

Классификация автодорожных знаков на основе свёрточной нейросети, обученной на синтетических данных

Моисеев Борис, Чигорин Александр

Факультет вычислительной математики и кибернетики

Московский Государственный Университет имени М.В.Ломоносова, Москва, Россия

mbd_mezhgorie@mail.ru

Аннотация

В данной статье представляется метод генерации синтетических изображений автодорожных знаков для машинного обучения свёрточной искусственной нейронной сети. За счёт этого метода можно получить большую тренировочную базу, не прилагая усилий для фотосъёмки и разметки полученных фотографий. Проанализированы различные приёмы, позволяющие получить наиболее реалистичные изображения. Также в данной работе рассматривается устройство свёрточной нейросети и анализируется эффективность её применения к задаче классификации дорожных знаков.

Ключевые слова: синтетические данные, свёрточная нейросеть, машинное обучение.

1. ВВЕДЕНИЕ

Системы распознавания автодорожных знаков имеют множество различных применений, таких как системы помощи водителю (ADAS - Advanced Driver Assistance Systems), управление автодорожной инфраструктурой и создание навигационных карт. На данный момент системы распознавания знаков включены в комплектацию автомобилей различных марок, но большинство из них имеют ряд существенных ограничений, сужающих спектр их применения; например, большинство систем способно распознавать лишь знаки ограничения скорости.

Одними из основных трудностей, с которыми сталкивается разработчик детектора и классификатора автодорожных знаков, являются жесткие ограничения на ресурсы, потребляемые программой. Бортовой автомобильный компьютер не может обладать большой вычислительной мощностью, равно как и большим объёмом оперативной памяти. Обе эти проблемы успешно решаются с помощью свёрточных нейросетей. Благодаря принципу разделяемых весов классификаторы на их основе не требуют большого объёма оперативной памяти, равно как и существенных вычислительных мощностей. Кроме того, как обучение, так и использование нейросети легко поддаётся распараллеливанию. Свёрточная нейросеть также может применяться и как детектор на основе скользящего окна, это применение оставим за рамками данной работы.

Однако при использовании нейросети разработчик может столкнуться с другой проблемой: качество классификации существенно зависит от объёма и репрезентативности тренировочной выборки. Для обучения эффективного классификатора необходимо иметь базу фотографий, содержащую знаки всех рассматриваемых классов в различных условиях освещения, снятых под различными углами и с различным зашумлением, возникающем при съёмке на камеру низкого разрешения на высокой скорости. Получение такой базы стоит существенных затрат времени и ресурсов.

Для устранения этой трудности предлагается метод генерации синтетических изображений. Используя пиктограммы знаков и фотографии природного и

городского пейзажа, с его помощью можно создать набор изображений, максимально приближенных к реальным фотографиям автодорожных знаков.

2. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

В работе [6] приводится сравнение результатов классификации при обучении на синтетических и реальных данных. Из графиков видно, что результаты классификатора, обученного на синтетических данных, существенно уступают результатам классификатора, обученного на реальных данных. Целью же этой работы является разработка алгоритма генерации синтетических изображений, таких, что обученный на них классификатор, пусть и на более массивной выборке, окажется не менее эффективным, чем классификатор, обученный на реальных данных.

3. СВЁРТОЧНАЯ НЕЙРОСЕТЬ

В качестве классификатора, с помощью которого будет проводиться оценка данного метода была выбрана реализация свёрточной нейросети [4]. Её преимуществами является высокая скорость работы за счет распараллеливания на графическом процессоре с архитектурой CUDA (Compute Unified Device Architecture), а также возможность легко адаптировать её к поставленной задаче.

Идея свёрточной нейросети основывается на трёх основных принципах:

- Локальные поля восприятия
- Разделяемые веса
- Пространственная субдискретизация

Эти принципы реализуются следующим образом. Нейросеть состоит из двух основных типов слоёв: свёрточных и субдискретизирующих.

Каждый нейрон свёрточного слоя соединён лишь с некоторой областью входной карты признаков (либо входного изображения, если этот слой — первый). При этом все нейроны этого слоя имеют одни и те же веса (так реализуется принцип разделяемых весов). Таким образом, карта признаков, полученная на выходе из свёрточного слоя, является, по сути, свёрткой входной карты с некоторым фильтром, коэффициенты которого и подвергаются обучению. Слой может иметь несколько таких фильтров, соответственно на выходе будет получено несколько карт признаков. Обучение может осуществляться методом обратного распространения ошибки или его модификациями, подробнее об этом можно прочитать в работе [5].

