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 №4
Object detection in images using deep neural networks

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 ***object detection***.

At the beginning of the lecture, the object detection problem is stated. An overview of well-known datasets for real-life object detection (PASCAL VOC 2007, 2012 [9, 10], MS COCO [11], Open Images Dataset [12]), for face detection and recognition (WIDER FACE [13], LFW [14], AFLW [15], IMDB-WIKI [16]), for pedestrian detection (Caltech [17], Wider Person [18]) is given. Examples of images and groundtruth are represented, as well as the main features of these datasets (sizes of train, test and validation datasets, the number of labeled objects). The most common quality metrics for object detection are introduced: true positive rate, false detection rate, average false positives per frame, average precision. Further, well-known deep neural networks for detecting objects in images are considered. Models were chosen due to the fact that they solve the problem of generating region proposals in different ways. First, a group of two-stage models is considered: R-CNN (Region-based Convolutional Neural Network) [1], Fast R-CNN [2], Faster R-CNN [3], R-FCN (Region-based Fully Convolutional Network) [4]. This group of models involves generating of region proposals (hypotheses about object location) using third-party methods, as well as the subsequent classification and refinement of the obtained bounding boxes. These models generate regions using the selective search algorithm or the special neural network RPN (Region Proposal Network). Also, we consider the one-stage models that provide the generation of regions and their classification using a single neural network: SSD (Single Shot Multibox Detector) [5], YOLOv1 (You Only Look Once) [6], YOLOv2 [7], YOLOv3 [8]. The architecture of deep networks and their features are given. At the moment, a significant number of neural networks that demonstrate good detection accuracy on public datasets are modifications of the models presented in this lecture. Therefore, an understanding of the structure and methods of constructing these models is necessary for the subsequent study of their modifications. In conclusion, a comparison of the quality and inference time of various deep models for object detection is represented.

Deep models for detecting objects of different classes in images are not limited to those considered in this lecture. Models are fundamentally different in the way they construct regions of possible object presence. The use of one or another method significantly affects the speed of model inference.

2 Literature

2.1 Books

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 References

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].