DIABETIC RETINOPATHY DETECTION BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS

Yi-Wei Chen[1], *Tung-Yu Wu*[2], *Wing-Hung Wong*[2,3] *and Chen-Yi Lee*[1]

[1]Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan
[2]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA
[3]Department of Statistics, Stanford University, Stanford, CA, USA

ABSTRACT

Diabetic retinopathy is the primary cause of blindness in the working-age population of the developed world. Diagnosing the disease heavily relies on imaging studies, which is a time consuming and a manual process performed by trained clinicians. Enhancing the accuracy and speed of the detection process can potentially have a significant impact on population health via early diagnosis and intervention. Motivated by this, we propose a recognition pipeline based on deep convolutional neural networks. In our pipeline, we design lightweight networks called SI2DRNet-v1 along with six methods to further boost the detection performance. Without any fine-tuning, our recognition pipeline outperforms state of the art on the Messidor dataset along with 5.26x fewer in total parameters and 2.48x fewer in total floating operations.

Index Terms— Diabetic Retinopathy Detection, Deep Convolutional Neural Networks, Image Classification

1. INTRODUCTION

The World Health Organization (WHO) estimates that 422 million adults had diabetes worldwide in 2014 [1]. Diabetic retinopathy (DR) is an eye disease associated with long-standing diabetes. Nearly all patients with type 1 diabetes and more than 60% of patients with type 2 diabetes have DR [2]. Although up to 98% of severe vision loss due to DR can be prevented with early detection and treatment, once it has progressed, vision loss is often permanent [3]. Diagnosing the severity scales often requires a trained clinician to examine digital color fundus photographs of the retina, which is a time-consuming and manual process [4]. Motivated by this, we propose an auxiliary system based on the deep convolutional neural networks (DCNN) to speed up the detection process and help clinicians screen DR in the early stages, which is an important step toward lowering the risk of vision loss associated with the disease.

Work supported by MOST of Taiwan under 105-2218-E-009 -001, 106-2218-E-009 -009 and NSF of USA under Grant DMS-1407557 and DMS-1721550.

Convolutional neural networks (CNN) became popular in the 1990s due to its success of handwritten character recognition tasks, but then fell out of fashion with the rise of support vector machines (SVM). In 2012, [5] rekindled interest in CNN by showing a substantial improvement in image classification accuracy on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [6]. Since then, CNN or DCNN have been adopted widely and kept finding new successes in various fields of image analysis, including medical image analysis [7], classification [8], and segmentation [9].

Early works on DR detection rely on the design of handcrafted features, which tends to be complicated [10]. Recently, CNN-based methods had significantly improved the DR detection accuracy. Vo and Verma proposed two networks, CKML Net and VNXK, which are based on the networks of GoogLeNet and VGG16 respectively, along with input feature selection [11]. Wang et al. proposed the Zoom-in-Net and dealt with classification and localization problems simultaneously [12]. When it comes to a new testing dataset with a different grading rule, like the Messidor dataset, the above methods require fine-tuning. Techniques such as SVM with linear [11] or RBF [12] kernel need to be implemented on top of the extracted features from their proposed network. The fine-tuning process would be expensive and even infeasible for large dataset. Moreover, the networks used in the above methods are complex. Zoom-in-Net [12] required the limit of Tesla K40 GPU card (12GB) for a single prediction.

In this paper, we introduce our lightweight networks, SI2DRNet-v1, which improves the detection performance in two aspects: (1) Without any fine-tuning, SI2DRNet-v1, combined with six other boosting methods achieved 0.959 and 0.965 area under curve (AUC) on the Messidor dataset for referable and non-referable screening, which outperforms state of the art (0.921 and 0.957) [12]. (2) SI2DRNet-v1 requires less than 700 MB GPU memory for a single prediction, which is much less than Zoom-in-Net (12GB) [12].

The remainder of this paper is organized as follows. Section 2 introduces the proposed system framework; Section 3 presents the experiment details and results; Section 4 concludes the report with a discussion on the results.