Каждый нейрон субдискретизирующего слоя также соединён лишь с некоторой областью входной карты признаков, но эти области обычно не пересекаются и нейроны не имеют весов. Вместо этого в нейроне вычисляется некая функция значения входов, часто используется максимум среди входных значений. В результате получается карта признаков меньшей

размерности, т.к. некоторая область входной карты заменяется одним значением.

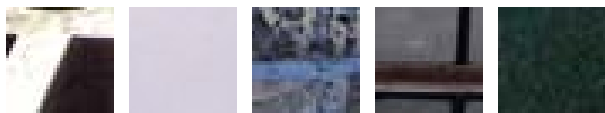
Свёрточные и субдискретизирующие слои в свёрточной нейросети чередуются, пока не будет получен набор карт признаков достаточно маленькой размерности, после чего все значения подаются на вход полносвязной одно- или двухслойной нейросети.

В результате нейросеть имеет сравнительно мало весов (свёрточный слой имеет число весов, пропорциональное числу обучаемых фильтров, субдискретизирующий слой не имеет весов), что уменьшает занимаемый объём памяти. Благодаря принципу локальных полей восприятия каждый нейрон соединён лишь с некоторой областью входной карты признаков, благодаря чему существенно повышается скорость обучения. Пространственная субдискретизация позволяет уменьшать размерность карт признаков без потери важной информации, а также обеспечивать инвариантность к масштабу классифицируемых изображений.

На вход классификатору подавались цветные изображения размером 32x32 пикселя в формате RGB, нормализованные с помощью алгоритма CLAHE (Contrast Limited Adaptive Histogram Equalization).

4. ГЕНЕРАЦИЯ СИНТЕТИЧЕСКИХ ИЗОБРАЖЕНИЙ

4.1 Экземпляры фона



Для получения примеров фона были использованы фотографии разрешения около 1000x1000 пикселей, из которых в случайных местах вырезались образцы 32x32. Фотографии мы брали разнообразные, сделанные в разное время года, в разное время суток и при различных погодных условиях, пейзажные и фотографии города.

4.2 Пиктограммы знаков



Пиктограммы знаков были обработаны таким образом, чтобы вся часть изображения вне знака была прозрачной, что было реализовано с помощью альфа-канала на изображениях в формате RGBA. Впоследствии значение альфа-канала используется в качестве маски для наложения на фон и блендинга. Кроме того, были добавлены некоторые пиктограммы, не соответствующие стандарту, но встречающиеся в реальных условиях.

4.3 Яркость и контраст



Яркость и контраст изображений меняется за счёт сужения гистограммы яркости пикселей. При этом смещается либо только левая граница («засвеченный» знак), либо только правая (затемнённый знак). Меняется только яркость знака, т.к. не всегда яркость знака и фона совпадают. Тем самым, мы одновременно и снижаем контраст, и изменяем

яркость изображений. Степень сжатия гистограммы принимает произвольное значение от 0% до 90%.

4.4 Изменение размера



Размер пиктограммы изменялся в пределах от 90% до 110%, что приблизительно соответствует точности выделения описываемого квадрата для знаков в базе GTSRB. Но кроме этого, в базе есть изображения, на которых границы выделены не совсем точно, т.е. знаки, которые сильно меньше, чем описанный вокруг них квадрат. Для таких знаков были добавлены примеры (не более 10% от общего числа синтетических изображений) с пиктограммой, уменьшенной в пределах от 60% до 90%.

4.5 Поворот



Поворот пиктограммы осуществляется на случайный угол, не превосходящий 30 градусов по часовой или против часовой стрелки. Ограничение на допустимый угол поворота накладывает тот факт, что некоторые знаки при повороте выглядят одинаково (например, знаки со стрелками).

4.6 Положение пиктограммы на изображении

Положение пиктограммы выбирается случайным образом, но ввиду того, что в большинстве случаев знак расположен в центре описанного квадрата, из двух случайно сгенерированных положений пиктограммы берётся среднее значение. Таким образом обеспечивается распределение, приближающееся к нормальному, при сохранении высокой производительности вычислений.

4.7 Проективные преобразования



Так как автодорожный знак со стороны проезжающего автомобиля виден обычно под некоторым углом, к генерируемым изображениям необходимо применить проективные преобразования. Ввиду различных условий установки знака и различных его положений относительно дороги и автомобиля, проезжающего по ней, бесполезно пытаться точно посчитать конкретные проективные преобразования, более эффективно попытаться покрыть весь класс таких преобразований (разумеется, не слишком искажающих изображение). Искомое проективное преобразование рассчитывалось как преобразование, ставящее в соответствие вершины двух четырёхугольников. В качестве первого четырёхугольника использовался квадрат, соответствующий знаку, т.е. со сторонами 32x32. Второй четырёхугольник получался из первого смещением вершин на случайное значение, не превосходящее 10% от размеров изображения.

