Training Deep Convolutional Neural Networks with Active Learning for Exudate Classification in Eye Fundus Images

Sebastian Otálora[2], Oscar Perdomo[1], Fabio González[1], and Henning Müller[2]

[1]Universidad Nacional de Colombia, Bogotá, Colombia,
[2]University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

Abstract. Training deep convolutional neural network for classification in medical tasks is often difficult due to the lack of annotated data samples. Deep convolutional networks (CNN) has been successfully used as an automatic detection tool to support the grading of diabetic retinopathy and macular edema. Nevertheless, the manual annotation of exudates in eye fundus images used to classify the grade of the DR is very time consuming and repetitive for clinical personnel. Active learning algorithms seek to reduce the labeling effort in training machine learning models. This work presents a label-efficient CNN model using the expected gradient length, an active learning algorithm to select the most informative patches and images, converging earlier and to a better local optimum than the usual SGD (Stochastic Gradient Descent) strategy. Our method also generates useful masks for prediction and segments regions of interest.

1 Introduction

Diabetes Mellitus is one of the leading causes of death according to statistics of the World Health Organization.[1] Diabetic Retinopathy (DR) is a condition caused by prolonged diabetes, causing blindness in persons at a relatively young age (20-69 years). The problem is that most persons have no symptoms and suffer the disease without a timely diagnosis. Because the retina is vulnerable to microvascular changes of diabetes and because diabetic retinopathy is the most common complication of diabetes, eye fundus imaging is considered a non-invasive and painless route to screen and monitor DR [6, 12].

In the earliest stage of DR, small areas of inflammation called *exudates* appear in the retinal blood vessels, the detection of these yellowish areas that grow along the retina surface is an important step for the ophthalmologist to grade the stage of DR. The manual segmentation of exudates in eye fundus images, is very time consuming and repetitive for clinical personnel [6].

In recent years, deep learning techniques have increased the performance of computer vision systems, deep convolutional neural networks (CNN) were used

[1] http://www.who.int/diabetes/en/

to classify natural images and recognize digits and are now being used successfully in biomedical imaging and computer-aided diagnosis (CADx) systems [3].

CNN models play a major role in DR grading showing superior performance in several settings and datasets compared to previous approaches. In 2015, the data science competition platform Kaggle launched a DR Detection competition[2]. The winner and the top participants used CNNs on more than 35,000 labeled images, demonstrating that for a successful training of such algorithms a significant amount of labeled data is required. In [4] the authors used more than 100,000 labeled eye fundus images to train a CNN with a performance comparable to an ophthalmologist panel. This presents a challenge, as the algorithms need to be fed with in the order of thousand of samples, which in practice is both time-consuming and expensive. It is important to make well-performing algorithms such as CNN less data intensive and thus able to learn with a few selected examples. This is more realistic in clinical practice, also because imaging devices change over time.

Active learning is an important area of machine learning research [10] where the premise is that a machine learning algorithm can achieve good accuracy with fewer training labels if the algorithm chooses the data from which it learns intelligently. This idea is key for building more efficient CADx systems and for reducing costs in building medical image datasets [13] where the expert annotations are costly and time-consuming.

In [14], an active learning algorithm for convolutional deep belief networks is presented with an application to sentiment classification of documents. In [2], the authors show how to formally measure the expected change of model outputs for Gaussian process regression showing an improvement in the area under the ROC curve with fewer queries to the model than the usual random selection. Active learning has also been applied to reduce the number of labeled samples in training CAD systems for DR. Sánchez et al. [9] compare two active leaning approaches, uncertainty sampling and query-by-bagging, showing that with the former, just a reduced number of labeled samples is necessary for the system to achieve a performance of 0.8 in area under the receiving operating characteristic curve. Nevertheless, this approach is computationally intensive for deep CNNs because it is based on building multiple committees of classifiers to choose the most informative sample, which translates into training multiple deep CNNs.