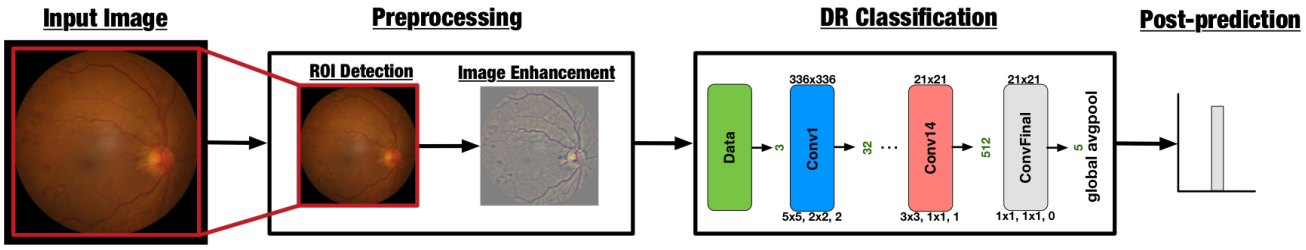

Fig. 1: System overview of our proposed DR recognition pipeline based on deep convolutional neural networks.

2. PROPOSED FRAMEWORK

Fig. 1 shows the overall DR detection system. It consists of three important components: the preprocessing step, the classification network, and the post-prediction. The preprocessing step can be further divided into the region of interest (ROI) detection and the image enhancement. The ROI detection filters out useless information, while the image enhancement emphasizes features relevant to DR symptoms, such as microaneurysms and hemorrhage. The classification network aims to classify the severity scales of DR. Finally, the post-prediction handles the problem of different DR grading systems efficiently. The details for each part are described in Section 2.1 - 2.4.

2.1. Region of Interest Detection

The raw fundus images contain useless background information. Based on the assumption that the background tends to be black, we perform Otsu's threshold [13] to find the regions of interest (ROI). Fig. 1 (ROI detection) shows the result after performing ROI detection.

2.2. Image Enhancement

Different contrast, color, and illumination inside the retinal fundus images are considered as main confounding factors in the detection of DR. [14] found that the visual appearance of images may be significantly improved by emphasizing its high-frequency contents to enhance the edges and detailed information within. The formulation of the classic linear unsharp masking (UM) is given by:

$$y(n,m) = x(n,m) + \lambda g(n,m) \tag{1}$$

where $y(n,m)$ is the enhanced image obtained from the input image $x(n,m)$, $g(n,m)$ is the correction signal computed as the output of a linear highpass filter and λ is the positive scaling factor that controls the level of contrast enhancement achieved at the output. Then, we adopt the method proposed by [15], and implement $g(n,m)$ as:

$$g(n,m) = 4[G(n,m,\sigma) * x(n,m) - x(n,m)] \tag{2}$$

where $G(n,m,\sigma)$ is a Gaussian filter with σ equals to $\frac{30}{r}$, r is the radius of ROI of the fundus image, $*$ denotes the convolution operator, and λ is set to 4. To further reduce contrast and illumination problems, the $x(n,m)$ is replaced with constant 128 to remove the unwanted DC component and map the background to gray color. Fig. 1 (image enhancement) shows the result after performing preprocessing.

2.3. The DR Classification Network: SI2DRNet-v1

The purpose of the classification network is to identify the severity scales of DR. Our proposed classification model is based on previous DCNN designs. Our model consists of 15 convolutional layers and 5 pooling layers with 10.4 million parameters. We use mostly 3 x 3 filters and double the number of channels after each pooling layer following the strategy used in [16]. To reduce the number of parameters and regularize the model, we use global average pooling and 1 x 1 filters to replace fully connected layers [17]. We also use batch normalization [18] after each convolutional layer to speed up convergence and reduce the generalization gap. We also found that scaling the kernel size of convolutional layer after each pooling layer from 3 x 3 to 5 x 5 increases the performance. The larger kernels provide larger receptive field and denser connections which are good for the final global averaging layer. Table 1 shows the detailed architecture of SI2DRNet-v1.

2.4. Post-prediction

Five probability values are extracted from the softmax layer, and summed up according to the following formula:

$$y_{pp} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 \tag{3}$$

where y_{pp} is the post-prediction value, p_0, p_1, p_2, p_3, and p_4 are the probabilities of normal, mild, moderate, severe , and proliferative DR. Then, we can decide new thresholds according to our objective function, such as quadratic weighted kappa which is more flexible than fine-tuning. For example, a five classification problem needs four thresholds, $t_{01}, t_{12}, t_{23}, t_{34}$ ($0 \leq t_{01} < t_{12} < t_{23} < t_{34} < 4$). If y_{pp} is larger than t_{01} but smaller than t_{12}, the predicted class would be class 1. The same rule applies to other thresholds. Noted that y_{pp} is bounded from 0 to 4, and we use Nelder-Mead algorithm [19] to solve the optimization problem.

