

Классификация дорожных знаков при помощи YOLO CNN на мобильной платформе NVIDIA Jetson

Н.С. Артамонов¹, П.Ю. Якимов¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Классификация объектов в видеопотоке с помощью глубинного обучения завоевала огромную популярность в современном мире. К большинству систем, решающих задачу классификации, предъявляется требование мобильности. В настоящей статье предложена реализация сверточной нейронной сети YOLO CNN (You Only Look Once) для решения задачи классификации дорожных знаков на мобильной платформе NVIDIA Jetson. Особенностью данной платформы является наличие мобильного графического процессора NVIDIA Tegra, который позволяет производить высокопроизводительные вычисления при низком энергопотреблении. Реализованный алгоритм нейронной сети YOLO CNN позволяет решать задачу классификации дорожных знаков в непрерывном видеопотоке с достойной точностью и скоростью, а платформа NVIDIA Jetson обеспечивает мобильность системы.

1. Введение

В настоящее время завоёвывают всё большую популярность системы распознавания образов, интегрированные в различные мобильные объекты. К таким объектам относятся различные летательные и наземные изменяющие положение в пространстве аппараты (такие как квадрокоптеры или автомобили). Интегрированные в автомобили системы решают самые разные задачи, но основной задачей таких систем является распознавание и классификация дорожных знаков. Одна из главных проблем – найти подходящее аппаратное обеспечение с низким энергопотреблением, но достаточной вычислительной мощностью для осуществления распознавания в режиме реального времени. Программное обеспечение играет не меньшую роль, потому что необходимо извлечь максимальную производительность из тех немногих ресурсов, что мобильная система в состоянии предоставить.

Решения в области распознавания изображений в режиме реального времени делятся на два основных семейства: Region Proposes (поочередно предлагаются и классифицируются регионы кадра) и Single Shot (на полученном изображении сразу детектируются все объекты). К первому семейству относятся такие нейронные сети, как R-CNN [4], Fast R-CNN [4], Faster R-CNN [4]. Ко второму же семейству относятся YOLO CNN [2], SSD [6]. Нейронные сети, использующие распознавание по регионам, имеют достаточно медленное время распознавания при качественном определении объектов. Однако для мобильных платформ более подходят Single Shot CNN, так как они более быстрые.

Одним из популярных аппаратных решений для распознавания образов в случаях, когда требуется низкое энергопотребление, является мобильная аппаратная платформа NVIDIA Jetson на базе процессора NVIDIA Tegra. NVIDIA Tegra – это семейство систем на кристалле

(SoC – System On Chip), объединяющих в себе ARM-процессор, графический процессор, медиа- и DSP-процессоры, контроллеры памяти и периферийных устройств, имея при этом низкое энергопотребление. Мобильная платформа NVIDIA Jetson TX2 в данный момент применяется компанией «Росэлэлектроника» в системах видео фиксации нарушений ПДД [8]. На рисунке 1 представлено изображение платформы NVIDIA Jetson TK1, которая работает на базе процессора Tegra K1.



Рисунок 1. NVIDIA Jetson TK1.

Настоящая статья описывает процесс установки, обучения и использования нейронной сети YOLO CNN на мобильной платформе NVIDIA Jetson TK1, а также анализ производительности этой нейронной сети на данном устройстве. Применение GPU на платформе, а также библиотек CUDA и OpenCV [5] позволяют задействовать всю вычислительную мощность системы.

2. CNN для детектирования объектов в реальном времени

2.1. Сверточные нейронные сети

Распознавание и классификация изображений достаточно успешно осуществляются с помощью сверточных нейронных сетей. Нейронная сеть – это некоторая математическая модель, которая состоит из соединенных между собой искусственных нейронов [1]. Сеть принимает в качестве входных данных вектор признаков, затем последовательно пропускает их через слои сети. На выходе получаются вероятности принадлежности объекта к заданным классам. Обычно нейронная сеть оперирует числовыми, а не символьными величинами.

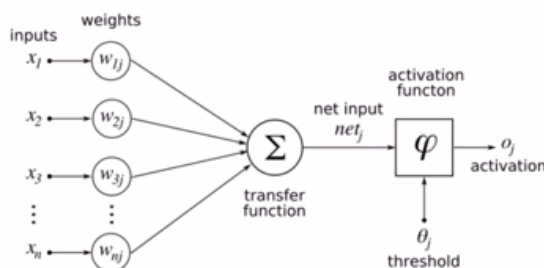


Рисунок 2. Схема искусственного нейрона.

