



Improving foveal avascular zone segmentation in fluorescein angiograms by leveraging manual vessel labels from public color fundus pictures

DOMINIK HOFER,¹ URSULA SCHMIDT-ERFURTH,¹ JOSÉ IGNACIO ORLANDO,^{1,2} FELIX GOLDBACH,¹ BIANCA S. GERENDAS,¹ AND PHILIPP SEEBOECK^{1,*}

¹Vienna Reading Center, Department of Ophthalmology and Optometry, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria

²Yatiris Group, PLADEMA Institute, CON-ICET, Universidad Nacional del Centro de la Provincia de Buenos Aires, Gral. Pinto 399, Tandil, Buenos Aires, Argentina

*philipp.seeboeck@meduniwien.ac.at

Abstract: In clinical routine, ophthalmologists frequently analyze the shape and size of the foveal avascular zone (FAZ) to detect and monitor retinal diseases. In order to extract those parameters, the contours of the FAZ need to be segmented, which is normally achieved by analyzing the retinal vasculature (RV) around the macula in fluorescein angiograms (FA). Computer-aided segmentation methods based on deep learning (DL) can automate this task. However, current approaches for segmenting the FAZ are often tailored to a specific dataset or require manual initialization. Furthermore, they do not take the variability and challenges of clinical FA into account, which are often of low quality and difficult to analyze. In this paper we propose a DL-based framework to automatically segment the FAZ in challenging FA scans from clinical routine. Our approach mimics the workflow of retinal experts by using additional RV labels as a guidance during training. Hence, our model is able to produce RV segmentations simultaneously. We minimize the annotation work by using a multi-modal approach that leverages already available public datasets of color fundus pictures (CFPs) and their respective manual RV labels. Our experimental evaluation on two datasets with FA from 1) clinical routine and 2) large multicenter clinical trials shows that the addition of weak RV labels as a guidance during training improves the FAZ segmentation significantly with respect to using only manual FAZ annotations.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Ophthalmologists routinely use multiple imaging modalities to analyze retinal biomarkers for diagnosing and monitoring diseases and therapeutic planning. Among the most commonly used en-face modalities are Color Fundus Photography (CFP) and Fluorescein Angiography (FA) [1,2].

CFP is a non-invasive photography of the so-called fundus of the eye, referring to the retinal surface. Normally, clinicians use CFPs for analyzing the Retinal Vasculature (RV) and the optic disc, or to assess the presence of lesions. For example, macular degeneration with drusen and signs of choroidal neovascularization in neovascular age-related macular degeneration (AMD) or hemorrhages, cotton wool spots or hard exudates in Diabetic Retinopathy (DR) can be viewed with this modality. Vascular changes are also investigated in case of ischemic diseases like Retinal Vein Occlusion (RVO), in particular distribution, connectivity and density changes of the RV [3].

These last mentioned changes can be best visualized with FA, which is an invasive modality that uses a fluorescent dye to increase the contrast of the vascular structure and visualize leakage from leaky pathological vessels. After injecting the dye intravenously, a series of images are

acquired while the dye distributes in the body's systemic bloodstream including the eye, allowing to see the retinal circulation in different phases and to discover pathological changes or vessel occlusions. As opposed to CFP, this technique also enables to observe the so-called Foveal Avascular Zone (FAZ), a vessel free area in the center of the macula that is responsible for sharp vision. The shape, size and perimeter of this region are known to be relevant biomarkers related to DR and RVO for diagnosis and disease monitoring, as the FAZ can be abnormally large in presence of these diseases [4–6].

Segmenting the FAZ is a key step to quantify these biomarkers in clinical routine. This process is naturally linked to the retinal vacuature, as the manual delineation of the vessel-free FAZ typically involves identifying the edges of the vessels around the macula.

The task of manual FAZ segmentation, however, is tedious, time consuming and subject to high intra- and inter-observer variability [7]. On the one hand, this can be explained by the varying quality and heterogeneity of data (see Fig. 1). In clinical routine, the images are often of poor quality in general, having low contrast or brightness, are blurry, contain noise or other image artifacts. Also the field of view can vary between or even within patients. On the other hand, the segmentation can be also difficult in FA scans with sufficient quality, as the FAZ can vary significantly in shape and size due to natural anatomical variations or pathological changes. For example, diseases such as DR can notably affect the identification of its contour [5].

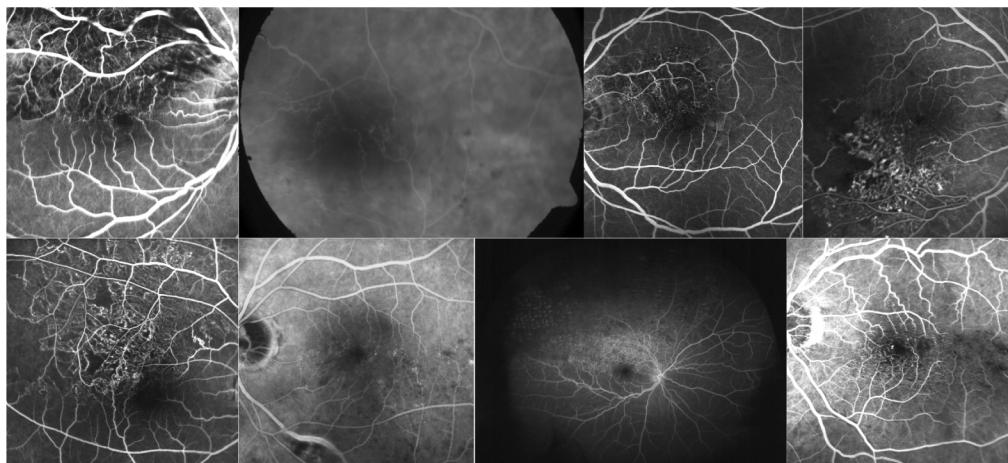


Fig. 1. FA scans showing the heterogeneity of clinical data. From top left to bottom right: A scan with (1) light motion blur and high contrast, (2) low contrast with almost no RV visible and black borders, (3) sufficient quality with the FAZ clearly visible, (4) a small FAZ and leaky RV, (5) non-centered fovea, (6) a scan with low contrast, (7) a huge FOV and (8) oversaturated brightness and contrast.

Computer-aided systems based on image processing techniques and deep learning (DL) became popular in recent years to automate segmentation tasks [8]. Due to their low cost and time-effectiveness, these approaches allow to process large amounts of data in a very efficient way, which can have significant impact on clinical practice and research. However, translating such models to real world scenarios is usually challenging, especially in the case of clinical FA due to the aforementioned heterogeneity of the data. Current approaches for FAZ segmentation in FA are tailored towards well curated datasets not reflecting the variability usually observed in real-world clinical data [9]. Furthermore, they usually rely on manual labels during application-time and cannot perform both FAZ and RV segmentation simultaneously [10–14], therefore not explicitly taking the relationship between these two anatomical regions into account. The latter is partially due to the lack of FA databases with manually annotated vessels. On the contrary, several

public CFP datasets with manual RV annotations are available, a setting that has favored the development of RV segmentation methods for this imaging modality [15].

