УДК 004.92

ПОДГОТОВКА И ОБРАБОТКА ДАННЫХ В ЗАДАЧАХ ГЕНЕРАЦИИ ФОТОРЕАЛИСТИЧНЫХ ИЗОБРАЖЕНИЙ

Шустов И.С.[1], Белов Ю.С.[1]

[1]Калужский филиал ФГБОУ ВО «Московский государственный технический университет им. Н.Э. Баумана (национальный исследовательский университет)», Калуга, e-mail: fn1-kf@mail.ru;

Проведён анализ работ по теме нейронных сетей в области работы с изображениями, а так же их непосредственной генерации. В частности, были проанализированы подходы к предварительной обработке данных перед проведением обучения искусственных нейронных сетей. Показано, что выбор правильного подхода к обработке данных может существенно повлиять на качество выполнения работы в целом. Продемонстрирована предварительная подготовка данных применительно к задачам генерации изображений при помощи свёрточных нейронных сетей. Предложен комбинированный подход, заключающийся в сочетании непрерывной последовательных кадров с разных углов зрения на объект и расширенного набора классов объектов и углов зрения. Так же приводится математическая модель обработки подобных данных. Модель представляет собой комбинацию обработки данных с помощью свёрточной нейронной сети, распознавание признаков при помощи многослойного перцептрона, сохранение характеристик в LSTM и развёртки с помощью разверточной нейронной сети. Даётся логическое и математическое обоснование предиктивному обучению сети при помощи последовательности, состоящей из изображений с соседних точек зрения, перебираемых в одном направлении, и сравнения сгенерированного предсказанного изображения с реальным. Делаетя предположение о большей эффективности подобного подхода.

Ключевые слова: нейронные сети, генерация изображений, подготовка данных, математическая модель

DATA PREPARATION AND PROCESSING IN TASKS OF PHOTOREALISTIC IMAGES GENERATION

Shustov I.S.[1], Belov Y.S.[1]

[1]Moscow State Technical University n.a. Bauman (National Research University), Kaluga Branch, Kaluga, e-mail: fn1kf@mail.ru

The analysis of works on a theme of neural networks in the field of image processing is carried out. In particular, there were analyzed the approaches to the preliminary processing of data before the training of neural networks. It is shown that the choice of the correct approach to data processing can significantly affect the quality of the work as a whole. There was demonstrated the data preparation for image generation tasks using convolutional neural networks. A combined approach consisting of a combination of continuous successive frames from different angles of view to an object and an extended set of classes of objects and angles of view is proposed. A mathematical model for processing such data is also given. The model is a combination of data processing using a convolutional neural network, futures recognition using MLP, saving characteristics in LSTM and deconvolution using deCNN. The rationale for predictive network learning is provided using a sequence, which contains a set of images from neighboring points of view, which are sorted in one direction, and comparing the generated predicted image with the real one. It is assumed that this approach is more effective.

Keywords: neural networks, image generation, data preparation, mathematic model

Введение. В настоящее время в задачах генерации фотореалистичных изображений при помощи компьютеров постепенно всё более значимое место занимают приложения с применением нейронных сетей. Особенно там, где стандартные алгоритмы приводят к относительно большой ошибке [1]. Однако далеко не последнюю роль в получении хороших результатов в генерации изображений при помощи нейросетей играет правильный подбор и обработка обучающих данных.

Качество обучения нейронной сети зависит не только от алгоритма обучения, но и от качества обучающей выборки [2] в частности и подготовки данных в целом [3]. В данной

работе предлагается подход к обучающим данным, является развитием подходов, представленных в работах [4, 5].

Анализ подходов к предобработке данных для обучения. В работе [4] входные данные представляют собой последовательность сгенерированных при помощи трёхмерного редактора изображений голов людей, используемых для предиктивного обучения. На каждом последующем изображении голова повёрнута на 30 градусов относительно предыдущего (рис. 1). Данный подход позволил с успехом выполнить цель работы [4].

