Pattern Recognition Letters 142 (2021) 58–64

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Multi-modality fusion learning for the automatic diagnosis of optic neuropathy

Zheng Cao[a,1], Chuanbin Sun[b,1], Wenzhe Wang[a], Xiangshang Zheng[a], Jian Wu[a,**], Honghao Gao[c,d,*]

[a] College of Computer Science and Technology, Zhejiang University, Hangzhou 310007, China
[b] Eye Center, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310009, China
[c] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[d] Gachon University, Gyeonggi-Do 461,701, South Korea

ARTICLE INFO

Article history:
Received 5 June 2020
Revised 13 November 2020
Accepted 15 December 2020
Available online 30 December 2020

Keywords:
Multi-modality
Deep learning
Optic neuropathy
Computer-aided diagnosis

ABSTRACT

Optic neuropathy is kind of common eye diseases, which usually causes irreversible vision loss. Early diagnosis is key to saving patients' vision. Due to the similar early clinical manifestations of common optic neuropathy, it may cause misdiagnosis and delays in treatment. Worse, most diagnoses rely on experienced doctors. In this paper, we proposed a novel deep learning architecture GroupFusionNet (GFN) to diagnose five normal optic neuropathy diseases, including Anterior Ischemic Optic Neuropathy (AION), papilledema, papillitis, Optic Disc Vasculitis (ODV), and optic atrophy (OA). Specifically, we combined multi-modalities in clinic examination such as fundus image, visual field tests and age of each patient. GFN utilized two ResNet pathways to extract and fuse both features of fundus image and visual field tests, and the information of structured data was embedded in the end. Experimental results demonstrate that multi-modality feature aggregation is effective for optic neuropathy diseases diagnosis, and GFN achieved a five-classes classification accuracy of 87.82% on the test dataset.

1. Introduction

Optic neuropathy means pathological abnormalities in the optic nerve. These abnormalities induced by various factors (e.g., inflammation, tumor, gene, infection, and ischemia) can lead to vision dysfunction or even vision loss [1]. The development of most optic neuropathy is acute, and extreme vision loss usually occurs. As a result, to avoid deteriorations like optic atrophy, patients must be treated timely [2]. Therefore, it is significant for ophthalmologists to diagnose optic neuropathy early and accurately, so as to treat patients timely and save their sight. But in fact it is difficult to do so because several kinds of optic neuropathy share similar clinical manifestation, especially Anterior Ischemic Optic Neuropathy (AION), papilledema, papillitis, Optic Disc Vasculitis (ODV), and Optic Atrophy (OA). The above five kinds of optic neuropathy manifest in optic nerve head swelling and visual field defects, making it hard for ophthalmologists to diagnose [3,4]. Despite similar clinical manifestation, the treatments of the five diseases are not the same. Once a misdiagnosis occurs, it will not only delay the condition but also cause complications.

For suspected cases of optic neuropathy, there are some common kinds of examinations. The optic disc is the starting part of the optic nerve, where gather optic nerve fibers on the retina, and the eyeballs are worn out to form the optic nerve. Fundus topography is the imaging of the posterior part of the eye, where the central and peripheral retina, optic disc, and macula can be seen [5]. The pathological changes of the optic disc can be observed in fundus images for optic nerve disease diagnosis [4]. In this paper, we mainly use color fundus image for the diagnosis of optic nerve neuropathy. However, diagnosis only rely on fundus images may not be enough. Furthermore, optic neuropathy diseases may cause local scotoma in the visual field or even a more extensive loss of vision. Different optic neuropathy diseases have different influences on the optic nerve, corresponding to different forms of visual field defects. The visual field test can be utilized to examine whether and how the visual field is affected by these diseases [6]. Both of the above examinations are easily-performed, low-cost, noninvasive, high in-patient acceptance, and relatively accurate in

* Corresponding author at: School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
** Corresponding author at: College of Computer Science and Technology, Zhejiang University, Hangzhou 310007, China
E-mail addresses: z.cao@zju.edu.cn (Z. Cao), wujian2000@zju.edu.cn (J. Wu), gaohonghao@shu.edu.cn (H. Gao).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.patrec.2020.12.009
0167-8655/

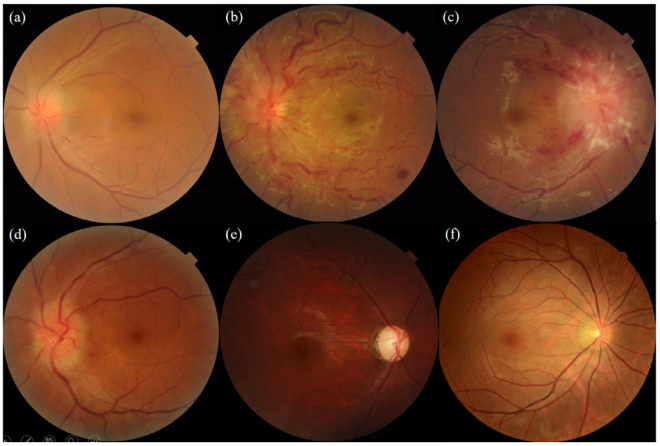

Fig. 1. Examples of fundus images of the five optic neuropathy diseases diagnosed in this paper: (a) Anterior Ischemic Optic Neuropathy (AION), (b) Optic Disc Vasculitis (ODV), (c) Papilledema, (d) Papillitis, (e) Optic Atrophy (OA), (f) Normal.