In this paper, we propose a DL-based model that accurately segments the FAZ in clinical routine FA by leveraging the RV as a guidance and exploiting the aforementioned anatomical relation between the FAZ and the RV during the learning process (Fig. 2). In addition, the final model also provides vessel segmentations in FA images. Instead of requiring laborious manual RV annotations on FA scans, the proposed method utilizes public datasets of CFPs which already come with manual RV labels. We trained and evaluated our method using a diverse dataset of FA images from large clinical multicenter trials, with RVO as the main pathology. Moreover, we performed an additional evaluation on a challenging external FA dataset retrospectively collected from clinical records, containing various pathologies such as diabetic macular edema (DME), DR, AMD and RVO. Experimental results on both datasets show that the proposed model outperforms the baseline approaches, with significant improvements observed in the challenging external FA dataset.

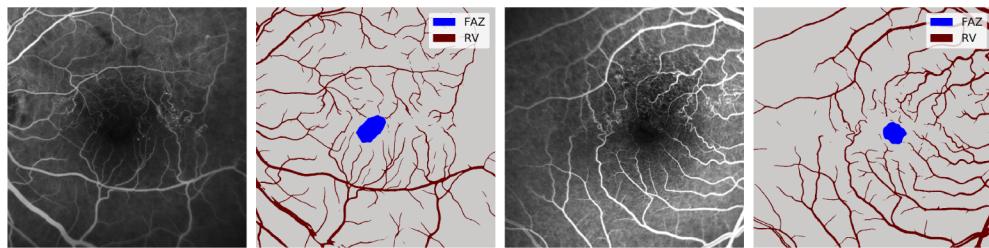


Fig. 2. FA scans with their corresponding FAZ and RV segmentations obtained by the proposed approach.

1.1. Related work

Although several methods have been introduced for FAZ segmentation in FA [9–14,16], the number of publications focusing on RV segmentation in this modality is scarce. On the contrary, this topic has been extensively researched primarily in CFP [15,17,18], in part due to the abundance of publicly available datasets with manual RV annotations.

The first approaches for FAZ segmentation in FA were mostly based on image processing techniques and statistical learning algorithms such as Bayesian methods [10,16] or Markov Random Fields models [12]. Semi-automatic approaches like region-growing [14], active contours [11] and level-sets [13] were also explored, as they mimic the process of delineating the region boundaries. However, they heavily rely on manual initialization, and their results might vary depending on the initial selection. On the other hand, all these early algorithms require careful hyperparameter tuning, which are usually tailored to specific datasets. As a result, they suffer from poor generalization to other sets if not re-calibrated, which takes time, computational resources and sometimes even require extra manual annotations. DL methods can alleviate these drawbacks by learning task-relevant features from data. However, they have not been used for FAZ segmentation in FA but in other imaging modalities such as OCT-A [19]. To the best of our knowledge, our previous multitask model [9] is the only DL approach published for FAZ segmentation in FA.

RV are observed in FA as a large, connected hyperfluorescent structure with high contrast to the surrounding retinal tissue. This property allows to identify the FAZ as a dark area centered in and around the fovea (Fig. 2). Based on this concept, Simó et. al [16] introduced a Bayesian model that simultaneously segments arteries, veins and the FAZ in FA, as three separate classes. However, they experimentally observed that integrating the vessels in the prediction

task compromised the FAZ segmentation performance compared to results from other existing approaches. A similar behavior was observed in [19] but in OCT-A.

Despite the fact that FA allows to easily assess the integrity and distribution of the RV because of its high contrast, CFP is more frequently used for this purpose. This is mainly due to the invasive characteristics of FA, which may lead like any intravenously applied medication to very rare allergic reactions and is thus not done on a regular basis repeatedly [1]. On the contrary, CFP shows the RV with less contrast, but it is non-invasive and easier to obtain. As a result, most of the public datasets with manual RV annotations and the majority of the RV segmentation models are based on it. Several approaches have been proposed to segment the RV in CFP, with the most recent DL-based methods showing outstanding performance [15]. In [17,18], for instance, a U-Net model [20] trained with an adversarial loss is able to produce cutting-edge results superior to the state-of-the-art. On the other hand, the literature in RV segmentation for FA is quite limited and mostly focused on simple image processing techniques, which allow to extract pixel-based features that are subsequently used for vessel/not vessel classification [21–26]. As previously stated, these techniques suffer from poor generalization ability and require an intensive labor to work on other datasets.

The vast majority of RV segmentation models are usually trained and evaluated using publicly available CFP datasets with manual RV labels, which are extensive and easy to access [27–32]. Leveraging labels produced in one image modality to segment targets in another one allows to not only benefit from existing datasets without requiring to manually create new annotations, but also to exploit the advantages of each modality in particular (e.g., the FAZ can only be segmented manually in FA or OCT-A). In general, this requires either an accurate multimodal registration algorithm or the usage of domain adaptation techniques. Registration based approaches [33–37] are in general applied when pairs of images from both modalities are available for each patient. Some of these approaches use paired, registered images so that labels from one modality can be used for the other. This has the disadvantage that often only limited data is available. Ding et. al. propose several methods that use registration between CFP and FA [33,34] and weak labels [35]. Hervella et. al. [36] also propose a registration-based transfer-learning approach between FA and CFP for RV segmentation in CFP. A similar approach is proposed by Noh et. al. [37], where they use weak labels from FA for training a segmentation in CFP. In contrast, domain adaptation methods translate images from one source to another, and can be applied even in unpaired sets for example by using CycleGANs. These transformed images can then be applied jointly with the source annotations to train a new model to automatically produce similar annotations but in the destination modality. Building on this idea, research has been conducted on how to use public RV annotations from CFP as weak labels for segmentation in FA or other modalities and vice versa, or to build models that can generally segment RV in both modalities [33–41]. Ju et. al. [38] propose a domain adaptation strategy using a CycleGAN model to bridge the gap between unpaired UWCF CFP and regular CFP. Rodrigues et. al. [39] describe a multi-modal RV segmentation approach that works in CFP, FA and Scanning Laser Ophthalmology (SLO). Similarly, Zhao et. al. [40] use a combination of image-processing methods and active contour segmentation to segment RV in CFP and FA. Brea et. al. [41] train a model on CFP and then apply and evaluate the result on SLO pictures and vice versa.

Research has also been conducted about modality conversion between CFP and FA. Several approaches [42–44] use a CycleGAN-based approach to synthesize FA from CFP. In contrast, Li et. al. [45] first structure the domain features into domain-shared and domain-independent features. Other methods [46,47] rely on paired images for the conversion. Hervella et. al [48] experimentally compare both approaches: generally, image translation learned from paired data seems to work better because it is easier to learn the mapping from one pixel to the other. However, paired images are not always available and therefore this approach is often not feasible.

Automatically segmenting retinal structures in clinical routine FA remains challenging due to datasets having variations in image resolution, pixel size, contrast, field of view and image artifacts that need to be covered by the segmentation model. Furthermore, pathologies can alter the size and shape of the retinal structures significantly, resulting in an even more difficult segmentation task.

