СИНТЕЗ ОБУЧАЮЩИХ ВЫБОРОК ДЛЯ КЛАССИФИКАЦИИ ДОРОЖНЫХ ЗНАКОВ С ПОМОЩЬЮ НЕЙРОСЕТЕЙ

В.И. Шахуро [1], А.С. Конушин [1,2]
[1] НИУ Высшая школа экономики, Москва, Россия,
[2] МГУ имени М.В. Ломоносова, Москва, Россия

Аннотация

В работе исследуется применимость порождающих конкурирующих нейронных сетей для синтеза обучающих выборок на примере задачи классификации дорожных знаков. Рассматриваются порождающие нейронные сети, обучаемые с помощью метрики Васерштейна. В качестве базового метода для сравнения используется метод генерации синтетических изображений дорожных знаков по иконке. Проводится экспериментальное сравнение нейросетевых классификаторов, обученных на реальных данных, двух видах синтетических данных, а также смеси реальных и синтетических данных. Эксперименты показывают, что современные порождающие нейронные сети позволяют создавать реалистичные обучающие выборки для классификации автодорожных знаков, которые превосходят по качеству методы генерации знаков по иконкам, но немного уступают по качеству реальным данным.

Ключевые слова: классификация дорожных знаков, синтетические обучающие выборки, порождающие нейронные сети.

Цитирование: **Шахуро, В.И.** Синтез обучающих выборок для классификации дорожных знаков с помощью нейросетей / В.И. Шахуро, А.С. Конушин // Компьютерная оптика. – 2018. – Т. 42, № 1. – С. 105-112. – DOI: 10.18287/2412-6179-2018-42-1-105-112.

Введение

В последние годы наблюдается стремительное развитие методов компьютерного зрения. Одним из ключевых факторов, обеспечивающих этот рост, является появление большего количества доступных обучающих выборок. Так, например, в задаче классификации изображений широкого набора классов этот рост опирается на коллекцию Imagenet, которая состоит сейчас из 14 миллионов изображений и 21 тысячи категорий, из которых обычно используется 1,5 миллиона изображений и 1 тысяча категорий для проведения соревнований [1]. Потом к этой коллекции добавилась коллекции MS COCO [2] (328 тысяч изображений с расширенной аннотацией – сегментацией индивидуальных объектов) и Cityscapes [3] (5000 изображений с плотной попиксельной разметкой изображений). Подготовка таких эталонных коллекций – очень сложная задача сама по себе, в первую очередь из-за необходимости ручной разметки. Дальнейшее развитие методов связано с развитием эталонных коллекций. При этом изменение задачи из-за появления новых объектов требует обновления коллекции, что может быть затруднительно.

Рассмотрим задачу распознавания дорожных знаков. Появление нового типа дорожных знаков требует съёмки этого знака с автомобиля в тех местах, где он появился. При этом нужно не только знать, что он появился, но и знать, где он установлен, и обеспечить доступ в эти места. Также знаков изначально может быть недостаточно для того, чтобы современные алгоритмы научились их распознавать. И это только один пример. Подобную проблему можно решить с помощью генерации синтетических обучающих выборок. Синтетические выборки позволяют как сократить затраты на сбор и разметку обучающих данных, так и решить проблему редких знаков. Широкому применению синтетических методов препятствуют два обстоятельства. Первое – высокая трудоёмкость синтеза фотореалистичных изображений. Второе – отсутствие метрики для оценки фотореализма получающихся изображений. Недавно появились модели синтеза изображений на основе порождающих конкурирующих нейросетей, которые потенциально позволяют решить обе проблемы. В этих моделях одновременно обучаются две нейронных сети: генератор и дискриминатор. Нейросеть-генератор на вход получает многомерный вектор шума из априорного распределения (например, нормальный) и преобразует этот шум в случайное изображение объекта требуемого класса. Нейросеть-дискриминатор по сути является метрикой фотореалистичности генерируемых изображений. Но обеспечивают ли эти модели качество, достаточное для обучения алгоритмов распознавания так, чтобы оно было сравнимо с обучением на реальных данных, неизвестно. Этот вопрос мы и исследуем в работе на примере задачи классификации дорожных знаков.

1. Обзор существующих работ

1.1. Использование синтетических обучающих выборок

Синтетические обучающие выборки можно применять для обучения алгоритмов машинного обучения в задачах, где получение достаточного объёма данных затруднительно или разметка слишком трудоёмка, и при этом обучение на синтетических данных позволяет достигнуть качества, сравнимого или превосходящего обучение на реальных данных. Генерация фотореалистичных изображений – сложный процесс, т.к. не существует метрики, с помощью которой можно оценить реалистичность изображения. Кроме того, в большинстве задач заранее неизвестно, какие свойства изобра-

жения необходимы для качественной обучающей выборки. Поэтому сейчас на практике используются два вида синтетических данных: дополнительные обучающие примеры, полученные из реальных путём размножения, и синтетические данные, полученные путём трёхмерного моделирования.