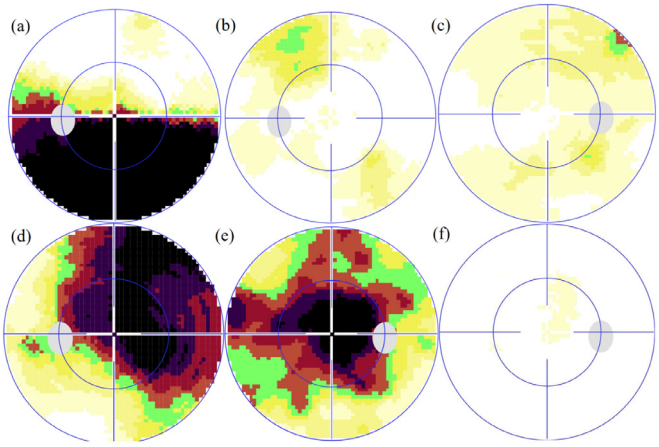

Fig. 2. Illustration of visual field tests of the five optic neuropathy diseases corresponding to Fig. 1. Note that the darker the color in the visual field, the more serious the visual field defects. (a) Anterior Ischemic Optic Neuropathy (AION), (b) Optic Disc Vasculitis (ODV), (c) Papilledema, (d) Papillitis, (e) Optic Atrophy (OA), (f) Normal.

diagnosing these five optic neuropathy diseases. The example fundus images and their corresponding visual field tests of the five optic neuropathy diseases along with normal people are shown in Fig. 1 and Fig. 2. Note that the darker the color in a visual field image, the more serious the visual field defects. Meanwhile, structured data such as age and gender is a potential factor for automatic optic neuropathy diagnosis. We collect patients ages as the input of our model. In total, three modalities (i.e., fundus images, visual field tests, and ages) are utilized in this paper for the automatic diagnosis of optic neuropathy.

With the rapid development of machine learning, especially deep learning, Computer-aided Diagnosis (CAD) systems can be built to assist disease diagnosis and to reduce the burden on ophthalmologists. However, most of the studies that conducted on the CAD of optic neuropathy are based on a single modality [7–26], and none of them can distinguish all the five optic neuropathy diseases. On the other hand, most existing work targeting on the multi-modality fusion learning of optical images utilized binary modalities [27–32], and most of them fused both of the modalities in a single point. In this paper, we assume that fusing the features extracted from multi-point could potentially improve the performance of multi-modality models. And we propose a novel deep learning based multi-modality fusion learning network, called

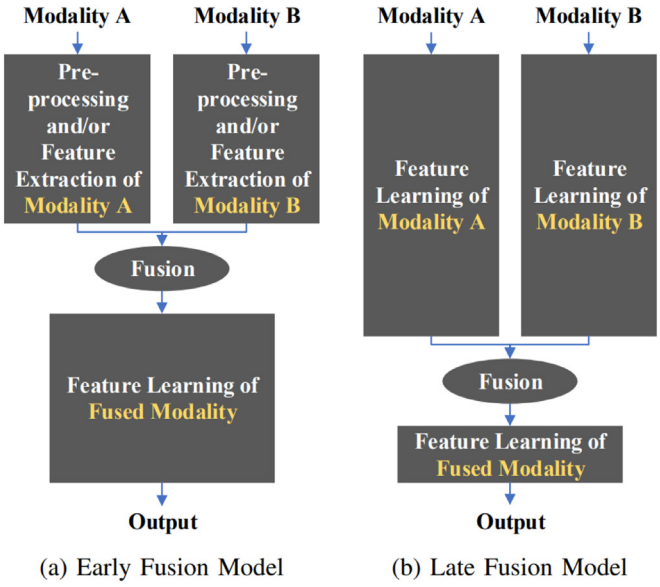

Fig. 3. Illustration of two types of commonly adopted single fusion models.

GroupFusionNet (GFN), that fuses three modalities by group. The proposed network inputs 5 optic neuropathy diseases classes, and outputs the diagnosis result. The overall architecture of our proposed GFN is shown in Fig. 4. Our contributions are three-fold.

1. Develop a CAD system for the automatic diagnosis of five optic neuropathy diseases (i.e., AION, papilledema, papillitis, ODV, and OA) at the same time;
2. We firstly use visual field tests for the automatic diagnosis of ophthalmic diseases;
3. Propose a multi-modality fusion learning network, called GroupFusionNet (GFN), to classify the five optic neuropathy diseases.

