DAISY descriptors combined with deep learning to diagnose retinal disease from high resolution 2D OCT images

Joshua Bridge[1], Simon P. Harding[1], and Yalin Zheng[1]

Department of Eye and Vision Science, University of Liverpool, L7 8TX, UK
{jbridge, sharding, yzheng}@liverpool.ac.uk
www.liv-cria.co.uk

Abstract. Optical Coherence Tomography (OCT) is commonly used to visualise tissue composition of the retina. Previously, deep learning has been used to analyse OCT images to automatically classify scans by the disease they display, however classification often requires downsampling to much lower dimensions. Downsampling often loses important features that may contain useful information. In this paper, a method is proposed which incorporates DAISY descriptors as 'intelligent downsampling'. By avoiding random downsampling, we are able to keep more of the useful information to achieve more accurate results. The proposed method is tested on a publicly available dataset of OCT images, from patients with diabetic macula edema, drusen, and choroidal neovascularisation, as well as healthy patients. The method achieves an accuracy of 76.6% and an AUC of 0.935, this is an improvement to a previously used method which uses InceptionV3 with an accuracy of 67.8% and AUC of 0.912. This shows that DAISY descriptors do provide good representations of the image and can be used as an alternative to downsampling.

Keywords: DAISY descriptors · Deep learning · Retina · OCT · GRU

1 Introduction

Optical Coherence Tomography (OCT) is an imaging method commonly used to analyse tissue composition [1]. It has recently been shown that deep learning has the ability to detect several retinal diseases from OCT images [10, 6]. One challenge encountered when analysing OCT data is the high resolution of the scans, passing these images straight to a deep learning network often results in an out of memory error. Current methods often require downsampling to much lower resolutions to make computation practical. Conventional downsampling methods often lose important features [7]. Previous methods often focus on making the image appear similar to a human observer, who may rely on different features to those which a computer may recognise. The method proposed here incorporates DAISY image descriptors to greatly reduce image dimension, followed by a three layer Gated Recurrent Unit (GRU) network [5] to classify the images according to disease (see Fig. 1). The use of image descriptors provides a data

efficient alternative to downsampling and allows us to use more of the useful information contained in each image. The aim of this study is to demonstrate that DAISY descriptors can successfully represent images, acting as 'intelligent downsampling', we aim to demonstrate that the method is a viable alternative to downsampling.

We demonstrated the method on a publicly available dataset of OCT scans [10]. The OCT images are split into four groups; normal, drusen, choroidal neovascularisation (CNV), and diabetic macular edema (DME). The normal group have no visible disease from the OCT scan and are classed as being healthy. The drusen group have small lipid deposits in the macula, which are commonly found in older people. Patients with drusen are more likely to develop age related macular degeneration (AMD) in the future [12] AMD is a leading cause of vision loss in older patients and greatly affects daily activities. The CNV group have signs of CNV which is indicative of wet AMD [12], this is when new blood vessels begin to form in the choroid [9]. CNV can often lead to blindness, although there are some treatments options available [3]. The final group consists of patients with DME. DME is a common result of diabetic retinopathy and causes vision problems in patients with diabetes [2].

