

# Семантическая сегментация дорожной разметки с использованием нейронных сетей

А. Р. Музалевский<sup>1</sup>, Е. В. Серых, М. М. Копычев, Е. В. Друян, М. А. Чернышев

*Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

<sup>1</sup>muzal2010@yandex.ru

**Аннотация.** В статье рассматриваются вопросы построения свёрточной нейронной сети, предназначенной для семантической сегментации горизонтальной дорожной разметки на основе данных камеры. Для обучения сети использовался набор данных ApolloScare.

**Ключевые слова:** семантическая сегментация; нейронная сеть; дорожная разметка.

## I. ВВЕДЕНИЕ

Семантическую сегментацию можно представить в виде последовательности действий:

- классификации определённого объекта на изображении;
- локализации – поиска объекта и определения его описывающего прямоугольника;
- сегментации – группирования пикселей в области описывающего прямоугольника путём создания сегментационной маски.

Задача семантической сегментации может быть отнесена к классификации определённого класса изображений и отделению его от остальных классов изображений путем наложения на него маски сегментации, что также можно рассматривать как классификацию изображений на уровне пикселей [1].

Сегментационные модели применяются в различных областях:

- Беспилотные транспортные средства. Автомобилю необходим модуль взаимодействия с окружающей средой, чтобы обнаруживать препятствия и планировать в соответствии с этой информацией дальнейший маршрут.
- Медицина. Определение злокачественных клеток, поражений органов, обнаружение частей органов.
- Промышленная робототехника. Работа по сортировке на конвейере, создание печатных плат.
- Системы распознавания и видеонаблюдения. Поиск людей, системы обнаружения дорожных пробок и аварий.

Целью данной работы является получение базовой модели для сегментации горизонтальной дорожной разметки.

## II. ОБЗОР РАБОТ СХОЖЕЙ ТЕМАТИКИ

Семантическая сегментация дорожной разметки является известной задачей машинного обучения, для решения которой существуют различные методы и подходы.

Аналитический подход. В типичной задаче обнаружения разметки первым шагом является цветовая сегментация исходного изображения, поскольку во всем мире разметка полос движения преимущественно окрашена в желтый или белый цвета. Вторым шагом – обнаружение границ. Затем используются различные техники уменьшения шума, после чего применяется преобразование Хафа [2], позволяющее выявить различные линии на изображении. В конце производится анализ выявленных линий в соответствии с задачей. Применимость такого подхода ограничена его небольшой устойчивостью к изменению изображения (например, изменение освещённости) [3].

Методы глубокого обучения. В работе [4] описана система детекции дорожной разметки на основе данных стереокамеры. В работе [5] использован фреймворк YOLO для решения задачи сегментации дорожной разметки в реальном времени. В работе [6] предложена оригинальная архитектура нейронной сети для сегментации дорожной разметки, отличающаяся использованием положения точки схода для данного кадра. В [7] предложен собственный подход к глубокому обучению – LaneNet, основанный на процедуре кодирования из E-net [8].

## III. ОПИСАНИЕ НАБОРА ДАННЫХ ДЛЯ ОБУЧЕНИЯ

Задача распознавания разметки дорожного полотна была сведена к задаче семантической сегментации. В работе была использована архитектура свёрточной нейронной сети U-Net [9] с ResNet34 [10]. В качестве функции потерь использовалась Cross Entropy Loss, оптимизатор – метод адаптивной оценки моментов (Adam).

Для работы был выбран набор данных с открытым доступом ApolloScare [11], который является частью проекта Apollo, направленного на создание беспилотных автомобилей, что требует решения задач представления окружающей среды, навигации, планирования и управления транспортными средствами. Из всего набора данных для семантической сегментации была создана выборка с разметкой дорожного покрытия, размер которой составил 160 тысяч кадров высокого разрешения, полученных при помощи стереокамер, установленных на крыше автомобилей, в различных городах. Использование стереокамеры позволяет получить два кадра одного и того же места с разных

ракурсов. В этом наборе данных размечены 35 различных классов, относящихся к дорожной разметке.

