

Использование ГПУ в задаче классификации дорожных знаков при помощи сверточных нейронных сетей

А.В. Шустанов^а, П.Ю. Якимов^а

^а Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Московское шоссе, 34, Самара

Аннотация

В последнее время широкое распространение для решения задач классификации получили методы глубинного обучения. В настоящей статье предложена реализация алгоритма классификации дорожных знаков при помощи сверточной нейронной сети. Обучение нейронной сети реализовано с использованием библиотеки TensorFlow и архитектуры многопоточного программирования CUDA. Классификация проходит в режиме реального времени на мобильном графическом процессоре. Экспериментальные результаты подтвердили высокую эффективность разработанной системы классификации дорожных знаков.

Ключевые слова: TensorFlow; сверточные нейронные сети; классификация дорожных знаков; обработка изображений; система компьютерного зрения; мобильные графические процессоры

1. Введение

Развитие технического уровня современных мобильных процессоров позволило многим автопроизводителям установить системы технического зрения в серийные автомобили. Среди прочих задач, которые решают такие системы, решается также и задача классификации дорожных знаков. Однако, основными проблемами таких систем является низкая точность обнаружения, а также невозможность некоторых систем классифицировать российские дорожные знаки.

Обнаружение дорожных знаков, как правило, выполняется в два этапа: локализация знака и последующая классификация. Есть много различных методов локализации [1–3]. В предыдущих работах [4], [5] авторы представили эффективную реализацию алгоритмов предварительной обработки кадров и локализации дорожных знаков. При этом, обработка выполнялась в реальном времени. Использованный модифицированный алгоритм Хафа позволил определять точные координаты дорожного знака, поэтому на этапе классификации использовался простой алгоритм сопоставления локализованного изображения с шаблоном. Данный алгоритм позволил достичь хороших результатов в сочетании с этапом локализации и показывал результат в 97,3 % точности распознавания дорожных знаков. Для обучения и тестирования алгоритмов была использована база данных дорожных знаков GTSDDB [6]. На рис. 1 приведены изображения для обучения алгоритмов распознавания дорожных знаков и для тестирования алгоритмов локализации.



Рис. 1. Изображения из немецкой базы данных дорожных знаков.

При последующих попытках тестирования разработанной технологии обнаружения и классификации дорожных знаков на эффективность работы на видеопоследовательностях с реальной дорожной обстановкой, полученных при помощи камеры, установленной в салоне автомобиля, было выявлено существенное снижение эффективности работы сквозной технологии. Исследования показали, что такое снижение возникло из-за слишком сильных вариаций в освещении, контрастности и угле поворота изображений локализованных дорожных знаков относительно камеры. Таким образом, простой алгоритм классификации – сопоставление с шаблоном – не мог справиться с задачей по причине ограниченного набора заранее заданных шаблонов. Было принято решение исследовать возможность применения получивших столь широкое применение сверточных нейронных сетей [7], [8] совместно с показавшим эффективную работу алгоритмом локализации.

Настоящая статья описывает обновленную сквозную технологию детектирования и распознавания дорожных знаков в реальном времени. Созданная система использует скорость, получаемую от транспортного средства. Это позволяет прогнозировать не только присутствие объекта, но также масштаб и его точные координаты на соседнем кадре. Таким образом, точность обнаружения повышается, в то время как вычислительная сложность остается прежней. Классификация локализованных объектов реализована с использованием сверточных нейронных сетей (Convolutional

Neurak Networks или CNN). Применение GPU позволяет осуществлять обработку кадров видеопоследовательности в режиме реального времени.

2. Обнаружение и отслеживание дорожных знаков

Разработанная технология обнаружения и различения дорожных знаков состоит из трех этапов: подготовка изображения, локализация дорожного знака и его классификация.

В процессе подготовки изображения цветное пространство HSV используется для извлечения из изображения красных и синих пикселей. Из-за ошибок матрицы видеокамеры или небольших цветных объектов на изображениях после применения порогового фильтра возникает точечный шум. Для устранения этого точечного шума применяется алгоритм, описанный в работе [4]. В работе [5] показана эффективная реализация алгоритма удаления шума на массивно-многопоточной архитектуре CUDA. Ускорение при использовании графических процессоров достигло 60–80 раз по сравнению с работой на обычном центральном процессоре. Размер кадра в видеопоследовательности 1920x1080 пикселей. Время выполнения обработки изображений на центральном процессоре 0,7-1 сек. Использование CUDA на базе NVIDIA GeForce 335M сократило время выполнения до 7–10 мс, что удовлетворяет требованиям обработки видео в режиме реального времени.