The accuracy of our proposed model reached 87.82% in classification of five classes of optic neuropathy diseases finally. The rest of the paper is organized as follows. Work related to the CAD of optic neuropathy and multi-modality fusion learning of optical images is introduced in Section 2. In Section 3, our method is presented. Experiments and results follow in Section 4. Finally, we conclude the paper in Section 5.

2. Related work

2.1. CAD of optic neuropathy

Many studies have been conducted on computer-aided optic neuropathy diagnosis using optical images based on a single modality such as fundus image. Yang et al. developed an automatic diagnosis system of optic neuropathy using logistic regression [7]. Alghamdi et al. and Liu et al. proposed to use a Convolutional Neural Network (CNN) for optic disc abnormality detection with digital fundus image [8,9]. Moreover, Eun-Oh et al. use fundus photographs to detect Retinal Nerve Fiber Layer (RNFL) defects [11]. But they did not classify images into detailed optic neuropathy diseases, making it still difficult for ophthalmologists to make decisions and develop treatment plans.

For automatic diagnosis of detailed optic neuropathy disease, Al-Naami et al. [10] used custom image processing algorithms for Ischemic Optic Neuropathy (ION) classification. Joshi et al. [12] used custom image processing algorithms and machine learning methods for automatic diagnosis of RNFL defects. As for papilledema automated analysis, a number of solutions based on sup-

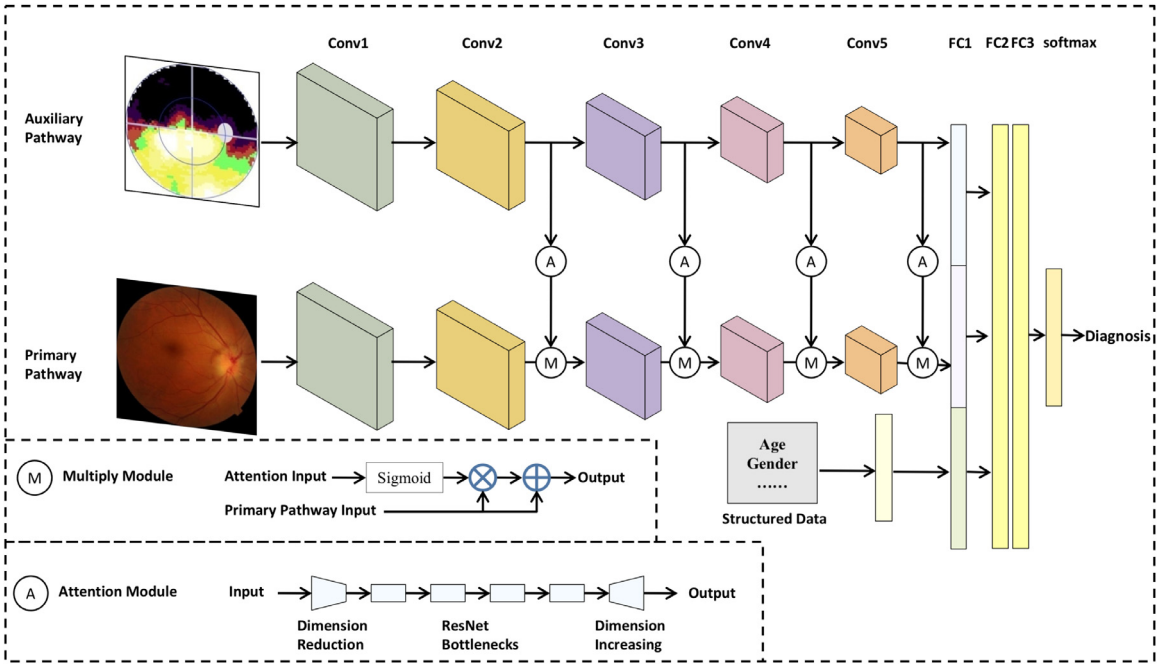

Fig. 4. The overall architecture of our proposed GroupFusionNet (GFN).

port vector machine (SVM) with different pre-processing feature extraction methods are proposed [13–19]. In the field of Glaucomatous Optic Neuropathy (GON) diagnosis, Twa et al. proposed to utilize Decision Tree for GON automatic diagnosis [20]. After that, Liu et al. further utilized CNNs to improve the performance of GON diagnosis models [23]. Some CNN methods are introduced based on different GON datasets [21–25]. Besides, new deep learning technology such as transfer learning is proposed to classify GON fundus image [26]. However, none of these studies can distinguish all the five optic neuropathy diseases, and none of them can diagnose optic papillitis and Optic Disc Vasculitis (ODV). In order to produce computer-aided optic neuropathy diagnosis, a network that can automatically classify optic neuropathy diseases in detail is urgently needed.