In this work we present a novel approach to detect exudates and highlight the most interesting areas of the eye fundus images using an active learning algorithm called *expected gradient length* (EGL) that works jointly with the CNN model parameters to select the most informative patches and images to train without a significant compromise in the model performance. Our method has the advantage of computing a single backward-forward pass in order to obtain the samples that lead to the most changes in the network parameters, i.e. the most informative images and patches to learn. To the best of our knowledge, this is the first time that an active learning method that uses the deep learning model parameters to select the most relevant samples is presented in the medical imaging field.

[2] `https://www.kaggle.com/c/diabetic-retinopathy-detection`

2 Deep Learning Model

Convolutional Neural Networks (CNN) are a particular kind of a supervised multi layer perceptrons inspired by the visual cortex. The CNNs are able to detect visual patterns with minimal preprocessing. They are trained with the robustness to respond to the distortion, variability and invariance to the exact position of the pattern and benefit from data augmentation that uses subtle transforms of the input for learning invariances. CNN models are one of the most successful deep learning models for computer vision. The medical imaging field is rapidly adapting these models to solve and improve a plethora of applications [3].

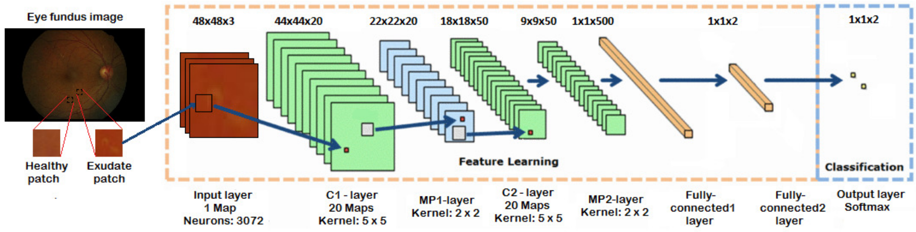

Fig. 1. Deep CNN architecture to classify between healthy and exudate patches.

Our deep learning model is based on a CNN architecture called LeNet [7] with 7 layers as shown in Figure 1, which is composed of a patch input layer followed by two convolutions and max pooling operations to finalize in a softmax classification layer that outputs the probability of a patch being healthy or exudate. We choose this architecture because of its good classification performance with small input and because (as seen in section 3) our model selects the samples by performing a forward-backward pass over the net. A deeper network would put a computational burden on our experiments.

3 EGL for Patch And Image Selection in Convolutional Neural Networks

Traditional supervised learning algorithms use whatever labeled data is provided to induce a model. Active learning, in contrast, gives the learner a degree of control by allowing to select which instances are labeled and added to the training set. A typical active learner begins with a small labeled set $\mathcal{L}$, selects one or more informative instances from a large unlabeled pool $\mathcal{U}$, learns from these labeled queries (that are added to $\mathcal{L}$), and repeats until convergence. The principle behind active learning is that a machine learning algorithm can achieve similar or even better accuracy when trained with few training labels than with the full training set if the algorithm is allowed to choose the data from which it learns [10].

An active learner may pose queries, usually in the form of unlabeled data instances to be labeled by an oracle (e.g. an ophthalmologist annotator). Active learning is well-motivated in many modern machine learning problems, where unlabeled data may be abundant or easily obtained but labels are not. This is an interesting direction for the so-called *deep learning in the small data regime*, where the objective is to train time-consuming and high sample complexity algorithms with fewer resources, as in the case of medical images.

Stochastic Gradient Descent (SGD) works by stochastically optimizing an objective function J with respect to the model parameters θ. This means to find the model parameters by optimizing with only one sample or sample batches instead of the full training dataset:

$$\theta_{t+1} = \theta_t - \eta \nabla J_i(\theta_t)$$

Where $J_i(\theta_t)$ is the objective function evaluated at the i-th sample tuple (x^i, y^i) at iteration t, η is the learning rate and ∇ is the gradient operator. For computing $\nabla J_i(\theta)$ we need the i-th sample representation and its corresponding label, if we measure the norm of this term, i.e. the gradient length term$\|\nabla J_i(\theta)\|$, this quantifies how much the i-th sample and its label contribute to each component of the gradient vector.

