Medical Image Analysis 60 (2020) 101593

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Multi-indices quantification of optic nerve head in fundus image via multitask collaborative learning

Rongchang Zhao[a,b], Shuo Li[c,*]

[a] School of Computer Science and Engineering, Central South University, Changsha, China
[b] Hunan Engineering Research Center of Machine Vision and Intelligent Medicine, Changsha, China
[c] Western University, London, ON, Canada

ARTICLE INFO

Article history:
Received 6 September 2019
Revised 13 October 2019
Accepted 25 October 2019
Available online 31 October 2019

Keywords:
Multi-indices quantification
Optic nerve head assessment
Collaborative learning
Glaucoma diagnosis

ABSTRACT

Multi-indices quantification of optic nerve head (ONH), measuring ONH appearance with multiple types of indices simultaneously from fundus images, is the most clinically significant tasks for accurate ONH assessment and ophthalmic disease diagnosis. However, no attempt has been reported due to its challenges of the large variation of fundus appearance across patients, heavy overlap and extremely weak contrast between optic nerve head areas. In this paper, we propose a multitask collaborative learning framework (MCL-Net) for multi-indices ONH quantification. The proposed MCL-Net, a two-branch neural network, first obtains expressive shared and task-specific representations with the backbone network and its two branches; then models the feature exchanges and aggregations between two branches with a well-designed feature interaction module (FIM) to promote each other collaboratively. After that, it estimates multiple types of ONH indices under a multitask ensemble module (MEM) that is capable of learning aggregation of multiple outputs automatically. Therefore, the proposed MCL-Net is consisted of the feature representation, inter-task feature interaction, dual-branch task-specific prediction, and multi-task quantification ensemble, which establish an effective framework which takes full advantages of segmentation and estimation tasks for multi-indices ONH quantification. Rather than the low-level feature sharing and individual prediction, the proposed MCL-Net collaboratively learns an optimal combination of shared and task-specific representation, as well as the aggregated prediction, therefore leads to accurate quantification of ONH with multiple types of indices.

Experimental results on the dataset of 650 fundus images show that MCL-Net successfully delivers accurate quantification of all the three types of ONH indices, with average mean absolute error of 0.98 ± 0.20, 0.97 ± 0.16, 1.19 ± 0.18, as well as average correlation coefficient of 0.699, 0.708 and 0.691, for diameters, whole areas and regional areas, respectively. In addition, the experiments demonstrate that quantitative indices obtained by our method provide more effective glaucoma diagnosis with AUC of 0.8698. This endows our proposed MCL-Net a great potential in clinical assessment from focal to global for ophthalmic disease diagnosis.

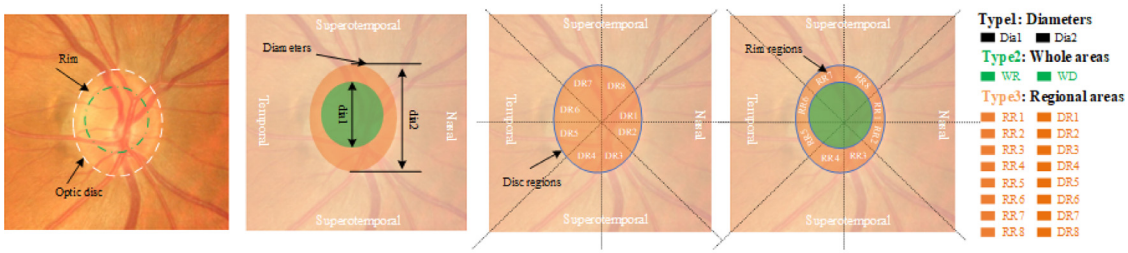
1. Introduction

Accurate quantification of optic nerve head (ONH) is one of the most clinically significant tasks in the diagnosis, treatment, and follow-up of many ophthalmic diseases, especially chronic glaucoma (Harizman et al., 2006; Maninis et al., 2016; Xu et al., 2014). It acts as an effective assessment tool in clinical practice to provide detailed quantitative information for the analysis of the anatomical structures and pathological features in retina. According to (Garway-Heath and Hitchings, 1998), multi-indices ONH quantification, measuring the ONH appearance from different views with 20 indices simultaneously, helps clinician to comprehensively evaluate the focal and global appearance of the ONH with multiple types of different indices, such as vertical diameters of optic disc (OD) and cup (OC), areas of OD and neuroretinal rim for the whole disc and individual 45 degree regions shown in Fig. 1(a).

The current manual ONH segmentation-and-measurement approach is very time-consuming and subjective, thus it is extremely useful to develop an automatic multi-indices ONH quantification method. However, no attempt has been reported to achieve automated multi-indices ONH quantification from fundus images due to three challenges: (1) Estimating multiple types of quantitative

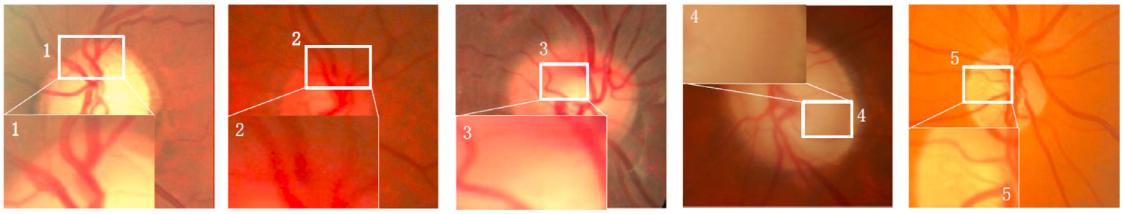
* Corresponding author.
E-mail addresses: zhaorc@csu.edu.cn (R. Zhao), slishuo@gmail.com (S. Li).

https://doi.org/10.1016/j.media.2019.101593
1361-8415/



(a) Multi-indices ONH quantification including three types of indices.

(b) Large variation of ONH appearance in fundus images across patients



(c) The extremely weak contrast between optic disc and neuroretinal rim

Fig. 1. (a) Multi-indices ONH quantification including three types of indices: diameters, whole areas and regional areas. RR: rim regions; DR: disc regions; WR: whole rim; WD: whole disc. (b) Examples of fundus images with diversity of ONH appearances which cannot be comprehensively described by single method. (c) The extremely weak contrast between optic disc and neuroretinal rim in appearance and shape, which leads to the highly ill-defined borders to separate optic cup and rim areas for the accurate ONH quantification.

indices directly from fundus image is very complicated and error-prone. The estimation often requires a complex nonlinear mapping from the fundus image to a multivariate vector, which is hard to model with existing methods. (2) Large variation of fundus appearance across patients heavily increases the difficulty of feature representation for multi-indices ONH quantification. ONH appearance changes in different ways with different pathology, e.g., cupping caused by thinning of neuroretinal rim and notch caused by the focal enlargement of the cup. Fig. 1(b) shows some fundus images with a diversity of ONH appearances. (3) There exist heavy overlap and extremely weak contrast between optic cup and neuroretinal rim in appearance and shape (as shown in Fig. 1(c)), which leads to the highly ill-defined borders to separate optic cup and rim areas from complex background for the accurate ONH quantification. Moreover, structure borders between optic cup and rim may not be clearly visible or different experts may have different style of annotating, which introduces ambiguities.

While many researches have been devoted to segmenting-and-measuring ONH automatically, existing methods address only single index ONH quantification (Cheng et al., 2015; Fu et al., 2018; Jiang et al., 2018; Zhao et al., 2019c), such as the vertical cup-to-disc ratio (CDR). These methods are generally categorized into two independent approaches: (1) segmentation-and-measurement method, which measures on the segmented OD and OC masks for quantification index; (2) directly estimation from fundus images, which learns a nonlinear mapping between fundus image and quantitative index. The most common method is the segmentation-and-measurement, where the robust and effective segmentation algorithm is a prerequisite to segregate optic disc/cup regions from the complex surroundings with clear borders, including statistical shape model (Xu et al., 2007; Cheng et al., 2011; Mary et al., 2015), multiview and multimodal approaches (Joshi et al., 2012; Miri et al., 2015), superpixel-based model (Cheng et al., 2013; Xu et al., 2014), and deep learning methods (Maninis et al., 2016; Guo et al., 2016). Then an ellipse fitting is adopted as the prior to measure the vertical diameters of optic cup and disc. Recently, direct estimation (Zhao et al., 2019) of the quantitative index from fundus images becomes popular, which tackle ONH quantification as a regression problem by bypassing the ill-conditioned segmentation. The ONH quantification is always formulated as the regression problem to learn an association between the given image and their corresponding prediction based on the expressive feature representation.

However, successful multi-indices ONH quantification has still not been reported due to the limitations of the segmentation-and-measurement or direct index estimation method individually. Those methods predict ONH indices based on the incomplete features only from the single task, but ignore the potential mutual benefits with each other, which is more essential since local spatial cues from segmentation task and global context from estimation task are complementary information for accurate ONH quantification. On one hand, segmentation-and-measurement methods provide pixel-wise spatial information for precise ONH localization, but it is easily disturbed by the great variability of shape and inhomogeneity in ONH appearance, especially the ambiguity optic cup borders, because it lacks end-to-end supervised information of global indices in the training procedure. On the other hand, direct estimation method provides the end-to-end ONH quantifica-

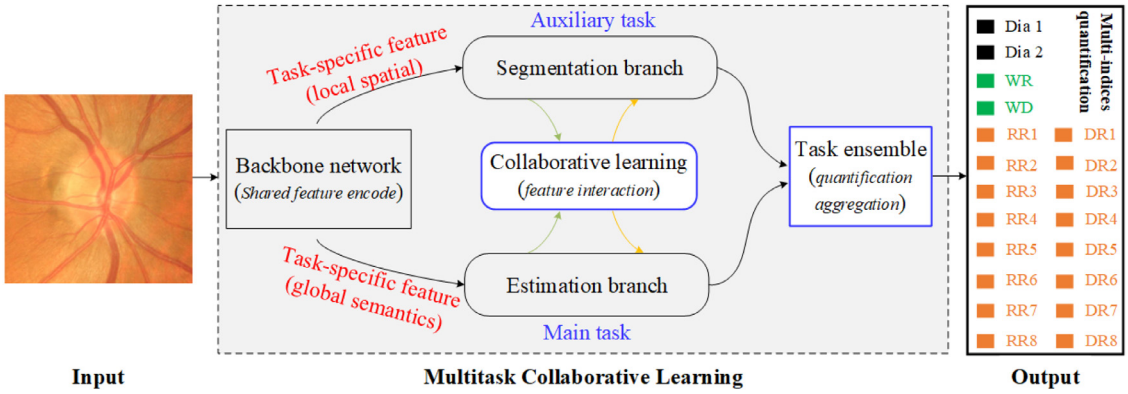


Fig. 2. The proposed multitask collaborative learning framework leverages the merit of spatial and semantic dependencies between ONH segmentation and indices estimation by multitask architecture and collaborative learning.

tion approach which directly learns the global context, shapes and structures to regress, yet it ignores pixel-level details of the ONH structures and visual interpretation for clinical ONH assessment.

