

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 №6  
Семантическая сегментация изображений  
с использованием методов глубокого обучения**

*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 изучить глубокие нейросетевые модели, используемые для решения задачи *семантической сегментации изображений* (semantic segmentation).

Вначале рассматривается постановка задачи семантической сегментации. Затем дается обзор широко известных наборов данных семантической сегментации объектов реальной жизни (PASCAL VOC 2012 [10], ADE20K [11], MS COCO'15 [12]), сегментации дорожных объектов (CamVid [13], Cityscapes [14], KITTI [15]), сегментации интерьеров с информацией о глубине (Sun-RGBD [16], NYUDv2 [17]). Приводятся примеры изображений и разметки, а также основные характеристики указанных наборов данных (размеры тренировочной и тестовой выборки, максимальное/минимальное/среднее разрешение изображений и другие). Вводятся наиболее распространенные показатели качества решения задачи семантической сегментации: попиксельная точность (pixel accuracy), средняя попиксельная точность по классам наблюдаемых объектов (mean pixel accuracy over classes), метрика IoU (Intersection over Union) или индекс Жаккара (Jaccard index), индекс Дайса (Dice index) или F1-score. Далее рассматриваются широко известные глубокие нейросетевые модели для семантической сегментации изображений. Выбор моделей обусловлен тем, что они по-разному решают проблему получения выхода, пространственные размеры которого совпадают с разрешением входного изображения. Изучение глубоких моделей для сегментации начинается с архитектуры полностью сверточных нейронных сетей FCN (Fully Convolutional Networks) [1], которые адаптируют классификационные модели посредством замены полносвязных слоев на полностью сверточные, а проблема с пространственным разрешением выходной карты сегментов решается с помощью обратных сверточных слоев. Далее вводится архитектура «кодировщик-декодировщик» на примере модели SegNet [2]. Рассматриваются топологии сетей U-Net [3] и PSPNet [4], которые учитывают признаки с разных уровней детализации, а также модель ICNet [5], основанная на построении каскада признаковых описаний для разных масштабов входного изображения. Затем приводится описание моделей DeepLab: DeepLab-v1 [6], построенная на базе сверточной нейронной сети для получения «грубой» карты сегментов и последующем применении условных случайных полей (Conditional Random Fields, CRF) для уточнения полученных результатов; DeepLab-v2 [7] и DeepLab-v3 [8], в которых вводится понятие пространственной пирамиды сверток с пропусками (Atrous Spatial Pyramid Pooling, ASPP) для интеграции признаков на разных масштабах; DeepLab-v3+ [9], основанная на применении архитектуры «кодировщик-декодировщик» к модели DeepLab-v3. В заключении приводится сравнение качества и времени работы различных глубоких моделей для семантической сегментации дорожных сцен набора данных Cityscapes [14] согласно [18]. Сравнение на других наборах данных можно найти по ссылке [19].

Множество глубоких моделей для семантической сегментации изображений не ограничивается рассмотренными в настоящей лекции. Модели различаются способом решения проблемы формирования выходной карты сегментов, пространственные размеры которой совпадают с разрешением входного изображения. Как правило, решение в значительной степени влияет на скорость работы модели. Поэтому построение оптимальной модели – поиск компромисса между качеством решения прикладной задачи и сложностью модели.

## 2 Литература

### 2.1 Основная литература

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](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](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 Ресурсы сети Интернет**

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](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>].