Exploratory Study on Direct Prediction of Diabetes using Deep Residual Networks

Samaneh Abbasi-Sureshjani, Behdad Dashtbozorg, Bart M. ter Haar Romeny, and François Fleuret

Abstract Diabetes is threatening the health of many people in the world. People may be diagnosed with diabetes only when symptoms or complications such as diabetic retinopathy start to appear. Retinal images reflect the health of the circulatory system and they are considered as a cheap and patient-friendly source of information for diagnosis purposes. Convolutional neural networks have enhanced the performance of conventional image processing techniques significantly by neglecting inconsistent feature extraction pipelines and learning informative features automatically from data. In this work we explore the possibility of using the deep residual networks as one of the state-of-the-art convolutional networks to diagnose diabetes directly from retinal images, without using any blood glucose information. The results indicate that convolutional networks are able to capture informative differences between healthy and diabetic patients and it is possible to differentiate between these two groups using only the retinal images. The performance of the proposed method is significantly higher than human experts.

Key words: Retinal images, Diabetes, Diabetic retinopathy, Deep learning, ResNet

S. Abbasi-Sureshjani, B. Dashtbozorg and B. M. ter Haar Romeny
Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB, The Netherlands, e-mail: {S.Abbasi,B.Dasht.Bozorg}@tue.nl

B. M. ter Haar Romeny
Department of Biomedical and Information Engineering, Northeastern University, Shenyang 110000, China, e-mail: Bart.Romeny@outlook.com

F. Fleuret
Computer Vision and Learning Group, Idiap Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland e-mail: Francois.Fleuret@idiap.ch

1 Introduction

Diabetes is a group of metabolic diseases resulting from defects in insulin secretion (type 1 diabetes), insulin action (type 2 diabetes), or both. If it is not controlled, it leads to long-term damages, dysfunctions, and failure of different organs, especially the eyes, kidneys, nerves, heart, and blood vessels [2]. One of the main long-term complications is retinopathy with potential loss of vision. Since visual loss may not be present in the earlier stages of retinopathy, regular screening of persons with diabetes is essential to enable early intervention. For decades, the diagnosis of diabetes has been based on glucose criteria [15]. However, non-invasive and easy-to-access approaches are favorable in large screening settings.

Retinal images are often used to develop automatic diagnosis systems and study the progression of different diseases such as diabetic retinopathy, age-related macular degeneration, glaucoma and retinopathy prematurity [12, 1]. The extraction of vascular biomarkers is not straightforward, and often relies on a series of preceding image processing tasks including vessel segmentation, vessel width measurement and artery-vein classification [4, 5]. Potential errors in the processing pipeline may accumulate, and the final extracted biomarkers may become unreliable. Due to contradictory and inconsistent results, it is still not clear to clinicians how the vasculature changes in patients are developing with diabetic retinopathy. Moreover, no studies have investigated whether it is possible to use retinal images to differentiate between the images of healthy and those of diabetic subjects before diagnosis of retinopathy.

Deep learning is rapidly becoming the state-of-the-art in various medical applications including image classification, segmentation, localization and registration [14, 7]. One of the main reasons behind the outstanding performance of Convolutional Neural Networks (CNNs) compared to conventional approaches is that features are learned from data automatically, instead of being handcrafted. In this work we use CNNs, in particular the deep residual nets (ResNets) [9], to investigate if the retinal images of healthy and diabetic subjects are differentiable. The residual networks are easier to optimize, and can gain accuracy from considerably increased depth, with much less complexity compared to other state-of-the-art architectures [9].

2 Material

We used a subset of a private dataset collected in the Maastricht Study[1]. This is a large phenotyping study focusing on type 2 diabetes, comprising subjects that live in the southern part of the Netherlands [13]. This subset includes 8924 good quality images of left and right eyes of 2336 subjects (1150 males and 1186 females, aged between 40 and 76), which are centered either on the fovea or on the optic disc. The

[1] https://www.demaastrichtstudie.nl/research

images are taken with a non-mydriatic autofocus fundus camera (Model AFC-230, Nidek). The images are categorized into two groups, either healthy (5791 images), or type 2 diabetic subjects (3133 images), based on blood sugar level tests.