Первый способ, размножение данных, повсеместно используется при обучении свёрточных нейронных сетей. Нейронные сети имеют большое количество параметров, и даже самых больших выборок недостаточно для их обучения. Поэтому к изображениям применяются различные преобразования: зеркальные отражения, повороты, сдвиги, масштабирования и т.п. Такими преобразованиями можно получить только близкие к обучающей выборке синтетические изображения.

Второй способ генерации синтетических данных – трёхмерное моделирование. В нём задается параметризованная трёхмерная модель. В качестве параметров используются случайные переменные. Сэмплируются параметры модели, и с фиксированными параметрами происходит рендеринг трёхмерной модели с помощью графического движка в изображение. Этот способ имеет ограниченное применение, т.к. для каждой задачи требуется своя трёхмерная модель с фотореалистичными текстурами. Рассмотрим несколько примеров применения трёхмерного моделирования для получения синтетических обучающих выборок. В [4] синтетические карты глубины используются для обучения случайного леса для регрессии позы человека. Использование трёхмерного моделирования в задаче было возможно, т.к. для карты глубины не требуется текстура, достаточно рендерить зашумлённое облако точек модели. В [5] графический движок из коммерческой компьютерной игры используется для получения сегментированных городских сцен. Эксперименты авторов показывают, что использование синтетических изображений уменьшает требования к количеству реальных данных в три раза, однако использование только синтетических данных показывает неудовлетворительное качество. В работах [6, 7] исследуются синтетические изображения автодорожных знаков, полученных путём преобразования иконок знаков. Этот метод требует задания преобразований и их параметров, которые меняются не только в зависимости от задачи, но даже от используемой тестовой выборки. Синтетические данные также используются в задаче вычисления оптического потока с помощью нейросетей [8]. В этой работе используется нереалистичный набор данных «Летающие стулья». Однако этих данных достаточно для качественного обучения нейросети, т.к. для вычисления оптического потока нейросеть учится сопоставлять области двух изображений и реалистичность изображений не требуется.

1.2. Порождающие нейросети

Как уже было отмечено в предыдущем параграфе, не существует метрики для оценки фотореалистичности изображения. Однако в последнее время был предложен подход, позволяющий оценивать реалистичность генерируемых изображений в зависимости от задачи. В работе [9] была сформулирована новая парадигма порождающих конкурирующих нейросетей. В этой модели одна нейросеть является генератором, т.е. пытается преобразовать случайный шум в реалистичное изображение. Вторая нейросеть – дискриминатор – пытается по изображению понять, является ли оно настоящим или сгенерированным. Нейросети обучаются попеременно. Метрикой потерь при этом является бинарная кроссэнтропия. Можно показать, что минимаксная формулировка оптимизационной задачи с кроссэнтропией эквивалентна минимизации дивергенции Йенсена–Шеннона двух распределений: реального и распределения сэмплов генератора. Таким образом, в виде нейросети-дискриминатора обучается сложная метрика реалистичности изображений, которую практически невозможно задать априорно.

С момента появления данный подход активно развивался. В [10] для генерации изображений предложена свёрточная архитектура нейронной сети с транспонированными свёртками для повышения разрешения. Использование свёрточных слоев, в сравнении с полносвязными слоями, позволяет обучать нейросеть быстрее (за счёт меньшего количества параметров) и повышает качество генерируемых изображений (за счёт более простого вида фильтров, которые проще обучать). В [11] лапласовская пирамида изображений и несколько пар нейросетей генератор-дискриминатор используются для генерации изображений высокого разрешения. В [12] рассматривается задача условной генерации изображений. На вход генератору подаётся не только случайный шум, но и метка класса объекта, который необходимо сгенерировать.

Конкурирующие нейронные сети ещё практически не используются для генерации обучающих выборок, однако в работе [13] показывается, что данные, генерируемые порождающими нейронными сетями, добавленные к реальным данным, помогают улучшить качество повторной идентификации людей в видео.

2. Описание метода генерации синтетических изображений

2.1. Постановка задачи

Имеется выборка изображений одного размера $W\times H\times C$ из распределения P_r. Под P_r имеется в виду вероятностное распределение реальных изображений. Необходимо обучить нейронную сеть $g_\theta(z)$, которая на вход получает многомерный шум $z\sim p(z)$ (например, нормальный) и преобразовывает этот шум в изображения размера $W\times H\times C$, похожие на реальные. Нейросеть имеет настраиваемые параметры θ. Формально говоря, нейросеть выдаёт независимые сэмплы из некоторого распределения P_g. Реалистичности изображений, выдаваемых нейросетью, можно добиться, если распределения P_r и P_g будут близки по какой-то метрике. Можно использовать различные метрики для оценки близости распределений. Мы рассмотрим две метрики: дивергенцию Йенсена–Шеннона [9] и метрику Васерштейна первого порядка [14].