В статье [5] рассматриваются алгоритмы обнаружения и отслеживания дорожных знаков. Метод локализации, являющийся модификацией обобщенного преобразования Хафа, разрабатывался с учетом ограничения по времени выполнения обработки одного кадра видеопоследовательности. Разработанный алгоритм показывает эффективные результаты и хорошо работает с предварительно обработанными изображениями. Отслеживание с использованием значения текущей скорости автомобиля позволило улучшить производительность системы, так как область поиска на соседних кадрах может быть значительно уменьшена. Кроме того, наличие знака в последовательности смежных кадров в предсказанных областях значительно повышает уверенность в правильности обнаружения. Классификация локализованных дорожных знаков гарантирует, что вся процедура обнаружения дорожных знаков успешно.

3. Классификация обнаруженных объектов

3.1. Сверточные нейронные сети

В задачах классификации и распознавании образов достаточно успешно используются нейронные сети. Нейронная сеть представляет собой математическую модель, построенную на основе соединяющихся между собой искусственных нейронов, по принципу организации биологических нейронных сетей. На вход сети подается вектор признаков, который, последовательно проходя через слои сети, перерабатывается в признаки более высоких уровней. Количество выходов такой сети равно количеству классов, к которым могут относиться входные векторы, и на каждом выходе будет формироваться вероятность принадлежности объекта к соответствующему классу.

Искусственный нейрон представляет собой взвешенный сумматор, выход которого можно описать следующей формулой:

$$a_j^i = \sigma(\sum_k a_k^{i-1} w_k^{ij}), \quad (1)$$

где a_j^i – j-ый нейрон на слое i, w_k^{ij} – вес синапса, соединяющего j-ый нейрон на слое i и k-ый нейрон на слое i-1. В качестве функции активации использовалась логистическая кривая, которая переключалась из логистической регрессии. Кроме того, стоит заметить, что отдельный нейрон выполняет функцию логистической регрессии.

На этапе обучения сети, обучающие данные передаются пакетами. Затем строится функция потерь и производится ее минимизация с помощью методов, основанных на градиентном спуске, общее название которых – обратное распространение ошибки. В качестве функции потерь в задачах классификации чаще всего используют перекрестную энтропию:

$$H(p, q) = -\sum_i Y(i) \log y(i). \quad (2)$$

В сетях с большим количеством и сигмоидной активационной функцией существует проблема под названием затухание градиентов. Для ее решения, вместо сигмоиды в качестве функции активации берут функцию ReLU:

$$ReLU(x) = \begin{cases} 0, & x \leq 0 \\ x, & \text{иначе} \end{cases} \quad (3)$$

На данный момент, самым лучшим методом классификации изображений являются сверточные нейронные сети. В отличие от традиционной нейронной сети, которой на вход подается одномерный вектор признаков, сверточной сети на вход подается двумерное изображение, которое последовательно обрабатывается сверточными слоями.

Суть сверточного слоя – применение к изображению набора фильтров-сверток для получения карты признаков. Так как свертки применяются к разным фрагментам изображения, то признаки, которые эти свертки обнаруживают, не

обязаны находиться в какой-то конкретной области изображения. Ядра сверток подбираются во время обучения. На рис. 2 продемонстрирован результат свертки входного изображения с 4мя ядрами.



Рис. 2. Свертка входного изображения.

3.2. Предлагаемая реализация

Для решения поставленной задачи использовалась библиотека глубокого обучения TensorFlow. Обучение и тестирование производилось на наборе данных The German Traffic Sign Recognition Benchmark [6]. В первой версии системы распознаются 16 наиболее популярных типов дорожных знаков.

В таблице 1 представлена архитектура сети. Сеть представляет собой последовательность соединенных между собой слоев, начинающуюся со сверточного слоя и заканчивающейся слоем softmax. Параметр сверточного слоя – шаг – определяет шаг скользящего окна свертки. В случае, если шаг больше 1, сверточный слой будет совмещать в себе кроме свертки еще и операцию пулинга. Слой softmax выполняет нормализацию результатов предыдущего слоя таким образом, что на его выходе будут формироваться вероятности отношения объекта к рассматриваемым классам.

Таблица 1. Архитектура нейронной сети

Слой
Сверточный с шагом 2 и ядром 7x7x4
Сверточный с шагом 2 и ядром 5x5x8
Сверточный с шагом 2 и ядром 3x3x16
Сверточный с шагом 2 и ядром 3x3x32
Сверточный с шагом 1 и ядром 2x2x16
Сверточный с шагом 1 и ядром 2x2x8
Сверточный с шагом 1 и ядром 2x2x4
Полносвязный с 64 выходами
Полносвязный с 16 выходами
Слой softmax

Модель можно условно поделить на 2 блока – сверточный блок и полносвязный блок. В состав фреймворка TensorFlow входят инструменты для отображения сети, которые позволяют визуализировать модель на разных уровнях абстракции, вплоть до низкоуровневых математических операций. На рис. 3 представлено графическое изображение сети, созданное с помощью инструментов TensorFlow.