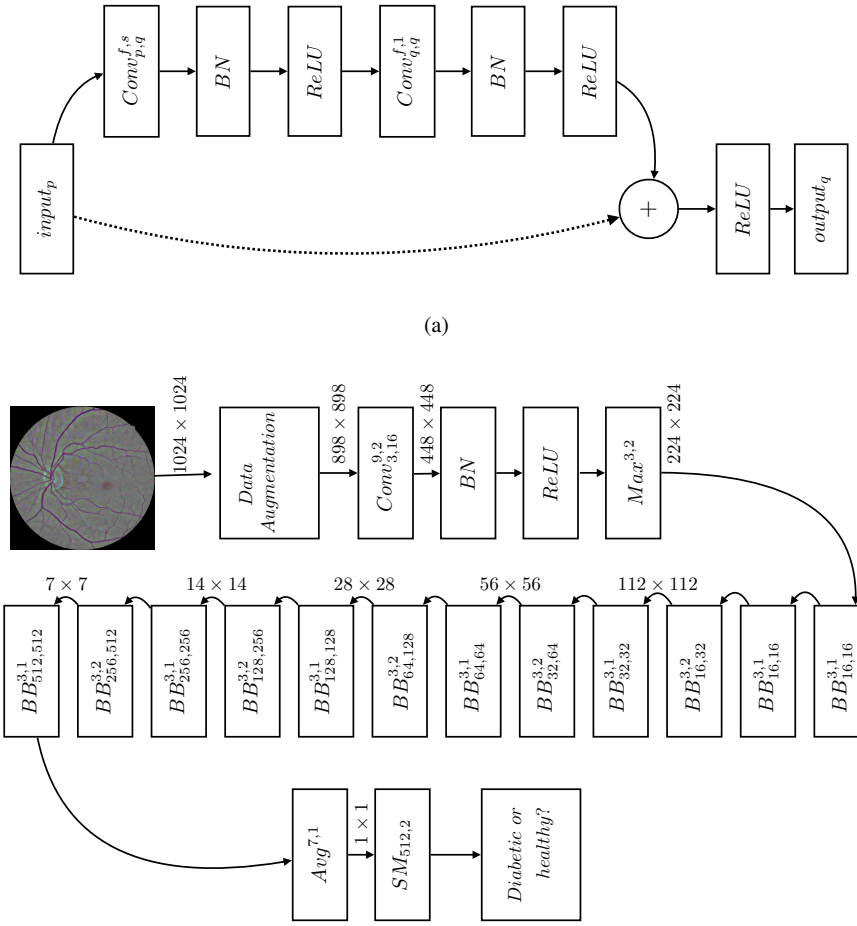
3 Methodology

The network follows the structure of the residual networks proposed by [9]. It chains several blocks, each consisting of two convolutional layers, batch normalization [10], rectified linear units [6], and a "pass-through" which adds the unchanged input.

One basic block ($BB^{f,s}_{p,q}$) is shown in Fig. 1a, and the complete model used in this work is depicted in Fig. 1b. In this figure, $Conv^{f,s}_{p,q}$ represents a 2D convolutional layer [11], where f, p, q and s represent respectively the filter size ($f \times f$), the number of input planes, the number of output planes, and the convolution step size. BN is a batch normalization layer [10] and $ReLU$ is the rectified linear unit [6]. $Avg^{k,s}$ and $Max^{k,s}$ apply a 2D average or max pooling operation in $k \times k$ regions by step size $s \times s$. Finally $SM_{p,q}$ represents the softmax classifier (linear unit with softmax operation at the end) with p and q as the input and output sizes. For each $BB^{f,s}_{p,q}$ if $p = q$ then $s = 1$ and the shortcut between input and output of the block is an identity map. Otherwise, the shortcut is implemented with a convolutional layer ($Conv^{f,s}_{p,q}$) with stride $s = 2$. This model has 26 weighted layers with 11,025,570 parameters. It needs an input image of size of 898×898, and it has a 2-class softmax classifier at the end. A cross entropy criterion is used for measuring the loss value.

Before feeding the images to the network, several pre-processing and data augmentation steps are applied. The images have varying resolutions, so the first step consists of cropping black borders and re-scaling all images to the resolution of 1024×1024 pixels, so that the aspect ratio remains unchanged. Data augmentations then include random affine transformation, random cropping to 85-95% of the initial size, horizontal flipping and random rotation between 0 to 360 degrees. When these transformations have been applied, images are rescaled to the desired model input size. At the end, a channel-wise global contrast normalization for each image (by subtracting mean and division over standard deviation) is used for normalizing the data.

We split the dataset into two 80/20% parts that we use as training and validation sets. The augmentation steps are applied to both sets similarly. Since there is more than one image per subject available, the average predictions of all images of each subject (besides testing each image 25 times) is used for evaluating the performance of the method. The weights are initialized as in [8] and training was done from scratch. We used SGD optimization technique with a weight decay of 0.0001 and a momentum of 0.9. The learning rate was fixed to 0.001 for 100 epochs, then decreased it to 0.0001 and trained the network for 200 more epochs.