4.8 Интерполяция

Важным моментом при проведении различного рода геометрических преобразований является интерполяция. При неаккуратной обработке на границах пиктограммы могут возникать артефакты, вызванные смешиванием

пиктограммы с окружающими её пустыми пикселями. Чтобы этого избежать, ещё перед всеми преобразованиями пиктограмма накладывается на фон. В результате на границах пиктограммы получается своего рода блендинг за счёт смешивания её с фоном. Интерполяция во всех описанных выше преобразованиях использовалась билинейная, как достаточно быстрая. Кроме того, ввиду низкого разрешения изображения кубическая интерполяция не сможет обеспечить существенного прироста в качестве. Маска прозрачности (альфа-канал) была приведена к формату черно-белого изображения и подвергалась таким же преобразованиям, как и пиктограмма, с использованием также билинейной интерполяции. Такой подход сразу обеспечивает необходимое размытие маски для этапа блендинга.

4.9 Блендинг

Наложение пиктограммы на фон осуществляется следующим образом:

- Маска переводится в представление с плавающей точкой;
- Маска размывается гауссовым фильтром (использовался фильтр размера 5x5 с параметром сигма, равным 0.5)
- Матрица пикселей пиктограммы умножается поэлементно на матрицу маски, а матрица фона — на матрицу из единиц минус маска.
- Полученные матрицы складываются

4.10 Точечный блик

В качестве одного из факторов, способных усложнить классификацию знака, было рассмотрено неравномерное освещение. В частности, на некоторых фотографиях наблюдается блик, из-за которого яркость различных частей фотографии существенно меняется и нормализация не может быть применена достаточно эффективно. Чтобы обучить классификатор распознавать и такие примеры входных данных, к части генерируемых синтетических изображений (20%) добавляется блик. Реализуется он следующим образом:

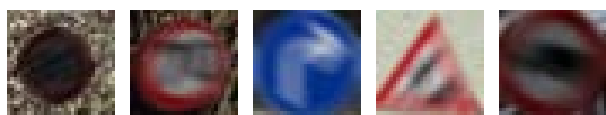


- Изображение переводится в формат YUV
- Выбирается произвольная точка
- Компонента изображения, соответствующая яркости (Y) домножается на гауссиану относительно расстояния пикселя до выбранной точки со случайно выбранным параметром сигма.

4.11 Шум

После всех действий к сгенерированному примеру применяются различные виды искажений, способные приблизить его к примерам реальных данных. Каждое из искажений применяется лишь к 20% от всех примеров.

4.11.1 «Смазанные» фотографии



Так как в большинстве случаев съёмка ведётся из движущегося автомобиля на обычную камеру, не обладающую способностью к высокоскоростной съёмке, то довольно большое количество изображений получаются смазанными. Для моделирования этого явления применяется фильтр motion, заключающийся в усреднении значений нескольких пикселей в одном направлении. Количество усредняемых пикселей не превосходит 25% от

размера знака, направление может принимать значение в пределах до 45% по часовой или против часовой стрелки

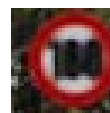
4.11.2 Фильтр Гаусса

Применяется фильтр Гаусса размером до 15% от размера изображения.



4.11.3 Морфологическое открытие/заккрытие

Осуществляется над компонентой яркости для изображения в формате YUV, моделирует эффект мокрого объектива. В качестве фильтра используется дисковый фильтр радиуса 1 (т.е. 3x3).



4.11.4 Пуассоновский и соль-перец шум

Применяется для повышения устойчивости обученного классификатора к мелким дефектам изображения.



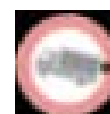
4.11.5 Уменьшение и увеличение изображения



Если при съёмке знак был достаточно удалён, то на фотографии он может занимать область, меньше чем 32x32, но несмотря на такое низкое разрешение, всё равно читаться (если не было других помех). Чтобы научить классификатор распознавать такие знаки, часть примеров уменьшаются до размера 16x16, а затем увеличиваются вновь.

4.11.6 «Битый пиксель»

Чёрный или белый квадрат накладывается в случайном месте изображения. Позволяет промоделировать различные помехи на знаке, а также избежать чрезмерной важности того или иного локального участка для классификации примера в целом.



