SUNET: A LESION REGULARIZED MODEL FOR SIMULTANEOUS DIABETIC RETINOPATHY AND DIABETIC MACULAR EDEMA GRADING

Zhi Tu[1] *Shenghua Gao*[1] *Kang Zhou*[1,2] *Xianing Chen*[1] *Huazhu Fu*[3]
Zaiwang Gu[5] Jun Cheng[5] Zehao Yu[1] Jiang Liu[2,4]

[1] ShanghaiTech University
[2] Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences
[3] Inception Institute of Artificial Intelligence
[4] Southern University of Science and Technology
[5] UBTech Research

ABSTRACT

Diabetic retinopathy (DR), as a leading ocular disease, is often with a complication of diabetic macular edema (DME). However, most existing works only aim at DR grading but ignore the DME diagnosis, but doctors will do both tasks simultaneously. In this paper, motivated by the advantages of multi-task learning for image classification, and to mimic the behavior of clinicians in visual inspection for patients, we propose a feature Separation and Union Network (SUNet) for simultaneous DR and DME grading. Further, to improve the interpretability of the disease grading, a lesion regularizer is also imposed to regularize our network. Specifically, given an image, our SUNet first extracts a common feature for both DR and DME grading and lesion detection. Then a feature blending block is introduced which alternately uses feature separation and feature union for task-specific feature extraction, where feature separation learns task-specific features for lesion detection and DR and DME grading, and feature union aggregates features corresponding to lesion detection, DR and DME grading. In this way, we can distill the irrelevant features and leverage features of different but related tasks to improve the performance of each given task. Then the task-specific features of the same task at different feature separation steps are concatenated for the prediction of each task. Extensive experiments on the very challenging IDRiD dataset demonstrate that our SUNet significantly outperforms existing methods for both DR and DME grading.

***Index Terms*—** Multi-disease diagnosis, Lesion regularization, Feature blending

1. INTRODUCTION

Diabetic retinopathy (DR), as a common complication of diabetes, has become the most common leading cause of preventable blindness in the working-age population of the world [1][2][3]. As the emergence of deep learning, the performance of automated DR grading system has been greatly improved. However, most works only focus on DR grading. Actually, DR is often accompanied by diabetic macular edema (DME), a common complication of DR, which is the most common cause of visual loss in both proliferative and nonproliferative diabetic retinopathy [4]. When diagnosing DR, clinicians would also evaluate the risk of DME at the same time. Therefore, it is desirable for an intelligent computer vision algorithm to grade DR and DME simultaneously. Thus this paper aims to consider the relation information of DR and DME and propose a framework for multi-disease detection.

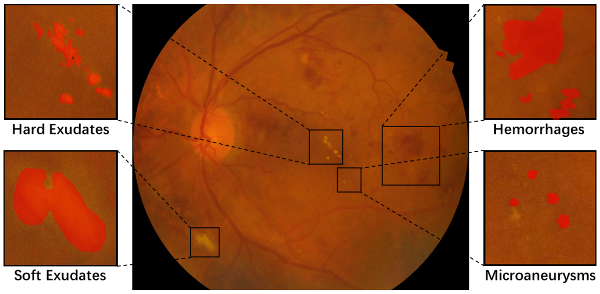

Fig. 1. Lesions of DR and DME, in which hard exudates are the cue for both DR and DME diagnosis.

Besides the aforementioned reason for simultaneous DR and DME grading, another reason comes from the possible performance boost introduced by the multi-task learning when conducting these two correlated tasks [5][6][7]. As shown in Fig.1, DR and DME share some similar patterns for some regions, for example, the presence of hard exudates is the

This work is supported by the ShanghaiTech-UnitedImaging Joint Lab, partially supported by ShanghaiTech-Megavii Joint Lab and supported in part by the National Natural Science Foundation of China (NSFC) under Grants No. 61932020.

cue for both DR and DME. In other words, these two tasks share common features. Many previous works have shown that jointly learning related tasks would improves the robustness of features and consequently improve the performance of both tasks for medical image analysis.