In this paper, we propose a novel Multitask Collaborative Learning method (MCL, Fig. 2) to achieve the multi-indices ONH quantification. Rather than obtaining ONH quantification with single task individually, MCL achieves multi-indices ONH quantification by leveraging the merit of spatial and context dependencies from ONH segmentation and indices estimation tasks with multitask architecture and collaborative learning. Specifically, a dual-branch multitask neural network with feature exchanges and aggregations (FIM) is innovatively designed to collaboratively learn the flexible combination of shared and task-specific features for the modeling of the dependencies. To learn the ensemble of results from two task-specific branches and model their consistency correlation, multitask ensemble module (MEM) is developed by designing a consistency loss function and two-stage aggregation strategy. Benefit from the multitask collaborative learning, the consensus of multiple views from different task-specific branches on the same data provides supplementary information and regularization to each other for accurate multi-indices ONH quantification.

The main contributions of this work are as follows:

- For the first time, Our proposed MCL method provides an accurate and comprehensive solution for ONH quantification and assessment. The proposed method obtains multiple views of quantitative assessment from focal to global to help clinician for diagnosis, treatment, and follow-up of many ophthalmic diseases.
- Our proposed multitask collaborative learning innovatively sets up an effective multitask training strategy for deep neural networks by automatically learning the flexible combination of shared and task-specific features for accurate prediction. The proposed method provides a generalization paradigm for other multitask learning models.
- Our proposed dual-branch multitask neural network (MCL-Net) constructs an effective feature interaction module (FIM) to exchange and aggregate fundus features between the two task-specific branches, and an effective multitask ensemble module (MEM) to leverage the merit of spatial and semantic dependencies between ONH segmentation and indices estimation tasks.

In this work, we advance our preliminary attempt on multi-indices ONH quantification (Zhao et al., 2019) in the following aspects: (1) conduct feature interaction module on the highly correlated tasks to boost the quantification performance by multitask collaborative learning; (2) automate the inter-task feature interaction and consistency correlations modeling with the multitask architecture and collaborative learning, instead of independently single task prediction; (3) carry out more extensive experiments on performance analysis and comparison.

The rest of this paper is organized as follows: In Section 2 we first introduce the related works, and then we give the detailed presentation of our proposed methodology and the algorithm in Section 3. Experimental configurations and dataset details are introduced in Section 4 and results analysis are presented in Section 5. Section 6 concludes the paper.

2. Related works

For the significant clinical application, the ONH quantification has attracted a wide of attention in the past decades. Exiting ONH quantification methods are mainly categorized into segmentation-and-measurement and direct index estimation approaches. In this section, we will review the two existing quantification approaches and multitask learning.

2.1. Segmentation-and-measurement quantification

The segmentation-and-measurement quantification is a type of two-stage method consisted of prerequisite segmentation algorithm and following measurement step. As the key component, the OD and OC segmentation algorithm from fundus images have been independently studied for years. It aims to separate the optic disc and cup with the clear borders or classify the fundus image pixels into OD, OC or background. Early works employ the hand-crafted features for OD and OC segmentation (Xu et al., 2014; Cheng et al., 2015; 2011; 2013; Joshi et al., 2011), however, the OC segmentation is more challenging due to the low contrast boundary. To increase robustness and accuracy, the segmentation algorithms often require the use of prior information from experienced experts and user interaction. That prior information usually comes from anatomical assumptions, such as the ellipse geometry of the OD and OC (Xu et al., 2014; Chen et al., 2015), blood vessel kinks (Wong et al., 2009; Joshi et al., 2011), and so on. Moreover, effective CNN architecture has been designed for OD and OC segmentation, and outperformed traditional hand-crafted feature based methods. Some works (Wang et al., 2019b; Fu et al., 2018; Jiang et al., 2018) were developed for joint OD and OC segmentation by utilizing the structure constraints between OD and OC layout in fundus image.

While the segmentation-and-measurement quantification provides pixel-wise details for ONH assessment, this two-stage quantification limits its effectiveness in the training phase, and the oversimplified prior information prevent leads to inaccurate ONH quantification and prevents OD/OC segmentation algorithms from accurate clinical application.

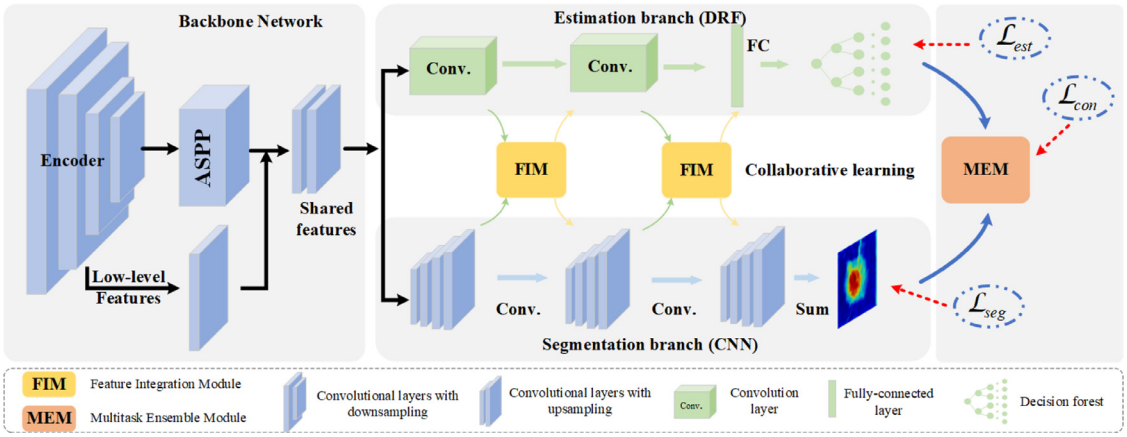


Fig. 3. The MCL-Net is a dual-branch neural network architecture implementing the multitask collaborative learning for multi-indices ONH quantification. For each branch, its intermediate results also serves as guidance for its counterpart, making both branches collaboratively guided. As a result, MCL-Net learns a flexible combination of shared and task-specific representation and model the dependencies between local spatial information and global context.

2.2. Direct index estimation

Direct index estimation becomes popular recently in anatomy quantification (Zhen et al., 2015; 2017; Afshin et al., 2012; Xue et al., 2018; Zhao et al., 2019c; Cheng et al., 2015), which obtains effective performance benefiting from machine learning algorithms. It aims to achieves end-to-end index estimation from the medical image without segmentation by employing the nonlinear regressor. Direct estimation usually follows a two-phase framework as fundus image representation and indices prediction. Fundus images are represented by hand-crafted features (Fernandez-Granero et al., 2017), deep features from multiscale feature pyramid pooling network (Zhao et al., 2019c), a set of reference fundus images (Cheng et al., 2015; 2017). Different clinical indices, such as CDR, disc diameter, compliance of ISNT rule, etc., are then estimated from those features with a robust regression model. Zhao et al., 2019 proposed a semi-supervised learning method to estimate the CDR values directly from fundus images, which learns the deep representation with neural network and regresses the CDR values with random forests. Cheng et al. (2015, 2017) computed the CDRs with the sparse dissimilarity-constrained coding (SDC) approach which considers both the dissimilarity constraint and the sparsity constraint from a set of reference discs with known CDRs.

2.3. Multitask learning

Multitask learning (MTL) is a learning paradigm in machine learning to leverage complementary information contained in multiple related tasks and improve the generalization of all the tasks (Zhang and Yang, 2017; Argyriou et al., 2007). In medical image analysis, many works incorporate multitask learning into deep neural network to deal with the challenges in segmentation (Argyriou et al., 2007; Wang et al., 2019a; Lu et al., 2019), detection (Rojas-Moraleda et al., 2017; López-Linares et al., 2018), localization (Roth et al., 2018) and quantification (Xu et al., 2018). Those methods leverage both powerful feature representation and effective multitask relationship learning to transfer knowledge among these related tasks. For ONH quantification, ONH segmentation and quantification are highly related tasks, and it is suitable well for improvement of multi-indices ONH quantification with the help of the related segmentation task.

Nowadays, existing multitask learning methods mainly deals with two challenges. (1) Some works learn task relationships based on the fixed form, which ignores the collaborative interaction among different tasks. Multitask learning architectures are constructed by sharing several low-level layers with auxiliary tasks including segmentation, detection and quantification, while the task-specific predictions are obtained individually. Xue et al. (2018) proposed a brand-new multitask relationship learning method for full quantification of cardiac left ventricle. Mu et al. (2018) proposed novel multitask strategy based on auxiliary training and geometric constraints for facial landmark detection. (2) Other works focus on the learning of task dependencies by designing flexible architectures of multitask models such as manually-tuned or soft parameter-sharing structures. To improve generalization and robustness, Song and Chai (2018) introduced collaborative learning to deep neural network in which multiple classifier heads are simultaneously and collaboratively trained on the same data. Xu et al. (2018) proposed a multitask generative adversarial network for joint segmentation and quantification of myocardial infraction. Ma et al. (2019); Misra et al. (2016) have shown that having more flexible architectures in multitask models to help improve the prediction accuracy.

Different from existing multitasks learning, we propose a flexible multitask framework of Multitask Collaborative Learning (MCL) to achieve multi-indices ONH quantification. The proposed MCL learns the merit of spatial and context dependencies between OD segmentation and indices estimation tasks. Rather than sharing the low-level layers and feature representation, MCL framework automatically learns an optimal combination of shared and task-specific representation from both tasks and aggregates those features using the proposed feature interaction module. Acquiring the advantages from collaborative training and flexible feature combination, MCL refines details and context semantics of ONH to achieve robust multi-indices ONH quantification.

