

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329355711>

Four Models for Automatic Recognition of Left and Right Eye in Fundus Images

Conference Paper · January 2019

DOI: 10.1007/978-3-030-05710-7_42

CITATIONS

2

READS

203

6 authors, including:



Xirong Li

Renmin University of China

102 PUBLICATIONS 1,966 CITATIONS

[SEE PROFILE](#)



Dayong Ding

Zhigu

31 PUBLICATIONS 107 CITATIONS

[SEE PROFILE](#)



Jieping xu

Renmin University of China

30 PUBLICATIONS 171 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



TRECVID 2014 [View project](#)



Collaboration and Productivity IT Research at Intel [View project](#)



Four Models for Automatic Recognition of Left and Right Eye in Fundus Images

Xin Lai^{1,2}, Xirong Li¹, Rui Qian¹, Dayong Ding², Jun Wu³, and Jieping Xu¹(✉)

¹ Key Lab of DEKE, Remin University of China, Beijing 100872, China
jieping@ruc.edu.cn

² Vistel AI Lab, Beijing 100872, China

³ Northwestern Polytechnical University, Xi'an 710072, China

Abstract. Fundus image analysis is crucial for eye condition screening and diagnosis and consequently personalized health management in a long term. This paper targets at left and right eye recognition, a basic module for fundus image analysis. We study how to automatically assign left-eye/right-eye labels to fundus images of posterior pole. For this under-explored task, four models are developed. Two of them are based on optic disc localization, using extremely simple max intensity and more advanced Faster R-CNN, respectively. The other two models require no localization, but perform holistic image classification using classical Local Binary Patterns (LBP) features and fine-tuned ResNet-18, respectively. The four models are tested on a real-world set of 1,633 fundus images from 834 subjects. Fine-tuned ResNet-18 has the highest accuracy of 0.9847. Interestingly, the LBP based model, with the trick of left-right contrastive classification, performs closely to the deep model, with an accuracy of 0.9718.

Keywords: Medical image analysis · Fundus images · Left and right eye recognition · Optic disc localization · Left-right contrastive classification · Deep learning

1 Introduction

Medical image analysis, either content-based or using multiple modalities, is crucial for both instant computer-aided diagnosis and longer-term personal health management. Among different types of medical images, fundus images are of unique importance for two reasons. First, fundus photography, imaging the retina of an eye including retinal vasculature, optic disc, and macula, provides an effective measure for ophthalmologists to evaluate conditions such as diabetic retinopathy, age-related macular degeneration, and glaucoma. These disorders are known to be sight threatening and even result in vision loss. Second, fundus photography is noninvasive, and the invention of non-mydriatic fundus cameras makes it even more patient-friendly and thus well suited for routinely health screening. This paper contributes to fundus image analysis. Different from current works that focus on diagnosis related tasks [3, 8], we aim for *left and right*

eye recognition, i.e., automatically determining if a specific fundus image is from a left or a right eye.

