Applications of Deep Learning in Fundus Images: A Review

Tao Li[a], Wang Bo[a], Chunyu Hu[a], Hong Kang[a], Hanruo Liu[b], Kai Wang[a,*], Huazhu Fu[c]

[a]*College of Computer Science, Nankai University, Tianjin 300350, China*
[b]*Beijing Tongren Hospital, Capital Medical University, Address, Beijing 100730 China*
[c]*Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, UAE*

A B S T R A C T

The use of fundus images for the early screening of eye diseases is of great clinical importance. Due to its powerful performance, deep learning is becoming more and more popular in related applications, such as lesion segmentation, biomarkers segmentation, disease diagnosis and image synthesis. Therefore, it is very necessary to summarize the recent developments in deep learning for fundus images with a review paper. In this review, we introduce 143 application papers with a carefully designed hierarchy. Moreover, 33 publicly available datasets are presented. Summaries and analyses are provided for each task. Finally, limitations common to all tasks are revealed and possible solutions are given. We will also release and regularly update the state-of-the-art results and newly-released datasets at https://github.com/nkicsl/Fundus_Review to adapt to the rapid development of this field.

1. Introduction

According to the World Vision Report[1] released by the World Health Organization in October 2019, more than 418 million people worldwide suffer from glaucoma, diabetic retinopathy (DR), age-related macular degeneration (AMD) or other eye diseases which can cause blindness. Patients with eye diseases are often unaware of the aggravation of asymptomatic conditions (Robinson, 2003), so early screening and treatment of eye diseases is particularly important.

A fundus image is a projection of the fundus captured by a monocular camera on a 2D plane. Unlike other eye scans, such as OCT images and angiographs, fundus images can be acquired in a non-invasive and cost-effective manner, making them more suitable for large-scale screening (Edupuganti et al., 2018). An example of a fundus image is presented in Fig. 1.

Many important biomarkers can be seen in the fundus image, such as optic disc (OD), optic cup (OC), macula, fovea, blood vessel, and some DR related lesions, such as microaneurysms (MAs), hemorrhages (HEs), hard exudates (EXs), and soft exudates (SEs). Fundus images can be used to diagnose a variety of eye diseases, including glaucoma, DR, AMD, cataract, retinopathy of prematurity (ROP), and diabetic macular edema (DME).

Recently, data-driven deep learning has been widely applied to ophthalmic disease diagnosis based on fundus images. Compared to traditional methods that use manually designed features, deep learning models can achieve better performance by automatically optimizing the features in an end-to-end manner.

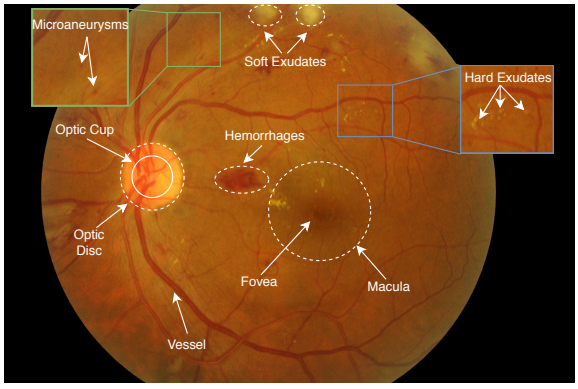


Fig. 1. A fundus image from the IDRiD dataset illustrating important biomarkers and lesions.

Most applications of deep learning in fundus images can be coarsely divided into classification, segmentation and synthesis tasks. For brevity, we only list widely used backbones in fundus image applications. Diagnosis and grading of ophthalmic diseases are two examples of classification tasks, VGG-Net (Simonyan and Zisserman, 2015), Inception (Szegedy et al., 2015, 2016, 2017), ResNet (He et al., 2016) and DenseNet (Huang et al., 2017) are the most widely used classification backbone networks. In terms of segmentation tasks, identifying lesions and biomarkers is of great importance in the diagnosis of diseases. In addition to those used for classification, other networks widely used for segmentation in fundus images include FCN (Long et al., 2015), SegNet (Badrinarayanan et al., 2017), U-Net (Ronneberger et al., 2015), MaskRCNN (He et al., 2017) and DeeplabV3+ (Chen et al., 2018). Finally, in the field of fundus image synthesis, generative adversarial network (GANs)

*Corresponding author: Tel.: +0-000-000-0000; fax: +0-000-000-0000;
[1]https://www.who.int/publications/i/item/world-report-on-vision

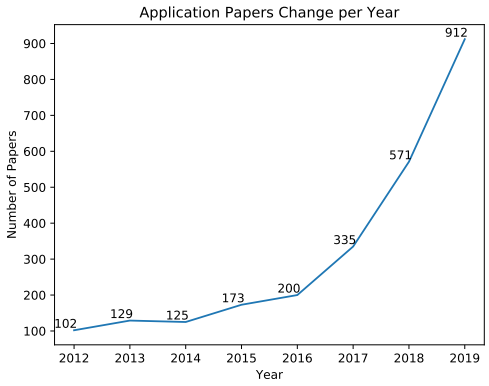


Fig. 2. Number of papers on fundus images and deep learning in recent years.

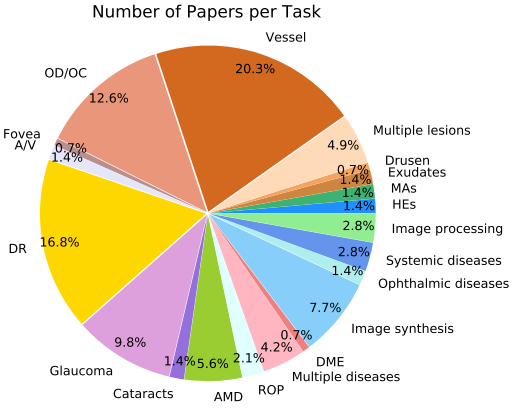
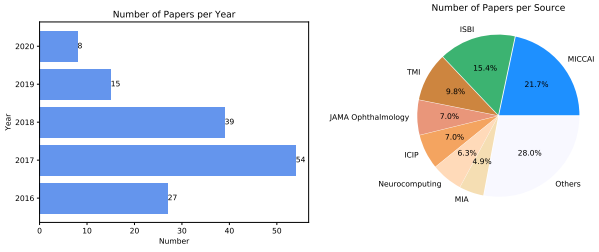


Fig. 3. The distribution of papers per task.



(a) Papers per year (b) Papers per source

Fig. 4. The distributions of papers per year and source.

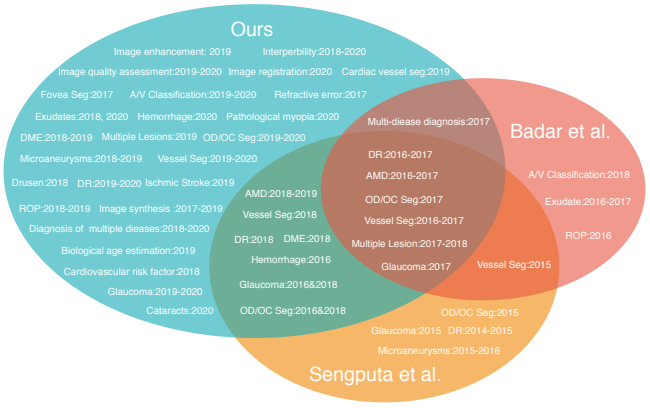


Fig. 5. Comparison with two recent reviews similar in view, style and year.

(Goodfellow et al., 2014) are the dominant architecture.

Motivation. The results in Fig. 2 show that the number of papers on fundus images and deep learning are increasing year by year. While several review papers already exist for these, they are all different from our review. For instance, Abramoff et al. (2010) and Zhang et al. (2014) only focus on classical machine learning methods; Salamat et al. (2019), Kanse and Yadav (2019) and Moccia et al. (2018) only consider specific individual diseases, such as DR or glaucoma; and Schmidt-Erfurth et al. (2018), Rahimy and Ehsan (2018) and Ting et al. (2018a,b, 2019) do not discuss specific deep learning methods or structures, but instead simply use "a deep learning method" and similar terms to refer to all methods. Therefore, it is necessary to provide a high-quality review that analyzes the trends and highlights the future directions for the applications of deep learning in fundus images.

Data. In this review, we focus on the successful application of deep learning methods in fundus images from January 2016 to August 2020. We collected 143 papers from the *DBLP*[2], *ScienceDirect*[3], *JAMA Network* [4], *Investigative Ophthalmology & Visual Science*[5] and *Web of Science*[6] databases using the following keywords: *retina*, *fundus*, *diabetes retinopathy*, *glaucoma*, *age-related macular degeneration*, *cataract*, *retinal vessel*, *optic disc / disk / cup*, *fundus / retinal + lesion / abnormal*, *hemorrhage*, *microaneurysm*, *exudate*, *neovascularization*, *drusen*, *fundus / retinal + synthesis / enhance*, *fundus / retinal + hypertension / stroke*, *fundus / retinal + kidney / brain / heart*, and *fundus / retinal + cardiovascular / cerebrovascular*.

The conference sources include *CVPR*, *AAAI*, *MICCAI*, *ISBI*, *ICMLA*, and *ICIP*, and the journal sources include *IEEE TIP*, *IEEE TMI*, *MIA*, *JAMA*, *JAMA Ophthalmology*, *Ophthalmology*, *Investigative Ophthalmology & Visual Science*, *Diabetes Care*, *Nature Biomedical Engineering*, *IEEE TBME*, *IEEE JBHI*, *Pattern Recognition*, *Neural Networks*, *Information Science*, *Knowledge Based Systems*, *Expert Systems with Applications*, *Neurocomputing* and *Future Generation Computer Systems*. The distribution of papers per task is shown in Fig. 3. Distributions of papers per year/source are shown in Fig. 4.

Two recent reviews, Badar et al. (2020) and Sengupta et al. (2020), are similar to ours in terms of view (fundus image, not ophthalmology), style (technical, not clinical), and method (deep learning, not artificial intelligence or machine learning). However, only 34 papers and 62 papers are reviewed in each, respectively, while we review 143 papers. Further, as shown in Fig. 5, the scopes are also different. Finally, as shown in Fig. 6, this review utilizes a carefully designed multi-layer hierarchy to organize the related works in a more intuitive manner.

[2]https://dblp.uni-trier.de/db/
[3]https://www.sciencedirect.com/
[4]https://jamanetwork.com/
[5]http://iovs.arvojournals.org/
[6]http://apps.webofknowledge.com/

Contributions. *First*, we give a comprehensive review of the applications of deep learning in fundus images. Compared to recent works, this review covers more recent papers, more eye diseases and more challenging tasks, especially including image synthesis and several interesting applications in Section 6. *Second*, we carefully design the taxonomy of our paper. A knowledge graph is summarized in Fig. 6. The lookup table for the references in the knowledge graph is presented in the *Appendix*. This can help readers to quickly find content of interest. *Third*, summaries and analyses are provided for each task. Limitations that are common to current approaches are also described and possible solutions given in Section 7. This may provide inspiring ideas for researchers in this field.

2. Lesion detection/segmentation

In this section we will review how deep learning methods have been applied to lesion detection/segmentation. The widely used datasets for this task are shown in Tab. 1. The *availability* column is linkable in the soft copy. Because of the correlation between lesion detection/segmentation and DR diagnosis, there is an overlap between datasets used for the two. The DIARETDB0 dataset (Kauppi et al., 2006) consists of 130 images, of which 110 contain signs of DR (EXs, SEs, MAs, HEs and neovascularization) and 20 are normal. DIARETDB1 (Kauppi et al., 2007) consists of 89 images, of which 84 contain at least mild non-proliferative signs of DR (MA) and five are normal. The RC-RGB-MA dataset (Dashtbozorg et al., 2018) contains 250 images. MAs were annotated by two experts at bounding-box level. Images in the RC-SLO-MA dataset (Dashtbozorg et al., 2018) were captured using scanning laser ophthalmoscopy (SLO). The dataset contains 58 images with MA labels. The Retinopathy Online Challenge (ROC) dataset (Niemeijer et al., 2010) consists of 100 images that have been divided into a training set and a test set, both containing 50 images. Center locations of MAs are labeled by experts. The e-ophtha dataset can be divided into two subsets, namely e-ophtha EX and e-ophtha MA. E-ophtha EX (Decenciere et al., 2013) provides pixel-level labels for EX segmentation. It consists of 47 images with exudates and 35 with no lesions. E-ophtha MA (Decenciere et al., 2013) consists of 148 images with MAs or small HEs and 233 healthy images. The Messidor dataset (Decenciere et al., 2014) consists of 1,200 images obtained from three ophthalmologic departments. 540 images are normal and 660 images are abnormal. Messidor is divided into three sets, one per department, with different resolutions. 800 images were acquired with pupil dilation and 400 without. Messidor-2 (Abràmoff et al., 2013) extended Messidor to 1,748 images. Unlike Messidor, images of Messidor-2 are all in pairs. The CLEOPATRA dataset (Sivaprasad et al., 2014) consists of 298 images obtained from 15 hospitals in the UK. It was acquired by different fundus cameras. Therefore, the images have different resolutions. Two experts were invited to annotate the ground truths for EXs, HEs and MAs. The first expert marked all images and the second marked 135 images. CLEOPATRA is not available online. The two names Kaggle and EyePACS (Foundation, 2015) refer to the same dataset which was provided by EyePACS and used in the "Diabetic Retinopathy Detection–Identify signs of diabetic retinopathy in eye images" Kaggle competition. Kaggle dataset consists of 35,126 training images graded into five DR stages and 53,576 test images of undisclosed stages. Images in the Kaggle dataset were obtained using multiple fundus cameras with diffrent fileds of view. The IDRiD dataset (Porwal et al., 2018) was used in the "Diabetic Retinopathy: Segmentation and Grading Challenge" held by ISBI in 2018. It consists of three tasks, namely segmentation, disease grading and localization, with official training and test sets provided. The segmentation task consists of 81 images with ground truths provided for lesions (MAs, HEs, EXs, SEs) and OD areas. The disease grading task consists of 516 images with severity grade for DR and DME. The localization task also consists of 516 images, with annotations for OD and fovea center localization. Note that images in IDRiD have relatively high resolution. The DDR dataset (Li et al., 2019b) consists of 13,673 images which were obtained from 147 hospitals, covering 23 provinces in China. Image level annotations with five classes of DR severity are provided for all images. In addition, 757 images are provided with pixel-level and bounding-box-level annotations for lesions (MAs, EXs, SEs and HEs).

Experimental results for lesion segmentation on the various datasets introduced in this section are provided in Tab. 2, 3, 4 and 5.

2.1. Hemorrhages

Hemorrhages (HEs) are one of the visible pathological signs of DR. Accurate detection or segmentation of HEs is important for DR diagnosis. In the task of lesion detection/segmentation, patch-based methods are quite popular because of the limited number of images in datasets and the need to reduce computational costs. Patch-based methods can generate tens of thousands of patches with only dozens of images, which can help improve performance and alleviate the problem of overfitting. However, HEs (as well as other lesions) are typically relatively small in size, with their pixels only making up a small proportion of the whole image. This leads to an imbalance problem, where only a few patches contain lesions and a large number do not contribute much to the lesion detection/segmentation task. Imbalance is also common in other lesion detection/segmentation tasks in this section, details of which will not be repeated for brevity. There are two main directions in the improvement of hemorrhage detection/segmentation; namely selective sampling and performing segmentation on coarsely-annotated datasets.

Selective sampling. van Grinsven et al. (2016) proposed a method called selective sampling to reduce the use of redundant data and speed up CNN training. They invited three experts to relabel the Messidor dataset and a subset of Kaggle. During the training process, weights of samples were dynamically adjusted according to the current iteration's classification results, so that the informative samples were more likely to be included in the next training iteration. Inspired by VGG, they designed a nine-layer CNN as the classifier. On the Kaggle competition and Messidor datasets, experimental results showed that the CNN with selective sampling (SeS) outperformed the CNN

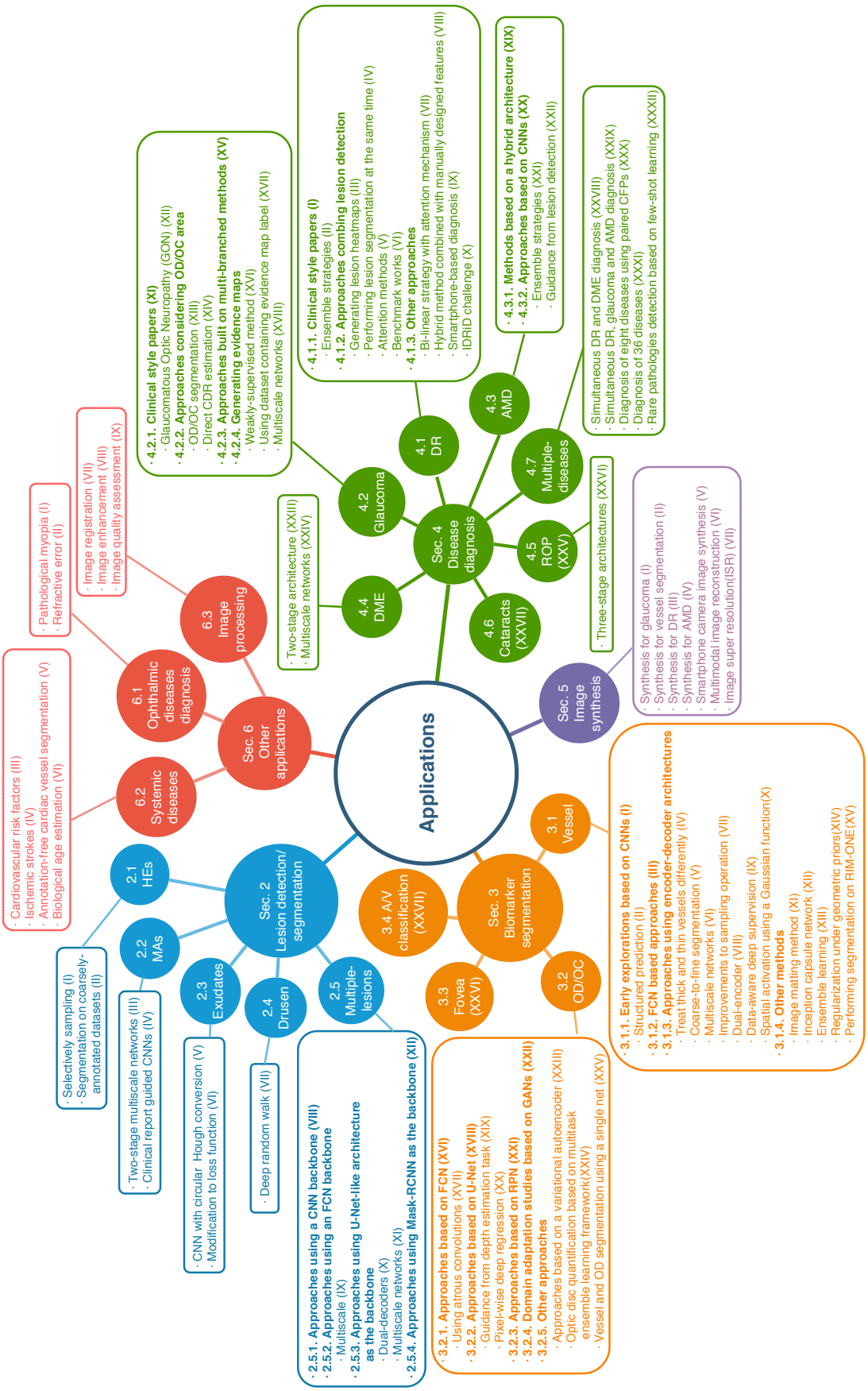


Fig. 6. The knowledge graph is summarized in this review. The lookup table for the references in the graph can be found in the *Appendix*.

Table 1. Widely used datasets of lesion detection/segmentation and DR diagnosis/grading

Dataset name	Number of images	Resolution	Camera	Availability
DIARETDB0	130 (110 DR, 20 Normal)	-	digital fundus cameras with unknown camera settings, FVO 50°	available online[1]
DIARETDB1	89 (84 DR, 5 Normal)	1500×1152	ZEISS FF 450plus fundus camera with Nikon F5 digital camera, FOV 50°	available online [2]
Retinopathy Online Challenge	100	-	a Topcon NW 100, a Topcon NW 200, or a CanonCR5-45NM, 2 differently shaped FOVs	available on registration[3]
RC-RGB-MA	250	2595×1944	a DRS non-mydriatic fundus camera, FOV45°	available online[4]
RC-SLO-MA	58	1024×1024	an EasyScan camera (i-Optics Inc., the Netherlands), FOV45°	available online[5]
IDRiD	516	4288×2848	a Kowa VX-10 alpha digital fundus camera, FOV 50°	available online[6]
Messidor	1200	1440×960, 2240×1488, 2304×1536	a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with FOV 45°	available on registration[7]
Messidor-2	1748	1440×960, 2240×1488, 2304×1536	a Topcon TRC NW6 non-mydriatic fundus camera with FOV 45°	available on registration[8]
e-ophtha EX	47 with 12,278 exudates, 35 healthy	ranging from 1440×960 to 2544×1696	-	available on registration[9]
e-ophtha MA	148 with 1306 MA, 233 healthy	ranging from 1440×960 to 2544×1696	-	available on registration[10]
DDR	13,673	mixed	42 types of fundus cameras with a 45°FOV	available online[11]
Kaggle/EyePACS	35,126 train, 53,576 test	-	multiple fundus cameras and different fields of views	available on registration[12]
CLEOPATRA	298	-	multiple fundus cameras	not available online

[1]https://www.it.lut.fi/project/imageret/
[2]https://www.it.lut.fi/project/imageret/
[3]http://webeye.ophth.uiowa.edu/ROC/
[4]http://www.retinacheck.org/datasets
[5]http://www.retinacheck.org/datasets
[6]https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid
[7]http://www.adcis.net/en/third-party/messidor/
[8]http://www.adcis.net/en/third-party/messidor2/
[9]http://www.adcis.net/en/third-party/e-ophtha/
[10]http://www.adcis.net/en/third-party/e-ophtha/
[11]https://github.com/nkicsl/DDR-dataset
[12]https://www.kaggle.com/c/diabetic-retinopathy-detection/data

Table 2. Summary of several results for lesion detection/segmentation on IDRiD dataset

Reference	Backbone	Loss	PR/%	SE/%	SP/%	ACC/%	AUPR/%	AUC/%	F1/%
Hemorrhage detection/segmentation									
Guo et al. (2019)	FCN	Top-k loss, Bin loss	-	-	-	-	-	**67.34**	-
Yan et al. (2019a)	U-Net	weighted CE	-	-	-	-	**70.3**	-	-
Microaneurysms detection/segmentation									
Sarhan et al. (2019)(geometric)	FCN	Dice loss, CE and Triplet loss	**61.128**	28.07	-	-	41.96	-	**38.4877**
Guo et al. (2019)	FCN	Top-k loss, Bin loss	-	-	-	-	-	**46.27**	-
Yan et al. (2019a)	U-Net	weighted CE	-	-	-	-	**52.5**	-	-
Xue et al. (2019)	Mask-RCNN	log loss, regression loss, CE loss	-	**76.4**	**99.8**	**99.7**	-	-	-
Hard exudate detection/segmentation									
Guo et al. (2020a)	HED	Top-k loss, Bin loss	-	**95.74**	-	-	-	**98.71**	**95.57**
Guo et al. (2019)	FCN	Top-k loss, Bin loss	-	-	-	-	-	79.45	-
Yan et al. (2019a)	U-Net	weighted CE	-	-	-	-	**88.9**	-	-
Xue et al. (2019)	Mask-RCNN	log loss, regression loss, CE loss	-	77.9	**99.6**	**99.2**	-	-	-
Soft exudate detection/segmentation									
Guo et al. (2019)	FCN	Top-k loss, Bin loss	-	-	-	-	-	**71.13**	-
Yan et al. (2019a)	U-Net	weighted CE	-	-	-	-	**67.9**	-	-

Table 3. Summary of several results for lesion detection/segmentation on E-ophtha dataset

Reference	Task	Backbone	Loss	PR/%	SE/%	SP/%	ACC/%	AUPR/%	AUC/%	F1/%
Carson et al. (2018)	MA classification	CNN	-	-	-	-	-	**86**	**94**	-
Guo et al. (2019)	MA segmentation	FCN	Top-k loss, Bin loss	-	-	-	-	-	**16.87**	-
Xue et al. (2019)	MA segmentation	Mask-RCNN	log loss, regression loss and CE loss	-	**67.2**	**99.8**	**99.7**	-	-	-
Carson et al. (2018)	Exudates classification	CNN	-	-	-	-	-	**64**	**95**	-
Guo et al. (2020a)	EX detection	HED	Top-k loss, Bin loss	-	**86.44**	-	-	-	**91.84**	**87.01**
Guo et al. (2019)	EX segmentation	FCN	Top-k loss, Bin loss	-	-	-	-	-	41.71	-
Xue et al. (2019)	EX segmentation	Mask-RCNN	log loss, regression loss and CE loss	-	84.6	**98.8**	**98.4**	-	-	-
Playout et al. (2019)	Bright Lesion segmentation	U-Net	loss based on Cohen's coefficient	**78.50**	**80.02**	**99.88**	**99.77**	-	-	**79.25**
Playout et al. (2019)	Red Lesion segmentation	U-Net	loss based on Cohen's coefficient	**75.26**	**75.62**	**99.99**	**99.88**	-	-	**75.44**

Table 4. Summary of several results for lesion detection/segmentation on DiaretDB1 dataset

Reference	Task	Backbone	Loss	PR/%	SE/%	SP/%	ACC/%	AUC/%	F1/%
Dai et al. (2018)	MA detection	CNN	-	**99.7**	**87.8**	-	**96.1**	**93.4**	-
Adem (2018)	Exudate detection	CNN	-	-	**99.2**	**97.97**	-	-	-
Playout et al. (2018)	Bright lesion segmentation	U-Net	loss based on Cohen's coefficient	-	75.35	99.86	-	-	-
Playout et al. (2019)	Bright lesion segmentation	U-Net	loss based on Cohen's coefficient	**81.70**	**88.29**	**99.93**	**99.89**	-	**84.87**
Playout et al. (2018)	Red lesion segmentation	U-Net	loss based on Cohen's coefficient	-	66.91	99.82	-	-	-
Playout et al. (2019)	Red lesion segmentation	U-Net	loss based on Cohen's coefficient	**78.96**	**85.18**	**99.89**	**99.83**	-	**81.95**

Table 5. Summary of several results for lesion detection/segmentation on other datasets

Reference	Task	Dataset	Backbone	Loss	SE/%	SP/%	AUC/%	mAP/%
van Grinsven et al. (2016)	HE detection	Kaggle	CNN	CE	**83.7**	**85.1**	**89.4**	-
van Grinsven et al. (2016)	HE detection	Messidor	CNN	CE	**91.9**	**91.4**	**97.2**	-
Huang et al. (2020)	HE segmentation	private	CNN	MSE, IoU, GIoU	-	-	-	**52.20**
Yan et al. (2018a)	Drusen segmentation	STARE, DRIVE	Encoder-decoder Network	-	**92.02**	**97.30**	-	-
Adem (2018)	Exudate detection	DiaretDB0	CNN	-	**100**	**98.41**	-	-
Adem (2018)	Exudate detection	DrimDB	CNN	-	**100**	**98.44**	-	-
Tan et al. (2017)	EX detection	CLEOPATRA	CNN	log-likelihood function	**87.58**	**98.73**	-	-
Tan et al. (2017)	HE detection	CLEOPATRA	CNN	log-likelihood function	**62.57**	**98.93**	-	-
Tan et al. (2017)	MA detection	CLEOPATRA	CNN	log-likelihood function	**46.06**	**97.99**	-	-
Guo et al. (2019)	EX segmentation	DDR	FCN	Top-k loss, Bin loss	-	-	**55.46**	-
Guo et al. (2019)	SE segmentation	DDR	FCN	Top-k loss, Bin loss	-	-	**26.48**	-
Guo et al. (2019)	HE segmentation	DDR	FCN	Top-k loss, Bin loss	-	-	**35.86**	-
Guo et al. (2019)	MA segmentation	DDR	FCN	Top-k loss, Bin loss	-	-	**10.52**	-

without selective sampling (NSeS), and SeS reduced the number of training epochs from 170 to 60.