2.2. Обучение конкурирующих порождающих нейросетей

Рассмотрим две нейросети, генератор $g_\theta(z)$ и дискриминатор $d_\phi(x)$. Генератор, как было описано ранее, принимает на вход вектор шума и преобразует его в (синтетическое) изображение. Дискриминатор принимает на вход изображение x и выдает одно число от 0 до 1 – вероятность того, что изображение является реальным изображением, а не синтетическим. Дискриминатор тренируется так, чтобы максимизировать ответ для реальных данных и минимизировать ответ для синтетических данных. Одновременно с ним тренируется генератор так, чтобы минимизировать величину $\log(1-d_\phi(g_\theta(z))$, т.е. минимизировать количество правильных ответов дискриминатора о том, что изображение является синтетическим.

В терминах теории игр нейросети g_θ и d_ϕ играют в минимаксную игру с функцией полезности

$$\min_{g_\theta}\max_{d_\phi} V(g_\theta, d_\phi) = \\ = \mathop{\mathrm{E}}_{x\sim P_r}\log d_\phi(x) + \mathop{\mathrm{E}}_{z\sim p(z)}\log\left(1-d_\phi\left(g_\theta\left(z\right)\right)\right). \quad (1)$$

В работе [9] показано, что такая игра эквивалентна минимизации дивергенции Йенсена–Шеннона распределений P_r и P_g:

$$\frac{1}{2}\left(KL\left(P_r \| \frac{1}{2}\left(P_r+P_g\right)\right)+KL\left(P_g \| \frac{1}{2}\left(P_r+P_g\right)\right)\right). \quad (2)$$

Таким образом, обучение нейросетей с помощью целевой функции (1) неявным образом минимизирует дивергенцию Йенсена–Шеннона, из чего следует сближение распределения P_g к P_r в процессе оптимизации.

2.3. Обучение порождающих нейросетей с помощью метрики Васерштейна первого порядка

Метрика Васерштейна первого порядка (также называется *Earth's mover distance*) сходства двух распределений реальных изображений P_r и синтетических изображений P_g определяется следующим образом:

$$W(P_r, P_g) = \inf_{\gamma\in\Pi(P_r,P_g)} \mathrm{E}_{(x,y)\in\gamma}\| x-y \|. \quad (3)$$

Здесь $\Pi(P_r, P_g)$ – множество всех совместных распределений $\gamma(x,y)$, маргинальные распределения которых равны соответственно P_r и P_g. Как было ранее замечено, $y = g_\theta(z)$, $z \sim p(z)$.

Опишем неформально устройство метрики. Для простоты рассмотрим одномерные дискретные распределения. Их можно представить в виде гистограмм различной формы. Тогда значение метрики – количество работы, которое нужно выполнить, чтобы преобразовать первую гистограмму во вторую. Работа определяется как количество перемещаемой массы из ячейки гистограммы, умноженное на перемещаемую дистанцию. В таком случае $\gamma(x,y)$ – двумерная гистограмма, которая задаёт план перемещения массы. Аналогичным образом метрика устроена и в многомерном случае.

Однако метрику в текущей формулировке (взятие инфимума по всем возможным распределениям $\Pi(P_r,P_g)$) сложно применить для обучения нейросетей. Используя двойственность Канторовича–Рубенштейна, метрику можно свести к следующему выражению:

$$W(P_r, P_g) = \\ = \sup_{\|f\|_L\le 1} \mathrm{E}_{x\sim P_r}\left[f(x)\right] - \mathrm{E}_{z\sim p(z)}\left[f(g_\theta(z))\right]. \quad (4)$$

Здесь супремум берётся по всем 1-липшицевым функциям. Представим это семейство функций приближённо семейством нейронных сетей с весами $w\in\mathcal{W}$. Тогда метрика будет выглядеть следующим образом:

$$W(P_r, P_g) = \\ = \max_{w\in\mathcal{W}} \mathrm{E}_{x\sim P_r}\left[f_w(x)\right] - \mathrm{E}_{z\sim p(z)}\left[f_w(g_\theta(z))\right]. \quad (5)$$

Метрику в таком виде уже можно использовать для обучения порождающей нейросети. Будем попеременно обучать две нейросети: нейросеть-критик, задающую функцию f_w и используемую при подсчёте метрики Васерштейна, и нейросеть-генератор, задающую функцию g_θ. Для того, чтобы гарантировать, что f_w – липшицева функция с фиксированной константой, веса нейросети-критика после каждого шага обучения ограничиваются сверху и снизу фиксированной константой c.

