

ГЕНЕРАЦИЯ ТЕКСТУР С ИСПОЛЬЗОВАНИЕМ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

*и. т. н. Бакунова О. М.,
м. т. н. Бакунов А. М.,
м. т. н. Калитеня И. Л.,
Бухта Д. В.
Петрик А. И.,
Лобковская В. В.*

Республика Беларусь, г. Минск, Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники

ARTICLE INFO

Received 05 March 2018

Accepted 19 March 2018

Published 12 April 2018

KEYWORDS

texture synthesis,
convolution,
neural network,
gram matrix

ABSTRACT

Nowadays textures are used in a wide range of tasks such as games or environment emulation software. However, creation of detailed textures is time-consuming. In such cases automatical means of texture synthesise could be used. Here we introduce an approach to synthesise textures based on feature representation of an original image in convolution neural network which takes into account the complex data dependency. An important aspect of the proposed approach is that, if we change a white noise image on the start of a texture synthesise process, the corresponding appearance of the generated textures will also be changed.

© 2018 The Authors.

В настоящее время текстуры используются в широком спектре задач. Основной областью применения текстур является текстурирование окружающей среды в различных демонстрационных, но чаще в игровых программных средствах. Именно благодаря игровым средствам текстуры получают наибольшее развитие в виде увеличения размеров и уровня детализации, так как качество игровой картинке и ее детализация во многом зависит от качества текстур объектов сцены. Однако не во всех приложениях требуется высокая детализация текстур. К примеру, текстуры часто являются фоном для других объектов, в играх, например, этими объектами может быть объемное представление травы, скрывающее за собой мало детализированные текстуры поверхности земли. Малая детализация текстур также применяется на мобильных устройствах, поскольку, мобильные устройства, как правило, обладают меньшей производительностью, а также хранение высокодетализированных текстур в оперативной памяти устройства может оказаться достаточно затратным решением.

Создание даже мало детализированных текстур занимает немало времени. В таких случаях используют автоматические средства генерации текстур. К наиболее широко известным методам относятся процедурная генерация текстур и генерация на основе ближайших соседей. Данные методы генерируют изображения на основе уже существующих изображений. Процедурная генерация текстур и генерация текстур на основе ближайших соседей быстро генерируют простые текстуры, однако данным методам достаточно сложно воспроизвести сложные зависимости в данных, поскольку они генерируют текстуры частями, без знания общей структуры и контекста оригинала.

В последнее время для задач связанных с генерацией правдоподобных изображений все больше используются искусственные нейронные сети. Это связано с тем, что искусственная нейронная сеть рассматривается как универсальный аппроксиматор.

Полносвязная искусственная нейронная сеть представляет собой линейную комбинацию входов сети, при этом сеть разделяется на слои, нейроны каждого слоя связаны со всеми нейронами последующего слоя. Завершает каждый слой нелинейность, или так называемая функция активации. Нелинейность в искусственных нейронных сетях помогает использовать сложную структуру данных для решения поставленных задач. Завершает модель искусственной нейронной сети функция потерь. Функция потерь – это функция которая измеряет степень различия между требуемым ответом и ответом, который выдает нейронная сеть. Функция потерь меняется в зависимости от решаемой задачи, к примеру, для задачи регрессии может использоваться квадратичная функция потерь, а для задач классификации – логистическая функция потерь.

Обучение нейронной сети представляет собой итерационный процесс, в ходе которого выбранный алгоритм оптимизации (чаще всего градиентные методы первого порядка) минимизирует значение функции потерь по отношению к коэффициентам сети.

Свёрточные нейронные сети активно применяются для решения задач, связанных с изображениями. Основой свёрточной нейронной сети является операция свёртки. Операция свёртки в обработке изображений определяется как скалярное произведение подматрицы изображения на матрицу ядра свёртки, при этом размеры матриц равны. Далее подматрица изображения смещается по выбранной оси на определенный шаг, чаще всего шаг равен единице. Операция свёртки повторяется по всем осям изображения. Размер ядра свёртки не ограничен, однако чаще всего используются ядра размером 3x3, 5x5, 7x7, ядра свёртки также называют фильтрами. Свёрточные нейронные сети лучше работают с изображениями, так как операция свёртки подразумевает использование пространственной информации, которую содержат в себе расположение и значение пикселей изображения. При этом общее количество весов свёрточной сети намного меньше количества весов полносвязной сети.

Далее рассматривается метод генерации текстур на основе свёрточной нейронной сети, который учитывает сложные зависимости в данных. Данный метод генерирует текстуру на основе другого изображения, при этом каждая полученная таким образом текстура уникальна.

Для генерации текстур на основе изображения x , изображение x пропускается через свёрточную нейронную сеть, после используются внутренние представления входного изображения (далее «карты активаций») для каждого слоя сети l . Слой с N_l различными фильтрами имеет N_l различных карт активаций, каждая карта активации размера M_l . Карты активаций могут храниться в матрице $F^l \in \mathbb{R}^{N_l \times M_l}$, где F_{jk}^l это карта активаций j -ого фильтра на позиции k в слое l . В картах активаций заложена пространственная информация о содержимом входного изображения, если использовать эту информацию, то сеть фактически каждый раз будет стремиться создать копию оригинального изображения. Для того чтобы не учитывать пространственную информацию используется матрица корреляции между различными картами активаций. Для приближения матрицы корреляций используется матрица Грама $G^l \in \mathbb{R}^{N_l \times N_l}$, где G_{ij}^l это скалярное произведение карт активаций i и j в слое l :

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (1)$$

Матрицы Грама, посчитанные на разных уровнях сети, описывают стиль входного изображения на разных уровнях абстракции.