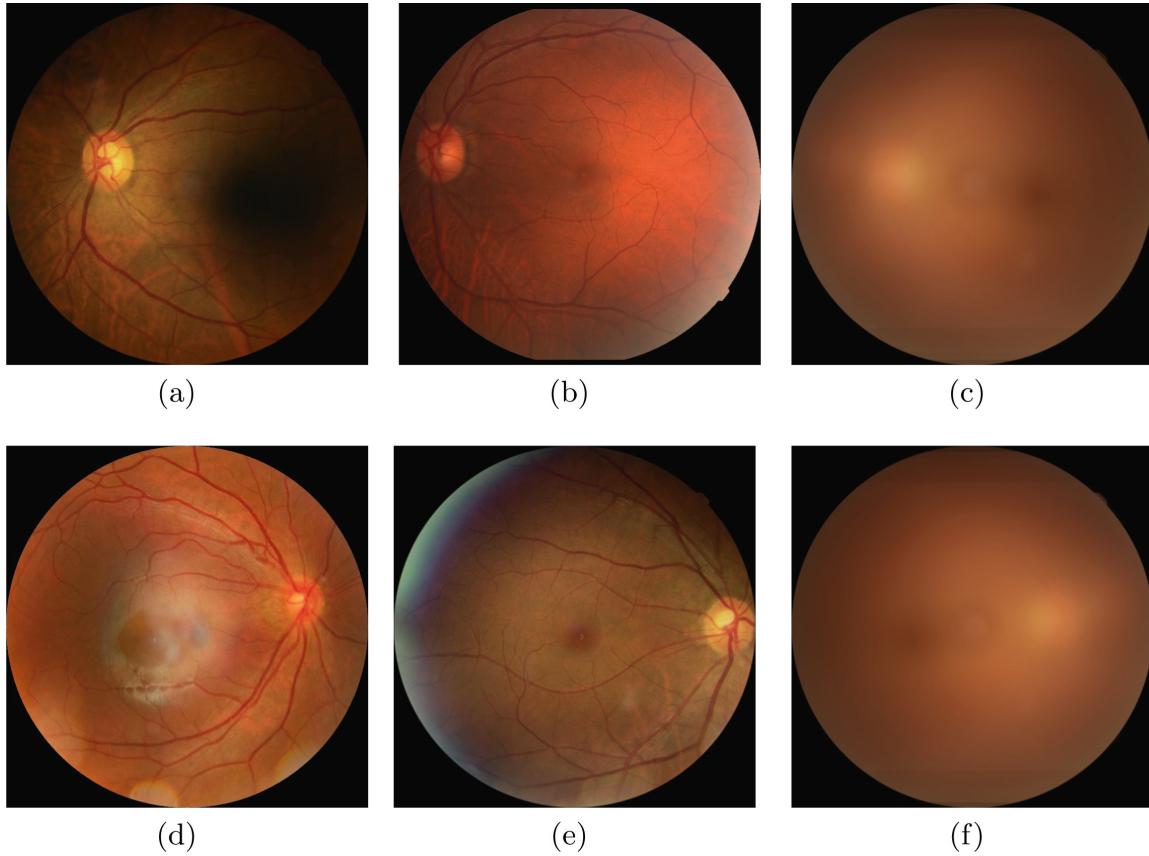


Fig. 1. Examples of fundus images, with the first row showing images taken from left eyes and the second row showing images from right eyes. These images capture the posterior pole, i.e., the retina between the optic disc and the macula. The two artificial images in the last column are generated by averaging many left-eye and right-eye images, respectively.

The left-eye and right-eye labels are basic contextual information that needs to be associated with a specific fundus image. The labels are necessary because the conditions of a subject's two eyes are not necessarily correlated. Diagnosis and health monitoring have to be performed and documented per eye. Besides, the labels are important for other applications such as fundus based person verification [6], where pairs of images to be compared have to be both left-eye or both right-eye. At present, this labeling task is manually accomplished, mainly by fundus camera operators.

Despite its importance, left and right eye recognition appears to be largely unexplored. Few efforts have been made [14, 15], both trying to leverage the location of the optic disc. The optic disc is the entry point for the major blood vessels that supply the retina [2]. Its area, roughly shaped as an ellipse, is typically the brightest in a fundus image, see Fig. 1. Tan *et al.* [14] develop their

left and right eye recognition based on optic disc localization and vessel segmentation. In particular, they first identify a region of interest (ROI) based on pixel intensities. Vessels within this ROI are segmented. A given image is classified as left-eye, if the left half of the ROI has less vessel pixels than the right half. They report a classification accuracy of 0.923 on a set of 194 fundus images. Later in [15], the same team improves over [14] by training an SVM classifiers based on segmentation-based features, reporting an accuracy of 0.941 on a set of 102 fundus images. Similar to the two pioneering works we also exploit the optic disc. Our major novelties are that we develop new models that require no optic disc localization, nor vessel segmentation, see Table 1. Moreover we investigate deep learning techniques that have not been considered for this task.

Table 1. Major characteristics of the two existing and four proposed models for left and right eye recognition in fundus images.

