

Нижегородский государственный университет им. Н.И. Лобачевского
Институт информационных технологий, математики и механики
Кафедра Математического обеспечения и суперкомпьютерных технологий

**Образовательный курс
«Современные методы и технологии глубокого обучения
в компьютерном зрении»**

**Лекция №2
Классификация изображений
с большим числом категорий
с использованием методов
глубокого обучения**

При поддержке компании Intel

Гетманская А.А., Кустикова В.Д.

Содержание

1	Аннотация	3
2	Литература	3
2.1	Основная литература	3
2.2	Ресурсы сети Интернет	4

1 Аннотация

Цель данной лекции состоит в том, чтобы изучить глубокие нейросетевые модели для решения задачи *классификации изображений с большим числом категорий* (image classification).

В начале лекции рассматривается постановка задачи классификации изображений. Затем дается краткая информация о широко известном конкурсе Large Scale Visual Recognition Challenge (ILSVRC) и наборе данных ImageNet [14], поскольку рассматриваемые далее в лекции модели решают задачу классификации на указанном наборе. Далее описываются широко известные глубокие нейросетевые модели: 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]. Приводятся основные особенности перечисленных архитектур. Если в первых моделях (до 2014 г., до моделей VGG включительно) авторы систематически наращивают количество сверточных слоев, то в последующих, – столкнувшись при обучении с *проблемой деградации*, они делают попытки решить эту проблему за счет введения специфичных строительных блоков. В связи с этим появляются *остаточные* (residual) и *inception-блоки* в различных модификациях. В лекции рассматривается структура типовых блоков. Достигнув достаточно высоких показателей точности, ставится вопрос о повышении эффективности глубоких моделей – снижении количества параметров при приемлемых показателях качества решения задачи. С 2017 г. (начиная с моделей MobileNet) разрабатываются классы моделей и подходы к их масштабированию для построения оптимальной модели с точки зрения соотношения «точность-сложность». Соответственно делается обзор существующих моделей и методов масштабирования. В заключении лекции проводится сравнение классификационных моделей на наборе данных ImageNet по показателям точности top-1 и top-5, а также по числу параметров [16]. За 5 лет в период с 2014 по 2019 гг. точность top-1 увеличивается на 10% (EfficientNet-B7 – 84.4% vs. VGG-16 – 74.4%), а количество параметров уменьшается примерно в 2 раза (EfficientNet-B7 – 66 млн. vs. VGG-16 – 138 млн.). При этом оптимальная модель – это всегда компромисс между приемлемой точностью и сложностью. Имеются «тяжеловесные» модели, которые демонстрируют высокие показатели качества, но работают медленно, а есть «легковесные» модели, которые дают более низкие показатели качества, но при этом работают в реальном времени.

Множество глубоких моделей для классификации изображений с большим числом категорий не ограничивается рассмотренными в настоящей лекции. Существует множество модификаций представленных архитектур [16]. В настоящее время большое количество моделей используют описанные архитектуры для решения задач из других областей за счет применения *переноса обучения* (transfer learning), либо за счет включения базовых строительных блоков из этих моделей.

2 Литература

2.1 Основная литература

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 Ресурсы сети Интернет

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>].