, T.; Soler, L. U-net transformer: Self and cross attention for medical image segmentation. In *Machine Learning in Medical Imaging*; Lian, C., Cao, X., Rekik, I., Xu, X., Yan, P., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 267–276.
48. Bhatkalkar, B.J.; Reddy, D.R.; Prabhu, S.; Bhandary, S.V. Improving the Performance of Convolutional Neural Network for the Segmentation of Optic Disc in Fundus Images Using Attention Gates and Conditional Random Fields. *IEEE Access* **2020**, *8*, 29299–29310. [CrossRef]
49. Fumero, F.; Alayon, S.; Sanchez, J.L.; Sigut, J.; Gonzalez-Hernandez, M. RIM-ONE: An open retinal image database for optic nerve evaluation. In Proceedings of the 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), Bristol, UK, 27–30 June 2011; pp. 1–6. [CrossRef]

50. Orlando, J.I.; Fu, H.; Breda, J.B.; van Keer, K.; Bathula, D.R.; Diaz-Pinto, A.; Fang, R.; Heng, P.A.; Kim, J.; Lee, J.; et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image Anal.* **2020**, *59*, 101570. [CrossRef]
51. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
52. Singh, V.K.; Abdel-Nasser, M.; Rashwan, H.A.; Akram, F.; Pandey, N.; Lalande, A.; Presles, B.; Romani, S.; Puig, D. FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access* **2019**, *7*, 130552–130565. [CrossRef]
53. Singh, V.K.; Rashwan, H.A.; Akram, F.; Pandey, N.; Sarker, M. .M.K.; Saleh, A.; Abdulwahab, S.; Maaroof, N.; Romani, S.; Puig, D. Retinal optic disc segmentation using conditional generative adversarial network. *arXiv* **2018**, arXiv:1806.03905.