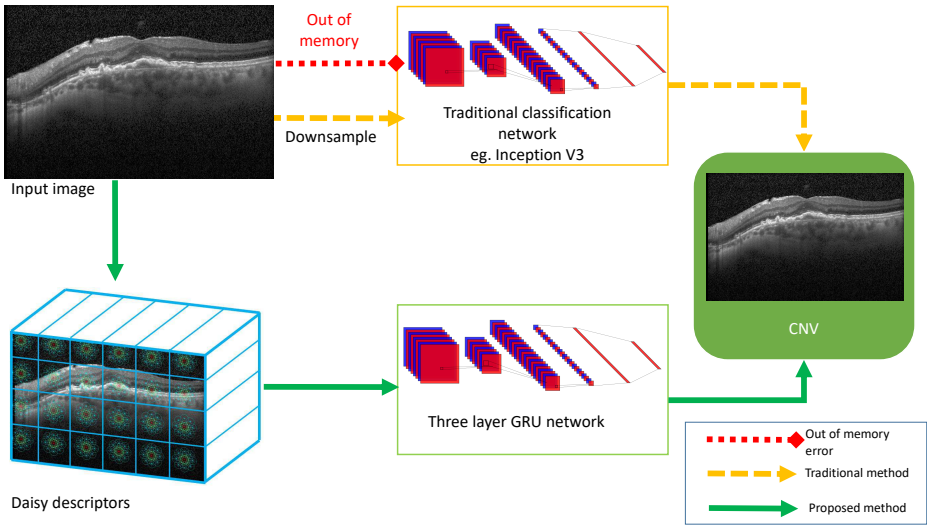


Fig. 1. The proposed framework aims to utilise DAISY descriptors as 'intelligent downsampling', followed by a GRU network to classify the images from the descriptors.

2 Methods

The method is extendible to many different types of images, here demonstrate the method using OCT, which has been likened to ultrasound, using light in-

stead of sound to produce a cross-sectional view of tissue composition [1]. Often OCT images are combined to produce a 3D representation of the tissue. OCT is commonly used to image the back of the eye (fundus) to diagnose eye disease.

2.1 Dataset

The data consists of OCT images, collected by Shiley Eye Institute and made publicly available [10]. In this dataset, a single OCT B-scan made up each image. Examples of OCT images are shown (see Fig. 3), depicting the 4 ocular diseases contained in this dataset. The original dataset contains 108,314 training images from 4,686 patients and 1,000 validation images from a separate 663 patients, split into 4 groups according to the disease type they display. In this preliminary work, we used a subset of this data to save time, 20,020 images were used for training, 4,112 images for validation, and the original 1,000 images were used for testing (Fig 2). Original images ranged in size from 512x496 to 1536x496. Images were first rescaled to 1500x1000 pixels, as all images must be the same size in this method.

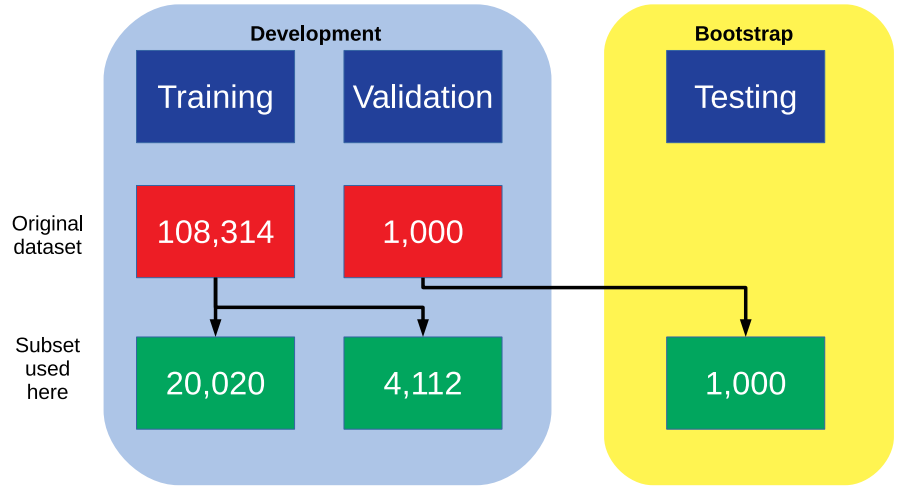


Fig. 2. Visual representation of the dataflow.

2.2 DAISY

DAISY is a method of image description which is similar to SIFT but is much faster to compute, due to it's implementation of Gaussian kernels. DAISY has previously been used for both classification [11] and matching problems[13, 14]. DAISY takes a greyscale image as an input, I, and creates orientation maps

