Dictionary Learning Informed Deep Neural Network with Application to OCT Images*

Joshua Bridge[1], Simon Harding[1,2], and Yalin Zheng[1,2]

[1] Department of Eye and Vision Science, University of Liverpool, L7 8TX, UK
[2] St. Pauls Eye Unit, Royal Liverpool University Hospital, UK
{joshua.bridge, s.p.harding, yalin.zheng}@liverpool.ac.uk

Abstract. Medical images are often of very high resolutions, far greater than can be directly processed in deep learning networks. These images are usually downsampled to much lower resolutions, likely losing useful clinical information in the process. Although methods have been developed to make the image appear much the same to human observers, a lot of information that is valuable to deep learning algorithms is lost. Here, we present a novel dictionary learning method of reducing the image size, utilizing DAISY descriptors and Improved Fisher kernels to derive features to represent the image in a much smaller size, similar to traditional downsampling methods. Our proposed method works as a type of intelligent downsampling, reducing the size while keeping vital information in images. We demonstrate the proposed method in a classification problem on a publicly available dataset consisting of 108,309 training and 1,000 validation grayscale optical coherence tomography images. We used an Inception V3 network to classify the resulting representations and to compare with previously obtained results. The proposed method achieved a testing accuracy and area under the receiver operating curve of 97.2% and 0.984, respectively. Results show that the proposed method does provide an accurate representation of the image and can be used as a viable alternative to conventional downsampling.

Keywords: Dictionary Learning · Deep Neural Network · DAISY descriptors · Improved Fisher Kernels · OCT

1 Introduction

Routinely collected medical images are usually high resolution, far exceeding computational capabilities, meaning that these images must be downsampled to much lower resolutions for most deep learning applications. Any downsampling will inevitably lose information [5]. There is a need for data efficient, intelligent downsampling techniques which can downsample images while keeping as much relevant information as possible.

* Supported by EPSRC Project 2110275 and Institute of Ageing and Chronic Disease, University of Liverpool

Traditional image downsampling techniques focus on making the image appear the same to a human observer (for example, bicubic interpolation [7]). However, there is often a difference between features that humans and algorithms perceive to be important. We aim to develop a new method that is data efficient and preserves the features that a deep neural network finds useful in challenging tasks.

This paper proposes a novel method largely inspired by Albarrak et al. [1], which aims to provide a better method of image downsampling. Albarrak et al. [1] proposed a method of volumetric image classification, which involved decomposing 3D volumetric images into homogeneous regions and then representing each region with a Histogram of Oriented Gradients (HOG). They then used Improved Fisher Kernels (IFK) [13] to create one feature vector for each image. Contrary to Albarrak et al. [1], our method utilizes DAISY descriptors to provide a 2D representation of images in a 3D space. IFK is then used to choose the best value in the third dimension of the DAISY feature maps, finally resulting in a 2D image. This method results in a 2D representation, which is of a significantly smaller size than the original image. The representation can then be passed through a deep learning network for classification. The method is illustrated in Figure 1. Our method is demonstrated on a publicly available dataset, consisting of Optical Coherence Tomography (OCT) images [6]. We compare our results with their published results obtained previously on this dataset and achieve improved accuracy.

Our main contribution can be summarised as follows. First, we propose a new dictionary learning approach for intelligent image downsampling based on DAISY descriptors and Fisher Vectors; second, we demonstrate that the proposed approach is compatible with deep learning algorithms; third, we show promising results in OCT images to improve disease classification.

The remainder of the paper is organized as follows. Section 2 gives a brief outline of previous work. Section 3 describes the methods used to create the image representation. In section 3, we apply our method to a dataset of OCT images and give results compared to previously obtained results. Finally, section 4 briefly discusses our findings and conclusions.

2 Previous work

There are a variety of well-established image downsampling techniques, including nearest neighbor, bilinear, and bicubic interpolation [7]. Downsampling techniques mainly focus on reducing image dimensions while still providing an accurate representation of the image. Other downsampling methods, such as adaptive downsampling [9], focus on enabling the image to be reconstructed to the original size. Previously, these methods have been evaluated based upon how similar they are to the ground truth, using performance measures such as peak signal to noise ratio and root mean square error. However, these performance measures fail to evaluate how well the method captures features that may be useful in a classification task.

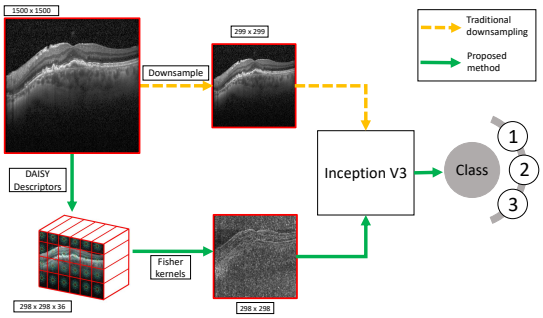

Fig. 1. The proposed framework. Previously, large images would need to be downsampled to avoid an out of memory error. The new framework aims to avoid random downsampling by introducing a more data efficient approach. The new image representation may not be more intuitive to a human observer; however, it may be more useful for a classification algorithm.