Segmentation on coarsely-annotated datasets. Huang et al. (2020) proposed a bounding box refining network (BBR-Net) which can generate more accurate bounding box annotations for coarsely-annotated data. Then they utilized a RetinaNet (Lin et al., 2017) to detect hemorrhage. Rather than using the finely annotated IDRiD dataset, they performed hemorrhage detection on a private dataset with coarsely annotated bounding box. They first established a dataset containing image pairs. For each pair, one image was taken from IDRiD and the other was obtained by simulating coarsely-annotated bounding boxes. BBR-Net took coarsely annotated patches as input and finely annotated patches as target. After training, the authors introduced their private data to obtain more accurate bounding box annotations, and then sent the results to the RetinaNet for hemorrhage detection.

Discussion. The selective sampling method alleviates the problem of data imbalance. Selective sampling is also used in other applications, which will be introduced in the following sections. The explorations made by Huang et al. (2020) also offer a promising direction. The generation of more accurate bounding box annotations can be seen as image synthesis, which will be discusses in more detail in Section 5.

However, there are still some limitations in the current HEs detection applications. First, the imbalance problem needs to be further studied. Second, compared to other lesions, less research has focused on HEs. More attention needs to be paid to this area for its importance in DR diagnosis. Third, pixel-level segmentation and detection are required. More datasets that provide pixel-level labels for HEs, like DDR, still need to be explored.

2.2. Microaneurysms

MAs are the earliest clinical sign of DR and have thus captured more research interests. There are several barriers affecting the segmentation of MAs, including the existence of other lesions with similar color, extremely low contrast, and variation in image lighting, clarity and background texture. Two-stage multiscale architectures and guidance from clinical reports are some successful strategies for MAs detection.

Two-stage multiscale networks. Sarhan et al. (2019) proposed a two-stage deep learning approach embedding a triplet loss for microaneurysm segmentation. The first stage is called the hypothesis generation network (HGN), in which multiscale FCNs are employed to generate a region of interest (ROI). The second stage is known as the patch-wise refinement network (PRN), in which patches extracted from around ROIs are passed to a modified ResNet-50 for classification. The authors introduced the triplet loss into the PRN to extract discriminative features. Further, the previously mentioned selective sampling method (van Grinsven et al., 2016) is utilized to reduce the computational cost and solve the data imbalance problem.

Clinical report guided CNNs. Dai et al. (2018) proposed a clinical report guided multi-sieving convolutional neural network (MS-CNN) for the detection of MAs. They first trained a weak image-to-text model from clinical reports and fundus images to generate a rough segmentation of microaneurysms. Then the proposed MS-CNN was used to generate final high-quality segmentation using the rough segmentation as guidance. In order to tackle the data imbalance problem, MS-CNN adopts a method similar to boosting. Specifically, MS-CNN is composed of multiple CNNs, where the false positives from the previous CNN are fed into the following CNN as negative examples.

Discussion. Several effective methods have been employed in MA segmentation, including multiscale networks, guidance from clinical reports and utilization of the triplet loss. The extraction of ROIs and cascaded architecture adopted in MS-CNN alleviate the imbalance problem. However, the two-stage architecture and cascaded architecture of MS-CNN lack efficiency. They use multiple base networks, leading to a huge number of

parameters to be trained. Thus, one promising direction in MA segmentation would be to reduce complexity of the networks while maintaining high performance.

2.3. Exudates

Soft and hard exudates are usually the basis for the diagnosis of DR. Accurate detection of SEs and EXs are thus crucial for timely treatment. Like other lesion detection/segmentation tasks, there are several challenges. The barriers include low contrast, varied sizes and similarity to other lesions. There are several approaches for exudate detection, most of which can be divided into CNN with circular Hough conversion and modifications to the loss function.

CNN with circular Hough conversion. Adem (2018) introduced a three-layer CNN architecture for the binary classification of exudated and exudate-free fundus images. During pre-processing, the OD region was removed by applying several methods, including adaptive histogram equalization, Canny edge detection and circular Hough conversion.

Modification to loss function. Guo et al. (2020a) proposed a top-k loss and a bin loss to enhance performance for exudate segmentation. The class balanced cross entropy (CBCE) loss (Xie and Tu, 2015) solved the class imbalance problem to some extent. However, this introduced the new problem of loss imbalance, where background similar to exudate tends to be misclassified. The main reason is that with the different weights for background and foreground pixels in CBCE loss, the loss for misclassifying a background pixel is much smaller than that for misclassifying a hard exudate pixel. To solve this loss imbalance problem, top-k loss is proposed, which considers all hard exudate pixels but only top-k background pixels with the larger loss. They also proposed a fast version of top-k loss named bin loss with consideration of efficiency.

Discussion. In exudate detection, some works have focused on modifying the loss function. The top-k loss and bin loss solved the loss imbalance problem caused by the use of CBCE. However, the misclassification problem still remains. Moreover, only baseline models have been tested and no innovative architecture has been proposed. Adem (2018)'s work is based on a CNN. Several more recent models, such as encoder-decoder networks, need to be utilized in exudate detection.

2.4. Drusen

Drusen, the main manifestations of the disease, can be used to assist in the diagnosis of AMD. There are four main challenges to drusen segmentation: their yellowish-white color is similar to the fundus image and OD; uneven brightness and interference from other biomarkers, such as blood vessels, is common; the drusen often have irregular shapes; and boundaries may be blurred.

Deep random walk. Yan et al. (2018a) proposed a deep random walk method to successfully segment drusen from fundus images. The proposed architecture is composed of three main parts. Fundus images are first passed into a deep feature extraction module, which consists of two branches: a SegNet-like network capturing deep semantic features and a three-layer CNN capturing low-level features. Then the captured features are fused together and passed into a named affinity learning module to obtain pixel-pixel affinities for formulating the transition matrix of the random walk. Finally, a deep random walk module is applied to propagate manual labels. This model achieved state-of-the-art performance on the STARE and DRIVE datasets.

Discussion. As can be seen, only one effective approach has been introduced so far. Other architectures and methods need to be explored. Further, the segmentation of drusen is closely related to the diagnosis of AMD. Therefore, one of the future works can be extending the original drusen segmentation to serve as an evidence for AMD diagnosis.

2.5. Multiple lesions

Most previous works only segment/detect one type of lesion or treat all lesions as a single group (usually red lesions or bright lesions). However, segmenting multiple lesions simultaneously is of more practical value. More and more researchers are thus focusing on multi-lesion segmentation/detection. The challenges found in the individual lesion detection/segmentation tasks, including imbalance, contrast, illumination, etc, still exist. Further the inter-class similarity for different lesions, such as HEs and MAs becomes more prominent. All these factors make multi-lesion segmentation a challenging task.

2.5.1. Approaches using a CNN backbone

Tan et al. (2017) conducted the first work to segment multiple lesions, including exudates, haemorrhages and microaneurysms, automatically and simultaneously using a 10-layer CNN, with the outputs evaluated at the pixel level. Their work demonstrated that it is possible to segment several lesions simultaneously using a single CNN architecture. Carson et al. (2018) used a CNN to perform five-class classification on image patches. The five classes consist of 1) normal, 2) microaneurysms, 3) dot-blot hemorrhages, 4) exudates or cotton wool spots, and 5) high-risk lesions such as neovascularization, venous beading, scarring, and so forth. They invited two ophthalmologists to verify and relabel a subset of the Kaggle dataset containing 243 images. The image patching method was used, proving that good performance can be obtained using such a method, even with limited training samples.

2.5.2. Approaches using an FCN backbone

Multiscale networks are important models that have been applied to many fields. Guo et al. (2019) proposed a small object segmentation network (L-Seg) which can segment four kinds of lesions, including microaneurysms, soft exudates, hard exudates and hemorrhages, simultaneously. The backbone network is a VGG-16, which has five groups of convolution layers and three fully connected (FC) layers. They removed all the FC layers and the fifth pooling layer and added a side extraction layer which consists of a 1×1 conv and upsampling to every conv group (except the first one) with deep supervision. The final output is obtained by multiscale weighted fusion of the side extraction layers, instead of simple element-wise sum. The bin loss (Guo et al., 2020a) was also used to solve the problem of class imbalance and loss imbalance.

2.5.3. Approaches using U-Net-like architecture as the backbone

There are two main directions explored in this section, namely dual-decoders and multiscale networks.

Dual-decoders. Playout et al. (2018) proposed an extension to U-Net which is capable of segmenting red and bright lesions simultaneously. They are the first to use fully convolutional approaches for joint lesion segmentation. Several novel developments were used in their decoder, including residual connections, global convolutions and mixed-pooling. They used two identical decoders, each specialized for one lesion category. Near the end of training, they also added two fully connected conditional random fields (CRFs) (Krähenbühl and Koltun, 2011). In their subsequent work, Playout et al. (2019), made several modifications. They proposed a novel unsupervised method to enhance segmentation performance by training the network at image-level labels when pixel-level annotations are limited. They introduced an exchange layer which aims to share parameters between two decoders softly, instead of employing hard parameter sharing as previously (Playout et al., 2018).

Multiscale networks. Yan et al. (2019a) combined local and global features to segment microaneurysms, soft exudates, hard exudates and hemorrhages. A GlobalNet was used to capture more context features, taking a downsampled version of original images as input. They also employed a LocalNet which takes cropped image patches as input, aiming to capture more detailed information. GlobalNet and LocalNet both use a U-Net-like encoder-decoder architecture as their backbone.

2.5.4. Approaches using Mask-RCNN as the backbone

Xue et al. (2019) proposed a deep membrane system for simultaneous MAs, EXs and OD segmentation. A hybrid structure, consisting of a dynamic membrane system and communication channels between cells, was designed. Three types of rules, i.e. T-rules, G-rules and D-rules were proposed for the computation and communication of the system, solving complex real applications in parallel. Mask-RCNN served as the computational cell of the membrane system.

2.5.5. Discussion

In this subsection, we have seen that various base networks have been applied to multi-lesion detection, including recent models like U-Net and Mask-RCNN. Multiscale methods have also been explored, and have been proven quite suitable for this task. Architectures modifications have also been introduced, with the dual-decoder being one notable example. Such a framework may work well on other similar scenarios. The proposed deep membrane system is quite innovative in fundus image analysis and is expected to be further explored.

However, there are still some limitations. First, compared to other segmentation and detection tasks like blood vessel segmentation and OD/OC segmentation, the performance, and in particular sensitivity of lesion segmentation/detection needs to be further improved. Second, there are still several works that focus on red or bright lesion segmentation instead of individual lesions. However, the specific segmentation and detection of individual lesions is more practical. Third, pixel-wise segmentation should be emphasized, and datasets with pixel-level lesion annotations deserve more attention.

3. Biomarker segmentation

3.1. Vessel segmentation

Segmentation of retinal blood vessels is of paramount importance in the diagnosis of various ophthalmic diseases including diabetic retinopathy and glaucoma (Abràmoff et al., 2010). With the use of powerful deep learning techniques such as CNNs, FCNs and recently U-Net, excellent performance has been achieved. However, there still remain some factors making retinal blood vessel segmentation a challenging task. These factors include varying contrast and intensity among different datasets, inter-vessel differences between thick and thin vessels, the presence of optic disc and lesions, limited annotated data and so on. We will discuss how these problems were addressed in the following subsections.

The most commonly used datasets in retinal blood vessel segmentation include DRIVE, STARE, CHASE_DB1 and HRF. The DRIVE dataset (Staal et al., 2004) consists of 40 images, seven of which show signs of mild early DR. DRIVE is officially divided into a training set and a test set, both containing 20 images. A single manual segmentation of vessels is provided in the training set and two manual segmentations are provided in the test set. Border masks are also available for all images. The STARE dataset (Hoover et al., 2000) consists of 400 images, 20 of which have two manual blood vessel segmentations annotated by two experts. Ten of the images contain pathologies. Coarsely annotated centerline-level artery/vein labels of 10 images are also provided. The CHASE_DB1 dataset (Owen et al., 2009) consists of 28 images, obtained from both eyes of 14 multi-ethnic school children. The HRF dataset (Budai et al., 2013) consists of 45 images, of which 15 are healthy, 15 have DR and 15 are glaucomatous. Compared to the other three datasets, images from HRF have a higher resolution (3504×2336). More details are shown in Tab. 6. Experimental results on different datasets are shown in Tab. 7, 8, 9 and 10.

3.1.1. Early explorations based on CNNs

Before fully convolutional networks were widely used, vessel segmentation was regarded as a pixel-by-pixel classification task and structured prediction was still a problem to be solved. The usual approach was to crop the images into patches as input, and uses CNNs whose last few layers are fully connected to predict the label of the center pixel of each patch. The approach proposed by Khalaf et al. (2016) is a typical one. They used a CNN containing three conv layers to perform vessel segmentation. The last FC layer of the proposed CNN contains three neurons, representing the probability of central pixels being large vessels, small vessels or background, respectively. Yu et al. (2020) also used a CNN whose last layers are FC layers for vessel segmentation. And they conducted further research based on the segmentation results. They first extracted vascular trees from the segmented vessels using a graph-based method.

Table 6. Widely used datasets for vessel segmentation

Dataset name	Number of images	Resolution	Camera	Availability
DRIVE	40 (33 healthy, 7 mild early DR)	768×584	a Canon CR5 non-mydriatic 3CCD camera, FOV 45°	available on registration[1]
STARE	400 (vessel segmentation labeling of 40 , A/V labeling of 10)	700 × 605	a TopCon TRV-50 fundus camera, FOV35°	available online[2]
CHASE_DB1	28	1280× 960	-	available online[3]
HRF	45, 15 each of healthy, DR and glaucomatous	3504 × 2336	a Canon CR-1 fundus camera with FOV 45°	available online[4]

[1]https://drive.grand-challenge.org/Download/
[2]http://cecas.clemson.edu/ ahoover/stare/
[3]https://blogs.kingston.ac.uk/retinal/chasedb1/
[4]http://www5.cs.fau.de/research/data/fundus-images/

Table 7. Summary of several results for vessel segmentation on DRIVE dataset

Reference	Backbone	Loss	SE/%	SP/%	ACC/%	AUC/%	F1/%
Khalaf et al. (2016)	CNN	-	83.97	95.62	94.56	-	-
Liskowski and Krawiec (2016)	CNN	CE	**91.60**	92.41	92.30	97.38	-
Yu et al. (2020)	CNN	-	76.43	98.03	95.24	97.23	-
Fu et al. (2016)	FCN	CBCE	76.03	-	95.23	-	-
Dasgupta and Singh (2017)	FCN	CE	76.91	98.01	95.33	97.44	-
Feng et al. (2017)	FCN	CBCE	78.11	98.39	95.60	97.92	-
Oliveira et al. (2018)	FCN	categorical CE	80.39	98.04	95.76	98.21	-
Zhang and Chung (2018)	U-Net	CE	87.23	96.18	95.04	97.99	
He et al. (2018)	U-Net	Focal loss	77.61	97.92	95.19	-	81.29
Yan et al. (2018b)	U-Net	Proposed segment-level loss	76.53	98.18	95.42	97.52	-
Yan et al. (2019b)	U-Net	CE	76.31	98.20	95.38	97.50	-
Wu et al. (2018)	U-Net	CE	78.44	98.19	95.67	98.07	-
Wu et al. (2020)	U-Net	CE	79.96	98.13	95.82	98.30	-
Wang et al. (2020)	U-Net	CE	78.49	98.13	95.67	97.88	82.41
Hu et al. (2018)	FCN	improved CE	77.72	97.93	95.33	97.59	-
Wu et al. (2019)	U-Net	CE	80.38	98.02	95.78	98.21	-
Soomro et al. (2019)	SegNet	CBCE	87	98.5	95.6	98.6	-
Zhang et al. (2019a)	U-Net	-	81.00	98.48	96.92	98.56	-
Wang et al. (2019a)	U-Net	CE and Jaccard loss	79.40	98.16	95.67	97.72	**82.70**
Ma et al. (2019)	U-Net	CE	79.16	98.11	95.70	98.10	-
Zhao et al. (2020a)	Dense U-Net	global pixel loss, local matting loss	83.29	97.67	-	-	82.29
Mishra et al. (2020)	U-Net	CE	89.16	96.01	95.40	97.24	-
Feng et al. (2020)	FCN	MSE	76.25	98.09	95.28	96.78	-
Cherukuri et al. (2020)	Residual FCN	MSE	84.25	**98.49**	**97.23**	**98.70**	-
Kromm and Rohr (2020)	CapsNet	margin loss	76.51	98.18	95.47	97.50	-
Liu et al. (2019a)	No-reference net	MSE	80.72	97.80	95.59	97.79	82.25

Table 8. Summary of several results for vessel segmentation on STARE dataset

Reference	Backbone	Loss	SE/%	SP/%	ACC/%	AUC/%	F1/%
Liskowski and Krawiec (2016)	CNN	CE	**93.07**	93.04	93.09	98.20	-
Yu et al. (2020)	CNN	-	78.37	98.22	96.13	97.87	-
Fu et al. (2016)	FCN	CBCE	74.12	-	95.85	-	-
Oliveira et al. (2018)	FCN	categorical CE	83.15	98.58	96.94	99.05	-
Zhang and Chung (2018)	U-Net	CE	76.73	99.01	97.12	98.82	-
He et al. (2018)	U-Net	Focal loss	81.20	98.95	97.04	-	85.53
Yan et al. (2018b)	U-Net	Proposed segment-level loss	75.81	98.46	96.12	98.01	-
Yan et al. (2019b)	U-Net	CE	77.35	98.57	96.38	98.33	-
Wu et al. (2020)	U-Net	CE	79.63	98.63	96.72	98.75	-
Wang et al. (2020)	U-Net	CE	90.24	**99.34**	**98.49**	**99.60**	**91.84**
Hu et al. (2018)	FCN	improved CE	75.43	98.14	96.32	97.51	-
Feng et al. (2020)	FCN	MSE	77.09	98.48	96.33	97	-
Soomro et al. (2019)	SegNet	CBCE	84.8	98.6	96.8	98.8	-
Cherukuri et al. (2020)	Residual FCN	MSE	86.64	98.95	98.03	99.35	-
Zhao et al. (2020a)	Dense U-Net	global pixel loss, local matting loss	84.33	98.57	-	-	83.51
Mishra et al. (2020)	U-Net	CE	87.71	96.34	95.71	97.42	-
Liu et al. (2019a)	No-reference net	MSE	77.71	98.43	96.23	97.93	80.36

Table 9. Summary of several results for vessel segmentation on CHASE_DB1 dataset

Reference	Backbone	Loss	SE/%	SP/%	ACC/%	AUC/%	F1/%
Fu et al. (2016)	FCN	CBCE	71.30	-	94.89	-	-
Oliveira et al. (2018)	FCN	categorical CE	77.79	98.64	96.53	98.55	-
Zhang and Chung (2018)	U-Net	CE	76.70	**99.09**	97.70	99.00	-
Yan et al. (2018b)	U-Net	Proposed segment-level loss	76.33	98.09	96.10	97.81	-
Yan et al. (2019b)	U-Net	CE	76.41	98.06	96.07	97.76	-
Wu et al. (2018)	U-Net	CE	75.38	98.47	96.37	98.25	-
Wu et al. (2020)	U-Net	CE	80.03	98.80	96.88	98.94	-
Wang et al. (2020)	U-Net	CE	79.48	98.42	96.48	98.47	82.20
Wu et al. (2019)	U-Net	CE	81.32	98.14	96.61	98.60	-
Soomro et al. (2019)	SegNet	CBCE	**88.6**	98.2	97.6	98.5	-
Zhang et al. (2019a)	U-Net		81.86	98.48	97.43	98.63	-
Cherukuri et al. (2020)	Residual FCN	MSE	80.17	99.08	**97.88**	98.64	-
Wang et al. (2019a)	U-Net	CE and Jaccard loss	80.74	98.21	96.61	98.12	80.37
Mishra et al. (2020)	U-Net	CE	88.05	96.51	96.01	97.63	-
Liu et al. (2019a)	No-reference net	MSE	87.69	98.43	97.42	**99.05**	**85.98**

Table 10. Summary of several results for vessel segmentation on HRF dataset

Reference	Backbone	Loss	SE/%	SP/%	ACC/%	AUC/%	F1/%
Soomro et al. (2019)	SegNet	CBCE	**82.9**	96.1	**96.2**	**98.5**	-
Zhao et al. (2020a)	Dense U-Net	global pixel loss, local matting loss	78.09	**98.18**	-	-	**78.13**

Then two algorithms were proposed for the hierarchical division of retinal vascular networks.

Structured prediction has been explored by several researchers. Liskowski and Krawiec (2016) used a CNN for segmentation. Their FC layer comprises two neurons representing the vessel and background. They also explored a structured prediction scheme, which can simultaneously predict labels of all pixels in an s×s window of an n×n patch. Their approach was to set the number of the last FC layer's neurons to s×s. Each neuron represents one pixel in the window, and the output is a set of two-dimensional vectors instead of scalars.

3.1.2. FCN based approaches

Fully convolutional networks provide an end-to-end solution, addressing the issue of structured prediction. Hence, they were quickly applied to vessel segmentation. Fu et al. (2016) proposed a fully convolutional network called DeepVessel. They employed a side-output layer to help the network learn multiscale features. At the end of the net, a CRF layer was used to further model non-local pixel correlations. Dasgupta and Singh (2017) also proposed a fully convolutional network. Their network contains six conv layers, one downsampling layer and one upsampling layer. Feng et al. (2017) proposed a fully convolutional network which can be considered as a simplified version of U-Net. Their network only upsamples and downsamples twice. In order to solve the class imbalance between background and blood vessels, they defined an entropy which measures the proportion of vessel pixels in the patch. During the training process, half of the patches are selected from the patches with the highest entropy, and the other half are randomly selected. Oliveira et al. (2018) proposed an FCN architecture which is similar to that of Feng et al. (2017). In the pre-processing phase, they utilized a stationary wavelet transform (SWT) to obtain additional channels for the input images. Hu et al. (2018) proposed a multiscale network inspired by RCF (Liu et al., 2017), which merges feature maps of every middle layer with the output. Similar to Guo et al. (2019), they also removed all the FC layers of VGG-16 as their backbone network. At the end of their net, fully connected CRFs are employed. An improved cross-entropy loss was also proposed to focus on hard examples.

3.1.3. Approaches using encoder-decoder architectures

Because of their excellent ability to extract features and extraordinary performance in practice, encoder-decoder architectures, especially U-Net, are still the most popular segmentation frameworks applied to fundus images up to now. There are many directions for improvement in this area, as will be discussed next.

Treating thick and thin vessels differently. In order to improve performance on capillaries, one possible solution is to treat thick and thin vessels differently. Zhang and Chung (2018) proposed a multi-label architecture. They used opening and dilation operations to expand the original vessel and background into five classes, namely 0 (other background pixels), 1 (background near thick vessels), 2 (background near thin vessels), 3 (thick vessels) and 4 (thin vessels). The proposed architecture uses a U-Net with residual connection as the backbone. A side-output layer was also introduced to capture multiscale features. He et al. (2018) introduced an operation named local de-regression (LODESS) to get additional labels. After the LODESS, the original binary labels (vessel and background) were further divided into five classes, specifically 0 (the center of big vessels), 1 (the edge of big vessels), 2 (the center and edge of small vessels), 3 (the center of background) and 4 (edge of background). Yan et al. (2018b) introduced a segment-level loss which assigns different weights to different segments according to their thickness. They first obtained vessel segments from the whole vessel tree based on skeletonization and then estimated the relative thickness of each segment. Then a weight assigning strategy was designed to give thinner segments higher weights. Yan et al. (2019b) proposed a three-stage model for vessel segmentation. They first applied a skeletonization method to extract the skeletons. For each skeleton pixel, the diameter of the maximum inscribed circle that is completely covered by vessel pixels is considered the thickness. A ThickSegmenter and a ThinSegmenter were utilized for thick and thin vessel segmentation respectively. Note that, when calculating the loss, only thick vessel pixels were counted for the ThickSegmenter and thin vessels for the ThinSegmenter. Finally, the results of the two segmenters were passed to a FusionSegmenter to get the final result.

Coarse-to-fine segmentation. This is another approach that employs two branches: the first takes fundus images as input to get a preliminary result and the second further refines it. Wu et al. (2018) proposed a multiscale network followed network (MS-NFN) to improve performance on capillaries. Input images are passed into two different branches, namely the 'up-pool' NFN and 'pool-up' NFN. The two branches both have identical U-Net-like structures. The first network converts input patches into a probability map, and the second performs further refinement. The difference between the two NFNs is that the 'up-pool' NFN upsamples before downsampling, and 'pool-up' is the opposite. Finally, probability maps of the two NFNs are averaged to generate the final prediction. In their subsequent work (Wu et al., 2020), they added some modifications to NFN to form a new network named NFN+. Compared to NFN, the main extensions include: introducing inter-network connections between the preceding and following networks; replacing the 'up-pool' and 'pool-up' networks with an identical U-Net-like architecture; and removing the ensemble operation. Wang et al. (2020) proposed a coarse-to-fine supervision network (CTF-Net) for vessel segmentation. Their CTF-Net consists of two U-shaped networks, namely the coarse segNet producing preliminary predicted map and the fine segNet further enhancing performance. They also proposed a feature augmentation module (FAM-residual block) to improve the ability of the network to extract features.