A natural choice for selecting the most informative patches for each batch iteration of SGD is to select the instances that give the highest values for the gradient length weighted by the probability of this sample having the y^i label. In other words, to select the instances that create the largest change to the current model if we knew their labels:

$$\Phi(x^i) = \sum_{j=1}^{c} p(y^i = j|x^i)\|\nabla J_i(\theta)\| \tag{1}$$

Where c is the total number of labels or classes. The Expected Gradient Length (EGL) works by sorting the Φ values from an unlabeled pool of samples and then adding them to the training dataset by asking an oracle to give the ground truth label of these samples. The EGL algorithm was first mentioned by Settles et. al. [11] in the setting of multiple-instance active learning. To the best of our knowledge this is the first time the approach is used in the selection of samples in CNN. For being able to select the most informative samples in a CNN architecture we have to compute the two terms involved in equation (1). For the probability of a sample having the j-th label we can perform a forward propagation through the network and obtain the corresponding probabilities from the softmax layer of the network. To measure the gradient length we can perform a backward propagation through the network to measure the frobenius norm of the gradient parameters. In a CNN architecture we have the flexibility to compute the backward/forward phases up to a certain layer. In our experiments we made the backward down to the first fully connected layer as experiments

Algorithm 1 EGL for Active Selection of patches in a CNN

```
Require: Patch Dataset $\mathcal{L}$, Initial Trained Model $\mathbf{M}$ with patches in $\mathcal{L}' \subset \mathcal{L}$, Number
    $k$ of most informative patches
while not converged do
    Create and shuffle batches from $\mathcal{L}$
    for each batch do
        Compute $\Phi(x)$ using $\mathbf{M}, \forall x \in$ batch
    end for
    Sort all the $\Phi$ values and return the highest $k$ corresponding samples $\mathcal{L}_k$
    Update $\mathbf{M}$ using $\mathcal{L}' \cup \mathcal{L}_k$
end while
```

showed no significant differences for in between layers. This process has to be done over all possible labels for each sample. Once we have computed the Φ values for all samples, we sort them and select the k samples with the highest EGL values.

We begin with a small portion of labeled samples $\mathcal{L}' \subset \mathcal{L}$ to train an initial model $\mathbf{M}$, and then incrementally adding the k samples to $\mathcal{L}'$ to update $\mathbf{M}$ parameters. We stop the training procedure when the algorithm converges i.e. when the training and validation errors do not decrease significantly or when the performance in terms of accuracy stays the same for more than one epoch. Since we are able to compute the most significant patches it is straightforward to extend the procedure to select not only the most informative patches but also the most informative images within the training set. The modification is that instead of computing the EGL values for all ground truth exudate and healthy patches we compute the *interestingness* of an image by *patchifying* the image with a given stride and then densely computing Φ. Then, images are sorted by their top EGL values and finally, the patches that belong to the most interesting image are added to the training set for further parameter updates using Algorithm 1 until convergence. We think that this is a more realistic scenario where an ophthalmologist does not have the time to manually annotate all images but only those that contain most information to train a label efficient system. The full algorithm is described in Algorithm 2

4 Experimental Setup

4.1 Ophtha Dataset

The e-ophtha database with color fundus images was used in this work. The database contains 315 images with a size ranging from 1440×960 to 2540×1690 pixels, 268 images have no lesion and 47 contain exudates that were segmented by ophthalmologists from the OPHDIAT Tele-medical network under a French Research Agency (ANR) project [1]. The labeled patch dataset was created with cropped 48×48 pixel patches that contain both exudate and healthy examples. We prevent over–fitting artificially creating new samples by generating artificially 7 new label-preserving samples using a combination of flipping and $90, 180$