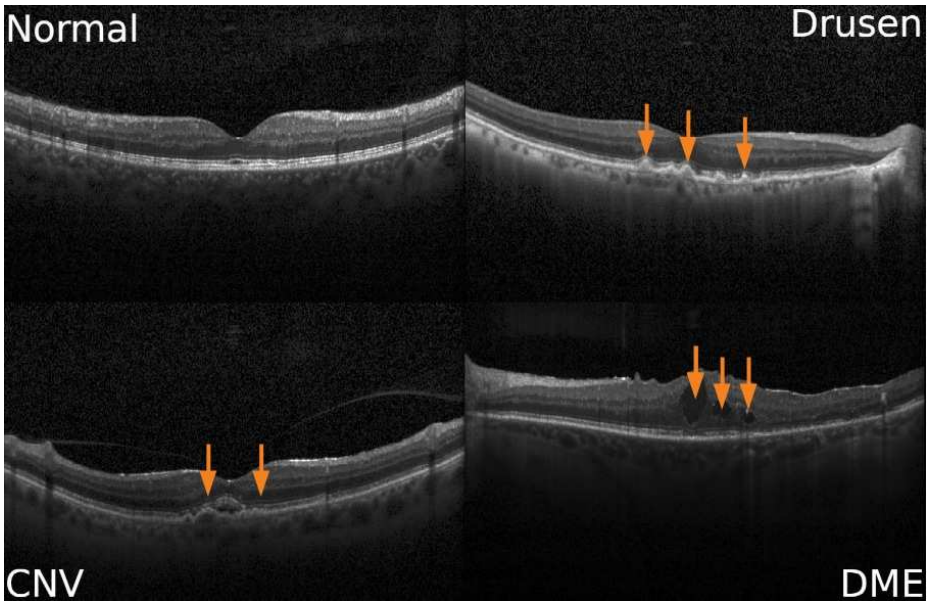


Fig. 3. Examples from each of the four groups included in the dataset. Arrows indicate areas highlighting the identified pathology.

using the gradient norm, G_O, for a specified number of directions, where O is the direction of the gradient. The orientation maps are calculated as:

$$G_O = max\left(0, \frac{\partial I}{\partial O}\right).$$

Gaussian kernels are then used to produce convolved orientation maps. The use of the Gaussian kernel in the convolutions makes the computation fast and efficient. Large Gaussian kernels can be calculated efficiently using many convolutions from much smaller kernels. For $\Sigma_1 < \Sigma_2$:

$$G_O^{\Sigma_2} = G_{\Sigma_2} * max\left(0, \frac{\partial I}{\partial O}\right) = G_{\Sigma} * G_{\Sigma_1} * max\left(0, \frac{\partial I}{\partial O}\right) = G_{\Sigma} * G_O^{\Sigma_1},$$

using $\Sigma = \sqrt{\Sigma_2^2 - \Sigma_1^2}$ [14].

At each pixel location, DAISY produces a vector of values obtained from the convolved orientation maps. The final output is a 3D tensor representation of the image, which is smaller than the original image. DAISY hyperparameters can be chosen to produce either sparse or dense descriptions of images, DAISY is mainly used for dense description as it is efficient compared to similar methods, such as SIFT and GLOH [14]. A visual representation of DAISY descriptors is displayed on an example image (Fig. 4).

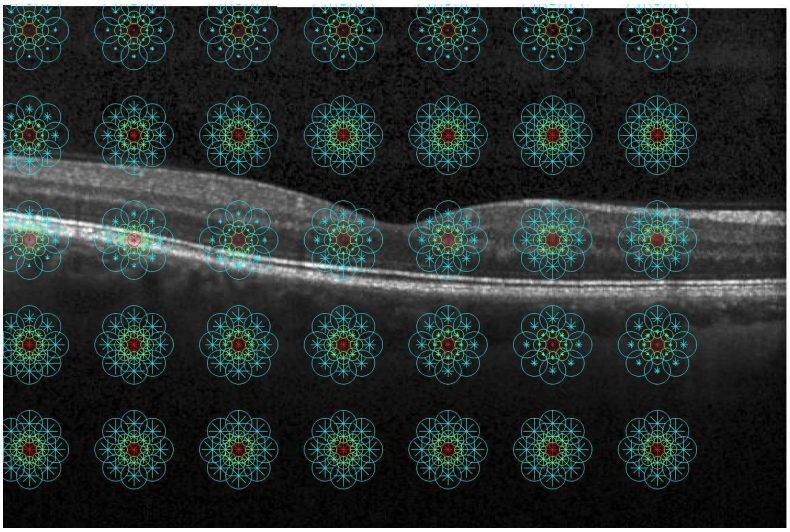


Fig. 4. Visual representation of DAISY descriptors on an example image. The DAISY hyperparameters were chosen to produce a sparse representation to clearly display the rings in DAISY. Each descriptor consists of a centre ring, 8 middle rings, and 8 outer rings, each with 8 directions. The output of the DAISY descriptors algorithm shown here is a $4 \times 6 \times 136$ tensor of descriptors. This example shows a sparse representation to highlight how DAISY describes the images.