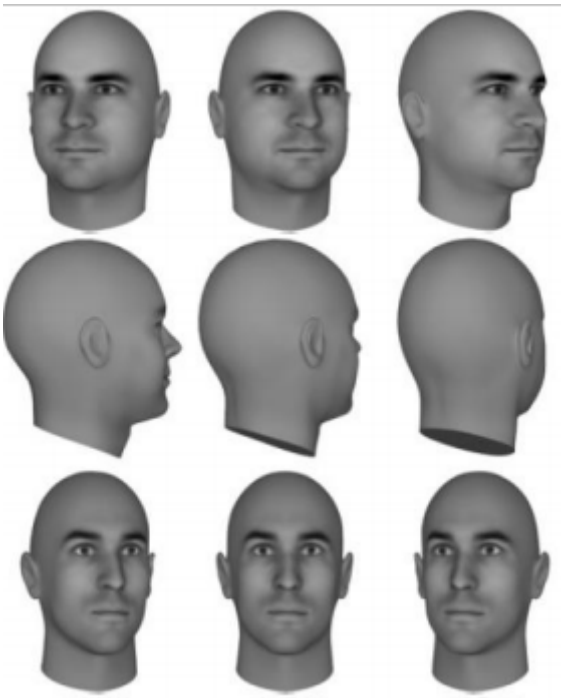

Рисунок 1 - Пример используемых в [4] изображений.

В работе [5] входные данные представляют собой набор из 809 поднаборов изображений различных стульев, каждый из которых содержит по 62 изображения объекта с разных сторон: 31 изображение по кругу с возвышением в 20 и 30 градусов. Изображения имеют разрешение 128 на 128 пикселей. Изображение стула расположено на белом фоне (рис. 2).

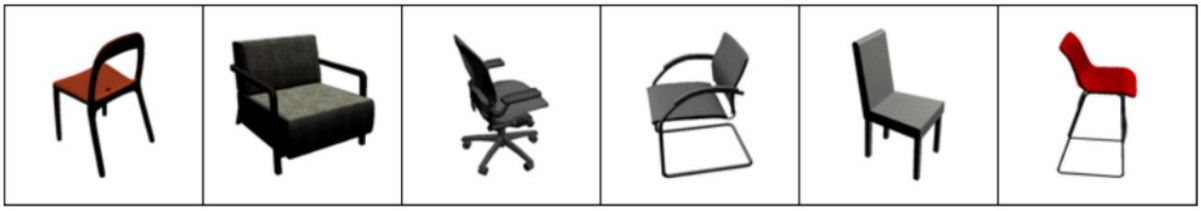

Рисунок 2 – Пример используемых в [5] изображений.

В данной работе предложен комбинированный подход к компоновке и обработке входных данных для задач генерации фотореалистичных изображений. Архитектурное решение нейронной сети, для которого предложен данный подход, изложен в [6]. В [6] предлагается создать генеративную сеть из трёх свёрточных подсетей для каждого канала

изображения – красного, синего и зелёного, в центральной части расположить долговременно-кратковременную память, а непосредственно для генерации изображения на выходе расположить четыре развёрточные нейронные сети, три для стандартных канала изображения и четвёртый для маски объекта на изображении. Так же для улучшенного механизма обучения в [6] предложено помимо генеративной сети создать дискриминирующую сеть, устанавливающую степень правдоподобности сгенерированного изображения и изменяющую тем самым переменные уравнений обучения генеративной сети.

Стандартное изображение входного множества имеет размерность 128 на 128 пикселей. На белом фоне в изображении расположен какой-либо объект, имеющий конечный объём. Изображения сгруппированы по классам, например, «стул», «стол», «кровать». Каждый класс делится на несколько типов, характеризуемых своими качественными характеристиками: текстурой, цветом, аспектами форм. Например, класс «стол» делится на типы «журнальный столик из красного дерева» и «светлый компьютерный стол». Каждому типу соответствует набор изображений с одним и тем же объектом, изображённым с разных углов зрения и при разном освещении (рис. 3).

Рисунок 3 – Пример изображений.

Изображения являются сгенерированными при помощи компьютерной трёхмерной графики. Шаг вращения объекта по горизонтали составляет 20 градусов, что даёт по 18 изображений объекта на одно значение широты. Угол возвышения камеры варьируется от -80 до 80 градусов с тем же шагом 20 градусов. Итого на одну освещённость на один тип приходится по 162 изображения одного объекта.

Общее количество классов и типов, а так же наборов свойств, применяемых для обучения, может меняться в зависимости от реакции сети на обучение. В среднем для обучения сети желательно использовать порядка 100 классов, каждый по десять типов. Итого 162 000 изображений.

При помощи таких данных можно сгенерировать изображение разными способами. В работах [4, 5] представлены сходные подходы, в которых генерация изображений достигается благодаря глубоким развёрточным нейросетям. Эти работы различаются методами обучения и возможностями генерации. В данной работе будет рассмотрено преобразование данных для архитектуры, представленной в работе [6].

