

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 №1
Цели и задачи, структура курса.
Схема решения задач компьютерного зрения
с использованием глубокого обучения**

При поддержке компании Intel

Кустикова В.Д.

Содержание

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

1 Аннотация

Цель данной лекции состоит в том, чтобы ознакомиться со структурой курса, рассмотреть постановки задач компьютерного зрения, которые решаются в процессе изучения курса, а также изучить общую схему решения задач компьютерного с использованием глубокого обучения.

Курс является практико-ориентированным. Он включает в себя 14 часов лекций (7 лекций по 2 академических часа) и 18 часов практических занятий с командным выполнением 4 заданий. Лекции проводятся в классической форме и сопровождаются примерами решения практических заданий с использованием инструмента Intel Distribution of OpenVINO Toolkit [31]. Выполнение практических заданий предполагает, что слушатели делятся на группы по 2-3 человека, выбирают задачу для последующего решения. Далее для групп проводятся индивидуальные консультации по ходу выполнения практических заданий.

Настоящая лекция является вводной. В лекции рассматриваются постановки классических задач компьютерного зрения: **классификация изображений** (image classification), **детектирование объектов на изображениях** (object detection), **семантическая сегментация изображений** (semantic segmentation) и **сопровождение объектов** на видео (tracking). Каждая из перечисленных задач описывается по следующей схеме. Вначале дается неформальная постановка задачи. Далее приводится строгая математическая постановка задачи и предлагается перечень широко известных данных по каждой задаче [7 – 21]. Вводятся наиболее распространенные показатели качества для дальнейшего сравнения моделей и методов решения поставленных задач.

Затем в лекции предлагается общая схема решения задач компьютерного зрения с использованием глубокого обучения. Данная схема включает несколько взаимосвязанных этапов: подготовительный этап, **обучение** (training) и **тестирование** (testing) модели, **внедрение** (deployment) модели. В заключении рассматриваются программные инструменты, которые могут использоваться на каждом из перечисленных этапов [22 – 32].

Далее в курсе рассматриваются современные глубокие нейросетевые модели для решения классических задач компьютерного зрения, поставленных в данной лекции [1 – 4]. Эти методы сравниваются в соответствии с описанными показателями качества. В заключении лекционной части курса рассматриваются глубокие модели, применяемые для генерации синтетических данных, которые могут использоваться на разных этапах решения задач. Выполнение практических заданий предполагает применение программного инструмента Intel Distribution of OpenVINO Toolkit [31], поэтому одна из промежуточных лекций посвящена вопросам компонентного состава инструмента и описанию программного интерфейса некоторых из его компонент.

2 Литература

2.1 Основная литература

1. Хайкин С. Нейронные сети. Полный курс. – М.: Издательский дом «Вильямс». – 2006. – 1104 с.
2. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика. – 2002. – 344 с.
3. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press. – 2016. – [<http://www.deeplearningbook.org>].
4. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. – Изд-во «Питер». – 2018. – 476 с.

2.2 Дополнительная литература

5. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge // International Journal of Computer Vision, 2015.

6. Bernardin K., Stiefelhagen R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics // Image and Video Processing. – 2008.

2.3 Ресурсы сети Интернет

7. MNIST [<http://yann.lecun.com/exdb/mnist>].
8. Fashion MNIST [<https://github.com/zalandoresearch/fashion-mnist>].
9. Cifar-10 / Cifar-100 [<http://www.cs.utoronto.ca/~kriz/cifar.html>].
10. Caltech-101 / Caltech-256 [http://www.vision.caltech.edu/Image_Datasets/Caltech101] / [http://www.vision.caltech.edu/Image_Datasets/Caltech256].
11. ImageNET [<http://www.image-net.org>].
12. PASCAL VOC [<http://host.robots.ox.ac.uk/pascal/VOC>].
13. MS COCO [<http://cocodataset.org>].
14. Open Images Dataset [<https://storage.googleapis.com/openimages/web/index.html>].
15. Street View House Numbers (SVHN) [<http://ufldl.stanford.edu/housenumbers>].
16. Stanford Dogs Dataset [<http://vision.stanford.edu/aditya86/ImageNetDogs>].
17. CamVid [<http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid>].
18. Cityscapes [<https://www.cityscapes-dataset.com>].
19. Multiple Object Tracking Benchmark (MOT Benchmark) [<https://motchallenge.net>].
20. TrackingNet [<https://tracking-net.org>].
21. Long-Term Visual Object Tracking Benchmark [<https://amoudgl.github.io/tlp>].
22. LabelImg [<https://github.com/tzutalin/labelImg>].
23. LabelMe [<https://github.com/wkentaro/labelme>].
24. CVAT [<https://github.com/open cv/cvat>].
25. Caffe [<https://caffe.berkeleyvision.org>], Intel Optimization for Caffe [<https://software.intel.com/en-us/frameworks/caffe>].
26. TensorFlow [<https://www.tensorflow.org>].
27. Keras [<https://keras.io>].
28. PyTorch [<https://pytorch.org>].
29. MXNet [<https://mxnet.apache.org>].
30. OpenCV [<https://opencv.org>].
31. Intel Distribution of OpenVINO Toolkit [<https://software.intel.com/en-us/openvino-toolkit>].
32. TensorRT [<https://developer.nvidia.com/tensorrt>].