

УДК 004.85

НЕЙРОСЕТЕВАЯ МОДЕЛЬ ГЕНЕРАЦИИ СТИЛИЗИРОВАННЫХ ИЗОБРАЖЕНИЙ

Нарбут А.П., Белов Ю.С.

Калужский филиал федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», г. Калуга (248000, г. Калуга, ул. Баженова, д. 2.), e-mail: narbutwork@mail.ru

В последнее время наблюдаются масштабные изменения в сфере машинного обучения. Данные технологии достаточно широко и разветвлено проникают в науку, образование, экономику, сферы экологической безопасности. Стремительно увеличивается спрос на различные рода систем, основанных на нейронных сетях. Одно из самых перспективных направлений – это обработка и преобразование изображений. Глубокие сверточные сети стали популярным инструментом для создания и восстановления изображений. Ряд научных работ, посвящены преобразованию изображений под определенный стиль, так называемая стилизация изображения. Но большинство существующих решений далеки в своих показателях выполнения от желаемых. Данная статья посвящена результатам спроектированной модели нейронной сети для генерации текстур и стилизации изображений. Нейронная сеть была построена на основе двух подходов: ограничение набора, генерируемых сетями образцов текстур и использование поэлементной нормализации вместо групповой. Выделены достоинства и недостатки представленной нейронной сети, а также приведены результаты работы.

Ключевые слова: стилизация изображений, машинное обучение, нейронные сети, ансамбль Julesz, поэлементная нормализация.

NEUROET NET MODEL OF GENERATION OF STYLIZED IMAGES

Narbut A.P., Belov Yu.S.

Recently, there have been large-scale changes in the field of machine learning. These technologies are widely and branched penetrate into science, education, economics, and environmental safety. The demand for various kinds of systems based on neural networks is rapidly increasing. One of the most promising areas is the processing and transformation of images. Deep convolutional networks have become a popular tool for creating and restoring images. A number of scientific works are devoted to the transformation of images under a certain style, the so-called stylization of the image. But most of the existing solutions are far from their desired performance. This article is devoted to the results of a modeled neural network model for generating textures and image styling. The neural network was built based on two approaches: the restriction of the set generated by the networks of sample textures and the use of elementwise normalization instead of the group one. Advantages and disadvantages of the network are described, as well as the results of the work.

Key words: image stylization, machine learning, neural networks, Julesz ensemble, instance normalization.

Данная статья предлагает рассмотреть результаты спроектированной модели нейронной сети, которая позволит создавать стилизованные изображения более высокого качества.

Существует несколько систем генерации изображений [3], поэтому критическим является момент определения эффективности работы тех систем, которые позволяли бы выбирать наиболее подходящие решения для конкретных задач.

Идея посвящена синтезу текстур и стилизации изображений. Задача синтеза текстур формулируется таким образом: дан образец текстуры и нужно предложить метод, который сможет по данному образцу генерировать новые образцы той же текстуры [6]. То есть задача синтеза текстур во многом упирается в задачу сравнения текстур. На данный момент один из самых качественных подходов – это выделение и сравнение матриц Грэма сгенерированного изображения и образца. Данный процесс был вынесен в отдельную нейросеть, чтобы,

настроив ее параметры, можно было получить желаемый результат. Недостатком данного подхода является долгосрочная стадия обучения.

Оптимизация нейросети производится методом стохастического градиентного спуска, а обучение происходит следующим образом: синтезируем вектор шума, пропускаем его через нейросеть, считаем матрицы Грэма, смотрим расхождения и обратным распространением через весь путь понимаем, какие надо поменять параметры нейросети.[1]

Для случая стилизации изображения нейросеть будет принимать на вход не только вектор шума, но и изображение, которое нужно стилизовать. Сеть обучается для произвольного стиля.

В данной задаче существует множество элементов, от которых зависит качество преобразования. Мы рассмотрим сгенерированные карты признаков у изображения со стилем, стохастическую оптимизацию и показатели энтропии, как основные элементы эффективности генерации изображений.

Для начала стоит рассмотреть архитектуру сети. Имея две архитектуры для генерации изображений, предложенных ранее в [4,5], была выбрана реализация [5] для всех экспериментов по переносу стиля. Нейронная сеть StyleNet с постфиксом BN обозначает групповую нормализацию, а IN -поэлементную.

Для того чтобы иметь достаточно высокие технические и визуальные различия в синтезе текстур мы сравниваем две архитектуры: многогранную полностью сверточную архитектуру [7] (NetV1) и архитектуру, которая была разработана нами (NetV2). NetV2 берет вектор шума размерностью 256 и сначала преобразует его с помощью двух полносвязных слоев, затем результат преобразуется в изображение 4x4 и повторно повышается дискретизация за счет фракционированной (незначительно-чередующейся) свертки.