Кроме всего перечисленного, на все примеры накладывается слабый гауссовский шум (сигма = 0.0002), иначе на части примеров могут возникнуть области в точности одного цвета.

5. ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДА И ЕГО ЧАСТЕЙ

5.1 Конфигурация сети

Используемая сеть состояла из 7 слоёв:

- Свёрточный с 64 фильтрами размера 5x5;
- Субдискретизирующий типа max-pooling (выбор максимального значения среди входов) с фильтром размера 3x3 и шагом 2;
- Свёрточный с 1 фильтром 5x5 на каждый канал;
- Субдискретизирующий типа max-pooling с фильтром размера 3x3 и шагом 2;
- Локальный слой с фильтрами 3x3;
- Локальный слой с фильтрами 3x3;
- Полносвязный слой из 43 нейронов (по числу классов).

Ранее не был рассмотрен локальный слой — слой нейросети, в котором каждый нейрон соединён с некоторой областью (в данном случае — квадрат 3x3)

входной карты признаков. Главное отличие от свёрточного слоя заключается в том, что веса эти связей — не разделяемые, и у каждого нейрона они свои.

Как уже указывалось выше, на вход нейросети подавались RGB изображения 32x32, нормализованные с помощью алгоритма CLAHE, на выходе сеть определяла его принадлежность к одному из классов.

Время обучения сети на выборке размером 240 000 изображений составляет 15-20 минут, время классификации — 4 000 примеров в секунду. Такая производительность может позволить не только проводить классификацию в режиме онлайн, но и использовать сеть в качестве детектора знаков.

5.2 Эксперименты с синтетическими данными

За счёт применения описанной выше конфигурации нейросети при обучении и тестировании данных из базы GTSRB [2] была достигнута точность 96.8%. Это не очень хороший результат по сравнению с результатами победителей соревнования GTSRB Challenge [3], но преимуществами данного подхода является сравнительная простота реализации и высокая скорость работы.

При обучении нейросети на выборке из 240 000 синтетических изображений, сгенерированных по описанному выше алгоритму, была достигнута точность классификации 97.1%. Тестирование проводилось на тестовой выборке базы GTSRB. Генерация такой выборки заняла 3.5 часа.

Генерация синтетики без использования нестандартных знаков (см. п. 4.2) привела к снижению точности на 0.2%;

Отсутствие изменения яркости изображений (см п.4.3) снизило точность на 10%;

Отсутствие колебаний размера снизило точность на 3%;

Отсутствие поворота пиктограмм — на 5.2%;

Отсутствие перспективных преобразований — на 1.8%

Без добавления точечного блика (п.4.10) сеть работает на 0,2% хуже.

Отсутствие шумов также снижает производительность, на 2.3%.

Таким образом, применение всех преобразований оправдано и приводит к повышению качества классификации.

6. ЗАКЛЮЧЕНИЕ

В результате работы была продемонстрирована как эффективность применения свёрточных нейросетей для классификации автодорожных знаков, так и продемонстрированы основные идеи генерации синтетических изображений для обучения классификатора.

В планах:

- Протестировать алгоритм на других базах данных;
- Применить свёрточную нейросеть не только для классификации, но и для детекции знаков.
- Создать более сложную структуру из свёрточных нейросетей, как, например, в работе [1]

7. ССЫЛКИ

[1] Dan Cireşan, Ueli Meier, Jonathan Masci and Jürgen Schmidhuber, "A Committee of Neural Networks for Traffic Sign Classification", IJCNN, 2011.

[2] Institut für Neuroinformatik, Dataset for The German Traffic Sign Recognition Benchmark. Available: <http://benchmark.ini.rub.de/>

[3] Institut für Neuroinformatik, The German Traffic Sign Recognition Benchmark Challenge. Available: <http://benchmark.ini.rub.de/>

[4] Alex Krizhevsky, "Convolutional Deep Belief Networks on CIFAR-10". Available: <http://code.google.com/p/cuda-convnet/>

[5] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.

[6] Paolo Medici, Claudio Caraffi, Elena Cardarelli, Pier Paolo Porta, Guido Ghisio. "Real Time Road Signs Classification." *Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety*. Columbus, OH, USA. September 22-24, 2008. pp. 253-258.

Об авторах

Моисеев Борис — студент ВМК МГУ им М.В.Ломоносова.

Чигорин Александр — аспирант ВМК МГУ им М.В.Ломоносова.