

ИСПОЛЬЗОВАНИЕ ДИСТИЛЛЯЦИИ МЕЖДУ СЛОЯМИ НЕЙРОННОЙ СЕТИ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА СЕГМЕНТАЦИИ ДОРОЖНОЙ РАЗМЕТКИ

Сиротко С. И., Фисько Д. В.

Кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь

E-mail: sergeyis@bsuir.by, dmitryfisko@gmail.com

Обучение нейронных сетей для распознавания дорожной разметки трудоёмкий процесс из-за разряженной структуры целевого класса. Без использования дополнительных приёмов при обучении модели, полученные модели зачастую сильно ошибаются в трудных условиях: слабом или слишком насыщенном освещении, при наличии изгибов дорог, объектов на дороге и т.д. Чтобы улучшить качество сегментации, предлагается в ходе обучения модели, после изначального предобучения, дополнить функцию потерь дополнительным слагаемым - метрикой схожести карт активации между последовательными слоями нейронной сети.

ВВЕДЕНИЕ

Распознавание дорожной разметки необходимо для понимания обстановки участниками транспортного движения, а также локализации положения транспортного средства на дорожном покрытии. Применение автоматизированного алгоритма для распознавания дорожной разметки затрудняется всей вариативностью возможного окружения: наличие машин вокруг, колебания освещения и плохих погодных условий, стёртой или полностью отсутствующей разметки.

Зачастую задачу сегментации дорожной разметки сводят к постановке в виде классической задачи сегментации изображения, используя нейросетевые модели с полностью связанными архитектурами. Обучение данных моделей сильно зависит от количества сигнала предсказываемого класса. Т.к. в обучающей выборке линии дорожной разметки зачастую представлены в виде тонких линий, то количество пикселей разметки сильно меньше, чем количество пикселей фона, что сильно затрудняет процесс обучения модели.

I. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Один из способов преодоления этого ограничения заключается в искусственном увеличении размеров дорожной разметки в обучающей выборке, но использование данного решения вносит неточности в качестве локализации разметки. Также в литературе были предложены методы многоклассовой сегментации - Multi Task Learning [2] и установка дополнительных соединений между слоями - Message Passing [3]. Метод MLT требует дополнительных усилий при сборе обучающей выборки, потому что помимо необходимого класса дорожной разметки, должны быть существенно представлены области объектов других классов. ML метод работает через добавление дополнительных соединений между слоями нейронной сети и значительно (на 35%)

замедляет производительность вычислительного графа нейронной сети.

II. ДИСТИЛЛЯЦИЯ КАРТ АКТИВАЦИЙ

Предлагаемый метод заключается в добавление нового штрафа в функцию потерь. Тем самым метод не будет затрагивать производительность нейронной сети при непосредственном использовании, а будет увеличивать вычислительную сложность только во время обучения. Цель добавления нового штрафа в функцию потерь заключается в следующем - стимулировать карты активаций нижних слоёв нейронной сети мимикрировать под карты активации верхних слоёв нейронной сети. Процесс обучения мимикрированию называется дистилляцией.

Интуиция этого метода исходит из следующего наблюдения. После предобучения нейронной сети, карты активаций в слоях нейронной сети накапливают достаточное количество разнообразной контекстной информации, которая позволяет визуальной определять разметку дороги. Эту информацию стоит усилить, через добавления дистилляции между соседними слоями. Т.е. карты активации первого слоя должны стремиться предсказывать карты активации второго слоя, а карты активации второго слоя предсказывают карты активации третьего слоя и т.д. Тем самым слои нижнего уровня стремятся извлекать больше контекстной информации.

Как показали эксперименты из таблицы 1, данный шаг даёт дополнительную информацию нейронной сети при обучении и позволяет при использовании нейросетевых моделей с меньшим количеством параметров получать аналогичное большим нейросетевым архитектурам показатели целевых метрик, например по сравнению с архитектурой модели SCNN [3].

В силу неполной изученности метода, его эффективное применение на данный момент достигнуто только на небольших нейросетевых моделях, таких как: ENet [4], ResNet-18 и ResNet-34

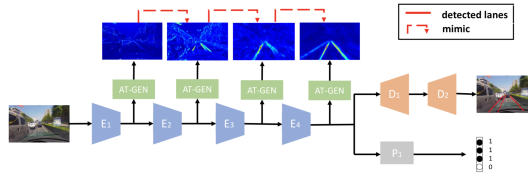


Рис. 1 – Архитектура сети ENet с добавлением дистилляции между блоками энкодеров

[5]. Более детально распишем методологию работы метода на примере модели с архитектурой ENet.