Table 1: SI2DRNet-v1

Type	Filters	Size/Stride	Output
Convolution	32	5x5/2	336x336
Convolution	32	3x3	336x336
Max Pooling		3x3/2	168x168
Convolution	64	5x5	168x168
Convolution	64	3x3	168x168
Convolution	64	3x3	168x168
Max Pooling		3x3/2	84x84
Convolution	128	5x5	84x84
Convolution	128	3x3	84x84
Convolution	128	3x3	84x84
Max Pooling		3x3/2	42x42
Convolution	256	5x5	42x42
Convolution	256	3x3	42x42
Convolution	256	3x3	42x42
Max Pooling		3x3/2	21x21
Convolution	512	5x5	21x21
Convolution	512	3x3	21x21
Convolution	512	3x3	21x21
Convolution	5	1x1	21x21
Global Avg. Pooling		21x21	1x1

3. EXPERIMENT

We implement our proposed system in Caffe [20]. In this section, we describe the experimental setup and results as follows: Section 3.1 and Section 3.2 introduce the dataset and evaluation metrics; Section 3.3 describes six methods to boost the recognition performance on the EyePACS dataset; Section 3.4 presents the testing results on the Messidor dataset.

3.1. Details of Dataset

We evaluate the proposed framework on two public datasets: EyePACS and Messidor.

EyePACS Dataset: The EyePACS dataset is sponsored by the California Healthcare Foundation and used in the Kaggles Diabetic Retinopathy Detection Challenge [4]. The competition organizer generously made the dataset public. It provides 35k, 11k, and 43k images for train, validation, and test set respectively. Based on the presence of diabetic retinopathy, a clinician labeled each image on a severity scale from 0 to 4, which represents normal, mild, moderate, severe, and proliferative DR respectively. We follow the same definition as [12], defining mild to proliferative DR as the DR class and moderate to proliferative DR as the referable DR (RDR) class for the better comparison with the previous works.

Messidor Dataset: The Messidor dataset is a public dataset provided by the Messidor research program [21]. It consists of 1200 retinal images and provides a retinopathy grade for each image from 0 to 3, which is different from EyePACS dataset. To better compare with the previous works, we adopt the same definition as [12], defining grade 1 to 3 as DR class and grade 2 to 3 as RDR class.

Table 2: Four basic evaluation measurement

Prediction	Actual class	
	$\in$ DR (RDR)	$\notin$ DR (RDR)
$\in$ DR (RDR)	TP	FP
$\notin$ DR (RDR)	FN	TN

Table 3: Five major evaluation metrics

Measure	Formula
Specificity (TNR)	$\frac{TN}{TN+FP}$
Sensitivity (TPR)	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$
AUC	$\int_{-\infty}^{\infty} TPR(T)FPR^{'}(T)dT$
κ	$1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,j}O_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,j}E_{i,j}}$

*FPR = 1 - Specificity, T = threshold
*N = number of classes, $w_{i,j}$ = weight matrices
*$E_{i,j}$ = expected matrices, $O_{i,j}$ = observed matrices

3.2. Evaluation Metrics

We utilize five metrics to evaluate the performance of our proposed framework: specificity, sensitivity, accuracy, area under curve (AUC) of receiver operating characteristic (ROC), and quadratic weighted kappa (κ). In the following, κ refers to quadratic weighted kappa for simplicity. Based on the definition of four basic measurements in Table 2, namely true positive (TP), false positive (FP), true negative (TN), and false negative (FN), we can derive four major evaluation metrics (specificity, sensitivity, accuracy, AUC) as listed in Table 3. Since the scale of DR severity has multiple levels, we introduce κ to compute a weighted measure for assessing classification accuracy.