Model	Optic disk localization?	Vessel segmentation?	Learning based?	Deep learning?
Tan <i>et al.</i> [14]	✓	✓	✗	✗
Tan <i>et al.</i> [15]	✓	✓	✓	✗
<i>This work:</i>				
ODL-MI	✓	✗	✗	✗
ODL-CNN	✓	✗	✓	✓
LRCC	✗	✗	✓	✗
FT-CNN	✗	✗	✓	✓

In ophthalmology, the posterior pole refers to the retina between the optic disc and the macula [1]. These two areas are crucial for examination and diagnosis. Hence, fundus images of posterior pole are most commonly used in practice. In such images, the optic disc is typically observed at the left half of a left-eye image, and at the right half of a right-eye image. This phenomenon is demonstrated by averaging left-eye images and right-eye images, respectively, see the last column of Fig. 1. The above observation leads to the following questions: *Is precise localization of the optic disc necessary? Can the problem of left and right eye recognition be effectively solved by determining at which half the optic disc appears?*

For answering these two questions, this paper makes contributions as follows:

1. We propose two types of models, according to their dependency on optic disc localization. For both types, we look into traditional image processing techniques and present-day deep learning techniques. The combination leads to four distinct models.
2. We show that with a proper design, the proposed non-deep learning model is nearly comparable to the ResNet-18 based model, yet is computationally light with no need of GPU for training and execution.
3. Experiments on a test set of 1,633 fundus images from 834 subjects, which are over 10 times larger than those reported in the literature, show the state-of-the-art performance of the proposed models.

The rest of the paper is organized as follows. The proposed models are described in Sect. 2, followed by experiments in Sect. 3. We conclude the paper in Sect. 4.

2 Four Models for Left and Right Eye Recognition

2.1 Problem Statement

Given a fundus image of posterior pole, the problem of left and right eye recognition is to automatically determine whether the fundus image was taken from the left or right eye of a specific person. A binary classification model is thus required. Without loss of generality, we consider fundus images from left eyes as positive instances. Accordingly, fundus images from right eyes are treated as negatives. Let x be a fundus image, and x_l and x_r denoting the left and right half of the image, respectively. Let y be a binary variable, where $y = 1$ indicates left eye and 0 otherwise. We use $p(y = 1|x) \in [0, 1]$ to denote a model that produces a probabilistic output of being left eye, and $h(x) \in \{0, 1\}$ as a model that gives a hard classification.

Next, we propose four models, where the first two models count on optic disc localization, while the last two models require no localization. Figure 2 conceptually illustrate the proposed models.

2.2 Model I. Optic Disc Localization by Max-Intensity (ODL-MI)

As noted in Sect. 1, the location of the optic disc is a strong cue for left and right eye recognition. In the meanwhile, the optic disc tends to be the brightest area in a normal fundus image. To exploit this priori knowledge, we propose a naive model which looks for the brightest region, i.e., the region with the maximum intensity. Hence, we term this model *Optic Disc Localization by Max-Intensity*, abbreviated as ODL-MI.

Given a color fundus image x , ODL-MI first converts x to gray-scale. The image is then uniformly divided into $s \times s$ regions, with s empirically set to 10. The intensity of each region is obtained by averaging the intensity of all pixels within the region. Accordingly, the region with the maximum intensity is localized, *i.e.*,

$$i^*, j^* = \underset{i,j \in \{1, \dots, s\}}{\operatorname{argmax}} \text{Intensity}(x, i, j), \quad (1)$$

where $\text{Intensity}(x, i, j)$ returns the averaged intensity of the region indexed by i and j . If the center of this region falls in x_l , the image will be classified as left eye. We formalize the above classification process as

$$h_{ODL-MI}(x) = \begin{cases} 1, & \text{if the center of region } (i^*, j^*) \text{ falls in } x_l \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The effectiveness of this fully intensity-driven model depends on image quality. For varied reasons including bad photography and bad eye conditions, some part of a fundus image might appear to be more brighter than the optic disc, see Fig. 1(e). So we consider learning-based optic disc localization as follows.

