Quick comparison of state-of-the-art Architectures for Face Classification

Ivan Gruber[1]

1 Introduction

Face recognition (FR) has been probably the most intensively studied topics in biometrics and computer vision for the last few years. It receives huge attention because compared to other biometric techniques (for example fingerprints), FR has the potential to non-intrusively recognize subject without any further cooperation of the subject. However, despite such big popularity, face recognition task can be hardly called solved. Due to the neural network improvements in recent years, most hand-crafted feature descriptors for face recognition (and image classification generally) become obsolete. The two most popular neural network architecture for image classification tasks are ResNet and DenseNet and compare them with baseline VGG16 architecture.

2 Casia-WebFace database

As a training and testing set for tested neural networks were chosen Casia-WebFace database (Yi (2014)). The database contains 494414 RGB images of 10575 subject with resolution 250×250 pixels. The database is very challenging, i.e. images cover various intrapersonal and interpersonal differences (including pose, illumination, occlusion, age variations, haircut changes, facial expressions, etc.). For training was used only identities, which have at least 100 images presented. This leaves me with 181901 images for 925 identities. These images were augmented and resized to the resolution 64×64 pixels, This leads to 908953 images in total. This database was split into three subsets - training, validation and testing set, in ratio 70:15:15.

3 Neural Network Architectures

As a baseline architecture, it was utilized VGG16 (Simonyan et al. (2014)), which belongs to the golden standard among classification networks nowadays. Its main drawback is a huge number of parameters and very slow training.

First tested architecture was based on Deep Residual network ResNet-101. This architecture was proposed by He et al. (2016) to address the problem of degradation during learning very deep neural networks. Authors address the degradation problem by introducing deep residual learning, which is realized by implementing shortcut connections as an elemental-wise addition.

The second tested architecture was based on Dense Convolutional network DenseNet-121 (Huang et al. (2017)). This architecture was presented by Huang et al. to improving the training of very deep neural networks. Instead of residual blocks, in this architecture is each

[1] Ph.D student, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, Computer Vision specialization & Ph.D student, ITMO University, Saint Petersburg, email: grubiv@kky.zcu.cz

layer connected to every other layer in a feed-forward fashion. These connections alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

4 Experiment and results

In my experiment, I evaluate the networks on the face classification of closed subset task, i.e. I want from the networks to classify the input into one of 925 trained classes.

ResNet-101 and DenseNet-121 architectures were both downloaded with pretrained weights (both networks were trained on ImageNet challenge). Then both architectures were fine-tuned with 120k iterations with mini-batch size 64. Comparison of results of classification is showed in Table 1.

Architecture	Development set	Test set	Number of parameters
VGG16	85.2%	84.4%	132863336
ResNet-50	91.5%	90.7%	31085632
DenseNet-121	96.6%	96.2%	7901056

Table 1: Comparison of classification recognition rates

DenseNet architecture decreased the recognition error by more than 5% on both, development and test set. This is approximately 60% of relative error decrease. These results are even more significant from a point of view of a number of parameters of the used architectures. While the older architecture has approximately 31 million parameters, the newer one has approximately 8 million only, which is almost four times less. This fact confirms the results from the original DenseNet article and it also shows us a huge boost of parameters efficiency. Both state-of-the-art architectures surpassed the baseline architecture results by a large margin while also spare a huge amount of parameters and computational time.

Acknowledgement

The work has been supported by the grant of the University of West Bohemia, project No. SGS-2019-027.

References

K. Simonyan and A. Zisserman (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, vol. abs/1409.1556.

K. He, X. Zhang, S. Ren, and J. Sun (2016) Deep Residual Learning for Image Recognition. *IEEE CVPR2016*, pp. 770–778.

G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger (2017) Densely Connected Convolutional Networks. *IEEE CVPR2017*, pp. 2261–2269.

D. Yi, Z. Lei, S. Liao, and S. Z. Li (2014) Learning Face Representation from Scratch. CoRR, vol. abs/1411.7923.