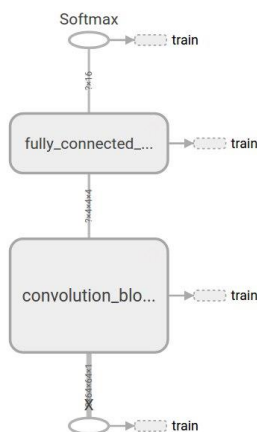


Рис. 3. Граф вычислений TensorFlow.

Для обучения модели исходный набор данных был случайно разделен на обучающую и тестовую выборки в соотношении 80 к 20. Во время обучения за одну итерацию сеть обрабатывала пачку из 50 изображений из обучающей выборки. Каждые 100 итераций подсчитывалась промежуточная точность классификации на пачке из 50 случайных изображений тестовой выборки. После завершения обучения точность была подсчитана на всей тестовой выборке. На

рис. 4 продемонстрирован рост точности классификации со временем. Уже на 2000 итерации сеть достигает точности более 0.90, а остальное время происходит более точная подстройка весов.

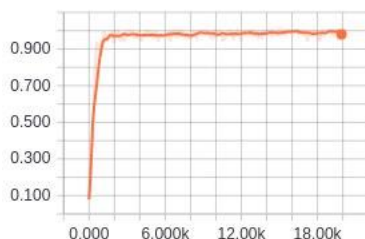


Рис. 4. Изменение точности классификации со временем.

4. Экспериментальные исследования

Эксперимент по классификации дорожных знаков проводился на трех вычислительных устройствах – видеокартах Nvidia GeForce GTX 650, Nvidia GeForce GT 650M и процессоре Intel Core i7. Оценивалось время обучения сети и время обработки отдельного кадра обученной сетью. Результаты представлены в таблице 2.

Таблица 2. Время обучение нейронной сети и время обработки одного объекта

Устройство	Обучение	Обработка одного кадра (64x64)
Nvidia GeForce GTX 650	7 мин	0.05 мс
Nvidia GeForce GT 650M	12 мин	0.14 мс
Intel Core i7	16 мин	0.37 мс

Для того, чтобы оценить точность алгоритма обнаружения, была использована немецкая база данных аннотированных изображений, содержащих дорожные знаки [6]. Она содержит более 50 000 изображений с дорожными знаками, зарегистрированными в различных условиях. Для оценки качества обнаружения было посчитано количество изображений с правильно локализованными и классифицированными дорожными знаками. При тестировании разработанных алгоритмов использовались только 9987 изображений, содержащих дорожные знаки требуемой формы и с красными контурами. Эксперименты показали 99,94% правильно локализованных и классифицированных запрещающих и предупреждающих дорожных знаков. В таблице 3 представлены результаты точности и быстродействия работы алгоритмов из [5], [6] и метода, описанного в настоящей статье.

Таблица 3. Показатели точности и быстродействия алгоритмов обнаружения и различения дорожных знаков

Алгоритм	Точность	Время обработки кадров в секунду
Обработка скользящим окном + метод опорных векторов	100 %	1
Модифицированный метод ОПХ с предобработкой + CNN	99,94 %	50
Модифицированный метод ОПХ с предобработкой	97,3 %	43
Модифицированный метод ОПХ без предобработки	89,3 %	25
Метод Виолы-Джонса	90,81 %	15
Метод гистограмм ориентированных градиентов (HOG)	70,33 %	20

Точность всех алгоритмов, указанных в таблице, была получена с использованием аннотированной базы данных изображений GTSDDB [6]. Алгоритм обработки скользящим окном [9] демонстрирует наилучший результат по точности обнаружения и классификации дорожных знаков – 100 %. Однако, наилучшим результатом по скорости работы является описанный выше модифицированный метод ОПХ с использованием в качестве метода классификации сверточной нейронной сети.

Реализация разработанного метода была протестирована на видеокдрах, полученных на улицах города Самары с использованием камеры GoPro Hero 3 Black Edition, установленной на автомобиле. На рис. 5 приведены фрагменты исходных изображений с отмеченными дорожных знаков на них. На рисунках видны отмеченные квадратным

контуром верно локализованные дорожные знаки. В верхнем левом углу каждого изображения представлен результат различения обнаруженного объекта.