1.2. Contributions

In this paper, we propose a DL-based model that accurately segments the FAZ in challenging FA data by using the RV as guidance. Hence, our model simultaneously produces segmentations of the vessels. To the best of our knowledge, this is the first approach to cope with this task by exploiting CFP based manual annotations. The main contributions of this work are fourfold:

- Performance improvement for FAZ segmentation in FA by using RV as additional guidance, mimicking the workflow of domain experts.
- Eliminating the need of obtaining manual RV annotations in FA, by exploiting public datasets from another modality (CFP) which already come with manual annotations.
- Simultaneous segmentation of FAZ and RV.
- Additional evaluation on a challenging external clinical routine FA dataset, containing the most common retinal diseases (AMD, DR, RVO, DME), various pathologies and diverse image quality.

2. Methods

An overview of the proposed framework is provided in Fig. 3. First, a CycleGAN-based model $F_{\text{CFP} \leftrightarrow \text{FA}}$ is trained on unpaired CFP and FA images to learn a conversion between these two modalities, omitting the need of paired training samples. The $F_{\text{CFP} \leftrightarrow \text{FA}}$ generator is then applied to convert public CFP into pseudo-FA (see Fig. 3, point 1). Second, an intermediate U-Net-based model $F_{\text{RV-SEG}}$ is trained on the pseudo-FA images to segment RV, using the manual annotations provided by the CFP datasets. Then, weak RV labels are obtained by applying this trained model on the real FA data (see Fig. 3 point 2). Third, the manual FAZ labels are combined with the weak RV labels as multi-class labels. Finally, a U-Net $G_{\text{FAZ\&RV-SEG}}$ is trained on the clinical FA data with the manual FAZ and weak RV labels as targets (see Fig. 3, point 3).

2.1. Domain adaptation for transforming CFP into pseudo-FA

We utilize a CycleGAN approach [49] to convert the CFPs X_{CFP} from dataset $S_{\text{PUBLIC-CFP}}$ into pseudo-FA. The CycleGAN consists of two GANs, each of them featuring a generator G and a discriminator D. Here, the first GAN $\text{GAN}_{\text{CFP} \rightarrow \text{FA}}$ converts CFP into pseudo-FA, while the second GAN $\text{GAN}_{\text{FA} \rightarrow \text{CFP}}$ converts pseudo-FA back into CFP.

Let $S_{\text{CFP} \leftrightarrow \text{FA}} = \{X_{\text{CFP}}, X_{\text{FA}}\}$ be the training data, with $X_{\text{CFP}} \in \mathbb{R}^{a \times b}$ being CFP images with a resolution of $a \times b$ pixels and $X_{\text{FA}} \in \mathbb{R}^{a \times b}$ being FA images. After training, the generator $G_{\text{CFP} \rightarrow \text{FA}}$ of $\text{GAN}_{\text{CFP} \rightarrow \text{FA}}$ has learned the mapping $G_{\text{CFP} \rightarrow \text{FA}} : X_{\text{CFP}} \rightarrow X_{\text{FA}}$.

The two generators are U-Nets with four encoding/decoding levels. Each encoding level consists of a convolution block with a 3×3 convolution with a stride of 2, followed by a Rectified Linear Unit (ReLU) activation function and instance normalization. The decoding part is similar to the convolution blocks in the encoding level, with each deconvolution block additionally having a skip connection and an upsampling layer with a stride of 2 after the convolution block. Both discriminators consist of four convolution-blocks with the same structures as above. The model is trained with the mean squared error loss function.

The final generator model $G_{\text{CFP} \rightarrow \text{FA}}$ is applied on dataset $S_{\text{PUBLIC-CFP}}$ to convert the public CFP images to pseudo-FA: $G_{\text{CFP} \rightarrow \text{FA}}(X_{\text{CFP}})$.

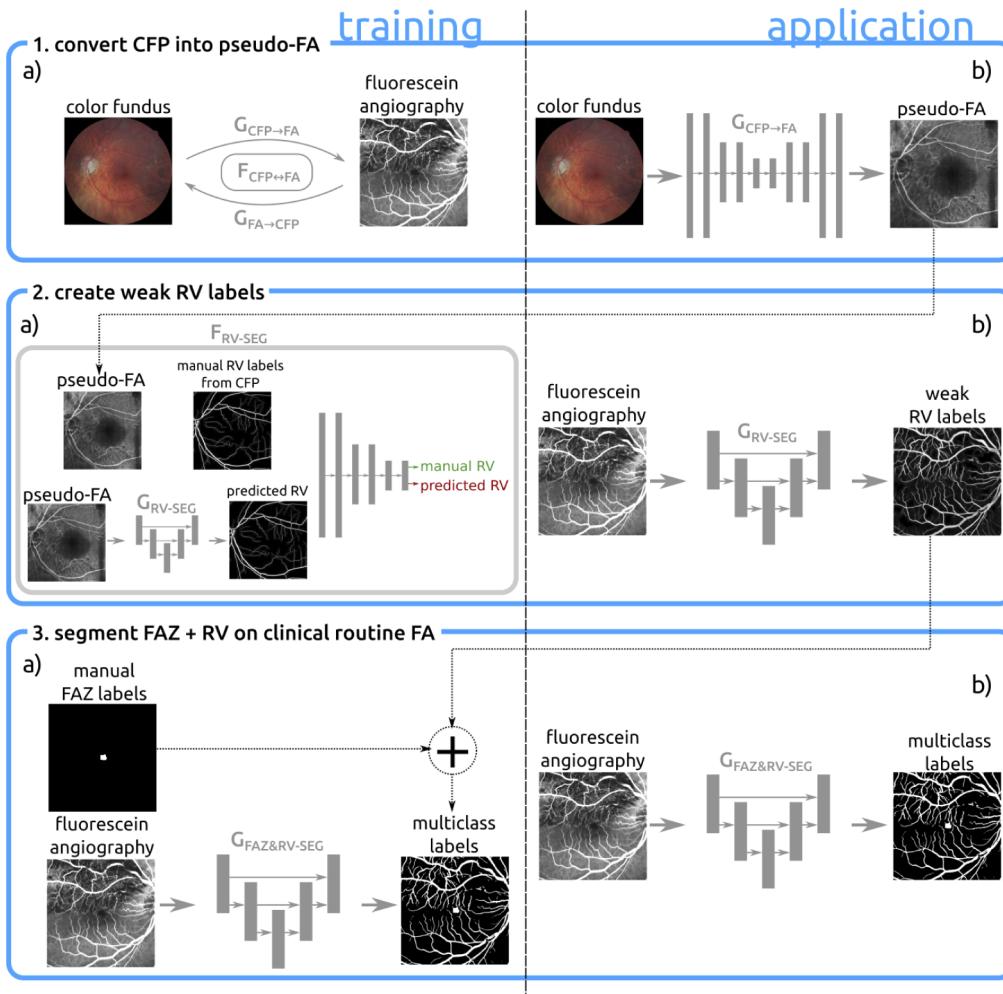


Fig. 3. Overview of the framework. 1.a) The CycleGAN learns the matching between CFP and FA 1.b) The CFP images are converted into pseudo-FA. 2.a.) F_{RV-SEG} is trained on the pseudo-FA and respective RV labels. 2.b.) G_{RV-SEG} is applied on dataset $S_{PUBLIC-CFP}$ to obtain the weak labels. 3.a) The manual FAZ and weakly-supervised RV labels are combined into two-class labels. 3.b.) $G_{FAZ\&RV-SEG}$ is trained on dataset $S_{REAL-FA}$ with the two class labels.