3. Proposed methodology

The proposed multitask collaborative learning (Fig. 3) is a dual-branch neural network framework (MCL-Net) with a backbone network and two branches for features representation and task-specific predictions. Different from existing models, two well-designed modules are inserted for collaborative learning and multi-indices ONH quantification. Specifically, the feature interaction module (FIM) is innovatively designed for feature exchange and aggregation between the two task-specific branches, and the multitask ensemble module (MEM) is developed for quantification aggregation by the consistency loss function and ensemble learning. By training two highly related branches collaboratively, the

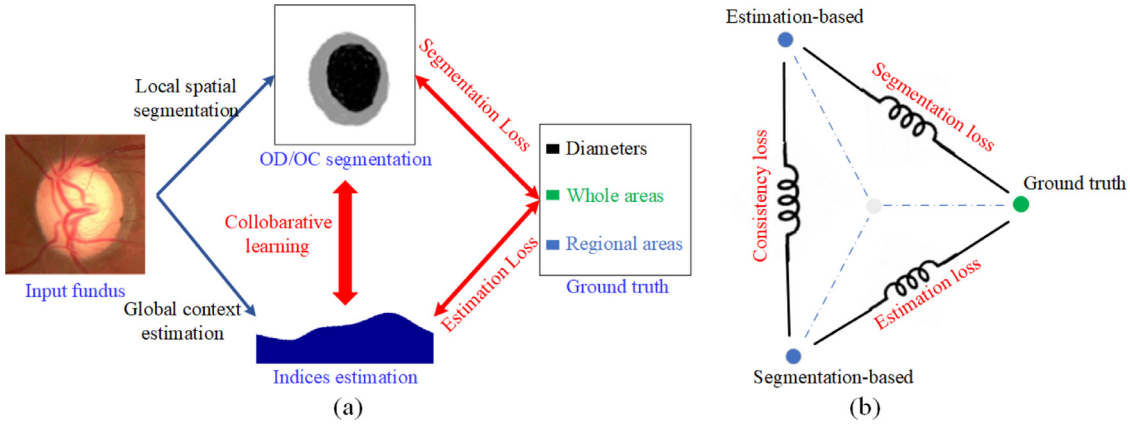


Fig. 4. Multitask collaborative learning between segmentation and estimation. (a) The consistent and complementary collaboration between individual segmentation and estimation tasks is modeled in an unified framework. (b) Multitask collaborative loss function is constructed to collaboratively learn optimal model that is capable of capture the more details of ONH quantification by providing supplementary information and regularization to each others.

MCL-Net learns the dependencies and complements from the two tasks to achieve accurate multi-indices ONH quantification.

3.1. Multitask collaborative learning (MCL)

The multitask collaborative learning (MCL) conforms to the feature-based multitask learning, which learns multiple task-specific branches collaboratively to improve the accuracy of multi-indices ONH quantification. The collaborative learning enhances and refines the quantification details with the flexible combination of shared and task-specific feature representations. To fully leverage the benefit of complementary information and improve the semantic consistency among different tasks, instead of independently capturing the task-specific feature, the proposed MCL conducts a collaborative feature exchange and aggregation framework by adding the module of FIM and MEM between segmentation and estimation branches. Taking advantages of FIM and MEM, MCL promotes performance improvement on accurate and robust multi-indices ONH quantification by multitask collaborative learning.

We study the task of multitask collaborative learning to deal with multi-indices ONH quantification, where the segmentation and estimation branches are trained in a unified multitask learning framework (as shown in Fig. 4(a)). The goal of our multitask collaborative learning is not only to train an optimized segmentation and estimation tasks, but also to learn the consistent and complementary collaboration of the two tasks for accurate multi-indices ONH quantification. Hence, the overall objective of the multitask collaborative learning is

$$\mathcal{L}_{total} = \alpha \underbrace{\mathcal{L}_{seg}(\mathcal{W})}_{segmentation\ \ loss} + \beta \underbrace{\mathcal{L}_{est}(\mathcal{W}, y_{est})}_{estimation\ \ loss} + \gamma \underbrace{\mathcal{L}_{con}(y_{est}, y_{seg})}_{consistency\ \ loss} \quad (1)$$

where $\mathcal{L}_{seg}$ is the segmentation loss to ensure the local spatial layout of optic disc and rim be encoded by the segmentation branch, $\mathcal{L}_{est}$ is the estimation loss to provide the global context information for the structure and shapes of optic disc by the estimation branch, and $\mathcal{L}_{con}$ denotes the consistency loss for measuring the prediction consistency between the segmentation and estimation branches, which imposes a complementary regularization on the multitask learning framework. Hence, $\mathcal{W}$ is the parameters of the model, y_{seg} is the derived indices vector based on the outputs of segmentation branch, and y_{est} is the predicted indices vector by the estimation branch.

The collaborative relationship among the three terms in Eq. (1) is demonstrated in Fig. 4(b), where segmentation loss aims to learn the pixel-level local spatial details of optic disc such as contours, edges, while estimation loss ensures the global structure and context to be encoded. The segmentation and estimation branches capture the complementary consensus of multiple views on the same data for accurate multi-indices ONH quantification. As shown in Fig. 4(b), dependencies and complementarity exists between local pixel-level segmentation and global index-level estimation. Although the two tasks have different optimization branches, they have the unified and consistent optimization results on multi-indices ONH quantification. Hence, our multitask collaborative learning is proposed to achieve accurate multi-indices ONH quantification by collaborative learning of the dependencies and complementary between the segmentation and estimation branches. To this end, we state the collaborative learning based on threefold: (1) The ground truth and segmentation-based results should be minimization, *i.e.,* $\min \mathcal{L}_{seg}$; (2) The ground truth and estimation-based results should be minimization, *i.e.,* $\min \mathcal{L}_{est}$; (3) The segmentation-based and estimation-based results should be consistent, *i.e.,* $\min \mathcal{L}_{con}$.

3.2. Dual-branch neural network architecture (MCL-Net)

The proposed MCL-Net (Fig. 3) contains an adapted DeepLab as backbone network for shared feature generation, but two task-specific decoder branches for task-specific feature selection and prediction. For each branch, its representation also serves as guidance for its counterpart, making both branches collaboratively guided. The MCL-Net allows for learning of the inter-task correlation by shared feature representation, whilst simultaneously allows for modeling of semantic dependencies and consistency for multi-indices ONH quantification between different task-specific branches by collaborative feature interaction.

3.2.1. Segmentation branch for ONH segmentation-based quantification

Segmentation branch formulates OD and OC (optic cup) segmentation as the multi-label pixel-level classification problem to learn the distribution of OD and OC region. To improve the segmentation for ONH quantification, we develop a novel distribution-aware segmentation loss to guide the segmentation decoder to capture the smoothness priors of the OD and OC region. The segmentation loss includes a dice coefficient loss $\mathcal{L}_{dice}$ measuring the overlap between the prediction and ground truth, and a distribution loss $\mathcal{L}_{dist}$ encouraging the predictive borders of OD and OC regions to be similar to the ground truth. Therefore, the distribution-

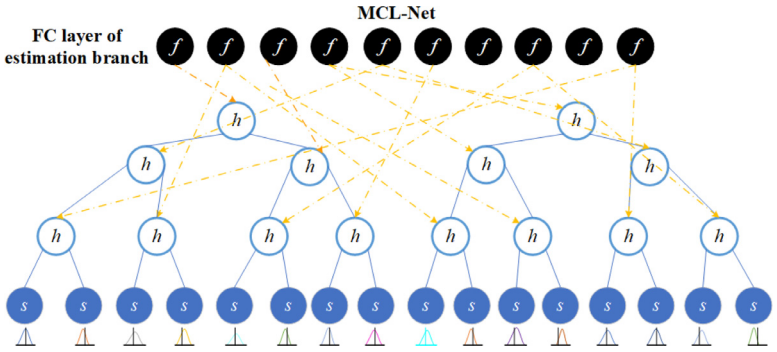


Fig. 5. Distribution regression forests(DRF). The top solid circles denote the output unit of FC layers of estimation branch, the solid blue circles is leaf nodes which hold a predication distribution **s** as shown its blow. The hollow blue circles denote split nodes, which define a split function *h* and randomly connected with the output of FC layers as the yellow dash. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

aware segmentation loss is defined as

$$\mathcal{L}_{seg}(\mathcal{W}) = \underbrace{1 - \frac{2\sum_i p_i y_i}{\sum_i p_i^2 + \sum_i y_i^2}}_{\mathcal{L}_{dice}} + \underbrace{\sum_c d_c \log(s_c)}_{\mathcal{L}_{dist}} \tag{2}$$

where p and y denote the predicted probability map and ground truth, respectively. s and d denote the predicted and ground truth of border pixels, and c is the number of the border pixels.

3.2.2. Estimation branch for regression-based quantification

We adopt a newly-designed model named Distribution Regression Forests (DRF, Fig. 5) in the estimation branch to achieve indices estimation by constructing a multitude of differentiable decision tree linked with the backbone network by convolutional and fully-connected layers. The estimation branch is consisted of two convolution layers, one fully-connected layer and one DRF module. The convolution layers are used to refine task-specific feature from the shared representation, meanwhile, exchange and aggregate information from segmentation branch. The fully-connected layer is used to transform the multi-channel feature maps into vectors for inputs of DRF module. The DRF not only allows for feature selection for the specific task from the shared representation, but also allows to be trained with other components as the end-to-end manner. The DRF module consists of a set of split nodes $\mathcal{N}$ and leaf nodes $\mathcal{L}$, which construct a multitude of decision trees at training time and output the prediction of the individual branch.

Mathematically, the aim of estimation branch (DRF) is to learn a mapping function $\mathbf{g}$**:** $\mathbf{x} \rightarrow \mathbf{d}$ from the shared features to its corresponding target distribution of the quantitative indices by a set of split nodes $\mathcal{N}$ and leaf nodes $\mathcal{L}$. Each split node $n \in \mathcal{N}$ defines a split function $h_n(\mathbf{x}; \mathbf{w}) : \mathcal{R}^m \times \mathcal{W} \rightarrow \{0, 1\}$, characterized by $\mathbf{w} \in \mathcal{W}$ and used to route the input into right or left sub-tree, while each leaf node $l \in \mathcal{L}$ holds a predication distribution $\mathbf{s}_l$. When a sample $\mathbf{x} \in \mathcal{X}$ reaches a split node n it will be sent to the left or right subtree based on the split function $h_n(\mathbf{x}; \mathbf{w})$. To enable the tree with differentiable split function and probabilistic routing, first a routing function $\mu_l(\mathbf{x}|\mathbf{w})$ is defined to provide the probability that sample $\mathbf{x}$ will reach leaf node l as

$$\mu_l(\mathbf{x}|\mathbf{w}) = \prod_{n \in \mathcal{N}} h_n(\mathbf{x}; \mathbf{w})^{\mathbb{1}(l \in \mathcal{L}_n^{left})} (1 - h_n(\mathbf{x}; \mathbf{w}))^{\mathbb{1}(l \in \mathcal{L}_n^{right})} \tag{3}$$

where $\mathbb{1}(\cdot)$ is an indicator function, $\mathcal{L}_n^{left}$ and $\mathcal{L}_n^{right}$ denote the sets of leaf nodes held by the subtrees rooted at the left and right children *left, right* of node *n*. $h_n(\mathbf{x}; \mathbf{w})$ indicates the probability that split node $n \in \mathcal{N}$ selects input feature $\mathbf{x}$ as its task-specific feature.