Methods such as Histogram of Oriented Gradients (HOG) [19], Scale Invariant Feature Transformation (SIFT) [10], dictionary learning [1], and deep learning [3], have been proposed as a method of reducing image dimension before classification. These methods have some success in producing excellent classification performance. A modified version of DAISY has previously been used in a logo classification problem [8], where it was used to produce edge maps. A linear support vector machine (SVM) was then used for classification. This method produced state-of-art results, which were superior to other SIFT-like methods. We differ from this method by using Fisher vectors to reduce the DAISY representation back to an image; this allows us to use already well-developed image classification algorithms.

3 Methods

The proposed new dictionary learning image representation consists of two steps. The first step utilizes DAISY descriptors to describe the image densely in three dimensions, creating a dictionary. The second step uses Fisher kernels to choose the most significant value in the third dimension of the DAISY features. This produces a much smaller image that resembles the original to some degree. While the generated image may not be insightful to human graders, its enhanced features may be more useful to a computer than the original image.

3.1 DAISY

DAISY is an efficient image descriptor [16], which works in a similar way to the more widely known algorithms, SIFT, and GLOH [15]. DAISY descriptors are named because of the flower like pattern they produce. DAISY begins by selecting a dense grid of uniformly spaced pixels, calculating orientation maps

for each of the chosen pixels and convolving them with Gaussian kernels. This produces a vector for each chosen pixel.The DAISY algorithm results in a 3D tensor, with the first two dimensions storing the horizontal and vertical location of the chosen pixels and the third dimension storing the descriptor values. A visual representation of DAISY is shown in Figure 2. The use of Gaussian kernels makes this algorithm efficient to compute, and it is well suited to providing dense representations [15].

For an image of size (I_1, I_2), and given parameters step size s, radius r, rings p, histograms h, and orientations o, the DAISY algorithm returns an array of size (D_1, D_2, D_3), where

$$D_1 = \left\lceil \frac{I_1 - r^2}{s} \right\rceil, \quad D_2 = \left\lceil \frac{I_2 - r^2}{s} \right\rceil, \quad D_3 = (p \times h + 1) \times o.$$

In our experiments, the DAISY algorithm was implemented using scikit-image 0.15.0[17], with a step size of 5, a radius of 5, one ring, 8 histograms, and 4 orientations. These parameters were chosen after some initial testing on a small subset of 5000 images, more intensive testing may provide better parameter choices. Each direction consisted of 298 pixels, resulting in a $298 \times 298 \times 36$ feature map.

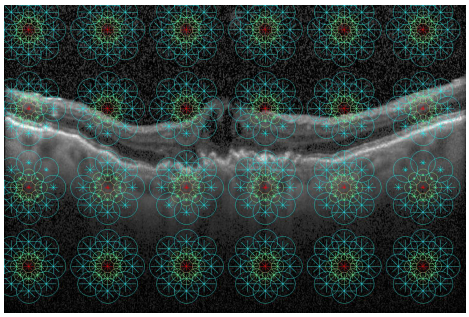

Fig. 2. Visual representation of DAISY descriptors on an example image. A sparse representation is used so that the pattern is easily observed.

3.2 Fisher vectors

The Fisher Kernel combines generative statistical models with discriminative methods, making it both generalizable and flexible [13]. It has previously been applied to a variety of problems, including image representation. Since its creation, the Fisher kernel has been improved to provide higher accuracy in real-life classification, using L2 normalization and changing the linear kernel with a non-linear additive kernel. [13]. This Improved Fisher Kernel (IFK) is used here to reduce the image dimension further. In brief, for a set of feature vectors

$F = (x_1, \ldots, x_N)$ in D dimensions, such as those extracted by DAISY, we fit a Gaussian Mixture Model (GMM) with K kernels, $\Gamma = (\mu_k, \Sigma_k, \pi_k; k = 1, \ldots, k)$. The GMM calculates posterior probabilities for each feature vector:

$$p_{nk} = \frac{\exp\left[-\frac{1}{2}(x_n - \mu_k)^T \Sigma_g^{-1}(x_n - \mu_k)\right]}{\sum_{i=1}^{K} \exp\left[-\frac{1}{2}(x_n - \mu_i)^T \Sigma_k^{-1}(x_n - \mu_i)\right]}.$$