2.2. Creating weak RV labels for FA images

We train an intermediate model $G_{\text{RV-SEG}}$ to segment RV on $G_{\text{CFP} \rightarrow \text{FA}}(X_{\text{CFP}})$. The model is based on the work of Son et. al. [17], which is a U-Net trained with an adversarial loss. Hence, our model $G_{\text{RV-SEG}}$ utilizes a discriminator trying to distinguish between real and generated segmentations.

Let $S_{\text{RV-SEG}} = \{G_{\text{CFP} \rightarrow \text{FA}}(X_{\text{CFP}}), Y_{\text{MANUAL-RV}}\}$ be the training data, with $Y_{\text{MANUAL-RV}} \in \mathcal{Y}^{a \times b}$ being the manual RV labels, originally annotated on X_{CFP} , with $\mathcal{Y} = \{0, 1\}$ being the possible classes for background and RV. The segmentation model has then trained to learn the mapping: $G_{\text{RV-SEG}} : G_{\text{CFP} \rightarrow \text{FA}}(X_{\text{CFP}}) \rightarrow Y_{\text{MANUAL-RV}}$.

Our U-Net based model $G_{\text{RV-SEG}}$ has five encoding/decoding levels, each level consisting of a convolution block with two 3×3 convolutions, each followed by a ReLU function and a batch normalization layer, concluding with a 2×2 max-pooling layer. The decoder is connected to the encoder via skip connections, using convolutional blocks of similar structure with additional upsampling layers. The discriminator used for the adversarial loss consists of five convolution blocks with the same layers as in the encoder of the segmentation network. The U-Net is trained with the cross entropy loss function. While the model is initially trained on pseudo-FA images, it is applied on the real FA scans from the target dataset $S_{\text{REAL-FA}}$ to obtain RV annotations that can be used as weak RV labels: $G_{\text{RV-SEG}}(X_{\text{REAL-FA}}) = Y_{\text{WEAK-RV}}$.

2.3. Simultaneous FAZ and RV segmentation using weak RV labels

We train $G_{\text{FAZ\&RV-SEG}}$ on clinical FA with the weak RV and manual FAZ labels as targets. Let $S_{\text{REAL-FA}} = \{X_{\text{REAL-FA}}, Y_{[\text{WEAK-RV}, \text{MANUAL-FAZ}]}\}$ be the training dataset for this model, with $X_{\text{REAL-FA}} \in \mathbb{R}^{a \times b}$ being FA images from large multicenter clinical trials. The target labels $Y_{[\text{WEAK-RV}, \text{MANUAL-FAZ}]} \in \mathcal{Y}^{a \times b}$ are obtained by combining the weak RV labels $Y_{\text{WEAK-RV}}$ with the manual FAZ labels $Y_{\text{MANUAL-FAZ}}$, with $\mathcal{Z} = \{0, 1, 2\}$ being the classes for background, RV and FAZ. To remove the borders that some FA images in the dataset have, the FA images $X_{\text{REAL-FA}}$ are cropped before training by calculating the bounding box of the weak RV labels $Y_{\text{WEAK-RV}}$. The bounding box crops the image according to the lower, upper, left and right bounds of the labels, resulting in an image without any black borders. $G_{\text{FAZ\&RV-SEG}}$ is a multi-class segmentation network based on a U-Net with six encoding/decoding levels. Each encoding level has a convolution block with two 3×3 -convolutions, each followed by an instance normalization layer, ReLU function, a dropout layer and a 2×2 max-pooling layer. The decoder consists of deconvolution blocks, with the addition of upsampling filters with a size of 2×2 and skip connections between encoder and decoder. The model is trained with the cross entropy loss function.

3. Experimental setup

We evaluated our method in the following terms: (1) quantitative and qualitative segmentation accuracy for FAZ segmentation in datasets $S_{\text{REAL-FA}}$ and S_{EVAL} for our proposed model and the baseline, (2) qualitative RV segmentation results along with AUC and Dice metrics for a subset of manually annotated images of $S_{\text{REAL-FA}}$ and (3) the Frechet Inception Distance (FID) and qualitative evaluation for the conversion from CFP to FA in dataset $S_{\text{PUBLIC-CFP}}$. Besides using the U-Net (Section 2.3 for $G_{\text{FAZ\&RV-SEG}}$), we performed an ablation experiment utilizing a corresponding U-Net++ [50,51] as architectural backbone. Moreover, we also compared the proposed approach using the weak RV labels as additional guidance to the baseline-strategy training solely on the binary FAZ annotations. For evaluation purposes, a grader additionally separated the $S_{\text{REAL-FA}}$ testset into regular and irregular FAZ shapes. The images have been split into 80% training and 20% validation data for datasets $S_{\text{PUBLIC-CFP}}$ and $S_{\text{CFP} \leftrightarrow \text{FA}}$. For $S_{\text{REAL-FA}}$

the data has been split into 60% training, 20% validation and 20% test data. S_{EVAL} is an external testset that is not associated with the training data and other datasets at all.

3.1. Materials

Table 1 summarizes the datasets used in our experiments. The (1) support dataset $S_{CFP \leftrightarrow FA}$ consists of 355 CFPs and FA images that are used for training the modality conversion, (2) support dataset $S_{PUBLIC-CFP}$ consists of 123 CFPs and their respective manual RV annotations from public datasets and (3) target dataset $S_{REAL-FA}$ consists of FA images from large multicenter clinical trials and their respective manual FAZ labels. For additional evaluation, we use (4) another dataset S_{EVAL} with 185 FA images from clinical routine.

Table 1. Overview of the datasets used for training and evaluating our proposed framework.

Dataset	Modality	Images	Annotations	Description	Pathology
$S_{REAL-FA}$	FA	494	FAZ	Large randomized multicenter clinical trials	RVO
$S_{CFP \leftrightarrow FA}$	CFP & FA	355	-	Modality conversion	DR, AMD, GL, hypertension, RVO
$S_{PUBLIC-CFP}$	CFP	123	RV	Public datasets	DR, AMD, GL, HT
S_{EVAL}	FA	185	FAZ	FA from clinical routine for evaluation	Miscellaneous

3.1.1. FA from large multicenter clinical trials ($S_{REAL-FA}$)

The target dataset $S_{REAL-FA}$ consists of 494 FA images and its respective manual FAZ segmentations. The images are taken from the Vienna Reading Center imaging database from RVO randomized clinical trials, which examined patients suffering from RVO. Images were acquired with devices from seven different vendors, with a resolution ranging from 4288×2848 to 512×512 pixels. The images vary significantly in terms of field of view, pixel size, contrast, RV and FAZ visibility, viewing angle, imaging artifacts and motion blur. The FAZ labels were manually annotated by an experienced grader (trained and certified by the Vienna Reading Center) and used as ground-truth. Furthermore, RV annotations in 18 images in the testset were produced by a grader.