На рисунке 2 представлена схема искусственного нейрона [7]. Он принимает на вход параметры $x_1, x_2, x_3, \dots, x_n$, которые являются либо исходными данными, либо выходными параметрами других нейронов. У каждого параметра есть вес, $w_{1j}, w_{2j}, w_{3j}, \dots, w_{nj}$, на который умножается значение параметра. Затем взвешенные параметры суммируются с помощью некоторой функции *transfer function*. Полученное значение отправляется в *activation function*, которая после вычисления результата принимает решение, передавать ли сигнал следующему нейрону. Выходное значение поступит в качестве одного из параметров в другой нейрон.

В настоящее время для обработки изображений наиболее эффективны сверточные нейронные сети. На вход свёрточной нейронной сети подается двумерное изображение,

которое затем обрабатывается свёрточными слоями. Свёрточные слои преобразуют фрагменты изображения в карту признаков.

2.2. YOLO

YOLO CNN [2] – это свёрточная нейронная сеть, позволяющая детектировать и классифицировать объекты в виде обрамляющих прямоугольников. YOLO работает по принципу Single Shot. Это значит, что архитектура сети устроена таким образом, что за один проход кадра, на нём детектируются все объекты.

На рисунке 3 представлена архитектура YOLO CNN. На вход YOLO подается трёхканальное изображение, у которого меняется размер до 448x448, над полученным изображением проводятся преобразования. Первое преобразование заключается в прогоне изображения через часть модифицированной архитектуры GoogLeNet. После этого преобразования получается feature maps размером 14x14x1024. Далее применяются две конволюции. После второй конволюции размерность уменьшается до 7x7x1024. Далее проводится ещё одна конволюция. Результат дважды прогоняется через полносвязный слой, изменяется до размерности 1470x1 и в итоге трансформируется в тензор размером 7x7x30. К полученному тензору применяется процедура детектирования, на выходе которой получается результирующее детектирование. Тензор представляет собой отображение сетки 7x7 на изображении. 30 значений несут информацию о ячейке: 10 значений для двух возможных рамок, 20 значений отношения к каждому из 20 доступных классов. Вся эта информация фильтруется, отфильтрованные данные отображаются.

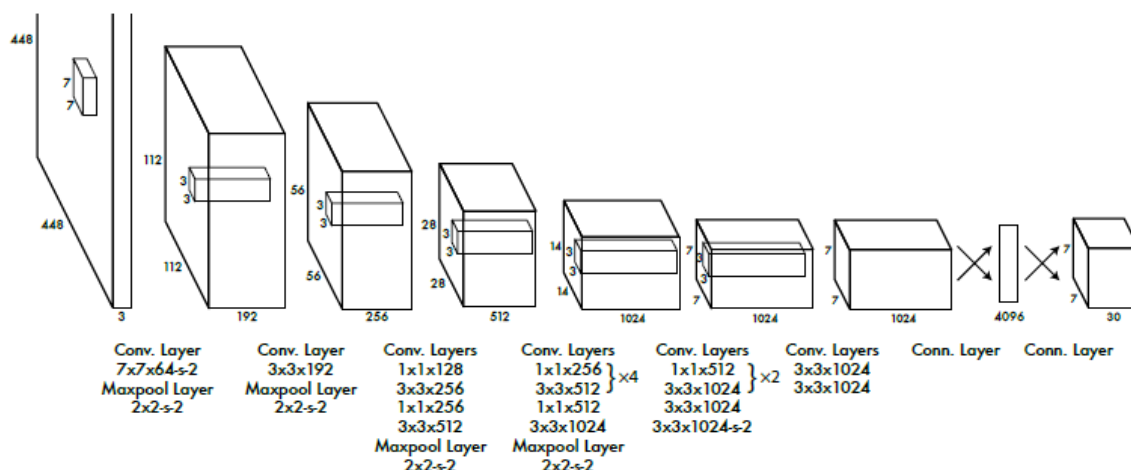


Рисунок 3. Архитектура YOLO CNN.

3. Реализация работы YOLO на NVIDIA Jetson

NVIDIA Jetson – это целое семейство мобильных платформ на базе процессоров NVIDIA Tegra. На сегодняшний день существуют три модели:

- NVIDIA Jetson TK1;
- NVIDIA Jetson TX1;
- NVIDIA Jetson TX2.

В таблице 1 приведены основные характеристики этих моделей.

В данной статье представлена реализация YOLO CNN на платформе NVIDIA Jetson TK1. Перед тем, как начинать работу с платформой, необходимо установить на неё всё необходимое программное обеспечение. Выполняется эта процедура с помощью Jetpack JDK – официальной утилиты NVIDIA. Для работы с нейронными сетями и распознаванием изображений на платформе предоставляются адаптированные под аппаратное обеспечение версии фреймворков CUDA 6.5, CUDNNv2, OpenCV.