2.2. Multi-modality fusion learning of medical images

For the multi-modality fusion learning of optical images, most existing works utilized binary modalities [27–31], where Optical Coherence Tomography (OCT) and fundus images are commonly used. And it is also very common that both of the modalities are fused in a single point [33]. Apart from the image data, Jordan et al. further consider the structured data such as age and gender in skin lesion diagnosis, and they prove the combination of such data can achieve a higher classification accuracy [32]. As for image modality, based on fusion locations, single fusion methods can be divided into early fusion [27,29,32] and late fusion [28,30,31], respectively fusing low-level features and prediction-level features [33]. The architecture of these two types of single fusion methods is shown in Fig. 3. As reported by Snoek et al. [34], late fusion is commonly largely outperformed by early fusion, while it performs slightly better in many cases. It is difficult to tell which one of the two types is better because they both have their advantages and disadvantages. In early fusion, low-level features or even original images are fused at the beginning of models and abundant low-level features can be learned, making it easier to extract prediction-level features. However, this may bring the under-fitting issue. In addition, in late fusion, prediction-level features belonging to two modalities are extracted individually and well. And the final prediction is conducted after fusion. However, the relationship between the two modalities in low-level is not well explored. In this paper, we propose a network that fuse features by group (GFN) and can take advantage of the two fusion types. Details are shown in Section 3.

3. Method

The overall architecture of our proposed GroupFusionNet (GFN) is illustrated in Fig. 4, and it is based on attention mechanisms. Two Residual Network (ResNet) pathways [35] are utilized as the backbone of our network, and features belonging to different modalities are fused by group along the ResNets. Then the structured data is embedded into the full connected layer. Each part in detail is described as follows.

3.1. Fusion of two image modalities

As described in Section 1 and shown in Figs. 1 and 2, fundus topography is a kind of imaging while visual field tests are logic matrixes that reflect the condition of visual fields. This suggests that fundus images may contain more information than visual fields. Based on this observation, we assume that the features in fundus images are more difficult to learn compared to those in visual fields. Therefore, we utilize a relative shallow network for the feature extraction of visual fields, while utilizing a relative deep network for that of fundus images. In order to learn features from 2D medical images, modular deep learning models like VGG [36], ResNet [35], ResNeXt [37], and DenseNet [38] are commonly used. Among the above methods, ResNet is easier to optimize with lower complexity [35]. Without loss of generality, in this paper, we utilize ResNet as the backbone of our network.

As described in Section 2.2, early fusion and late fusion are two important and commonly adopted fusion types for multi-modality fusion learning of optical images. It would be natural to combine the two types in one model, so that their advantages can be combined and their disadvantages can be alleviated. Further, we propose a network that can fuse features more dense. Note that since the two ResNets that we utilized contain different layers, fusing

features at every layer is hard. So we divide ResNets into basic groups. A typical ResNet can be divided into 6 groups, which are the first 7×7 convolutional layer, 4 groups of ResBlocks and the final fully-connected layer [35]. After each of the convolutional groups (i.e., the first five groups), we designed the attention module to extract the information from each ResBlock output of the auxiliary pathway, as shown in Fig. 4. The attention module is similar as a ResBlock, which firstly reshape the dimension of input data to 256×256. Then it is fed into four bottleneck building blocks as described in He et al. [35], which contains 1×1, 3×3 and 1×1 three convolution layers within a residual function. After that, the dimension shape of the attention data is recovered. To fuse the attention features in primary convolutional pathway, we design a multiply module as well, in which the attention map is normalized by sigmoid function and multiplied by the primary pathway map. Eq. (1) concludes the fusion process:

$$P_{i'} = (Sigmoid(f_{attention}(A_i) + 1)) \times P_i \forall i \in [1, 5] \tag{1}$$

where P and A indicate the primary pathway and auxiliary pathway respectively; $f_{attention}$ is the function of attention module; the subscripts i and i' represent the sequence number of current group and next group respectively.

3.2. Fusion of structured data

Apart from fundus image and view field test, some structured data like age and gender are considered to assist diagnosis during the clinic. For example, AION is normal in people aged 57–65 [39] and the female aged 15–44 are more susceptible to papilledema [40]. Therefore, the further fusion these information can have a better representation and improve the classification accuracy. In GFN, we categorize the age into four groups according to different generation: 0–18, 19–40, 41–60 and 61–100. And the structured data is represented by a 6-dimensional vector with one-hot form.

3.3. Learning from the fused data

After the group fusion of two image modalities, the structured data can be fused as a late fusion type. The fully connected layers are applied to merge the information and further learn the correlations across different modalities. The fully connected layer is computed by Eq. (2):

$$z_i = f(W_i \hat{x}_{i-1} + b_i) \tag{2}$$

where z_i means the activations in the ith layer; W_i and b_i indicate the weights and bias learned for the ith layer; $\hat{x}_{i-1}$ are the normalized output of the previous layer; $f(x) = \max(0, x)$ is the ReLU activation function.