3.1.2. Public CFP datasets ($S_{PUBLIC-CFP}$)

The dataset $S_{PUBLIC-CFP}$ consists of public datasets of CFP images and their respective RV annotations [27–32]. CHASE_DB1 [27] consists of 28 CFPs with a resolution of 1280×960 pixels. As with all other datasets described in this section, each image has manual RV labels. DR HAGIS [28] was originally extracted from a DR screening program and consists of 39 high-resolution CFP images. The image resolution ranges from 2816×1880 to 4752×3168 pixels. The dataset has images that show signs from four different diseases: glaucoma (GL), hypertension, DR and AMD. DRIVE images [29] were also collected from a DR screening program and consists of 40 CFP images with a pixel size of 768×584 . 33 of the images show healthy eyes and 7 images show signs of mild early DR. HRF [30] consists of 45 CFP images, of which 15 show healthy eyes, 15 show DR and 15 show GL. The images have a resolution of 3504×2336 pixels. LES-AV [31] consists of 22 images with resolutions of 1444×1620 and 1958×2196 pixels from GL and healthy subjects. STARE [32] consists of 20 CFP images with manual RV annotations, of which 10 show healthy eyes and 10 show eyes with pathologies that obscure the RVs, with a resolution of 605×700 pixels.

3.1.3. CFP and FA for modality conversion ($S_{CFP \leftrightarrow FA}$)

This dataset consists of 123 CFP images from $S_{PUBLIC-CFP}$ and 204 FA images without borders from the training-set of $S_{REAL-FA}$.

3.1.4. External FA testset from clinical routine for additional evaluation (S_{EVAL})

The FA images of dataset S_{EVAL} come from clinical routine. Similar to the images from the target dataset, the image properties differ widely. The image dimensions range from 768×768 to 1536×1536 pixels and a grader annotated the FAZ on all images. As the data comes directly from patients that visit the clinic, the images cover a large range of retinal pathologies but can also show healthy retinas.

3.2. Pre-processing

To accelerate the training process, the CFPs of $S_{CFP \leftrightarrow FA}$ were converted to grayscale and cropped to the same field of view as the FA images. Then, both CFP and FA were resized to an image size of 256×256 pixels due to memory limitations. The same steps were applied during test time to convert the CFPs of $S_{PUBLIC-CFP}$ to pseudo-FA. The images of $S_{REAL-FA}$ were cropped to the bounding box of the weak RV labels and resized to a resolution of 800×800 pixels.

3.3. Evaluation metrics

The performance for the FAZ and RV segmentations was quantitatively evaluated using standard metrics such as the Dice index, the Area Under the precision-recall-Curve (AUC), precision and recall. A Wilcoxon signed-rank test with a significance level of $\alpha = 0.05$ was used to study the statistical significance of the differences in performance between our proposed model and the baseline. The AUC was computed from score maps. Dice, precision and recall were calculated from binary maps using a threshold of 0.5 for baseline, while for our model each pixel gets assigned the class that has the highest probability.

3.4. Training details

We trained $F_{CFP \leftrightarrow FA}$ for 800 epochs using Glorot uniform initialization, Adam optimization, and learning rates of 2×10^{-3} and 2×10^{-4} for the discriminator and GAN, respectively, with a batch-size of 1. F_{RV-SEG} was trained for 100 epochs, using Glorot uniform initialization, Adam optimization and a learning rate of 2×10^{-2} , with a batch-size of 1. The images were randomly flipped vertically and rotated by up to 25 degrees during training. The model with the highest Dice on the validation set was selected as the final model. $G_{FAZ\&RV-SEG}$ was trained for 600 epochs with He normal initialization and Adam optimization, and a learning rate of 10^{-4} . During training, data augmentation in the form of random shifting up to the maximum of 6% of the image size, rotation up to 25 degrees, zooming up to 40% of the image size, shearing up to 20%, as well as random horizontal flipping was used. Due to hardware limitations we trained the network with a batch size of 1. The model with the highest Dice on the validation set was selected for final evaluation. Both baselines were trained with the exact same settings as their respective multiclass models.

4. Results

4.1. FAZ segmentation in FA from clinical routine (S_{EVAL})

Quantitative results for the U-Net and U-Net++ based models for S_{EVAL} are depicted in Table 2. Both models show an improvement in AUC and Dice values when using RV as guidance, with statistically significant improvements for both the U-Net and U-Net++ based models ($p=0.035$ for AUC and Dice for the UNet based model, $p=0.003$ for AUC and $p \approx 0$ for Dice for the UNet++

based model). The proposed multi-class model using the U-Net architecture as backbone shows the best overall performance with improvements of 1.1% and 1.7% in terms of AUC and Dice compared to its single-class baseline, respectively. Figure 4 illustrates the distribution of Dice values for our proposed and the baseline approach in this best model.

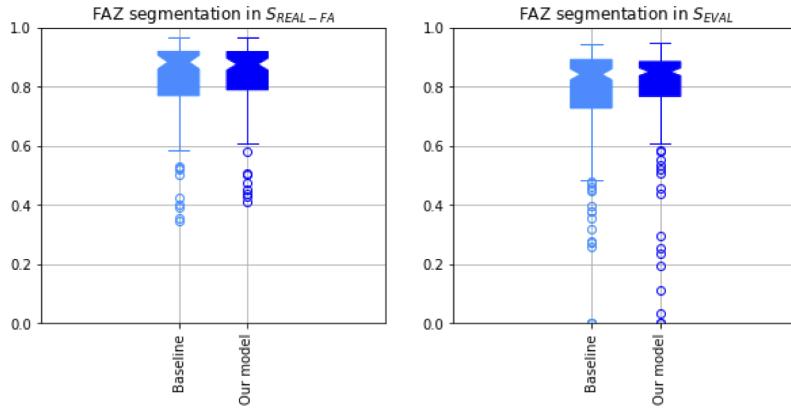


Fig. 4. Boxplots showing the Dice values for FAZ segmentation in testsets $S_{REAL-FA}$ and S_{EVAL} from the standard U-Net model.

Table 2. Results for FAZ segmentation in the additional external testset S_{EVAL} , showing the Area Under the precision-recall-Curve (AUC), precision (PR), recall (RE) and Dice.

Model	RV guidance	AUC	PR	RE	Dice
U-Net++	✗	0.807 ± 0.228	0.430 ± 0.216	0.9375 ± 0.174	0.530 ± 0.210
U-Net++	✓	0.822 ± 0.231	0.654 ± 0.232	0.801 ± 0.280	0.656 ± 0.227
U-Net	✗	0.881 ± 0.182	0.615 ± 0.228	0.942 ± 0.151	0.776 ± 0.181
U-Net	✓	0.892 ± 0.178	0.733 ± 0.221	0.893 ± 0.158	0.793 ± 0.170

4.2. FAZ segmentation in FA from multicenter clinical trials ($S_{REAL-FA}$)

Table 3 shows AUC, precision, recall and Dice values for our proposed method and the baseline approaches on the testset of $S_{REAL-FA}$. In line with the results on the external test set, the additional weak RV supervision improves the FAZ segmentation performance for both architectures. The best-performing multi-class U-Net architecture outperforms the baseline by 1.4% and 1.0% in terms of AUC and Dice, respectively. Moreover, the Dice values of this model on the $S_{REAL-FA}$ testset are illustrated in the left boxplot of Fig. 4.

