From a simple neural network to transfer learning for Caltech-UCSD Bird-200-2011

Pierre-Louis Guhur
ENS Paris-Saclay
Master MVA
pguhur@ens-paris-saclay.fr

Abstract

Learning to classify images from a labelled dataset is a well-known research problem. In twenty years, convolutional neural networks have completely overthrown the field, in particular with the emergence of deep learning architectures. Nonetheless, many questions remain still opened, for example about depth or optimizer. Structures evolved into complexity. This report explains I modified my first simple neural network to a complex architecture based on transfer learning.

1. Introduction

Convolutional neural networks are becoming deeper and deeper with AlexNet [5], the first emerging deep neural network to Highway Networks [1] and Residual Networks [2] achieving more than 100 layers. An important challenge on the proposed dataset is its size with at most 60 images per classes.

2. Going deeper in the convolutional neural networks

2.1. Increasing diversity with data augmentation

I started from the proposed convolutional neural network and I tried to increase the dataset size with data augmentation methods by randomly flipping, cropping but also changing brightness, contrast, saturation and hue. I also tuned the image resolution to 300x300, which is a trade-off between the information available in an image and the computational power available. I refer to this architecture as *AugmentedNet*.

2.2 Increasing depth with AlexNet

Because results were unsatisfactory, I tested well-known structures. AlexNet is a easy to implement, but to accelerate the training, diminishing the batch size and decreasing learning rate are necessary. Hence, I used a batch size of 64 images and an exponential decreasing learning rate from 0.75 to 0.25 in 35 epochs.

Having deeper networks means that the information might vanish before reaching the end of the network. To address this issue, DenseNet [4] offers new blocks whose layers are fully connected in a feed-forward fashion. DenseNet provided much better results, but to prevent over-fitting, some neurons are dropped out during the training phase with a probability of 0.2. As a consequence, the average loss is more fluctuant, and more epochs (300) are required for training.

3. Training the dataset with a much larger dataset

To go further, I had to address the dataset size issue. Because data augmentation was not efficient enough, I chose to pre-train the network on a much larger dataset, ImageNet, with its 1.2 million labelled images. Because Kaggle's GPU does not allow to perform such large training, I used pre-trained models available online.

I used first ResNet-18 [2] with provided satisfying results, but best results were obtained with Inception v4 [6].

Transfer learning can be reached from initializing the weights with the pre-trained values or fixing the weights and learning additional layers. I used each time the method providing best results.

4. Results

To compare results, I include the score given by Kaggle on the test set and the loss function obtained with the cross entropy.

Neural Network	Loss function	Kaggle's score
AugmentedNet	0.918	0.13548
AlexNet	0.459	0.22580
DenseNet	0.120	0.47741
TL-ResNet-18	0.053	0.70967
TL-Inception-v4	0.014	0.80000

Table 1: comparison of neural network's performances

5. Conclusion

This report shows acceptable performances of an image classifier among 20 classes with a dataset of 1082 images. However, improvements can be brought from several manners: enhancing diversity as in Polynet [7], or training the networks to learn the locations of bird and other binary attributes using labelled data from the original dataset and a region-convolutional neural network [8].

6. References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[2] R. K. Srivastava, K. Greff, and J. Schmidhuber. Training very deep networks. In NIPS, 2015.

[3] Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013, February). On the importance of initialization and momentum in deep learning. In International conference on machine learning (pp. 1139-1147).

[4] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In CVPR (Vol. 1, No. 2, p. 3).

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105)

[6] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI (Vol. 4, p. 12) Zhang, X., Li, Z., Loy, C. C., & Lin, D. (2017, July).

[7] Polynet: A pursuit of structural diversity in very deep networks. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on (pp. 3900-3908). IEEE. Ren, S., He, K., Girshick, R., & Sun, J. (2015).

[8] Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).