Использование метрики Васерштейна для обучения порождающих нейросетей имеет два ключевых преимущества по сравнению с дивергенцией Йенсена–Шеннона, используемой в обычных конкурирующих порождающих нейросетях:

1. Сходимость по метрике Васерштейна эквивалентна сходимости по распределению, тогда как сходимость по дивергенции Йенсена–Шеннона эквивалента сходимости по вариации; т.е. метрика Васерштейна допускает более широкий класс сходящихся распределений. Это свойство улучшает сходимость процесса обучения.
2. Метрика Васерштейна почти везде непрерывна и дифференцируема, значение метрики коррелирует с визуальным качеством генерируемых изображений (что неверно для дивергенции Йенсена–Шеннона). Это свойство делает процесс обучения понятным для исследователя.

Поскольку метрика Васерштейна имеет вышеназванные преимущества по сравнению с дивергенцией Йенсена–Шеннона, в экспериментах мы используем именно метрику Васерштейна.

2.4. Генерация условных изображений

Как было упомянуто во введении, на вход нейросеть-генератор получает многомерный вектор шума, а на выход выдаёт случайное изображение дорожного знака. При этом нейросеть должна уметь генерировать изображения нужных дорожных знаков. Таким образом, распределение $P_r(x|c)$ выходных изображений должно быть условным, т.е. зависеть от метки класса c. В работе [12] порождающие нейронные сети обучаются сэмплировать из условных распределений.

На вход генератору, помимо случайного шума, подаётся класс изображения, которое необходимо сгенерировать. Класс кодируется бинарным вектором методом one-hot: одна компонента вектора содержит 1, остальные – 0.

К сожалению, такой подход нельзя напрямую переложить на метрику Васерштейна, т.к. функция потерь отличается от обычных конкурирующих нейросетей и на практике нейросеть игнорирует метку класса. Поэтому было решено упростить задачу и вместо одной условной нейросети обучать N нейросетей, где N – количество классов дорожных знаков. Каждая нейросеть обучается на одном узком классе дорожных знаков. Как результат, качество генерируемых примеров повысилось (рис. 1, 2).

3. Экспериментальная оценка

3.1. База изображений и классификатор дорожных знаков

Для экспериментальной оценки генератора синтетических изображений была использована немецкая база автодорожных знаков GTSRB [15]. База содержит 43 класса знаков и 52 тысячи изображений, разделённых на тренировочную и тестовую выборки в пропорции 3:1. Размер изображений – от 15×15 пикселей до 250×250 пикселей. Возможно было использовать также базу RTSD [16], однако она содержит большее количество классов знаков (120) и время, требуемое на обучение порождающих нейросетей, увеличилось бы в 3 раза, до двух недель.

Рис. 1. Примеры изображений, сгенерированных нейросетью, обученной на всей выборке изображений дорожных знаков