Table 3. Results for FAZ segmentation in target dataset $S_{REAL-FA}$.

Model	RV guidance	AUC	PR	RE	Dice
U-Net++	✗	0.855 ± 0.170	0.696 ± 0.231	0.848 ± 0.181	0.716 ± 0.168
U-Net++	✓	0.884 ± 0.173	0.774 ± 0.201	0.848 ± 0.174	0.778 ± 0.167
U-Net	✗	0.907 ± 0.131	0.885 ± 0.163	0.840 ± 0.177	0.817 ± 0.145
U-Net	✓	0.921 ± 0.112	0.891 ± 0.141	0.809 ± 0.165	0.827 ± 0.113

For irregular FAZ shapes, we observed mean Dice values of 0.814 for the proposed U-Net based model and 0.812 for the corresponding baseline. For regular FAZ shapes the mean Dice

values are 0.852 for the proposed method and 0.825 for the baseline, respectively. These results indicate the improved generalizability of the proposed model to varying types of FAZ shapes.

Figure 5 depicts qualitative results of the U-Net based model on two exemplary samples, illustrating the original FA image, manual FAZ segmentation, single-target baseline FAZ segmentation as well as FAZ and RV predictions from our proposed approach. The first row in Fig. 5 shows a case with a pathologically large FAZ. Here, the additional RV segmentations help to improve the accuracy by forcing the network to provide a more coherent result, following the anatomical characteristics better. The second row shows a FAZ with a normal, round shape. In this case there is a slight improvement in accuracy compared to the baseline, where the RV supervision helps to remove a small outlier on the top of the FAZ segmentation.

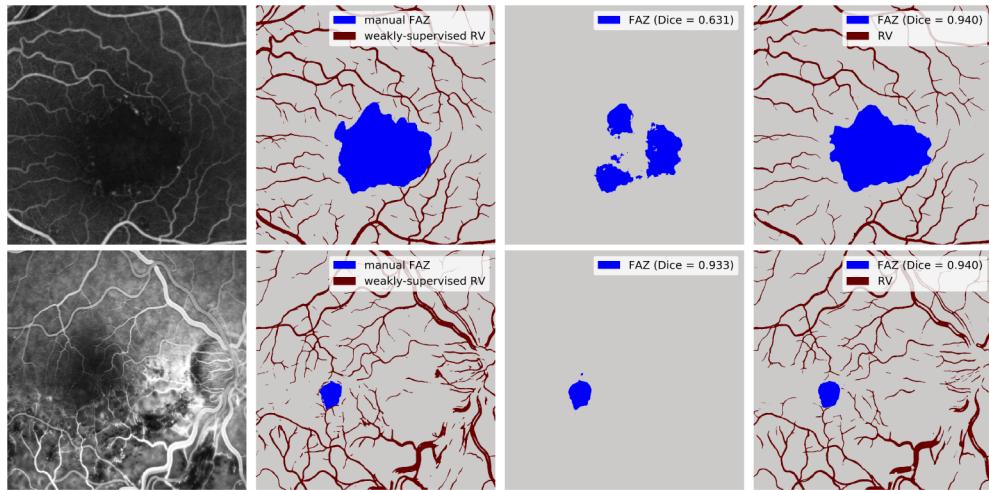


Fig. 5. Qualitative examples of the results of our segmentation model in the testset from clinical trials. From left to right: FA input image, manually annotated FAZ and automated vessel segmentations used for training the model, FAZ segmentation produced by the baseline, FAZ and RV segmentation results produced by the proposed model.

4.3. Converting CFP to pseudo-FA

For the conversion from CFP to pseudo-FA we observed an FID of 101.198. Figure 6 shows qualitative examples of how the CFP images are transformed into pseudo-FA images by our CycleGAN-model $G_{\text{CFP} \rightarrow \text{FA}}$. It is worth noting that the vessels are turned white and at the same time the tissue in the area of the macula stays black. This indicates that the CycleGAN learned a semantically meaningful transformation, as simple inversion of the input image would result in a bright fovea.

4.4. RV segmentation in clinical FA

The AUC and Dice for RV segmentation are 0.74 and 0.67, respectively. As shown in Fig. 7, the model predicts main RV structures in an accurate manner. However, we empirically observed that the segmentation performance of small bifurcations or vessels decreases in low-contrast images and images with motion artifacts.

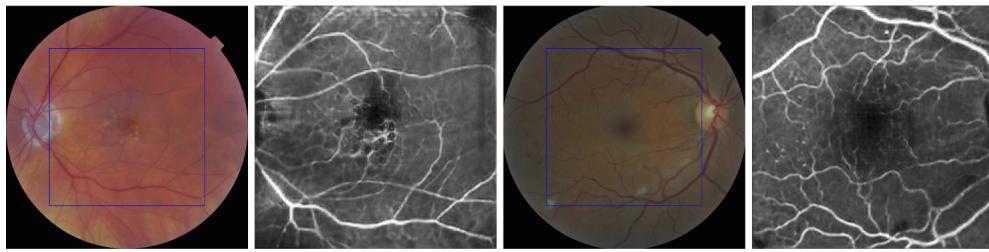


Fig. 6. Qualitative results obtained for converting CFP to pseudo-FA using our CycleGAN model. Left: input CFP. Right: pseudo-FA image converted from the CFP.

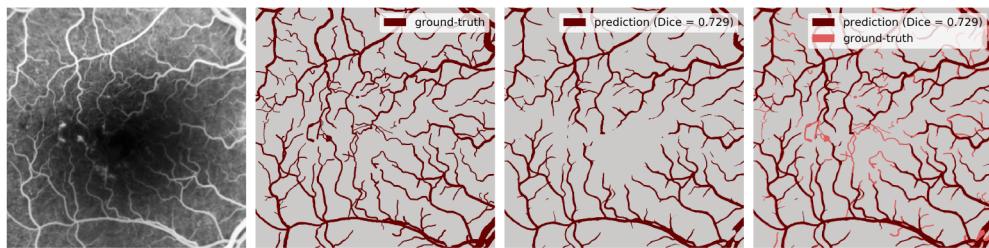


Fig. 7. Qualitative results of RV segmentation on real FA images. From left to right: input image, manual annotations, prediction of our weakly trained model comparison.

5. Discussion

In this paper, we proposed a framework that can accurately segment the FAZ in challenging FA images taken from large multicenter clinical trials and clinical routine. When done manually, the grader normally follows the edges of the RV around the macula to annotate the FAZ. Mimicking the annotation workflow of human experts, we propose that incorporating RV labels as extra supervision during training to guide the algorithm to improve detection of the FAZ contour. Hence, our model can simultaneously segment the RV in addition to the FAZ. The need for labourious manual annotation of RV ground-truth is mitigated by a multi-modal approach that leverages already existing labels from public CFP datasets. To the best of our knowledge, this is the first framework to use weak RV labels from another modality to improve FAZ segmentation in FA.