В ходе анализа набора данных было выявлено большое количество ошибок в разметке, что негативно сказалось на метриках работы модели (рис. 1). Для всех кадров с правой камеры область автомобиля, показанная белым цветом на маске, выделена некорректно – в правом углу вместо левого.

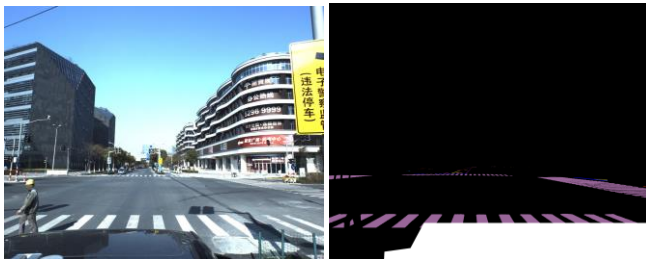


Рис. 1. Пример кадра с некорректной маской

Из-за ограниченных вычислительных мощностей невозможно использовать все доступные данные, поэтому в работе использовался 1% от исходного набора данных (1044 кадра). Данный объем данных был разделён на три выборки:

- обучающая выборка: 522 кадра (50 %);
- валидационная выборка: 224 кадра (21,5 %);
- тестовая выборка: 298 кадров (28,5 %).

При разделении выборок во избежание утечек информации, данные с записью одного и того же места (дороги) с разных камер попадали в одну выборку.

Чтобы повысить степень устойчивости модели к изменениям входных данных, при обучении использовались различные аугментации:

- преобразование кадра в оттенки серого (с вероятностью 30 %);
- отражение по горизонтали (с вероятностью 50 %);
- добавление шума Гаусса (с вероятностью 50 %);
- добавление мультипликативного шума (с вероятностью 30 %);
- сжатие изображения (с вероятностью 50 %);
- добавление медианного размытия (с вероятностью 20 %);
- добавление размытия Гаусса (с вероятностью 20 %);
- добавление размытия в движении (с вероятностью 20 %);
- добавление стандартного размытия (с вероятностью 20 %);
- CLAHE (с вероятностью 50 %);
- добавление зернистости (с вероятностью 50 %);
- Emboss (с вероятностью 50 %);

- случайное изменение яркости и контраста (с вероятностью 50%);
- случайное изменение насыщенности (с вероятностью 30 %);
- изменение масштаба, сдвиг и поворот на 15 градусов (с вероятностью 50 %);
- CoarseDropout (с вероятностью 50 %).

Пример входного кадра с целевой маской представлен на рис. 2 и 3 соответственно.



Рис. 2. Пример изображения с камеры из набора данных ApolloScape

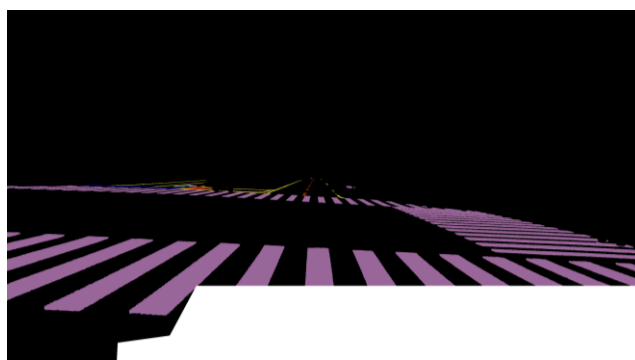


Рис. 3. Пример размеченной маски для кадра с рис. 1

Каждый кадр представлен RGB-изображением, размером (2710, 3384) пикселя. Чтобы использовать нейронную сеть каждый кадр был разделён на сегменты размером (512, 512) пикселей. Таким образом, 1044 кадра были преобразованы в 43848 изображений, на вход сети подавалось следующее количество кадров:

- обучающая выборка из 21924 изображений (50%);
- валидационная выборка из 9408 изображений (21,5%);
- тестовая выборка из 12516 изображений (28,5%).

