

Модели переноса стиля

Практикум на ЭВМ 2017/2018

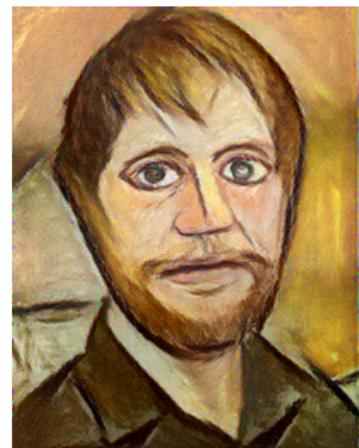
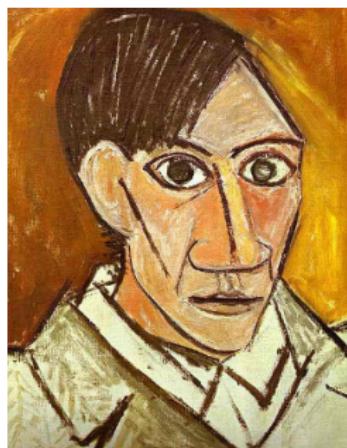
Дженаков Дмитрий Даниилович

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

17 мая 2018 г.

Коротко о задаче

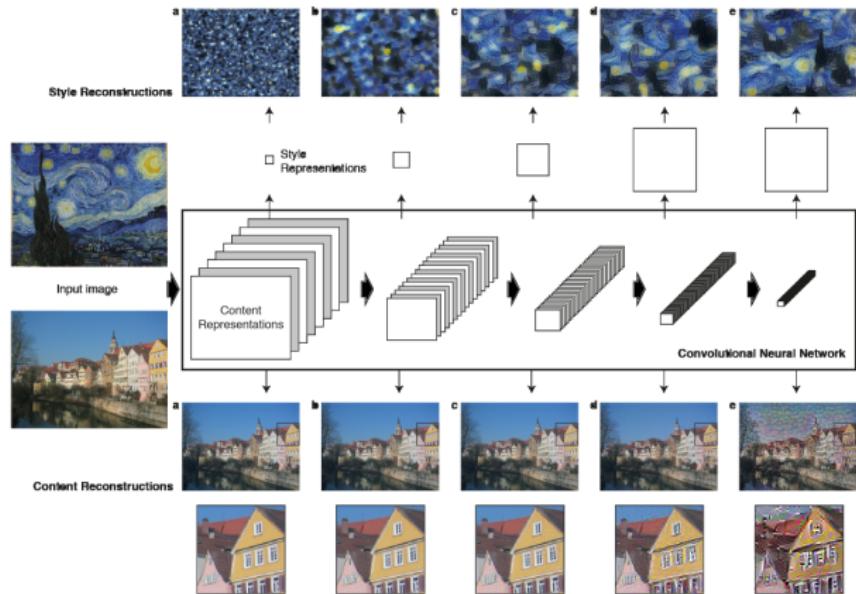
- Изображение представляется в виде композиции содержания и стиля
- Необходимо перенести стиль одного из изображений на другое, сохранив содержание



Общая идея модели Neural style

- Сверточная нейронная сеть работает с содержанием изображения
- Разделимость содержания и стиля
- Описать стиль как корреляцию различных фильтров на каждом слое
- Выбрать изображение, наиболее близкое исходным по стилю и содержанию

