

Методы повышения эффективности обучающих выборок путем дополнения их генерированными графическими данными

А.И. Ефимов (МАИ)

Проанализирован ряд известных методов решения проблемы недостаточной репрезентативности обучающей выборки. Предложен новый метод, заключающийся в использовании генерированных графических данных для формирования обучающих выборок. Определены возможности использования и границы применимости данного подхода, сформулированы его достоинства и недостатки. Предложены способы решения выявленных проблем и методы повышения их эффективности.

Ключевые слова: нейронные сети, глубокое обучение, аугментация графических данных, генерация графических данных, формирование обучающей выборки

Системы компьютерного зрения находят все более широкое применение в современном мире, начиная от технологических процессов и медицины, заканчивая развлекательными приложениями. Можно выделить основные области их применения: детектирование, распознавание, классификация, сегментация изображения, построение объемной модели окружения или объекта и т. д. На данный момент наиболее динамично развивающейся областью данной технологии являются алгоритмы, основанные на нейронных сетях глубокого обучения. Данные системы позволяют решать широкий круг задач путем выявления сложных зависимостей между входными/выходными данными. Эти действия производятся на обучающей выборке в автоматическом режиме, после чего результаты этого обучения используются для решения аналогичной задачи, в частности, задачи анализа изображения на новых данных [1]. Обучающая выборка представляет собой набор записей, каждая из которых состоит из входных и сопоставленных с ними эталонных выходных данных. В рамках задач компьютерного зрения структуры входных/выходных данных можно условно разделить на содержащие относительно малый объем числовой информации (< 100 выходных числовых параметров) и представляющие большие массивы числовой информации, чаще всего организованные в виде изображения. Для удобства изложения далее будут введены термины классификации нейронных сетей по типу входных/выходных данных:

1) изображение->числовая информация (image-number network — INN) используется в рамках за-

дач детектирования, классификации, верификации (рис. 1 (1));

2) изображение->изображение (image-image network — IIN) используется в рамках сегментации, преобразования сжатия и восстановления изображений (рис. 1(2));

3) числовая информация->изображение (number-image network — NIN) используется для генерации изображений, исходя из набора числовых параметров (рис. 1 (3)).

Существуют значительные сложности в реализации данных систем. Одной из них является необходимость наличия репрезентативной обучающей выборки. Важно отметить, что обеспечение репрезентативности выборки часто является крайне сложной задачей. Сложность может заключаться в недостатке объема и разнообразия графических данных, а также сложности и трудоемкости их разметки. Относительно причин сложности разметки нужно отметить, что в INN для успешного обучения необходимо большое число записей (часто > 1 млн. Ед.), а IIN, несмотря на то, что требует значительно меньшее их число, имеет на порядок большую трудоемкость разметки каждого отдельного изображения и сложность (в некоторых случаях невозможность) обеспечения необходимой точности.

Частным случаем решения проблемы недостатка обучающей выборки является аугментация данных, заключающаяся в простейших преобразованиях изображения из тестовой выборки перед его отправлением на вход нейронной сети [2]. В качестве преобразо-