2.3 Classification network

The classification network consists of a 3 layer GRU network [4]. GRU networks are a recurrent neural network, which uses gated units to control the flow of information, update and reset gates. GRU has been shown to perform better than older recurrent units such as tanh and is at least as good as LSTM [5].

This network was trained using Keras 2.2.4 on an Ubuntu 18.04 machine with a Titan X 12GB GPU and 32GB of memory. Training was performed for 500 epochs of 200 steps each using the Nesterov Adam optimiser, batch size was set to 32, with categorical cross-entropy used as the loss function. Parametric Rectified Linear Unit (PReLU) was used as the activation function for the hidden GRU layers and Softmax was used in the output layer. Early stopping with a patience of 10 epochs was used to prevent overfitting, model checkpoints were used to select the best classification model, based on the loss in the validation dataset.

2.4 Model performance

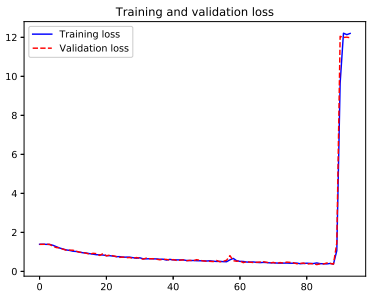
Model performance was assessed using loss, accuracy, and the Area Under the receiver operating Curve (AUC), with categorical cross-entropy used as the loss function. The dataset included a testing dataset with patients independent to

those in the training and validation dataset [10]. The model was assessed on this testing dataset, with bootstrapping used to construct confidence intervals.

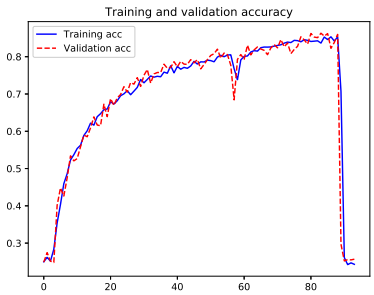
3 Results

3.1 Training and validation

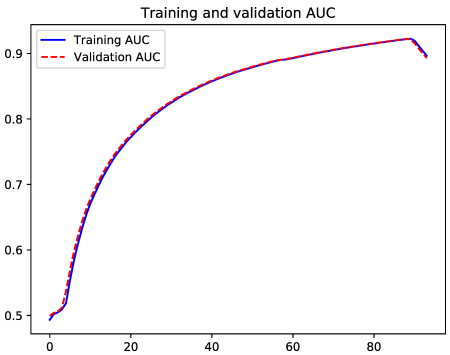
The output of the DAISY algorithm was a $16 \times 24 \times 200$ tensor which is a suitable size for computation. This took an average of 4.9 seconds per image to produce, training the deep learning model then took 29s per epoch. The model was trained until convergence (see Fig. 5), at this point the best model based on validation accuracy was chosen. In the training dataset we achieved an overall accuracy of 93.7% and AUC of 0.9358, in the validation dataset accuracy was 76.5% and AUC was 0.9359.



(a) Categorical cross-entropy loss



(b) Categorical accuracy



(c) AUC

Fig. 5. Model performance at each epoch in the testing and validation datasets, for loss 5(a), accuracy 5(b), and AUC 5(c). After around 90 epochs, model performance appears to have converged and no further epochs were required.

3.2 Testing dataset performance

Bootstrapping[8] was performed on 1,000 samples, the median values and 95% confidence intervals were 77.1% (74.3%, 79.7%) for accuracy and 0.928 (0.928, 0.929) for AUC. Bootstrapping required no further model training and was very quick to calculate confidence intervals.

3.3 Comparisons

To assess the usefulness of our method, we compared our results to results using Inception V3, which has previously been used on the full dataset[10]. To provide a fair comparison we use exactly the same conditions as before, only changing the network itself. The results are presented in Table 1.

Table 1. Our method shows improved performance over the previously used Inception V3 network. All performance measures show improved performance, with the exception of loss in the testing dataset which showed a non-statistically significant improvement. Confidence intervals were calculated using a 1000 iteration bootstrap procedure.

