DEEP CONVOLUTIONAL NETWORKS FOR OCT IMAGE CLASSIFICATION

Branislav Hesko
Doctoral Degree Programme (4), FEEC BUT
E-mail: hesko@vutbr.cz

Supervised by: Vratislav Harabiš
E-mail: harabis@feec.vutbr.cz

Abstract: In this work, OCT (optical coherence tomography) images are classified according to the present pathology into four distinct categories. Three different neural network models are used to classify images, each model is recent and we are achieving exceptional results on the testing dataset, which was unknown to the network during the training. Accuracy on the testing set is higher than 98% and only a few of images are classified into the wrong category. This makes our approach perspective for future automatic use. To further improve results, all three models are using transfer learning.

Keywords: OCT, deep learning, classification, retina

1 INTRODUCTION

OCT is a noninvasive, unharmful method for retina imaging. A low-coherence light source emits photons, which are directed towards the fibre-optic coupler. The role of the coupler is to split the incident beam evenly into the retina and towards the reference path. Light is focused on the retina and is reflected from different retina underlying structures. Both reflected and reference light are focused on the detector, while both interfering. This principle represents a simple Michelson interferometer. This way, a single A-scan is created. Applying lateral scanning results into the B-scan acquisition, which contain both lateral and depth information. Various pathological retinal changes may be detected, resulting in two leading blindness causes diagnostics: age-related macular degeneration (AMD) and diabetic macular oedema. [5, 6]

Each year, approximately 30 million OCT scans are performed worldwide, giving rise to an automatic scan processing. One of the leading domains in automatic medical image processing is image classification, especially pathological changes detection in OCT scans [6]. Various methods are nowadays available, but deep convolutional networks are the best performing subclass of algorithms applied to image classification of 1000 category task with millions of images [1]. Architectures are evolving constantly adding more and more convolutional layers. In the beginning, AlexNet being one of the first architectures with only 8 layers beat until then used methods by almost 10% [1]. From then, layers have grown from 19 for the VGG (very deep network) model [2], to recent DenseNet with 169 layers [3]. This evolution allows to use these models with medical data, where is a strong emphasis on the accuracy.

In this work, OCT images are classified into 4 different classes representing a pathology, with one class designed for healthy scans, using deep convolutional networks. Overall, 3 different networks are used for this classificatation problem: AlexNet [1], VGG-19 [2] and DenseNet [3]. Results suggest excellent performance of each network, having more than 98% accuracy.

2 DATASET

A large OCT image dataset is used [4] in this work. The goal is to classify images into four distinct classes: normal (NORMAL), choroidal neovascularization (CNV), diabetic macular oedema (DME) or drusen (DRUSEN) present in early age macular degradation. Training set consists of 83488 images, validation set is small, it contains only 32 images and finally testing set is composed of 968 images. While training set classes are not equally distributed, each class in the test set has 242 images. An example of an image of each class is included as Figure 1. Finally, images are having different sizes, which is solved by resizing each image to 224×224 pixels.

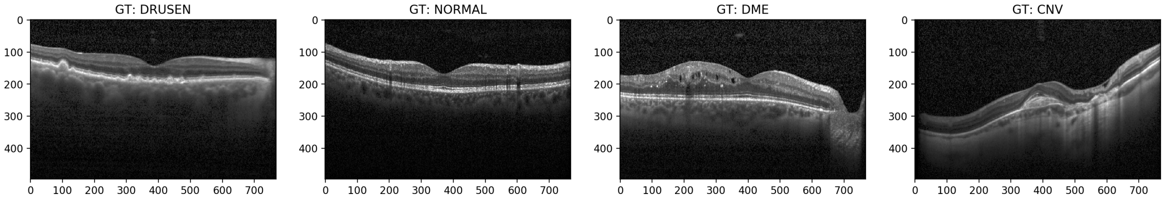

Figure 1: Example of an image of each class. [4]

3 METHODS

3.1 ALEXNET

As previously mentioned, AlexNet was the first model with high training capacity, improving result on ImageNet subset by a large margin. It introduced several new concepts improving training speeds and performance. First, instead of the previously used hyperbolic tangent which added nonlinearity into the network, ReLU activation started to be used. Next, normalization preventing from saturating the network is applied. Finally, overlapping pooling, where a single element enters the pooling multiple times (kernels are overlapping) is employed to further improve results. By summarizing, AlexNet is composed of 8 convolutional layers and two linear, fully connected layers. The main architecture was originally implemented with two parallely used graphics cards. Architecture is shown on Figure 2.