To enable the forest with the capability to be optimized end-to-end together with the backbone network, the differentiable split function is defined as

$$h_n(\mathbf{x}; \mathbf{w}) = \sigma(f_{\varphi(n)}(\mathbf{x}; \mathbf{w})) \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function, $f_{\varphi(n)}(\mathbf{x}; \mathbf{w})$ is outputs of the network, adopted as the shared feature extraction function to end-to-end learn the expressive representation of fundus image. $\varphi(\cdot)$ is an index function to assign the connection between the output of function $f(\mathbf{x}; \mathbf{w})$ and split node n. In this work, the index function $\varphi(\cdot)$ is a random function to link split nodes with the shared root network randomly. Therefore, the final mapping function $\mathbf{g(y|x; w)}$ is obtained by weighting all the holding distribution of the probability of reaching the leaf

$$\mathbf{g}(\mathbf{y}|\mathbf{x}; \mathbf{w}) = \sum_{l \in \mathcal{L}} \mu_l(\mathbf{x}|\mathbf{w}) \mathbf{s}_l \tag{5}$$

Given the training set $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{d}_{\mathbf{x}_i})\}_{i=1}^N$, where $\mathbf{d}_{\mathbf{x}_i}$ is the distribution generated by Gaussian distribution whose mean is the chronological indices, our goal is to learn a distribution regression forests which can output the distribution $\mathbf{g}(\mathbf{y}|\mathbf{x}_i; \mathbf{w})$ similar to $\mathbf{d}_{\mathbf{x}_i}$ for each sample *i*. In this work, the Kullback-Leibler (K-L) divergence is adopted to measure the similarity between predicted distribution $\mathbf{g}(\mathbf{y}|\mathbf{x}_i; \mathbf{w})$ and ground truth $\mathbf{d}_{\mathbf{x}_i}$. Therefore, the learning procedure is minimizing the following cross-entropy loss

$$\mathcal{L}_{est}(\mathcal{W}, y_{est}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} d_{\mathbf{x}_i}^c log(\sum_{l \in \mathcal{L}} \mu_l(\mathbf{x}_i|\mathbf{w}) \mathbf{s}_l) \tag{6}$$

where y_{est} denotes the estimated indices held by all the leaf nodes of this forest, $d_{\mathbf{x}_i}^c$ is the probability of the sample $\mathbf{x}_i$ having the c-th ground truth.

3.3. Feature interaction module for inter-task features exchange and aggregation

Given that multiple indices quantitatively describe the corresponding ONH appearance from local to global, it is intuitive that local spatial cues and global semantics of ONH have high dependencies, which is helpful for decoders to refine details and focus on learning informative patterns. Collaboratively, the neural activation from the specific kernel in segmentation branch can be considered as an extra spatial cue for regressing ONH indices. In this regard, we propose a new unit, *Feature Interaction Module (FIM)*, that utilizes features from the highly related branches to model the dependencies by learning an aggregation of the task-specific features.

Given feature maps x_S, x_E from segmentation and estimation branches, we learn the aggregation $\tilde{x}_S$, $\tilde{x}_E$ and then feed these aggregated features as input to the next layers (Fig. 3). The exchange and aggregation are parameterized using a nonlinear transformation matrix Λ, which represents the feature transformation between different branches and different layers of the same branch. Specifically, at location (i, j), the feature exchange and aggregation

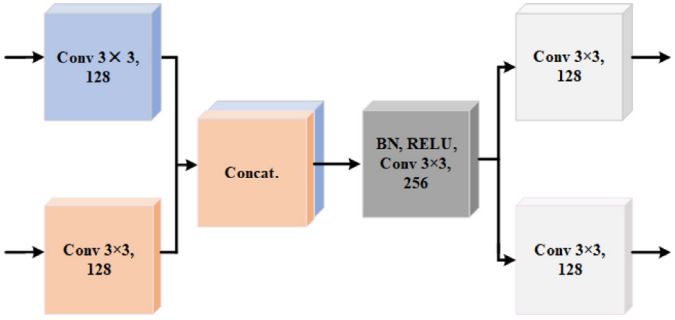


Fig. 6. Detail structure of Feature Interaction Module in the our MCL-Net, where a CNN-based module is implemented containing convolution and concatenation unit to achieve the function of nonlinear transformation Λ represented in Eq. (8).

are represented as

$$\begin{bmatrix} \tilde{x}_S^{ij} \\ \tilde{x}_E^{ij} \end{bmatrix} = \begin{bmatrix} \Lambda_{SS} & \Lambda_{SE} \\ \Lambda_{ES} & \Lambda_{EE} \end{bmatrix} \begin{bmatrix} x_S^{ij} \\ x_E^{ij} \end{bmatrix} \tag{7}$$

We aim to exchange and aggregate the different information between the two branches by learning the optimizated matrix Λ, where Λ_{SS}, Λ_{EE}, Λ_{SE}, Λ_{ES} represents the feature transformation from segmentation to segmentation branch, from estimation to estimation branch, from segmentation to estimation branch, and from estimation to segmentation branch, respectively.

In the training procedure, the partial derivatives for loss function can be computed as

$$\begin{bmatrix} \frac{\partial L}{\partial x_S^{ij}} \\ \frac{\partial L}{\partial x_E^{ij}} \end{bmatrix} = \begin{bmatrix} \Lambda_{SS} & \Lambda_{SE} \\ \Lambda_{ES} & \Lambda_{EE} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \tilde{x}_S^{ij}} \\ \frac{\partial L}{\partial \tilde{x}_E^{ij}} \end{bmatrix} \tag{8}$$

$$\frac{\partial L}{\partial \Lambda_{SE}} = \frac{\partial L}{\partial \tilde{x}_E^{ij}} x_S^{ij}, \quad \frac{\partial L}{\partial \Lambda_{SS}} = \frac{\partial L}{\partial \tilde{x}_S^{ij}} x_S^{ij} \tag{9}$$

Because of the linear combination of feature aggregation in Eq. (7), we train the FIM with the MCL-Net with the backpropagate algorithm for the optimized Λ.

We use the FIM unit for collaborative learning in the dual-branch neural network. Fig. 6 shows the details of FIM structure, it takes features from two branches as inputs, and then implements the convolution and concatenation operators as Λ shown in Eq. (8) to exchange and aggregate informative features for each task. In this way, the two branches interact and exchange to directly supervise their counterpart. In other words, the FIM unit servers as a regularization to help regularize both branches to obtain consistent ONH quantification.

3.4. Multitask ensemble module for final ONH quantification

To learn the ensemble of results from two branches and model their consistency correlation, multitask ensemble module is developed, which contains a consistency loss function to impose the penalty for the consistent multi-indices ONH quantification between segmentation and estimation branches, and a two-stage aggregation for final ONH quantitative indices. Consistency loss is designed to minimize the prediction difference between two branches, i.e., ONH segmentation and indices estimation tasks. Ideally, indices predicted by the two branches are the same. To ensure the indices from different branches as consistent as possible, the consistency loss is defined as the difference between the indices vectors

$$\mathcal{L}_{con} = \frac{1}{2}(y_{est} - y_{seg})^2 \tag{10}$$

where y_{est} and y_{seg} denote indices vectors coming from estimation branch and ONH segmentation branch, respectively.

To integrate predictions from each leaf node of the two task branches, two-stage aggregation is adopted. 1) Intra-task aggregation: with the prediction on each leaf nodes of DRF, the task-specific quantitative indices are obtained by aggregating those leaf nodes predictions into a single coherent output followed ensemble learning as $\sum_{l\in\mathcal{L}} \mu_l(\mathbf{x}|\mathbf{w})\mathbf{s}^l$, where $\mu_l(\mathbf{x}|\mathbf{w})$ is the probability that feature $\mathbf{x}$ be selected by leaf node l and defined in Eq. (3), $\mathbf{s}^l$ is the predicted indices vector on leaf node l. Note that $\mathbf{s}^l$ is measured based on the segmented mask when node l belongs to ONH segmentation branch, while directly regressed when belongs to indices estimation branch. 2) Inter-task aggregation: with the prediction of each task-specific branch, the final quantitative indices are build based on the simple yet effective adaptive weighting method. MCL-Net learns to average task weighting by considering the loss for each task, and the task weighting for segmentation and estimation tasks are defined as:

$$\alpha_{seg} = \frac{\exp(\mathcal{L}_{seg}^t)}{\exp(\mathcal{L}_{seg}^t) + \exp(\mathcal{L}_{est}^t)}, \quad \alpha_{est} = \frac{\exp(\mathcal{L}_{est}^t)}{\exp(\mathcal{L}_{seg}^t) + \exp(\mathcal{L}_{est}^t)} \tag{11}$$

where $\mathcal{L}_{seg}^t$ and $\mathcal{L}_{est}^t$ are the average loss from segmentation and estimation task branches in the *t-th* epoch over several iterations.

3.5. Algorithm of MCL-Net

Given a training set $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, learning a multitask collaborative learning model described in Section 3.1 leads to minimizing the overall objective (Eq. (1)). The training procedure requires estimating the parameters $\mathcal{W}$ both in two branches and shared backbone. We optimize the segmentation branch and estimation branch in an alternate way. To optimize the individual branch, we minimize the objective function in Eqs. (2) and (6), respectively. To optimize the MCL-Net collaboratively, we calculate the consistency loss $\mathcal{L}_{con}$, $\mathcal{L}_{seg}$ and $\mathcal{L}_{est}$ in Eq. (1), *i.e.*,

$$(\mathcal{W}^*, \Lambda^*) = \arg\min_{\mathcal{W},\Lambda} \mathcal{L}_{total} \tag{12}$$

Algorithm 1 summarizes the detailed procedure of training for our MCL-Net. The algorithm follows the two-stage optimization

Algorithm 1 Training procedure of MCL-Net.

Input: Fundus images: $\mathcal{X}$ and its corresponding ground truth (segmentation mask M ONH indices $\mathcal{Y}$); learning rate η, constants α, β, γ;
Output: Parameters of MCL-Net: $\mathcal{W}$; the quantitative ONH indices: $\hat{\mathcal{Y}}$

```
Initialize Λ, W;
/* first stage
while iter < max-iter do
    while iter1 < max-iter1 do
        Update W as W ← W − η∇wL_seg by optimizing the segmentation objective (Eq. (2));
    end while
    while iter2 < max-iter2 do
        Update W as W ← W − η∇wL_est by optimizing the estimation objective (Eq. (6));
    end while
end while
/* Second stage
while L_total not converged do
    Update W with back propagation from L_total (Eq. (1));
    Update Λ with (Eq. (9));
end while
Update Y with (Eq. (11))
return W, Λ
```

strategy which contains two steps: (1) fixing Λ and optimizing individual branches $\mathcal{W}$ alternatively; (2) jointly optimizing overall objective Λ, $\mathcal{W}$ and updating quantitative indices, until convergence or a maximum number of iterations is reached. In both stages, the parameters learning of the neural network can be optimized through the usually utilized iterated stochastic gradient descent (SGD) procedure.

4. Experiments

4.1. Datasets

In this section, we evaluate the effectiveness of our MCL-Net on two widely used datasets, ORIGA (Cheng et al., 2017) and REFUGE challenge dataset.[1] The ORIGA contains 650 images (168 glaucomatous and 482 normal eyes) with manual labeled optic disc mask, divided into 325 training and 325 testing images. REFUGE challenge dataset contains 400 images with manual pixel-wise annotations of the optic disc and cup contours with seven independent glaucoma specialists from Zhongshan Ophthalmic Center, divided into 200 training and 200 testing images.