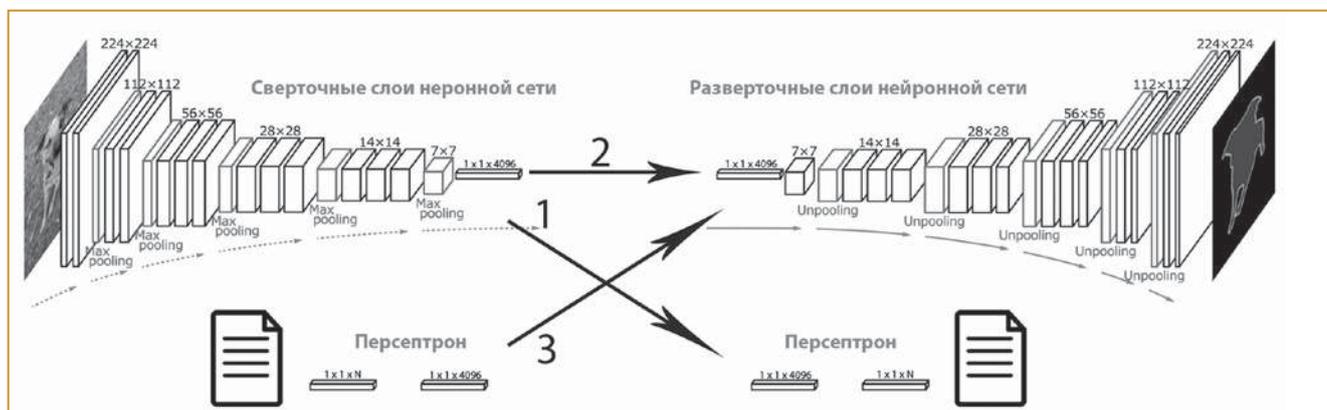


Рис. 1. Классификация нейронных сетей по типу входных/выходных данных



Рис. 2. Аугментация графических данных

ваний используют сдвиг на некоторое число пикселей, зеркалирование, поворот на угол, выделение области, скрытие части изображения закрасиванием определенной области, изменение яркости, контрастности и гаммы изображения или его элементов (рис. 2). Данные преобразования позволяют уменьшить привязанность признаков к местоположению на изображении и к цвету отдельных элементов, текстура которых может меняться, новая разметка при этом либо не меняется (в случае классификации и верификации), либо тривиально получается из старой (в случае сегментации и определении позиции объекта на фото).

Более сложный вариант заключается в комбинировании частей изображений тестовой выборки в один коллаж. В чистом виде он редко может быть эффективно применен в связи с сильной разницей полученных изображений с естественными. Однако, используя PIN модификации Generative Adversarial Networks (GAN), можно значительно сократить разницу генерированных и реальных изображений. Принцип действия GAN заключается в попеременном обучении двух соревновательных нейросетей, одна из которых тренируется классифицировать изображения на сгенерированные и реальные, а другая

генерировать данные с реалистичностью, достаточной для “обмана” первой. Использование GAN может значительно повысить реалистичность коллажей. Данный подход сейчас активно разрабатывается в медицине для обучения систем диагностики по рентген снимкам [3]. Реже для генерации изображений возможен стандартный PIN вариант GAN с построением изображения по заданным параметрам. Редкость применения обусловлена недостатком разнообразия сгенерированных изображений, постоянным следованием определенным паттернам при генерации и нетривиальным получением разметки изображения.

Однако перечисленные выше решения не всегда являются достаточными для расширения обучающей выборки в рамках конкретной задачи. Для подобных случаев можно предложить вариант генерации входных и эталонных изображений, а также сопутствующих данных путем использования полноценной математической модели виртуальной сцены (рис. 3). Применение данного метода в настоящее время ограничено обучением алгоритмов принятия решений в рамках систем управления транспортными средствами (при этом обучение модулей компьютерного зрения производится отдельно на реальных

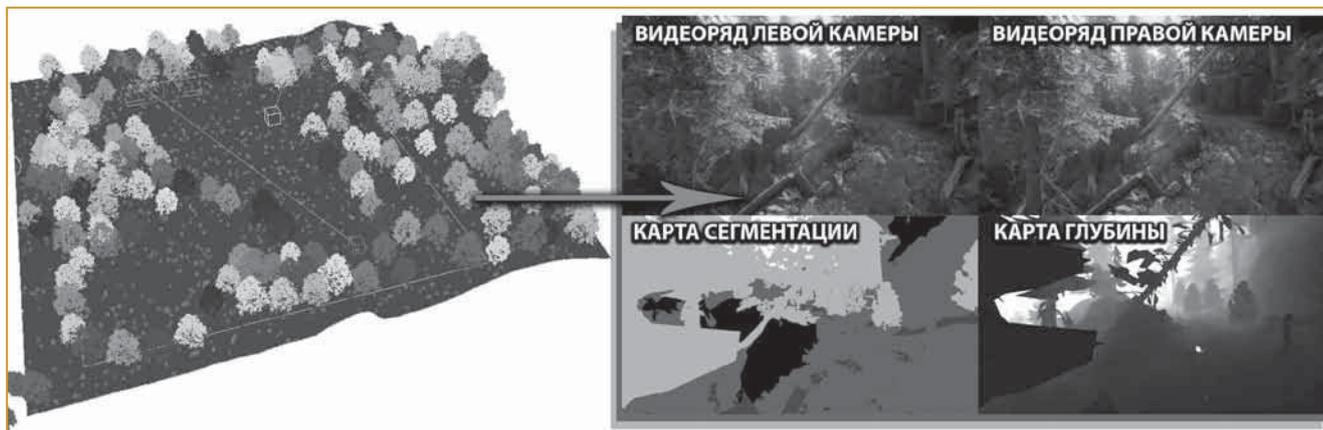


Рис. 3. Генерация обучающей выборки изображений из полной математической модели локации

данных) и практически не рассматривается использование обученных на виртуальных графических данных нейронных сетей компьютерного зрения в рамках реального мира. Для построения предложенной модели виртуальной сцены подходит множество инструментариев для работы с 3D графикой, имеющих поддержку использования сценариев (3DsMax, Maya, Blender), а также инструментариев создания 3D приложений (Unity3D, Unreal Engine, Cryengine). Преимущество первой категории заключается в более высоком уровне графики, преимущества второй — в более высоких гибкости функциональности и скорости работы. Для экспериментов в рамках данной работы использовался Unity3D по причине широких возможностей и функциональности, свободного пространства, наличия большого числа инструментов и готового контента в магазине дополнений.