3.3. Ablation Study on The EyePACS Dataset

We train our baseline network on EyePACS train dataset for 80 epochs using stochastic gradient descent with a starting learning rate of 0.001, step rate decay for every 20 epochs, gamma of 0.5, weight decay of 0.0005 and momentum of 0.9 using the SI2DRNet-v1 network with input size 224x224. During training, we only use random crops for data augmentation. Then, we use the central crop of the image to get the testing result as the baseline. We evaluate six methods to further boost the recognition accuracy as below and compare the results from different settings in Table 4.

Use a pre-train model: The SI2DRNet-v1 is first trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [6] dataset, then fine-tuned on the EyePACS train dataset. With a pre-training, the SI2DRNet-v1 converges faster and increases κ by 31% for the validation set and 27% for the test set.

Image enhancement: Using the image enhancement method described in Section 2.2 can improve κ by 9% for the validation set and 7% for the test set.

Table 4: Ablation study of the proposed recognition pipeline on the EyesPACS dataset

	Baseline								SI2DRNet-v1
Use pre-train model		√	√	√	√	√	√	√	√
Image enhancement			√	√	√	√	√	√	√
More data augmentation				√	√	√	√	√	√
L1 norm					√	√	√	√	√
Post-prediction						√	√	√	√
Scale input resolution (2x)							√		
Scale input resolution (3x)								√	√
10 crops for testing									√
EyesPACS validation set (κ)	0.466	0.611	0.671	0.707	0.723	0.754	0.801	0.808	0.808
EyesPACS test set (κ)	0.471	0.601	0.654	0.704	0.709	0.742	0.796	0.802	0.804

More data augmentation: Besides random crop, we include random rotation (60°), random zoom in (0.2), and random shear (0.2), which improve κ by 5% for the validation set and 8% for the test set.

L1 norm: L1-norm imposes more sparsity on neuron activation than weight decay (L2-norm), which is more consistent with the properties of DR symptoms. With L1-norm penalty 0.000074, the κ increases by 2% for the validation set and 0.7% for the test set.

Post-prediction: As described in Sec 2.4, by using κ as the objective function and findind four thresholds based on 10% of the EyePACS train dataset, we can improve κ by 4% for both the validation and test set.

Scale input resolution: Scaling up the input resolution can preserve more information contained in the images. If we double the input resolution (448x448), κ increases by 6% and 7% for the validation and test set respectively. If we triple the input resolution (672x672), κ increases by 7% and 8% for the validation and test set respectively. The performance gain from resolution scale-up saturates at 672x672.

With the above six methods and 10 crops for testing, we arrive at optimal performance of SI2DRNet-v1.

3.4. Experiment Results on The Messidor Dataset

We follow the same evaluation procedure as [12] and conduct two binary classifications for fair comparison with previous works. A key advantage of our method is that we do not need to train an extra SVM on top of the softmax layer; All we need is to compute new thresholds based on the given dataset. Then, we use AUC to quantify the performance. Table 5 compares the results of our method with previous works. To the best of our knowledge, we achieve the highest AUC for both DR and RDR classification on the Messidor dataset. At a specificity of 0.5, the sensitivity of SI2DRNet-v1 is 0.978 and 0.984 for the DR and RDR tasks, which outperform state of the art (0.960 and 0.978) [12].

3.5. Model Complexity Analysis

In addition to the detection performance, we also compare the number of learned parameter and floating point operations (FLOPs) per forwarding of our SI2DRNet-v1 with other networks. Table 6 shows the comparison results. We approximate VNXK with one VGG16 [11], CKML Net with three GoogLeNet [11], and Zoom-in-Net with one Inception-resnet-v2 [12]. Note that the details of above networks are not released, hence our approximation may be an underestimation. However, compared with CKML Net [11], VNXK [11], and Zoom-in-Net [12], our SI2DRNet-v1 is 6.74x, 47.86x, and 5.26x fewer in total parameters, and 1.24x, 4.11x, and 2.48x fewer in total FLOPs.