In the end, a softmax layer is utilized after the full connected layers to predict the diagnosis of five categories of optic neuropathy, by Eq. 3:

$$p(y = j|\hat{x}; W, b) = \frac{exp(W_j \hat{x} + b_j)}{exp(\sum_{i=1}^{5} W_i \hat{x} + b_i)} \tag{3}$$

where $p(y = j)$ represents the probability of the input data belonging to the jth category, and $\forall j \in [1, 5]$ here; $\hat{x}$ indicates the normalized output of full connected layers; W and b are weights and bias learned in softmax layer.

4. Experiments and results

4.1. Experimental datasets

We collect 1032 cases in total from 1032 distinct eyes at the Eye Centre, Second Affiliated Hospital of Zhejiang University School of Medicine. Each eye represents a case and each of these cases contains a fundus image with a resolution of 1500×1500 pixels, a visual field with a resolution of 512×512 pixels, and the age and gender of the patient. Fundus images, visual fields and ages associated with a specific eye are assigned with the same class. For each case, six ophthalmologists jointly classify its condition as normal, Anterior Ischemic Optic Neuropathy (AION), papilledema, papillitis, Optic Disc Vasculitis (ODV), and Optic Atrophy (OA), by examining the three corresponding modalities. The dataset is almost evenly distributed, and each disease class contains about 200 cases. Our dataset is shuffled and equally divided into 6-fold, so that we can evaluate our method by 6-fold cross validation.

4.2. Performance evaluations

To evaluate our model, we report both per-class and overall performance using four metrics, i.e., Average Top-1 accuracy, sensitivity, specificity, and F1-score. The definition of the evaluation metric can be defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$\text{F1-Score} = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{7}$$

in which TP, FP, TN and FN represents true positive, false positive, true negative and false negative prediction respectively.

4.3. Implementation details

For network training, fundus images are resized to 512×512 as the input, so that they have the same size as visual fields and the two modalities can be fused directly in our network. In order to accelerate convergence and achieve better results, we firstly pre-train the two ResNet pathways respectively until they get converge. After that the whole network is trained jointly with the pre-trained weights. When training of the two ResNets and our whole network, Cross-Entropy loss is utilized for back-propagation. Rotation and horizontal flipping are utilized as data augmentation for both of the image modalities. Note that the two modalities are augmented simultaneously when our network is trained, so that the two modalities are still relatively registered. Each of the networks is trained for 200 epochs with batch size of 32 and learning rate of 1e4. And Adam is utilized as the optimizer of the networks, which combines the ability of AdaGrad to deal with sparse gradients and the ability of RMSProp to deal with non-stationary objectives [41]. Our experiments are performed on a workstation platform with Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz, 256 GB RAM and 8x NVIDIA Titan Xp GPU with 12 GB GPU memory. The code is implemented with PyTorch 0.4.1 in Ubuntu 18.04.

4.4. Result and discussion

For our final model, we implement ResNet-101 as a backbone in GFN for both primary pathway and auxiliary pathway. The detailed performance of GFN is shown in Table 2, in which the accuracy, recall, precision and F1-score are reported for each class. Overall, the model can distinguish the five optical neuropathy diseases well. With the embedding of structured data, the model achieved a mean classification accuracy of 87.82% on a held-out test set.

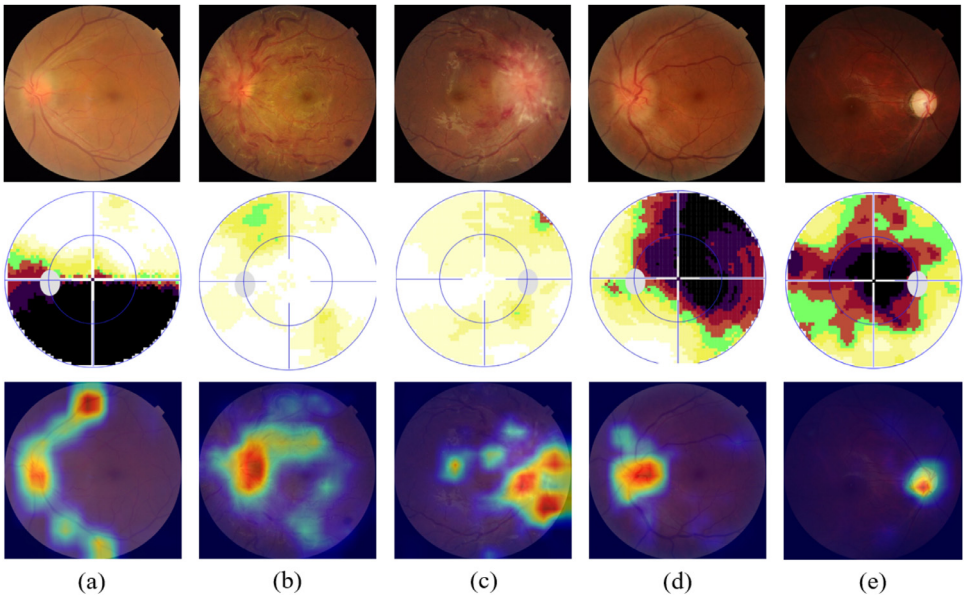

Fig. 5. Feature map captured in Conv5 for five diseases, row 1 and 2 give the input fundus image and view field test image respectively, and row 3 shows the feature map of (a) Anterior Ischemic Optic Neuropathy (AION), (b) Optic Disc Vasculitis (ODV), (c) Papilledema, (d) Papillitis, (e) Optic Atrophy (OA).