Algorithm 2 EGL for Active Selection of images in a Convolutional Neural Network.

```
Require: Training Image Set $\mathcal{T}$, Patch Dataset $\mathcal{L}$, Number $\mu$ of initial images to look at
    Select an initial set $\mathcal{T}_\mu$ of images randomly
    Train initial model $\mathbf{M}$ using the ground truth patches from the $\mu$ images
    while not converged do
        for each image in $\mathcal{T} \setminus \mathcal{T}_\mu$ do
            Patchify image and compute $\sigma_{image} = \sum_{patch \in image} \Phi(patch)$, using $\mathbf{M}$
        end for
        Sort all the $\sigma_{image}$ values and return $\mathcal{I}_{max}$, the image with higher sum
        $\mathcal{T}_\mu = \mathcal{T}_\mu \cup \mathcal{I}_{max}$
        $\mathcal{L}_\mu = \{$ patch $\in \mathcal{L}_\mathcal{I}, \forall \mathcal{I} \in \mathcal{T}_\mu\}$
        Update $\mathbf{M}$ with $k$ selected patches using Algorithm 1 and the patches in $\mathcal{L}_\mu$
    end while
```

and 270 degree rotations. After the preprocessing steps of cropping and data augmentation, the dataset splits were built with randomly selected patches of each class as follows: a training split with 8760 patches for each class, a validation split with 328 per class and a test split with 986. Images of a given patient could only belong to a single group according to the described dataset distribution. At test time, only patches of unseen patient images are evaluated.

4.2 Evaluation

The technique of Decencieriere et al. [1] was chosen as our baseline. The base LeNet model was trained using stochastic gradient descent (SGD) from scratch without any transfer learning from other datasets. The learning rate and batch size were explored in a grid search and showed robustness in the range of 32-64 in terms of batch size with a learning rate of 0.01 when trained with all the training patches. In our final experiments we set the batch size to 32 and 0.01 for the learning rate, using 30 as the number of epochs to train the model. The model **M** is the LeNet CNN model described in Figure 1 and initially trained with 5 batches of 32 samples.

The proposed approach was implemented in Python 2.7 and the Caffe deep learning framework [5] that allows for efficient access to parameters and data in memory. We use an NVIDIA GTX TITAN X GPU for our experiments. During all the experiments, training loss, validation loss, as well as the accuracy over the validation set were monitored.

5 Results

We test our algorithm 2 in the scenario where an ophthalmologist selects only a few important or relevant images instead of patches to annotate and train

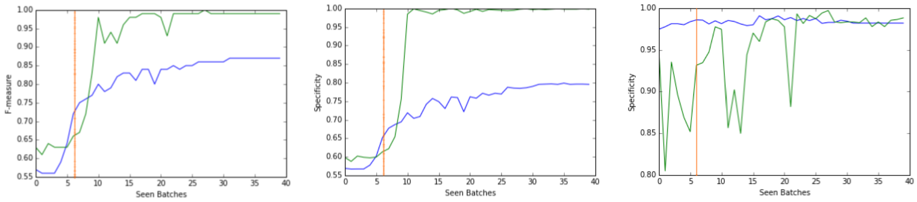

Fig. 2. Results for F-Measure, sensitivity and specificity, using the random strategy (blue) and active learning using EGL (green) for algorithm 2. In this setup only the patches of the 4 initial training images were used for training the model in the first 6 SGD iterations, after this (orange line) we add the patches from the images with maximum EGL value to the training set.

the model. In Figure 2 the left side of the orange line is when the initial model training is performed. Then, the Algorithm 2 is used to select the most interesting image for the model and subsequently to update the model. In our approach, the convergence is reached at an earlier stage. As few as 15 batches are enough for the model convergence, showing that in this more realistic scenario our strategy also outperforms the standard way of training deep CNN models.

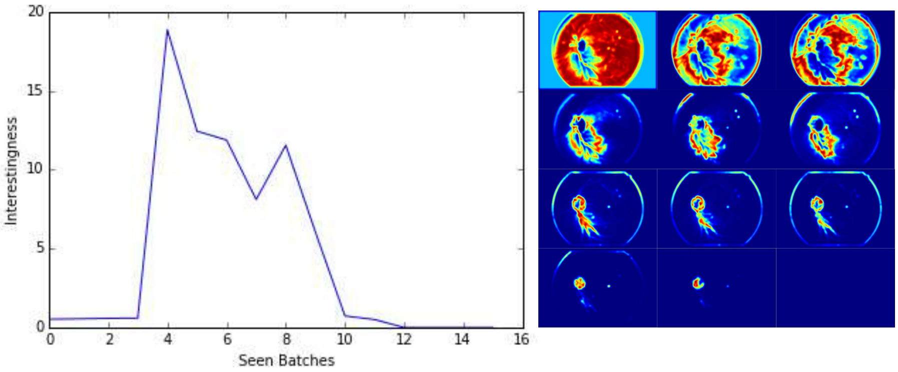

Fig. 3. Interestingness over training time. After the model converges the interestingness value decays to 0 because the norm of the gradient is close to 0.