All the fundus images undergo several preprocessing steps, including ROI cropping and resizing. The input images of MCL-Net are aligned with the dimension of 512×512 centering at OD. Then the ground truth values of 20 indices are obtained by measuring the borders of optic disc, cup and rim. During the evaluation, the obtained results are evaluated by the pixel number.

To leverage the powerful representation for the circle-shaped ONH appearance, followed as (Fu et al., 2018), all input fundus images are pixel-wisely converted into the polar coordinate system and then re-sampled. Let $X(\theta,\ \rho)$ denotes the point on converted fundus image, where $\theta,\ \rho$ are the directional angle and radius in the polar coordinates, respectively. Pixels in optic disc region are re-sampled along the angular and radius dimension, therefore resulting in the regions of OC, OD and background in the ordered layout. In this work, ρ equals half of the disc region size, θ is 360 for the fundus image.

4.2. Configurations

The adapted DeepLabv3+ (Chen et al., 2018) is adopted as the backbone network of the MCL-Net where the Xception is replaced by the lightweight MobileNetV2 (Sandler et al., 2018) to reduce the number of parameters. Then we add the segmentation and estimation branches, FIM and MEM modules into the backbone network. The estimation branch consists of two convolution layers with the output channel of {256, 256} and one fully connected layers followed by a DRF unit. The DRF contains two binary trees and sets the hyper-parameters of DRF as the previous work (Shen et al., 2018). $\varphi(\cdot)$ in Eq. (4) assigning the connection between the output of fully connected layer and the split node is randomly generated before forest learning, which means that the connection between the neuron of FC layer from neural network and split nodes are randomly generated. The segmentation branch is a decoder neural work with three blocks of convolutional layers followed by ReLU and batch normalization.

Our MCL-Net is implemented with the Caffe library, using NVIDIA Tesla P100 Server Graphics Cards. Our network is initialized with the pretrained wighted. We adopt the alternating optimization strategy to obtain the optimistic parameters of DRF and the SGD algorithm for the whole MCL-Net. Other hyper-parameters are fixed as follows: $\alpha = 0.8,\ \beta = 0.8,\ \gamma = 0.5$.

[1] https://refuge.grand-challenge.org/Home/

4.3. Evaluation criteria

In this work, ONH indices obtained from manually annotated borders by ophthalmologists are adopted as the gold standard (ground truth) for ONH quantification. We evaluate the proposed MCL-Net in terms of correlation coefficient ρ and mean absolute error (MAE) between the ground truth values and the computed one as

$$MAE(y, \bar{y}) = \frac{1}{N}\sum_{i=1}^{N}|y_i - \bar{y}_i| \tag{13}$$

$$\rho(y, \bar{y}) = \frac{2\sum_{i=1}^{N}(y_i - m)(\bar{y}_i - \overline{m})}{\sum_{i=1}^{N}(y_i - m)^2 + \sum_{i=1}^{N}(\bar{y}_i - \overline{m})^2} \tag{14}$$

where $m = \frac{1}{N}\sum_{i=1}^{N} y_i,\ \overline{m} = \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i$

Statistical significance of the proposed method versus ground truth is examined by the paired t-test with significance level of 0.1%. The p-value for each pair of measurements are computed to demonstrate the significance improvement of our proposed MCL-Net. A lower p-value than 0.001 indicates that the method achieves the quantification of the corresponding ONH index with no significant differences.

In addition, the performance for glaucoma screening is evaluated based on the computed ONH indices. We report the receiver operating characteristic (ROC) curve and area under the curve (AUC) as the overall measure of the screening strength. To plot the ROC curve and compute AUC value, the obtained indices are first inputted into a simple classifier (such as SVM) and the subjects with a larger output are regarded as glaucomatous suspect. Then we calculate the false positive rate and true positive rate. By changing the threshold value, the ROC curve is obtained and AUC value is computed.

5. Results and analysis

The effectiveness of the proposed MCL-Net for multi-indices ONH quantification is validated in three folds. (1) The quantification performance is examined for the three types of ONH indices. (2) The effectiveness of each component in our MCL-Net is probed to demonstrate its capacity in robust multi-indices ONH quantification. (3) The advantages of the proposed MCL-Net over existing methods on ONH quantification and glaucoma screening are revealed compared with state-of-the-art methods, including the single segmentation-and-measurement methods and ensemble-learning-based method.

5.1. Performance of multi-indices ONH quantification

As shown in the last row of Table 1, MCL-Net successfully delivers accurate quantification of all the three types of ONH indices by leveraging the merit of segmentation and regression tasks with multitask collaborative learning. Specifically, MCL-Net achieves average MAE of 0.98 ± 0.20, 0.97 ± 0.16, 1.19 ± 0.18, as well as average correlation coefficient of 0.699, 0.708 and 0.691, for diameters, whole areas and regional areas, respectively, which is more accurate than other single-task-based approaches with the lowest average MAE over all the 20 quantitative indices. Experimental results on different datasets (shown in Table 2) demonstrated that our MCL-Net achieves robust multi-indices ONH quantification with a low average MAE for diameters, regional areas and whole areas. We have the observation that the MCL-Net has better performance on ORIGA dataset than those on REFUGE dataset. One possible reason is that the larger size of training set on ORIGA dataset promotes the learning capability of deep learning.

The areas indices (both regional and whole) involves both the optic disc and cup contours, which are challenged to be localized,

Table 1
Performance of MCL-Net under different configurations for multi-indices ONH quantification versus ground truth. Average Mean Absolute Error (MAE), correlation coefficient ρ and p-value for paired t-test with significance level of 0.1% are used for the quantification evaluation criterion. Here each cell contains MAE on the top and ρ/p-value on the bottom.

Method	Segmentation branch	Estimation w/ DRF	Estimation w/o DRF	two-task Ensemble	collaborative + Ensemble
Diameter					
Dia1	0.97 ± 0.25	0.97 ± 0.29	1.23 ± 0.64	0.97 ± 0.29	0.94 ± 0.23
	0.605/<.001	0.6089/<.001	0.5214/<.001	0.6104/<.001	0.6408/<.001
Dia2	1.07 ± 0.21	1.05 ± 0.25	1.51 ± 0.60	1.05 ± 0.25	1.03 ± 0.18
	0.7158/<.001	0.7192/.001	0.5185/<.001	0.7203/<.001	0.7563/<.001
Average	1.02 ± 0.23	1.01 ± 0.27	1.37 ± 0.62	1.01 ± 0.27	**0.98 ± 0.20**
	0.6604	0.6640	0.5199	0.6809	**0.6985**
Whole areas					
WR	0.32 ± 0.16	0.32 ± 0.15	0.57 ± 0.29	0.32 ± 0.15	0.31 ± 0.08
	0.3921/<.001	0.3901/.002	0.2497/.003	0.3930/.001	0.4561/<.001
WD	1.64 ± 0.17	1.65 ± 0.20	1.89 ± 0.42	1.65 ± 0.20	1.63 ± 0.18
	0.9425/<.001	0.9442/<.001	0.7957/.005	0.9597<.001	0.9607/<.001
Average	0.98 ± 0.16	0.98 ± 0.17	1.23 ± 0.35	0.98 ± 0.17	**0.97 ± 0.13**
	0.6673	0.6671	0.5227	0.6763	**0.7084**
Regional areas					
RR1	0.73 ± 0.25	0.75 ± 0.23	0.96 ± 0.49	0.73 ± 0.21	0.72 ± 0.20
	0.3114/.239	0.3121/.254	0.1944/.302	0.3189/.284	0.4210/.240
RR2	0.35 ± 0.14	0.34 ± 0.13	0.47 ± 0.25	0.34 ± 0.10	0.31 ± 0.12
	0.3584/.107	0.3658/.114	0.2151/.300	0.3763/.127	0.4852/.103
RR3	0.11 ± 0.11	0.12 ± 0.10	0.21 ± 0.19	0.11 ± 0.09	0.09 ± 0.10
	0.3188/<.001	0.3159/<.001	0.2654/.022	0.3189/<.001	0.5012/<.001
RR4	0.55 ± 0.35	0.55 ± 0.32	0.68 ± 0.48	0.55 ± 0.29	0.52 ± 0.28
	0.3792/<.001	0.3756/<.001	0.2978/.004	0.3763/<.001	0.5615/<.001
RR5	0.65 ± 0.31	0.68 ± 0.34	0.89 ± 0.41	0.69 ± 0.21	0.65 ± 0.20
	0.4254/<.001	0.4210/.001	0.3001/.001	0.4215/<.001	0.5110/<.001
RR6	0.28 ± 0.20	0.31 ± 0.14	0.45 ± 0.29	0.30 ± 0.13	0.27 ± 0.13
	0.3793/.010	0.3721/.010	0.2674/.015	0.3800/.011	0.5232/.003
RR7	0.12 ± 0.31	0.15 ± 0.25	0.22 ± 0.40	0.13 ± 0.24	0.11 ± 0.25
	0.3856/.001	0.3854/.001	0.3321/.002	0.3910/.001	0.5998/.001
RR8	0.25 ± 0.30	0.25 ± 0.22	0.31 ± 0.38	0.25 ± 0.18	0.24 ± 0.18
	0.4564/.025	0.4555/.034	0.4220/.039	0.4590/.030	0.5184/.026
Average	0.38 ± 0.24	0.39 ± 0.21	0.52 ± 0.36	0.38 ± 0.18	**0.35 ± 0.18**
	0.3765	0.3754	0.2870	0.3802	**0.5151**
DR1	1.27 ± 0.12	1.25 ± 0.23	1.52 ± 0.33	1.26 ± 0.12	1.24 ± 0.10
	0.8625/<.001	0.8605/<.001	0.6957/<.001	0.8619/<.001	0.8694/<.001
DR2	3.18 ± 0.54	3.29 ± 0.53	3.64 ± 0.87	3.21 ± 0.46	3.16 ± 0.33
	0.8399/<.001	0.8385/<.001	0.7354/.001	0.8402/<.001	0.8457/<.001
DR3	3.74 ± 0.41	3.78 ± 0.39	4.02 ± 0.84	3.78 ± 0.29	3.74 ± 0.27
	0.8345/<.001	0.8321/<.001	0.7289/<.001	0.8343/<.001	0.8541/<.001
DR4	3.27 ± 0.23	3.29 ± 0.22	3.99 ± 0.85	3.30 ± 0.20	3.25 ± 0.18
	0.8701/<.001	0.8694/<.001	0.7021/<.001	0.8733/<.001	0.8944/<.001
DR5	2.58 ± 0.34	2.57 ± 0.23	3.01 ± 0.68	2.57 ± 0.18	2.55 ± 0.17
	0.8754/<.001	0.8715/<.001	0.6649/<.001	0.8768/<.001	0.8841/<.001
DR6	1.73 ± 0.21	1.77 ± 0.32	2.34 ± 0.89	1.76 ± 0.18	1.73 ± 0.18
	0.8521/<.001	0.8500/<.001	0.7001/<.001	0.8520/<.001	0.8610/<.001
DR7	0.64 ± 0.21	0.60 ± 0.24	0.99 ± 0.55	0.61 ± 0.17	0.60 ± 0.17
	0.8433/<.001	0.8409/<.001	0.7118/<.001	0.8441/<.001	0.8498/<.001
DR8	0.05 ± 0.18	0.06 ± 0.06	0.18 ± 0.19	0.06 ± 0.05	0.05 ± 0.05
	0.8721/<.001	0.8659/<.001	0.7456/<.001	0.8718/<.001	0.8821/<.001
Average	2.05 ± 0.28	2.07 ± 0.27	2.46 ± 0.65	2.06 ± 0.20	**2.04 ± 0.18**
	0.8562	0.8536	0.7105	0.8568	**0.8675**

Table 2
Performance of MCL-Net with different datasets.