(a)

(b)

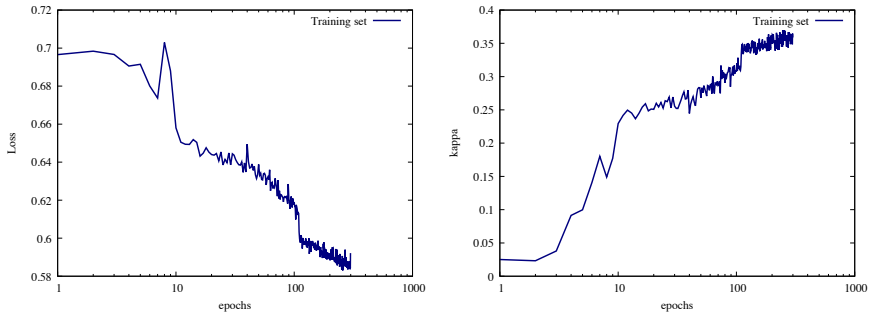
Fig. 1 The basic block ($BB_{p,q}^{f,s}$) of residual nets (1a), and the complete structure of the network used in this work (1b).

4 Results

For evaluating the performance of the method we use the weighted Cohen's kappa with quadratic weights and the F1-score. κ statistic compares the accuracy of the system to the accuracy of a random system. Complete agreement corresponds to $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance, then $\kappa \leq 0$ [3].

The evolution of training loss and kappa score during training time is depicted in Fig. 2. By testing the model on the validation set, our trained model is capable of

predicting the diabetes status directly from the retinal images (without any further information about other glucose measurements) with a κ score of 0.458 and F1-score of 0.758. In order to compare this performance, we asked an ophthalmologist expert to predict the diabetes status of a subset of 32 images (16 diabetic and 16 healthy). Only 2 out of 16 diabetic images could be detected by the expert and the rest was labeled as healthy, which corresponds to a κ score of 0.125 and F1-score of 0.222. The kappa score of the validation set is a bit higher than the training set, because the final predictions over multiple eyes per subjects are averaged during test time. This makes the prediction more accurate.



(a) (b)

Fig. 2 The evolution of the training loss and kappa score per epoch

5 Conclusion

Currently, diagnosis of diabetes is based on glucose measurements, and retinal images are useful to observe the early signs of diabetic retinopathy such as micro-aneurysms and retinal haemorrhages. When these signs do not exist in the image, it is difficult for experts to differentiate between healthy and diabetic subjects. The preliminary results in this work show that CNNs are able to differentiate between these two groups only by using the retinal images, with a significantly better performance compared to ophthalmologists. This indicates that the retinal vasculature start to change in patients with diabetes in early stages and CNNs are able to capture these changes. By proper visualization techniques, extending the experiments on larger data-sets and including additional meta-data it may be possible to discover new biomarkers, which can be used for diagnosis of diabetes non-invasively. This is very advantageous in large screening settings and early blindness prevention.

Acknowledgements This project has received funding from the European Union's Seventh Framework Programme, Marie Curie Actions-Initial Training Network, under grant agreement No.

607643, "Metric Analysis For Emergent Technologies (MAnET)". The authors would like to thank Dr. Tos Berendschot and Dr. Jan Schouten from University Eye Clinic Maastricht for providing the fundus images and clinical data used during this research.

References

1. Abràmoff, M.D., Garvin, M.K., Sonka, M.: Retinal imaging and image analysis. IEEE reviews in biomedical engineering **3**, 169–208 (2010)
2. American Diabetes Association: Diagnosis and classification of diabetes mellitus. Diabetes care **33**(Supplement 1), S62–S69 (2010)
3. Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin **70**(4), 213 (1968)
4. Dashtbozorg, B., Abbasi-Sureshjani, S., Zhang, J., Huang, F., Bekkers, E., ter Haar Romeny, B.M.: Infrastructure for retinal image analysis. In: Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop (OMIA 2016) Held in Conjunction with MICCAI 2016 (2016)
5. Dashtbozorg, B., Zhang, J., Abbasi-Sureshjani, S., Huang, F., ter Haar Romeny, B.M.: Retinal health information and notification system (RHINO). In: SPIE Medical Imaging, pp. 10,313,437–1013,437–6. International Society for Optics and Photonics (2017)
6. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Aistats, vol. 15, p. 275 (2011)
7. Greenspan, H., van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging **35**(5), 1153–1159 (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034 (2015)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning (ICML-15), pp. 448–456 (2015)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
12. Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yogesan, K., Constable, I.J.: Retinal image analysis: concepts, applications and potential. Progress in retinal and eye research **25**(1), 99–127 (2006)
13. Schram, M.T., Sep, S.J., van der Kallen, C.J., Dagnelie, P.C., Koster, A., Schaper, N., Henry, R.M., Stehouwer, C.D.: The maastricht study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. European journal of epidemiology **29**(6), 439–451 (2014)
14. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering **19**(1) (2017)
15. Wilkinson, C., Ferris, F.L., Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J.T., et al.: Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology **110**(9), 1677–1682 (2003)