Then for each kernel k and dimension d, the mean and covariance are calculated:

$$m_{dk} = \frac{1}{N\sqrt{\pi_k}} \sum_{n=1}^{N} p_{nk} \frac{x_{dn} - \mu_{dk}}{\sigma_{dk}},$$

$$c_{dk} = \frac{1}{N\sqrt{2\pi_k}} \sum_{n=1}^{N} p_{dk} \left[\left(\frac{x_{dn} - \mu_{dk}}{\sigma_{dk}} \right)^2 - 1 \right].$$

In order to reduce computation, we apply Fisher vectors with one kernel, giving $q_i = 1$. A Fisher Vector is calculated per image, meaning that $m_{d1} = 0, \forall d$. Hence, for our application, Fisher Vectors can be described with one vector:

$$c_{d1} = \frac{1}{N\sqrt{2\pi_1}} \sum_{n=1}^{N} \left[\left(\frac{x_{dn} - \mu_{d1}}{\sigma_{d1}} \right)^2 - 1 \right].$$

This vector is then used to reconstruct a grayscale image. This is a special case of IFK; otherwise, we can use more kernels and which would produce a longer feature vector.

4 Experimental and Results

4.1 Data

The proposed method is demonstrated on a publicly available dataset consisting of OCT images [6]. OCT is a similar concept to ultrasound; however, it uses light instead of sound to produce a cross-sectional view of tissue composition with micrometre resolution [2]. In this dataset, each image is labeled as either normal, Choroidal Neovascularisation (CNV), Diabetic Macular Edema (DME), or drusen, corresponding to the disease that they display. CNV is a leading retinal disease that can cause irreversible sight loss. There are various causes of CNV, with the main form of CNV being wet age-related macular degeneration(AMD) [4, 18]. DME, also called Diabetic Macula Oedema (DMO), is a common cause of vision loss in patients with diabetes. DME is characterized by intraretinal fluid, causing a thickening in the retinal layers [11]. Drusen are subretinal lipid deposits and are indicative of AMD [12]. Examples of each of the four classes are shown in Figure 3.

The training dataset consists of 37,205 CNV images, 11,348 DME images, 8,616 drusen images, and 51,140 healthy images. The testing dataset comprised of 250 images from each class and was collected from patients separate from the training dataset[6]. Same as used by Kermany et al. we use this as a validation dataset to evaluate performance.

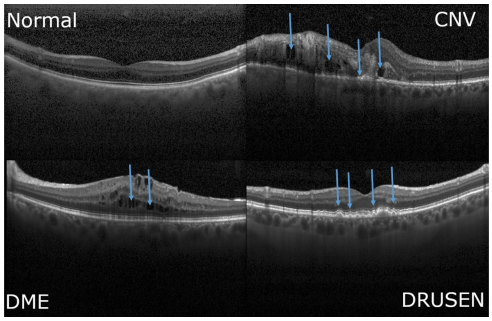

Fig. 3. Examples of the four classes in the OCT dataset. Arrows indicate the prominent features that lead to the diagnosis.

4.2 Experiments setup

All experiments were conducted on a Linux machine using Ubuntu 18.04, with a Titan X 12GB GPU and 32GB of memory. Python 3.6 and Keras 2.2.4 were used to implement the method and in the deep learning network. For comparison, we followed the work of Kermany et al. [6] who used an Inception V3 network [14], pretrained on Imagenet, with the Adam optimizer to classify the images according to the disease they displayed. Inception V3 is a popular deep learning network consisting of 159 layers. The Inception V3 is based on previous Inception networks, introducing new design principles to increase accuracy while reducing computational complexity, such as. During the training of the classification model, early stopping, with a patience of 10 epochs, and model checkpoints were used to prevent overfitting and to select the best model. Class weights were used to balance the training dataset.

4.3 Results

The output of the DAISY algorithm was a 298x298x36 tensor. A Fisher vector generalized mixed model was then applied in the third dimension, with kernel dimension 1. The final output resulted in a final size of 298x298, which is close to our target of 299x299. Each image took an average of 2.1s to process offline in this way. We used the classification method described by Kermany et al. [6], which allowed us to compare our results directly with theirs.

To assess the performance of the proposed method, accuracy, macro-average multi-class area under the receiver operating characteristic (AUC), sensitivity, and specificity were calculated and compared with previously reported results by Kermany et al. [6], shown in Table 4.3.

5 Discussion and Conclusions

This novel method aims to reduce image size while keeping important information useful to the classification. DAISY descriptors provide a dense representa-

Table 1. Multiclass classification performance metrics in the testing dataset. Values in bold indicate the best score for that metric.