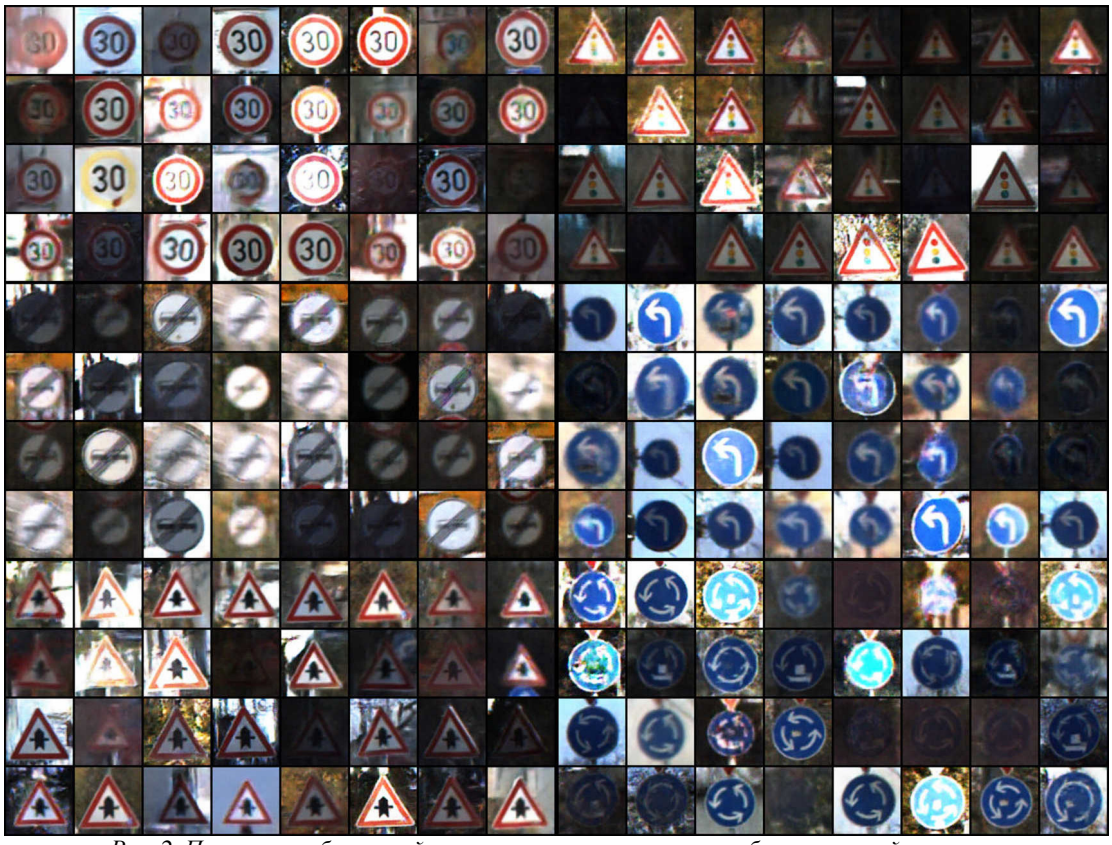

Рис. 2. Примеры изображений, сгенерированных поклассово обученными нейросетями

Существует два наиболее популярных метода классификации изображений дорожных знаков. Первый, на основе гистограмм ориентированных градиентов (HOG) и машины опорных векторов (SVM), подробно описан, например, в [17]. Второй метод, на основе свёрточных нейронных сетей, описан в [18]. Для экспериментов был выбран второй метод, т.к. свёрточные нейросети существенно превосходят другие методы в задаче

классификации изображений. Архитектура нейронной сети показана в табл. 1 и взята из статьи [18]. В качестве функций активаций после свёрток и предпоследнего полносвязного слоя используется ReLU.

Табл. 1. Архитектура свёрточной нейронной сети, использовавшейся для классификации знаков

№	Тип	Количество карт в слое и нейронов	Ядро
0	Входной	3 карты по 48×48 нейронов	
1	Свёрточный	100 карт по 100×100 нейронов	7×7
2	Max pooling	100 карт по 21×21 нейронов	2×2
3	Свёрточный	150 карт по 18×18 нейронов	4×4
4	Max pooling	150 карт по 9×9 нейронов	2×2
5	Свёрточный	250 карт по 6×6 нейронов	4×4
6	Max pooling	250 карт по 3×3 нейронов	2×2
7	Полносвязный	300 нейронов	1×1
8	Полносвязный	43 нейрона (классы)	1×1

Для размножения использовались следующие преобразования: повороты в пределах 10 градусов, сдвиги по вертикали и горизонтали на 10 % от размеров изображения и масштабирования с коэффициентом от 0,8 до 1,2. Параметры преобразований изображений для обучения нейросетей в различных работах отличаются, но обычно преобразования должны быть не очень сильные.

3.2. Генерация синтетических изображений

Архитектура нейросети-генератора взята из работы [10]. Её схема изображена на рис. 3. За счёт транспонированных свёрток эта нейросеть повышает разрешение 100-компонентного вектора шума до трёхканального изображения размером 64×64 пикселя. При подаче изображений в нейросетевой классификатор они масштабируются до размера 48×48 пикселей.

Нейросети-генераторы для разных классов знаков обучаются градиентным спуском батчами по 64 примера. В качестве метода градиентного спуска используется RMSProp с темпом обучения 0,00005. На обучение одной нейросети требуется 50–100 тысяч эпох. Суммарно обучение 43 порождающих нейросетей потребовало 4 дня вычислений на компьютере с процессором Intel Core i7-3770K, 32 Гб памяти и видеокартой Nvidia GeForce GTX 1070.

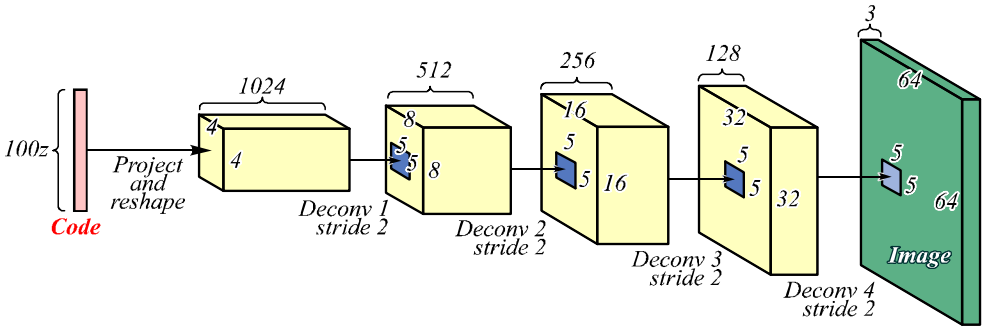

Рис. 3. Архитектура нейросети-генератора синтетических изображений дорожных знаков

В качестве метода для сравнения был выбран метод генерации изображений дорожных знаков по иконке из [6]. К иконке дорожного знака применяются следующие преобразования:

- размытие изображения по Гауссу,
- добавление независимого гауссовского шума,
- изменение яркости компонент изображения в пространстве HSV,
- применение размытия движения к изображению,
- добавление фона, вырезанного из реальных изображений с видеорегистратора.