III. МЕТОДОЛОГИЯ

Добавление дистилляции между слоями нейронной сети может происходить в произвольный момент обучения нейронной сети, но добавление должно происходить когда нейронная сеть достаточно обучилась. На примере архитектуры ENet (рисунок 1), добавление штрафа на точность дистилляции происходит следующим образом: между блоками E1-E4 добавляется вычитывание агрегата над активациями нейронной сети и берётся сумма евклидова расстояния между агрегатами последовательных слоёв. За генерирование агрегата на основе карт активации слоя, отвечает блок называемый AT-GEN. Обозначим блок AT-GEN как $\Psi(\cdot)$. Он принимает на вход матрицу активаций m по счёту слоя - матрицу A_m и преобразует её по следующей формуле:

$$\Psi(A_m) = \Phi(\mathcal{B}(\mathcal{G}_{sum}^2(A_m))), \quad (1)$$

где преобразование $\mathcal{G}(\cdot)$ суммирует поканально возведённые в квадрат значения карт активаций слоя, преобразование $\mathcal{B}(\cdot)$ масштабирует размер полученного агрегата через билинейную дискретизацию, а преобразование $\Phi(\cdot)$ является поканальным softmax преобразованием.

Используя данное для \mathcal{G} определение, итоговый вид функция штрафа основанной на точности дистилляции между последовательными слоями нейронной сети будет иметь следующий вид:

$$\begin{aligned} \mathcal{L}_{distill}(A_m, A_{m+1}) \\ = \sum_{m=1}^{M-1} (L_2(\Psi(A_m), \Psi(A_{m+1}))), \end{aligned} \quad (2)$$

где A_m - это матрица активаций слоя m , M это количество слоёв на которых добавляется дистилляция по картам активаций, L_2 это функция евклидова расстояния, а Ψ это функция описанная ранее в формуле 1.

IV. ЭКСПЕРИМЕНТЫ

Для сравнения результатов был выбран метод называемый Deep Supervision [6]. Он заключается в том, чтобы карты активации предсказывали аналогичные целевому классу силуэты. Также тестирование метода проводилось на различных обучающих выборках: CuLane [3] и BDD100 [7].

Таблица 1 – Сравнение результатов

Алгоритм	TuSimple	BDD100K	
	Accuracy	Accuracy	IoU
ENet	93.02%	34.12%	14.64
ENet-Deep	94.69%	35.61%	15.38
ENet-SAD	96.64%	36.56%	16.02
ResNet-34	92.84%	34.12%	12.24
ResNet-34-Deep	94.52%	31.72%	13.59
ResNet-34-SAD	96.24%	32.68%	14.56

ЗАКЛЮЧЕНИЕ

Из полученной таблицы 1 делаем вывод, что добавление дистилляции между слоями нейронной сети для решения задачи распознавания дорожной разметки оправдано как минимум для небольших нейросетевых архитектур. Применимость метода для нейросетевых архитектур большего размера предстоит изучить.

1. Learning Lightweight Lane Detection CNNs by Self Attention Distillation – ICCV 2019. – Режим доступа: <https://arxiv.org/abs/1908.00821>. – Дата доступа: 01.10.2019.
2. VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition – arXiv.org, 2017. – Режим доступа: <https://arxiv.org/abs/1710.06288>. – Дата доступа: 03.10.2019.
3. Spatial as Deep: Spatial CNN for Traffic Scene Understanding – AAAI, 2018. – Режим доступа: <https://arxiv.org/abs/1712.06080>. – Дата доступа: 03.10.2019.
4. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation – arXiv.org, 2016. – Режим доступа: <https://arxiv.org/abs/1606.02147>. – Дата доступа: 07.10.2019.
5. Deep Residual Learning for Image Recognition – arXiv.org, 2015. – Режим доступа: <https://arxiv.org/abs/1512.03385>. – Дата доступа: 07.10.2019.
6. Holistically-Nested Edge Detection – arXiv.org, 2015. – Режим доступа: <https://arxiv.org/abs/1504.06375>. – Дата доступа: 07.10.2019.
7. BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling – arXiv.org, 2018. – Режим доступа: <https://arxiv.org/abs/1805.04687>. – Дата доступа: 01.10.2019.