

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 №4
Детектирование объектов
на изображениях**

При поддержке компании Intel

Гетманская А.А., Кустикова В.Д.

Нижний Новгород
2020

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 рассмотреть модели глубокого обучения для задачи *детектирования объектов на изображениях* (object detection).

Вначале рассматривается постановка задачи детектирования объектов. Затем дается обзор широко известных наборов данных для детектирования объектов реальной жизни (PASCAL VOC 2007, 2012 [9, 10], MS COCO [11], Open Images Dataset [12]), для детектирования и распознавания лиц (WIDER FACE [13], LFW [14], AFLW [15], IMDB-WIKI [16]), для детектирования пешеходов (Caltech [17], Wider Person [18]). Приводятся примеры изображений и разметки, а также основные характеристики указанных наборов данных (размеры тренировочной, тестовой и валидационной выборок, количество размеченных объектов). Вводятся наиболее распространенные показатели качества решения детектирования объектов: показатель числа истинных срабатываний (true positive rate), показатель числа ложных срабатываний (false detection rate), количество ложных срабатываний, в среднем приходящихся на изображение (average false positives per frame), средняя точность предсказания (average precision). Далее рассматриваются широко известные глубокие нейросетевые модели для детектирования объектов на изображениях. Выбор моделей обусловлен тем, что они по-разному решают задачу генерации областей потенциального наличия объектов. Вначале рассматривается группа двухстадийных моделей: R-CNN (Region-based Convolutional Neural Network) [1], Fast R-CNN [2], Faster R-CNN [3], R-FCN (Region-based Fully Convolutional Network) [4]. Данная группа моделей предполагает генерацию областей наличия объектов (гипотез) с помощью сторонних методов, а также последующую классификацию и уточнение границ полученных областей. Приведенные модели генерируют области с использованием метода выборочного поиска (selective search algorithm) или специальной нейронной сети RPN (Region Proposal Network). Далее рассматриваются одностаийные модели, которые обеспечивают генерацию областей и их классификацию средствами одной нейронной сети: SSD (Single Shot Multibox Detector) [5], YOLOv1 (You Only Look Once) [6], YOLOv2 [7], YOLOv3 [8]. Приводится архитектура глубоких сетей и их особенности. На данный момент значительное количество нейронных сетей, которые демонстрируют хорошие результаты детектирования на открытых наборах данных, являются модификациями приведенных в настоящей лекции моделей. Поэтому понимание структуры и способов формирования перечисленных моделей является необходимым для последующего осознания их модификаций. В заключении приводится сравнение качества и времени работы различных глубоких моделей для детектирования объектов.

Множество глубоких моделей для детектирования объектов разных классов на изображениях не ограничивается рассматриваемыми в настоящей лекции. Модели принципиально отличаются способом формирования областей возможного наличия объектов. Использование того или иного способа в значительной степени влияет на скорость работы моделей.

2 Литература

2.1 Основная литература

1. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. – 2014. – [<https://arxiv.org/pdf/1311.2524.pdf>], [<https://ieeexplore.ieee.org/abstract/document/6909475>].
2. Girshick R. Fast R-CNN. – 2015. – [<https://arxiv.org/pdf/1504.08083.pdf>], [<https://ieeexplore.ieee.org/document/7410526>].
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. – 2016. – [<https://arxiv.org/pdf/1506.01497.pdf>], [<https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>].
4. Dai J., Li Y., He K., Sun J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. – 2016. – [<https://arxiv.org/pdf/1605.06409.pdf>], [<https://papers.nips.cc/paper/6465-r-fcn-object-detection-via-region-based-fully-convolutional-networks.pdf>].

5. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: Single Shot MultiBox Detector. – 2016. – [<https://arxiv.org/pdf/1512.02325.pdf>], [https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2].
6. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection. – 2015. – [<https://arxiv.org/pdf/1506.02640.pdf>], [<https://ieeexplore.ieee.org/document/7780460>].
7. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger. – 2016. – [<https://arxiv.org/pdf/1612.08242.pdf>], [<https://pjreddie.com/darknet/yolo>].
8. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. – 2018. – [<https://pjreddie.com/media/files/papers/YOLOv3.pdf>].

2.2 Ресурсы сети Интернет

9. PASCAL VOC 2007 [<http://host.robots.ox.ac.uk/pascal/VOC/voc2007>].
10. PASCAL VOC 2012 [<http://host.robots.ox.ac.uk/pascal/VOC/voc2012>].
11. MS COCO [<http://cocodataset.org>].
12. Open Images Dataset [<https://storage.googleapis.com/openimages/web/index.html>].
13. WIDER FACE [<http://shuoyang1213.me/WIDERFACE>].
14. LFW [<http://vis-www.cs.umass.edu/lfw>].
15. AFLW [<https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw>].
16. IMDB-WIKI [<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki>].
17. Caltech [http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians].
18. Wider Person [<http://www.cbsr.ia.ac.cn/users/sfzhang/WiderPerson>].