The majority of datasets that have been utilized in previous works consist of images with good quality only and are trained and tested in the same data distribution. In contrast, our model is trained on clinical data with a high variability (see Fig. 1), and we provide an extensive evaluation including a separate external testset from clinical routine with challenging images. Our results show that extra RV supervision aids the FAZ segmentation task, resulting in a significant improvement of the accuracy in terms of AUC and Dice (see Fig. 4 and Table 2).

Typical examples of wrong results in challenging clinical data are FAZ predictions that have an incorrect shape, additional disconnected segmentations or false positive regions lying outside of the fovea in the periphery (e.g. where the image has similar intensity values as the FAZ area). Addressing these outliers and wrong contours is a key aspect in FAZ detection. We empirically observed that our proposed model improves generalizability to varying types of shapes and reduces the number of outliers in the periphery. We hypothesize that this is due to the RV map, providing spatial context and aiding to find the location of the fovea.

Interestingly, Li et. al. [19] come to a different conclusion, stating that extra RV supervision using a multi-task model does not improve the FAZ segmentation accuracy in OCT-A projection

maps. We assume that this might be due to OCT-A data showing the RV with high contrast and in a quality where additional supervision becomes unnecessary for segmenting the FAZ.

Additionally to the FAZ detection, we prove that it is possible to get accurate RV segmentations in FA images without the need of manual annotations by leveraging public CFP datasets. However, we empirically observed that small bifurcations and vessels are sometimes missed by our model, which may be due to the low resolution of the pseudo-FA images, as the task gets harder with decreasing image-quality. Nonetheless, the extra RV supervision improves the FAZ segmentation significantly, raising the question regarding the importance of small vessels even in the manual annotation process by clinicians.

Other methods in literature that use modality conversion to obtain weak labels use simple image processing techniques like inversion of the grayscale CFP to imitate FA. However, our CycleGAN-based model yields a more realistic result with the fovea remaining black, while the RV turns white (see Fig. 6). With respect to the realism of this modality conversion, we obtained an FID value of 101.198, which is higher than those reported in related literature by paired image-to-image translation models (e.g. 43 by Tavakkoli et al. [47]). On the one hand, this can be explained by our use of unpaired CFP and FA data for training the CycleGAN in contrast to the paired training setting in [47], and on the other hand by our challenging dataset consisting of heterogeneous and often low-quality images. In general, however, there is no standard method to quantitatively evaluate the performance of a given GAN that is generally agreed upon in literature. Therefore we also provide qualitative samples, allowing manual inspection. Furthermore, the fact that the overall segmentation accuracy of the framework is improved indicates that the quality of modality conversion is sufficient.

A potential improvement of our approach might be the retinal vasculature segmentation accuracy in difficult cases. While our approach already improves the performance of the FAZ segmentation significantly, it would be interesting to see if refining those weak RV labels would result in an even better delineation of the FAZ. For example, the model could be fine-tuned on the few public FA images that are available or by manually annotating RV on clinical FA. Related to this, future work could also focus on investigating and improving the CFP-FA translation model. This might improve the RV segmentation result and consequently also the multi-class segmentation performance of the overall framework. Another research question worth investigating would be to apply this framework the other way around from FA to CFP, in order to see if the FAZ could be segmented in CFP. Currently, this task is even impossible to do for retina specialists. It would also be interesting to see if other modalities or labels could be included in this framework, and especially if the model could be transferred to OCT-A.

Finally, there is an increased interest in improving data pre-processing and engineering to get good results, as opposed to implementing complicated architectural designs. Here, we showed that providing RV as additional guidance (where all of the steps in our framework are based on straightforward DL-based models) can improve the segmentation result significantly, regardless of the backbone used in our experiments. Nevertheless, empirical evaluations using other architectures are still necessary to ensure that this holds general for any chosen network.

6. Conclusion

We proposed a segmentation model that simultaneously segments the FAZ and RV in heterogeneous FA images from clinical routine with various pathologies. We evaluated our algorithm in two datasets, coming from clinical trials and clinical routine, respectively. We showed that our model significantly improves the FAZ segmentation performance by using weak RV labels as extra supervision. Future work might include additional improvement of the RV labels by refining the modality conversion, segmenting the FAZ in CFP, or including further modalities such as OCT-A, or other pathologies.

Disclosures. DH, PS, JIO and FG declare no conflicts of interest. BSG: Roche (C), Bayer (C), Novartis (C), Digital Diagnostics (F). US-E: Genentech (C), Novartis (C), Roche (C), Heidelberg Engineering (C), Kodiak (C), RetInSight (C).

Data availability. Data underlying the final results presented in this paper are not publicly available at this time. The public datasets used in the framework are available at [27–32].