In practice, a clinician makes his/her decision by identifying the suspicious regions. To mimic the behavior of a clinician in diagnosing the patient, we propose to introduce a segmentation regularizer into our simultaneous DR and DME grading framework. There are two advantages for such segmentation regularizer: firstly, by using the common features of the two tasks for lesion segmentation, we can guarantee the features learnt concentrate more on the lesions thus help the subsequent DR and DME grading tasks; secondly, the lesion map estimated from the image would help clinicians better understand the classification results of the algorithm and makes the final decision.

Most previous deep learning based multi-task learning framework learns a common feature for all tasks first, then the common feature is fed into different sub-networks corresponding to different tasks to learn task-specific features [6][8], and task-specific features will not interact with each other any mores. Though such network architecture has shown its success, it is not easy to determine where is the optimal position for splitting the shared backbone network into different subnetworks. Splitting the backbone too early will not fully use the dependence between different tasks, and splitting the backbone too late will make the task-specific feature less discriminative for each individual task. To avoid this issue, a feature Separation and Union Network (SUNet) is proposed. In SUNet, a feature blending block alternately uses feature separation step and feature union step, where feature separation learns task-specific features for lesion detection and DR/DME grading, and feature union aggregates features corresponding to lesion detection and DR/DME grading. It is worth noting that the task-specific features at different feature separation steps contain knowledge of different tasks extracted at different levels. Then we concatenate all task-specific features of the same task for the prediction of each task. In this way, our SUNet alleviates the issue in determining the location for splitting the backbone network, meanwhile leverages the dependence of different tasks at different levels for task-specific feature extraction, thus improves the performance.

The contributions of our work are summarized as follows: i) we propose a multi-disease grading network (SUNet) for simultaneous grading DR and its complication DME. As far as we know, this is the first work for simultaneous DR and DME grading. ii) we introduce a lesion regularizer into the disease detection network which enforces network concentrates on those lesions, meanwhile segmentation also provides a cue for a doctor to better understand the prediction results; iii) we design a SUNet to interleave feature maps for the multi-disease diagnosis and lesion regularization, and it is a novel multi-task learning framework, which can be readily applied to other multi-task learning scenarios; iv) extensive experiments validate the effectiveness of our approach for simultaneous DR and DME grading.

2. PROPOSED METHOD

The proposed SUNet learns disease diagnosis from image-level supervision, meanwhile with pixel-level lesion regularization. It mimics the behavior of clinicians in disease diagnosis and lesion localization.

Our SUNet includes 4 parts as shown in Fig.2, 1) A feature extracting network to extract feature map for all tasks. The architecture of this subnetwork adopts Resnet-34 [9]. 2) A Feature Blending Block, which takes the encoded feature from the feature extracting block as input and blend the features for both disease diagnosis task and lesion regularization task. 3) A multi-disease diagnosis block to grade the severity of each disease. 4) A lesion regularizer block (LR-Net) to inspire the network to learn features of various kinds of lesions.

2.1. Feature Blending Block

The Feature Blending Block (F-Block) is a CNN block which takes the extracted feature map $\mathbf{M}_0$ as input. F-Block processes the feature map with a sequence of feature separation and feature union layers and generate two feature maps $\mathbf{L}, \mathbf{D}$ for the Lesion Regularize Net (LR-Net) and the Multi-Disease Diagnosis Block (MD-Block) respectively. Each feature separation layer S_i consists of a lesion layer (SL_i) and a diagnosis layer (SD_i). For any S_j, given a feature map $\mathbf{M}_j$ as input, we use SL_j and SD_j to generate two feature maps $\mathbf{L}_j$ and $\mathbf{D}_j$ respectively and pass them to the following feature union layer. Each feature union layer takes two feature maps and obtain a blended feature map $\mathbf{M}_{j+1}$ for the following feature separation layer. Every feature union layer, the two input feature maps are composed of covolutional layers with 3×3 kernel, batch normalization (BN) and Rectified Linear Units(ReLU).

$$(\mathbf{L}_j, \mathbf{D}_j) = [SL_j(\mathbf{L}_j), SD_j(\mathbf{M}_j)] = S_j(\mathbf{M}_j), \qquad (1)$$

$$\mathbf{M}_{j+1} = U(\mathbf{L}_j, \mathbf{D}_j), \qquad (2)$$

where U denotes feature union operation. All of SM_j, SL_j and S_j are covolutional layers with 3×3 kernel, batch normalization (BN) and Rectified Linear Units(ReLU).