Author Proof

Automatic Recognition of Left and Right Eye in Fundus Images 5

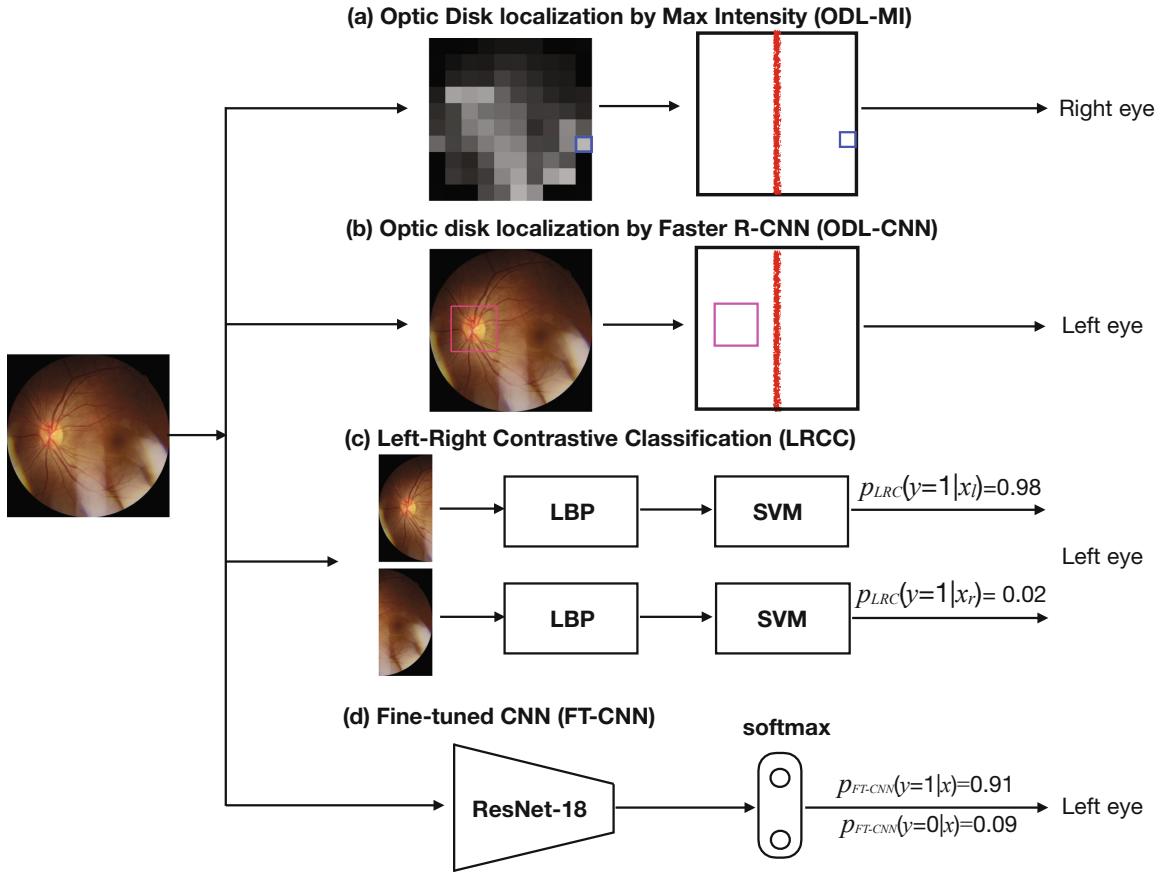


Fig. 2. A conceptual diagram of the four proposed models for left and right eye recognition. The first two models, *i.e.*, ODL-MI and ODL-CNN, are based on optic disc localization, while the LRCC and FT-CNN models are *localization free*, resolving the recognition problem by holistic classification. Note that For LRCC, the two sub images, corresponding to the left and right half of the input image, go through the same LBP feature extraction module and the same SVM classification module.