Были выбраны 6 классов для сегментации:

- фон (background): цвет RGB (0, 0, 0);
- разделительная сплошная жёлтая линия (SYD): цвет RGB (60, 15, 67);
- белая прерывистая линия (BWG): цвет RGB (142, 35, 8);
- разделительная сплошная белая линия (SWD): цвет RGB (180, 173, 43);

- белая стоп-линия (SWS): цвет RGB (0, 0, 192);
- пешеходный переход (CWYZ): цвет RGB (153, 102, 153).

Модель обучалась в течение 80 эпох, для визуализации процесса обучения использовался ClearML.

На рис. 4 представлено распределение пикселей классов в процентном отношении, без класса фона.

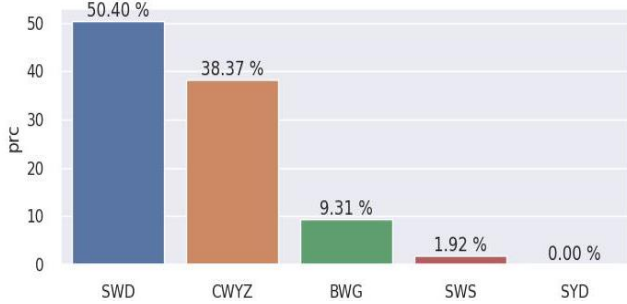


Рис. 4. Соотношение площадей масок целевых классов в наборе данных

В данных присутствует дисбаланс, мажорными классами являются: «SWD» (разделительная белая сплошная линия) и «CWYZ» (пешеходный переход). Класс «SYD» является минорным классом и в данных, использованных в данной работе, почти не представлен. В дальнейших исследованиях планируется расширить информацию по этому классу.

Для решения проблемы дисбаланса данных в процесс обучения были введены весовые коэффициенты при расчёте функции потерь такие, что модель штрафовалась сильнее за ошибку в классах разметки, чем за ошибку в определении фона.

#### IV. РЕЗУЛЬТАТЫ

На рис. 5–8 представлены графики обучения и валидации нейронной сети. По графикам видно, что модель не переобучилась.