To guarantee every task-specific feature maps to contain knowledge of corresponding tasks, we compress all the intermediate feature maps $\mathbf{L}_j$ with 1×1 convolution to obtain $\mathbf{L}$ and by the same means, we can calculate $\mathbf{D}$ with all the $\mathbf{D}_j$.

2.2. Multi-Disease Diagnosis Block

The Multi-Disease Diagnosis Block takes a shared feature map $\mathbf{D}$ from F-Block as input and processes it by two separate disease classifiers respectively and outputs a probability

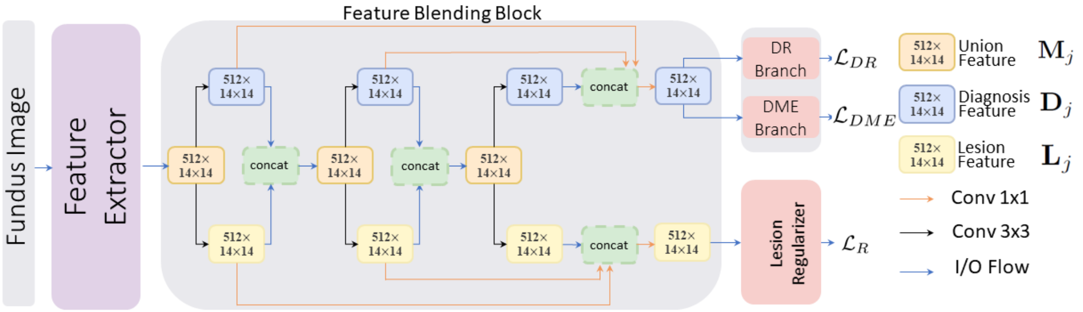

Fig. 2. The pipeline of SUNet. Our pipeline contains a feature extracting block, a lesion regularizer block, a multi-disease diagnosis block and a feature blending block to balance the feature for both lesion regularization and multi-disease diagnosis task.

distribution vector $\hat{y}$ for each disease. The vector $\hat{y}$ indicates the probabilities of the feature map belonging to each disease level. Each disease classifier is composed by convolution, max pooling and ReLU. At the end of the Multi-Disease Diagnosis Block, we use the weighted sum of cross entropy for all the disease's prediction

$$\mathcal{L}_{DR} = -\sum_{i=0}^{N-1} y_i \log(\hat{y}_i), \tag{3}$$

where y and $\hat{y}_i$ represent the disease severity level and the prediction of possibility of the corresponding level, N denotes the number of grade levels of DR. By the same means, we can calculate $\mathcal{L}_{DME}$.

2.3. Lesion Regularizer and alternately training

As aforementioned, we utilize pixel-level labels of lesion to guide the network with Lesion Regularizer Net. Given a feature map $\mathbf{L}$, we repeatedly upsample and compress the intermediate feature maps until we get a predicted lesion mask $\mathbf{M}_R$ in the end. Then we calculate lesion regularizer using cross entropy loss

$$\mathcal{L}_R = -\frac{1}{N}\sum\sum_{i=0}^{N_L-1} y_i \log(\hat{y}_i) \tag{4}$$

where y_i and $\hat{y}_i$ represent the lesion category and the prediction of possibility of the corresponding lesion.

However, as is stated afore, in the field of medical image analysis, pixel-level labels of lesion are hard to acquire. Therefore, we focus on utilizing a small scale of images with lesion mask labels. In our training set, every images would obtain diagnosis grades for all the diseases but only small percentage (φ) of the images have lesion mask labels. So an alternate training method is proposed. We divide the training set into two sets by whether they have lesion segmentation labels and use the two sets to train our model alternately. When the model is being trained by images with lesion masks, the total loss is calculated by the weighted sum of $\mathcal{L}_R$, $\mathcal{L}_{DR}$ and $\mathcal{L}_{DME}$.

$$\mathcal{L}_{total} = \mathcal{L}_R + \alpha\mathcal{L}_{DR} + \beta\mathcal{L}_{DME} \tag{5}$$

where α and β are hyper-parameters. For our experiments, we choose $\alpha = 1, \beta = 0.5$. When we train the model with the other training set without lesion masks, $\mathcal{L}_R$ would be set to zero.