Table 1
Comparative and ablation study results. Average Top-1 accuracy, weighted F1-Score, mean recall along with the mean inference time of each case are considered here.

Algorithm	Acc	F1	Recall	Time (ms)
Single Modality: Fundus Image				
VGG-16 [36]	77.38	78.06	77.27	151.52
DenseNet-121 [38]	74.47	74.10	73.48	66.43
ResNet-50 [35]	80.61	80.47	80.49	53.09
ResNet-101 [35]	81.38	81.84	80.70	84.52
ResNeXt-101 [37]	81.25	81.28	81.32	139.62
Single Modality: View Field Test Image				
VGG-16 [36]	50.58	50.20	55.29	139.62
ResNet-50 [35]	52.07	51.22	57.01	52.34
ResNet-101 [35]	53.59	53.44	55.61	137.40
Multi-Modality: ResNet-50 as backbone				
Early Fusion [34]	80.62	80.58	80.37	56.42
Late Fusion [30]	81.74	82.17	82.19	114.81
Proposed GFN	86.84	86.57	86.89	150.32
Multi-Modality: ResNet-101 as backbone				
Early Fusion [34]	81.87	81.90	79.95	89.23
Late Fusion [30]	83.16	82.07	85.71	187.61
Proposed GFN	**87.82**	**87.73**	**88.95**	216.32
Ablation Study: ResNet-101 as backbone				
GFN (1-layer)	84.23	83.80	84.01	191.45
+ (4-layer)	85.73	85.31	85.50	201.52
+ Attention	86.91	86.88	86.59	213.27
+ Structured data	87.82	87.73	88.95	216.32

The recall, precision and F1-Score performance of GFN is 88.95%, 86.68%, and 87.73% respectively.

Comparative experiments The comparative experiments results of state-of-the-art methods are shown in Table 1, which reports the values of average Top-1 accuracy, weighted F1-Score and mean recall for different experiments. The average rum time of model inference on the test set with a single Titan Xp is shown in Table 1 as well. As shown in Table 1, when only using single modality, ResNet-101 has a best classification accuracy of 81.38% based on fundus images, which has a slightly better performance than ResNet-50 and ResNeXt-101 (80.61% and 81.25% accuracy respectively). As a baseline method, VGG-16 [36] has a 77.38% classification accuracy but consumes relatively longer time in inference. Besides, DenseNet-121 [38] performs not as good as other methods, which only has a 73.48% recall. Due to less information contained in view field test image, either ResNet or VGG can only classify the five diseases with the accuracy of around 50%. The experiment results of multi-modality model indicate that the fusion of the two modalities can significantly improve the classification performance of the model. Similar as the results obtained in single modality test, GFN utilized ResNet-101 performs slight better than which with ResNet-50. The experiments demonstrates that our GFN can capture more information than either early fusion or late fusion methods, at an increasing accuracy of 5–7%. Moreover, the complexity of GFN is acceptable because it only consumes 15% more inference time than late fusion methods.

Ablation experiments Apart from the above comparison with state-of-art method, we conduct several extensive experiments to validate the effectiveness of our method as well, as reported in Table 1. Firstly, we only fuse the information of last layer in each pathway directly with function $P_i' = (A_i + 1) \times P_i$. GFN can obtain approximately 2% improvement in accuracy. Secondly, when fuses 4-layers, GFN can obtain a 85.80% classification F1 Score, and it prove more fusion layers can capture more effective information.

Table 2
Performance of GFN in diagnosis five normal optic neuropathy on a held-out test set. Values depict mean from six-fold cross-validation.

EvaluationAccuracy(%)Class	Papillitis	AION	Papilledema	ODV	OA	Overall
Accuracy	89.64	97.75	88.65	98.20	91.44	87.82
Recall	84.66	90.24	86.46	93.94	89.53	88.95
Precision	88.69	86.05	86.36	83.78	88.51	86.68
F1 Score	86.63	88.10	86.71	88.57	89.02	87.73

Thirdly, the attention module is tested, which show its potential in feature extraction with the result of 1.18% more accuracy. Finally, with the embedding of structured data, GFN achieves 87.82% accuracy in five optical neuropathy disease diagnoses.

Visualization results In our proposed GFN, the view field test image pathway is designed to assist fundus image path, which can draw the correct attention in deep learning model and eventually help the multi-view feature learning. We visualize the feature maps of the Conv5 group result in Fig. 5, following the methods introduced in Zhou et al. [42]. As shown in Fig. 5, the feature map can capture more attention on the regions that view field test image point out, which also can repress the noise of fundus image.