Принцип генерации изображений состоит из нескольких итераций, на каждой из которых выбирается один из диапазонов высот съемки (2...10 м, 10...70 м, 70...500 м), исходя из которого строится непосредственно сама локация нужного размера и степени детализации. Принцип построения локации может быть как полностью случайным, так и опираться на векторную топологическую карту. Далее на построенной локации производится создание отдельных элементов обучающей выборки, каждый из которых содержит набор данных в конкретный момент времени, в который могут входить изображения с различных камер, значения датчиков, а также эталонные изображения и числовые свойства (в рамках экспериментов использовалось одно изображение с камеры и карта сегментации). После генерации определенного числа записей начинается следующая итерация, и это повторяется до создания обучающей выборки нужного размера.

Эксперименты в рамках данной работы осуществлены с использованием ПК со следующими характеристиками: процессор Intel Core i7—6700 3,4 ГГц; RAM 8 ГБ; ЦПУ NVIDIA GeForce GTX 950. Разрешение генерируемых изображений 1920x1080, размер изображений на входе/выходе нейронных сетей зависит от используемой топологии сети и находится в диапазоне от 256x256 до 960x540. В рамках эксперимента рассматривались нейросети, решающие задачу сегментации изображения на пяти составляющих (небо, поверхность земли, крупные естественные объекты (деревья, крупные кустарники), крупные искусственные объекты (здания, дороги, транспортные средства), водоемы). В качестве данных использованы изображения аэросъемки природных ландшафтов средней полосы России и схожих климатических зон с различных ракурсов в диапазоне высот 2...500 м. Результаты оценки на PIN архитектурах Unet, ResNet и VGG показали разницу абсолютных погрешностей в пределах 3% от аналогичных, обученных на реальных фотографиях, причем данная разница становилась еще меньше при увеличении числа слоев сети. В частности, в рамках Unet архитектуры с 19 сверточными слоями

удалось достигнуть погрешности 18,6% при обучении на реальных данных и 19,4% на генерированных с временем обучения около 2 ч. Зафиксировано отсутствие разницы в скорости обучения за исключением условий п. 4, расположенного ниже списка.

Подход с обучением нейросетей на генерированных данных имеет ряд преимуществ.

1. Возможность быстрой генерации практически неограниченного объема графического контента, обладающего широким разнообразием при обеспечении достаточной вариативности виртуальной сцены. Скорость генерации может варьироваться от заданного качества изображения и системных ресурсов. В рамках эксперимента при максимальном качестве была достигнута скорость в 72 изображения в минуту без учета времени на пересоздание локации.

2. Обеспечение полноты и точности эталонной информации за счет использования математической модели виртуальной сцены.

3. Обеспечение возможности эмуляции любого числа датчиков, среди которых могут быть дополнительные камеры, лидар, GPS трекер, акселерометр, гироскоп и т. д. В частности, данный подход позволяет обучить нейронную сеть, строящую карту глубины на основе нескольких изображений, а также обучить алгоритмы построения трехмерной модели окружающего пространства, что очень сложно реализовать на реальных данных из-за крайне сложной и неточной разметки данных обучающей выборки.

4. Генерация данных разной сложности и детализованности, что позволяет ускорить выбор и тестирование различных топологий нейронной сети. Достигается это проверкой упрощенной топологии на более простых изображениях. Это особенно полезно в условиях сильной ограниченности в вычислительных средствах. Кроме того, постепенное усложнение изображений полезно при использовании ResNet сетей в связи с последовательным обучением их слоев, эксперименты показали, что данный подход позволяет сокращать время обучения вплоть до 40%.

5. В случае обучения поведения системы управления перемещающимся устройством, вышеописанный подход позволяет погрузить его полностью в виртуальную среду, что позволит эффективно обучить и провести первичное тестирование логики перемещения, не создавая опасность для устройства и объектов окружения [4]. Эффективность обучения обусловлена большим разнообразием обрабатываемых ситуаций и локаций, ускорением времени в рамках виртуальной среды и большой свободой в принятии решений.

Однако помимо вышеописанных преимуществ у этого подхода существует ряд серьезных, но вполне разрешаемых сложностей.