3. EXPERIMENTS

3.1. Dataset and metrics

For experimentation, we use datasets of both challenge-1 (segmentation task) and challenge-2 (classification task) of IDRiD [10], which contains 516 images collected by retinal fundus camera Model Kowa VX-10α. The classification task dataset contains all of the 516 images, 413 for training and 103 for testing and the segmentation task dataset contains 81 images, 54 for training and 27 for testing. In the classification dataset, each image has grading groundtruth for two diseases, DR (5 classes) and risk of DME (3 classes).

In our multi-disease grading task, we evaluate the performance of models by the accuracy of predicting DR, DME respectively and the total accuracy of the two diseases. The total accuracy is defined as follows:

$$\textit{Total accuracy} = \frac{\mathrm{card}(R \cap E)}{\mathrm{card}(T)} \tag{6}$$

where R represents the set of images with correct DR grade prediction; E is to denote the set of images with correct DME grade prediction; T is the set of total 103 images in test set and card() is a function that returns the number of elements in a set.

3.2. Ablation study and comparison with other models

Ablation study of SUNet is shown in Table. 1. The multi-disease model outperforms single DR result but not the DME result. This is acceptable because in the multi-disease model, DR diagnosis accuracy becomes the bottleneck. For both DR and DME, lesion regularization improves the accuracy

Table 1. Ablation study on the design of our SUNet.

Method	total accuracy	DR accuracy	DME accuracy
DR	-	0.5049	-
DME	-	-	0.7961
DR+DME	0.5243	0.5825	0.7670
DR+LR[a]	-	0.6117	-
DME+LR	-	-	0.8155
SUNet	**0.6116**	**0.6506**	**0.8155**

[a] LR is short for Lesion Regularizer.

of disease grading by about 0.1 and 0.02 respectively. Combining the multi-disease classifier and lesion regularizer, our model achieves a total accuracy of 0.6116, nearly 0.09 higher than the model without lesion regularizer. The ablation study shows the effectiveness of our design combining multi-disease diagnosis and feature blending block.

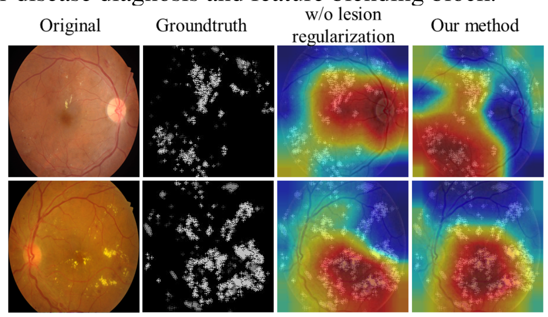

Fig. 3. Comparison of CAM between our model and a model without lesion regularization. (Best viewed with colors.)

Moreover, we use Class Activation Mapping (CAM) [11] generated by SUNet to obtain the lesion guidance map for clinicians to refer to, as shown in Fig.3. As we can see, without lesion regularization, the attention of the classifier sometimes can be misled to the position where optic disc or vessels lie due to the structure of retinal fundus images, as the first row of Fig.3 shows. Besides, in other cases, our model is able to focus on lesion areas more precisely as the second row shows. The comparison of CAM can also prove that lesion regularizer can help our model focus on main lesion areas and keep it from distractions caused by some special structures in fundus images.

The results of the comparison experiments are shown in Table. 2, in which all the experimental results are measured by the highest performance of all the models trained with training set. The multi-disease diagnosis results of VGG-Net [12], Res-Net34 [9] are obtained using two softmax functions to get the probability distribution vectors for both disease. Our method can outperform these baselines and lie in the second position in the ranking of the grand-challenge competitors. Besides, we compare our model with three multi-task learning baselines [9], in which, feature maps for different tasks separate at three different stages (early stage, in the middle and late stage, respectively) in the network. The performance of these three baselines reveals that the position for splitting the common features for different tasks, which affects the extent of blending features for related tasks, greatly affects the performance, but it is hard to be decided, while our feature blending block alleviates this issue.

Table 2. Comparison study of our model.