5. Conclusions

In summary, we proposed a novel multi-modality deep learning architecture GroupFusionNet (GFN) to guide optic neuropathy diagnosis, which makes full use of clinic information including fundus image, view field test and any non-image data. The model can finally achieve a classification percentage of about 88% among five normal optic neuropathy diseases. The experimental result has proved that the fusion of multi-modality in deep learning can achieve much better performance. Due to the limitation of the number of patient cases, it will be possible to further improve the modal performance with additional training data. In addition, considering the complex pathology of optic neuropathy, more modality in clinic examination such as optical coherence tomography (OCT) can be integrated in fusion model in the future work.

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

We acknowledge the contribution of Miss Yuqing Fan for her support and valuable comments. This research was partially supported by the National Research and Development Program of China under grant nos. 2019YFB1404802, 2019YFC0118802, and 2018AAA0102102, the National Natural Science Foundation of China under grant no. 61672453, the Zhejiang University Education Foundation under grants nos. K18-511120-004, K17-511120-017, and K17-518051-02, the Zhejiang public welfare technology research project under grant no. LGF20F020013, and the Key Laboratory of Medical Neurobiology of Zhejiang Province. Dr. Sun's research was supported by Ophthalmology Star Program of Shenyang Ophthalmic Industrial Technology Institute under grant no. QMX2019-01-001.