2.3 Model II: Optic Disc Localization by CNN (ODL-CNN)

In this model, we improve the optic disc localization component of ODL-MI, by substituting an object detection CNN for the intensity-based rule. Notice that for left and right eye recognition, knowing the precise boundary of the optic disc is unnecessary. A bounding box centered around the optic disc is adequate. In that regard, we adopt Faster R-CNN [12], a well-performed CNN for object detection and localization. In particular, we use Faster R-CNN we trained for the task of joint segmentation of the optic disc and the optic cup in fundus images. Given a test image, the network proposes 300 candidate regions of interest. The proposed regions are then fed into the classification block of Faster R-CNN. Consequently, each region is predicted with a probability of covering the optic disc region. After non-maximum suppression, the best region is selected as the final proposal. We consider the center of this region as the coordinate of the optic disc. Subsequently, a similar decision rule as described in Sect. 2.2 is applied, *i.e.*,

$$h_{ODL-CNN}(x) = \begin{cases} 1, & \text{if the center of the proposed region falls in } x_l \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The ODL-MI and ODL-CNN models both heavily rely on precise localization of the optic disc. Note that in order to determine whether an image is left-eye or right eye, the horizontal position of the optic disc is far more important than its vertical position. This means precision localization might be unnecessary. Following this hypothesis, we develop in Sects. 2.4 and 2.5 two models that are localization free.

2.4 Model III. Left-Right Contrastive Classification (LRCC)

As the horizontal position of the optic matters, we propose to reformulate the recognition problem as to determine which half of a given image x contains the optic disc. As this is essentially the left half x_l versus the right half x_r , we term the new model *Left-Right Contrastive Classification (LRCC)*.

As we cannot assume a priori which sub image, x_l or x_r , contains the optic disc, they have to be treated equally. Hence, we need a visual feature that is discriminative to capture the visual appearance of the optic disc against its background. In the meanwhile, the feature should be robust against moderate rotation, low contrast and illumination changes often present in fundus images. In that regard, we employ the rotation-invariant Local Binary Pattern (LBP) feature [7].

Obtained by comparing every pixel with its surrounding pixels, an LBP descriptor can recognize bright and dark spots and edges of curvature at a given scale [7]. Specifically, for each pixel in an image, a circle of radius R centered on this pixel is first formed. Every pixel on the circle is compared against the central pixel, with the comparison result encoded as 1 if larger and 0 otherwise. This results in a binary pattern of length $8 \times R$. Note that, the pattern changes with respect to the choice of the starting point and the rotation of the image. The rotation-invariant LBP cancels out such changes by circling the pattern end to end, and categorizing it into a fixed set of classes based on the number of bitwise 0/1 changes in the circle. In this work the radius R is empirically set to 3, resulting in a 25-dimensional LBP feature per sub image. The feature has been l_1 normalized in advance to the subsequent supervised learning.

To train the LRCC model, we construct training instances as follows. For a left-eye image, its left sub image is used as a positive instance with its right sub image as negative. While for a right-eye image, its left sub image will be treated as a negative instance. In this context, $p_{LRCC}(y = 1|x_l)$ and $p_{LRCC}(y = 1|x_r)$ indicate the probability of the optic disc occurring in the left half and the right half, respectively. We train a linear SVM to produce the two probabilities. Accordingly, the LRCC model is expressed as

$$h_{LRCC}(x) = \begin{cases} 1, & p_{LRCC}(y = 1|x_l) > p_{LRCC}(y = 1|x_r) \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Note that the left-right contrastive strategy allows LRCC to get rid of thresholding.

2.5 Model IV. Classification by Fine-Tuned CNN (FT-CNN)

We aim to build a deep CNN model that directly categorizes an input image into either left or right eye. Note that the relatively limited availability of our training data makes it difficult to effectively learn a new CNN from scratch. We therefore turn to fine tuning [11, 17]. The main idea of this training strategy is to initialize the new CNN with its counterpart pre-trained on the large-scale ImageNet dataset.