References

1. D. A. Salz and A. J. Witkin, "Imaging in diabetic retinopathy," *Middle East Afr. J. Ophthalmol.* **22**(2), 145 (2015).
2. A. Bajwa, R. Aman, and A. K. Reddy, "A comprehensive review of diagnostic imaging technologies to evaluate the retina and the optic disk," *Int. Ophthalmol.* **35**(5), 733–755 (2015).
3. M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.* **3**, 169–208 (2010).
4. M. B. Parodi, F. Visintin, P. D. Rupe, and G. Ravalico, "Foveal avascular zone in macular branch retinal vein occlusion," *Int. Ophthalmol.* **19**(1), 25–28 (1995).
5. J. Conrath, R. Giorgi, and D. Raccah, "Foveal avascular zone in diabetic retinopathy: quantitative vs qualitative assessment," *Eye* **19**(3), 322–326 (2005).
6. G. H. Bresnick, R. Condit, S. Syrigala, M. Palta, A. Groo, and K. Korth, "Abnormalities of the foveal avascular zone in diabetic retinopathy," *Arch. Ophthalmol.* **102**(9), 1286–1293 (1984).
7. L. de Sisterne, G. Jonna, M. A. Greven, Q. Chen, T. Leng, and D. L. Rubin, "Individual Drusen segmentation and repeatability and reproducibility of their automated quantification in optical coherence tomography images," *Trans. Vis. Sci. Tech.* **6**(1), 12 (2017).
8. U. Schmidt-Erfurth, A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunovic, "Artificial intelligence in retina," *Prog. Retin. Eye Res.* **67**, 1 (2018).
9. D. Hofer, J. I. Orlando, P. Seeböck, G. Mylonas, F. Goldbach, A. Sadeghipour, B. S. Gerendas, and U. Schmidt-Erfurth, "Foveal avascular zone segmentation in clinical routine fluorescein angiographies using multitask learning," in *Proc. OMIA* (Springer, 2019), pp. 35–42.
10. M. Iba nez and A. Simó, "Bayesian detection of the fovea in eye fundus angiographies," *Pattern Recog. Lett.* **20**(2), 229–240 (1999).
11. L. Ballerini, "Genetic snakes for medical images segmentation," in *Proc. EVolASP/EuroEcTel* (Springer, 1999), pp. 59–73.
12. A. Haddouche, M. Adel, M. Rasigni, J. Conrath, and S. Bourennane, "Detection of the foveal avascular zone on retinal angiograms using markov random fields," *Digit. Signal Process.* **20**(1), 149–154 (2010).
13. Y. Zheng, J. S. Gandhi, A. N. Stangos, C. Campa, D. M. Broadbent, and S. P. Harding, "Automated segmentation of foveal avascular zone in fundus fluorescein angiography," *Invest. Ophthalmol. Vis. Sci.* **51**(7), 3653–3659 (2010).
14. J. Conrath, O. Valat, R. Giorgi, M. Adel, D. Raccah, F. Meyer, and B. Ridings, "Semi-automated detection of the foveal avascular zone in fluorescein angiograms in diabetes mellitus," *Clin. Exp. Ophthalmol.* **34**(2), 119–123 (2006).
15. M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, R. M. Anjana, A. S. Doney, C. N. Palmer, and E. Trucco, "A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification," *Med. Image Anal.* **68**, 101905 (2021).
16. A. Simó and E. de Ves, "Segmentation of macular fluorescein angiographies. a statistical approach," *Pattern Recog.* **34**(4), 795–809 (2001).
17. J. Son, S. J. Park, and K.-H. Jung, "Retinal vessel segmentation in fundoscopic images with generative adversarial networks," ArXiv abs/1706.09318 (2017).
18. J. Son, S. J. Park, and K.-H. Jung, "Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks," *J. Digit. Imaging* **32**(3), 499–512 (2019).
19. M. Li, Y. Zhang, Z. Ji, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "IPN-V2 and OCTA-500: Methodology and dataset for retinal image segmentation," arXiv:2012.07261 (2020).
20. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (2015), pp. 234–241.
21. L. Ding, A. Kuriyan, R. Ramchandran, and G. Sharma, "Quantification of longitudinal changes in retinal vasculature from wide-field fluorescein angiography via a novel registration and change detection approach," in *Proc. ICASSP* (IEEE, 2018), pp. 1070–1074.
22. H. Jelinek, M. Cree, J. Leandro, J. Soares, R. Cesar Junior, and A. Luckie, "Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy," *J. Opt. Soc. Am. A* **24**(5), 1448–1456 (2007).
23. M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. B. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Med. Image Anal.* **11**(1), 47–61 (2007).
24. A. Perez-Rovira, E. Trucco, P. Wilson, and J. Liu, "Deformable registration of retinal fluorescein angiogram sequences using vasculature structures," in *Proc. EMBC* (IEEE, 2010), pp. 4383–4386.
25. A. Perez-Rovira, K. Zutis, J.-P. Hubschman, and E. Trucco, "Improving vessel segmentation in ultra-wide field-of-view retinal fluorescein angiograms," in *Proc. EMBC* (IEEE, 2011), pp. 2614–2617.
26. L. Ding, A. Kuriyan, R. Ramchandran, and G. Sharma, "Multi-scale morphological analysis for retinal vessel detection in wide-field fluorescein angiography," in *Proc. WNYISPW* (2017), pp. 1–5.

27. M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.* **59**(9), 2538–2548 (2012).
28. S. I. Holm, G. Russell, V. Nourrit, and N. McLoughlin, "DR HAGIS—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients," *SPIE Journal of Medical Imaging* **40**, 014503 (2017).
29. J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004).
30. A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *Int. J. Biomed. Imaging* **2013**, 1–11 (2013).
31. J. Orlando, J. Barbosa-Breda, K. Van Keer, M. Blaschko, P. Blanco, and C. Bulant, "Towards a glaucoma risk index based on simulated hemodynamics from fundus images," in *Proc. MICCAI* (Springer, 2018), pp. 65–73.
32. A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging* **19**(3), 203–210 (2000).
33. L. Ding, A. Kuriyan, R. Ramchandran, and G. Sharma, "Retinal vessel detection in wide-field fluorescein angiography with deep neural networks: A novel training data generation approach," in *Proc. ICIP* (IEEE, 2018).
34. L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "A novel deep learning pipeline for retinal vessel detection in fluorescein angiography," *IEEE Trans. on Image Process.* **29**, 6561–6573 (2020).
35. L. Ding, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning," *IEEE Transactions on Medical Imaging* **40** 2748 (2020).
36. Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Self-supervised deep learning for retinal vessel segmentation using automatically generated labels from multimodal data," in *Proc. IJCNN*, (2019), pp. 1–8.
37. K. J. Noh, S. J. Park, and S. Lee, "Fine-scale vessel extraction in fundus images by registration with fluorescein angiography," in *Proc. MICCAI*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, eds. (Springer International Publishing, Cham, 2019), pp. 779–787.
38. L. Ju, X. Wang, X. Zhao, P. Bonnington, T. Drummond, and Z. Ge, "Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling," *IEEE Transactions on Medical Imaging* **40** 2911 (2021).
39. E. O. Rodrigues, A. Conci, and P. Liatsis, "Element: Multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach," *IEEE J. Biomed. Health Inform.* **24**(12), 3507–3519 (2020).
40. Y. Zhao, Y. Liu, X. Wu, S. P. Harding, and Y. Zheng, "Retinal vessel segmentation: An efficient graph cut approach with retinex and local phase," *PLoS One* **10**(4), e0122332 (2015).
41. L. Sanchez Brea, D. A. De Jesus, S. Klein, and T. van Walsum, "Deep learning-based retinal vessel segmentation with cross-modal evaluation," in *Proc. MIDL*, vol. 121 of *Proceedings of Machine Learning Research* T. Arbel, I. Ben Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, and C. Pal, eds. (PMLR, 2020), pp. 709–720.
42. F. Schiffrers, Z. Yu, S. Arguin, A. Maier, and Q. Ren, *Synthetic Fundus Fluorescein Angiography using Deep Neural Networks* (Springer, Berlin Heidelberg, 2018), pp. 234–238.
43. W. Li, W. Kong, Y. Chen, J. Wang, Y. He, G. Shi, and G. Deng, "Generating fundus fluorescence angiography images from structure fundus images using generative adversarial networks," in *Proc. MIDL*, (2020).
44. Z. Cai, J. Xin, J. Wu, S. Liu, W. Zuo, and N. Zheng, "Triple multi-scale adversarial learning with self-attention and quality loss for unpaired fundus fluorescein angiography synthesis," in *Proc. EMBC*, (2020), pp. 1592–1595.
45. K. Li, L. Yu, S. Wang, and P.-A. Heng, "Unsupervised retina image synthesis via disentangled representation learning," in *Proc. SASHIMI*, (Springer International Publishing, Cham, 2019), pp. 32–41.
46. Á. Hervella, J. Rouco, J. Novo, and M. Ortega, "Retinal image understanding emerges from self-supervised multimodal reconstruction," in *Proc. MICCAI*, (Springer International Publishing, Cham, 2018), pp. 321–328.
47. A. Tavakkoli, S. A. Kamran, K. F. Hossain, and S. L. Zuckerbrod, "A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs," *Sci. Rep.* **10**(1), 21580 (2020).
48. Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Deep multimodal reconstruction of retinal images using paired or unpaired data," in *Proc. IJCNN*, (2019), pp. 1–8.
49. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, (2017).
50. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging* **39**, 1856 (2019).
51. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: a nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (Springer, 2018), pp. 3–11.