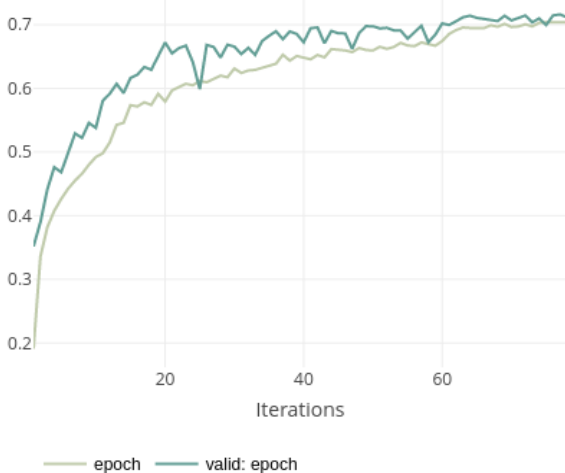


Рис. 5. График изменения dice\_macro

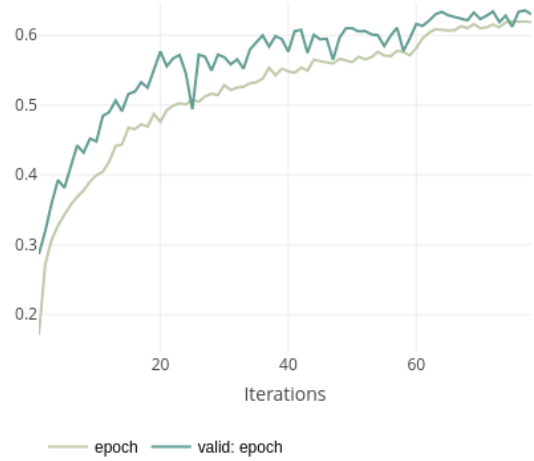


Рис. 6. График изменения iou\_macro

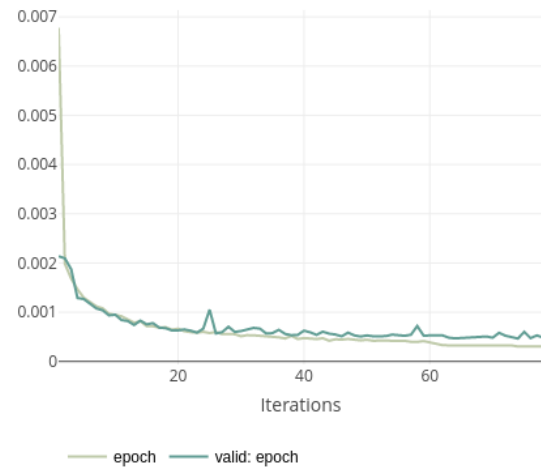


Рис. 7. График изменения потерь

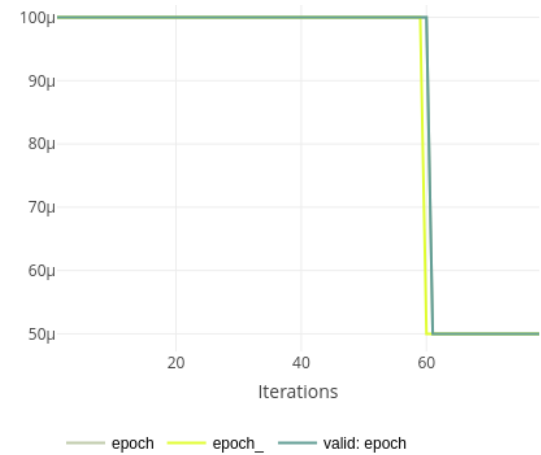


Рис. 8. График изменения скорости обучения

В табл. 1 представлены метрики качества работы обученной модели на тестовой выборке.

ТАБЛИЦА I. МЕТРИКИ МОДЕЛИ

Целевой класс	Показатели	
	iou	dice
background	0.998	0.999
SYD	0.986	0.987
BWG	0.693	0.775
SWD	0.794	0.881
SWS	0.114	0.162
CWYZ	0.386	0.496

Целевой класс	Показатели	
	iou	dice
micro	0.995	0.998
macro	0.662	0.717

Низкие значения метрик для класса SWS (белая стоп-линия) можно объяснить тем, что в выборках мало таких примеров. Выравнивание распределения пикселей классов в обучающей выборке позволит повысить качество сегментации.

На рис. 9 и 11 представлены результаты работы модели сегментации в виде исходного изображения с наложенной предсказанной маской, а на рисунках 10 и 12 показаны размеченные маски для исходных кадров.



Рис. 9. Результат работы модели на участке дороги с поворотом

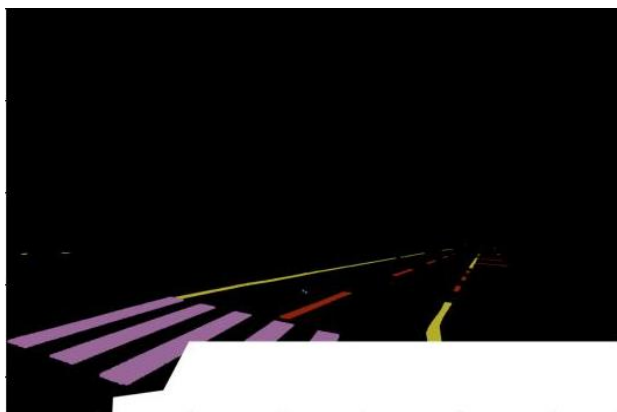


Рис. 10. Размеченная маска для тестового изображения с рис. 9



Рис. 11. Результат работы модели на прямом участке дороги с различным освещением.