Dataset	ORIGA	REFUGE
Diameter (10^2 pixel)		
Average	0.98 ± 0.20	0.99 ± 0.22
	0.6985	0.6854
Whole areas (10^4 pixel)		
Average	0.97 ± 0.16	1.01 ± 0.15
	0.7084	0.6825
Regional areas (10^4 pixel)		
Average	1.19 ± 0.18	1.24 ± 0.22
	0.6913	0.6784

especially optic cup border. Especially, the regional areas (Fig. 1) in the glaucomatous subject are often very small and easy to be misestimated. In clinical assessment of glaucoma, there exists large variation of regional areas in different direction across patients, which often disturb accurate ONH quantification, therefore, it is more difficult to achieve than others. However, MCL-Net can still obtain a low average MAE of 1.19 ± 0.18 for regional areas and 0.97 ± 0.16 for whole areas, as well as remain a high correlation coefficient of 0.691 and 0.708.

Fig. 7 shows the correlations between the computed indices values with the ones obtained from the manual labels. Ideally, the distribution of indices plots as close as possible to the blue straight line. Despite the challenges in multi-indices ONH quantification, the proposed MCL-Net achieves a high correlation between the

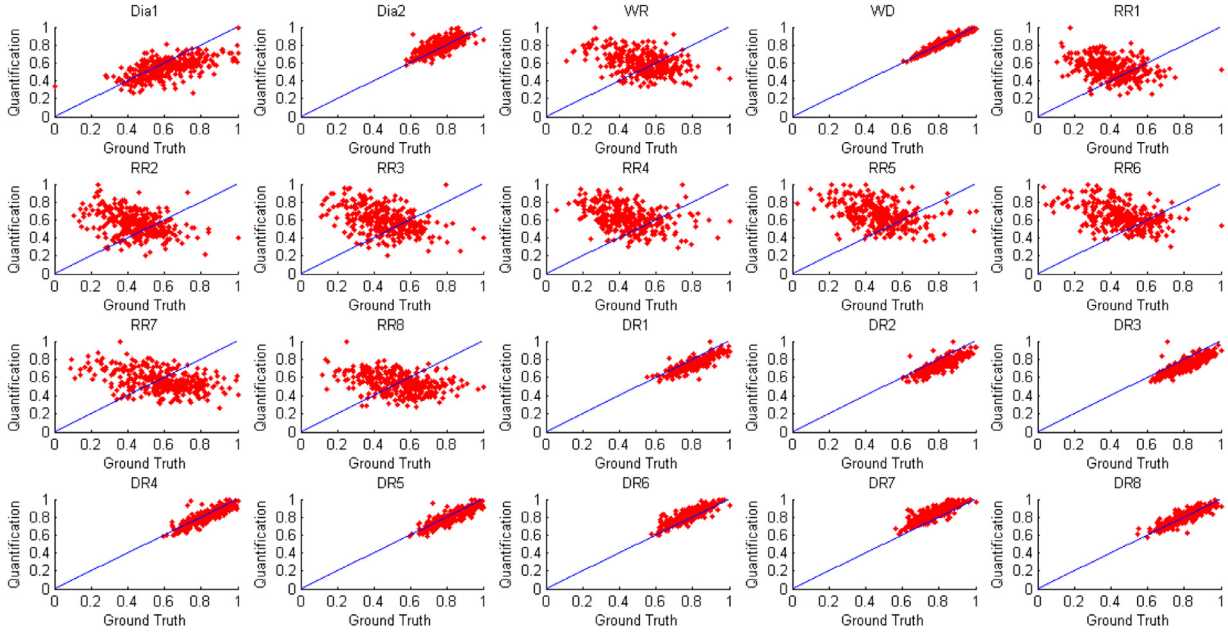


Fig. 7. The correlations between the computed indices values with the ones obtained from the manual labels. Despite the challenges in ONH quantification, especially in rim region, the proposed method achieves a high correlation between the computed indices and manual ones. It should be noted that the plots in 5–12 subfigure is not good as the others because of the challenges in rim quantification, however, the correlations obtained by our proposed method still higher than the existing methods.

computed indices and manual ones. The plots shown in Fig. 7 are distributed around the blue straight line which means the computed indices by our method are highly correlated with the ground truth. It should be noted that the plots in 5–12 subfigure is not good as the others because of the challenges in rim quantification, however, the correlations obtained by our proposed method still higher than the existing methods.

Furthermore, experimental results on glaucoma screening show that the quantitative indices provide more effective assessment tools (with 0.8698 AUC) for ophthalmic diseases screening. The effectiveness of glaucoma screening benefits from comprehensive quantification of 20 indices, which provides both the detailed and semantic information for ONH assessment.

5.2. Ablation study

We conduct some ablation experiments by comparing against the baselines to analyze the effectiveness of each components in our MCL-Net. From the ONH quantification results shown in Table 1, we have the following observations:

Effectiveness of DRF. We conduct a contrast experiment to demonstrate the effective of our newly-designed DRF module for multi-indices ONH quantification. In the contrast experiment, out proposed MCL-Net under different configurations (with and without DRF module) are evaluated to show the contribution of DRF module. The third and fourth columns obviously indicate that the DRF module improves the average 22.3% of 20 indices compared with the baseline model without benefits of the DRF module. The results show that the DRF module contributes to the 3 types and 20 indices ONH quantification due to its advantages of feature selection and ensemble learning.

Effectiveness of multitask ensemble. The first three columns of Table 1 demonstrate that the multitask ensemble produces a competitive performance (MAE=1.13 ± 0.18), which is better than single task solutions (MAE $=1.18 \pm 0.23$). Two single-task models serve as baselines without benefits of multitask ensemble. First we evaluate a single network trained on only one task (denoted as *segmentation branch* and *estimation branch*). The model settings are the same as the ones described in Section.3, whereas the other branch, FIM and MEM are discarded. Indices shown in the first and second columns of Table 1 are independently obtained by the single-task baselines. The results clearly indicate that multitask ensemble improves average 0.5%, 0.8% of 20 indices compared with the single segmentation and estimation task branch, respectively. Compared with the single task, multitask ensemble obtains the smallest bias overall indices and lowest average MAE overall 20 indices. The average MAE and bias show multitask ensemble learning brings clearly improvements for all the indices quantification. This proves that multitask ensemble of segmentation and estimation improves the segmentation and direct estimation results, as well as provides the best quantification results for the multi-indices ONH quantification.

Effectiveness of collaborative learning. To demonstrate the effectiveness of our collaborative learning, the DMTFS proposed in (Zhao et al., 2019c) serves as a baseline without benefits of FIM module. As shown in the last column of Table 1, the MCL-Net with collaborative learning obtains higher scores than DMTFs and single task solutions with MAE of 0.88%, 1.77% and correlation coefficient of 9.26%, 11.2% which demonstrates the effectiveness of our collaborative learning on multi-indices ONH quantification. The multitask ensemble method proposed in the previous work (Zhao et al., 2019) provides a multi-indices quantification by only using multitask ensemble, which obtains MAE of 1.01 ± 0.27, 0.98 ± 0.17, 1.22 ± 0.19 on ORIGA dataset. By contrast, our MCL-Net obtains the best performance of ONH quantification on all the 20 indices, which improves the regional areas such as RR4, RR5, DR1, DR2, DR5 and DR7. It should be noticed that these results prove the MCL potential and generalization ability. Besides, our MCL-Net also outperforms other single task solutions on multi-indices ONH quantification.

To demonstrate the effectiveness of our collaborative learning on segmentation task, the segmentation branch without collaboration is individually tested for ONH segmentation as a baseline, which is compared with our segmentation branch after collaboration with estimation branch. Qualitative results shown in Fig. 10 illustrate that the segmentation branch after collaboration achieves

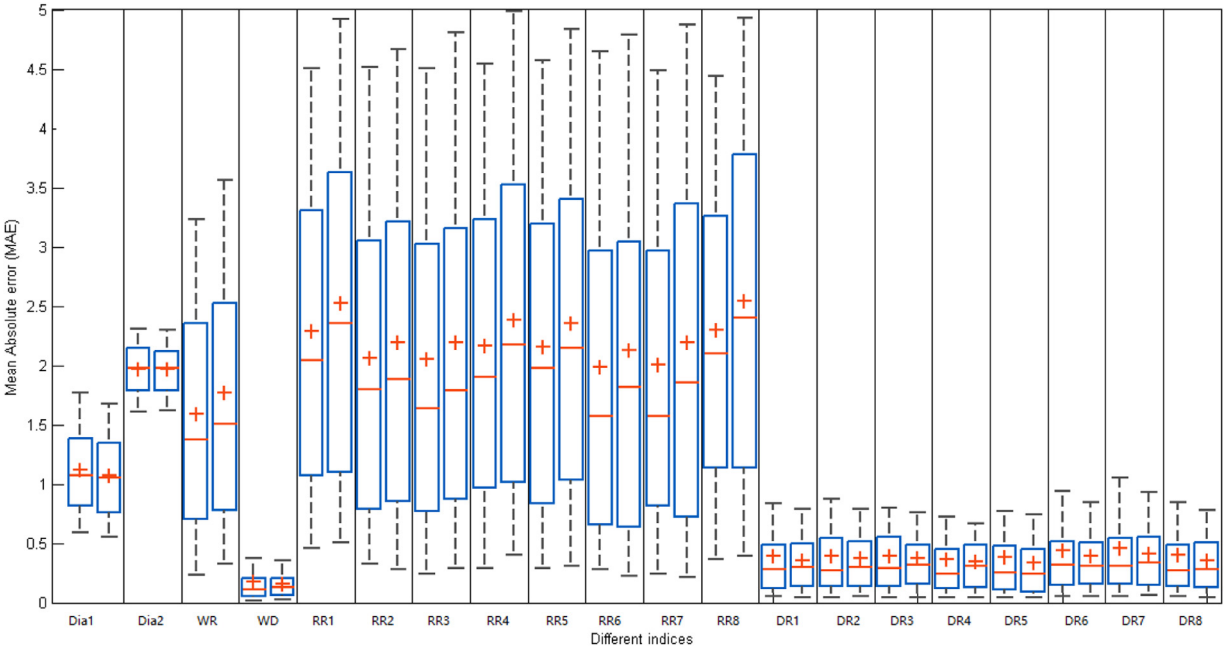


Fig. 8. Box-and-whisker demonstrates the accuracy of our proposed method for multiple ONH indices. Each whisker represents the range of ONH indices and its errors with all the test images, and the bottonm end of each whisker indicates the optimal performance and the horizontal red line shows the median value of the MAE. Each column indicates each of the ONH indices and the leaf whisker is with our MCL-Net while right one is MNet. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