Once we have an initial training of the model we can measure the interestingness of a full image computing the sum of its EGL values. This was the criterion to select images for the results of algorithm 2. An example image with its interestingness values over different training times is shown in Figure 2. We can plot this value and see how this evolves as the model sees more batches. These values are illustrated in Figure 3. Here we can see how the interestingness value decays after the model has converged, so when the loss function does not decrease anymore and the norm of the parameters is nearly 0.

6 Discussion

This paper presents for the first time an active learning strategy to select the most relevant samples and images for a sample efficient training of a deep convolutional neural network to classify exudate patterns in eye fundus images. The proposed strategy was able to achieve a similar performance compared to the model trained with the full dataset [8] but only using an informative portion of the training data. Besides the speed-up for convergence, our algorithm also brings an additional interpretation layer for deep CNN models that locates the regions of the image that the ophthalmologist should label, improving the interaction between the model and the specialist that conventional CNN models lack. Our approach presents a computational drawback when the number of unlabeled data–samples to check is large, but we think that this could be overcome with traditional sampling techniques. Despite our results showing good performance using only a portion of the data, we would like to do further experimentation using only the initially labeled portion and involving large–scale datasets where the combination of our sample selection techniques with transfer learning could lead to a performance boost. We think that active learning techniques have a promising application landscape in the challenging tasks of medical imaging using deep learning because of their potential to relief the need for large amounts of labeled data. This will allow the usage of deep learning models in a broader set of medical imaging tasks like detection and segmentation of structures in specialized domains such as histopathology image analysis or computed tomography scans where the labels are costly.

Acknowledgments This work was supported by the Administrative Department of Science, Technology and Innovation of Colombia (Colciencias) through the grant Jóvenes Investigadores 2014 in call 645 and by Nvidia with a TitanX GPU.

References

[1] Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.C., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M., Danno, R., et al.: Teleophta: Machine learning and image processing methods for teleophthalmology. IRBM 34(2), 196–203 (2013)

[2] Freytag, A., Rodner, E., Denzler, J.: Selecting influential examples: Active learning with expected model output changes. In: European Conference on Computer Vision. pp. 562–577. Springer (2014)

[3] 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)

[4] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22), 2402–2410 (2016)

[5] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. pp. 675–678. ACM (2014)

[6] Kauppi, T., et al.: Eye fundus image analysis for automatic detection of diabetic retinopathy. Lappeenranta University of Technology (2010)

[7] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)

[8] Oscar Perdomo, Sebastian Otalora, F.R.J.A.F.A.G.: A novel machine learning model based on exudate localization to detect diabetic macular edema. In: Ophthalmic Medical Image Analysis Third International Workshop (OMIA 2016). pp. 137–144. University of Iowa (2016)

[9] Sánchez, C.I., Niemeijer, M., Abràmoff, M.D., van Ginneken, B.: Active learning for an efficient training strategy of computer-aided diagnosis systems: application to diabetic retinopathy screening. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 603–610. Springer (2010)

[10] Settles, B.: Active learning literature survey. University of Wisconsin, Madison 52(55-66), 11 (2010)

[11] Settles, B., Craven, M., Ray, S.: Multiple-instance active learning. In: Advances in neural information processing systems. pp. 1289–1296 (2008)

[12] Stitt, A.W., Lois, N., Medina, R.J., Adamson, P., Curtis, T.M.: Advances in our understanding of diabetic retinopathy. Clinical science 125(1), 1–17 (2013)

[13] Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., Xiao, J.: Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv:1506.03365 (2015)

[14] Zhou, S., Chen, Q., Wang, X.: Active semi-supervised learning method with hybrid deep belief networks. PloS one 9(9), e107122 (2014)