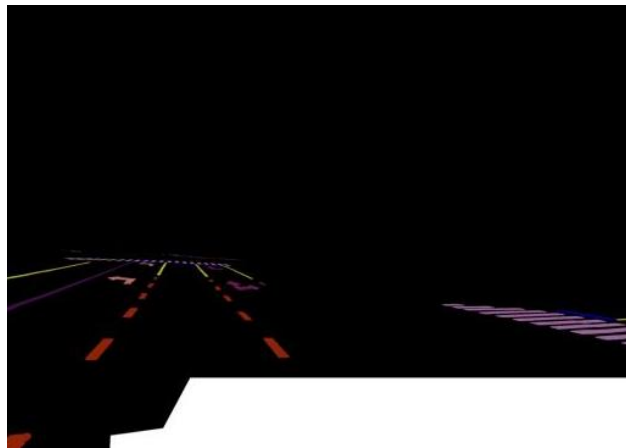


Рис. 12. Размеченная маска для тестового изображения с рис. 11

Как видно на рис. 9, модель успешно справилась с обнаружением пешеходного перехода, прерывистой линии и сплошной линии. Однако на более контрастном кадре (рис. 11) в тени наблюдаются артефакты на месте пешеходного перехода, а также не обнаружены линии на крайней левой полосе, поскольку в процессе поворота эта полоса оказалась левее привычного положения в сравнении с кадрами прямолинейного движения, которые составляют большую часть обучающей выборки. Повышение разнообразия условий окружающей среды и дорожных ситуаций может повысить качество работы модели.

## V. ЗАКЛЮЧЕНИЕ

Несмотря на неравномерность распределения целевых классов в обучающем наборе и ошибки разметки, модель справляется с задачей сегментирования разметки в кадре.

В качестве развития работы в дальнейших исследованиях планируется:

- увеличить объём данных для обучения, валидации и тестирования;
- увеличить разнообразие данных, чтобы повысить степень устойчивости модели;
- добавить масштабирование исходного кадра к меньшему размеру, чтобы уменьшить время на обучение сети;
- добавить настройку гиперпараметров модели;
- добавить нормализацию значений пикселей при обучении сети;
- уточнить целевые маски (особенно в области автомобиля);
- рассмотреть альтернативные архитектуры нейронных сетей (например, DeepLab).

## СПИСОК ЛИТЕРАТУРЫ

- [1] Онлайн-учебник по семантической сегментации. Онлайн-ресурс. URL: <https://www.v7labs.com/blog/semantic-segmentation-guide>
- [2] Hough Transform. Онлайн-ресурс. URL: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm>
- [3] Девяткин А.В., Музалевский А.Р., Морозов А.С. Система компьютерного зрения для распознавания показаний стрелочных приборов // 2022 XXV International Conference on Soft Computing and Measurements (SCM), 2022, С. 275-278.

- [4] W. Song, Y. Yang, M. Fu, Y. Li, M. Wang, Lane detection and classification for forward collision warning system based on stereo vision, *IEEE Sensors Journal*, 2018, 18 (12) pp. 5151–5163.
- [5] Nugraha B.T., Su S.-F. Towards self-driving car using convolutional neural network and road lane detector, in: 2017 2nd international conference on automation, cognitive science, optics, micro-electro-mechanical system, and information technology (ICACOMIT), 2017, pp. 65–69.
- [6] S. Lee, J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T.- H. Lee, H. S. Hong, S.-H. Han, I. S. Kweon, Vpnet: Vanishing point guided network for lane and road marking detection and recognition, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 1947–1955.
- [7] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans, L. V. Gool, Towards end-to-end lane detection: an instancesegmentation approach, in: 2018 IEEE intelligent vehicles symposium (IV), 2018, pp. 286–291.
- [8] Paszke A., Chaurasia A., Kim S., Culurciello E. Enet: A deep neural network architecture for real-time semantic segmentation, arXiv preprint arXiv: 2016, 1606.02147.
- [9] Shelhamer, Evan et al. “Fully convolutional networks for semantic segmentation.” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014): 3431-3440.
- [10] He, Kaiming et al. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 770-778.
- [11] Описание набора данных ApolloScape. Онлайн-ресурс. URL: [https://apolloscape.auto/lane\\_segmentation.html](https://apolloscape.auto/lane_segmentation.html)