Для синтеза текстур, аналогично [7], было замечено полезным нормализовать градиент потери текстуры, когда тот проходит через сеть VGG-19. Это позволяет быстро сходиться стохастической оптимизации, но неявно изменяет целевую функцию и требует регулировки параметров.

Стоит заметить, что для текстур с молнией (рис.1) высокий вес энтропии приводит к изменению яркости изображения. Было предположено, что данную проблему можно решить, используя более рациональное расстояние для оценки энтропии или ранее введенное изображение. [2]

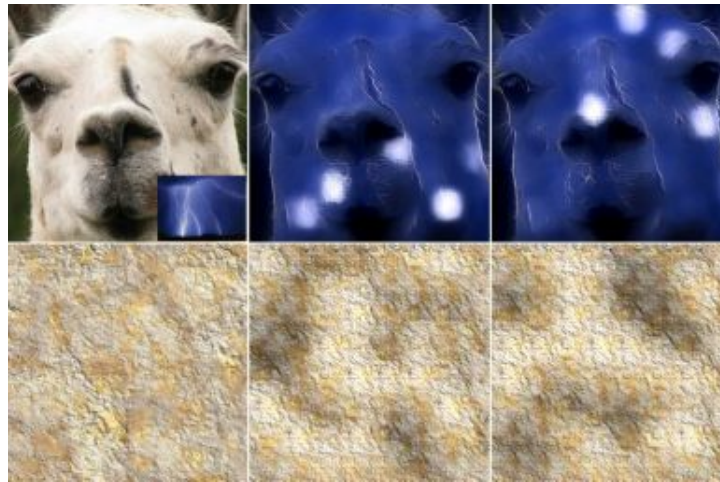


Рис. 1 – Отрицательный пример. Если параметр энтропии слишком высокий для обучающего стиля, генератор сети стремится создавать артефакты (предметы), в которых яркость изменяется локально вместо изменения самой структуры изображения.

Сравнение показателей

Цель обучения для StyleNet IN и StyleNet BN – функция итераций SGD (стохастический градиентный спуск). Целевая функция идентичная, но StyleNet IN сходится намного быстрее. Это подтверждается резким различием качественных результатов (рис.2)

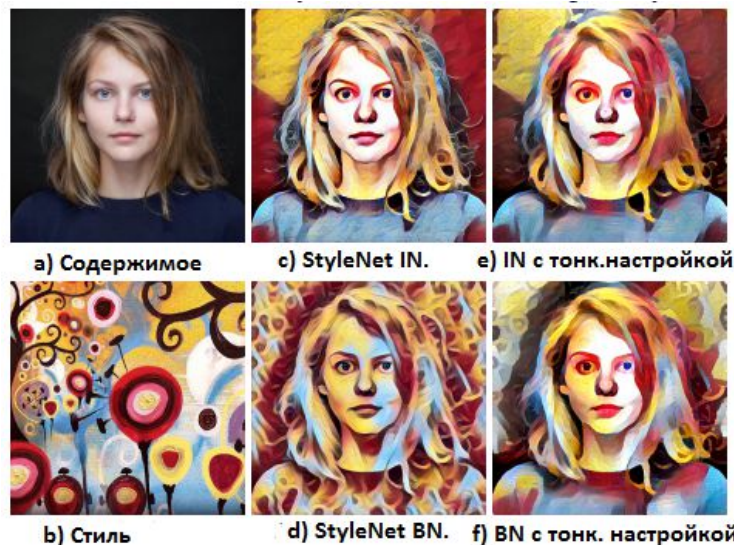


Рис.2 - Сравнение методов нормализации

Поскольку сети StyleNets обучаются минимизировать ошибку расхождения, что и итеративная оптимизация Gatys, они могут быть использованы и в его оптимизационном алгоритме. График (рис.3) показывает применение оптимизации Gatys, начиная со случайной инициализации и вывода двух StyleNets. Становится очевидным: обе сети гораздо ближе к оптимальному результату, чем случайный (Random) и IN точнее, чем BN.

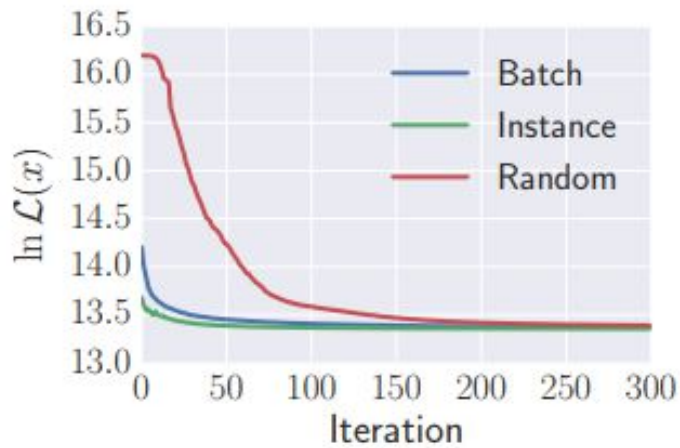


Рис. 3 – Прямая оптимизация Gatys для изображения, представленного выше.