Mutliscale networks. This is another important direction that has been explored. Wu et al. (2019) proposed Vessel-Net, which is based on the multiscale method. They first implemented an Inception-Residual (IR) block inspired by Inception and ResNet that can be embedded into U-Net. Four supervision paths were introduced to the net, including: a traditional supervision path; a richer feature supervision path, which resizes

all stages of the encoder's output to the same size as the input patches (48×48) and then concatenates them; and two multiscale supervision paths, where feature maps generated by the encoder with size 12×12 and 24×24 are passed into a 1×1 conv layer with Relu and softmax. Feng et al. (2020) proposed a cross-connected convolutional neural network (CcNet) for vessel segmentation, which also utilizes the multiscale method. The CcNet had two paths. The first is the primary path, which has more convolutional kernels than the other path to extract more features. The other is called the secondary path. Each conv layer of the primary path is connected to all conv layers of the secondary path to learn multiscale features.

Improvements to sampling operation. The downsampling and upsampling operations will change the resolution of the feature maps, which is not ideal for the segmentation task. Several works thus tried to improve or replace these two operations. Soomro et al. (2019) proposed a strided-CNN model to improve the sensitivity. They first performed pre-processing including morphological mappings and principal component analysis (PCA). Then the processed images were passed to a SegNet-like encoder-decoder architecture. The pooling operation was replaced with a strided-conv inspired by Springenberg et al. (2015). Zhang et al. (2019a) proposed the Attention Guided Network (AG-Net) for vessel segmentation. An attention guided filter inspired by He et al. (2013) was proposed. Specifically, it takes high-resolution feature maps from the encoder and low-resolution feature maps from the lower stage of the decoder as input and produces high-resolution feature maps as output. The attention guided filter can preserve edge and structural information. Note that AG-Net can also perform OD/OC segmentation.

Dual-encoder. Wang et al. (2019a) proposed the Dual Encoding U-Net (DEU-Net). DEU-Net consists of two encoders. The first, inspired by the global convolutional network (Peng et al., 2017), has a spatial path with larger kernels to capture more spatial information. The second, inpired by Inception, has a context path with multiple kernels to get more context features. A feature fusion module was proposed to fuse the features extracted by the two encoders at the top stage. Channel attention was used to replace the skip-connection in the original U-Net.

Data-aware deep supervision. Mishra et al. (2020) added a data-aware deep supervision path to a U-Net-like network. Based on the concept of effective receptive field (EFT) proposed by Luo et al. (2016), where the output-affecting region is actually smaller than the theoretical receptive field (RF), they proposed the concept of layer-wise effective receptive fields (LERFs), which are calculated by the gradient of the loss function using back-propagation. The average vessel width was taken as the target object size. The convolutional layer with the smallest absolute difference between its LERF and vessel width was selected as the target layer, and considered as the preeminent layer. Deep supervision was used in the target layer.

Spatial activation using a Gaussian function. Ma et al. (2019) proposed a multitask network which can perform vessel segmentation and A/V classification simultaneously. In view of the observation that the value of capillary vessels and boundary vessels in a probability map is close to 0.5, they proposed a spatial activation module that assigns higher weights to the thin vessels by a Gaussian function. Deep supervision was also utilized.

3.1.4. Other methods

Several other methods have been proposed to improve model performance. They not only achieve good experimental performance, but are also inspiring.

Image matting method. Zhao et al. (2020a) transformed the segmentation problem to a related matting problem. A trimap was first obtained using a bi-level thresholding of the score map. Then the retinal images and corresponding trimaps were sent to an end-to-end matting network to get the foreground matte. They proposed a local matting loss together with a global pixel loss for training. The final segmentation map was obtained by applying a threshold to all pixels of the matte.

Inception capsule network. Kromm and Rohr (2020) combined the Capsule network (Sabour et al., 2017) with the Inception architecture for vessel segmentation and centerline extraction. Their Inception Capsule network has a shallow architecture with fewer parameters and does not need data augmentation.

Ensemble learning. Liu et al. (2019a) proposed a novel and simple unsupervised ensemble strategy for vessel segmentation. They multiplied the output results of the best performing recent networks by the weights to obtain a result. The weights of results were then trained and they finally obtained better results than a single network.

Regularization under geometric priors. Cherukuri et al. (2020) proposed a domain enriched deep network for vessel segmentation. A representation network was first employed. Two geometrical regularizers, including an orientation diversity regularizer and a data adaptive noise regularizer, were added to the loss function to learn specific geometric features. After that they introduced a network containing residual blocks with no downsampling/upsampling steps, instead of using U-Net-like most other works.

Performing segmentation on RIM-ONE. Nasery et al. (2020) performed vessel segmentation on the RIM-ONE dataset. Compared to the DRIVE dataset, RIM-ONE is of a lower quality and does not have any vessel annotations. Instead of performing image synthesis to get high-quality images, they transformed high-quality images with expert labels from the DRIVE dataset to resemble poor-quality target images. To accomplish this, substantial vignetting masks were used. Then a U-Net was trained using the resulting images and their corresponding labels. Once trained, the net could be used to obtain vessel masks of images from the RIM-ONE dataset.

3.1.5. Discussion

From the model discussed in this section, we can see the development of base networks used, for vessel segmentation, from CNNs to FCNs to U-Net-like architectures. The use of CNNs and FCNs has become less common recently, while U-Net-like architectures are very popular. However, the feature extraction ability of U-Net is inadequate. U-Net only has 10 convolution operations in the encoder, which is even less than that in

VGG-16. Therefore, several works have focused on how to improve the feature extraction ability. Alternatives include Dense U-Net, Residual U-Net, and a dual-encoder network. Another disadvantage of U-Net is that there are four paired sampling operations (downsampling and upsampling), which is not ideal for the segmentation task. Several studies have tried to alleviate this problem by, for instance, using a shallower version of U-Net that has two or three paired sampling operations. Multiscale methods can also be utilized to improve the performance of the segmentation task. Low-level spatial features and high-level semantic features can both be focused on by the network. In order to solve the problem of poor performance on thin and edge vessels, method for treating thick and thin vessels differently have been explored. It is worth noting that there are several other inspiring and also interesting studies, including data-aware deep supervision, spatial activation, image matting, using a Inception capsule network, ensemble learning and performing segmentation on RIM-ONE.

However, there are still several limitations. First, there is still room for improvement in thin and edge vessels segmentation. Specifically, sensitivity and accuracy need to be further improved while maintaining specificity and AUC. Second, there are only three commonly used datasets for vessel segmentation, namely DRIVE, STARE and CHASE_DB1. And they contain fewer images than datasets for other tasks. On the one hand, more experiments need to be carried on the high-resolution HRF dataset. On the other hand, more images should be collected and annotated. Researchers can also employ image synthesis methods. Third, the imbalance problem also exists in the vessel segmentation task and it is even more challenging to solve than in other tasks like lesion and OD/OC segmentation for the irregular shape of blood vessels. The typical approach is to use a class-balancing loss. It is worth noting that a selective sampling method based on entropy could be effective. More attention should thus be paid to the imbalance problem.

3.2. OD/OC Segmentation

Cup-to-Disc ratio (CDR) is a widely accepted and used standard for the diagnosis of glaucoma. It is calculated as the ratio of vertical cup diameter (VCD) and vertical disc diameter (VDD) (Phene et al., 2019). The segmentation of the optic cup (OC) and optic disc (OD) is therefore very important for the diagnosis of glaucoma. Compared to OD segmentation, OC segmentation is a more challenging task for its subtle boundaries. Further, there is an imbalance problem for OC, as the OC region only accounts for a low proportion of extracted ROIs.

The datasets used in this field are shown in Tab. 11. Similar to lesion segmentation/detection, there is also an overlap between datasets used in OD/OC segmentation and glaucoma diagnosis. The ONHSD dataset (Lowell et al., 2004) consists of 99 images obtained from 50 patients of various ethnic backgrounds. Further, 96 images have discernable ONH. The Drions-DB dataset (Carmona et al., 2008) consists of 110 images, belonging to 55 patients with glaucoma (23.1%) and eye hypertension (76.9%). The images were obtained from a hospital in Spain. The ORIGA dataset (Zhang et al., 2010) consists of 650 images, of which 168 are glaucomatous and 482 are normal. The boundaries of OD and OC, CDR value and a label indicating whether glaucoma exists or not are provided for each image. The RIM-ONE-r3 dataset (Fumero et al., 2011) consists of 169 ONH images, of which 118 are normal, 11 have ocular hypertension (OHT) and 40 are glaucomatous. Five-class labels were provided by five experts. The ACHIKO-K dataset (Zhang et al., 2013) consists of 258 images, which were obtained from 67 glaucomatous patients from Korea. 144 images are of glaucomatous eyes and 114 are normal. The Drishti-GS dataset (Sivaswamy et al., 2014) contains 101 images, which are officially divided into 50 training images and 51 test images. Images were obtained from a hospital in India. The SCES dataset (Baskaran et al., 2015) consists of 1,676 images, each from a single subject, which only provide clinical diagnoses. 46 images of SCES are glaucomatous. The RIGA dataset (Almazroa et al., 2018) is made up of three parts, namely 460 images from MESSIDOR, 195 images from the Bin Rushed Ophthalmic Center and 95 images from the Magrabi Eye Center, making the total number of images 750. Each image was manually annotated by six ophthalmologists. The LAG dataset (Li et al., 2020b) contains 11,760 images, of which 6,882 do not have glaucoma and 4,878 are suspicious. 5,824 images were further annotated with attention labels, in which 2,392 display glaucoma and the remaining 3,432 do not.

Experimental results are shown in Tab. 12, 13, 14, 15 and 16. "A" in the tables means balanced accuracy,"E" is the widely used overlapping error, and δ denotes absolute CDR error.

3.2.1. Approaches based on FCN

Similar to blood vessel segmentation, fully convolutional networks were widely used in early OD/OC segmentation. Edupuganti et al. (2018) used FCN-8s to perform OD/OC segmentation. They also explored various strategies, such as assigning higher weights to edges in the loss function, for further improvement.

Using atrous convolutions is quite common in this field. Mohan et al. (2018) proposed a structure named Fine-Net which has a symmetrical encoder-decoder architecture. Inspired by full-resolution residual networks (FRRNs) (Pohlen et al., 2017), they used several full-resolution residual units (FRRU) in Fine-Net. An atrous convolution was also introduced to alleviate the memory cost while ensuring reliable performance at the same time. In their subsequent work, Mohan et al. (2019) proposed P-Net. P-Net is used to obtain preliminary segmentation results by taking downscaled images as input. The output of P-Net is upscaled and then sent to Fine-Net as guidance for further segmentation. DenseBlock and atrous convolution are combined as the dense atrous block (DB) to formulate P-Net. DBs with different dilation rates are used to capture multiscale features, inspired by atrous spatial pyramid pooling (ASPP) (Chen et al., 2017). Liu et al. (2019e) proposed a spatial-aware neural network which adopts a multiscale method. First, an atrous CNN model is used to extract spatially denser feature maps. Then, the extracted features are fed to a pyramid filtering module to obtain multiscale features. Finally, the multiscale features are passed to a spatial-aware segmentation network to get the final result.

Table 11. Widely used datasets for OD/OC segmentation and glaucoma diagnosis/grading

Dataset name	Number of images	Resolution	Camera	Availability
ONHSD	100	640×480	a Canon CR6 45MNf fundus camera, FOV 45°	available online[1]
Drishti-GS	101	2896×1944	a fundus camera with FOV 30°	available online[2]
Drions-DB	110	600×400	a colour analogical fundus camera	available online[3]
ORIGA	650 (168 glaucomatous, 482 normal)	3072×2048	-	not available online
RIGA	750	ranging from 2240×1488 to 2743×1936	multiple fundus cameras with different FOV	available online[4]
RIM-ONE	169 ONH	-	a fundus camera Nidek AFC-210 with a body of a Canon EOS 5D Mark II of 21.1 megapixels	not available online
ACHIKO-K	258 (144 glaucomatic)	640×480; 2144×1424; 3216×2136, etc	NIKON D80, NIKON D90	available online[5]
SEED	235 (43 glaucoma)	-	-	not available online
REFUGE	1200	2124×2056, 1634×1634	a Zeiss Visucam 500 fundus camera and a Canon CR-2 device	available online[6]
SCES	1676	3072×2048	-	not available online
SINDI	5783	3072×2048	-	not available online
LAG	11,760 (6882 glaucoma)	ranging from 582×597 to 3456×5184	3 types of devices: Topcon, Canon and Carl Zeiss	available online[7]

[1]http://www.aldiri.info/Image%20Datasets/ONHSD.aspx
[2]http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php
[3]https://www.researchgate.net/publication/326460478_Glaucoma_dataset_-_DRIONS-DB
[4]https://deepblue.lib.umich.edu/data/concern/data_sets/3b591905z/
[5]https://oar.a-star.%20edu.sg/jspui/handle/123456789/1080?mode=full
[6]https://refuge.grand-challenge.org/
[7]https://github.com/smilell/AG-CNN

Table 12. Summary of several results for OD/OC segmentation on Drishiti-GS dataset

Reference	Backbone	Loss	OD		OC		δ
			Dice/%	IoU/%	Dice/%	IoU/%	
Edupuganti et al. (2018)	FCN	weighted CE	-	69.58	-	81.22	-
Mohan et al. (2018)	FCN	bootstrapped CE and Dice loss	96.4	-	-	-	-
Mohan et al. (2019)	FCN	bootstrapped CE and Dice loss	97.13	-	-	-	-
Liu et al. (2019e)	FCN	spatial-aware error function	**98**	-	89	-	-
Shankaranarayana et al. (2019)	Encoder-decoder net	multi-class CE	96.3	-	84.8	-	0.1045
Shah et al. (2019)(PSBN)	U-Net	logarithmic dice loss	95	91	88	**80**	-
Shah et al. (2019)(WRoIM)	U-Net	logarithmic dice loss	96	**93**	89	**80**	-
Wang et al. (2019c)	Deeplab, GAN	dice coefficient loss, smoothness loss and adversarial loss	97.4	-	**90.1**	-	**0.048**
Wang et al. (2019b)	DeeplabV3+, GAN	CE, MSE, Adversarial loss	96.1	-	86.2	-	-

Table 13. Summary of several results for OD/OC segmentation on ORIGA dataset

Reference	Backbone	loss	OD		OC		Rim		δ
			A/%	E	A/%	E	A/%	E	
Liu et al. (2019e)	FCN	spatial-aware error function	-	0.059	-	**0.208**	-	**0.215**	-
Fu et al. (2018a)	U-Net	proposed multi-label loss	98.3	0.071	93.0	0.230	94.1	0.233	0.071
Shankaranarayana et al. (2019)	Encoder-decoder net	multi-class CE	97.4	**0.051**	92.8	0.212	-	-	0.067
Yin et al. (2019)	RPN	Multi-label CE	**98.6**	0.066	**94.2**	**0.208**	**94.9**	0.224	**0.065**
Jiang et al. (2020)	atrous CNN and RPN	Smooth L_1 loss and BCE	-	0.063	-	0.209	-	-	0.068

Table 14. Summary of several results for OD/OC segmentation on RIM-ONE-r3 dataset

Reference	Backbone	loss	OD				OC				δ
			A/%	E	Dice/%	IoU/%	A/%	E	Dice/%	IoU/%	
Shankaranarayana et al. (2019)	Encoder-decoder net	multi-class CE	**97.5**	**0.058**	**97.0**	-	**92.0**	**0.284**	**87.6**	-	0.066
Shah et al. (2019)(PSBN)	U-Net	logarithmic dice loss	-	-	91	84	-	-	75	60	-
Shah et al. (2019)(WRoIM)	U-Net	logarithmic dice loss	-	-	94	**90**	-	-	82	**71**	-
Wang et al. (2019c)	Deeplab, GAN	dice coefficient loss, smoothness loss, adversarial loss	-	-	96.8	-	-	-	85.6	-	**0.049**
Wang et al. (2019b)	DeeplabV3+, GAN	CE, MSE, Adversarial loss	-	-	89.8	-	-	-	81.0	-	-

Table 15. Summary of several results for OD/OC segmentation on REFUGE dataset

Reference	Backbone	Loss	OD			OC			Rim		δ
			A/%	E	Dice/%	A/%	E	Dice/%	A/%	E	
Wang et al. (2019f)	RPN	Weighted CE, regression loss	-	-	95.3	-	-	87.2	-	-	0.047
Yin et al. (2019)	RPN	Multi-label CE	**97.9**	**0.088**	-	**98.0**	**0.223**	-	**93.6**	**0.204**	0.048
Wang et al. (2019c)	Deeplab, GAN	dice coefficient loss, smoothness loss and adversarial loss	-	-	**96.02**	-	-	**88.26**	-	-	**0.0450**
Liu et al. (2019d)	GAN	dice segmentation loss, adversarial loss and MSE loss	-	-	94.16	-	-	86.27	-	-	0.0481

Table 16. Summary of several results for OD/OC segmentation on other datasets

Reference	Dataset	Backbone	Loss	OD		OC		δ
				E	Dice/%	E	Dice/%	
Mohan et al. (2018)	DrionsDB	FCN	bootstrapped CE, Dice loss	-	95.5	-	-	-
Mohan et al. (2019)	DrionsDB	FCN	bootstrapped CE, Dice loss	-	**96.6**	-	-	-
Mohan et al. (2018)	MESSIDOR	FCN	bootstrapped CE, Dice loss	-	95.7	-	-	-
Mohan et al. (2019)	MESSIDOR	FCN	bootstrapped CE, Dice loss	-	**96.8**	-	-	-
Jiang et al. (2020)	SCES	atrous CNN, RPN	Smooth L_1 loss, BCE	**0.063**	-	**0.209**	-	**0.068**
Sedai et al. (2017a)	EyePACS	VAE	negative KL-divergence, BCE	-	-	-	-	**0.80**

3.2.2. Approaches based on U-Net

There are also several works using U-Net as the baseline. Fu et al. (2018a) proposed M-Net for OD/OC segmentation. M-Net contains a multiscale input layer. Images are down-sampled to form an image pyramid as U-Net's input, to obtain a multi-level receptive field. In order to segment OD and OC simultaneously, the authors proposed a multi-label loss based on the dice loss, which is intended to solve the problem of multi-label and data imbalance. Moreover, a polar transformation was introduced to obtain spatial consistency and increase the proportion of the OD/OC in a patch. Their approach greatly inspired subsequent work.

Shah et al. (2019) proposed two different methods named the Parameter-Shared Branched Network (PSBN) and Weak Region of Interest Model-based segmentation (WRoIM). With fewer parameters, they obtained comparable performance to state-of-the-art approaches. PSBN has two branches, which are used to generate masks for the OD and OC, respectively. Encoders of the two branches share parameters, and the OC branch uses cropped activations from the OD branch. WRoIM first obtains a coarse OD area through a small U-Net structure (one conv block for downsampling and one conv block for upsampling), and uses the extracted ROI as another U-Net's input to perform fine segmentation.

Guidance from depth estimation task can boost performance of OD/OC segmentation. Shankaranarayana et al. (2019) proposed a fully convolutional network for retinal depth estimation and used the results to guide OD/OC segmentation. They proposed a Dilated Residual Inception (DRI) module utilizing convolution kernels with different dilation rates in the way of an Inception Block to extract multiscale features. In order to ensure that the retinal depth estimation branch guides OD/OC segmentation, they proposed a multi-modal feature fusion module to fuse feature maps from the depth estimation branch and the OD/OC segmentation branch.

Pixel-wise deep regression is a novel direction under exploration. Meyer et al. (2018) reformulated the segmentation task as a pixel-wise regression task to perform OD and fovea detection simultaneously. A bi-distance map is first obtained, illustrating the distance between every pixel and its nearest landmarker, namely OD and fovea. Then a U-Net-like deep network is utilized for distance regression and obtaining a globally consistent prediction map.

3.2.3. Approaches based on RPN

The RPN method from Faster-RCNN (Ren et al., 2015) and Mask R-CNN (He et al., 2017) has also been applied in the segmentation of OD/OC. Inspired by RPN, Wang et al. (2019f) proposed the Ellipse Proposal Network (EPN) to detect ellipse regions. They used $\{X_0, Y_0, F_1, F_2\}$ as parameters of elliptical anchors, where (X_0, Y_0) denotes the center coordinate of the ellipse and F_1, F_2 denote the major and minor axis, respectively. Two EPN branches were used to detect OD and OC. In addition, a spatial attention path was introduced between the OD to OC branches to guide the detection of OC. Yin et al. (2019) proposed a PM-Net, which is also inspired by RPN. They introduced a segmentation branch to RPN to provide more accurate proposals for localization of the optic nerve head (ONH) area. A pyramid RoIAlign module was also proposed to capture multiscale features. Jiang et al. (2020) proposed JointRCNN for OD and OC segmentation. An atrous-VGG16 is first used for feature extraction. The extracted features are passed to two parallel branches, i.e., a disc proposal network (DPN) and a cup proposal network (CPN), for OD and OC region proposal. A disc attention module is employed between the DPN and CPN to determine where the OC is located based on the result of the DPN.

3.2.4. Domain adaptation studies based on GANs

Domain adaptation is another direction that has been explored. Wang et al. (2019c) proposed the patch-based Output Space Adversarial Learning (pOSAL) framework. Images from the source and target domains are first passed into a lightweight network to extract ROIs. The extraction network is based on Deeplabv3+ (Chen et al., 2018) but utilizes MobileNetV2 (Sandler et al., 2018) as the backbone. The authors also designed a morphology-aware segmentation loss for the network. Then the ROIs extracted are passed to a patch discriminator (Patch GAN) for adversarial learning, which can learn abstract spatial and shape features from the label distribution of the source domain. In their following work, Wang et al. (2019b) proposed an unsupervised domain adaptation network named Boundary and Entropy-Driven Adversarial Learning (BEAL). Based on the observation that predictions in the target domain made by

the network trained on the source domain tend to contain ambiguous and inaccurate boundaries and the corresponding entropy map is noisy with high-entropy outputs, they introduced two segmentation branches focused on the boundary and entropy map, respectively. An adversarial learning method was introduced to encourage predictions of the two branches to be domain-invariant. Liu et al. (2019d) proposed an unsupervised domain adaptation architecture named Collaborative Feature Ensembling Adaptation (CFEA). Their framework consists of three parts; namely, the source domain network (SN), which learns from the source domain with labels, the target domain student network (TSN) and the target domain teacher network (TTN), which learn from target domains without labels. Adversarial learning was introduced between SN and TSN, where the supervised SN enables the segmentation network obtain more precise prediction, and the unsupervised TSN introduces a perturbation to the training of the network. The MSE between TSN and TTN was calculated to help the student network's learning.

3.2.5. Other approaches

There are serveral other approaches for addressing vessel and OD segmentation using a variational autoencoder (VAE) as the backbone, tackling a novel task named optic disc quantification and using a single net.

Approaches based on a variational autoencoder. Note that some works were built on VAE. Sedai et al. (2017a) proposed a semi-supervised method to perform OC segmentation using limited labeled training samples. First, a generative variational autoencoder (GVAE) is trained using a large amout of unlabeled data to learn the feature embedding. Then a segmentation variational autoencoder (SVAE) is used to predict the OC mask by leveraging the feature embedding provided by the GVAE.

OD quantification based on multitask ensemble learning framework. Unlike traditional OD/OC segmentation, optic disk quantification refers to the simultaneous quantification of a series of medical indicators, namely two vertical diameters (OD, OC), two complete regions (Disc, Rim), and 16 local regions (Garway-Heath and Hitchings, 1998). Accurate optic disc quantification provides effective help in the diagnosis and treatment of many eye diseases such as chronic glaucoma (Maninis et al., 2016). Zhao et al. (2019c) proposed a multitask ensemble learning framework (DMTFs) to perform optic disc quantification. To the best of our knowledge, they were the first to use deep learning to accomplish this task. In their following work (Zhao and Li, 2020), they made several modifications to their original model, including incorporating a conduct feature interaction module for highly correlated tasks.

Vessel and OD segmentation using a single net. Maninis et al. (2016) proposed Deep Retinal Image Understanding (DRIU) for vessel and OD segmentation. They used VGG-16 as "base network" and removed its FC layers. The feature maps of different levels from the base network are resized and fused to form two task-specific "specialized" layers, which are used to perform vessel segmentation and OD segmentation simultaneously.

3.2.6. Discussion

Compared to other segmentation tasks like vessel or lesion segmentation, the segmentation of the OD/OC is more similar to natural image segmentation due to its ellipse shape. Therefore, several architectures taken from natural image segmentation are used in this task, including Deeplabv3+ and Mask-RCNN. Such networks are rarely seen in vessel segmentation or lesion segmentation. Further, compared to the other two segmentation tasks, the research on OD/OC segmentation is the most complete. Architectures of FCN, U-Net, Deeplabv3+ and Mask-RCNN have been used, and methods like multiscale and polar transformations have been tried.

Future work on OD/OC segmentation may lie in the following directions. First, the more accurate task of OD quantification may be promising. Second, the problems of domain shift and poor generation performance still exist. Therefore, domain adaption should be paid more attention to. It is also worth noting that the REFUGE2 challenge is ongoing. Researchers can follow the competition to see the latest methods and research directions.

3.3. Fovea segmentation

The fovea is one of the most significant landmarks in fundus images. Segmenting the fovea can help define the risk of a particular lesion in retinal diagnosis. However, due to the lack of publicly-available datasets, few works focus on fovea segmentation.

Sedai et al. (2017b) proposed a two-stage framework for fovea segmentation, using a subset of EyePACS as training data. The first stage is a coarse network, which performs coarse segmentation to localize the fovea region. The authors discarded the FC layers of VGG-16 to make it a fully convolutional network. Feature maps at different levels are upsampled to the same size as the input images and fused together to get the output. The second stage is a fine network, which takes the ROI regions obtained by the coarse network to generate the final result. The only difference between the fine network and coarse network is that the fine network only uses the last two blocks of VGG-16 to get the segmentation output.