Математическая модель преобразования данных. Данные преобразуются следующим образом. Изначально белый фон перед подачей в сеть заменяется чёрным для того, чтобы избежать излишних значений возбуждений нейронов в сети. Затем изображение разбивается на три потока, т. е. по факту его можно представить как трёхмерную матрицу размерностью 128x128x3, где каждая ячейка содержит значение от 0 до 1, при чём 0 означает отсутствие компонента, а 1 – максимальное его присутствие в соответствующем пикселе.

Далее эта трёхмерная матрица проходит преобразование с использованием свёрток. Пусть входное изображение является

$$Input = x^i = (x_r^i, x_g^i, x_b^i) \quad (1)$$

где x^i - изображение в каналах RGB, x_r^i, x_g^i, x_b^i - изображения, состоящие исключительно из каналов R, G, B соответственно, i – номер изображения в выборке. Тогда карта свойств, полученная после прохождения двумерной матрицы через операцию свёртки, будет

$$A_{ij}^{lh} = \sum_{q=i-2}^{i+2} \sum_{w=j-2}^{j+2} F(W_{qwk}^{lh'} A_{qw}^{(l-1)h'}) + bias_{ij}^{lh} \quad (2)$$

где i, j, q, w – положение нейрона в карте свойств по вертикали и горизонтали, l – номер слоя, h – номер карты свойств, A – активация соответствующего нейрона, W – весовой коэффициент в ядре свёртки, bias – смещение соответствующего нейрона, F – функция активации нейрона. В итоге после прохождения всех операций свёртки и макс-пулинга, на выходе многослойного персептрона образуется многокомпонентный вектор, состоящий из 1162 компонент. Сюда входит 1000-элементный вектор классификации и 162-х элементный вектор угла зрения.

Следующий этап не является преобразованием данных как таковым. Это этап обучения свёрточной подсети по классическому алгоритму обратного распространения ошибки. Модификация весовых коэффициентов и ядер происходит по формулам:

$$\sigma_{ij}^{l_{out}} = \left(d_{ij} - y_{ij}^{l_{out}}\right) * f'\left(s_{ij}\right) \quad (3)$$

ошибка выхода нейрона последнего слоя MLP свёрточной сети, где d_{ij} - целевое значение выхода нейрона, $y_{ij}^{l_{out}}$ – действительное значение выхода нейрона, $f'\left(s_{ij}\right)$ - производная активационной функции нейрона, s_{ij} - взвешенный вход нейрона.

$$\Delta w_{ijk}^{l_{out}} = \lambda \sigma_{ij}^{l_{out}} A_{qw}^{(l_{out}-1)h} \quad (4)$$
$$\Delta w_{bias}{}_{ij}^{l_{out}} = \lambda \sigma_{ij}^{l_{out}}$$

где $\Delta w_{ijk}^{l_{out}}$ - изменение веса k дендрита, связывающего текущий выходной нейрон и нейрон $A_{qw}^{(l_{out}-1)h}$, $\Delta w_{bias}{}_{ij}^{l_{out}}$ - изменение смещения текущего выходного нейрона, λ – скорость обучения. Ошибка $\sigma_{ij}^{l_{out}}$ передаётся на связанные с текущим нейроны предыдущего слоя. Далее каждый нейрон скрытых слоёв высчитывает свою ошибку как:

$$\sigma_{ij}^{lh} = F'(s_{ij}^{lh}) \sum_{n=1}^{N} \sigma_n^{(l+1)h\prime} w_{ijn}^{(l+1)h\prime} \quad (5)$$

где N – количество связанных с данным нейроном синапсов. Ошибка $\sigma_n^{(l+1)h\prime}$ соответствует нейрону, с которым связан текущий нейрон синапсом n, который имеет вес $w_{ijn}^{(l+1)h\prime}$. Величины изменений весов связи и корректировки смещения высчитывается, соответственно:

$$\Delta w_{ijk}^{lh} = \lambda \sigma_{ij}^{lh} A_k^{(l-1)h} \quad (6)$$
$$\Delta w_{bias}{}_{ij}^{lh} = \lambda \sigma_{ij}^{lh}$$

Результирующий вектор подаётся на вход долговременной-кратковременной памяти для накопления знаний – long short term memory - LSTM. По примеру работа [4, 6], LSTM служит для предиктивного обучения на основе пяти последовательных кадров. На вход LSTM подаётся пять полученных векторов, и данная подсеть экстраполирует эти вектора в шестой, который при развёртке преобразуется в выходное изображение.