Реализация метода генерации с параметрами преобразований была предоставлена автором статьи [6].

Примеры изображений, сгенерированных с помощью поклассовых порождающих нейросетей, показаны на рис. 2. Примеры изображений, сгенерированных по иконке, показаны на рис. 4.

Можно заметить, что порождающие нейросети, обучаемые с помощью метрики Васерштейна, позволяют генерировать фотореалистичные изображения, которые неотличимы от реальных для человеческого глаза.

При этом изображения, сгенерированные по иконке, существенно не похожи на реальные, и требуются дополнительные преобразования для получения, например, бликов и теней на изображениях.

Рис. 4. Примеры изображений, сгенерированных по иконке

3.3. Результаты экспериментов

Экспериментальное сравнение проводилось на трёх типах обучающих выборок: реальных данных из базы GTSRB, синтетических данных, сгенерированных с помощью нейросети или по иконке, смеси синтетических и реальных данных. Рассматривались выборки размером 39 тысяч примеров (размер реальной выборки), 215 тысяч примеров (по 5 тысяч изображений знаков на класс) и выборки, полученные размножением из двух предыдущих.

Результаты численных экспериментов показаны в табл. 2.

Табл. 2. Результаты тестирования классификатора на различных выборках знаков

	39 тыс. трен. без размножение	215 тыс. трен. без размножения	39 тыс. / 215 тыс. трен. с размножением
Реальные данные	96,6	–	98,4 / –
WGAN синтетика	95,3	96,1	97,6 / 98,1
Реальные данные + WGAN синтетика	–	97,7	– / 98,4
Синт. данные по иконке	46,5	53,7	67,8 / 69,7
Реальные данные + синтетика по иконке	–	96,5	– / 97,9

Из результатов можно сделать следующие выводы:

1. Классификатор, обученный на синтетических данных, показывает качество, близкое к классификатору, обученному на реальных данных, но, тем не менее, разница составляет 0,3 %. Заметим также, что для получения качества классификации, близкого к реальным данными, требуется больше примеров синтетических изображений.
2. Синтетические изображения можно использовать в дополнение к реальным данным для получения лучшего качества классификации, однако как метод размножения данных они работают хуже, чем эмпирическое размножение с помощью поворотов, сдвигов, масштабирований.

В сравнении с синтетическими изображениями, сгенерированными по иконке, нейросетевая синтетическая обучающая выборка существенно более качественная. При одинаковых объёмах выборок качество классификатора, обученного на нейросетевых синтетических данных, существенно выше.

Заключение

В данной работе был исследован вопрос применимости синтетических обучающих выборок, сгенерированных с помощью порождающих нейросетей, для обучения классификаторов изображений. Для примера была рассмотрена задача классификации изображений дорожных знаков. Экспериментальная оценка показала, что порождающие сети, обученные с помощью метрики Васерштейна на узких классах знаков, способны генерировать реалистичные изображения, подходящие для обучения нейросетевого классификатора, но достигаемое качество классификации всё ещё уступает, пусть и незначительно, качеству классификатора, обученного на реальных данных.

В дальнейшем планируется распространить подход на генерацию обучающих выборок для детектора дорожных знаков, а также развить его для генерации обучающих выборок классов объектов, для которых примеров изображений не существует.

Благодарности

Работа выполнена при поддержке гранта РНФ 17-71-20072 «Нейробайесовские методы в задачах машинного обучения, масштабируемой оптимизации и компьютерного зрения».

Литература

1. **Russakovsky, O.** ImageNet large scale visual recognition challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei // International Journal of Computer Vision. – 2015. – Vol. 115, Issue 3. – P. 211-252. – DOI: 10.1007/s11263-015-0816-y.
2. **Lin, T.** Microsoft COCO: Common objects in context / T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll'ar, L. Zitnick. – In book: Computer Vision – ECCV 2014 / ed. by D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars. – Switzerland: Springer International Publishing; 2014. – P. 740-755. – ISBN: 978-3-319-10592-5.
3. **Cordts, M.** The cityscapes dataset for semantic urban scene understanding / M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 3213-3223. – DOI: 10.1109/CVPR.2016.350.
4. **Shotton, J.** Real-time human pose recognition in parts from single depth images / J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore // Communications of the ACM. – 2013. – Vol. 56, Issue 1. – P. 116-124. – DOI: 10.1145/2398356.2398381.
5. **Richter, S.R.** Playing for data: Ground truth from computer games / S.R. Richter, V. Vineet, S. Roth, V. Koltun. – In book: Computer Vision – ECCV 2016 / ed by B. Leibe, J. Matas, N. Sebe, M. Welling. – Cham, Switzerland: Springer, 2016. – P. 102-118. – DOI: 10.1007/978-3-319-46475-6_7.
6. **Moiseyev, B.** Evaluation of traffic sign recognition methods trained on synthetically generated data / B. Moiseyev, A. Konev, A. Chigorin, A. Konushin // Proceedings of the 15th International Conference on Advanced Concepts for Intelligent Vision Systems. – 2013. – P. 576-583. – DOI: 10.1007/978-3-319-02895-8_52.
7. **Chigorin, A.** A system for large-scale automatic traffic sign recognition and mapping / A. Chigorin, A. Konushin // CMRT13 – City Models, Roads and Traffic 2013. ISPRS

Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2013. – Vol. II-3/W3. – P. 13-17. – DOI: 10.5194/isprsannals-II-3-W3-13-2013.
8. **Fischer, P.** Flownet: Learning optical flow with convolutional networks / P. Fischer, A. Dosovitskiy, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, T. Brox // arXiv preprint arXiv:1504.06852. – 2015.
9. **Goodfellow, I.** Generative adversarial nets / I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // Proceedings of the 27th International Conference on Neural Information Processing Systems. – 2014. – Vol. 2. – P. 2672-2680.
10. **Radford, A.** Unsupervised representation learning with deep convolutional generative adversarial networks [Electronical Resource] / A. Radford, L. Metz, S. Chintala // arXiv preprint arXiv:1511.06434. – 2015. – URL: https://arxiv.org/abs/1511.06434. (date request 24.11.2017).
11. **Denton, E.L.** Deep generative image models using a Laplacian pyramid of adversarial networks / E.L. Denton, S. Chintala, A. Szlam, R. Fergus // Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS). – 2015. – Vol. 1. – P. 1486-1494.
12. **Mirza, M.** Conditional generative adversarial nets [Electronical Resource] / M. Mirza, S. Osindero // arXiv preprint arXiv:1411.1784. – 2014. – URL: https://arxiv.org/abs/1411.1784 (date request 24.11.2017).
13. **Zheng, Z.** Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [Electronical Resource] / Z. Zheng, L. Zheng, Y. Yang // arXiv preprint arXiv:1701.07717. – 2017. – URL: https://arxiv.org/abs/1701.07717 (date request 24.11.2017).
14. **Arjovsky, M.** Wasserstein gan [Electronical Resource] / M. Arjovsky, S. Chintala, L. Bottou // arXiv preprint arXiv:1701.07875. – 2017. – URL: https://arxiv.org/abs/1701.07875 (date request 24.11.2017).
15. **Stallkamp, J.** Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition / J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel // Neural networks. – 2012. – Vol. 32. – P. 323-332. – DOI: 10.1016/j.neunet.2012.02.016.
16. **Шахуро, В.И.** Российская база изображений автодорожных знаков / В.И. Шахуро, А.С. Конушин // Компьютерная оптика. – 2016. – Т. 40, № 2. – С. 294-300. – DOI: 10.18287/2412-6179-2016-40-2-294-300.
17. **Лисицын, С.О.** Распознавание дорожных знаков с помощью метода опорных векторов и гистограмм ориентированных градиентов / С.О. Лисицын, О.А. Байда // Компьютерная оптика. – 2012. – Т. 36, № 2. – С. 289-295.
18. **Ciresan, D.** Multi-column deep neural network for traffic sign classification / D. Ciresan, U. Meier, J. Masci, J. Schmidhuber // Neural Networks. – 2012. – Vol. 32. – P. 333-338. – DOI: 10.1016/j.neunet.2012.02.023.

Сведения об авторах

Шахуро Владислав Игоревич, 1993 года рождения, в 2015 году окончил МГУ имени М.В. Ломоносова. Аспирант НИУ Высшая школа экономики. Научные интересы: обработка изображений, компьютерное зрение, машинное обучение, программирование. E-mail: *vlad.shakhuro@hse.ru* .

Конушин Антон Сергеевич, 1980 года рождения, в 2002 году окончил МГУ имени М.В. Ломоносова. В 2005 году защитил кандидатскую диссертацию в ИПМ имени М.В. Келдыша РАН. Работает доцентом на ВМК МГУ имени М.В. Ломоносова. Научные интересы: компьютерное зрение, машинное обучение. E-mail: *anton.konushin@graphics.cs.msu.ru* .

ГРНТИ: 28.23.15.
Поступила в редакцию 19 июля 2017 г. Окончательный вариант – 1 декабря 2017 г.

IMAGE SYNTHESIS WITH NEURAL NETWORKS FOR TRAFFIC SIGN CLASSIFICATION

V.I. Shakhuro [1], A.S. Konushin [1,2]
[1] NRU Higher School of Economics, Moscow, Russia,
[2] Lomonosov Moscow State University, Moscow, Russia

Abstract

In this work, we research the applicability of generative adversarial neural networks for generating training samples for a traffic sign classification task. We consider generative neural networks trained using the Wasserstein metric. As a baseline method for comparison, we take image generation based on traffic sign icons. Experimental evaluation of the classifiers based on convolutional neural networks is conducted on real data, two types of synthetic data, and a combination of real and synthetic data. The experiments show that modern generative neural networks are capable of generating realistic training samples for traffic sign classification that outperform methods for generating images with icons, but are still slightly worse than real images for classifier training.