It is clear that only one remarkable work has been introduced for fovea segmentation. More researches are thus called for to obtain better performance on fovea segmentation. Moreover, the framework used by Sedai et al. (2017b) is a two-stage architecture. More architectures need to be explored.

3.4. A/V classification

Subdividing the blood vessels in fundus images into arteries and veins is of vital importance for the early diagnosis of many diseases. For example, a low ratio between arteriolar and venular width (AVR) can predict diabetes and many cardiovascular diseases (Niemeijer et al., 2011). The widely used datasets for this task are as follows. Compared to the DRIVE dataset used in vessel segmentation, the DRIVE-AV dataset (Hu et al., 2013) further provides pixel-level artery/vein labels. As mentioned before, it has 20 images in the training set and 20 images in the test set. DRIVE-AV is also called RITE in

some papers. The LES-AV dataset[7] (Orlando et al., 2018) consists of 22 images with pixel-level labels. The INSIPRE-AVR dataset[8] (Dashtbozorg et al., 2014) consists of 40 images with only centerline-level annotation. The private IOSTAR dataset (Abbasi-Sureshjani et al., 2015) consists of 24 images annotated by two experts.

Galdran et al. (2019) regarded the A/V classification task as a four-class segmentation problem, with categories including background, artery, vein and uncertain. They used a U-Net-like structure to classify the arteries and veins directly without segmenting the vessel tree first. To the best of our knowledge, they are the first to focus on pixel-level uncertainty in the task of vascular segmentation and classification. Raj et al. (2020) proposed an Artery-Vein Net (AV-Net) for A/V classification. The backbone network is ResNet-50 and squeeze-excitation (SE) blocks are used. Feature maps of different scales are upsampled and fused to the same size as the input image to get the segmentation map. AV-Net does need a segmented vasculature map as input, instead only requiring a single wavelength, color fundus image. Finally, as introduced previously, Ma et al. (2019) performed A/V classification while segmenting blood vessels.

Discussion. From the above methods we can see that A/V classification is a promising direction. The general tendency is to directly perform A/V classification without performing vessel segmentation first. However, A/V classification is an even more challenging task than vessel segmentation. Further in existing works, there is still the problem of arteries and veins appearing in a single vessel segment, which is not common in reality.

4. Disease diagnosis/grading

4.1. Diabetic retinopathy

Diabetic retinopathy (DR) is a vascular disease that affects normal blood vessels in the eye and is the leading cause of preventable blindness worldwide (Wilkinson et al., 2003). There is a unified standard for DR classification, namely the International Clinical Diabetic Retinopathy Scale (ICDRS). According to this standard, the severity of DR can be graded into five classes, namely 0 (no apparent DR), 1 (mild DR), 2 (moderate DR), 3 (severe DR), 4 (poliferative DR). The most commonly used datasets are shown in Tab. 1. All of them have been introduced in Section 2. The experimental results are shown in Tab. 17.

4.1.1. Clinical style papers

There are several clinical style papers, usually found in clinical journals such as JAMA and Diabetic Care. These papers usually pay more attention to actual clinical meaning rather than network architecture improvements. Most of the training datasets were collected by the authors, rather than using public datasets.

David et al. (2016) proposed a system that can automatically detect DR, called IDx-DR X2.1. It applies a set of CNN-based detectors for each image in the detection. Their CNN structure is inspired by AlexNet and VGG and is able to predict four labels, namely negative (no or mild DR), referable DR (rDR), vision-threatening DR (vtDR), and low exam quality (protocol errors or low-quality images). CNN-based anatomy detectors can further detect hemorrhages, exudates, and other lesions. Gargeya and Leng (2017) also used a CNN to perform DR binary classification. They used a CNN containing five residual blocks to extract image features. The features extracted by the deep CNN and metadata information were fed into a decision tree model for binary classification. Li et al. (2018b) used a deep learning algorithm (DLA) for the detection of referable DR. For the training and validation set, they collected 71,043 images from a website named LabelMe and invited 27 ophthalmologists to annotate them. They used four Inception-v3 networks for different tasks, namely 1) classification of vision-threatening referable DR, 2) classification of DME, 3) evaluation of image quality for DR, and 4) assessment of image quality and of the availability of the macular region of DME.

Ensemble strategies are commonly used in this area. Gulshan et al. (2016) used a CNN for binary classification of with/without DR. They used a dataset of 128,175 images, which were annotated three to seven times by 54 experts. The specific network uses the structure of Inception-v3, and an ensemble of ten networks trained with the same data. The final result is the average of all network outputs. Krause et al. (2017) used a CNN for the five-class classification of DR. Their improvements over Gulshan et al. (2016) include: using Inception-v4 instead of Inception-v3, using a larger dataset during training, and using higher-resolution input images. Their network structure is also an ensemble of ten networks. Zhang et al. (2019b) established a high-quality labeled dataset, and adopted an ensemble strategy to perform two-class and four-class classifications. Features extracted from different CNN models are passed through the corresponding SDNN modules, which are defined as component classifiers. Then, the features are fused and fed into a FC layer to generate the final results.

4.1.2. Approaches combining lesion detection

Considering the internal correlation between the diagnosis of DR and detection of hemorrhages, exudates and other lesions, many works also generate heatmaps of lesions while performing DR diagnostic grading. These methods consist of: generating lesion heatmaps, lesion segmentation, attention method and two benchmark works.

Generating lesion heatmaps. Yang et al. (2017) proposed a two-stage DCNN that can simultaneously delineate the lesions and perform DR severity grading. The first stage is a local network, which extracts local features for lesion detection. The second stage is a global network for the grading of DR. A weighted lesion map is obtained from the local network and the original fundus images. An imbalanced weighting scheme was introduced to pay more attention to lesion patches while performing DR grading. Gondal et al. (2017) adopted unsupervised learning to perform DR grading and generate lesion

[7] https://ignaciorlando.github.io/#publications_selected
[8] http://www.retinacheck.org/datasets

Table 17. Summary of several results for DR diagnosis/grading

Reference	Dataset	Category	Backbone	Loss	SE/%	SP/%	AUC/%	Kappa/%
David et al. (2016)	Messidor-2	4	CNN	-	**96.8**	**87.0**	**98.0**	-
Gulshan et al. (2016)	Messidor-2	2	Inception-v3	-	87.0	**98.5**	**99.0**	-
Gargeya and Leng (2017)	Messidor-2	2	CNN	2-class categorical CE	**93**	87	94	-
Wang et al. (2017)	Messidor	5	CNN	-	-	-	95.7	-
Lin et al. (2018)	Messidor	5	CNN	-	-	-	**96.8**	-
Gulshan et al. (2016)	EyePACS	2	Inception-v3	-	90.3	**98.1**	**99.1**	-
Gargeya and Leng (2017)	EyePACS	2	CNN	2-class categorical CE	**94**	98	97	-
Gargeya and Leng (2017)	E-Ophtha	2	CNN	2-class categorical CE	**90**	**94**	**95**	-
Quellec et al. (2017)	E-Ophtha	2	CNN	-	-	-	94.9	-
Wang et al. (2017)	Kaggle	5	CNN	-	-	-	85.4	-
Lin et al. (2018)	Kaggle	5	CNN	-	-	-	-	85.9
Roy et al. (2017)	Kaggle	5	CNN	-	-	-	-	**86**
Yang et al. (2017)	Kaggle	4	CNN	-	-	-	**95.90**	-
Quellec et al. (2017)	Kaggle	2	CNN	-	-	-	**95.5**	-
Gondal et al. (2017)	DiaretDB1	2	CNN	-	**93.6**	**97.6**	**95.4**	-
Foo et al. (2020)	SiDRP14-15	5(No DR here)	U-Net, VGG16	binary CE	-	-	**78.56**	-
Foo et al. (2020)	IDRiD	5(No DR here)	U-Net, VGG16	binary CE	-	-	**99.00**	-
Lin et al. (2018)	private	5	CNN	-	-	-	-	**87.5**
Krause et al. (2017)	private	5 (moderate or worse DR here)	Inception-v4	-	**97.1**	**92.3**	**98.6**	**84**
Li et al. (2018b)	private	2	Inception-v3	-	**92.5**	**98.5**	**95.5**	-
Zhang et al. (2019b)	private	2	CNN	CE	**97.5**	**97.7**	**97.7**	-
Zhang et al. (2019b)	private	4	CNN	CE	**98.1**	**98.9**	-	-
Gulshan et al. (2019)	hospital in Sankara	2	CNN	-	**92.1**	**95.2**	**98.0**	-
Gulshan et al. (2019)	hospitals in Aravind	2	CNN	-	**88.9**	**92.2**	**96.3**	-

heatmaps using only image-level labels. Their main network uses the o_O solution (Mathis Antony, 2015), replacing the last dense layer with a global average pooling (GAP) layer. Their way of generating heatmaps was mainly inspired by Zhou et al. (2016). Quellec et al. (2017) proposed a solution to generate a heatmap that shows what roles the pixels in an image play in image-level prediction. They can detect both image-level referable DR and pixel-level biomarkers. Their network's baseline is also the o_O solution, and a method called backward-forward propagation was proposed to optimize the parameters.

Performing lesion segmentation at the same time. Foo et al. (2020) used an encoder-decoder network for DR grading and lesion segmentation. They replaced the encoder of U-Net with VGG-16, which has five groups of conv layers. Correspondingly, the decoder is modified as a mirror of the encoder. This architecture can perform lesion segmentation naturally. Then, for DR grading, they attached a GAP layer to the saddle layer of the network for classification. They further proposed a semi-supervised approach to increase the number of training images.

Attention methods. Attention mechanisms are also commonly used in DR diagnosis and grading. Wang et al. (2017) proposed a Zoom-in-Net that can simultaneously perform five-class DR grading and generate attention maps highlighting lesions. Zoom-in-Net consists of three parts, namely a main net (M-Net) using Inception-Resnet as the backbone, which aims to extract features and can output diagnostic results; an A-Net, which can generate attention maps using only image-level supervision; and a C-Net, which simulates the zoom-in operation when clinicians examine images. Lin et al. (2018) proposed a framework based on anti-noise detection and attention-based fusion which can perform five-class DR grading. They first extract the features using a CNN, then feed them into a designed center-sample detector to generate lesion maps. Lesion maps and original images are sent to the proposed attention fusion network (AFN), which can learn the weights of the original images and lesion maps to reduce the influence of unnecessary information.

Benchmark works. Although, according to priori knowledge, detecting related lesions is helpful for the diagnosis/grading of DR, lesion detection is actually a complex and difficult task, and there exists a trade-off between lesion detection and DR grading.

Li et al. (2019b) built a dataset called DDR. DDR is the only dataset considering both DR and lesion detection; it is the largest dataset for lesion detection and second largest for DR grading. The authors evaluated ten state-of-the-art deep learning models on this dataset, including five classification models, two segmentation models, and three detection models. Although these methods achieved a maximum acc of 0.8284 in DR grading, their performance in lesion segmentation and detection was particularly poor, indicating that the detection or segmentation of lesions is a very challenging task. Ahmad et al. (2019) performed a benchmark work on Messidor-2. They evaluated eight state-of-the-art deep learning classification models and generated class activation maps (CAMs) of lesions at the same time. The results showed that there is a trade-off between classification and localization. As the networks' depth and parameters increased, they performed better in classification, while performing worse in localization.

4.1.3. Other approaches

There are several other approaches for DR grading, including a bi-linear strategy, a hybrid method and the IDRiD challenge.

Bi-linear strategy with attention mechanism. Zhao et al. (2019e) proposed a BiRA-Net to perform DR grading. In the introduced RA-Net, features extracted from ResNet were fed to a proposed attention net to pay more attention to the decisive areas for grading. A bilinear strategy was adopted to train two RA-Nets for more fine-grained classification.

Hybrid method combined with manually designed features. Roy et al. (2017) proposed a strategy that combines CNN and dictionary-based strategies for DR severity assessment. The activation value of the second fully connected layer (FC2) of the CNN was converted to a discriminative pathology histogram (DPH) and generative pathology histogram (GPH), which consist of manually designed features with specific concerns. The two histogram feature vectors and the original-size image were fused with the CNN's FC2 response, and finally a decision tree classifier was used to obtain the final result.

Smartphone-based diagnosis. Natarajan et al. (2019) proposed an offline DR screening system on a smartphone to detect referable DR. Users can download the app and get DR diagnosis results instantly. It is unbelievable to imagine that the diagnosis of DR can be performed by such a low-cost device in such a convenient way. Such an offline system is of great significance to areas with limited medical resources.

IDRiD challenge. Porwal et al. (2020) described the IDRiD dataset and outlined the setup of the challenge "Diabetic Retinopathy Segmentation and Grading" held at ISBI2018. They also discussed a variety of deep learning models that were outstanding in the competition, as well as lessons learned from analyzing of the submissions.

4.1.4. Discussion

The diagnosis/grading of DR has been widely studied. There are several clinical style papers in this field. In these works, a large number of images were typically collected and labeled and the significance of using deep learning in actual clinical diagnosis was assessed. From a technical point of view, the diagnosis/grading of DR is a classification task. All we need to do is to predict a number indicating the stage of DR. However, only providing a single number may confuse clinicians. They also need to know why the network makes certain decisions, and what are deemed decisive regions. Therefore, many works have focused on generating heatmaps or performing lesion segmentation at the same time. Other effective methods like attention mechanisms and hybrid methods have also been explored. It is also worth noting that a remarkable smartphone-based offline diagnosis system has been created.

However, the existing researches still face several shortcomings. Rather than fine segmentation, the heatmap is typically generated coarsely and cannot provide lesion labels. Therefore, performing lesion segmentation and DR diagnosis/grading at

the same time is a promising direction. However, as discussed previously, there is a trade-off between DR diagnosis/grading and lesion segmentation. This is mainly because the high-level semantic features needed for the classification task tend to lack the spatial information required for segmentation. Reaching this trade-off is important for this multi-task problem.

4.2. Glaucoma

Glaucoma is one of the major causes of blindness worldwide. The number of glaucoma infections is expected to grow to 112 million by 2040. Because of its irreversibility, early screening of glaucoma is extremely important (Bourne et al., 2013; Tham et al., 2014). The datasets used in this field are shown in Tab. 11. The SINDI, REFUGE and SEED datasets were not introduced in Section 3.2. The SINDI dataset (Fu et al., 2018b) was established to assess the risk factors of visual impairment in the Singapore-Indian community. It consists of 5,783 images, of which 5,670 are normal and 113 are glaucomatous. The REFUGE dataset (Orlando et al., 2020) was used in the REFUGE challenge. It contains 1,200 fundus images with segmentation ground truth and clinical glaucoma labels. The SEED study (Zheng et al., 2013) was conducted in southwestern Singapore between 2004 and 2011. The population included 3,353 Chinese, 3,280 Malays and 3,400 Indian adults aged 40 and older. Experimental results are shown in Tab. 18.

4.2.1. Clinical style papers

Similar to DR, there are several clinical applications of glaucoma diagnosis. Raghavendra et al. (2018) utilized an 18-layer CNN containing five conv layers to perform glaucoma classification. They obtained 1,426 images from Kasturba Medical College, Manipal, India. Liu et al. (2019c) proposed a deep learning system (DLS) named Glaucoma Diagnosis with Convoluted Neural Networks (GD-CNN), based on ResNet. They established a dataset named FIGD consisting of 241,032 images. They further proposed an online deep learning (ODL) system to improve the generalization ability of GD-CNN.

Glaucomatous Optic Neuropathy (GON) diagnosis is also widely studied. Li et al. (2018a) used deep learning to perform binary classification of GON. They downloaded 70,000 fundus images from the online dataset LabelMe[9], and selected 48,116 images for annotation. They invited 27 qualified ophthalmologists for labeling and used Inception-v3 as the classification network. Phene et al. (2019) collected 86,618 images from several sources, including EyePACS, Inoveon[10], AREDS, UK Biobank[11] and three hospitals in India. They invited 43 graders to perform image-level and feature-level labeling. They also used Inception-v3 and trained an ensemble of 10 networks. Their network can predict referable GON and the presence/absence of various ONH features at the same time.

[9]http://www.labelme.org/

[10]http://www.inoveon.com/

[11]https://www.ukbiobank.ac.uk/aboutbiobankuk

4.2.2. Approaches considering OD/OC area

Based on the priori knowledge that the OD and OC area can be helpful to diagnose glaucoma, many methods have paid attention to these two areas. Applications can be subdivided into OD/OC segmentation and direct CDR estimation.

OD/OC segmentation. dos Santos Ferreira et al. (2018) designed a texture descriptor with a CNN to diagnose glaucoma. They first used a U-Net to segment the OD area, then inspired by the domain knowledge of biology, designed a module called phylogenetic diversity indexes to extract semantic features, and finally used a CNN-based classifier for the diagnosis of glaucoma. Pal et al. (2018) designed a G-EyeNet for the classification of glaucoma, which performs particularly well when the dataset is small. They first perform OD segmentation, then the extracted ROIs are fed to a U-Net-like architecture for image reconstruction. Finally, an FC layer followed by a softmax classifier were incorporated into to the encoder for glaucoma classification.

Direct CDR estimation. Zhao et al. (2020b) abandoned the intermediate step of segmenting the OD and OC area and decided to directly estimate CDR from fundus images. In the proposed MFPPNet, fundus images are passed through three DenseBlocks, and then the extracted features go through a feature pyramid pooling module and a fully connected feature fusion module to learn and fuse multiscale features. Finally, random forest regression is used to perform CDR regression.

4.2.3. Approaches built on multi-branched methods

Multi-branched methods are also widely explored for glaucoma diagnosis. The results of multiple networks with different focuses are fused together to achieve higher accuracy. Fu et al. (2018b) proposed the Disc-aware Ensemble Network (DENet) which contains four branches. The global image stream learns the image-level global features, employing a ResNet-50 as the backbone and using the original images as input. The second stream is a segmentation-guided network using a U-Net to segment the OD area as guidance for the other two branches. FC layers are connected to the saddle layer of U-Net to output classification results. The local disc region stream and the disc polar transformation stream both take ResNet-50 as the classifier, with the former taking the disc region crop as input and the other take polar transformed version. Chai et al. (2018) designed a multi-branch neural network (MB-NN) combining domain knowledge. MB-NN takes three branches as input. The first is a set of original images. The second branch is the optical disc region generated by Faster-RCNN. The third branch contains domain knowledge features, which include image features such as CDR and PPA size, and non-image features such as age, intraocular pressure and eye sight.

4.2.4. Generating evidence maps

There are also some inspiring studies that generate evidence maps when performing glaucoma diagnosis. The approaches include a weakly supervised method, using a LAG dataset containing evidence label and a multiscale method.

Weakly supervised method. Zhao et al. (2019d) proposed a weakly-supervised multi-task Learning method (WSMTL) to

Table 18. Summary of several results for glaucoma diagnosis/grading

Reference	Dataset	Backbone	Loss	SE/%	SP/%	ACC/%	BACC/%	AUC/%
Li et al. (2019a)/Li et al. (2020b)	RIM-ONE	CNN	K-L divergence function and CE	84.8	85.5	85.2	-	91.6
dos Santos Ferreira et al. (2018)	RIM-ONE, DRISHTI-GS	U-Net, CNN	-	**100**	**100**	**100**	-	**100**
Zhao et al. (2019d)	ORIGA	CNN	contrastive loss and hinge loss	-	-	-	-	**92**
Liao et al. (2020)	ORIGA	CNN	-	-	-	-	-	88
Li et al. (2019a)	LAG	CNN	K-L divergence function and CE	**95.4**	95.2	95.3	-	97.5
Li et al. (2020b)	LAG	CNN	K-L divergence function and CE	**95.4**	**96.7**	**96.2**	-	**98.3**
Pal et al. (2018)	DRIONS-DB	Encoder-decoder network	Reconstruction loss and CE	-	-	-	-	**92.3**
Fu et al. (2018b)	SCES	U-Net, ResNet50	Dice coefficient loss and CE	**84.78**	**83.80**	-	**84.29**	**91.83**
Fu et al. (2018b)	SINDI	U-Net, ResNet50	Dice coefficient loss and CE	**78.76**	**71.15**	-	**74.95**	**81.73**
Raghavendra et al. (2018)	Private	CNN	-	**98.00**	**98.30**	**98.13**	-	-
Li et al. (2018a)	Private	Inception-v3	-	**95.6**	**92.0**	-	-	**98.6**
Phene et al. (2019)	Private	Inception-v3	-	-	-	-	-	**94.5**
Chai et al. (2018)	Private	FCN, CNN, Faster-RCNN	CE	**92.33**	**90.90**	**91.51**	-	-
Liu et al. (2019c)	Private FIGD	ResNet	CE	**96.2**	**97.7**	-	-	**99.6**

perform accurate evidence identification, optic disc segmentation and automated glaucoma diagnosis simultaneously. First a skip and densely connected CNN is used to capture multiscale features. Then, the extracted features are fed to the proposed pyramid integration structure to generate high-resolution evidence maps. These evidence maps are passed to a constrained clustering branch which clusters pixels with relational constraints. The evidence maps are also fed to a fully-connected discriminator to diagnose glaucoma.

Using dataset containing evidence map label. Li et al. (2019a) established a large-scale attention-based glaucoma (LAG) dataset. LAG contains 5,824 fundus images and attention maps provided by ophthalmologists. They also proposed an AG-CNN to diagnose glaucoma. First, an attention prediction subnet was introduced to generate attention maps. In this subnet, multiscale and channel attention methods are utilized. Then, a pathological area localization subnet was designed to locate the pathological area, in which attention maps are embedded to feature maps at each stage. Finally, the located pathological areas and predicted attention maps are concatenated together and fed to a glaucoma classification subnet to predict the binary label of glaucoma. In their subsequent work (Li et al., 2020b), they extended their LAG dataset to 11,760 fundus images. They also proposed a weakly supervised learning strategy for AG-CNN.

Multiscale networks. Liao et al. (2020) introduced a clinically interpretable ConvNet architecture (EAMNet) for glaucoma diagnosis. They first used a CNN as backbone network with several residual blocks to extract useful features. Then a method named Multi-Layers Average Pooling (M-LAP) was proposed to bridge the gap between low-level localization information and high-level semantic information. Moreover, evidence activation maps (EAMs) were obtained by weighted summation of feature maps.

4.2.5. Discussion

Like in DR grading, there are several glaucoma diagnosis papers that care more about clinical applications, as discussed in Section 4.2.1. Further, Sections 4.2.2, 4.2.3, and 4.2.4 can all be regarded as focusing on the OC area from different aspects. Section 4.2.2 describes methods that perform OD/OC segmentation or CDR estimation and glaucoma diagnosis simultaneously. In Section 4.2.3, OD/OC segmentation serves as a branch to guide the glaucoma diagnosis task. Finally, in Section 4.2.4, heatmaps are generated to highlight decisive regions for glaucoma diagnosis.

However, there is still room for improvement in the diagnosis of glaucoma. First, just like in DR diagnosis, while heatmaps can be used to provide guidance for diagnosis, the more accurate task of OD/OC segmentation should be emphasized. Second, the diagnosis of glaucoma does not only lie in CDR estimation; there are several other factors that can affect the result, such as age, race and family history. However, few works focus on these factors.

4.3. AMD

Age-related macular degeneration (AMD) is the leading cause of vision loss among people aged 50 and above. 6.2 million people worldwide suffered from AMD in 2015 (Vos et al., 2016). The datasets used in this field are shown in Tab. 19. AREDS is widely used in AMD diagnosis. The AREDS dataset (Group, 1999) consists of over 206,500 images acquired from 5,208 participants. iChallenge-AMD was used as the dataset of the iChallenge competition. It consists of 1,200 images, of which 77% are from non-AMD subjects and 23% are from AMD patients. Labels for AMD/non-AMD, disc boundaries and fovea locations and lesion boundaries are provided. The KORA dataset (Brandl et al., 2016) was acquired from 2,840 individuals aged 25 to 74 years old from South Germany. Experimental results are shown in Tab. 20.

4.3.1. Methods based on a hybrid architecture

Burlina et al. (2016) are one of the very first to use deep learning for AMD diagnosis. They used a pre-trained OverFeat DCNN to map original images into a 4,096-dimensional feature vector. Then the vectors were passed through a linear SVM classifier to output accurate AMD binary classification results, namely disease-free/early stages and referable intermediate/advanced stages. In their following work (Burlina et al., 2017), they expanded the previous method by using datasets about 10 to 20 times larger. Horta et al. (2017) made several modifications to Burlina et al. (2016). They added some side channel features such as sunlight, education and gender. Further, they used two inscribed rectangles of fundus images at different scales as input, and thus obtained 8,192-dimensional feature vectors. However, the dimension of the side channel features was much smaller than 8,192 dimensions. In order to alleviate this imbalance, they used PCA for dimension reduction. These features were then fused together to train a random forest classifier for final AMD classification.

4.3.2. Approaches based on CNNs

Govindaiah et al. (2018) evaluated the performance of deep learning networks in two-class (no or early AMD and intermediate or advanced AMD) and four-class (no AMD, early AMD, intermediate AMD and advanced AMD) classification for AMD. The networks evaluated include VGG-16 with transfer learning, VGG-16 without transfer learning, and ResNet-50. The experimental results showed that, whether on two-class or four-class classifications, VGG-16 without transfer learning performs best. Tan et al. (2018) designed a 14-layer CNN for early AMD diagnosis. Their data was obtained from the Ophthalmology Department of Kasturba Medical College (KMC), included 402 normal images, 583 retinal images with early, intermediate AMD, or GA and 125 retinal images with evidence of wet AMD. Burlina et al. (2018) used deep learning for detailed severity characterization and estimation of five-year risk among patients with AMD. The classification network used was ResNet-50. For AMD severity scales, they employed four-step and nine-step scales, respectively. For the estimation of five-year risk of progression to advanced AMD, they evaluated three deep learning-based strategies, namely Soft Prediction, Hard Prediction and Regressed Prediction.

Ensemble strategies have also been used in AMD diagnosis. Grassmann et al. (2018) used an ensemble network for 13-class AMD classification. The pre-processed images were independently trained using six different CNNs (AlexNet, VGG-16, GoogLeNet, Inception-v3, ResNet, Inception-ResNet-v2). The results of the six networks were fused using random forests.

Guidance from lesion detection is helpful to AMD diagnosis. Peng et al. (2018) proposed a DeepSeeNet to grade the severity of AMD (0-5). Their network consists of three parts, Drusen Net (D-Net) for detecting drusen in 3 sizes (none/small, medium and large), Pigment-Net (P-Net) for detecting pigment abnormalities (hypopigmentary or hyperpigmentary) and Late AMD-Net (LA-Net) to detect the presence of late AMD (neovascular AMD or central GA). The three subnetworks all use an Inception-v3 structure.