Иллюстрация модели



VGG - сеть

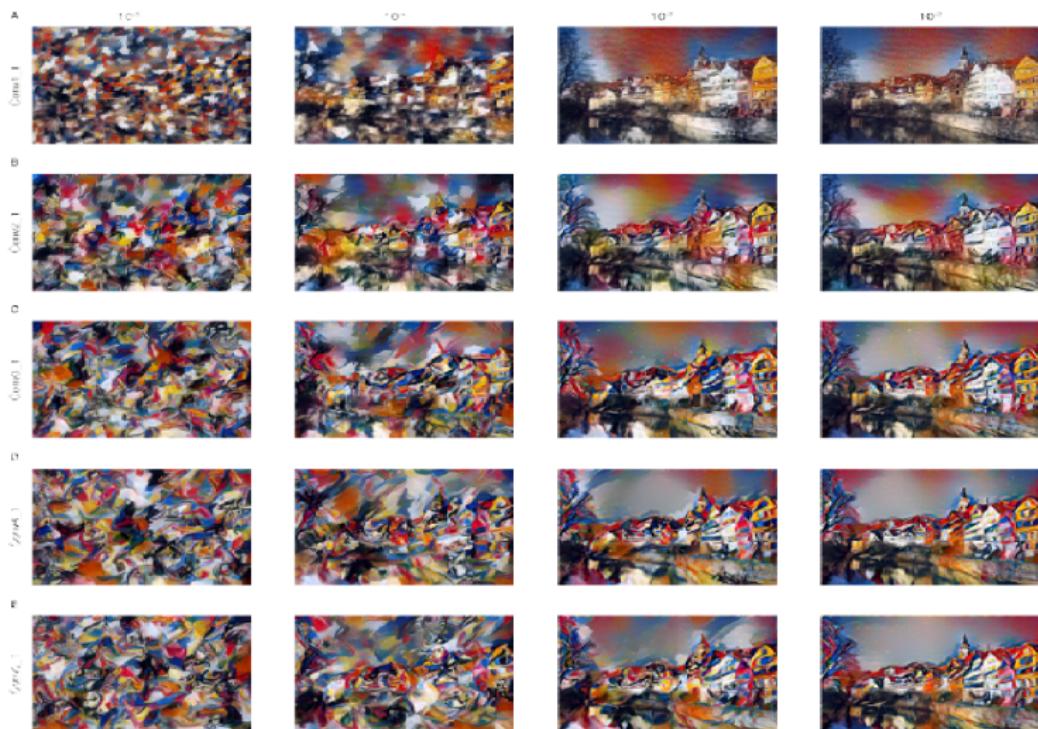
Данный вид сверточной нейросети используется во всех описанных моделях.



Детали реализации

В качестве основы используется VGG - сеть. Для синтеза изображения используется average-pooling При обучении производится градиентный спуск по функции потерь, равной сумме функций потерь по содержанию и по стилю, взятых в некотором отношении.

Влияние параметров на результат



Функция потерь содержания

N^l - количество фильтров в слое l , M^l - размер карты признаков на слое l . F^l - матрица размера $N^l \times M^l$, F_{ij}^l - значение функции активации i -го фильтра на позиции j . p, x - исходное и сгенерированное изображение, P^l, F^l - их признаковые представления на слое l . Тогда функция потерь содержания на слое l будет равна

$$L_{content}(p, x, l) = \frac{1}{2} * \sum_{i,j} (F_{i,j}^l - P_{i,j}^l)^2$$

Суммарная функция потерь содержания:

$$L_{content} = \sum_{l=0}^L w_l L_{content}(p, x, l)$$

Функция потерь стиля

G^l - матрица Грама слоя l , $G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$ Это - представление признаков стиля на данном слое. a, x - исходное и сгенерированное изображение, A^l, G^l - их матрицы Грама на слое l . Тогда вклад слоя в общую функцию потерь будет равен $L_{style}(p, x, l) = \frac{1}{4N_l^2 M_l^2} * \sum_{i,j} (G_{i,j}^l - A_{i,j}^l)^2$

Суммарная функция потерь содержания:

$$L_{content} = \sum_{l=0}^L w_l L_{style}(p, x, l)$$

Общая функция потерь:

$$L_{total} = \alpha * L_{content} + \beta * L_{style}$$

Улучшение модели Neural style

- Разработано компанией Adobe совместно с Корнелльским университетом
- Цель - сохранить фотorealизм исходного изображения
- Метод - добавление регуляризации

Регуляризация для фотorealизма

Идея регуляризации - ограничивать изменения изображения. В идеале переходы должны быть аффинными. Регуляризация идейно основана на алгоритме matting'a. Из этого алгоритма берется аффинное выражение grayscale - маски через каналы RGB.

M_I - матрица размера $N \times N$, $V_c[O]$ - вытянутый в вектор канал c изображения O .

$$L_m = \sum_{c=1}^3 V_c^T[O] M_I V_c[O]$$

Функция потерь имеет вид

$$L_{total} = \alpha * L_{content} + \beta * L_{style} + \lambda * L_m$$

Сравнение трех моделей



Исходное изображение

Изображение для стиля

Neural Style result

CNNMRF result

Adobe result

Использованные материалы

- <https://arxiv.org/pdf/1508.06576.pdf>
- <https://arxiv.org/pdf/1601.04589.pdf>
- <https://arxiv.org/pdf/1703.07511.pdf>
- <http://webee.technion.ac.il/people/anat.levin/papers/Mathematics.pdf>