Table 5: Performance comparison on the Messidor dataset

Method	DR		RDR	
	Acc.	AUC	Acc.	AUC
Fisher Vector [10]	-	-	-	0.863
VNXK [11]	0.871	0.870	0.893	0.887
CKML Net [11]	0.857	0.862	0.897	0.891
Comprehensive CAD [22]	-	0.876	-	0.91
Expert A [22]	-	0.922	-	0.94
Expert B [22]	-	0.865	-	0.92
Zoom-in-Net [12]	0.905	0.921	0.911	0.957
SI2DRNet-v1	**0.905**	**0.959**	**0.912**	**0.965**

Table 6: Model complexity comparison

Network	Input size	Params	FLOPs
CKML Net [11]	451x451	71.5M	19.2G
VNXK [11]	449x449	507.4M	63.4G
Zoom-in-Net [12]	492x492	55.8M	38.2G
SI2DRNet-v1	672x672	**10.6M**	**15.4G**

4. CONCLUSION

In this paper, we present a framework based on DCNN for the DR detection. Along with six useful methods, the proposed framework achieves 0.959 and 0.965 AUC for DR and RDR cases on the Messidor dataset which outperform state of the art (0.921 and 0.957) [12]. Furthermore, we are able to achieve this performance with a lightweight model. Compared with CKML Net [11], VNXK [11], and Zoom-in-Net [12], SI2DRNet-v1 is more memory efficient with at least 5.26x fewer in total parameters and requires lower computation cost with at least 1.24x fewer in total FLOPs.

5. REFERENCES

[1] "World health organization: Global report on diabetes," http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf?ua=1, Accessed: 2017-06-27.

[2] Donald S. Fong, Lloyd Aiello, Thomas W. Gardner, George L. King, George Blankenship, Jerry D. Cavallerano, Fredrick L. Ferris, and Ronald Klein, "Retinopathy in diabetes," *Diabetes Care*, vol. 27, no. suppl 1, pp. s84–s87, 2003.

[3] Lisa Crossland, Deborah Askew, Robert Ware, Peter Cranstoun, Paul Mitchell, Andrew Bryett, and Claire Jackson, "Diabetic Retinopathy Screening and Monitoring of Early Stage Disease in Australian General Practice: Tackling Preventable Blindness within a Chronic Care Model," *J. Diabetes Res.*, vol. 2016, pp. 8405395, dec 2016.

[4] "Kaggle: Diabetic retinopathy detection," https://www.kaggle.com/c/diabetic-retinopathy-detectionua=1, Accessed: 2017-06-29.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Adv. Neural Inf. Process. Syst. 25*, F Pereira, C J C Burges, L Bottou, and K Q Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li, "ImageNet: A large-scale hierarchical image database.," in *CVPR*. 2009, pp. 248–255, IEEE Computer Society.

[7] T Liu, S Xie, J Yu, L Niu, and W Sun, "Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features," in *2017 IEEE Int. Conf. Acoust. Speech Signal Process.*, 2017, pp. 919–923.

[8] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-Excitation Networks," *arXiv Prepr. arXiv1709.01507*, 2017.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B Girshick, "Mask r-cnn," *CoRR*, vol. abs/1703.0, 2017.

[10] Ramon Pires, Sandra Avila, Herbert F Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha, "Beyond lesion-based diabetic retinopathy: a direct approach for referral," *IEEE J. Biomed. Heal. informatics*, vol. 21, no. 1, pp. 193–200, 2017.

[11] H H Vo and A Verma, "New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space," in *2016 IEEE Int. Symp. Multimed.*, 2016, pp. 209–215.

[12] Zhe Wang, Yanxin Yin, Jianping Shi, Wei Fang, Hongsheng Li, and Xiaogang Wang, "Zoom-in-Net: Deep Mining Lesions for Diabetic Retinopathy Detection," *CoRR*, vol. abs/1706.0, 2017.

[13] N Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[14] A Polesel, G Ramponi, and V J Mathews, "Image enhancement via adaptive unsharp masking," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 505–510, 2000.

[15] B. Graham, "Kaggle diabetic retinopathy detection competition report," https://kaggle2.blob.core.windows.net/forum-message-attachments/88655/2795/competitionreport.pdf, 2015, Accessed: 2017-07-06.

[16] Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1, 2014.

[17] Min Lin, Qiang Chen, and Shuicheng Yan, "Network In Network," *CoRR*, vol. abs/1312.4, 2013.

[18] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[19] John A Nelder and Roger Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[20] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimed.* 2014, pp. 675–678, ACM.

[21] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein, "Feedback on a publicly distributed database: the Messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[22] Clara I Sánchez, Meindert Niemeijer, Alina V Dumitrescu, Maria S A Suttorp-Schulten, Michael D Abramoff, and Bram van Ginneken, "Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 7, pp. 4866–4871, 2011.