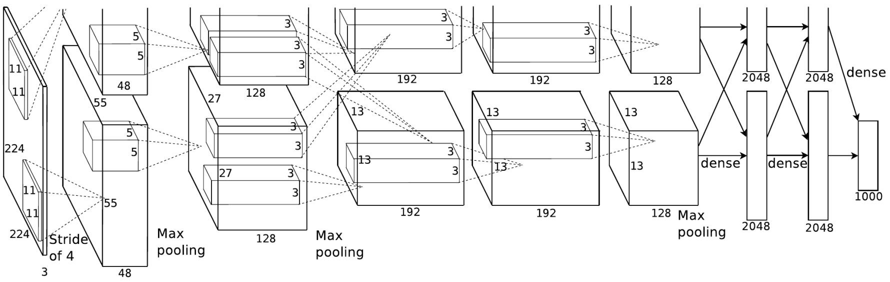

Figure 2: AlexNet architecture. Original implementation using two graphics cards may be visible. [1]

3.2 VGG-19

Very deep convolutional networks are a natural evolution of the AlexNet. It added more layers, but also adapted features from the AlexNet, both ReLU activation and also normalization. They provided a bunch of models, with VGG-19 that we pick for the OCT classification task and it contains 16 convolutional layers and 3 fully connected layers [2, 7]. The architecture is depicted on Figure 3:

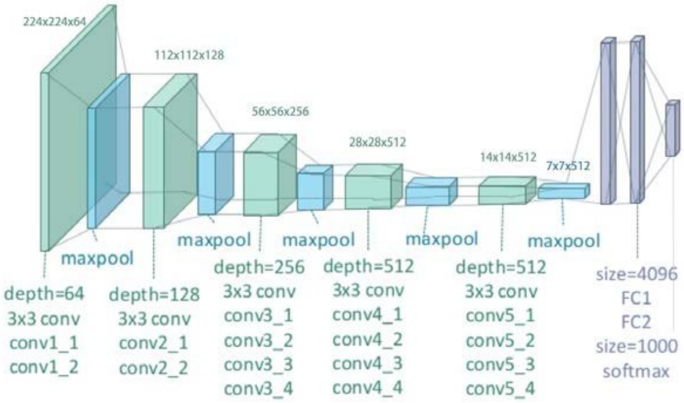

Figure 3: VGG-19 architecture, for a 1000 class segmentation. [7]

3.3 DENSENET

When augmenting the number of convolutional layers, two main problems arise, both are connected to gradient backpropagation. First, each convolutional layers decreases the gradient value, therefore vanishing gradient is a problem for very deep networks. On the other side, error gradients may accumulate and result in very large local gradients causing the network is unstable. This problem was solved by using ReLU activation and shortcut introduced in the ResNet. DenseNet introduced feature reuse by concatenating output from different dense blocks. The DenseNet is composed of several DenseBlocks in a series. By using a shortcut (feature reuse), DenseNet may have more than one hundred of layers. In this work, we employ DenseNet with 169 layers, which is shown in Figure 4 [3, 8].

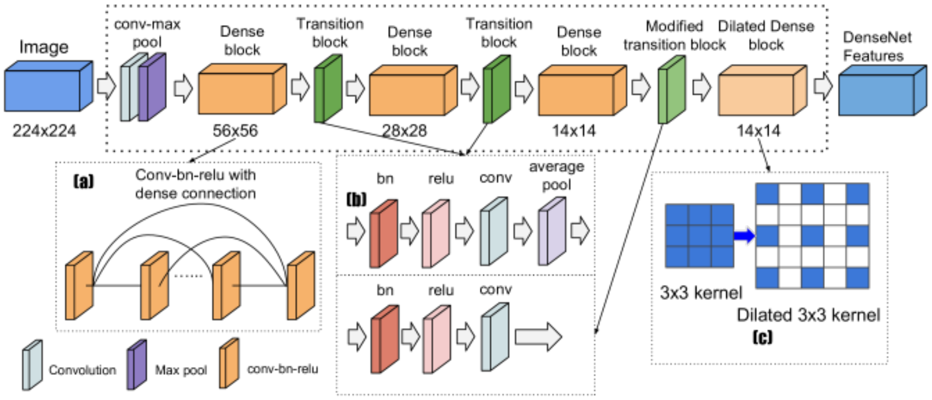

Figure 4: DenseNet architecture. There are 169 layers in total, while maintaining reasonable training speeds and excellent performance. [8]

3.4 TRANSFER LEARNING

The biggest performance influencing setting is transfer learning. Instead of learning our networks from scratch, we employ already pretrained weights from the ImageNet dataset classification, with only output linear classifier being changed to classify images into 4 classes instead of originally set 1000.

3.5 NETWORK SETTINGS

For the training, we are using the following settings:

1. **Number of epochs** - 10 for each network.
2. **Loss function** - Cross-entropy.
3. **Optimizer** - Stochastic gradient descent with momentum set to 0.9.
4. **Learning rate** - Adaptive learning rate.
5. **Batch size** - 16 images are passed in a single batch.
6. **Data augmentation** - For the training, only randomly resized crop and random vertical flip are used.

4 RESULTS

Each network is trained for 10 epochs with a validation step after each epoch. Afterwards, the output is predicted on the testing set. Each network is initialized with pretrained weights. Results are inserted into Table 1.

