

A Neural Algorithm of Artistic Style

Сверточные Глубинные нейронные сети являются наиболее мощной моделью в обработке и анализе картинок. Сверточные нейронные сети состоят из слоев выполняющих небольшие вычисления обрабатывающие визуальную информацию иерархически от первого слоя к последнему. Каждый слой может представлять из себя набор фильтров, каждый из которых достает конкретную фичу из входящего изображения. Таким образом выход каждого слоя состоит из мапа фич – поразному отфильтрованных вариантов входного изображения.

Когда сверточная нейронная сеть обучается на распознавание изображений, она собирает некоторое представление изображения, которое делает информацию об объекте более и более явной в иерархическом процессе обработки. Поэтому, в течении иерархического процесса обучения нейронной сети, входящее изображение трансформируется в некоторое представление которое в первую очередь хранит в себе "содержание" изображения, а не детальное значение пикселей. Мы можем напрямую визуализировать трансформированное изображение с каждого уровня восстановив изображение только с мапа фич с данного слоя (Fig1 content reconstruction). Высшие слои сети извлекают из изображения высокоуровневую – более обобщенную информацию о содержании (контенте) изображения, о расположении объектов на входном изображении, практически без какой либо привязки к конкретным значениям пикселей на исходном изображении. В отличие от изображения восстановленного с нижних слоев в которых не сложно воспроизвести практически точные значения пикселей исходного изображения.

Что бы получить представление о стиле исходного изображения, надо использовать признаковое пространство, изначально предназначенное для сбора информации про текстуру изображения. Данное пространство построено на выходах фильтров с каждого слоя сети и состоит из корреляции между различными выходами фильтров. Включая корреляцию между слоями мы получаем стационарное, многомасштабное представление входного изображения, захватывающего текстуры.

<Пичка с сетью>

Мы можем визуализировать информацию полученную из этого признакового пространства текстуры изображения построенного с разных слоев сети путем построения изображения, которое соответствует стилю конкретного входного. Действительно, реконструкция стиля производит новую версию исходного изображения которое захватывает основные особенности стиля с точки зрения структур и цвета.

Основной вывод данной статьи заключается в том, что представление о содержании и стиле изображения в сверточной нейронной сети делимо. То есть мы можем манипулировать обоими представлениями независимо для получения новых изображений. Изображения получаются с помощью поиска картинки одновременно соответствующей по содержанию первому изображению и по стилю второму.

Анализ третьей пикчи

Методы построения:

Все приведенные в статье результаты были получены с использованием WGG-Network. Признакомое пространство было получено с помощью 16 сверточных слоев и 5 пулингов данной сети, состоящей из 19 слоев. Полносвязная часть сети не использовалась. Max пулинг был заменен на average пулинг для улучшения градиентного спуска и получения более явного результата.

Каждый слой сети представляет из себя набор нелинейных фильтров сложность которых увеличивается вместе с номером слоя сети.

Пусть входное изображение x было получено пропуская через слои сети. Слой l с N_l различных фильтров имеет N_l мапов фичей каждый размера M_l значит отвеы слоя l могут быть сохранены в

матрице $F \in R^{N_l \times M_l}$, где F_{ij}^l выход i -ого фильтра на j -ом слое.

Для визуализации информации с каждого слоя мы будем запускать градиентный спуск на белом шуме что бы найти изображение которому соответствуют такие же выходы слоев (мары фич)

Пусть p, x оригинальная картинка и сгенеренная, тогда P^l, F^l их представления на l слое тогда определим квадратичную ошибку между двумя представлениями и ее производную с помощью которого градиент будет сходиться к исходному изображению x используя квадратичную ошибку. Пример на первой пикче

При построении пространства представлений на каждом уровне мы находим корреляцию между различными выходами фильтров с помощью построения матрицы Грама $G^l \in R^{N_l \times N_l}$, где G_{ij}^l – корреляция между векторизованными мапами фич i и j на слое l

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$$

Для генерации текстуры которая соответствует стилю исходного изображения мы будем использовать градиентный спуск с белого шума для поиска другого изображения которое имеет такой же стиль. Будем делать это с помощью минимизации среднеквадратичного отклонения матрицы грама от оригинального. Пусть a, x оригинальное изображение и белый шум, а A^l, G^l представления на слое l . Тогда определим ошибку для конкретного слоя и общую ошибку, где w это вес вложения данного слоя в исходную ошибку. определим производную.

Для генерации изображения которое соответствует контенту одного изображению и стилю другого мы будем минимизировать расстояния белого шума к первой картинке по контенту и к второй картинку по стилю