In this work we adapt a ResNet-18 network [4], which strikes a good balance between classification accuracy and GPU footprint. The network has been pre-trained predict 1,000 visual objects defined in the ImageNet Large Scale Visual Recognition Challenge [13]. For our binary classification task, we replace the task layer, *i.e.*, the last fully connected layer, of ResNet-18 by a new fully connected layer consisting of two neurons. Accordingly, our CNN-based model is expressed as

$$p_{cnn}(y|x) := \text{softmax}(\text{ResNet-18}(x)), \quad (5)$$

where *softmax* indicates a softmax layer converting the output of ResNet-18 network into probabilist output. Accordingly, FT-CNN makes a decision as

$$h_{FT-CNN}(x) = \underset{\hat{y} \in \{0,1\}}{\operatorname{argmax}} p_{cnn}(y = \hat{y}|x). \quad (6)$$

The model is re-trained to minimize the cross entropy loss by stochastic gradient descent with a momentum of 0.9. The learning rate is initially set to 0.001, and decays every 7 epochs. The number of epochs is 150 in total. The model scoring the best validation accuracy is retained.

3 Experiments

3.1 Experimental Setup

Datasets. We use the public Kaggle fundus image dataset [5] as our training data. While originally developed for diabetic retinopathy detection, the left and right eye information of the Kaggle images can be extracted from their filenames. Nevertheless, we observe incorrect labels, *e.g.*, images with their filename indicating left eye might actually be right eye, and vice versa. We improve label quality by manually verifying and correcting the original annotations. To make manual labeling affordable, we took a random subset of around 12 K images. During the labeling process, images that cannot be categorized, *e.g.*, those with optic discs invisible, were removed. This results in a set of 11,126 images, 60% of which is used for training and the remaining 40% is used as an validation set for optimizing hyper parameters. We constructed a test set of 1,633 images collected through eye screening programmes performed in local sites. Therefore, the test set is completely independent of our training and validation sets. The

Table 2. Basic statistics of datasets used in our experiments. We use a random subset of the Kaggle DR dataset [5] for training and validation, and an independent set of 1,633 fundus images for testing.

	Training set	Validation set	Test set
Left-eye images	3,286	2,194	778
Right-eye images	3,369	2,277	845
Total	6,655	4,471	1,633

test images come from 834 subjects, with 474 females and 360 males. Table 2 presents basic statistics of the three datasets.

Preprocessing. Note that a fundus image is captured under a specific spatial extend of a circular field-of-view, visually indicated by a round mask. As there is no relevant information outside the mask, each image has been automatically cropped as follows. We use a square bounding box tangent to the round mask, so that the cropped image is a square one containing only the field-of-view. The bounding box is estimated by fitting a circle from candidate points detected on the boundary of the mask.

Implementations. We use the scikit-image toolbox [16] to extract the LBP features, and scikit-learn [10] to train the SVM models. The penalty parameter C is selected to maximize the model accuracy on the validation set. For deep learning we use PyTorch [9].

Evaluation Criterion. As the two classes are more or less balanced, we report accuracy, *i.e.*, the rate of test images correctly predicted.