Таблица 1. Основные характеристики платформ NVIDIA Jetson.

	Jetson TK1	Jetson TX1	Jetson TX2
GPU	NVIDIA Kepler GPU с 192 ядрами CUDA	NVIDIA Maxwell™, 256 ядер CUDA	NVIDIA Pascal™, 256 ядер CUDA
Процессор	Четырехъядерный процессор ARM® Cortex™-A15 с архитектурой NVIDIA 4-Plus-1™	Четырехъядерный ARM® A57/2 МБ L2	HMP Dual Denver 2/2 МБ L2 + Четырехъядерный ARM® A57/2 МБ L2
Память	2 ГБ памяти x16, 64-битная шина	4 ГБ памяти LPDDR4, 64-bit 25,6 Гбит/с	8 ГБ памяти LPDDR4, 128-bit 59,7 Гб/с

В YOLO CNN предусмотрена возможность компиляции и запуска сети с помощью CPU, GPU, CUDA, CUDNN, OpenCV, OpenMP и их комбинаций. В ходе запуска YOLO на Jetson TK1 выявился ряд проблем:

- Платформа поддерживает слишком старую версию CUDNN.
- Платформа не поддерживает OpenMP.
- Малый объем оперативной памяти необходимо компенсировать swar-файлом (1ГБ).

Все эти проблемы отрицательно сказываются на производительности нейронной сети. Тем не менее, компиляция с комбинацией GPU, CUDA, OpenCV проходит успешно. На рисунке 4 представлен результат работы YOLO на тестовом изображении.

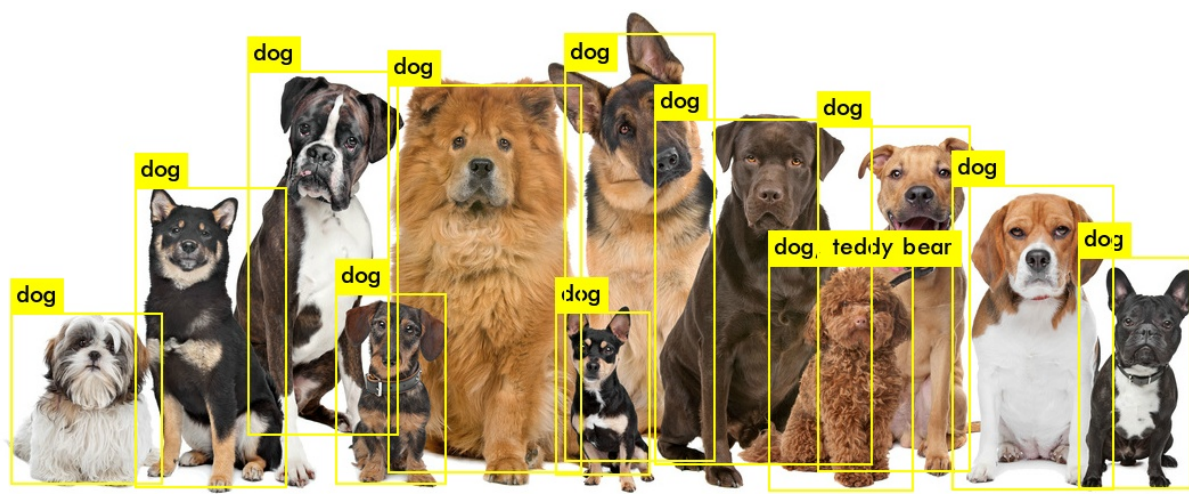


Рисунок 4. Результат работы YOLO.

Для запуска YOLO необходимы файл конфигурации (пример yolo.cfg), файл весов для всех необходимых классов (пример yolo.weights) и медиа-файл (изображение или видео, также есть возможность запуска видеопотока с веб-камеры). [3] Пример команды:

```
./darknet detect cfg/yolo.cfg yolo.weights data/dogs.png
```

Нейронная сеть имеет большую базу уже обученных классов, на которых можно протестировать все возможности YOLO, а также производительность аппаратного обеспечения. Но нейронную сеть можно обучить любому классу изображений, если грамотно подобрать исходные данные. Для обучения YOLO необходимо большое количество изображений. К каждому изображению должен прилагаться текстовый файл с размеченными регионами тренируемого класса объектов. Также необходим файл с начальными весами и файл с системной информацией, в котором указаны пути к изображениям и путь, по которому будут записываться точки восстановления. Точки восстановления представляют из себя файлы весов на определенном шаге обучения. Файлы весов записываются в постоянную память каждые 100

итераций, что позволяет в любое время прервать обучение, а затем продолжить с последним полученным весом. Чем дольше учится сеть, тем качественнее будет детектирование. Команда для обучения YOLO выглядит следующим образом:

```
./darknet detector train cfg/voc.data cfg/yolo-voc.cfg darknet.conv.23
```

Для обучения нейронной сети YOLO не обязательно использовать мобильную платформу. Для этой цели больше подходит вычислитель с мощным GPU. Таким образом, платформа Jetson будет использована для непосредственно применения обученной нейронной сети.

