Nizhny Novgorod State University
Institute of Information Technologies, Mathematics and Mechanics
Department of Computer Software and Supercomputer Technologies

Educational course «Modern methods and technologies of deep learning in computer vision»

Lecture №2
Image classification with a large number of categories using deep learning

Supported by Intel

Getmanskaya A.A., Kustikova V.D.

Nizhny Novgorod
2020

Content

1 Abstract

The goal of this lecture is to study deep neural networks for solving the problem of ***image classification with a large number of categories***.

At the beginning of the lecture, the image classification problem is stated. A brief information about the well-known ImageNet Large Scale Visual Recognition Challenge (ILSVRC) contest and the ImageNet dataset [14] are given, since the models discussed later in the lecture solve the classification problem on the specified set. Further we consider the most popular deep models: AlexNet (2012) [1], OverFeat (2013) [2], VGG-16 и VGG-19 (2014) [3], GoogLeNet (2014) [4], ResNet [5], Inception-v2 [6] и v3 [7] (2015), DenseNet (2016) [8], Xception (2016) [9], MobileNet (2017) [10], ResNeXT (2017) [11], MobileNetV2 (2018) [12], EfficientNet (2019) [13]. We discuss the main features of these architectures. If in the first models (up to 2014, up to the VGG models inclusive), the authors systematically increase the number of convolutional layers, then in the subsequent ones, encountering ***the problem of model degradation***, they try to solve this problem by introducing specific building blocks. In this regard, ***residual*** and ***inception*** blocks in various modifications appear. The lecture discusses the structure of typical blocks. Having achieved sufficiently high accuracy, the question about increasing the efficiency of deep models (reducing the model complexity with acceptable accuracy of solving the problem) has been stated. Since 2017 (starting with the MobileNet models), classes of models and approaches to their scaling are being developed to build the optimal model in terms of the "accuracy-complexity" ratio. An overview of existing models and methods of scaling is given. The lecture concludes with a comparison of classification models on the ImageNet dataset in terms of top-1 and top-5 accuracy, as well as the number of parameters [16]. For 5 years from 2014 to 2019 top-1 accuracy is increased by 10% (EfficientNet-B7 – 84.4% vs. VGG-16 – 74.4%), and the number of parameters is reduced by about 2 times (EfficientNet-B7 – 66 millions vs. VGG-16 – 138 millions) Moreover, the optimal model is always a compromise between acceptable accuracy and complexity. There are "heavy" models that demonstrate high accuracy, but inference is very slowly, and there are "light" models that give lower accuracy, but at the same time inference works in real time.

Deep models for image classification are not limited to those discussed in this lecture. There are many modifications of basic architectures [16]. Nowadays a large number of models for solving problems from other problem areas use the described architectures based on ***transfer learning*** approach, or use the basic building blocks of the considered models.

2 Literature

2.1 Books

1. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems. – 2012. – [http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf].
2. Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. – 2013. – [https://arxiv.org/pdf/1312.6229.pdf].
3. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – 2014. – [https://arxiv.org/pdf/1409.1556.pdf].
4. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. – 2014. – [https://arxiv.org/pdf/1409.4842.pdf].
5. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – 2015. – [https://arxiv.org/pdf/1512.03385.pdf].
6. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. – 2015. – [https://arxiv.org/pdf/1502.03167.pdf].
7. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf], [https://www.cv-

foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf].
8. Huang G., Liu Z., Maaten L., Weinberger K.Q. Densely Connected Convolutional Networks. – 2016. – [https://arxiv.org/pdf/1608.06993.pdf].
9. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].
10. Howard A.G., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].
11. Xie S., Girshick R., Dollar P., Tu Z., He K. Aggregated Residual Transformations for Deep Neural Networks. – 2017. – [https://arxiv.org/pdf/1611.05431v2.pdf], [https://ieeexplore.ieee.org/document/8100117].
12. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. – 2018. – [https://arxiv.org/pdf/1801.04381.pdf], [https://ieeexplore.ieee.org/document/8578572].
13. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. – 2019. – [https://arxiv.org/pdf/1905.11946.pdf].

2.2 References

14. ImageNET [http://www.image-net.org].
15. WordNet [https://wordnet.princeton.edu].
16. Image Classification on ImageNet [https://paperswithcode.com/sota/image-classification-on-imagenet].