Table 1: Achieved results for all of our three classifier networks.

10 epochs	Alexnet		VGG-19		DenseNet	
	Loss / image	Accuracy	Loss / image	Accuracy	Loss / image	Accuracy
Training set	0.020	0.890	0.012	0.927	0.012	0.928
Validation set	0.40	1.00	0.19	1.00	0.80	0.97
Test set	0.014	0.985	0.017	0.987	0.006	0.990

Some successfully classified images are attached as Figure 5. These images were randomly selected from the testing dataset.

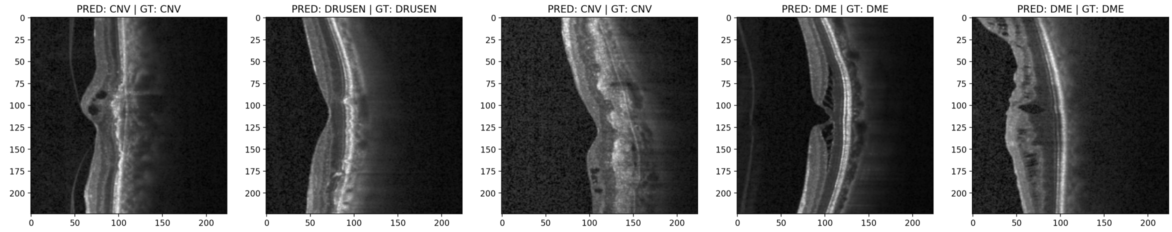

Figure 5: Successfully classified images. Each image has in its title ground truth and its predicted classification. Prediction network: DenseNet.

The test set was previously unknown to the network. Even with our outstanding performance, some images are not classified correctly. We also attach some of the misclassified images 6:

Only few images are classified into the wrong category. This may be probably caused by uneven distribution of training images. In this work, each category contains at least 10000 images and we are not modifying our models with category weighting. Inferring a single resized image lasts about 0.5 second. Otherwise networks were able to generalise pathology properties in images very well, while being faster than human and sufficiently accurate.

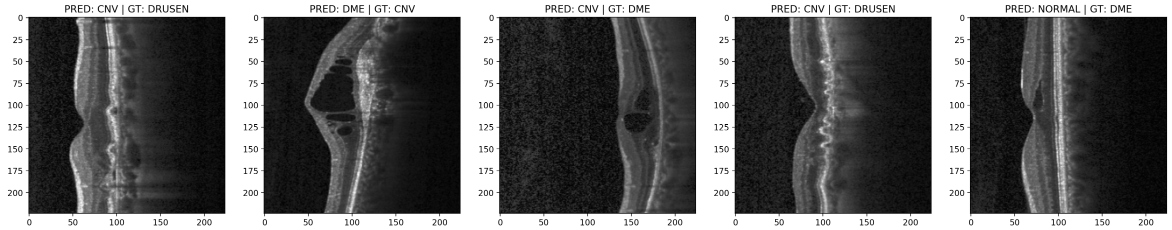

Figure 6: Misclassified images. Each image has in its title ground truth and its predicted classification. Prediction network: AlexNet.

5 CONCLUSION

We have applied the most recent classification models to OCT image dataset. All of these models are fully convolutional except the classifier part, which is used to classify features detected by the convolutional part of the network. Performance of these models is related to the year when published, the most modern method - DenseNet is achieving the best results while having less computation time than VGG-19 and fewer parameters. Evaluation on the testing dataset differs only slightly between these networks, with results over 98.5%, which means that more than 963 from 976 images are correctly classified. Results suggest possibility for our methods to be used for automatic pathology detection without needing human assistance.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

REFERENCES

[1] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: *ImageNet Classification with Deep Convolutional Neural Networks*. Advances in neural information processing systems. 2012.

[2] Simonyan, K. and Zisserman, A. *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556 (2014).

[3] Huang, G. et al. *Densely connected convolutional networks*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4700-4708.

[4] Kermany, D. S., et al. *Identifying medical diagnoses and treatable diseases by image-based deep learning*. Cell, 172(5), 1122-1131.

[5] Popescu, D. P., et al. *Optical coherence tomography: fundamental principles, instrumental designs and biomedical applications*. Biophysical reviews, 3(3), 155. 2011.

[6] Fercher, A. F., Drexler, W., Hitzenberger, C. K., and Lasser, T. *Optical coherence tomography-principles and applications*. Reports on progress in physics, 66(2), 239. 2003

[7] Chansung, P. *Transfer Learning in Tensorflow (VGG19 on CIFAR-10): Part 1*. online [03 / 2019]: `https://towardsdatascience.com/transfer-learning-in-tensorflow-9e4f7eae3bb4`.

[8] Zhou, B. *A Weakly Supervised Adaptive DenseNet for Classifying Thoracic Diseases and Identifying Abnormalities*. ArXiv. 2018.