Method	total accuracy	DR accuracy	DME accuracy
VGG19	0.5631	0.5728	0.8350
ResNet34	0.5728	0.5728	0.8058
1st[a]	0.6311	-	-
2nd[a]	0.5534	-	-
3rd[a]	0.5146	-	-
early[b]	0.4951	0.5243	0.7670
middle[b]	0.5631	0.5728	0.7767
late [b]	0.5243	0.6019	0.7864
SUNet	**0.6116**	**0.6505**	**0.8155**

[a] The 1st, 2nd and 3rd are results of the top three participants in leaderboard of grand-challenge IDRiD.

[b] The early, middle and late correspond to the position of separating features for different tasks in multi-task learning framework at early state, middle stage and late stage.

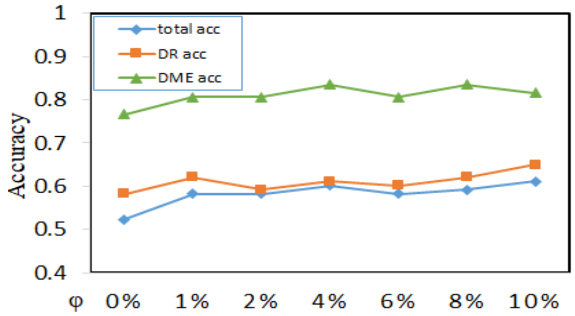

Fig. 4. Accuracy of DR, DME and total accuracy under different φ. φ is the percentage of data with lesion segmentation labels over all training data

3.3. Experiments with different scales of data for lesion regularization

To test the sensitivity of the model for the scale of data used in lesion regularization, we perform series of controlled trials testing our SUNet with different scales of data, measured by φ, the percentage of data with lesion segmentation labels over all training data. As can be seen in Fig.4, our SUNet can be improved by lesion regularization with a relatively small number of data. This property of our model would in practice be adopted to the situation of lacking lesion segmentation labels to better help with clinical disease diagnosis.

4. CONCLUSION

We propose a SUNet architecture to simultaneously grade DR and DME. Meanwhile, our SUNet predicts the lesion map for clinicians to refer to. Thus the results of our method are more explainable. Furthermore, our SUNet is a novel multi-task learning framework which alleviates the difficulty in determining the position of feeding common features into task-specific subnetworks. It can be readily used for the other multi-disease detection tasks.

5. REFERENCES

[1] Tien Y Wong, Jennifer Sun, et al., “Guidelines on diabetic eye care: the international council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings,” *Ophthalmology*, vol. 125, no. 10, pp. 1608–22, 2018.

[2] Ryan Lee, Tien Y Wong, et al., “Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss,” *Eye and vision*, vol. 2, no. 1, pp. 17, 2015.

[3] Martin M Nentwich and Michael W Ulbig, “Diabetic retinopathy-ocular complications of diabetes mellitus,” *World journal of diabetes*, vol. 6, no. 3, pp. 489, 2015.

[4] Rohit Varma, Neil M Bressler, et al., “Prevalence of and risk factors for diabetic macular edema in the united states,” *JAMA ophthalmology*, vol. 132, no. 11, pp. 1334–1340, 2014.

[5] C. Playout, R. Duval, et al., “A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images,” *TMI*, pp. 1–1, 2019.

[6] Clément Playout, Renaud Duval, et al., “A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images,” in *MICCAI*. Springer, 2018, pp. 101–108.

[7] Kang Zhou, Zaiwang Gu, Wen Liu, Weixin Luo, Jun Cheng, Shenghua Gao, and Jiang Liu, “Multi-cell multitask convolutional neural networks for diabetic retinopathy grading,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2724–2727.

[8] Rajeev Ranjan, Vishal M Patel, et al., “Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,” *TPAMI*, vol. 41, no. 1, pp. 121–135, 2019.

[9] Kaiming He, Xiangyu Zhang, et al., “Deep residual learning for image recognition,” in *CVPR*, 2016, pp. 770–778.

[10] Prasanna Porwal, Samiksha Pachade, et al., “Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research,” *Data*, vol. 3, no. 3, 2018.

[11] Bolei Zhou, Aditya Khosla, et al., “Learning deep features for discriminative localization,” in *CVPR*, 2016, pp. 2921–2929.

[12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.