Среднее время обработки одного кадра при показателе 50 кадра в секунду – около 20 мс. Из них в среднем 19 мс – это время, затрачиваемое на этапы предобработки и обнаружения дорожного знака. При ограничении области поиска на последующих кадрах время локализации снижается до 10-15 мс.



Рис. 5. Примеры обнаружения дорожных знаков в видеопоследовательностях.



Рис. 6. а) Изображение с отмеченной областью поиска дорожного знака на следующем кадре;
б) Следующий кадр с отмеченной областью поиска других не отслеживаемых дорожных знаков.

Для того, чтобы не пропускать дорожные знаки, которые ещё не были обнаружены, поиск осуществляется не только в рассчитанной области интереса (рис. 6а), но и в той части изображения, которая не пересекается с отслеживаемой областью (рис. 6б).

Помимо снижения вычислительной сложности, отслеживание обнаруженных объектов в видеопоследовательности добавляет уверенности в верной классификации. То есть при обнаружении объекта такого же класса в рассчитанной области интереса на следующем кадре подтверждается предположение о наличии на текущем участке дорожного знака определённого класса.

5. Заключение

В статье описывается обновленная сквозная технология обнаружения и классификации дорожных знаков. Метод локализации, являющийся модификацией обобщенного преобразования Хафа, разрабатывался с учетом ограничения по времени выполнения обработки одного кадра видеопоследовательности. Разработанный алгоритм показывает эффективные результаты и хорошо работает с предварительно обработанными изображениями. Отслеживание с использованием значения текущей скорости автомобиля позволило улучшить производительность системы, так как область поиска на соседних кадрах может быть значительно уменьшена. Кроме того, наличие знака в последовательности смежных кадров в предсказанных областях значительно повышает уверенность в правильности обнаружения. Классификация реализована при помощи сверточной нейронной сети. Была использована реализация из библиотеки TensorFlow.

Тестирование описанных алгоритмов на видеопоследовательностях, полученных при съёмке реальных дорожных знаков, подтвердило возможность обрабатывать видеопоток высокого разрешения в реальном времени – 50 кадров в секунду. Дальность обнаружения дорожных знаков при использовании FullHD видеокамеры достигает 50 м.

Для ускорения выполнения описанных методов были использованы графические процессоры и технология массивно-многопоточного программирования CUDA.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 16-37-60106 мол_а_дк и 16-37-00362 мол_а.

Литература

- [1] Shneier, M. Road sign detection and recognition // Proc. IEEE Computer Society Int. Conf. on Computer Vision and Pattern Recognition. – 2005 – P. 215–222.
- [2] Ruta. A. A New Approach for In-Vehicle Camera Traffic Sign Detection and Recognition / A. Ruta, F. Porikli, Y. Li, S. Watanabe, H. Kage, K. Sumi // IAPR Conference on Machine Vision Applications (MVA), Session 15: Machine Vision for Transportation. – 2005.
- [3] Belaroussi, R. Road Sign Detection in Images / R. Belaroussi, P. Foucher, J.P. Tarel, B. Soheilian, P. Charbonnier, N. Paparoditis // A Case Study, 20th International Conference on Pattern Recognition (ICPR). – 2010. – P. 484-488.
- [4] Фурсов, В.А. Локализация контуров объектов на изображениях при вариациях масштаба с использованием преобразования Хафа / С.А. Бибииков, В.А. Фурсов, П.Ю. Якимов // Компьютерная оптика. – 2013. – Т. 37, № 4. – С. 502-508.
- [5] Якимов, П.Ю. Отслеживание дорожных знаков в видеопоследовательности с использованием скорости автомобиля / П.Ю. Якимов // Компьютерная оптика. – 2015. – Т. 39, № 5. – С. 795-800.
- [6] Houben, S. Detection of Traffic Signs in Real-World Images: The German Traffic Sign Detection Benchmark / S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, C. Igel // International Joint Conference on Neural Networks. – 2013.
- [7] Zhu, Z. Traffic-Sign Detection and Classification in the Wild / Zhe Zhu, Dun Liang, Songhai Zhang, Xiaolei Huang, Baoli Li, Shimin Hu // proceedings of CVPR. – 2016. – P. 2110-2118.
- [8] LeCun, Y. Traffic Sign Recognition with Multi-Scale Convolutional Networks / Yan LeCun, Pierre Sermanet // Proceedings of International Joint Conference on Neural Networks (IJCNN'11). – 2011.
- [9] Mathias, M. Traffic sign recognition - how far are we from the solution? / M. Mathias, R. Timofte, R. Benenson, L. V. Gool // Proceedings of IEEE International Joint Conference on Neural Networks. – 2013. – P. 1-8. – ISSN 2161-4393.