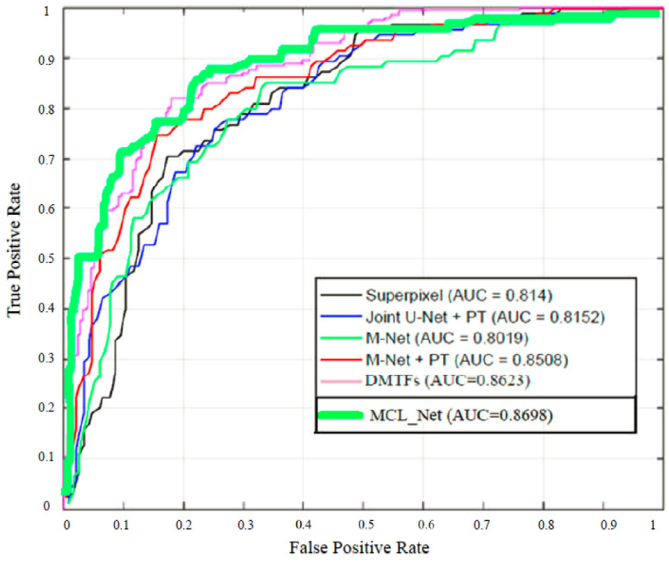


Fig. 9. The ROC curves with AUC scores for glaucoma screening based on the quantitative multi-types indices for our MCL-Net while only CDRs for others. Source: Fu et al. (2018) with our results added.

more precise segmentation of ONH compared with the baseline, benefiting from the collaborative information of estimation task with high-level semantics.

Statistical significance. Statistical differences from ground truth is assessed using paired *t*-test and *p*-values are shown in Table 1 to demonstrate the paired *t*-test for our proposed MCL-Net for 20 ONH quantitative indices. The test results reveal the extent of difference between our MCL-Net and its different configurations, which indicates that the proposed MCL-Net significantly improves the performance of multi-indices ONH quantification.

5.3. Performance comparison

MCL-Net reveals great advantages for multi-indices ONH quantification over existing segmentation-and-measurement method such as weakly-supervised (Zhao et al., 2019c), DeepLab (Chen et al., 2018), spatial-aware (Liu et al., 2019), MNet (Fu et al., 2018) and ensemble learning based method such as DMTFs (Zhao et al., 2019), as shown in Table 3.

Compared results show that MCL-Net obtains the average improvement of 3.5% on 20 indices, which achieves the best performance on multi-indices ONH quantification. Furthermore, the MCL-Net significantly improves the performance of glaucoma screening by evaluating the obtained multiple ONH indices. From the compared results we have the following observations.

(1) MCL-Net outperforms significantly the segmentation-and-measurement methods. Comparing the last column with others of Table 3, it clearly shows MCL-Net obtains more accurate multi-indices ONH quantification than single segmentation-and-measurement methods, which demonstrates the remarkable advantages in more detailed ONH quantification. As far as we know, MNet (Fu et al., 2018) and spatial-aware joint (Liu et al., 2019) achieve the best performance of ONH segmentation based on the single task model, while perform poorly for multi-indices ONH quantification, especially for regional areas. Our MCL-Net outperforms those segmentation-and-measurement methods with average MAE of 5.1%, 2.1% and 3.3% for diameters, regional areas and whole areas, respectively. We can see from Table 3 that multitask collaborative learning reduces the diversity of the quantitative indices with 11.6%. It is easy to understand that MCL-Net captures the detailed local and global semantic quantitative information simultaneously to boost the quantification performance.
Fig. 8 demonstrates the performance variations of different methods with the statistical box-and-whisker diagrams. Each whisker represents the range of ONH indices and its

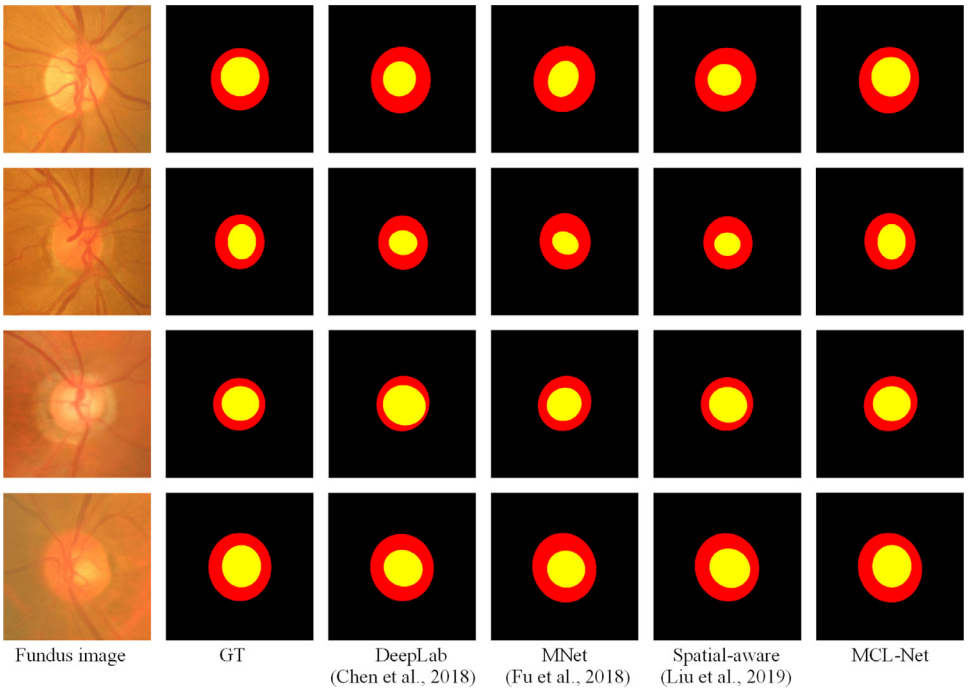


Fig. 10. Visual examples of the segmented mask of optic disc and cup, where the last row on challenging case shows the advantages of our MCL-Net due to the collaboration from estimation task. From the left to right: fundus image, ground truth (GT), DeepLab, M-Net, Spatial-aware method and our MCL-Net.

Table 3
Compassion with state-of-the-art methods for multi-indices ONH quantification.

Method	Weakly Zhao et al. (2019c)	DeepLab Chen et al. (2018)	Spatial-aware Liu et al. (2019)	MNet Fu et al. (2018)	DMTFs Zhao et al., 2019	MCL-Net Proposed
Diameter						
Dia1	–	1.12 ± 0.60	0.98 ± 0.21	0.97 ± 0.23	0.96 ± 0.21	0.94 ± 0.23
Dia2	1.86 ± 0.68	1.75 ± 0.43	1.08 ± 0.20	1.09 ± 0.28	1.05 ± 0.19	1.03 ± 0.18
Average	1.86 ± 0.68	1.43 ± 0.51	1.03 ± 0.20	1.03 ± 0.25	1.00 ± 0.20	**0.98 ± 0.20**
Whole areas						
WR	–	0.43 ± 0.23	0.32 ± 0.08	0.32 ± 0.09	0.32 ± 0.12	0.31 ± 0.08
WD	1.82 ± 0.25	1.72 ± 0.16	1.66 ± 0.17	1.67 ± 0.16	1.65 ± 0.15	1.63 ± 0.18
Average	1.82 ± 0.25	1.07 ± 0.19	0.99 ± 0.12	0.99 ± 0.12	0.98 ± 0.13	**0.97 ± 0.13**
Regional areas						
RR1	–	0.56 ± 0.34	0.74 ± 0.23	0.74 ± 0.21	0.73 ± 0.21	0.72 ± 0.20
RR2	–	0.55 ± 0.38	0.39 ± 0.12	0.39 ± 0.13	0.34 ± 0.10	0.31 ± 0.12
RR3	–	0.54 ± 0.37	0.13 ± 0.08	0.11 ± 0.08	0.11 ± 0.09	0.09 ± 0.10
RR4	–	0.53 ± 0.37	0.57 ± 0.24	0.55 ± 0.25	0.55 ± 0.29	0.52 ± 0.28
RR5	–	0.54 ± 0.29	0.72 ± 0.30	0.70 ± 0.31	0.69 ± 0.21	0.65 ± 0.20
RR6	–	0.55 ± 0.33	0.32 ± 0.14	0.30 ± 0.15	0.30 ± 0.13	0.27 ± 0.13
RR7	–	0.56 ± 0.30	0.16 ± 0.23	0.15 ± 0.26	0.13 ± 0.24	0.11 ± 0.25
RR8	–	0.56 ± 0.28	0.28 ± 0.18	0.26 ± 0.17	0.25 ± 0.18	0.24 ± 0.18
Average	–	0.54 ± 0.33	0.41 ± 0.19	0.40 ± 0.19	0.38 ± 0.18	**0.35 ± 0.18**
DR1	1.61 ± 0.64	1.44 ± 0.56	1.26 ± 0.13	1.28 ± 0.19	1.26 ± 0.12	1.24 ± 0.10
DR2	3.48 ± 0.66	3.47 ± 0.64	3.15 ± 0.48	3.17 ± 0.54	3.21 ± 0.46	3.16 ± 0.33
DR3	3.54 ± 0.69	3.45 ± 0.65	3.78 ± 0.36	3.80 ± 0.38	3.78 ± 0.29	3.74 ± 0.27
DR4	3.44 ± 0.58	3.39 ± 0.61	3.31 ± 0.30	3.32 ± 0.36	3.30 ± 0.20	3.25 ± 0.18
DR5	2.59 ± 0.54	2.41 ± 0.52	2.54 ± 0.28	2.55 ± 0.31	2.57 ± 0.18	2.55 ± 0.17
DR6	1.79 ± 0.41	1.49 ± 0.54	1.76 ± 0.25	1.77 ± 0.29	1.76 ± 0.18	1.73 ± 0.18
DR7	0.96 ± 0.25	0.94 ± 0.19	0.66 ± 0.23	0.66 ± 0.24	0.61 ± 0.17	0.60 ± 0.17
DR8	0.19 ± 0.11	0.89 ± 0.21	0.08 ± 0.06	0.09 ± 0.08	0.06 ± 0.05	0.05 ± 0.05
Average	2.20 ± 0.48	2.18 ± 0.49	2.06 ± 0.26	2.08 ± 0.29	2.06 ± 0.20	**2.04 ± 0.18**

errors with all the test images, the bottom end of each whisker represents the optimal performance and the horizontal red line shows the median value of the performance. From the plots we can see that the variety of the whisker's bottom ends are obvious high with the MCL-Net, and all the ranges of our whisker is smaller than others.