References

[1] L.K. Gordon, Optic nerve, in: Pharmacologic Therapy of Ocular Disease, Springer, 2016, pp. 369–386.
[2] W. Lagrèze, Treatment of optic neuropathies–state of the art, Klinische Monatsblatter Augenheilkd. 226 (11) (2009) 875–880.
[3] L.A. Mesentier-Louro, C. Zaverucha-do Valle, P.H. Rosado-de Castro, A.J. Silva-Junior, P.M. Pimentel-Coelho, R. Mendez-Otero, M.F. Santiago, Bone marrow-derived cells as a therapeutic approach to optic nerve diseases, Stem Cells Int. 2016 (2016).
[4] N.R. Miller, F.B. Walsh, W.F. Hoyt, Walsh and Hoyt's Clinical Neuro-Ophthalmology, 2, Lippincott Williams & Wilkins, 2005.
[5] P.J. Saine, M.E. Tyler, Ophthalmic Photography: Retinal Photography, Angiography, and Electronic Imaging, 132, Butterworth-Heinemann Boston, 2002.
[6] J. Smythies, A note on the concept of the visual field in neurology, psychology, and visual neuroscience, Perception 25 (3) (1996) 369–371.
[7] H.K.Yang, Development of an Automated Diagnosis System of Optic Neuropathy in Digital Fundus Photographs, Seoul National University, 2018 Ph.D. thesis.
[8] H.S. Alghamdi, H.L. Tang, S.A. Waheeb, T. Peto, Automatic optic disc abnormality detection in fundus images: a deep learning approach (2016).
[9] T.A. Liu, D.S. Ting, H.Y. Paul, J. Wei, H. Zhu, P.S. Subramanian, T. Li, F.K. Hui, G.D. Hager, N.R. Miller, Deep learning and transfer learning for optic disc laterality detection: implications for machine learning in neuro-ophthalmology, J. Neuro-Ophthalmol. (2019).
[10] B. Al-Naami, N. Gharaibeh, T. Hayajneh, B.J. Mohd, A.-A. Kheshman, Classification of ischemic optic neuropathy using custom image processing algorithm-statistical based analysis, Int. J. Digit. Content Technol. Appl. 7 (10) (2013) 95.
[11] J.E. Oh, H.K. Yang, K.G. Kim, J.-M. Hwang, Automatic computer-aided diagnosis of retinal nerve fiber layer defects using fundus photographs in optic neuropathy, Investig. Ophthalmol. Vis. Sci. 56 (5) (2015) 2872–2879.
[12] G.D. Joshi, J. Sivaswamy, R. Prashanth, S. Krishnadas, Detection of peri-papillary atrophy and RNFL defect from retinal images, in: International Conference Image Analysis and Recognition, Springer, 2012, pp. 400–407.
[13] S. Akbar, M.U. Akram, M. Sharif, A. Tariq, U. ullah Yasin, Arteriovenous ratio and papilledema based hybrid decision support system for detection and grading of hypertensive retinopathy, Comput. Methods Prog. Biomed. 154 (2018) 123–141.
[14] K. Yousaf, M.U. Akram, U. Ali, S.A. Sheikh, Assessment of papilledema using fundus images, in: 2016 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2016, pp. 476–481.
[15] S. Echegaray, G. Zamora, H. Yu, W. Luo, P. Soliz, R. Kardon, Automated analysis of optic nerve images for detection and staging of papilledema, Investig. Ophthalmol. Vis. Sci. 52 (10) (2011) 7470–7478.
[16] C. Muramatsu, Y. Hatanaka, A. Sawada, T. Yamamoto, H. Fujita, Computerized detection of peripapillary chorioretinal atrophy by texture analysis, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 5947–5950.
[17] S. Akbar, M.U. Akram, M. Sharif, A. Tariq, U. ullah Yasin, Decision support system for detection of papilledema through fundus retinal images, J. Med. Syst. 41 (4) (2017) 66.
[18] K.N. Fatima, T. Hassan, M.U. Akram, M. Akhtar, W.H. Butt, Fully automated diagnosis of papilledema through robust extraction of vascular patterns and ocular pathology from fundus photographs, Biomed. Opt. Express 8 (2) (2017) 1005–1024.
[19] K.N. Fatima, M.U. Akram, S.A. Bazaz, Papilledema detection in fundus images using hybrid feature set, in: 2015 5th International Conference on IT Convergence and Security (ICITCS), IEEE, 2015, pp. 1–4.
[20] M.D. Twa, S. Parthasarathy, C.A. Johnson, M.A. Bullimore, Morphometric analysis and classification of glaucomatous optic neuropathy using radial polynomials, J. Glaucoma 21 (5) (2012) 302.
[21] M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini, R. Klette, A comparison of transfer learning techniques, deep convolutional neural network and multilayer neural network methods for the diagnosis of glaucomatous optic neuropathy, in: International Computer Symposium, Springer, 2018, pp. 627–635.
[22] Y. Xi, Y. Zhang, S. Ding, S. Wan, Visual question answering model based on visual relationship detection, Signal Process. 80 (2020) 115648.
[23] H. Liu, L. Li, I.M. Wormstone, C. Qiao, C. Zhang, P. Liu, S. Li, H. Wang, D. Mou, R. Pang, et al., Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs, JAMA Ophthalmol. 137 (12) (2019) 1353–1360.
[24] S. Ding, S. Qu, Y. Xi, S. Wan, Stimulus-driven and concept-driven analysis for image caption generation, Neurocomputing 398 (2020) 520–530.
[25] L.A. Al-Aswad, R. Kapoor, C.K. Chu, S. Walters, D. Gong, A. Garg, K. Gopal, V. Patel, T. Sameer, T.W. Rogers, et al., Evaluation of a deep learning system for identifying glaucomatous optic neuropathy based on color fundus photographs, J. Glaucoma 28 (12) (2019) 1029–1034.
[26] M. Christopher, A. Belghith, C. Bowd, J.A. Proudfoot, M.H. Goldbaum, R.N. Weinreb, C.A. Girkin, J.M. Liebmann, L.M. Zangwill, Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs, Sci. Rep. 8 (1) (2018) 1–13.
[27] M.S. Miri, M.D. Abràmoff, K. Lee, M. Niemeijer, J.-K. Wang, Y.H. Kwon, M.K. Garvin, 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) (2015) 1854–1866.
[28] Z. Hu, M.D. Abràmoff, M. Niemeijer, M.K. Garvin, Automated multimodality concurrent classification for segmenting vessels in 3D spectral OCT and color fundus images, in: Medical Imaging 2011: Image Processing, 7962, International Society for Optics and Photonics, 2011, p. 796204.
[29] M.S. Miri, A multimodal machine-learning graph-based approach for segmenting glaucomatous optic nerve head structures from SD-OCT volumes and fundus photographs (2016).
[30] W. Wang, Z. Xu, W. Yu, J. Zhao, J. Yang, F. He, Z. Yang, D. Chen, D. Ding, Y. Chen, et al., Two-stream CNN with loose pair training for multi-modal AMD categorization, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 156–164.
[31] M. Arikan, A. Sadeghipour, B. Gerendas, R. Told, U. Schmidt-Erfurt, Deep learning based multi-modal registration for retinal imaging, in: Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support, Springer, 2019, pp. 75–82.
[32] J. Yap, W. Yolland, P. Tschandl, Multimodal skin lesion classification using deep learning, Exp. Dermatol. 27 (11) (2018) 1261–1267.
[33] J.-M. Pérez-Rúa, V. Vielzeuf, S. Pateux, M. Baccouche, F. Jurie, MFAS: multimodal fusion architecture search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 6966–6975.

[34] C.G. Snoek, M. Worring, A.W. Smeulders, Early versus late fusion in semantic video analysis, in: Proceedings of the 13th Annual ACM International Conference on Multimedia, 2005, pp. 399–402.
[35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556(2014).
[37] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
[38] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[39] S. Berry, W.V. Lin, A. Sadaka, A.G. Lee, Nonarteritic anterior ischemic optic neuropathy: cause, effect, and management, Eye Brain 9 (2017) 23.
[40] M. Rigi, S.J. Almarzouqi, M.L. Morgan, A.G. Lee, Papilledema: epidemiology, etiology, and clinical management, Eye Brain 7 (2015) 47.
[41] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980(2014).
[42] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.