4.3.3. Discussion

In this section, we have discussed approaches for AMD diagnosis. All of them are based on a CNN architecture. Hybrid methods, ensemble strategies and guidance from lesion detection have also been explored. However, there are still several limitations. First, the attention paid to AMD does not match its prevalence and severity. There is much less research on AMD diagnosis than on DR and glaucoma. Second, the datasets and the number of images used for AMD diagnosis are far fewer than those for DR and glaucoma. Finally, the actual amount of data in the website is inconsistent with what is claimed in the original paper.

4.4. DME

Diabetic macular edema (DME) is the most common complication of DR and may cause severe vision loss (Ciulla et al., 2003). Two approaches addressing this task use a two-stage architecture and multiscale method respectively.

Two-stage architecture. Mo et al. (2018) proposed the cascaded deep residual networks for DME diagnosis. The datasets used include e-ophtha EX and Hamilton Eye Institute Macular Edema Dataset (HEI-MED). The HEI-MED dataset[12] (Giancardo et al., 2012) consists of 169 images, of which 115 are healthy and 54 contain exudates. Their framework consists of two stages. The first stage is an exudate segmentation network adopting a deep fully convolutional residual network (FCRN). Then a fixed-size region is cropped, where the center pixel has the maximal probability value. The cropped region is fed to the second stage, which is a deep residual network performing binary classification.

Multiscale networks. He et al. (2019) proposed a DME-Net based on the multiscale method for DME classification. They used IDRiD and MESSIDOR as their datasets. They first passed fundus images through a U-Net to generate fovea and hard exudate region masks. Then a multiscale feature extraction module using VGG-16 as the backbone was designed, in which a GAP

[12] http://www.vibot.u-bourgogne.fr/luca/heimed.php

Table 19. Widely used datasets for AMD diagnosis/grading

Dataset name	Number of images	Resolution	Camera	Availability
AREDS	Over 206,500 images	-	-	available online[1]
iChallenge-AMD	1200	-	-	available on registration[2]
KORA	images from 2840 individuals	-	-	available online[3]

[1]https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1
[2]http://ai.baidu.com/broad/introduction?dataset=amd
[3]https://epi.helmholtz-muenchen.de/

Table 20. Summary of several results for AMD diagnosis/grading

Reference	Dataset	Backbone	Loss	Category	SE/%	SP/%	ACC/%	AUC/%	Kappa/%
Burlina et al. (2016)	AREDS	CNN with SVM	-	2(1vs.3,4)	**93.4**	**95.6**	**95.0**	-	-
Burlina et al. (2017)	AREDS	CNN with SVM	-	2	-	-	88.4~91.6	**94~96**	-
Horta et al. (2017)	AREDS	CNN with RF	-	2	66.34	88.95	79.04	84.76	-
Govindaiah et al. (2018)	AREDS	CNN	-	2	-	-	92.5	-	-
Govindaiah et al. (2018)	AREDS	CNN	-		-	-	**83**	-	-
Burlina et al. (2018)	AREDS	ResNet50	Regression loss	4	-	-	75.7	-	-
Peng et al. (2018)	AREDS	Inception-v3	-	6	**59.0**	**93.0**	**67.1**	-	**55.8**
Burlina et al. (2018)	AREDS	ResNet50	Regression loss	9	-	-	**59.1**	-	-
Grassmann et al. (2018)	AREDS, KORA	CNN	weighted k metric	13	-	-	**63.3**	-	-
Tan et al. (2018)	Collected	CNN	-	2	**96.43**	**93.75**	**95.45**	-	-

operation was applied to feature maps of each stage. Then the features were concatenated to obtain multiscale features. They passed the original fundus images, obtained fovea and hard exudate region masks, and the macular region cropped from fundus images through the proposed multiscale feature extraction modules respectively. The features were fused together and fed to a XGBoost classifier (Chen and Guestrin, 2016) to output the final results.

Discussion. The two approaches introduced above both detect exudates as guidance for the DME diagnosis task. Therefore, the architectures used are both multi-stage. Possible future work for this task may be to design a more lightweight architecture and decrease the parameters to be trained. Another promising direction is to detect DME and DR at the same time. Such works can be seen in Section 4.7.

4.5. ROP

Retinopathy of prematurity (ROP) is an eye disease that often occurs in infants with low birth weight or premature birth. It is the main cause of childhood blindness (Tasman et al., 2006). Brown et al. (2018) used deep CNNs to diagnose plus disease in ROP. The plus disease is defined as arterial tortuosity and venous dilation of the posterior retinal vessels that is greater than or equal to that found in a standard published retinal photograph (for Retinopathy of Prematurity Cooperative Group, 1988). The presence of plus disease is the most critical feature of severe, treatment-requiring ROP. Their training dataset contains 5,511 images, and formed part of the multicenter Imaging and Informatics in Retinopathy of Prematurity (i-ROP) cohort study. They first used a U-Net for preprocessing, then an Inception-v1 architecture was employed to diagnose plus disease. Experimental results showed that the deep CNNs outperformed six of eight ROP experts invited. Taylor et al. (2019) used deep learning to objectively monitor ROP progression. Their data was also from the i-ROP study. In their work, a quantitative ROP vascular severity score was developed using previous work (Brown et al., 2018). Tracking the quantitative severity score may be an effective method for identifying patients at risk of disease progression.

Three-stage architectures have been shown to be suitable frameworks for ROP diagnosis. Hu et al. (2019) used deep learning to classify ROP. To solve the problem of insufficient labeled data, they collected 2,668 examinations obtained from 720 infants. Each examination consists of several fundus images from different views. Because only a few of the n images in one ROP examination may contain features that can diagnose ROP, it is necessary to extract features from all the images in one ROP examination and fuse these features. To this end, they designed a three-stage network which is divided into three parts: feature extraction, feature fusion and classification. In the feature extraction stage, there are a total of n CNN networks with identical structures, which are used to extract features from the n images in one examination. In the feature fusion stage, they considered two ways of max and mean, to fuse the n h×w×c features from the feature extraction stage into one h×w×c feature map. The final classification stage has a convolutional layer and a GAP layer for the final binary classification of ROP. For the network structure, they considered VGG-16, Inception-v2 and ResNet-50, and also experimented with different image resolutions.

Discussion. From Hu et al. (2019)'s work we can see that only a few images of a ROP examination contain useful features. This situation is similar to MRI and CT image processing. Researchers can thus borrow some inspiration from these tasks. However, publicly available datasets are still limited in number. More finely-annotated datasets are called for.

4.6. Cataracts

Cataracts can cause severe vision loss and are one of the most serious eye diseases that can cause blindness. Zhou et al. (2020) used deep learning for cataract diagnosis (non-cataract/cataract) and cataract grading (non-cataract, mild cataract, moderate cataract and severe cataract). They first established a dataset containing 1,335 images from Beijing Tongren hospital, China. 433 images of the dataset are non-cataract and 922 images have cataracts. They proposed discrete state transition (DST) and empirical DST (EDST) strategies. In the DST strategy, the weights and activation values were restricted in a unified discrete space, while in EDST they were restricted in an exponential discrete space. DST and EDST can reduce the networks' energy consumption and prevent overfitting. When using priori knowledge, they extracted the improved Haar wavelet features and visible structure features from fundus image as the input of DST-MLP and EDST-MLP. When not using priori knowledge, they used DST-ResNet and EDST-ResNet as classification networks. Xu et al. (2020) introduced a hybrid global-local representation CNN for cataract grading. They established a dataset consisting of 8,030 images, which were manually annotated by ophthalmologists. They first used AlexNet to learn global features and then used a deconvolutional network (DN) in each CNN layer to analyze which pixel contributes most to the classification, and explain the misclassification cases. Then a hybrid model which is an ensemble of several AlexNets, was employed combining global and local features.

Discussion. Cataract diagnosis using fundus images is a promising direction. However, there are several limitations remaining. First, there are no publicly available datasets. Second, there is no unified grading standard like in DR diagnosis. There limitations make it difficult to compare different works.

4.7. Diagnosis of multiple diseases

The diagnoses of different eye diseases may affect each other. For example, for a patient who has both glaucoma and cataracts, it may be difficult to diagnose the glaucoma because of the unclear biomarkers caused by the cataracts. Therefore, the diagnosis of multiple diseases may be a possible solution to this problem. Moreover, the diagnosis of multiple diseases simultaneously is more convenient and helpful to clinicians. The diagnosis of multiple diseases can be divided into simultaneous DR and DME diagnosis, simultaneous DR, glaucoma and AMD diagnosis, the diagnosis of eight diseases using paired CFPs, the diagnosis of 36 diseases and rare pathologies detection.

Simultaneous DR and DME diagnosis. Li et al. (2020c) proposed a cross-disease attention network (CANet), which can simultaneously diagnose DR and DME. CANet contains two different types of attention modules. Disease-specific attention module is used to selectively learn useful features for diagnosing the specific disease. Disease-dependent attention module can further learn the internal relationship between the two diseases. Features extracted from ResNet-50 are passed through two disease-specific attention modules and two disease-dependent attention modules successively. They used IDRiD and MESSIDOR as their datasets. The diagnosis of the two diseases can be mutually enhanced. Tu et al. (2020) proposed a multi-task network named feature Separation and Union Network (SUNet) for simultaneous DR and DME grading. Experiments were carried on the IDRiD dataset. They first used ResNet-34 to extract features for all tasks. Then a feature blending block was proposed, which contains a sequence of feature separation and feature union layers. The feature separation layers learn task-specific features, i.e., diagnosis features for the multi-disease diagnosis block (MD-Block) and lesion features for lesion regularize net (LR-Net). The feature union layers are able to learn useful union features for both branches. Finally, the MD-Block is used to predict results of DR and DME grading.

Simultaneous DR, glaucoma, and AMD diagnosis. Ting et al. (2017) used a CNN network to perform referable DR, vision-threatening DR, possible referable glaucoma and referable AMD diagnosis simultaneously. Their dataset contains 494,661 retinal images, which were obtained from the ongoing Singapore National Diabetic Retinopathy Screening Program (SIDRP) (Nguyen et al., 2016). The backbone network used is VGG-Net.

Diagnosis of eight diseases using paired CFPs. Li et al. (2020a) proposed a Dense Correlation Network (DCNet) to diagnose eight diseases using paired color fundus photographs (CTF) from the ODIR dataset[13]. The ODIR dataset consists of 10,000 paired images from 5,000 Chinese patients. Eight kinds of labels denoting the stages of specific diseases are provided for each image. Here, paired refers to images of the left eye and the right eye from the same patient. DCNet consists of a shared CNN feature extractor for paired CFPs, ResNet in this case, a spatial correlation module (SCM) and a final classifier. The SCM is utilized to capture dense correlations between extracted features and fuse relevant ones.

Diagnosis of 36 diseases. Wang et al. (2019d) used multi-task learning to diagnose 36 diseases simultaneously. To achieve this, they collected and relabeled 200,817 images with 36 categories, of which 17,385 images have more than one label. Their proposed network structure is divided into two stages. The first stage has a modified YOLO-v3 (Redmon and Farhadi, 2018) as the main structure, which is used to detect the macula and the OD/OC area. The second level has three branches, namely the general task stream, macular task stream, and optic-disc task stream, which use original images, the macula area, and the OD and OC area as inputs, respectively. The general task stream uses Inception-ResNet-v2 as the backbone network to detect general retinal diseases, fusing features from the other two streams. The macular task stream is used to detect macular diseases. It uses Inception-v3 as the backbone network and fuses features from the general task stream. The optic-disc

[13] https://github.com/nkicsl/OIA-ODIR

task stream also uses Inception-v3 as the backbone network, but it is independent and does not fuse features from the other branches.

Rare pathologies detection based on few-shot learning. Quellec et al. (2020) used few-shot learning to perform rare pathologies detection. They used the OPHDIAT dataset for training. This dataset (Massin et al., 2008) consists of 763,848 images acquired from the Ile-de-France area. DR grading is provided for every image. Moreover, the ophthalmologists also indicated his or her findings in free-form text. The images contain 41 conditions, some of which are rare pathologies. Based on the observation that CNNs trained to detect frequent conditions, such as DR, also cluster many other unrelated conditions in the feature space, the authors trained a CNN classifier and derived several simple probabilistic models from its feature space to detect rare conditions, solving the few-shot learning problem.

Discussion. The diagnosis of multiple diseases is very significance in clinical practice. In this subsection, we first discussed the simultaneous diagnosis of DR and DME and the diagnosis of three different diseases. Then we introduced three works that focus on multiple disease diagnosis for up to 41 classes. The methods used vary but are all inspiring. In conclusion, this developing direction is very promising and deserves more attention, since more comprehensive systems are needed in practice. Therefore, more experiments should be carried on newly built datasets such as ODIR.

5. Image synthesis

As mentioned before, the training datasets for medical imaging often consist of a fewer number of images than in other deep learning tasks. Further, high-quality annotated datasets are often costly to obtain. One possible solution is image synthesis. Image synthesis can increase the number of fundus images, help us to better understand the images and improve model performance.

Synthesis for glaucoma. Deshmukh and Sivaswamy (2019) proposed a deep learning based method to synthesize the ONH region of fundus images. Given the OD, OC, and blood vessel segmentation masks from arbitrary fundus images, their method can generate high-quality images with vessels bending at the edges of the OC, like in real images. The generator contains four U-Nets. Three parallel branches take an OC, OD and vessel mask as input, respectively, and the outputs are jointly passed through another U-Net to generate RGB images. The synthetic and original images along with their corresponding OC, OD and vessel masks compose the input of the discriminator, which employs a five-layer FCN as the backbone. For the datasets, they used Drishti-GS and DRIVE. Diaz-Pinto et al. (2019) used deep convolutional generative adversarial networks (DCGANs) to obtain a fundus image synthesizer and used a semi-supervised method for glaucoma assessment. Their system can not only generate synthetic images, but also provide labels automatically. They collected 14 datasets as the training set, including an unprecedented 86,926 images. Their two systems, DCGAN and SS-DCGAN, have similar structures, and both contain a generator and a discriminator. The difference between them is that the DCGAN only performs image synthesis, while the SS-DCGAN can predict glaucoma by changing the final output layer of the discriminator of DCGAN. Wang et al. (2019e) proposed a pathology-aware visualization strategy for glaucoma classification and a pathology-based GAN (Patho-GAN) for image synthesis. For brevity, the pathology-aware visualization net will not be described here. Different from the usual GAN network, the synthetic images generated by the generator and the original images are passed through the pathology-aware visualization net and the pathological loss is calculated. Specific pathological areas can be enhanced by optimizing the pathological loss. They used LAG as their dataset.

Synthesis for vessel segmentation. Costa et al. (2018) utilized a GAN to perform retinal image synthesis. An adversarial autoencoder is first trained to reconstruct vessel maps and learn a latent space associated with a normal distribution. Then the generated vessel maps are passed to a GAN. The generator of GAN is used to generate synthetic retinal images that can fool the discriminator. Synthetic pairs and the real pairs are then passed to the discriminator. Once trained, a synthetic retinal image can be generated using decoder of the adversarial autoencoder and generator of the GAN, using a normal distribution as input. In their implementation, the vessel annotation-free dataset Messidor-1 and DRIVE dataset were used. Zhao et al. (2018) proposed a Tub-GAN and a Tub-sGAN for retinal and neuronal image synthesis, which work well on small datasets. They aimed to learn a mapping from a tubular structured annotation to a synthetic image. In Tub-GAN, a GAN is employed with a vessel map ground truth and random noise as input. In their Tub-sGAN, style transfer is incorporated using VGG-Net learning style features and content features. In Zhao et al.'s following work (Zhao et al., 2019a), they proposed a R-sGAN to perform image synthesis for further segmentation of unannotated fundus images. Their framework consists of two stages. In the first stage, vessel maps from finely-annotated datasets and retinal images from unannotated images are passed to R-sGAN as input to generate retinal images that have the same style as the unannotated datasets. R-sGAN is a non-linear variant of GRU (Chung et al., 2014). Then, image pairs of vessel maps and generated retinal images are passed to a segmentation network for training. Once trained, the segmentation network can perform segmentation of unannotated images. Both works used the DRIVE, STARE and HRF datasets.

Synthesis for DR. Zhou et al. (2019) proposed a diabetic retinopathy generative adversarial network (DR-GAN). It can generate high-resolution images given arbitrary DR grading and lesion information. They designed a generator conditioned on vessel and lesion masks to generate high-resolution images. They introduced a fine-grained design which aims to learn better and more realistic local details. A multiscale discriminators framework was also employed, containing three identical discriminators. The only difference between the discriminators is the resolution of input images. For the dataset, they used EyePACS, IDRiD and DRIVE.

Synthesis for AMD. Burlina et al. (2019) utilized a GAN to perform image synthesis for AMD. The GAN model was trained using 133,821 fundus images from AREDS as input.

They invited two ophthalmologists to diagnose AMD on real images and synthetic images. The results obtained are similar for real and synthetic images. Moreover, a classification network trained only on the synthetic images showed similar performance to the network trained only on real images.

Smartphone camera image synthesis. It is relatively easy as well as cost-effective to collect fundus images using a smartphone camera (SC). However, the images are often low-quality, have uneven illumination and other problems. V and Sivaswamy (2019) proposed a ResCycleGAN for image synthesis using SC images. Their modifications over CycleGAN (Zhu et al., 2017) are two-fold; namely, they introduced a residual connection and proposed a structure similarity based loss function. The dataset used consists of 540 images acquired using an iPhone6.

Multimodal image reconstruction. Different modalities provide complementary views of the same real-world object. Image reconstruction from one modality to another is a self-supervised task. On the one hand, important general features can be learned in the reconstruction process. On the other hand, images which are obtained invasively in practice can be obtained using non-invasive modal images. Hervella et al. (2018) employed U-Net to perform image reconstruction from retinography to angiography. They used the publicly available Isfahan MISP dataset[14] (Alipour et al., 2012), which contains 59 retinography/angiography pairs.

Image super resolution (ISR). ISR takes low-resolution images as input and outputs super-resolved (SR) images. This is useful to several downstream, such as small or blurred lesion and biomarker detection. Mahapatra et al. (2017) proposed an ISR method based on GANs. A local saliency map is obtained by combining abstraction, element distribution and uniqueness. Then a local saliency loss is calculated and added to the cost function. Entropy filtering is performed to highlight compact regions. They used DRIVE, STARE and CHASE_DB1 as their datasets.

Discussion. Image synthesis based on deep learning is a relatively new task in fundus image processing. Synthetic images can be used to help training, which can improve performance and alleviate overfitting. In terms of architecture, nearly all approaches used GANs. The latest powerful variant, CycleGAN, can also be seen in some applications. Image synthesis will likely be a very popular direction in the near future. It is hard to imagine what kinds of explorations could be done using GANs.

6. Other applications

6.1. Ophthalmic disease diagnosis

There are several works focusing on rare pathologies, such as pathological myopia and refractive error.

Pathological myopia is a common disease that can cause loss of vision. Guo et al. (2020b) introduced a lesion-aware segmentation network (LSN) to perform atrophy and detachment segmentation, which is related to pathological myopia. The architecture is a U-Net-like encoder-decoder network. The authors added a classification branch to the saddle layer to predict the existence of lesions. A feature fusion module is used in the decoder, which is designed as a multiscale network. To further boost the sensitivity to lesion edges, they added a loss function named edge overlap rate (EOR). The training set used was taken from from PALM challenge[15] in ISBI 2019, consisting of 400 images.

Refractive error is one of the leading causes of visual impairment. Varadarajan et al. (2017) used deep learning to diagnose this. They used UK Biobank and AREDS as their datasets. The architecture is a combination of ResNet and soft attention (Xu et al., 2015). Their network can also generate attention maps. Results showed that the foveal region is one of the most import regions for making predictions.

Discussion. It is good to see that these diseases have received some attention, despite not being as prevalent as other diseases like DR, glaucoma, etc. Diagnoses of rare pathologies are also important. It is expected that the successful experiences in other diseases can easily be well extended to the rare pathologies diagnoses. Current challenges lie in the lack of data.

6.2. Systemic diseases

As shown in previous studies, the condition of the retina may reflect other diseases. In fact, many ophthalmologists have used ophthalmoscopes to diagnose systemic diseases such as hypertension, sarcoidosis and CMV infection. (Schmidt-Erfurth et al., 2018). We thus discuss studies on systemic disease diagnosis as follows.

Cardiovascular risk factors. Poplin et al. (2018) used a deep learning method to predict multiple cardiovascular risk factors including age, gender, smoking, systolic pressure (SBP) and so on. Their training dataset includes the data of 284,335 patients, and was collected from the UK Biobank and EyePACS. Inception-v3 was used as their classification network. Moreover, to help clinicians better understand the decision process of the CNN network, they used a mechanism named soft attention to generate heatmaps, which can highlight decisive regions in the process of CNN classification.

Ischemic strokes. Lim et al. (2019) utilized deep learning methods to predict strokes from fundus images. Images positive for ischemic stroke were obtained from the MCRS study (Silva et al., 2009) while negative images were taken from five other fundus image datasets. Their classification network is VGG-16. They also adopted the feature isolation method. First, a U-Net was used to segment the vessel tree from original images. Then, vascular maps were used as the input of the classification network, providing additional information.

Annotation-free cardiac vessel segmentation. Yu et al. (2019) proposed a knowledge transfer based shape-consistent generative adversarial network (SC-GAN) and a simpler Add U-Net for cardiac vessel segmentation. In SC-GAN, an average fundus image and digital subtraction angiography (DSA) image were passed to a generator to obtain a synthetic image that had

[14] https://misp.mui.ac.ir/data/eye-images.html

[15] https://palm.grand-challenge.org/

both retinal vessels and coronary arteries. A shape-consistent loss was proposed to ensure shape-consistency. A discriminator was then trained using synthetic images and real DSA images as input. Finally, a U-Net was trained using synthetic DSA images with synthetic labels for cardiac vessel segmentation. In Add U-Net, a U-Net was trained using an average fundus image and DSA image as input and a combination of fundus image annotations and the Frangi segmentation (Frangi et al., 1998) results of DSA images as labels. They used DRIVE as the source domain and collected 1,092 coronary angiographies (DSA) with no annotations.

Biological age estimation. Biological age (BA) is a widely used aging biomarker. Liu et al. (2019b) developed a CNN classifier to estimate BA based on retinal images. Two datasets named the Yangxi Dataset and Shenzhen Dataset were collected, containing 5,825 and 2,911 adults aged 50 years or older, respectively. They employed a detail manipulation method to enhance the global details of the non-specific global anatomical and physiological features related to aging. Then a VGG-19 network was used to estimate BA. They also proposed a joint loss to boost performance. Results showed that their method outperforms existing 'brain age' models.

Discussion. The diagnosis of systemic diseases using fundus images is an encouraging direction. With the successful studies using multi-disease and smartphone-based offline diagnosis systems, there is promise of predicting systematic diseases in a remote, non-invasive, offline, convenient way.

6.3. Image processing

Here we discuss approaches for two aspects of image processing, namely image registration and image quality assessment. Both of these are important for the processing and selection of images.

Image registration. Zou et al. (2020) proposed an unsupervised architecture for non-rigid retinal image registration. They formulated the image registration task as a parameterized deformation function. Thus, the aim is to regress the non-linear spatial correspondence between a pair of images. For this regression task, they proposed the Structure-Driven Regression Network (SDRN) framework, which utilizes a multiscale method to focus on global and local features simultaneously. They used the publicly available Fundus Image Registration (FIRE) dataset[16] (Hernandez-Matas et al., 2017), which consists of 129 retinal images forming 134 image pairs.

Image enhancement. Image enhancement is another way to improve performance on existing datasets. Zhao et al. (2019b) proposed a data-driven strategy to enhance blurred fundus images in a weakly supervised manner. Their strategy uses two unpaired datasets for training. Their method is the first end-to-end deep generative model for blurred retinal image enhancement. They designed two generators with the same structure; one to enhance low-quality images to high-quality ones, and the other to convert high-quality images to low-quality ones for training reference. Similarly, they designed two corresponding discriminators with the same structure. They also introduced a dynamic retinal image feature limit to guide the generator to improve performance and avoid the over-enhancement of extremely blurred areas. They used a private dataset which consists of 550 blurry images and 550 high-quality images for training and 60 blurry images for testing. The blurry images are from cataract patients and the high-quality ones are from normal people.

Image quality assessment. Retinal image quality assessment (RIQA) is important for ensuring the quality of images used by clinicians and deep learning systems. Fu et al. (2019) proposed the Multiple Color-space Fusion Network (MCF-Net) for RIQA. They first re-annotated an Eye-Quality (EyeQ) dataset with 28,792 images from EyePACS. Compared with other datasets, they extended binary labels ('Accept' and 'Reject') to ternary labels ('Good', 'Usable' and 'Reject'). For the model architecture, they first transferred the original RGB color-space to HSV and LAB color-space. Then, retinal images with different color-spaces were passed to their corresponding base network to extract features. Feature fusion was performed at a feature level and prediction level. Shen et al. (2020) proposed a domain-invariant interpretable fundus IQA system. In order to improve the interpretability, they added three clinically accepted aspects (artifact, clarity and field definition) to the output and a visual feedback. A coarse-to-fine architecture was introduced to locate landmarks, including OD and fovea, for robustness. In order to generalize well on different datasets, they adopted a semi-tied adversarial discriminative domain adaption model. They collected their dataset from patients who participated in the Shanghai Diabetic Retinopathy Screening Program (SDRSP).

Discussion. The above three directions are all significant in supporting other tasks. And the use of deep learning make these image processings effective. Shen et al. (2020)'s work is particularly inspiring. Domain invariance and visualization, which are useful for clinicians, are also goals common to almost all other tasks. It is clear that deep learning is on the frontier of research, and more architectures and methods should be explored.