(2) MCL-Net outperforms the best of existing multitask ensemble method. Contrast to the existing multitask ensemble method (DMTFs) (Zhao et al., 2019) that integrate segmentation and estimation tasks into a unified framework, our MCL-Net leverages the merit of spatial and semantic dependencies between OD segmentation and indices estimation by collaborative learning, which brings obvious improvement of

quantification performance with 3.06%, 1.03% and 8.57% for diameters, regional areas and whole areas, respectively.

(3) MCL-Net even helps improve the performance of glaucoma screening. Fig. 9 shows the success of the proposed MCL-Net on glaucoma screening based on the quantitative 20 indices. Evidenced by ROC curves and AUC value (0.8698), the glaucoma screening results indicate that our multi-indices ONH quantification achieves a competitive performance using the 20 quantitative indices compared with the other methods only using the CDR value. Firstly, the multitask ensemble method obtains higher AUC scores than the one of MNet (Fu et al., 2018), which indicates the accurate quantitative indices improve the performance of glaucoma screening. Secondly, our MCL-Net achieves the best performance of glaucoma screening based on the multiple ONH indices. Those demonstrate that contrasted with the classical CDR values, multi-indices ONH quantification provides a more effective assessment of glaucomatous changes of ONH, which significantly improves clinical glaucoma diagnosis.

(4) MCL-Net helps improve the performance of ONH segmentation. To demonstrate the superiority of the proposed algorithm on ONH segmentation, visual examples of the segmented mask of optic disc and cup from our segmentation branch and other methods are shown in Fig. 10, where the first two rows are normal eyes and the rest rows are glaucoma cases, and the first two columns are the original fundus images and its manual ground truth. The last row in Fig. 10 is a challenging case with blurred and glaucomatous fundus, which is hard to find the precise OD/OC boundary. From the visual comparison we can see, our method addresses this case more accurate than other methods based on the collaboration of the high-level ONH indices from estimation branch.

6. Conclusion

In this work, we proposed a novel multitask collaborative learning (MCL-Net) for multi-indices ONH quantification. MCL-Net achieves multi-indices ONH quantification by leveraging the merit of spatial and context dependencies from ONH segmentation and indices estimation tasks collaboratively. It is the first time that the dependencies between two task-specific branches are investigated to obtain accurate multi-indices quantification by collaborative learning. Benefit from the consensus of multiple views, MCL-Net provides supplementary information and regularization to each other to promote accurate multi-indices ONH quantification. Experimental results on 650 fundus images show that MCL-Net not only achieves accurate multi-indices ONH quantification, but also obtains high AUC when the quantitative indices be used to glaucoma diagnosis. It reveals the great potential of focal and regional ONH assessment in clinical practice.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Science Foundation of China [61702558]; the Key Research and Development Plan of Hunan Province [2017WK2074]; and the National Key Research and Development Project of China [2017YFC0840104]. Computations were performed based on the heterogeneous GPU cluster (graham.computecanada.ca).

References

Afshin, M., Ayed, I.B., Islam, A., Goela, A., Peters, T.M., Li, S., 2012. Global assessment of cardiac function using image statistics in MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 535–543.

Argyriou, A., Evgeniou, T., Pontil, M., 2007. Multi-task feature learning. In: Advances in Neural Information Processing Systems, pp. 41–48.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Chen, X., Xu, Y., Yan, S., Wong, D.W.K., Wong, T.Y., Liu, J., 2015. Automatic feature learning for glaucoma detection based on deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 669–677.

Cheng, J., Liu, J., Wong, D.W.K., Yin, F., Cheung, C., Baskaran, M., Aung, T., Wong, T.Y., 2011. Automatic optic disc segmentation with peripapillary atrophy elimination. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 6224–6227.

Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D.W.K., Tan, N.-M., Tao, D., Cheng, C.-Y., Aung, T., Wong, T.Y., 2013. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imaging 32 (6), 1019–1032.

Cheng, J., Yin, F., Wong, D.W.K., Tao, D., Liu, J., 2015. Sparse dissimilarity-constrained coding for glaucoma screening. IEEE Trans. Biomed. Eng. 62 (5), 1395–1403.

Cheng, J., Zhang, Z., Tao, D., Wong, D.W.K., Liu, J., Baskaran, M., Aung, T., Wong, T.Y., 2017. Similarity regularized sparse group lasso for cup to disc ratio computation. Biomed. Opt. Express 8 (8), 3763–3777.

Fernandez-Granero, M., Sarmiento, A., Sanchez-Morillo, D., Jiménez, S., Alemany, P., Fondón, I., 2017. Automatic CDR estimation for early glaucoma diagnosis. J. Healthc. Eng. 2017.

Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging 37 (7), 1597–1605.

Garway-Heath, D., Hitchings, R., 1998. Quantitative evaluation of the optic nerve head in early glaucoma. Br. J. Ophthalmol. 82 (4), 352–361.

Guo, Y., Zou, B., Chen, Z., He, Q., Liu, Q., Zhao, R., 2016. Optic cup segmentation using large pixel patch based CNNs.

Harizman, N., Oliveira, C., Chiang, A., Tello, C., Marmor, M., Ritch, R., Liebmann, J.M., 2006. The ISNT rule and differentiation of normal from glaucomatous eyes. Arch. Ophthalmol. 124 (11), 1579–1583.

Jiang, Y., Xia, H., Xu, Y., Cheng, J., Fu, H., Duan, L., Meng, Z., Liu, J., 2018. Optic disc and cup segmentation with blood vessel removal from fundus images for glaucoma detection. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 862–865.

Joshi, G.D., Sivaswamy, J., Krishnadas, S., 2011. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. Med. Imaging 30 (6), 1192–1205.

Joshi, G.D., Sivaswamy, J., Krishnadas, S., 2012. Depth discontinuity-based cup segmentation from multiview color retinal images. IEEE Trans. Biomed. Eng. 59 (6), 1523–1531.

Liu, Q., Hong, X., Li, S., Chen, Z., Zhao, G., Zou, B., 2019. A spatial-aware joint optic disc and cup segmentation method. Neurocomputing.

López-Linares, K., Aranjuelo, N., Kabongo, L., Maclair, G., Lete, N., Ceresa, M., García-Familiar, A., Macía, I., Ballester, M.A.G., 2018. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using deep convolutional neural networks. Med. Image Anal. 46, 202–214.

Lu, D., Heisler, M., Lee, S., Ding, G.W., Navajas, E., Sarunic, M.V., Beg, M.F., 2019. Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. Med. Image Anal. 54, 100–110.

Ma, J., Zhao, Z., Chen, J., Li, A., Hong, L., Chi, E., 2019. SNR: sub-network routing for flexible parameter sharing in multi-task learning. In: The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19).

Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2016. Deep retinal image understanding. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 140–148.

Mary, M.C.V.S., Rajsingh, E.B., Jacob, J.K.K., Anandhi, D., Amato, U., Selvan, S.E., 2015. An empirical study on optic disc segmentation using an active contour model. Biomed. Signal Process. Control 18, 19–29.

Miri, M.S., Abràmoff, M.D., Lee, K., Niemeijer, M., Wang, J.-K., Kwon, Y.H., Garvin, M.K., 2015. Multimodal segmentation of optic disc and cup from SD-OCT and color fundus photographs using a machine-learning graph-based approach. IEEE Trans. Med. Imaging 34 (9), 1854–1866.

Misra, I., Shrivastava, A., Gupta, A., Hebert, M., 2016. Cross-stitch networks for multi-task learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3994–4003.

Mu, G., She, Q., Tian, Z., Gan, H., Jiang, P., 2018. A multi-task collaborative learning method based on auxiliary training and geometric constraints. In: 2018 IEEE Industrial Cyber-Physical Systems (ICPS). IEEE, pp. 79–84.

Rojas-Moraleda, R., Xiong, W., Halama, N., Breitkopf-Heinlein, K., Dooley, S., Salinas, L., Heermann, D.W., Valous, N.A., 2017. Robust detection and segmentation of cell nuclei in biomedical images based on a computational topology framework. Med. Image Anal. 38, 90–103.

Roth, H.R., Lu, L., Lay, N., Harrison, A.P., Farag, A., Sohn, A., Summers, R.M., 2018.

Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Med. Image Anal. 45, 94–107.
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.
Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., Yuille, A.L., 2018. Deep regression forests for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2304–2313.
Song, G., Chai, W., 2018. Collaborative learning for deep neural networks. In: Advances in Neural Information Processing Systems, pp. 1832–1841.
Wang, M., Zhang, D., Shen, D., Liu, M., 2019a. Multi-task exclusive relationship learning for alzheimers disease progression prediction with longitudinal data. Med. Image Anal. 53, 111–122.
Wang, S., Yu, L., Yang, X., Fu, C.-W., Heng, P.-A., 2019b. Patch-based output space adversarial learning for joint optic disc and cup segmentation. arXiv:1902.07519.
Wong, D., Liu, J., Lim, J., Li, H., Wong, T., 2009. Automated detection of kinks from blood vessels for optic cup segmentation in retinal images. In: Medical Imaging 2009: Computer-Aided Diagnosis, 7260. International Society for Optics and Photonics, p. 72601J.
Xu, C., Xu, L., Brahm, G., Zhang, H., Li, S., 2018. MuTGAN: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 525–534.
Xu, J., Chutatape, O., Chew, P., 2007. Automated optic disk boundary detection by modified active contour model. IEEE Trans. Biomed. Eng. 54 (3), 473–482.
Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D.W.K., Wong, T.Y., Liu, J., 2014. Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 788–795.
Xue, W., Brahm, G., Pandey, S., Leung, S., Li, S., 2018. Full left ventricle quantification via deep multitask relationships learning. Med. Image Anal. 43, 54–65.
Zhang, Y., Yang, Q., 2017. A survey on multi-task learning. arXiv:1707.08114.
Zhao, R., Chen, X., Liu, X., Chen, Z., Guo, F., Li, S., 2019. Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning doi:10.1109/JBHI.2019.2934477.
Zhao, R., Chen, Z., Liu, X., Zou, B., Li, S., 2019. Multi-index optic disc quantification via multitask ensemble learning. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 11764. Springer, pp. 21–29.
Zhao, R., Liao, W., Zou, B., Chen, Z., Li, S., 2019. Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis. The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 33. AAAI, pp. 809–816.
Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S., 2015. Direct and simultaneous four-chamber volume estimation by multi-output regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 669–676.
Zhen, X., Yu, M., He, X., Li, S., 2017. Multi-target regression via robust low-rank learning. IEEE Trans. Pattern Anal. Mach. Intell. 40 (2), 497–504.