Method	Accuracy	AUC	Sensitivity	Specificity
Kermany et al. [6]	96.6%	-	**97.8%**	97.4%
Our model	**97.2%**	**0.984**	97.1%	**99.1%**

tion of the image and Fisher kernels further reduce the size of that representation of that image. The results presented here achieve improved accuracy and specificity over the previous results by Kermany et al. [6], which utilized traditional downsampling methods. This demonstrates that the proposed method successfully captures useful information in images and may provide a better alternative to traditional downsampling methods. Our method achieves improved specificity at the expense of some sensitivity. Principal Component Analysis was used as an alternative to the Fisher kernels, however we were not able to obtain good results and these are not presented here. DAISY has the advantage of providing dense representations, while also being computationally inexpensive. The use of DAISY descriptors may also provide a more robust representation against both photometric and geometric transformations compared to other descriptor algorithms, as observed by Tola et al. [15].

The biggest current limitation of this method is the time taken to process the images; speed increases may be possible if DAISY descriptors can be calculated on a GPU. The effect of DAISY hyperparameters such as step size and radius are yet to be fully explored, and improved results may be possible. More work needs to be carried out to confirm if the proposed method generalizable to other imaging modalities such as color fundus, which will contain more features than OCT.

In conclusion, we have successfully demonstrated that our new method may provide a viable alternative to downsampling images before training a deep learning network. Our method has increased accuracy over previous work when tested on a publicly available dataset. Future work will concentrate on further optimizing the model in terms of speed and optimal parameters and in seeking other applications.

References

1. Albarrak, A., Coenen, F., Zheng, Y.: Volumetric image classification using homogeneous decomposition and dictionary learning: A study using retinal optical coherence tomography for detecting age-related macular degeneration. Computerized Medical Imaging and Graphics **55**, 113–123 (2017)
2. Bezerra, H.G., Costa, M.A., Guagliumi, G., Rollins, A.M., Simon, D.I.: Intracoronary optical coherence tomography: a comprehensive review clinical and research applications. JACC. Cardiovascular interventions **2**(11), 1035–1046 (2009)
3. Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P.E., Verrill, C., Walliander, M., Lundin, M., Haglund, C., Lundin, J.: Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports **8**(1), 3395 (2018)

4. Cohen, S.Y., Laroche, A., Leguen, Y., Soubrane, G., Coscas, G.J.: Etiology of choroidal neovascularization in young patients. Ophthalmology **103**(8), 1241–1244 (1996)
5. Dodge, S.F., Karam, L.J.: Understanding how image quality affects deep neural networks. CoRR **abs/1604.04004** (2016), http://arxiv.org/abs/1604.04004
6. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M.K., Pei, J., Ting, M.Y.L., 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.N., 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), 1122–1131.e9 (2018)
7. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing **29**(6), 1153–1160 (1981)
8. Lei, B., Thing, V.L.L., Chen, Y., Lim, W.: Logo classification with edge-based daisy descriptor. In: 2012 IEEE International Symposium on Multimedia. pp. 222–228 (2012)
9. Lin, W., Li Dong: Adaptive downsampling to improve image compression at low bit rates. IEEE Transactions on Image Processing **15**(9), 2513–2521 (2006). https://doi.org/10.1109/TIP.2006.877415
10. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. vol. 2, pp. 1150–1157 vol.2 (1999). https://doi.org/10.1109/ICCV.1999.790410
11. Otani, T., Kishi, S., Maruyama, Y.: Patterns of diabetic macular edema with optical coherence tomography. American Journal of Ophthalmology **127**(6), 688–693 (1999)
12. Pedersen, H.R., Gilson, S.J., Dubra, A., Munch, I.C., Larsen, M., Baraas, R.C.: Multimodal imaging of small hard retinal drusen in young healthy adults. British Journal of Ophthalmology **102**(1), 146–152 (2018)
13. Perronnin, F., Snchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision ECCV 2010. pp. 143–156. Springer Berlin Heidelberg
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. CoRR **abs/1512.00567** (2015), http://arxiv.org/abs/1512.00567
15. 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)
16. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching (2008 2008)
17. van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., the scikit-image contributors: scikit-image: image processing in Python. PeerJ **2**, e453 (6 2014). https://doi.org/10.7717/peerj.453, https://doi.org/10.7717/peerj.453
18. Wong, T.Y., Ohno-Matsui, K., Leveziel, N., Holz, F.G., Lai, T.Y., Yu, H.G., Lanzetta, P., Chen, Y., Tufail, A.: Myopic choroidal neovascularisation: current concepts and update on clinical management. British Journal of Ophthalmology **99**, 289–296 (2015)
19. Zhao, Y., Zhang, Y., Cheng, R., Wei, D., Li, G.: An enhanced histogram of oriented gradients for pedestrian detection. IEEE Intelligent Transportation Systems Magazine **7**(3), 29–38 (2015)