7. Conclusions and Discussions

7.1. Conclusions

As demonstrated in Sections 2 to 5, the performance of deep learning for fundus image diagnostic tasks is quite impressive. In fact, deep learning methods have even achieved better performance than experienced humans in some cases. Specifically, deep learning can provide helpful suggestions for ophthalmologists. It can detect and segment important biomarkers, such as lesions, blood vessels, OD/OC, and provide evidence for physicians to diagnose specific diseases. It can also directly predict whether a patient has an ophthalmic disease, and can serve as a powerful assistant for physicians in the screening of glaucoma, DR, AMD and other ophthalmic diseases. Gulshan et al. (2019) compared the performance of deep learning with that of a human expert and a trained grader in DR diagnosis in two hospitals in India. Their results showed that the deep learning

[16] https://projects.ics.forth.gr/cvrl/fire/

system is well generalized to the actual data of India. Further, for the first time in ophthalmology, it was verified in practice rather than on datasets that deep learning achieves comparable or even better performance than human experts. The excellent performance of deep learning makes it a promising replacement for traditional computer-aided diagnostic systems (CADs). In fact, IDx (David et al., 2016) has already been approved by the US FDA for practical use. We believe that more deep learning methods will further be deployed as stable, efficient, and robust diagnostic systems for practical clinical diagnosis.

In terms of network structure, the classification backbone network has evolved from VGG and Inception-v1 to Inception-v2, Inception-v3, ResNet and DenseNet, while the segmentation backbone network has evolved from manually designed CNNs to FCNs and then to U-Net, Mask-RCNN, DeeplabV3+, etc. However, using deep learning in fundus image analysis is more than simply applying the backbone networks to specific tasks. There are many practical problems emerging. For example, the number of pixels of biomarkers such as lesions, OD/OC, and blood vessels is much smaller than that of the background. Further, the number of pixels of the unique curved structure of blood vessels, especially capillaries, makes it a hard sample problem. To solve these problems, specific methods are required according to the characteristics of each task. Advances in methods range from simple applications of deep learning to multi-branch, multiscale, and coarse-to-fine networks, as well as attention mechanisms, and so on. We have summarized the approaches in Sections 2 to 5 according to tasks and methods, and provided an overall summary in Fig. 6.

7.2. Limitations and possible solutions

Although the application of deep learning in the field of fundus image analysis has achieved gratifying performance, it is worth noting that it still has limitations in many other aspects. The problems which restrict the performance have not been solved, and the inherent limitations of deep learning also remain unresolved. Studying how to solve these limitations will be a key issue in the future of this field. We have discussed limitations for specific tasks in each section. Thus, here we will only list limitations which are common to all tasks, and provide possible solutions to these.

7.2.1. Lack of high-quality labeled data

As mentioned previously, deep learning is data-driven. There are many large-scale datasets in the field of natural image processing. For instance, ImageNet (Deng et al., 2009) has more than 14 million images. Fundus image datasets, however, are quite limited, as with other medical fields. Unlike natural images, the labeling of fundus images needs to be completed by experts, and is very difficult. For example, because of the lack of depth information, an expert typically requires eight minutes to label a fundus image for OD/OC segmentation and glaucoma diagnosis (Lim et al., 2015). These limitations have resulted in a lack of high-quality labeled fundus images. The smaller the size of the dataset, the more likely it will lead to lower accuracy and overfitting.

In addition to waiting for ophthalmologists to label more data, researchers can also seek measures to help alleviate this problem:

Weakly supervised learning. Weakly supervised learning can solve the lack of high-quality labeled data to a certain extent. We saw in Sections 2 to 5 that weakly supervised methods have already been applied to the field of fundus images. Weakly supervised learning can be divided into incomplete, inexact and inaccurate supervision. Incomplete supervision means that part of the data is labeled, while the other part is not; that is, the labels are incomplete. Active and semi-supervised learning can be used to solve this. Inexact supervision means the granularity of the annotation does not match the problem to be solved. Multi-instance learning can be used to address this. Inaccurate supervision means that the annotation is not completely accurate, so there are samples with incorrect annotations. One can consider learning with noisy labels to solve this. For specific relevant methods please refer to Zhou (2018).

Image synthesis and enhancement. Unlike weakly supervised learning, which addresses the lack of high-quality annotations, image enhancement can improve the quality of the image, while image synthesis can directly generate realistic images, even with labels. The excellent performance of image generation is due to the use of powerful GAN models. In fact, GANs have become mainstream for many image generation problems, such as style transfer, image inpainting, super resolution, and so on. Section 5 introduced several specific examples of GAN-based image synthesis. We believe that as the quality of the generated images improves, there will be more approaches that use these for training, and there will be in turn more research to further improve the quality of the synthesized images.

Federated learning. Creating a high-quality annotated dataset involves more than simply inviting ophthalmologists to annotate the data. It is a complicated matter, and there are many other considerations, including data privacy, competition in various research institutes and hospitals, and relevant laws and regulations. Note that many of the fundus image datasets are also private. How to achieve data sharing while satisfying diverse research groups, complying with regulations and not infringing upon user privacy is an urgent problem to be solved. In 2016, Google proposed federated learning to solve the "data islands" problem. Federated learning can be divided into horizontal federated learning, vertical federated learning and federated transfer learning. Horizontal federated learning means that the data features of two datasets overlap more and users overlap less. Vertical federated learning means that the two datasets have more user overlap and less user feature overlap. Federated transfer learning means that both user and data features overlap little. Federated learning is still a relatively new field to be explored. Readers can refer to Kairouz et al. (2019) for more information about federated learning.

7.2.2. Imbalance

The imbalance in fundus images is mainly an imbalance in foreground and background, or number of samples in different classes. A large portion of the methods introduced in Sections 2 to 5 were proposed to solve this problem. For in-

stance, imbalance between foreground and background occurs in many fundus image tasks, with the number of pixels in lesions, blood vessels, OD and OC, being much smaller than the number of pixels in their respective backgrounds. This imbalance will directly increase the difficulty of training. For lesions and OD/OC, using the detection network to extract the ROI as the input of the network can increase the proportion of the foreground. Such an approach can be seen in Chai et al. (2018), Fu et al. (2018b), Sarhan et al. (2019), Shah et al. (2019) and Wang et al. (2019c). Using spatial attention, as done by Wang et al. (2017) and Zhao et al. (2019e), is also very common and allows the network to focus on areas that are more decisive for solving tasks. It is worth noting that Fu et al. (2018a) performed polar coordinate transformation to alleviate the imbalance between foreground and background based on the unique ellipse shape of the OD/OC region. An imbalance in the number of samples in different classes is also very common in fundus images. One solution is to use a class balance loss function, such as cross-entropy loss and focal loss. Selective sampling is also a direction that has been explored. It uses a carefully designed sampling strategy to maintain a certain proportion of samples in different classes during each training epoch, thereby avoiding imbalance. This strategy was used in van Grinsven et al. (2016), Gondal et al. (2017), Dai et al. (2018) and Sarhan et al. (2019).

7.2.3. Poor generalization performance

There are certain differences between the various fundus image datasets, including acquisition camera, resolution, light source intensity, parameter settings, and so on. The differences between the datasets pose a challenge to the generalization performance of deep learning models. In fact, even some state-of-the-art models only perform well on certain datasets and degrade on others. This problem is mainly caused by the distribution difference between different datasets, that is, domain shift (Ghafoorian et al., 2017). Domain adaption, introduced in Section 3.2.4, can be used to enhance the model's performance on the target domain and solve the problems caused by the domain shift. This strategy was used in Wang et al. (2019b), Wang et al. (2019c), and Liu et al. (2019d) to enhance the generalization performance of the optic disc segmentation model. Domain adaption is still an area being explored, and there are various diverse methods for it. Readers can learn more from Wang and Deng (2018).

7.2.4. High consumption of deep learning

The excellent performance of deep learning comes at the cost of very high consumption, since the size of the parameters are much larger than those of traditional machine learning. This not only means that the models require significant computational resources and time during training, but also prevents them from being deployed on portable devices, such as binocular indirect ophthalmoscopes (Hajabdollahi et al., 2018). One direction to solve the high-consumption problem of deep learning is to design more novel network structures and explore operations or layers with low computational load and low memory consumption. However, this is mainly the work of basic network researchers. Researchers in this field can directly apply or learn from mature models, for example using a lightweight network to decrease complexity. For instance, MobileNet (Sandler et al., 2018) could be used instead of ResNet as the backbone. The models could also be compressed by quantizing the network weight to reduce the complexity. One approach is to quantize the weight and activation values in a CNN from a 32-bit float number to a low-bit number, or constrain the weights and activations to binary values and so on. The model pruning method can also compress the models. This strategy changes some parameters in the model to zero and skips the calculation. In addition, there are methods such as Huffman coding for the weights of the model. Some researchers have also explored how to reduce the consumption of fundus image network models specifically. For example, Hajabdollahi et al. (2018) used the method of quantifying weights and pruning to reduce the complexity of a blood vessel segmentation model.

7.2.5. Lack of interpretability

An important issue in the application of deep learning to actual medical systems is to what extent doctors accept its "black box". This lack of interpretability is an inherent defect in deep learning. Fortunately, several studies have focused on this issue. Approaches for solving this can be divided into generating heatmaps, clinical meaning of heatmaps and other explorations.

Generating heatmaps. The basic idea of several studies is to merge the feature maps of each layer of the deep network to generate a heatmap, called a class activation map (CAM) or evidence map. The generated heatmap shows which part of the image the deep network referred to when making its final judgment. Keel et al. (2019) tried to generate heatmaps for DR and GON diagnosis systems. They used a threshold strategy when generating the final probability map to visualize decisive regions for the prediction. The application of this idea can also be seen in Yang et al. (2017), Gondal et al. (2017), Quellec et al. (2017), Li et al. (2019a), Zhao et al. (2019d) and Li et al. (2020b).

Clinical meaning of heatmaps. The heatmap generated not only provides a cue that how deep learning makes decisions, but also provides guidance and assistance to the diagnostic process. Meng et al. (2020) explored how the process of generating heatmaps can improve performance of disease diagnosis and explainability of the net. They first generated heatmaps using a gradient-based classification activation map (Grad-CAM) (Selvaraju et al., 2017). Then the network was fine-tuned by several designed losses and ophthalmologist intervention. Experimental results on a private dataset showed performance improvement for the classification task. Sayres et al. (2019) evaluated the role of deep learning in guiding diagnosis. They invited 10 experts to grade DR in 796 fundus images. Each image had three forms: the original image without auxiliary information, the image with only grading results and the images with grading results and heatmaps. Images were randomly assigned to different ophthalmologists. The results showed that the assistance of deep learning diagnostic results improves the accuracy and confidence of experts in diagnosing DR, especially with heatmaps.

Other explorations. Note that there are several other studies

on the topic of interpretability. Araújo et al. (2020) proposed a deep learning-based grading system named DR|GRADUATE. In addition to the grading of DR, it can also estimate how uncertain the prediction is. de La Torre et al. (2020) proposed a deep learning-based interpretable classifier for DR grading. In their classifier, a score similar to the concept of relevance was assigned to every point of the input and hidden spaces. The scores indicated the contributions to the final prediction. Niu et al. (2019) explored interpretability in the diagnosis of DR and borrowed some ideas from Koch's law in infectious diseases.

Acknowledgments

This work is partially supported by the National Natural Science Foundation (61872200), the Natural Science Foundation of Tianjin (19JCZDJC31600, 18YFYZCG00060) and the Open Project Fund of State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences No. CARCH201905

References

Abbasi-Sureshjani, S., Smit-Ockeloen, I., Zhang, J., ter Haar Romeny, B.M., 2015. Biologically-inspired supervised vasculature segmentation in SLO retinal fundus images, in: Kamel, M., Campilho, A.J.C. (Eds.), Image Analysis and Recognition - 12th International Conference, ICIAR 2015, Niagara Falls, ON, Canada, July 22-24, 2015, Proceedings, Springer. pp. 325–334. URL: https://doi.org/10.1007/978-3-319-20801-5_35, doi: 10.1007/978-3-319-20801-5_35.

Abramoff, M.D., Garvin, M.K., Sonka, M., 2010. Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering 3, 169–208.

Abràmoff, M.D., Folk, J.C., Han, D.P., Walker, J.D., Williams, D.F., Russell, S.R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D.C., Quellec, G., Niemeijer, M., 2013. Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy. JAMA Ophthalmology 131, 351–357. URL: https://doi.org/10.1001/jamaophthalmol.2013.1743, doi: 10.1001/jamaophthalmol.2013.1743.

Abràmoff, M.D., Garvin, M.K., Sonka, M., 2010. Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering 3, 169–208.

Adem, K., 2018. Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks. Expert Syst. Appl. 114, 289–295. URL: https://doi.org/10.1016/j.eswa.2018.07.053, doi: 10.1016/j.eswa.2018.07.053.

Ahmad, M., Kasukurthi, N., Pande, H., 2019. Deep learning for weak supervision of diabetic retinopathy abnormalities, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 573–577. URL: https://doi.org/10.1109/ISBI.2019.8759417, doi: 10.1109/ISBI.2019.8759417.

Alipour, S.H.M., Rabbani, H., Akhlaghi, M., 2012. Diabetic retinopathy grading by digital curvelet transform. Comput. Math. Methods Medicine 2012, 761901:1–761901:11. URL: https://doi.org/10.1155/2012/761901, doi: 10.1155/2012/761901.

Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlaim, M., Alkatee, M., Raahemifar, K., Lakshminarayanan, V., 2018. Retinal fundus images for glaucoma analysis: the RIGA dataset, in: Zhang, J., Chen, P.H. (Eds.), Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, International Society for Optics and Photonics. SPIE. pp. 55 – 62. URL: https://doi.org/10.1117/12.2293584, doi: 10.1117/12.2293584.

Araújo, T., Aresta, G., Mendonça, L., Penas, S., Maia, C., Carneiro, Â., Mendonça, A.M., Campilho, A., 2020. Dr|graduate: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. Medical Image Anal. 63, 101715. URL: https://doi.org/10.1016/j.media.2020.101715, doi: 10.1016/j.media.2020.101715.

Badar, M., Haris, M., Fatima, A., 2020. Application of deep learning for retinal image analysis: A review. Comput. Sci. Rev. 35, 100203. URL: https://doi.org/10.1016/j.cosrev.2019.100203, doi: 10.1016/j.cosrev.2019.100203.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495. URL: https://doi.org/10.1109/TPAMI.2016.2644615, doi: 10.1109/TPAMI.2016.2644615.

Baskaran, M., Foo, R.C., Cheng, C., Narayanaswamy, A., Zheng, Y., Wu, R., Saw, S., Foster, P.J., Wong, T.Y., Aung, T., 2015. The prevalence and types of glaucoma in an urban chinese population: The singapore chinese eye study. JAMA Ophthalmology 133, 874–880.

Bourne, R.R.A., Stevens, G.A., White, R.A., Smith, J.L., Flaxman, S.R., Price, H., Jonas, J.B., Keeffe, J., Leasher, J., Naidoo, K., Pesudovs, K., Resnikoff, S., Taylor, H.R., 2013. Causes of vision loss worldwide, 1990–2010: a systematic analysis. The Lancet Global Health 1, e339 – e349. URL: http://www.sciencedirect.com/science/article/pii/S2214109X1370113X, doi: https://doi.org/10.1016/S2214-109X(13)70113-X.

Brandl, C., Breinlich, V.A., Stark, K., Enzinger, S., Asenmacher, M., Olden, M., Grassmann, F., Graw, J., Heier, M., Peters, A., et al., 2016. Features of age-related macular degeneration in the general adults and their dependency on age, sex, and smoking: Results from the german kora study. PLOS ONE 11.

Brown, J.M., Campbell, J.P., Beers, A., Chang, K., Ostmo, S., Chan, R.V.P., Dy, J., Erdogmus, D., Ioannidis, S., Kalpathy-Cramer, J., Chiang, M.F., for the Imaging, in Retinopathy of Prematurity (i ROP) Research Consortium, I., 2018. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. JAMA Ophthalmology 136, 803–810. URL: https://doi.org/10.1001/jamaophthalmol.2018.1934, doi: 10.1001/jamaophthalmol.2018.1934.

Budai, A., Bock, R., Maier, A.K., Hornegger, J., Michelson, G., 2013. Robust vessel segmentation in fundus images. Int. J. Biomed. Imaging 2013, 154860:1–154860:11. URL: https://doi.org/10.1155/2013/154860, doi: 10.1155/2013/154860.

Burlina, P., Freund, D.E., Joshi, N., Wolfson, Y., Bressler, N.M., 2016. Detection of age-related macular degeneration via deep learning, in: 13th IEEE International Symposium on Biomedical Imaging, ISBI 2016, Prague, Czech Republic, April 13-16, 2016, IEEE. pp. 184–188. URL: https://doi.org/10.1109/ISBI.2016.7493240, doi: 10.1109/ISBI.2016.7493240.

Burlina, P.M., Joshi, N., Pacheco, K.D., Freund, D.E., Kong, J., Bressler, N.M., 2018. Use of Deep Learning for Detailed Severity Characterization and Estimation of 5-Year Risk Among Patients With Age-Related Macular Degeneration. JAMA Ophthalmology 136, 1359–1366. URL: https://doi.org/10.1001/jamaophthalmol.2018.4118, doi: 10.1001/jamaophthalmol.2018.4118.

Burlina, P.M., Joshi, N., Pacheco, K.D., Liu, T.Y.A., Bressler, N.M., 2019. Assessment of Deep Generative Models for High-Resolution Synthetic Retinal Image Generation of Age-Related Macular Degeneration. JAMA Ophthalmology 137, 258–264. URL: https://doi.org/10.1001/jamaophthalmol.2018.6156, doi: 10.1001/jamaophthalmol.2018.6156.

Burlina, P.M., Joshi, N., Pekala, M., Pacheco, K.D., Freund, D.E., Bressler, N.M., 2017. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. JAMA Ophthalmology 135, 1170–1176. URL: https://doi.org/10.1001/jamaophthalmol.2017.3782, doi: 10.1001/jamaophthalmol.2017.3782.

Carmona, E.J., Rincón, M., García-Feijoó, J., Martínez-de-la-Casa, J.M., 2008. Identification of the optic nerve head with genetic algorithms. Artif. Intell. Medicine 43, 243–259. URL: https://doi.org/10.1016/j.artmed.2008.04.005, doi: 10.1016/j.artmed.2008.04.005.

Carson, L., Caroline, Y., Laura, H., Daniel, R., 2018. Retinal lesion detection with deep learning using image patches. Investigative Ophthalmology & Visual Science 59, 590–596.

Chai, Y., Liu, H., Xu, J., 2018. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowl. Based Syst. 161, 147–156. URL: https://doi.org/10.1016/j.knosys.2018.07.043, doi: 10.1016/j.knosys.2018.07.043.

Chen, L., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. CoRR abs/1706.05587. URL: http://arxiv.org/abs/1706.05587.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmenta-

tion, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII, Springer. pp. 833–851. URL: https://doi.org/10.1007/978-3-030-01234-2_49, doi: 10.1007/978-3-030-01234-2_49.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Krishnapuram, B., Shah, M., Smola, A.J., Aggarwal, C.C., Shen, D., Rastogi, R. (Eds.), Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, ACM. pp. 785–794. URL: https://doi.org/10.1145/2939672.2939785, doi: 10.1145/2939672.2939785.

Cherukuri, V., G, V.K.B., Bala, R., Monga, V., 2020. Deep retinal image segmentation with regularization under geometric priors. IEEE Trans. Image Process. 29, 2552–2567. URL: https://doi.org/10.1109/TIP.2019.2946078, doi: 10.1109/TIP.2019.2946078.

Chung, J., Gülçehre, Ç., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR abs/1412.3555. URL: http://arxiv.org/abs/1412.3555.

Ciulla, T.A., Amador, A.G., Zinman, B., 2003. Diabetic retinopathy and diabetic macular edema: Pathophysiology, screening, and novel therapies. Diabetes Care 26, 2653–2664.

Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abramoff, M., Mendonça, A.M., Campilho, A.J.C., 2018. End-to-end adversarial retinal image synthesis. IEEE Trans. Medical Imaging 37, 781–791. URL: https://doi.org/10.1109/TMI.2017.2759102, doi: 10.1109/TMI.2017.2759102.

Dai, L., Fang, R., Li, H., Hou, X., Sheng, B., Wu, Q., Jia, W., 2018. Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. IEEE Trans. Med. Imaging 37, 1149–1161. URL: https://doi.org/10.1109/TMI.2018.2794988, doi: 10.1109/TMI.2018.2794988.

Dasgupta, A., Singh, S., 2017. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation, in: 14th IEEE International Symposium on Biomedical Imaging, ISBI 2017, Melbourne, Australia, April 18-21, 2017, IEEE. pp. 248–251. URL: https://doi.org/10.1109/ISBI.2017.7950512, doi: 10.1109/ISBI.2017.7950512.

Dashtbozorg, B., Mendonça, A.M., Campilho, A.J.C., 2014. An automatic graph-based approach for artery/vein classification in retinal images. IEEE Trans. Image Process. 23, 1073–1083. URL: https://doi.org/10.1109/TIP.2013.2263809, doi: 10.1109/TIP.2013.2263809.

Dashtbozorg, B., Zhang, J., Huang, F., ter Haar Romeny, B.M., 2018. Retinal microaneurysms detection using local convergence index features. IEEE Trans. Image Process. 27, 3300–3315. URL: https://doi.org/10.1109/TIP.2018.2815345, doi: 10.1109/TIP.2018.2815345.

David, A.M., Lou, Y., Ali, E., Warren, C., Ryan, A., Folk, J.C., Meindert, N., 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative Ophthalmology & Visual Science 57, 5200–.

Decenciere, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.C., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M., Danno, R., et al., 2013. Teleophta: Machine learning and image processing methods for teleophthalmology. Irbm 34, 196–203.

Decenciere, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonezvarela, J., Massin, P., Erginay, A., et al., 2014. Feedback on a publicly distributed image database: The messidor database. Image Analysis & Stereology 33, 231–234.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Li, F., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA, IEEE Computer Society. pp. 248–255. URL: https://doi.org/10.1109/CVPR.2009.5206848, doi: 10.1109/CVPR.2009.5206848.

Deshmukh, A., Sivaswamy, J., 2019. Synthesis of optical nerve head region of fundus image, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 583–586. URL: https://doi.org/10.1109/ISBI.2019.8759414, doi: 10.1109/ISBI.2019.8759414.

Diaz-Pinto, A., Colomer, A., Naranjo, V., Morales, S., Xu, Y., Frangi, A.F., 2019. Retinal image synthesis and semi-supervised learning for glaucoma assessment. IEEE Trans. Med. Imaging 38, 2211–2218. URL: https://doi.org/10.1109/TMI.2019.2903434, doi: 10.1109/TMI.2019.2903434.

Edupuganti, V.G., Chawla, A., Kale, A., 2018. Automatic optic disk and cup segmentation of fundus images using deep learning, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 2227–2231. URL: https://doi.org/10.1109/ICIP.2018.8451753, doi: 10.1109/ICIP.2018.8451753.

Feng, S., Zhuo, Z., Pan, D., Tian, Q., 2020. Ccnet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features. Neurocomputing 392, 268–276. URL: https://doi.org/10.1016/j.neucom.2018.10.098, doi: 10.1016/j.neucom.2018.10.098.

Feng, Z., Yang, J., Yao, L., 2017. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation, in: 2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, September 17-20, 2017, IEEE. pp. 1742–1746. URL: https://doi.org/10.1109/ICIP.2017.8296580, doi: 10.1109/ICIP.2017.8296580.

Foo, A., Hsu, W., Lee, M., Lim, G., Wong, T.Y., 2020. Multi-task learning for diabetic retinopathy grading and lesion segmentation, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, AAAI Press. pp. 13267–13272. URL: https://aaai.org/ojs/index.php/AAAI/article/view/7035.

Foundation, C.H., 2015. Diabetic retinopathy detection - identify signs of diabetic retinopathy in eye images. https://www.kaggle.com/c/diabetic-retinopathy-detection/overview.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering, in: Wells, W.M., Colchester, A., Delp, S. (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI'98, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 130–137.

Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018a. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging 37, 1597–1605. URL: https://doi.org/10.1109/TMI.2018.2791488, doi: 10.1109/TMI.2018.2791488.

Fu, H., Cheng, J., Xu, Y., Zhang, C., Wong, D.W.K., Liu, J., Cao, X., 2018b. Disc-aware ensemble network for glaucoma screening from fundus image. IEEE Trans. Med. Imaging 37, 2493–2501. URL: https://doi.org/10.1109/TMI.2018.2837012, doi: 10.1109/TMI.2018.2837012.

Fu, H., Wang, B., Shen, J., Cui, S., Xu, Y., Liu, J., Shao, L., 2019. Evaluation of retinal image quality assessment networks in different color-spaces, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 48–56. URL: https://doi.org/10.1007/978-3-030-32239-7_6, doi: 10.1007/978-3-030-32239-7_6.

Fu, H., Xu, Y., Lin, S., Wong, D.W.K., Liu, J., 2016. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Ünal, G.B., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II, pp. 132–139. URL: https://doi.org/10.1007/978-3-319-46723-8_16, doi: 10.1007/978-3-319-46723-8_16.

Fumero, F., Alayón, S., Sánchez, J.L., Sigut, J.F., González-Hernández, M., 2011. RIM-ONE: an open retinal image database for optic nerve evaluation, in: Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems, 27-30 June, 2011, Bristol, United Kingdom, IEEE Computer Society. pp. 1–6. URL: https://doi.org/10.1109/CBMS.2011.5999143, doi: 10.1109/CBMS.2011.5999143.

Galdran, A., Meyer, M.I., Costa, P., Mendonça, A.M., Campilho, A., 2019. Uncertainty-aware artery/vein classification on retinal images, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 556–560. URL: https://doi.org/10.1109/ISBI.2019.8759380, doi: 10.1109/ISBI.2019.8759380.

Gargeya, R., Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. Ophthalmology , S0161642016317742.

Garway-Heath, D.F., Hitchings, R., 1998. Quantitative evaluation of the optic nerve head in early glaucoma. British Journal of Ophthalmology 82, 352–361.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., III, W.M.W., 2017.

Transfer learning for domain adaptation in MRI: application in brain lesion segmentation, in: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III, Springer. pp. 516–524. URL: https://doi.org/10.1007/978-3-319-66179-7_59, doi: 10.1007/978-3-319-66179-7_59.

Giancardo, L., Meriaudeau, F., Karnowski, T.P., Li, Y., Garg, S., Tobin, K.W., Chaum, E., 2012. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Medical Image Analysis 16, 216 – 226. URL: http://www.sciencedirect.com/science/article/pii/S1361841511001010, doi: https://doi.org/10.1016/j.media.2011.07.004.

Gondal, W.M., Köhler, J.M., Grzeszick, R., Fink, G.A., Hirsch, M., 2017. Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images, in: 2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, September 17-20, 2017, IEEE. pp. 2069–2073. URL: https://doi.org/10.1109/ICIP.2017.8296646, doi: 10.1109/ICIP.2017.8296646.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y., 2014. Generative adversarial nets, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.