Keywords: traffic sign classification, synthetic training sample, generative neural network.
Citation: Shakhuro VI, Konushin AS. Image synthesis with neural networks for traffic sign classification. Computer Optics 2018; 42(1): 105-112. DOI: 10.18287/2412-6179-2018-42-1-105-112.
Acknowledgements: This work was supported by the Russian Science Foundation (RSF) grant 17-71-20072 "Deep Bayesian Methods in Machine Learning, Scalable Optimization and Computer Vision Problems".

References

[1] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg A, Fei-Fei L. ImageNet large scale visual recognition challenge. International Journal of Computer Vision 2015; 115(3): 211-252. DOI: 10.1007/s11263-015-0816-y.

[2] Lin T, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Doll'ar P, Zitnick L. Microsoft COCO: Common objects in context. In Book: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. Computer Vision – ECCV 2014. Switzerland: Springer International Publishing; 2014: 740-755. ISBN: 978-3-319-10592-5.

[3] Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition 2016: 3213-3223. DOI: 10.1109/CVPR.2016.350.

[4] Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M, Moore R. Real-time human pose recognition in parts from single depth images. Communications of the ACM 2013; 56(1): 116-124. DOI: 10.1145/2398356.2398381.

[5] Richter SR, Vineet V, Roth S, Koltun V. Playing for data: Ground truth from computer games. In book: Leibe B, Matas J, Sebe N, Welling M, eds. Computer Vision – ECCV 2016. Cham, Switzerland: Springer; 2016: 102-118. DOI: 10.1007/978-3-319-46475-6_7.

[6] Moiseyev B, Konev A, Chigorin A, Konushin A. Evaluation of traffic sign recognition methods trained on synthetically generated data. ACIVS 2013: 576-583. DOI: 10.1007/978-3-319-02895-8_52.

[7] Chigorin A, Konushin A. A system for large-scale automatic traffic sign recognition and mapping. CMRT13 – City Models, Roads and Traffic 2013: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2013; II-3/W3: 13-17. DOI: 10.5194/isprsannals-II-3-W3-13-2013.

[8] Fischer P, Dosovitskiy A, Ilg E, Hausser P, Hazirbas C, Golkov V, van der Smagt P, Cremers D, Brox T. Flownet: Learning optical flow with convolutional networks. arXiv preprint arXiv:1504.06852 2015.

[9] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Proc NIPS 2014; 2: 2672-2680.

[10] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 2015. Source: ⟨https://arxiv.org/abs/1511.06434⟩.

[11] Denton EL, Chintala S, Fergus R. Deep generative image models using a Laplacian pyramid of adversarial networks. Proc NIPS 2015; 1: 1486-1494.

[12] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 2014. Source: ⟨https://arxiv.org/abs/1411.1784⟩.

[13] Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. arXiv preprint arXiv:1701.07717 2017. Source: ⟨https://arxiv.org/abs/1701.07717⟩.

[14] Arjovsky M, Chintala S, Bottou L. Wasserstein gan. arXiv preprint arXiv:1701.07875 2017. Source: ⟨https://arxiv.org/abs/1701.07875⟩.

[15] Stallkamp J, Schlipsing M, Salmen J, Igel C. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Networks 2012; 32: 323-332. DOI: 10.1016/j.neunet.2012.02.016.

[16] Shakhuro VI, Konushin AS. Russian traffic signs images dataset [In Russian]. Computer Optics 2016; 40(2): 294-300. DOI: 10.18287/2412-6179-2016-40-2-294-300.

[17] Lisitsyn SO, Bayda OA. Road sign recognition using support vector machines and histogram of oriented gradients [In Russian]. Computer Optics 2012; 36(2): 289-295.

[18] Ciresan D, Meier U, Masci J, Schmidhuber J. Multi-column deep neural network for traffic sign classification. Neural Networks 2012; 32: 333-338. DOI: 10.1016/j.neunet.2012.02.023.

Author's information

Vladislav Igorevich Shakhuro, (b. 1993), graduated from Lomonosov Moscow State University in 2015. Currently graduate student at NRU Higher School of Economics. Research interests are image processing, computer vision, machine learning, and programming. E-mail: *vlad.shakhuro@hse.ru* .

Anton Sergeevich Konushin, (b. 1980), graduated from Lomonosov Moscow State University in 2002. In 2005 he successfully defended his PhD thesis in M.V. Keldysh Institute for Applied Mathematics RAS. He is currently associate professor at Lomonosov Moscow State University. Research interests are computer vision and machine learning. E-mail: *anton.konushin@graphics.cs.msu.ru* .

Received July 19, 2017. The final version – December 1, 2017.