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 №6 Semantic segmentation of images 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 ***semantic segmentation***.

At the beginning of the lecture, the semantic segmentation problem is stated. An overview of well-known datasets of real-life images (PASCAL VOC 2012 [10], ADE20K [11], MS COCO'15 [12]), on-road scenes (CamVid [13], Cityscapes [14], KITTI [15]), and indoor scenes (RGBD [16], NYUDv2 [17]) is given. Examples of images and groundtruth are represented, as well as the main features of these datasets (sizes of train and test datasets, minimum/average/maximum image resolution). The most common quality metrics for semantic segmentation are introduced: pixel accuracy, mean pixel accuracy over classes, IoU metric (Intersection over Union) or Jaccard index, Dice index or F1-score. Further, well-known deep neural networks for semantic segmentation are considered. Models were chosen due to the fact that they solve the problem of obtaining an output whose spatial dimensions coincide with the resolution of the input image in different ways. We begin to study deep models for semantic segmentation with FCNs (Fully Convolutional Networks) [1], which adapt classification models by replacing the fully connected layers with the fully convolutional ones, and the problem of spatial resolution is solved using deconvolutional layers. Further, the encoder-decoder architecture is introduced on the example of SegNet [2]. We consider the U-Net [3] and PSPNet [4] models which combine features from different scales of details. Also, the ICNet model [5] is described, it is based on constructing the cascade of feature maps for different scales of the original image. Further, we consider the family of DeepLab models. DeepLab-v1 [6] is based on the constructing of a deep convolutional model to obtain a coarse map of segments and the subsequent using of conditional random fields (CRF) to refine the coarse map. DeepLab-v2 [7] and DeepLab-v3 [8] introduce the concept of Atrous Spatial Pyramid Pooling for combining features at different scales. DeepLab-v3+ [9] is based on the applying encoder-decoder architecture to the DeepLab-v3 model. In conclusion, a comparison of the quality and inference time of various deep models for semantic segmentation of on-road scenes on the Cityscapes dataset [14] is represented [18]. Results of semantic segmentation for another datasets are available by link [19].

Deep models for semantic segmentation are not limited to those considered in this lecture. Models differ the way of solving the problem of obtaining an output whose spatial dimensions coincide with the resolution of the input image. As a rule, the decision greatly affects the model performance. Therefore, constructing of an optimal model is a compromise between the segmentation quality and the model complexity.

2 Literature

2.1 Books

1. Long J., Shelhamer E., Darrel T. Fully Convolutional Networks for Semantic Segmentation. – 2015. – [https://arxiv.org/pdf/1411.4038.pdf], [https://ieeexplore.ieee.org/document/7298965].
2. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. – 2015. – [https://arxiv.org/pdf/1511.00561.pdf], [https://ieeexplore.ieee.org/document/7803544].
3. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. – 2015. – [https://arxiv.org/pdf/1505.04597.pdf], [https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28].
4. Zhao H., Shi J., Qi X., Wang X., Jia J. Pyramid scene parsing network. – 2016. – [https://arxiv.org/pdf/1612.01105.pdf], [https://ieeexplore.ieee.org/document/8100143].
5. Zhao H., Qi X., Shen X., Shi J., Jia J. ICNet for Real-Time Semantic Segmentation on High-Resolution Images. – 2017. – [https://arxiv.org/pdf/1704.08545.pdf], [https://link.springer.com/chapter/10.1007/978-3-030-01219-9_25].
6. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. – 2014. – [https://arxiv.org/pdf/1412.7062.pdf].

7. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. – 2017. – [https://arxiv.org/pdf/1606.00915.pdf], [https://ieeexplore.ieee.org/document/7913730].
8. Chen L.-C., Papandreou G., Schroff F., Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. – 2017. – [https://arxiv.org/pdf/1706.05587.pdf].
9. Chen L.-C., Zhu Y., Papandreou G., Schoff F., Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. – 2018. – [https://arxiv.org/pdf/1802.02611.pdf].

2.2 References

10. PASCAL VOC 2012 [http://host.robots.ox.ac.uk/pascal/VOC/voc2012].
11. ADE20K [http://groups.csail.mit.edu/vision/datasets/ADE20K].
12. MS COCO’15 [http://mscoco.org].
13. CamVid [http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid].
14. Cityscapes [https://www.cityscapes-dataset.com].
15. KITTI [http://www.cvlibs.net/datasets/kitti].
16. Sun-RGBD [http://rgbd.cs.princeton.edu].
17. NYUDv2 [http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html].
18. Real-Time Semantic Segmentation on Cityscapes test [https://paperswithcode.com/sota/real-time-semantic-segmentation-on-cityscapes].
19. Semantic Segmentation [https://paperswithcode.com/task/semantic-segmentation/latest].