Для проведения такого обучения в работе [4] используется метод градиентного спуска, и в качестве функции потери используется сумма среднеквадратичной ошибки и соревновательной ошибки:

$$\sigma_G^{tot} = \sigma_G^{MSE} + \lambda \sigma_G^{AL} \quad (7)$$

При чём соревновательная ошибка рассчитывается исходя из выхода сети-дискриминатора из работ [4, 6] по следующим формулам:

$$\sigma_G^{AL} = \frac{1}{n} \sum_{i=1}^{n} \log(1 - D(G(x_{1:t}^i), x_{1:t}^i)) \quad (8)$$

где $x_{1:t}^i$ - обучающая последовательность кадров, $G(x_{1:t}^i)$ - предсказанное генеративной сетью изображение на основе обучающей последовательности, $D(*, x_{1:t}^i)$ – выход дискриминатора, n – количество кадров в обучающей последовательности.

Для обучения дискриминатора используется функция ошибки:

$$\sigma_D^{AL} = -\frac{1}{2n}\sum_{i=1}^{n}\left[\log\left(D\left(x_{t+1}^i, x_{1:t}^i\right)\right) + \log\left(1 - D(G\left(x_{1:t}^i, x_{1:t}^i\right))\right)\right] \quad (9)$$

Для обычной генерации изображения после обучения непосредственно на вход генеративной части следует подать вектора класса, вида и транформации, которые описаны в [6]. Над вектором проводятся операции развёртки для всех каналов таким образом, что значение выхода нейрона (I, j) слоя l карты свойств h представляет собой

$$Y_{ij}^{lh} = F\left(\sum_{n=1}^{N} w_n Y_n^{(l-1)h}\right) + bias_{ij}^{lh} \quad (10)$$

Или же, если выделить матрицу фильтра, то:

$$Y_{qw}^{(l+1)h} = F\left(w_{nm_k}^{\ \ lh} Y_{ij}^{lh}\right) + bias_{qw}^{(l+1)h} \quad (11)$$

где $w_{nm_k}^{\ \ lh}$ - элемент (n,m) транспонированной матрицы ядра фильтра k.

Анпулинг происходит по формуле

$$Y_{qw}^{(l+1)h} = F\left(w_{ij}{}_{nm}^{\ \ lh} Y_{ij}^{lh}\right) + bias_{qw}^{(l+1)h} \quad (12)$$

где $w_{ij}{}_{nm}^{\ \ lh}$ - весовой коэффициент в карте пулинга для соответствующего нейрона из свёрточной сети. В обычном случае карта анпулинга представляет собой матрицу 2 на 2 с одним элементом, отличным от нуля. Этот элемент проставляется в момент пулинга и указывает на местоположение максимального элемента в пулинге и, соответственно, на местоположение максимального элемента при анпулинге.

На выходе всех четырёх потоков, описанных в [6], получаются матрицы 128 на 128 элементов, которые следует сложить для получения финального изображения. Для получения более понятной картинки в качестве постобработки изображения можно заменить чёрный фон на белый.

Вывод. Таким образом предложенный в данной статье подходя является развитием и в некотором роде совокупностью подходов, предложенных в [4] и [5]. Увеличение количества изображений с разных углов увеличивает возможности к генерации изображений, а применение комбинированного подхода к обучению позволяет добиться более качественного обучения.

Список литературы

[1] Вершинин В.Е., Гришунов С.С., Логинова М.Б. Моделирование процессов распознавания и классификации многомерных объектов пересекающихся классов на основе представлений теории нечётких множеств. Электронный журнал: наука, техника и образование. 2016. № 1 (5). С. 120-133.

[2] Гришанов К.М., Рыбкин С.В. Тестирование свёрточной нейронной сети в задачах машинного зрения. Электронный журнал: наука, техника и образование. 2017. № 2 (12). С. 186-193.

[3] Редько А.В., Молчанов А.Н., Белов Ю.С. Использование алгоритмов определения ключевых точек изображения в задаче реконструкции трёхмерных сцен. Электронный журнал: наука, техника и образование. 2017. № СВ2 (13). С. 59-66.

[4] William Lotter, Gabriel Kreiman & David Cox. Unsupervised learning of visual structure using predictive generative networks. 2016. URL: https://arxiv.org/pdf/1511.06380.pdf

[5] Alexey Dosovitskiy, Jost Tobias Springenberg, Thomas Brox. Learning to generate chairs with convolutional neural networks. CVPR2015.

[6] Шустов И.С. Белов Ю.С. Архитектура генеративной нейронной сети для создания фотореалистичных изображений. // Сборник статей Международной научно-практической конференции «Технологии XXI века: проблемы и перспективы развития», часть 2 (Уфа, 13.06.2017). – Уфа: Аэтерна, 2017. – 162-166 с.