Method	Measure	Training	Validation	Testing (95% Confidence Interval)
Inception V3	Loss	0.901	0.830	0.729 (0.668, 0.790)
	Accuracy	62.4%	68.1%	67.8% (65.1, 70.9)
	AUC	0.705	0.709	0.912 (0.912, 0.914)
Our method	Loss	**0.433**	**0.347**	0.614 (0.559, 0.677)
	Accuracy	**83.7%**	**86.3%**	**76.6% (74.0, 79.1)**
	AUC	**0.918**	**0.918**	**0.935 (0.935, 0.936)**

4 Discussion and conclusions

This paper has briefly outlined a two-stage method which combines DAISY descriptors with a deep learning network. This has provided a more data efficient alternative to downsampling. On an example dataset of OCT images, the two-stage method achieved an overall multi-class accuracy of 76.6%, an improvement over a previously used method. With large volumes of high dimensional OCT images, the proposed method has given a quick method of classifying disease type. Acting as a type of 'intelligent downsampling', DAISY descriptors have enabled our method to maintain more of the information contained within images, while only using a three layer recurrent neural network. In future this method can provide a better alternative to random downsampling.

The main limitation of this study is that DAISY descriptors give a 3D output, there are very few pretrained deep neural networks for 3D data such as those for 2D images. Model performance is expected to greatly increase with further exploration of the classification network.

References

1. Adhi, M., Duker, J.S.: Optical coherence tomography-current and future applications (2013)
2. Beck, R.W., Edwards, A.R., Aiello, L.P., Bressler, N.M., Ferris, F., Glassman, A.R., Hartnett, M.E., Ip, M.S., Kim, J.E., Kollman, C.: Three-year follow-up of a randomized trial comparing focal/grid photocoagulation and intravitreal triamcinolone for diabetic macular edema. Archives of Ophthalmology **127**(3), 245–251 (2009)
3. Chen, Y., Wong, T.Y., Lai, T.Y., Yu, H.G., Lanzetta, P., Leveziel, N., Ohno-Matsui, K., Holz, F.G., Tufail, A.: Myopic choroidal neovascularisation: current concepts and update on clinical management. British Journal of Ophthalmology **99**(3), 289–296 (2014)
4. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation (2014)
5. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling (2014)
6. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C.O., Raine, R., Hughes, J., Sim, D.A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P.T., Suleyman, M., Cornebise, J., Keane, P.A., Ronneberger, O.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine **24**(9), 1342–1350 (2018)
7. Díaz García, J., Brunet Crosa, P., Navazo Álvaro, I., Pérez, F., Vázquez Alcocer, P.P.: Feature-preserving downsampling for medical images. EuroVis 2015: The EG/VGTC Conference on Visualization: Posters track (2015)
8. Efron, B.: Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics **7**(1), 1–26 (1979)
9. Green, W.R., Wilson, D.J.: Choroidal Neovascularization. Ophthalmology **93**(9), 1169–1176 (1986)
10. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M.K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., Shi, A., Zhang, R., Zheng, L., Hou, R., Shi, W., Fu, X., Duan, Y., Huu, V.A., Wen, C., Zhang, E.D., Zhang, C.L., Li, O., Wang, X., Singer, M.A., Sun, X., Xu, J., Tafreshi, A., Lewis, M.A., Xia, H., Zhang, K.: Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell **172**(5), 1122–1124.e9 (2018)
11. Lei, B., Thing, V.L., Chen, Y., Lim, W.Y.: Logo classification with Edge-based DAISY descriptor. In: Proceedings - 2012 IEEE International Symposium on Multimedia, ISM 2012. pp. 222–228. IEEE (2012)
12. Parodi, M.B., Evans, J.R., Virgili, G., Michelessi, M., Bacherini, D.: Laser treatment of drusen to prevent progression to advanced age-related macular degeneration. Cochrane Database of Systematic Reviews (2015)
13. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching. In: 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR. pp. 1–8. IEEE (2008)

14. Tola, E., Lepetit, V., Fua, P.: DAISY: An efficient dense descriptor applied to wide-baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(5), 815–830 (2010)