3.2 Results

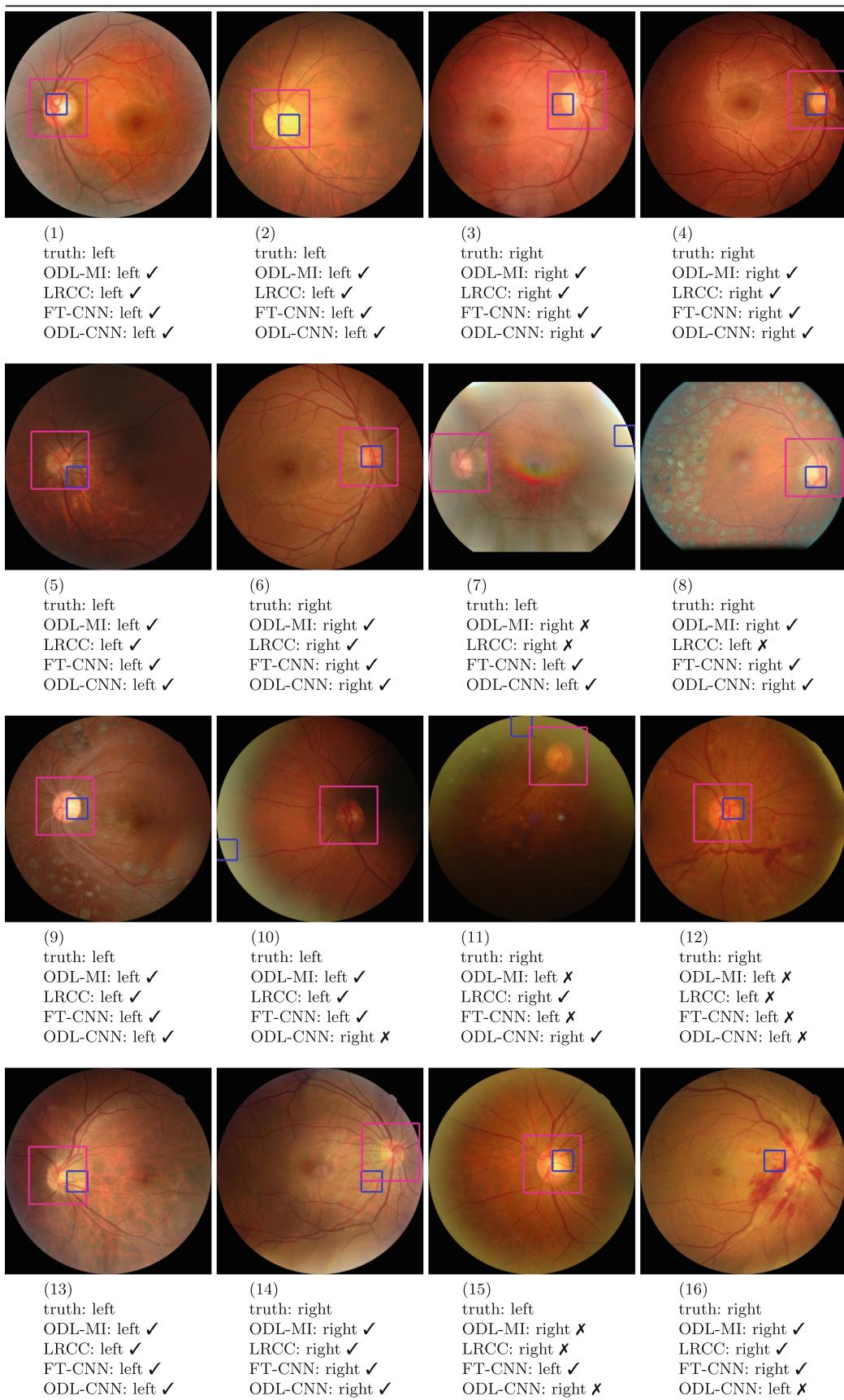
Table 3 summarizes the performance of the four models on the test set. FT-CNN, with an accuracy of 0.9847, performs the best. It is followed by ODL-CNN (0.9767), LRCC (0.9718) and ODL-MI (0.9314).

Misclassification by ODL-MI is mainly due to its incorrect localization of the optic disc, see examples #7 and #11 in Table 4. For ODL-CNN, it performs

Table 3. Performance of the four proposed models on the test set, sorted in ascending order according to their recognition accuracy. The notation $X \rightarrow Y$ means X is predicated as Y . Fine-tuned CNN (FT-CNN) is the best.

Proposed model	Correct prediction		Incorrect prediction		Accuracy
	$Left \rightarrow Left$	$Right \rightarrow Right$	$Left \rightarrow Right$	$Right \rightarrow Left$	
ODL-MI	733	788	55	57	0.9314
LRCC	772	815	16	30	0.9718
ODL-CNN	783	812	5	33	0.9767
FT-CNN	786	822	2	23	0.9847

Table 4. Some results of left and right eye recognition produced by the four models, with correct and incorrect prediction marked by ✓ and ✗. Optic disk regions found by ODL-MI and ODL-CNN are highlighted by the small blue and larger purple squares, respectively. Best viewed in color.



quite well when the optic disc can be located, scoring an accuracy of 0.9851. However, for 24 test images, ODL-CNN gives no object proposal. Consider the test image #16 in Table 4 for instance. This image shows the symptom of optic disc edema, which makes the boundary of the optic disc mostly invisible. Due to the 24 failures, the accuracy of ODL-CNN is dropped to 0.9767.