Govindaiah, A., Hussain, M.A., Smith, R.T., Bhuiyan, A., 2018. Deep convolutional neural network based screening and assessment of age-related macular degeneration from fundus images, in: 15th IEEE International Symposium on Biomedical Imaging, ISBI 2018, Washington, DC, USA, April 4-7, 2018, IEEE. pp. 1525–1528. URL: https://doi.org/10.1109/ISBI.2018.8363863, doi: 10.1109/ISBI.2018.8363863.

Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Weber, B.H.F., 2018. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology 125, 8280209.

van Grinsven, M.J.J.P., van Ginneken, B., Hoyng, C.B., Theelen, T., Sánchez, C.I., 2016. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. IEEE Trans. Med. Imaging 35, 1273–1284. URL: https://doi.org/10.1109/TMI.2016.2526689, doi: 10.1109/TMI.2016.2526689.

Group, T.A.R.E.D.S.R., 1999. The age-related eye disease study (areds): Design implications areds report no. 1. Controlled Clinical Trials 20, 573 – 600. URL: http://www.sciencedirect.com/science/article/pii/S0197245699000318, doi: https://doi.org/10.1016/S0197-2456(99)00031-8.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., Webster, D.R., 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316, 2402–2410. URL: https://doi.org/10.1001/jama.2016.17216, doi: 10.1001/jama.2016.17216.

Gulshan, V., Rajan, R.P., Widner, K., Wu, D., Wubbels, P., Rhodes, T., Whitehouse, K., Coram, M., Corrado, G., Ramasamy, K., Raman, R., Peng, L., Webster, D.R., 2019. Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India. JAMA Ophthalmology 137, 987–993. URL: https://doi.org/10.1001/jamaophthalmol.2019.2004, doi: 10.1001/jamaophthalmol.2019.2004.

Guo, S., Li, T., Kang, H., Li, N., Zhang, Y., Wang, K., 2019. L-seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. Neurocomputing 349, 52–63. URL: https://doi.org/10.1016/j.neucom.2019.04.019, doi: 10.1016/j.neucom.2019.04.019.

Guo, S., Wang, K., Kang, H., Liu, T., Gao, Y., Li, T., 2020a. Bin loss for hard exudates segmentation in fundus images. Neurocomputing 392, 314–324. URL: https://doi.org/10.1016/j.neucom.2018.10.103, doi: 10.1016/j.neucom.2018.10.103.

Guo, Y., Wang, R., Zhou, X., Liu, Y., Wang, L., Lv, C., Lv, B., Xie, G., 2020b. Lesion-aware segmentation network for atrophy and detachment of pathological myopia on fundus images, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1242–1245. URL: https://doi.org/10.1109/ISBI45749.2020.9098669, doi: 10.1109/ISBI45749.2020.9098669.

Hajabdollahi, M., Esfandiarpoor, R., Najarian, K., Karimi, N., Samavi, S., Soroushmehr, S.M.R., 2018. Low complexity convolutional neural network for vessel segmentation in portable retinal diagnostic devices, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 2785–2789. URL: https://doi.org/10.1109/ICIP.2018.8451665, doi: 10.1109/ICIP.2018.8451665.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, IEEE Computer Society. pp. 2980–2988. URL: https://doi.org/10.1109/ICCV.2017.322, doi: 10.1109/ICCV.2017.322.

He, K., Sun, J., Tang, X., 2013. Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1397–1409. URL: https://doi.org/10.1109/TPAMI.2012.213, doi: 10.1109/TPAMI.2012.213.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society. pp. 770–778. URL: https://doi.org/10.1109/CVPR.2016.90, doi: 10.1109/CVPR.2016.90.

He, Q., Zou, B., Zhu, C., Liu, X., Fu, H., Wang, L., 2018. Multi-label classification scheme based on local regression for retinal vessel segmentation, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 2765–2769. URL: https://doi.org/10.1109/ICIP.2018.8451415, doi: 10.1109/ICIP.2018.8451415.

He, X., Zhou, Y., Wang, B., Cui, S., Shao, L., 2019. Dme-net: Diabetic macular edema grading by auxiliary task learning, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 788–796. URL: https://doi.org/10.1007/978-3-030-32239-7_87, doi: 10.1007/978-3-030-32239-7_87.

Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., Douma, S., Argyros, A., 2017. Fire: Fundus image registration dataset. Journal for Modeling in Opthalmology (to appear) .

Hervella, Á.S., Rouco, J., Novo, J., Ortega, M., 2018. Retinal image understanding emerges from self-supervised multimodal reconstruction, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I, Springer. pp. 321–328. URL: https://doi.org/10.1007/978-3-030-00928-1_37, doi: 10.1007/978-3-030-00928-1_37.

Hoover, A.W., Kouznetsova, V., Goldbaum, M.H., 2000. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. IEEE Trans. Medical Imaging 19, 203–210. URL: https://doi.org/10.1109/42.845178, doi: 10.1109/42.845178.

Horta, A., Joshi, N., Pekala, M., Pacheco, K.D., Kong, J., Bressler, N.M., Freund, D.E., Burlina, P., 2017. A hybrid approach for incorporating deep visual features and side channel information with applications to AMD detection, in: Chen, X., Luo, B., Luo, F., Palade, V., Wani, M.A. (Eds.), 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, Cancun, Mexico, December 18-21, 2017, IEEE. pp. 716–720. URL: https://doi.org/10.1109/ICMLA.2017.00-75, doi: 10.1109/ICMLA.2017.00-75.

Hu, J., Chen, Y., Zhong, J., Ju, R., Yi, Z., 2019. Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Trans. Med. Imaging 38, 269–279. URL: https://doi.org/10.1109/TMI.2018.2863562, doi: 10.1109/TMI.2018.2863562.

Hu, K., Zhang, Z., Niu, X., Zhang, Y., Cao, C., Xiao, F., Gao, X., 2018. Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. Neurocomputing 309, 179–191. URL: https://doi.org/10.1016/j.neucom.2018.05.011, doi: 10.1016/j.neucom.2018.05.011.

Hu, Q., Abràmoff, M.D., Garvin, M.K., 2013. Automated separation of binary overlapping trees in low-contrast color retinal images, in: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013 - 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II, Springer. pp. 436–443. URL: https://doi.org/10.1007/978-3-642-40763-5_54, doi: 10.1007/978-3-642-40763-5_54.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely

connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society. pp. 2261–2269. URL: https://doi.org/10.1109/CVPR.2017.243, doi: 10.1109/CVPR.2017.243.

Huang, Y., Lin, L., Li, M., Wu, J., Cheng, P., Wang, K., Yuan, J., Tang, X., 2020. Automated hemorrhage detection from coarsely annotated fundus images in diabetic retinopathy, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1369–1372. URL: https://doi.org/10.1109/ISBI45749.2020.9098319, doi: 10.1109/ISBI45749.2020.9098319.

Jiang, Y., Duan, L., Cheng, J., Gu, Z., Xia, H., Fu, H., Li, C., Liu, J., 2020. Jointrcnn: A region-based convolutional neural network for optic disc and cup segmentation. IEEE Trans. Biomed. Engineering 67, 335–343. URL: https://doi.org/10.1109/TBME.2019.2913211, doi: 10.1109/TBME.2019.2913211.

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Rouayheb, S.E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., Zhao, S., 2019. Advances and open problems in federated learning. CoRR abs/1912.04977. URL: http://arxiv.org/abs/1912.04977.

Kanse, S.S., Yadav, D.M., 2019. Retinal fundus image for glaucoma detection: A review and study. J. Intelligent Systems 28, 43–56. URL: https://doi.org/10.1515/jisys-2016-0258, doi: 10.1515/jisys-2016-0258.

Kauppi, T., Kalesnykiene, V., Kamarainen, J., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Uusitalo, H., Kälviäinen, H., Pietilä, J., 2007. The DIARETDB1 diabetic retinopathy database and evaluation protocol, in: Rajpoot, N.M., Bhalerao, A.H. (Eds.), Proceedings of the British Machine Vision Conference 2007, University of Warwick, UK, September 10-13, 2007, British Machine Vision Association. pp. 1–10. URL: https://doi.org/10.5244/C.21.15, doi: 10.5244/C.21.15.

Kauppi, T., Kalesnykiene, V., kristian Kamarainen, J., Lensu, L., Sorri, I., Uusitalo, H., Kälviäinen, H., Pietilä, J., 2006. Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms.

Keel, S., Wu, J., Lee, P.Y., Scheetz, J., He, M., 2019. Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma. JAMA Ophthalmology 137, 288–292. URL: https://doi.org/10.1001/jamaophthalmol.2018.6035, doi: 10.1001/jamaophthalmol.2018.6035.

Khalaf, A.F., Yassine, I.A., Fahmy, A.S., 2016. Convolutional neural networks for deep feature learning in retinal vessel segmentation, in: 2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016, IEEE. pp. 385–388. URL: https://doi.org/10.1109/ICIP.2016.7532384, doi: 10.1109/ICIP.2016.7532384.

Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected crfs with gaussian edge potentials, in: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain, pp. 109–117. URL: http://papers.nips.cc/paper/4296-efficient-inference-in-fully-connected-crfs-with-gaussian-edge-potentials.

Krause, J., Gulshan, V., Rahimy, E., Karth, P., Webster, D.R., 2017. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology 125.

Kromm, C., Rohr, K., 2020. Inception capsule network for retinal blood vessel segmentation and centerline extraction, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1223–1226. URL: https://doi.org/10.1109/ISBI45749.2020.9098538, doi: 10.1109/ISBI45749.2020.9098538.

de La Torre, J., Valls, A., Puig, D., 2020. A deep learning interpretable classifier for diabetic retinopathy disease grading. Neurocomputing 396, 465–476. URL: https://doi.org/10.1016/j.neucom.2018.07.102, doi: 10.1016/j.neucom.2018.07.102.

Li, C., Ye, J., He, J., Wang, S., Qiao, Y., Gu, L., 2020a. Dense correlation network for automated multi-label ocular disease detection with paired color fundus photographs, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1–4. URL: https://doi.org/10.1109/ISBI45749.2020.9098340, doi: 10.1109/ISBI45749.2020.9098340.

Li, L., Xu, M., Liu, H., Li, Y., Wang, X., Jiang, L., Wang, Z., Fan, X., Wang, N., 2020b. A large-scale database and a CNN model for attention-based glaucoma detection. IEEE Trans. Med. Imaging 39, 413–424. URL: https://doi.org/10.1109/TMI.2019.2927226, doi: 10.1109/TMI.2019.2927226.

Li, L., Xu, M., Wang, X., Jiang, L., Liu, H., 2019a. Attention based glaucoma detection: A large-scale database and CNN model, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, Computer Vision Foundation / IEEE. pp. 10571–10580. URL: http://openaccess.thecvf.com/content_CVPR_2019/html/Li_Attention_Based_Glaucoma_Detection_A_Large-Scale_Database_and_CNN_Model_CVPR_2019_paper.html, doi: 10.1109/CVPR.2019.01082.

Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., Kang, H., 2019b. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Inf. Sci. 501, 511–522. URL: https://doi.org/10.1016/j.ins.2019.06.011, doi: 10.1016/j.ins.2019.06.011.

Li, X., Hu, X., Yu, L., Zhu, L., Fu, C., Heng, P., 2020c. Canet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE Trans. Med. Imaging 39, 1483–1493. URL: https://doi.org/10.1109/TMI.2019.2951844, doi: 10.1109/TMI.2019.2951844.

Li, Z., He, Y., Keel, S., Meng, W., Chang, R.T., He, M., 2018a. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology , S0161642017335650.

Li, Z., Keel, S., Liu, C., He, Y., Meng, W., Scheetz, J., Lee, P.Y., Shaw, J., Ting, D., Wong, T., Taylor, H., Chang, R., He, M., 2018b. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. Diabetes Care URL: https://care.diabetesjournals.org/content/early/2018/09/27/dc18-0147, doi: 10.2337/dc18-0147.

Liao, W., Zou, B., Zhao, R., Chen, Y., He, Z., Zhou, M., 2020. Clinical interpretable deep learning model for glaucoma diagnosis. IEEE J. Biomed. Health Informatics 24, 1405–1412. URL: https://doi.org/10.1109/JBHI.2019.2949075, doi: 10.1109/JBHI.2019.2949075.

Lim, G., Cheng, Y., Hsu, W., Lee, M., 2015. Integrated optic disc and cup segmentation with deep learning, in: 27th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2015, Vietri sul Mare, Italy, November 9-11, 2015, IEEE Computer Society. pp. 162–169. URL: https://doi.org/10.1109/ICTAI.2015.36, doi: 10.1109/ICTAI.2015.36.

Lim, G., Lim, Z.W., Xu, D., Ting, D.S.W., Wong, T.Y., Lee, M., Hsu, W., 2019. Feature isolation for hypothesis testing in retinal imaging: An ischemic stroke prediction case study, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press. pp. 9510–9515. URL: https://doi.org/10.1609/aaai.v33i01.33019510, doi: 10.1609/aaai.v33i01.33019510.

Lin, T., Goyal, P., Girshick, R.B., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, IEEE Computer Society. pp. 2999–3007. URL: https://doi.org/10.1109/ICCV.2017.324, doi: 10.1109/ICCV.2017.324.

Lin, Z., Guo, R., Wang, Y., Wu, B., Chen, T., Wang, W., Chen, D.Z., Wu, J., 2018. A framework for identifying diabetic retinopathy based on anti-noise detection and attention-based fusion, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 74–82. URL: https://doi.org/10.1007/978-3-030-00934-2_9, doi: 10.1007/978-3-030-00934-2_9.

Liskowski, P., Krawiec, K., 2016. Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging 35, 2369–2380. URL: https://doi.org/10.1109/TMI.2016.2546227, doi: 10.1109/TMI.2016.2546227.

Liu, B., Gu, L., Lu, F., 2019a. Unsupervised ensemble strategy for retinal ves-

sel segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 111–119. URL: https://doi.org/10.1007/978-3-030-32239-7_13, doi: 10.1007/978-3-030-32239-7_13.

Liu, C., Wang, W., Li, Z., Jiang, Y., Han, X., Ha, J., Meng, W., He, M., 2019b. Biological age estimated from retinal imaging: A novel biomarker of aging, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 138–146. URL: https://doi.org/10.1007/978-3-030-32239-7_16, doi: 10.1007/978-3-030-32239-7_16.

Liu, H., Li, L., Wormstone, I.M., Qiao, C., Zhang, C., Liu, P., Li, S., Wang, H., Mou, D., Pang, R., Yang, D., Zangwill, L.M., Moghimi, S., Hou, H., Bowd, C., Jiang, L., Chen, Y., Hu, M., Xu, Y., Kang, H., Ji, X., Chang, R., Tham, C., Cheung, C., Ting, D.S.W., Wong, T.Y., Wang, Z., Weinreb, R.N., Xu, M., Wang, N., 2019c. Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. JAMA Ophthalmology 137, 1353–1360. URL: https://doi.org/10.1001/jamaophthalmol.2019.3501, doi: 10.1001/jamaophthalmol.2019.3501.

Liu, P., Kong, B., Li, Z., Zhang, S., Fang, R., 2019d. CFEA: collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part V, Springer. pp. 521–529. URL: https://doi.org/10.1007/978-3-030-32254-0_58, doi: 10.1007/978-3-030-32254-0_58.

Liu, Q., Hong, X., Li, S., Chen, Z., Zhao, G., Zou, B., 2019e. A spatial-aware joint optic disc and cup segmentation method. Neurocomputing 359, 285–297. URL: https://doi.org/10.1016/j.neucom.2019.05.039, doi: 10.1016/j.neucom.2019.05.039.

Liu, Y., Cheng, M., Hu, X., Wang, K., Bai, X., 2017. Richer convolutional features for edge detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society. pp. 5872–5881. URL: https://doi.org/10.1109/CVPR.2017.622, doi: 10.1109/CVPR.2017.622.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, IEEE Computer Society. pp. 3431–3440. URL: https://doi.org/10.1109/CVPR.2015.7298965, doi: 10.1109/CVPR.2015.7298965.

Lowell, J., Hunter, A., Steel, D., Basu, A., Ryder, R., Fletcher, E., Kennedy, L., 2004. Optic nerve head segmentation. IEEE Trans. Medical Imaging 23, 256–264. URL: https://doi.org/10.1109/TMI.2003.823261, doi: 10.1109/TMI.2003.823261.

Luo, W., Li, Y., Urtasun, R., Zemel, R.S., 2016. Understanding the effective receptive field in deep convolutional neural networks, in: Lee, D.D., Sugiyama, M., von Luxburg, U., Guyon, I., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain, pp. 4898–4906. URL: http://papers.nips.cc/paper/6203-understanding-the-effective-receptive-field-in-deep-convolutional-neural-networks.

Ma, W., Yu, S., Ma, K., Wang, J., Ding, X., Zheng, Y., 2019. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 769–778. URL: https://doi.org/10.1007/978-3-030-32239-7_85, doi: 10.1007/978-3-030-32239-7_85.

Mahapatra, D., Bozorgtabar, B., Hewavitharanage, S., Garnavi, R., 2017. Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis, in: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III, Springer. pp. 382–390. URL: https://doi.org/10.1007/978-3-319-66179-7_44, doi: 10.1007/978-3-319-66179-7_44.

Maninis, K., Pont-Tuset, J., Arbeláez, P.A., Gool, L.V., 2016. Deep retinal image understanding, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Ünal, G.B., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II, pp. 140–148. URL: https://doi.org/10.1007/978-3-319-46723-8_17, doi: 10.1007/978-3-319-46723-8_17.

Massin, P., Chabouis, A., Erginay, A., Viens-Bitker, C., Lecleire-Collet, A., Meas, T., Guillausseau, P.J., Choupot, G., André, B., Denormandie, P., 2008. Ophdiat: A telemedical network screening system for diabetic retinopathy in the ile-de-france. Diabetes & Metabolism 34, 227 – 234. URL: http://www.sciencedirect.com/science/article/pii/S1262363608000426, doi: https://doi.org/10.1016/j.diabet.2007.12.006.

Mathis Antony, d., 2015. Team o_o solution summary. https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15617.

Meng, Q., Hashimoto, Y., Satoh, S., 2020. How to extract more information with less burden: Fundus image classification and retinal disease localization with ophthalmologist intervention, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1373–1377. URL: https://doi.org/10.1109/ISBI45749.2020.9098600, doi: 10.1109/ISBI45749.2020.9098600.

Meyer, M.I., Galdran, A., Mendonça, A.M., Campilho, A., 2018. A pixel-wise distance regression approach for joint retinal optical disc and fovea detection, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 39–47. URL: https://doi.org/10.1007/978-3-030-00934-2_5, doi: 10.1007/978-3-030-00934-2_5.

Mishra, S., Chen, D.Z., Hu, X.S., 2020. A data-aware deep supervised method for retinal vessel segmentation, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1254–1257. URL: https://doi.org/10.1109/ISBI45749.2020.9098403, doi: 10.1109/ISBI45749.2020.9098403.

Mo, J., Zhang, L., Feng, Y., 2018. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. Neurocomputing 290, 161–171. URL: https://doi.org/10.1016/j.neucom.2018.02.035, doi: 10.1016/j.neucom.2018.02.035.

Moccia, S., Momi, E.D., Hadji, S.E., Mattos, L.S., 2018. Blood vessel segmentation algorithms - review of methods, datasets and evaluation metrics. Comput. Methods Programs Biomed. 158, 71–91. URL: https://doi.org/10.1016/j.cmpb.2018.02.001, doi: 10.1016/j.cmpb.2018.02.001.

Mohan, D., Kumar, J.R.H., Seelamantula, C.S., 2018. High-performance optic disc segmentation using convolutional neural networks, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 4038–4042. URL: https://doi.org/10.1109/ICIP.2018.8451543, doi: 10.1109/ICIP.2018.8451543.

Mohan, D., Kumar, J.R.H., Seelamantula, C.S., 2019. Optic disc segmentation using cascaded multiresolution convolutional neural networks, in: 2019 IEEE International Conference on Image Processing, ICIP 2019, Taipei, Taiwan, September 22-25, 2019, IEEE. pp. 834–838. URL: https://doi.org/10.1109/ICIP.2019.8804267, doi: 10.1109/ICIP.2019.8804267.

Nasery, V., Soundararajan, K.B., Galeotti, J.M., 2020. Learning to segment vessels from poorly illuminated fundus images, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1232–1236. URL: https://doi.org/10.1109/ISBI45749.2020.9098694, doi: 10.1109/ISBI45749.2020.9098694.

Natarajan, S., Jain, A., Krishnan, R., Rogye, A., Sivaprasad, S., 2019. Diagnostic Accuracy of Community-Based Diabetic Retinopathy Screening With an Offline Artificial Intelligence System on a Smartphone. JAMA Ophthalmology 137, 1182–1188. URL: https://doi.org/10.1001/jamaophthalmol.2019.2923, doi: 10.1001/jamaophthalmol.2019.2923.

Nguyen, H.V., Tan, G.S.W., Tapp, R.J., Mital, S., Ting, D.S.W., Wong, H.T., Tan, C.S., Laude, A., Tai, E.S., Tan, N.C.a., 2016. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in singapore. Ophthalmology , 2571–2580.

Niemeijer, M., van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G.,

Sánchez, C.I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., Wu, X., Cazuguel, G., You, J., Mayo, A., Li, Q., Hatanaka, Y., Cochener, B., Roux, C., Karray, F., García, M., Fujita, H., Abràmoff, M.D., 2010. Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans. Medical Imaging 29, 185–195. URL: https://doi.org/10.1109/TMI.2009.2033909, doi: 10.1109/TMI.2009.2033909.

Niemeijer, M., Xu, X., Dumitrescu, A.V., Gupta, P., van Ginneken, B., Folk, J.C., Abràmoff, M.D., 2011. Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs. IEEE Trans. Med. Imaging 30, 1941–1950. URL: https://doi.org/10.1109/TMI.2011.2159619, doi: 10.1109/TMI.2011.2159619.

Niu, Y., Gu, L., Lu, F., Lv, F., Wang, Z., Sato, I., Zhang, Z., Xiao, Y., Dai, X., Cheng, T., 2019. Pathological evidence exploration in deep retinal image diagnosis, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press. pp. 1093–1101. URL: https://doi.org/10.1609/aaai.v33i01.33011093, doi: 10.1609/aaai.v33i01.33011093.

Oliveira, A., Pereira, S., Silva, C.A., 2018. Retinal vessel segmentation based on fully convolutional neural networks. Expert Syst. Appl. 112, 229–242. URL: https://doi.org/10.1016/j.eswa.2018.06.034, doi: 10.1016/j.eswa.2018.06.034.

Orlando, J.I., Breda, J.B., van Keer, K., Blaschko, M.B., Blanco, P.J., Bulant, C.A., 2018. Towards a glaucoma risk index based on simulated hemodynamics from fundus images, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 65–73. URL: https://doi.org/10.1007/978-3-030-00934-2_8, doi: 10.1007/978-3-030-00934-2_8.

Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Bogunovic, H., 2020. REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Anal. 59. URL: https://doi.org/10.1016/j.media.2019.101570, doi: 10.1016/j.media.2019.101570.

Owen, C.G., Rudnicka, A.R., Mullen, R.J., Barman, S., Monekosso, D., Whincup, P.H., Ng, J., Paterson, C., 2009. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. Investigative Ophthalmology & Visual Science 50, 2004–2010.

Pal, A., Moorthy, M.R., Shahina, A., 2018. G-eyenet: A convolutional autoencoding classifier framework for the detection of glaucoma from retinal fundus images, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 2775–2779. URL: https://doi.org/10.1109/ICIP.2018.8451029, doi: 10.1109/ICIP.2018.8451029.

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters - improve semantic segmentation by global convolutional network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society. pp. 1743–1751. URL: https://doi.org/10.1109/CVPR.2017.189, doi: 10.1109/CVPR.2017.189.

Peng, Y., Dharssi, S., Chen, Q., Keenan, T.D., Agrón, E., Wong, W.T., Chew, E.Y., Lu, Z., 2018. Deepseenet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. Ophthalmology .

Phene, S., Dunn, R.C., Hammel, N., Liu, Y., Krause, J., Kitade, N., Schaekermann, M., Sayres, R., Wu, D.J., Bora, A., Semturs, C., Misra, A., Huang, A.E., Spitze, A., Medeiros, F.A., Maa, A.Y., Gandhi, M., Corrado, G.S., Peng, L., Webster, D.R., 2019. Deep learning and glaucoma specialists: The relative importance of optic disc features to predict glaucoma referral in fundus photographs. Ophthalmology 126, 1627 – 1639. URL: http://www.sciencedirect.com/science/article/pii/S0161642019318755, doi: https://doi.org/10.1016/j.ophtha.2019.07.024.

Playout, C., Duval, R., Cheriet, F., 2018. A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 101–108. URL: https://doi.org/10.1007/978-3-030-00934-2_12, doi: 10.1007/978-3-030-00934-2_12.

Playout, C., Duval, R., Cheriet, F., 2019. A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images. IEEE Trans. Med. Imaging 38, 2434–2444. URL: https://doi.org/10.1109/TMI.2019.2906319, doi: 10.1109/TMI.2019.2906319.

Pohlen, T., Hermans, A., Mathias, M., Leibe, B., 2017. Full-resolution residual networks for semantic segmentation in street scenes, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society. pp. 3309–3318. URL: https://doi.org/10.1109/CVPR.2017.353, doi: 10.1109/CVPR.2017.353.

Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M., Corrado, G., Peng, L., Webster, D.R., 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering 2, 158–164.

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Mériaudeau, F., 2018. Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. Data 3, 25. URL: https://doi.org/10.3390/data3030025, doi: 10.3390/data3030025.

Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., Wu, T., Xiao, J., Wang, F., Yin, B., Wang, Y., Danala, G., He, L., Choi, Y.H., Mériaudeau, F., 2020. Idrid: Diabetic retinopathy - segmentation and grading challenge. Medical Image Anal. 59. URL: https://doi.org/10.1016/j.media.2019.101561, doi: 10.1016/j.media.2019.101561.

for Retinopathy of Prematurity Cooperative Group, T.C., 1988. Multicenter trial of cryotherapy for retinopathy of prematurity: Preliminary results. Archives of ophthalmology 106, 471–479.

Quellec, G., Charrière, K., Boudi, Y., Cochener, B., Lamard, M., 2017. Deep image mining for diabetic retinopathy screening. Medical Image Anal. 39, 178–193. URL: https://doi.org/10.1016/j.media.2017.04.012, doi: 10.1016/j.media.2017.04.012.