Следует также заметить, что производительность NVIDIA Jetson TX1 и TX2 значительно превышает NVIDIA Jetson TK1, использованной при проведении настоящего исследования. Однако, большая программная совместимость разных поколений аппаратного обеспечения продуктов NVIDIA позволяет отработать методы и алгоритмы на имеющемся в наличии оборудовании с целью последующего запуска на более новых платформах.

4. Экспериментальные исследования

Для решения задачи классификации дорожных знаков был подготовлен dataset из 130 изображений дорожного знака «Пешеходный переход». Нейронная сеть совершила 900 итераций обучения в течении 60 минут, затем полученные веса были использованы для детектирования изображения знака. На рисунке 5 представлен результат детектирования. Обучение прошло удачно, не смотря на малое количество исходных данных и итераций при обучении.



Рисунок 5. Результат детектирования знака «Пешеходный переход».

В таблице 2 приведены результаты работы YOLO на NVIDIA Jetson TK1 и на ПК с GPU NVIDIA GeForce GTX 1050ti.

Таблица 2. Результаты работы YOLO.

	NVIDIA Jetson TK1	NVIDIA GeForce GTX 1050ti
Скорость распознавания одного изображения (с)	1.0000	0.0034
Количество кадров в секунду при обработке видео (fps)	1.5	20.0

5. Заключение

В статье описывается реализация работы нейронной сети YOLO CNN на мобильной платформе NVIDIA Jetson TK1. Нейронная сеть успешно запускается на платформе, её возможно обучить

для дальнейшего использования в решении задачи классификации дорожных знаков. Jetson TK1 работает с нейронной сетью медленно, что вряд ли подходит для задачи детектирования объектов в непрерывном видеопотоке, но сама возможность запуска YOLO открывает перспективы для новых поколений платформы, например Jetson TX2, который значительно превышает по характеристикам первую версию. Главной же особенностью платформ NVIDIA Jetson остается низкое энергопотребление.

Нейронная сеть YOLO легко обучилась на небольшом количестве данных за короткое время и выдала приемлемый результат.

6. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-60106 мол_а_дк.

7. Литература

- [1] LeCun, Y. Traffic Sign Recognition with Multi-Scale Convolutional Networks / Y. LeCun, P. Sermanet // Proceedings of International Joint Conference on Neural Networks (IJCNN'11), 2011.
- [2] Redmon J. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // You Look Only Once, 2015. – 10 p.
- [3] YOLO: Real-Time Object Detection [Electronic resource]. – Access mode: <https://pjreddie.com/darknet/yolo/> (01.11.2017).
- [4] Ren, S. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / S. Ren, K. He, R. Girshick, J. Sun. – 2016. – 14 p.
- [5] Библиотека обработки изображений OpenCV [Электронный ресурс]. – Режим доступа: <http://opencv.org> (04.11.2017).
- [6] Liu, W. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A. Berg. – 2016. – 15 p.
- [7] Заенцев, Е.В. Нейронные сети: основные модели / Е. В. Заенцев. – Воронеж, 1999. – 76 с.
- [8] 3D News. Daily Digital Digest [Electronic resource]. – Access mode: <https://3dnews.ru/959458> (01.11.2017).

Classification of traffic signs with YOLO CNN using NVIDIA Jetson mobile platform

N.S. Artamonov¹, P.Y. Yakimov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. Classification of objects in the video stream with the help of deep learning has gained immense popularity in the modern world. To the majority of systems solving the classification problem, mobility is required. This article proposes the implementation of the YOLO CNN (You Only Look Once) convolutional neural network to solve the problem of the classification of road signs on the mobile platform NVIDIA Jetson. A particular feature of this platform is the availability of mobile graphics processor NVIDIA Tegra, which allows high-performance computing with low power consumption. The implemented algorithm of the YOLO CNN neural network allows solving the problem of the classification of road signs in a continuous video stream with decent accuracy and speed, and the NVIDIA Jetson platform provides mobility of the system.

Keywords: NVIDIA Jetson, YOLO CNN, convolutional neural networks, classification of road signs, image processing, computer vision system, mobile graphics processors.