Despite its simplicity, LRCC works quite well, with a relative loss of 1.3% compared to FT-CNN. Moreover, LRCC is computationally light, with no need of GPU resources for training and execution. The left-right contrastive strategy is found to be effective. Simply using the 8-dimensional intensity histogram gives an accuracy of 0.9357. Using LBP alone gives an accuracy of 0.9706. Their concatenation brings in a marginal improvement, reaching an accuracy of 0.9718. Using the same feature but without using the contrastive strategy would make the accuracy drop to 0.5266. These results allow us to attribute the effectiveness of LRCC to its left-right contrastive strategy.

3.3 Discussion

As we mentioned in Sect. 1, [14] and [15] are the two initial attempts for left and right eye recognition in fundus images. However, both their code and data are not publicly accessible. Moreover, their models involve a number of hyper parameters that are not clearly documented. Consequently, it is difficult to replicate the two peer works with the same mathematical precision as intended by their developers. We therefore do not compare them in our experiments. Taking their recognition accuracy and the test set size into account, *i.e.*, 0.923 on 194 images [14] and 0.941 on 102 images [15], we are confident that the proposed models, with accuracy of over 0.97 on 1,633 images from real scenarios, are the state-of-the-art.

4 Conclusions

For automatic recognition of left and right eye in fundus images, we develop four models, among which ODL-MI and ODL-CNN require optic disc localization, while LRCC and FT-CNN perform holistic classification. Experiments using a set of 11,126 Kaggle images as training data and a new set of 1,633 images as test data support conclusions as follows. Precise localization of optic disc is unnecessary. Moreover, left and right eye recognition can be effectively resolved by determining which half of a fundus image contains the optic disc using the LRCC model. For the state-of-the-art performance, we recommend FT-CNN, which obtains an accuracy of 0.9847 on our test set. When striking a balance between recognition accuracy and computational resource, we recommend LRCC, which has an accuracy of 0.9718.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 61672523), the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 18XNLG19).

References

1. Cassin, B., Solomon, S.: Dictionary of Eye Terminology. Triad Publishing Company, Gainesville (1990)
2. Gamm, D.M., Albert, D.M.: Blind spot (2011). <https://www.britannica.com/science/blind-spot>. Accessed 30 July 2018
3. Gargiela, R., Leng, T.: Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* **124**(7), 962–969 (2017)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the CVPR (2016)
5. Kaggle: Diabetic retinopathy detection (2015). <https://www.kaggle.com/c/diabetic-retinopathy-detection>
6. Oinonen, H., Forsvik, H., Ruusuvuori, P., Yli-Harja, O., Voipio, V., Huttunen, H.: Identity verification based on vessel matching from fundus images. In: Proceedings of the ICIP (2010)
7. Ojala, T., Pietikäinen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *T-PAMI* **24**(7), 971–987 (2002)
8. Orlando, J., Prokofyeva, E., del Fresno, M., Blaschko, M.: Convolutional neural network transfer for automated glaucoma identification. In: Proceedings of the ISMIPA (2017)
9. Paszke, A., et al.: Automatic differentiation in PyTorch. In: NIPS-W (2017)
10. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *JMLR* **12**, 2825–2830 (2011)
11. Pittaras, N., Markatopoulou, F., Mezaris, V., Patras, I.: Comparison of fine-tuning and extension strategies for deep convolutional neural networks. In: Proceedings of the MMM (2017)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *T-PAMI* **39**, 1137–1149 (2017)
13. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *IJCV* **115**(3), 211–252 (2015)
14. Tan, N.M., et al.: Automatic detection of left and right eye in retinal fundus images. In: Lim, C.T., Goh, J.C.H. (eds.) ICBME 2009, vol. 23, pp. 610–614. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92841-6_150
15. Tan, N.M., et al.: Classification of left and right eye retinal images. In: Proceedings of the SPIE (2010)
16. van der Walt, S., et al.: the scikit-image contributors: scikit-Image: image processing in Python. *PeerJ* **2**, e453 (2014)
17. Wei, Q., Li, X., Wang, H., Ding, D., Yu, W., Chen, Y.: Laser scar detection in fundus images using convolutional neural networks. In: Proceedings of the ACCV (2018)