Quellec, G., Lamard, M., Conze, P., Massin, P., Cochener, B., 2020. Automatic detection of rare pathologies in fundus photographs using few-shot learning. Medical Image Anal. 61, 101660. URL: https://doi.org/10.1016/j.media.2020.101660, doi: 10.1016/j.media.2020.101660.

Raghavendra, U., Fujita, H., Bhandary, S.V., Gudigar, A., Tan, J.H., Acharya, U.R., 2018. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf. Sci. 441, 41–49. URL: https://doi.org/10.1016/j.ins.2018.01.051, doi: 10.1016/j.ins.2018.01.051.

Rahimy, Ehsan, 2018. Deep learning applications in ophthalmology. Current Opinion in Ophthalmology , 1.

Raj, P.K., Manjunath, A., Kumar, J.R.H., Seelamantula, C.S., 2020. Automatic classification of artery/vein from single wavelength fundus images, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1262–1265. URL: https://doi.org/10.1109/ISBI45749.2020.9098580, doi: 10.1109/ISBI45749.2020.9098580.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. CoRR abs/1804.02767. URL: http://arxiv.org/abs/1804.02767.

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, pp. 91–99. URL: http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.

Robinson, B., 2003. Prevalence of asymptomatic eye disease prévalence des maladies oculaires asymptomatiques. Revue Canadienne D'Optométrie , 175–180.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Navab, N., Hornegger, J., III, W.M.W., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Mu-

nich, Germany, October 5 - 9, 2015, Proceedings, Part III, Springer. pp. 234–241. URL: https://doi.org/10.1007/978-3-319-24574-4_28, doi: 10.1007/978-3-319-24574-4_28.

Roy, P.K., Tennakoon, R.B., Cao, K., Sedai, S., Mahapatra, D., Maetschke, S., Garnavi, R., 2017. A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images, in: 14th IEEE International Symposium on Biomedical Imaging, ISBI 2017, Melbourne, Australia, April 18-21, 2017, IEEE. pp. 1078–1082. URL: https://doi.org/10.1109/ISBI.2017.7950703, doi: 10.1109/ISBI.2017.7950703.

Sabour, S., Frosst, N., Hinton, G.E., 2017. Dynamic routing between capsules, in: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, pp. 3856–3866. URL: http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules.

Salamat, N., Missen, M.M.S., Rashid, A., 2019. Diabetic retinopathy techniques in retinal images: A review. Artif. Intell. Medicine 97, 168–188. URL: https://doi.org/10.1016/j.artmed.2018.10.009, doi: 10.1016/j.artmed.2018.10.009.

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L., 2018. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. CoRR abs/1801.04381. URL: http://arxiv.org/abs/1801.04381.

dos Santos Ferreira, M.V., de Carvalho Filho, A.O., de Sousa, A.D., Silva, A.C., Gattass, M., 2018. Convolutional neural network and texture descriptor-based automatic detection and diagnosis of glaucoma. Expert Syst. Appl. 110, 250–263. URL: https://doi.org/10.1016/j.eswa.2018.06.010, doi: 10.1016/j.eswa.2018.06.010.

Sarhan, M.H., Albarqouni, S., Yigitsoy, M., Navab, N., Eslami, A., 2019. Multi-scale microaneurysms segmentation using embedding triplet loss, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 174–182. URL: https://doi.org/10.1007/978-3-030-32239-7_20, doi: 10.1007/978-3-030-32239-7_20.

Sayres, R., Taly, A., Rahimy, E., Blumer, K., Coz, D., Hammel, N., Krause, J., Narayanaswamy, A., Rastegar, Z., Wu, D., Xu, S., Barb, S., Joseph, A., Shumski, M., Smith, J., Sood, A.B., Corrado, G.S., Peng, L., Webster, D.R., 2019. Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology 126, 552 – 564. URL: http://www.sciencedirect.com/science/article/pii/S0161642018315756, doi: https://doi.org/10.1016/j.ophtha.2018.11.016.

Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B.S., Waldstein, S.M., Bogunović, H., 2018. Artificial intelligence in retina. Progress in Retinal and Eye Research 67, 1 – 29. URL: http://www.sciencedirect.com/science/article/pii/S1350946218300119, doi: https://doi.org/10.1016/j.preteyeres.2018.07.004.

Sedai, S., Mahapatra, D., Hewavitharanage, S., Maetschke, S., Garnavi, R., 2017a. Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder, in: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II, Springer. pp. 75–82. URL: https://doi.org/10.1007/978-3-319-66185-8_9, doi: 10.1007/978-3-319-66185-8_9.

Sedai, S., Tennakoon, R.B., Roy, P.K., Cao, K., Garnavi, R., 2017b. Multi-stage segmentation of the fovea in retinal fundus images using fully convolutional neural networks, in: 14th IEEE International Symposium on Biomedical Imaging, ISBI 2017, Melbourne, Australia, April 18-21, 2017, IEEE. pp. 1083–1086. URL: https://doi.org/10.1109/ISBI.2017.7950704, doi: 10.1109/ISBI.2017.7950704.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, IEEE Computer Society. pp. 618–626. URL: https://doi.org/10.1109/ICCV.2017.74, doi: 10.1109/ICCV.2017.74.

Sengupta, S., Singh, A., Leopold, H.A., Gulati, T., Lakshminarayanan, V., 2020. Ophthalmic diagnosis using deep learning with fundus images - A critical review. Artif. Intell. Medicine 102, 101758. URL: https://doi.org/10.1016/j.artmed.2019.101758, doi: 10.1016/j.artmed.2019.101758.

Shah, S., Kasukurthi, N., Pande, H., 2019. Dynamic region proposal networks for semantic segmentation in automated glaucoma screening, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 578–582. URL: https://doi.org/10.1109/ISBI.2019.8759171, doi: 10.1109/ISBI.2019.8759171.

Shankaranarayana, S.M., Ram, K., Mitra, K., Sivaprakasam, M., 2019. Fully convolutional networks for monocular retinal depth estimation and optic disc-cup segmentation. IEEE J. Biomed. Health Informatics 23, 1417–1426. URL: https://doi.org/10.1109/JBHI.2019.2899403, doi: 10.1109/JBHI.2019.2899403.

Shen, Y., Sheng, B., Fang, R., Li, H., Dai, L., Stolte, S., Qin, J., Jia, W., Shen, D., 2020. Domain-invariant interpretable fundus image quality assessment. Medical Image Anal. 61, 101654. URL: https://doi.org/10.1016/j.media.2020.101654, doi: 10.1016/j.media.2020.101654.

Silva, D.A.D., Liew, G., Wong, M.C., Chang, H.M., Wong, T.Y., 2009. Retinal vascular caliber and extracranial carotid disease in patients with acute ischemic stroke: the multi-centre retinal stroke (mcrs) study. Stroke; a journal of cerebral circulation 40, 3695–9.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. URL: http://arxiv.org/abs/1409.1556.

Sivaprasad, S., Arden, G., Prevost, A.T., Crosby-Nwaobi, R., Holmes, H., Kelly, J., Murphy, C., Rubin, G., Vasconcelos, J., Hykin, P., 2014. A multicentre phase iii randomised controlled single-masked clinical trial evaluating the clinical efficacy and safety of light-masks at preventing dark-adaptation in the treatment of early diabetic macular oedema (cleopatra): study protocol for a randomised controlled trial. Trials .

Sivaswamy, J., Krishnadas, S.R., Joshi, G.D., Jain, M., Tabish, A.U.S., 2014. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation, in: IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014, April 29 - May 2, 2014, Beijing, Chin, Beijing, China, IEEE. pp. 53–56. URL: https://doi.org/10.1109/ISBI.2014.6867807, doi: 10.1109/ISBI.2014.6867807.

Soomro, T.A., Afifi, A.J., Gao, J., Hellwich, O., Zheng, L., Paul, M., 2019. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. Expert Syst. Appl. 134, 36–52. URL: https://doi.org/10.1016/j.eswa.2019.05.029, doi: 10.1016/j.eswa.2019.05.029.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.A., 2015. Striving for simplicity: The all convolutional net, in: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings. URL: http://arxiv.org/abs/1412.6806.

Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B., 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Medical Imaging 23, 501–509. URL: https://doi.org/10.1109/TMI.2004.825627, doi: 10.1109/TMI.2004.825627.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: Singh, S.P., Markovitch, S. (Eds.), Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA, AAAI Press. pp. 4278–4284. URL: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, IEEE Computer Society. pp. 1–9. URL: https://doi.org/10.1109/CVPR.2015.7298594, doi: 10.1109/CVPR.2015.7298594.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society. pp. 2818–2826. URL: https://doi.org/10.1109/CVPR.2016.308, doi: 10.1109/CVPR.2016.308.

Tan, J.H., Bhandary, S.V., Sivaprasad, S., Hagiwara, Y., Bagchi, A., Raghaven-

dra, U., Rao, A.K., Raju, B., Shetty, N.S., Gertych, A., Chua, K.C., Acharya, U.R., 2018. Age-related macular degeneration detection using deep convolutional neural network. Future Gener. Comput. Syst. 87, 127–135. URL: https://doi.org/10.1016/j.future.2018.05.001, doi: 10.1016/j.future.2018.05.001.

Tan, J.H., Fujita, H., Sivaprasad, S., Bhandary, S.V., Rao, A.K., Chua, K.C., Acharya, U.R., 2017. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. Inf. Sci. 420, 66–76. URL: https://doi.org/10.1016/j.ins.2017.08.050, doi: 10.1016/j.ins.2017.08.050.

Tasman, W., Patz, A., Mcnamara, J.A., Kaiser, R.S., Trese, M.T., Smith, B.T., 2006. Retinopathy of prematurity: The life of a lifetime disease. American Journal of Ophthalmology 141, 0–174.

Taylor, S., Brown, J.M., Gupta, K., Campbell, J.P., Ostmo, S., Chan, R.V.P., Dy, J., Erdogmus, D., Ioannidis, S., Kim, S.J., Kalpathy-Cramer, J., Chiang, M.F., for the Imaging, in Retinopathy of Prematurity Consortium, I., 2019. Monitoring Disease Progression With a Quantitative Severity Scale for Retinopathy of Prematurity Using Deep Learning. JAMA Ophthalmology 137, 1022–1028. URL: https://doi.org/10.1001/jamaophthalmol.2019.2433, doi: 10.1001/jamaophthalmol.2019.2433.

Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C., 2014. Global prevalence of glaucoma and projections of glaucoma burden through 2040 a systematic review and meta-analysis. Ophthalmology 121, 2081–2090.

Ting, D., Pasquale, L., Peng, L., Campbell, J., Lee, A., Raman, R., Tan, G., Schmetterer, L., Keane, P., Wong, T., 2018a. Artificial intelligence and deep learning in ophthalmology. British Journal of Ophthalmology 103, bjophthalmol–2018. doi: 10.1136/bjophthalmol-2018-313173.

Ting, D.S., Peng, L., Varadarajan, A.V., Keane, P.A., Burlina, P.M., Chiang, M.F., Schmetterer, L., Pasquale, L.R., Bressler, N.M., Webster, D.R., Abramoff, M., Wong, T.Y., 2019. Deep learning in ophthalmology: The technical and clinical considerations. Progress in Retinal and Eye Research 72, 100759. URL: http://www.sciencedirect.com/science/article/pii/S1350946218300909, doi: https://doi.org/10.1016/j.preteyeres.2019.04.003.

Ting, D.S.W., Cheung, C.Y.L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I.Y., Lee, S.Y., Wong, E.Y.M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N.C., Finkelstein, E.A., Lamoureux, E.L., Wong, I.Y., Bressler, N.M., Sivaprasad, S., Varma, R., Jonas, J.B., He, M.G., Cheng, C.Y., Cheung, G.C.M., Aung, T., Hsu, W., Lee, M.L., Wong, T.Y., 2017. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 318, 2211–2223. URL: https://doi.org/10.1001/jama.2017.18152, doi: 10.1001/jama.2017.18152.

Ting, D.S.W., Pasquale, L.R., Peng, L., Campbell, J.P., Lee, A.Y., Raman, R., Tan, G.S.W., Schmetterer, L., Keane, P.A., Wong, T.Y., 2018b. Artificial intelligence and deep learning in ophthalmology. British Journal of Ophthalmology .

Tu, Z., Gao, S., Zhou, K., Chen, X., Fu, H., Gu, Z., Cheng, J., Yu, Z., Liu, J., 2020. Sunet: A lesion regularized model for simultaneous diabetic retinopathy and diabetic macular edema grading, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1378–1382. URL: https://doi.org/10.1109/ISBI45749.2020.9098673, doi: 10.1109/ISBI45749.2020.9098673.

V, S.A., Sivaswamy, J., 2019. Matching the characteristics of fundus and smartphone camera images, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 569–572. URL: https://doi.org/10.1109/ISBI.2019.8759381, doi: 10.1109/ISBI.2019.8759381.

Varadarajan, A.V., Poplin, R., Blumer, K., Angermüller, C., Ledsam, J., Chopra, R., Keane, P.A., Corrado, G., Peng, L., Webster, D.R., 2017. Deep learning for predicting refractive error from retinal fundus images. CoRR abs/1712.07798. URL: http://arxiv.org/abs/1712.07798.

Vos, T., Allen, C., Arora, M., Barber, R.M., Bhutta, Z.A., Brown, A., Carter, A., Casey, D.C., et al., F.J.C., 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. The Lancet 388, 1545 – 1602. URL: http://www.sciencedirect.com/science/article/pii/S0140673616316786, doi: https://doi.org/10.1016/S0140-6736(16)31678-6.

Wang, B., Qiu, S., He, H., 2019a. Dual encoding u-net for retinal vessel segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 84–92. URL: https://doi.org/10.1007/978-3-030-32239-7_10, doi: 10.1007/978-3-030-32239-7_10.

Wang, K., Zhang, X., Huang, S., Wang, Q., Chen, F., 2020. Ctf-net: Retinal vessel segmentation via deep coarse-to-fine supervision network, in: 17th IEEE International Symposium on Biomedical Imaging, ISBI 2020, Iowa City, IA, USA, April 3-7, 2020, IEEE. pp. 1237–1241. URL: https://doi.org/10.1109/ISBI45749.2020.9098742, doi: 10.1109/ISBI45749.2020.9098742.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153. URL: https://doi.org/10.1016/j.neucom.2018.05.083, doi: 10.1016/j.neucom.2018.05.083.

Wang, S., Yu, L., Li, K., Yang, X., Fu, C., Heng, P., 2019b. Boundary and entropy-driven adversarial learning for fundus image segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 102–110. URL: https://doi.org/10.1007/978-3-030-32239-7_12, doi: 10.1007/978-3-030-32239-7_12.

Wang, S., Yu, L., Yang, X., Fu, C., Heng, P., 2019c. Patch-based output space adversarial learning for joint optic disc and cup segmentation. IEEE Trans. Med. Imaging 38, 2485–2495. URL: https://doi.org/10.1109/TMI.2019.2899910, doi: 10.1109/TMI.2019.2899910.

Wang, X., Ju, L., Zhao, X., Ge, Z., 2019d. Retinal abnormalities recognition using regional multitask learning, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 30–38. URL: https://doi.org/10.1007/978-3-030-32239-7_4, doi: 10.1007/978-3-030-32239-7_4.

Wang, X., Xu, M., Li, L., Wang, Z., Guan, Z., 2019e. Pathology-aware deep network visualization and its application in glaucoma image synthesis, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 423–431. URL: https://doi.org/10.1007/978-3-030-32239-7_47, doi: 10.1007/978-3-030-32239-7_47.

Wang, Z., Dong, N., Rosario, S.D., Xu, M., Xie, P., Xing, E.P., 2019f. Ellipse detection of optic disc-and-cup boundary in fundus images, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 601–604. URL: https://doi.org/10.1109/ISBI.2019.8759173, doi: 10.1109/ISBI.2019.8759173.

Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H., Wang, X., 2017. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection, in: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III, Springer. pp. 267–275. URL: https://doi.org/10.1007/978-3-319-66179-7_31, doi: 10.1007/978-3-319-66179-7_31.

Wilkinson, C.P., Iii, F.L.F., Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J.T., 2003. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 110, 0–1682.

Wu, Y., Xia, Y., Song, Y., Zhang, D., Liu, D., Zhang, C., Cai, W., 2019. Vessel-net: Retinal vessel segmentation under multi-path supervision, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 264–272. URL: https://doi.org/10.1007/978-3-030-32239-7_30, doi: 10.1007/978-3-030-32239-7_30.

Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W., 2018. Multiscale network followed network model for retinal vessel segmentation, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, Septem-

ber 16-20, 2018, Proceedings, Part II, Springer. pp. 119–126. URL: https://doi.org/10.1007/978-3-030-00934-2_14, doi: 10.1007/978-3-030-00934-2_14.

Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W., 2020. Nfn +: A novel network followed network for retinal vessel segmentation. Neural Networks 126, 153–162. URL: https://doi.org/10.1016/j.neunet.2020.02.018, doi: 10.1016/j.neunet.2020.02.018.

Xie, S., Tu, Z., 2015. Holistically-nested edge detection, in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1395–1403.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A.C., Salakhutdinov, R., Zemel, R.S., Bengio, Y., 2015. Show, attend and tell: Neural image caption generation with visual attention, in: Bach, F.R., Blei, D.M. (Eds.), Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015, JMLR.org. pp. 2048–2057. URL: http://proceedings.mlr.press/v37/xuc15.html.

Xu, X., Zhang, L., Li, J., Guan, Y., Zhang, L., 2020. A hybrid global-local representation CNN model for automatic cataract grading. IEEE J. Biomed. Health Informatics 24, 556–567. URL: https://doi.org/10.1109/JBHI.2019.2914690, doi: 10.1109/JBHI.2019.2914690.

Xue, J., Yan, S., Qu, J., Qi, F., Qiu, C., Zhang, H., Chen, M., Liu, T., Li, D., Liu, X., 2019. Deep membrane systems for multitask segmentation in diabetic retinopathy. Knowl. Based Syst. 183. URL: https://doi.org/10.1016/j.knosys.2019.104887, doi: 10.1016/j.knosys.2019.104887.

Yan, F., Cui, J., Wang, Y., Liu, H., Liu, H., Wei, B., Yin, Y., Zheng, Y., 2018a. Deep random walk for drusen segmentation from fundus images, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 48–55. URL: https://doi.org/10.1007/978-3-030-00934-2_6, doi: 10.1007/978-3-030-00934-2_6.

Yan, Z., Han, X., Wang, C., Qiu, Y., Xiong, Z., Cui, S., 2019a. Learning mutually local-global u-nets for high-resolution retinal lesion segmentation in fundus images, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019, IEEE. pp. 597–600. URL: https://doi.org/10.1109/ISBI.2019.8759579, doi: 10.1109/ISBI.2019.8759579.

Yan, Z., Yang, X., Cheng, K., 2018b. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. IEEE Trans. Biomed. Engineering 65, 1912–1923. URL: https://doi.org/10.1109/TBME.2018.2828137, doi: 10.1109/TBME.2018.2828137.

Yan, Z., Yang, X., Cheng, K., 2019b. A three-stage deep learning model for accurate retinal vessel segmentation. IEEE J. Biomed. Health Informatics 23, 1427–1436. URL: https://doi.org/10.1109/JBHI.2018.2872813, doi: 10.1109/JBHI.2018.2872813.

Yang, Y., Li, T., Li, W., Wu, H., Fan, W., Zhang, W., 2017. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks, in: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III, Springer. pp. 533–540. URL: https://doi.org/10.1007/978-3-319-66179-7_61, doi: 10.1007/978-3-319-66179-7_61.

Yin, P., Wu, Q., Xu, Y., Min, H., Yang, M., Zhang, Y., Tan, M., 2019. Pm-net: Pyramid multi-label network for joint optic disc and cup segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 129–137. URL: https://doi.org/10.1007/978-3-030-32239-7_15, doi: 10.1007/978-3-030-32239-7_15.

Yu, F., Zhao, J., Gong, Y., Wang, Z., Li, Y., Yang, F., Dong, B., Li, Q., Zhang, L., 2019. Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part II, Springer. pp. 714–722. URL: https://doi.org/10.1007/978-3-030-32245-8_79, doi: 10.1007/978-3-030-32245-8_79.

Yu, L., Qin, Z., Zhuang, T., Ding, Y., Qin, Z., Choo, K.R., 2020. A framework for hierarchical division of retinal vascular networks. Neurocomputing 392, 221–232. URL: https://doi.org/10.1016/j.neucom.2018.11.113, doi: 10.1016/j.neucom.2018.11.113.

Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., Tan, M., Xu, Y., 2019a. Attention guided network for retinal image segmentation, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 797–805. URL: https://doi.org/10.1007/978-3-030-32239-7_88, doi: 10.1007/978-3-030-32239-7_88.

Zhang, W., Zhong, J., Yang, S., Gao, Z., Hu, J., Chen, Y., Yi, Z., 2019b. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowl. Based Syst. 175, 12–25. URL: https://doi.org/10.1016/j.knosys.2019.03.016, doi: 10.1016/j.knosys.2019.03.016.

Zhang, Y., Chung, A.C.S., 2018. Deep supervision with additional labels for retinal vessel segmentation task, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Springer. pp. 83–91. URL: https://doi.org/10.1007/978-3-030-00934-2_10, doi: 10.1007/978-3-030-00934-2_10.

Zhang, Z., Liu, J., Yin, F., Lee, B., Wong, D.W.K., Sung, K.R., 2013. Achikok: Database of fundus images from glaucoma patients, in: 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), pp. 228–231.

Zhang, Z., Srivastava, R., Liu, H., Chen, X., Duan, L., Wong, D.W.K., Kwoh, C.K., Wong, T.Y., Liu, J., 2014. A survey on computer aided diagnosis for ocular diseases. BMC Med. Inf. & Decision Making 14, 80. URL: https://doi.org/10.1186/1472-6947-14-80, doi: 10.1186/1472-6947-14-80.

Zhang, Z., Yin, F.S., Liu, J., Wong, W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y., 2010. Origa-light: An online retinal fundus image database for glaucoma analysis and research, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 3065–3068.

Zhao, H., Li, H., Cheng, L., 2020a. Improving retinal vessel segmentation with joint local loss by matting. Pattern Recognit. 98. URL: https://doi.org/10.1016/j.patcog.2019.107068, doi: 10.1016/j.patcog.2019.107068.

Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L., 2018. Synthesizing retinal and neuronal images with generative adversarial nets. Medical Image Anal. 49, 14–26. URL: https://doi.org/10.1016/j.media.2018.07.001, doi: 10.1016/j.media.2018.07.001.

Zhao, H., Li, H., Maurer-Stroh, S., Guo, Y., Deng, Q., Cheng, L., 2019a. Supervised segmentation of un-annotated retinal fundus images by synthesis. IEEE Trans. Medical Imaging 38, 46–56. URL: https://doi.org/10.1109/TMI.2018.2854886, doi: 10.1109/TMI.2018.2854886.

Zhao, H., Yang, B., Cao, L., Li, H., 2019b. Data-driven enhancement of blurry retinal images via generative adversarial networks, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 75–83. URL: https://doi.org/10.1007/978-3-030-32239-7_9, doi: 10.1007/978-3-030-32239-7_9.

Zhao, R., Chen, X., Liu, X., Chen, Z., Guo, F., Li, S., 2020b. Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning. IEEE J. Biomed. Health Informatics 24, 1104–1113. URL: https://doi.org/10.1109/JBHI.2019.2934477, doi: 10.1109/JBHI.2019.2934477.

Zhao, R., Chen, Z., Liu, X., Zou, B., Li, S., 2019c. Multi-index optic disc quantification via multitask ensemble learning, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 21–29. URL: https://doi.org/10.1007/978-3-030-32239-7_3, doi: 10.1007/978-3-030-32239-7_3.

Zhao, R., Li, S., 2020. Multi-indices quantification of optic nerve head in fundus image via multitask collaborative learning. Medical Image Anal. 60. URL: https://doi.org/10.1016/j.media.2019.101593, doi: 10.1016/j.media.2019.101593.

Zhao, R., Liao, W., Zou, B., Chen, Z., Li, S., 2019d. Weakly-supervised simultaneous evidence identification and segmentation for automated glau-

coma diagnosis, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press. pp. 809–816. URL: https://doi.org/10.1609/aaai.v33i01.3301809, doi: 10.1609/aaai.v33i01.3301809.

Zhao, Z., Zhang, K., Hao, X., Tian, J., Chua, M.C.H., Chen, L., Xu, X., 2019e. Bira-net: Bilinear attention net for diabetic retinopathy grading, in: 2019 IEEE International Conference on Image Processing, ICIP 2019, Taipei, Taiwan, September 22-25, 2019, IEEE. pp. 1385–1389. URL: https://doi.org/10.1109/ICIP.2019.8803074, doi: 10.1109/ICIP.2019.8803074.

Zheng, Y., Cheng, C., Lamoureux, E.L., Chiang, P.P., Anuar, A.R., Wang, J.J., Mitchell, P., Saw, S., Wong, T.Y., 2013. How much eye care services do asian populations need? projection from the singapore epidemiology of eye disease (seed) study. Investigative Ophthalmology & Visual Science 54, 2171–2177.

Zhou, B., Khosla, A., Lapedriza, À., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society. pp. 2921–2929. URL: https://doi.org/10.1109/CVPR.2016.319, doi: 10.1109/CVPR.2016.319.

Zhou, Y., He, X., Cui, S., Zhu, F., Liu, L., Shao, L., 2019. High-resolution diabetic retinopathy image synthesis manipulated by grading and lesions, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019 - 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I, Springer. pp. 505–513. URL: https://doi.org/10.1007/978-3-030-32239-7_56, doi: 10.1007/978-3-030-32239-7_56.

Zhou, Y., Li, G., Li, H., 2020. Automatic cataract classification using deep neural network with discrete state transition. IEEE Trans. Med. Imaging 39, 436–446. URL: https://doi.org/10.1109/TMI.2019.2928229, doi: 10.1109/TMI.2019.2928229.

Zhou, Z., 2018. A brief introduction to weakly supervised learning. National Science Review 5, 44–53.

Zhu, J., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251.

Zou, B., He, Z., Zhao, R., Zhu, C., Liao, W., Li, S., 2020. Non-rigid retinal image registration using an unsupervised structure-driven regression network. Neurocomputing 404, 14–25. URL: https://doi.org/10.1016/j.neucom.2020.04.122, doi: 10.1016/j.neucom.2020.04.122.