Можно заметить, что разница в качестве изображений велика, так изображения ((e) и (f) рис. 2) показывают изменение StyleNets после тонкой настройки с помощью итеративной оптимизации потери, которая имеет небольшой эффект для изображений IN.

Аналогичные результаты были получены и на других изображениях. Почти везде IN намного превосходит BN, также он ближе к результатам, полученным с помощью гораздо более медленного итерационного метода Gatys.

В то время как поэлементная нормализация работает намного лучше, чем групповая, для синтеза текстур эти два метода работают одинаково хорошо (рис. 4).



Рис. 4 – Результаты генерации текстуры для произвольного шума

Из всего вышесказанного можно сделать вывод, что для формирования изображения, заменяя групповую нормализацию на поэлементную, можно значительно улучшить производительность некоторых глубоких нейронных сетей.

Влияние энтропии

Проверяя архитектуру на основе StyleNet IN, постараемся оценить влияние термина разнообразия на основе энтропии в целевой функции (1).

$$E(g) = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T} \mathcal{L}(g(z_i)) - \lambda \ln \min_{j \neq i} \|g(z_i) - g(z_j)\| \right] \quad (1)11$$



Рис. 5 – Результаты текстурогенерации

Эксперимент (рис. 5) начинается с рассмотрения проблемы генерации текстуры. Необходимо сравнить высокопроизводительные NetV2 и NetsV1 с малой пропускной способностью. Модель с малой пропускной способностью такая же, как [7]. Данная модель использовалась для обучения сети нетривиальным зависимостям от входного шума, тем самым создавая разнообразные результаты, несмотря на то, что задача обучения имеет тенденцию подавлять разнообразие. Результаты эксперименты разнообразны, но неоднозначны, так как иногда изображение на выходе низкого качества. Это следует противопоставить NetV2: ее визуальная точность намного выше, но, используя аналогичную целевую функцию, сеть обучается генерировать одно изображение, как и ожидалось. NetV2 с новой настройкой ($\lambda > 0$) является лучшим из всех вариантов как по качеству, так и по уникальности. [8]

Что касается стилизации, результаты аналогичны тем, которые создаются при синтезе текстур. Разнообразие побуждает сеть создавать разные результаты на основе входного шума. Трудность с нейронной сетью текстур и стилизации состоит в том, что вес потери энтропии λ должен быть настроен для каждой модели обученной текстуры отдельно. Выбрав слишком маленький λ , невозможно обучить разнообразный генератор, а слишком высокий уровень может создать артефакты.

Выводы

На основе подхода по замене групповой нормализации на поэлементную были замечены визуальные изменения качества стилизации в лучшую сторону, хотя скорость в некоторых случаях незначительно упала.

Что касается разнообразия и ансамбля Julesz, то здесь было выявлено следующее: для синтеза текстур важно нормализовать градиент потери текстуры при прохождении через сеть

распознавания. Это позволяет быстрее сходиться стохастической оптимизации, но неявно изменяет целевую функцию и требует корректировки параметров. Также было замечено, что для текстур с равномерной яркостью высокий вес энтропии приводит к изменениям яркости изображения. Было предположено, что эта проблема может быть решена, если будет использована либо более рациональное расстояние для оценки энтропии, либо ранее введенное изображение.

Список использованной литературы:

1. Гришанов К.М., Белов Ю.С. Модель сверточной нейронной сети в задачах машинного зрения. // Наука, техника и образование. 2017. № СВ1. 100 с.
2. Гришанов К.М., Белов Ю.С. Морфологические операции для уменьшения шума на изображении. // Наука, техника и образование. 2016. № 2 (6). С. 90-95.
3. Либеров Р.В., Нарбут А.П. Методы стилизации изображений на основе свёрточной и глубокой нейронной сети // Электронный журнал: наука, техника и образование. 2017. №1(10). С. 27-30. Режим доступа: <http://nto-journal.ru/uploads/articles/025875f0dc01384290ab7ff9df88636a.pdf> (дата обращения: 13.04.2018)
4. Нестеров А.Ю., Белов Ю.С. Распознавание образов по уникальным точкам на примере дорожных знаков // Наука, техника и образование. 2016. № 4 (9). С. 113-119.
5. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS) .2014. С. 2672–2680.
6. J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Computer Vision. // ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II. 2016. С. 694–711.
7. B. Julesz. Textons, the elements of texture perception, and their interactions. // Nature, 290(5802). 1981. С. 91–97.
8. D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In Proceedings of the 33rd International Conference on Machine Learning // ICML 2016, New York City, NY, USA. 2016. С. 1349–1357.