1. Трудоемкость создания разнообразных и проработанных виртуальных сцен. Данная проблема значительно сокращается проработанностью современных графических инструментов. Важно также отметить

наличие при большинстве из инструментариев магазинов, распространяющих адаптированный графический контент, включающий модели, текстуры, материалы, анимацию, дополнительные инструменты отрисовки, редактирования и генерации виртуальных объектов и графических эффектов.

2. Отличие генерированных изображений от фотографий является не менее важной сложностью. Развитие современных аппаратных и программных средств позволяет добиться сверхвысокой точности имитации реальных объектов. Особенно это касается инструментов с длительным расчетом изображения, однако инструменты, адаптированные под рендеринг¹ в реальном времени могут обеспечивать хоть и меньшую, но достаточную реалистичность изображения со значительно меньшими затратами вычислительных ресурсов и времени. Эксперименты в рамках данной работы доказали эффективность использования обоих подходов для обучения ПН с разницей точности в пределах статистической погрешности. В рамках INN эффективен только подход с длительным временем рендеринга.

3. При использовании генерированных изображений для обучения нужно помнить про увеличение вероятности переобучения нейронной сети. В первую очередь это связано с отличием генерированных изображений на уровне отдельных пикселей, однако данная проблема в рамках ПН практически полностью решается генерированием изображений в 4 раза большего разрешения, чем подаваемое на вход нейронной сети, после чего производится гауссово размытие на 4...10 пикселей.

Нужно отдельно отметить возможность улучшения фотореалистичности изображений за счет использования модифицированной ПН версии GAN, схожей с ранее упомянутой сетью для генерации изображений рентген снимков. В частности, можно выделить модификации SRGAN [5]. Данный вариант крайне ресурсоемкий, однако позволяет получать изображения практически неотличимые от реальных. Основным применением этого подхода является адаптивное обучение данных для обучения INN.

Помимо генерации фотореалистичных датасетов, важно отметить полезность генерированных изображений в более узких задачах. Подобной задачей может быть анализ формы труб по карте проходящего сквозь них света. В данной задаче, зная геометрию трубы и карту входного свечения, можно построить пример-

ную выходную карту освещенности для данных условий. Исходя из этого для обучения алгоритма можно создать множество подобных комбинаций как для корректных, так и для деформированных вариантов труб.

Все вышеперечисленное показывает, что использование генерированных данных как в сочетании с реальными данными, так и без них в рамках обучающей выборки может быть крайне полезно при обучении ПН в широком диапазоне прикладных задач, а также в частных случаях INN. Среди прикладных задач можно в первую очередь выделить задачи с недостатком данных для обучения, например, в случае обучения системы управления наземным транспортом присутствует множество условий и ситуаций, которое сложно заполнить только реальными данными. Важно отметить максимальную адаптивность данного подхода к обучению систем управления воздушными, наземными, надводными транспортными средствами, управления станками и роботизированными системами в связи с возможностью генерации данных, зависящих от принимаемых системой решений [6]. Кроме того, описанный способ применим в случае возможности сведения задачи к построению детерминированной математической модели, в частности, для определения форм труб по карте проходящего сквозь них света.

Список литературы

1. *He Wang, Wen-Hao Zhang, K. Y. Michael Wong, Si Wu.* Advances in Neural Networks ISNN 2017. Sapporo, Hakodate, Hokkaido, Japan, 2017.
2. *Jason Wang, Luis Perez.* The Effectiveness of Data Augmentation in Image Classification using Deep Learning Stanford University, Stanford, USA, 2017.
3. *Shahrokh Valaee.* Training artificial intelligence with artificial X-rays. University of Toronto Faculty of Applied Science & Engineering. Toronto, Canada. 2018.
4. *Jennifer Chu.* Researchers develop virtual-reality testing ground for drones. MIT News Office. <http://news.mit.edu/2018/virtual-reality-testing-ground-drones-0517>.
5. *Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi* Twitter. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. IEEE Conference on Computer Vision and Pattern Recognition. 2017.
6. *Ефимов А.И., Балилый Н.А.* Методика обучения систем управления беспилотными летательными аппаратами путем погружения их в виртуальную реальность // Кибернетика и программирование. 2019. № 2. С.17-22.

*Ефимов Александр Иванович — аспирант Московского авиационного института (национального исследовательского университета).
Контактный телефон 8(967)285-60-17.
E-mail: alhimic23@yandex.ru*

¹ Рендеринг (rendering — визуализация) — термин в компьютерной графике, обозначающий процесс получения изображения по модели с помощью компьютерной программы. Здесь модель — это описание любых объектов или явлений на строго определенном языке или в виде структуры данных. Такое описание может содержать геометрические данные, положение точки наблюдателя, информацию об освещении, степени наличия какого-то вещества, напряженность физического поля и пр.