

УДК 004.93'12

СЕМАНТИЧЕСКАЯ СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ С КАМЕР БЕСПИЛОТНЫХ АВТОМОБИЛЕЙ В РЕАЛЬНОМ ВРЕМЕНИ

М.О. Доброхвалов (*modobrokhvalov@gmail.com*)

А.Ю. Филатов (*ant.filatov@gmail.com*)

Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ», Санкт-Петербург

Аннотация. Семантическая сегментация - ключевая для понимания окружающих сцен технология автономных транспортных средств. В данной работе предлагается модификация существующей нейронной сети с использованием блоков, предложенных в других исследованиях. Скорость работы предложенной модификации превышает скорость работы рассмотренных аналогов.

Ключевые слова: семантическая сегментация, нейронная сеть, архитектура сети.

Введение

В современном мире значительное внимание уделяется области автоматизации управления транспортными средствами. Для взаимодействия с окружающей средой автономным транспортным средствам нужна информация с различных датчиков, таких как камера и др. В данных, полученных с обычной камеры, сложно выделить классы объектов. Обработка данных должна происходить в реальном времени (не менее 30 кадров в секунду). Таким образом, решаемой задачей является семантическая сегментация изображений в реальном времени. В настоящее время для решения данной задачи используются нейронных сетей.

В работе будут рассмотрены одни из лучших решений задачи семантической сегментации изображений в реальном времени с использованием нейронных сетей. После этого будет предложена модификация, построенная на основе нескольких работ и описана скорость работы предложенной модификации.

1 Существующие работы

Существует множество разных реализаций нейронных сетей, решающих задачу семантической сегментации. Данная работа основана на одних из лучших работ (по метрике среднее пересечение над объединением на момент написания статьи).

Первая работа [Hong et al., 2021] описывает двухветвевую архитектуру сети DDRNet, в которой одна ветвь выделяет контекстную информацию, а другая – пространственную.

В работе [Gao S. et al. 2019] рассматривается модуль Res2Net, который может использоваться в различных архитектурах нейронных сетей. В данном модуле разбиения обрабатываются в многомасштабном режиме, что способствует извлечению как глобальной, так и локальной информации.

Работа [Chao P. et al., 2019] предлагает новую архитектуру блока, называемого HarDBlock. Идея в том, что слой k соединен со слоем $k - 2^n$, если 2^n делит k , где n - неотрицательное целое число и $k - 2^n \geq 0$; в частности, слой 0 является входным слоем.

2 Модификация и первичные оценки

Идея работы заключается в том, что использование двухветвевой архитектуры DDRNet с замещением основных блоков этой сети на блоки HarDBlock, а также использование блока Res2Net в качестве бутылочного горлышка, позволит увеличить значение метрики “среднее пересечение над объединением”.

Использование двухветвевой архитектуры предполагает выделение контекстной и пространственной информации, которые являются представлением информации разных уровней. Это позволяет получать большее представление как об общей картине, так и о точных границах, при слиянии информации двух данных представлений. Что касается использования блока HarDBlock, то его архитектура позволяет учитывать каждый слой свертки внутри блока [Fan et al. 2021], а не только последний, что является важным в задаче сегментации. Использование блока Res2Net позволяет использовать преимущества блока бутылочного горлышка, при этом дополняя стандартную архитектуру локальной многомасштабной обработкой.

В работе [Fan et al. 2021] представлена модификация процесса обучения нейронной сети. Генерируется бинарная достоверная информация (ground-truth) из семантической сегментации. Данная информация используется для определения точных границ при вычислении функции потерь для последующей корректировки границ классов. Использование описанной модификации обучения также предполагает улучшение метрики среднее пересечение над объединением.

Обучение производится на датасете Cityscapes. В качестве оптимизатора используется стохастический градиентный спуск (SGD) с импульсом (momentum) 0.9, сокращением весов (weight decay) $5e-4$. Также выбрана полиномиальное планирование скорости обучения (learning rate) со степенью 0.9.

Другой важной метрикой является количество кадров в секунду. Все измерения проводились на видеокарте NVIDIA GeForce GTX 1050Ti Mobile. Оригинальная сеть DDRNet показывала 8.69 FPS, HarDNet68 - 7.57 FPS предлагаемая модификация - 10.70 FPS. Как можно заметить, предложенная модификация показала более быструю работу, чем аналоги, на основе которых строится изучаемая модель. Измерение скорости производилось на необученных сетях.

Заключение

В рамках работы была предложена модификация архитектуры сети DDRNet с использованием блоков HarDBlock и Res2Net. Также предложено при обучении использовать модификацию функции потерь. На видеокарте 1050Ti модификация показала более быструю работу, чем аналоги, на основе которых строится изучаемая модель. В качестве дальнейшего развития работы планируется произвести обучение модели, выявить влияние каждого блока. Также планируется выполнить сравнение предложенной модификации с оригинальными моделями по точности (метрика среднее пересечение над объединением).

Список литературы

- [Chao et al., 2019] Chao P. et al. Hardnet: A low memory traffic network //Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2019. – С. 3552-3561.
- [Fan et al. 2021] Fan M. et al. Rethinking BiSeNet For Real-time Semantic Segmentation //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2021. – С. 9716-9725.
- [Gao et al. 2019] Gao S. et al. Res2net: A new multi-scale backbone architecture //IEEE transactions on pattern analysis and machine intelligence. – 2019.
- [Hong et al. 2021] Hong Y. et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes //arXiv preprint arXiv:2101.06085. – 2021.