Для генерации новой текстуры используется изображение из белого шума, после чего, используя метод оптимизации, представление изображения белого шума в матрицах Грама приближается к представлению оригинального изображения в матрицах Грама. Это приближение представляет собой минимизацию среднеквадратической ошибки между матрицами Грама оригинала и изображением белого шума.

Пусть \vec{x} – оригинальное изображение, а $\hat{\vec{x}}$ – изображение генерируемой текстуры, и G^l и \hat{G}^l – соответствующие представления в матрицах Грама в слое l .

Вклад слоя l в функцию потерь представляет собой:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{i,j}^l - \hat{G}_{i,j}^l)^2 \quad (2)$$

Функция потерь:

$$L(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^L w_l E_l \quad (3)$$

где w_l – вес ошибки E_l в итоговой функции потерь, w_l подбирается опытным путем.

Таким образом нейронная сеть пытается приблизить стилевое содержание генерируемой текстуры к оригинальному изображению. Однако полученное изображение при использовании функции потерь из уравнения 3 получается «зернистым». В настоящих изображениях значение одного пикселя сильно коррелирует со значениями соседствующих. Используя это знание добавим в функцию потерь дополнительное слагаемое:

$$S(x) = \sum_c^C \sum_h^H \sum_w^W (2x_{h,w,c} - x_{h,w+1,c} - x_{h+1,w,c})^2 \quad (4)$$

где $x_{h,w,c}$ – пиксель в h строке, w столбце канала c матрицы изображения x .

H – высота изображения

W – ширина изображения

C – количество каналов (в цветных изображениях как правило равно 3)

Итоговая функция потерь:

$$L(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^L w_l E_l + \alpha S(\hat{\vec{x}}) \quad (5)$$

где α – коэффициент сглаживания, α подбирается опытным путем.

В данной работе для выделения карт активаций используется свёрточная нейронная сеть VGG19. Для получения представлений следует выбрать слои, чем дальше слой от входного слоя, тем более абстрактные представления в нем содержатся, также при выборе слоев важно учитывать, что на более «глубоких» слоях теряются мелкие детали изображения. Для этой работы были выбраны следующие слои: «conv1_1», «conv2_1», «conv3_1», «conv4_1», «conv5_4». В качестве алгоритма оптимизации был выбран алгоритм адаптивного градиентного спуска Adam. Градиентный спуск осуществлялся в течении 300 итераций. Скорость обучения $\lambda = 0.02$. Коэффициент сглаживания $\alpha = 1.5$. Веса ошибок $w_l = 0.1/L$. Использовалось отсечение значений градиента на интервале $[-1, 1]$. Начальные изображения белого шума генерировались из равномерного распределения. Некоторые результаты представлены на рисунке 1.

Генерация одного изображения размером 110x110 пикселей, на процессоре Intel Core i7-6700 занимает около 1 минуты.

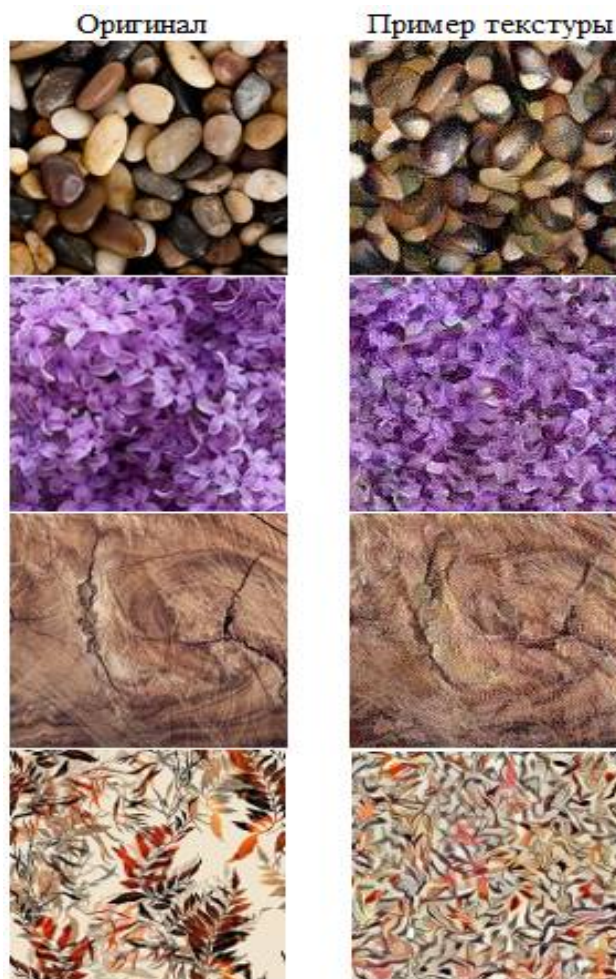


Рис. 1. Пример оригинальных изображений и текстур на их основе

ЛИТЕРАТУРА

1. M. Cimpoi, S. Maji, and A. Vedaldi. Deep convolutional filter banks for texture recognition and segmentation
2. J. Portilla and E. P. Simoncelli. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. International Journal of Computer Vision, 40(1):49–70, October 2000.
3. K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition