УДК 519.2, УДК 519.6, ГРНТИ 28.23.15: РАСПОЗНАВАНИЕ ОБРАЗОВ. ОБРАБОТКА ИЗОБРАЖЕНИЙ

**Программная реализация нейросетевых алгоритмов для переноса худождественных стилей в изображениях.**

***Бендер С.А.,*** канд. техн. наук, доцент

***Горбов Г. В.*,** студент

Ижевский государственный технический университет им. М.Т. Калашникова

***Аннотация***

В данной работе поставлена задача переноса художественного стиля с одного изображения на другое, с первичной сегментацией исходного изображения. Предложена математическая интерпретация понимания стиля изображения и схожести стилей в двух изображениях.

***Ключевые слова****:* нейросети, сверточные нейронные сети, перенос стиля, сегментация, машинное обучение.

Введение

Задачи и предложенные к ним алгоритмы CV(computer vision) области машинного обучения имеют в нынешнее время огромную популярность. Такие алгоритмы очень часто применяются в различных междисциплинарных областях в которых предполагается, что алгоритмы должны обладать человеческой интеллектуальностью и творчеством. В ходе работы были предложены математические определения таких интуитивных понятий как ‘стиль изображения’, вычисление расстояний между стилями изображений. В ходе данной работы поставлена задача сегментации, переноса художественного стиля. Произведено обучение разработанных моделей, при помощи средств Google Colab.

# Математическая постановка задачи

# Математическая постановка работы состоит из двух частей. Первая часть – это математическая постановка задачи сегментации. Вторая часть – это математическая постановка задачи переноса стиля.

**Математическая постановка для задачи сегментации**

Пусть имеется множество изображений каждый элемент которой является (RGB изображением), а так же множество ответов *Y* где , каждый ответ является тензором, каждый канал которого отвечает за свой класс. Пусть так же имеется функция

Требуется по заранее заданной выборке определить функцию

Которая бы наилучшим образом аппроксимировала функцию *f* При помощи рассмотрения семейства функций где , (*m,n* – размер входных изображений RGB) и нахождения , которая бы разделяла пиксили входного изображения *x* в соответствии с заранее заданными классами. Где поиск точки осуществляется при помощи определения функции ошибок , от параметров и матриц и решения задачи где *Q* – эмпирический риск.

**Математическая постановка для задачи переноса стиля**

Требуется по заданному изображению составить алгоритм определения из множества параметрических моделей:

где (*m,n* – размер входных изображений RGB)

Функции

, которая бы наилучшим образом аппроксимировала функцию S.

Где поиск точки осуществляется при помощи определения функции ошибок *L* и решения задачи минимизации

## Выбор модели для решения задачи сегментации

В качестве модели для решения задачи сегментации были рассмотрены уже известные алгоритмы, которые приведены в следующей таблице.

|  |  |  |
| --- | --- | --- |
| Network | mean IoU | global pixelwise acc |
| FCN ResNet50 | 60.5 | 91.4 |
| FCN ResNet101 | 63.7 | 91.9 |
| DeepLabV3 ResNet50 | 66.4 | 92.4 |
| DeepLabV3 ResNet101 | 67.4 | 92.4 |
| DeepLabV3 MobileNetV3-Large | 60.3 | 91.2 |
| LR-ASPP MobileNetV3-Large | 57.9 | 91.2 |

Таблица 1. Модели семантической сегментации

Предлагается использование модели ResNet101. Так как она удовлетворяет временным потребностям решения задачи и выдает приемлемый результат.

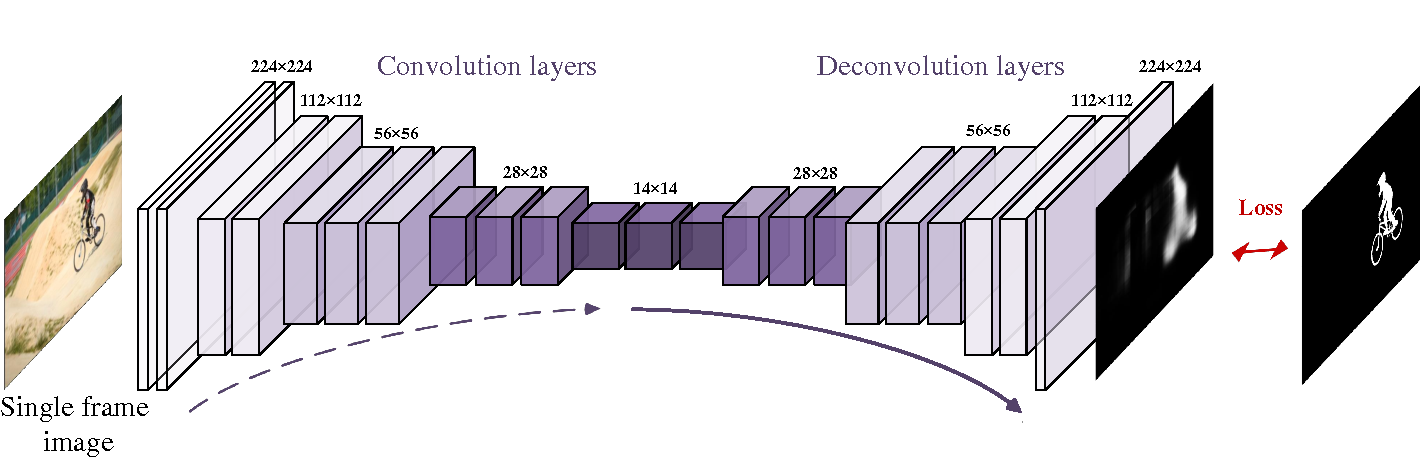
****

Рисунок 1. Архитектура ResNet101

## Функция потерь

Функция ошибки L будет состоять из двух отдельных компонент, а более конкретно представляться следующей формулой.

Так, мы должны будем следить за тем, чтобы наше изображение не отставало от исходного стилевого изображения по стилю и от исходного контент изображения по контенту.

Теперь дадим математическое определение понятия стиля изображения. Согласно описанному выше принципу распознавания структур в изображениях, которые производят Convolutions neural layers, мы имеем, что при рассмотрении произвольного слоя сети, предположим это будет слой переводящий 64 карт активации в 128 карт. Тогда каждая из 128 карт – это будет матрица, которая реагирует на проявление какого – либо конкретного признака в изображении. Поэтому для удобства преобразуем тензор с количеством каналов 128 в матрицу, путем развертывания каждой матрицы в вектор, а затем составлением из них двумерной матрицы. Затем при помощи стандартного скалярного произведения между векторами в Евклидовом пространстве, будем находить корреляцию между каждой парой признаков, для обнаружения того, насколько эти признаки одновременно часто встречаются на нашем изображении. По итогу этой операции будет получена так называемая Грамм матрица(скалярных произведений).

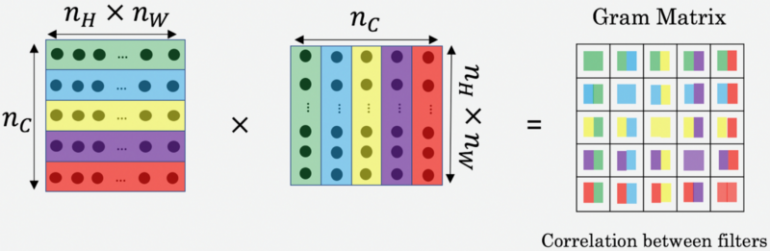
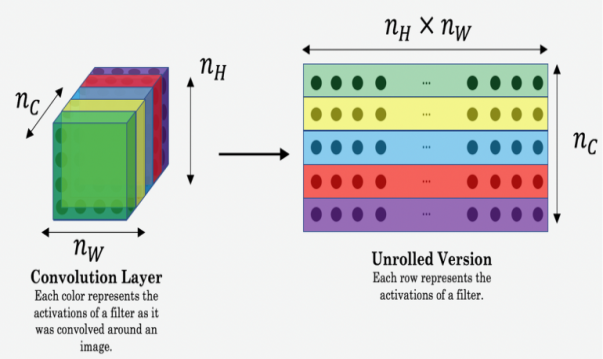


Рисунок 2. Преобразование карт активаций с произвольного слоя и получение Грамм матрицы

Собственно именно Грамм матрица и будем отвечать за стиль произвольного изображения – совокупность корреляции признаков, которые встречаются в этом изображении. Теперь остается лишь указать то как будет определяться схожесть между стилями изображений. Получаем итоговую формулу.

Расстояние между контентом двух изображений , будем определять следующим обазом.

Где *a(C), a(G) –* результаты преобразования в двумерные матрицы выхода сверточного слоя для *input* и *style* изображений.

И того добавляя параметры , которые выступают в качестве гиперпараметров модели и которые можно изменять, чтобы определять вклад каждой из компонент в общее значение функции потери.

## Интеграция сегментации в алгоритм переноса стиля

При помощи алгоритма сегментации получаем тензор глубины 21 с матрицами размера , который затем преобразуем в 2D изображение, путем изменения каждого элемента заранее взятой матрицы на соответствующий наибольший элемент по каждому из каналов. *.* Затем заменяя каждый не нулевой элемент матрицы MASK на 1, получаем решение задачи бинарной классификации, то есть матрицу, которая будет отвечать за разделение обьекта изображения от фона.

Так как в общем случае функция потерь имеет вид

Соответствующее вычисление следует изменить так, чтобы стиль на исходное изображение *input*, инициализируемое изображением content, накладывался только на его часть, а именно та матрица *MASK*, которая получается по итогу сегментации будет участвовать в роли шаблона. Таким образом получаем формулы

Приведем визуализацию процесса.

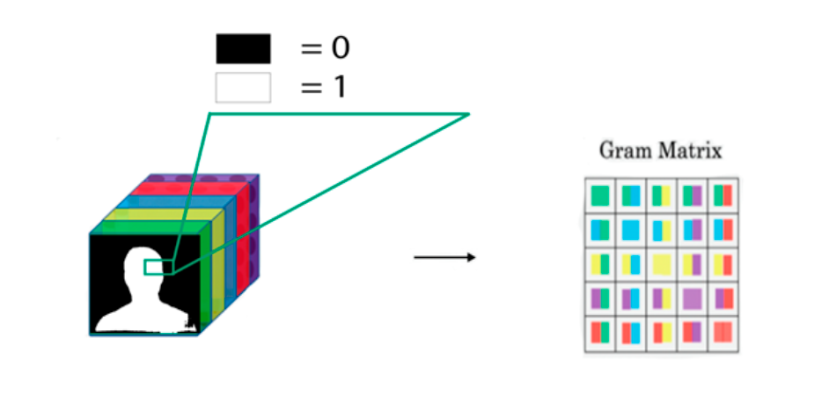


Рисунок 5. Преобразование карт активаций с произвольного слоя

Далее как и в исходном алгоритме происходит вычисление функции потерь, но уже на новой матрице Грамма, и решается задача оптимизации

Где параметры функции служат пиксели изображения input.

По полученной точке минимума строится изображение, которое и будет соответствовать переносу стиля на фон либо на обьект изображения таким образом, что стиль полученной картинки наилучшим образом схож со стилевым изображением в смысле матрицы Грамма

**Определение гиперпараметров модели**

В этой части работы опишем то, как именно мы будем изменять выбранный нами алгоритм классификатора, который нам необходимо адаптировать под нашу задачу. Начнем с указания гиперпараметров модели, которые были получены как результат многократных обучений модели с различными вариациями этих параметров. Первый предложенный гиперпараметр это произведение над архитектурой классификатора VGG19 следующих изменений: вначале избавимся от той части нейронной сети которая содержит перцептроны аналогичные простой полносвязной сети. Визуально теперь модель можно представить следующим образом.

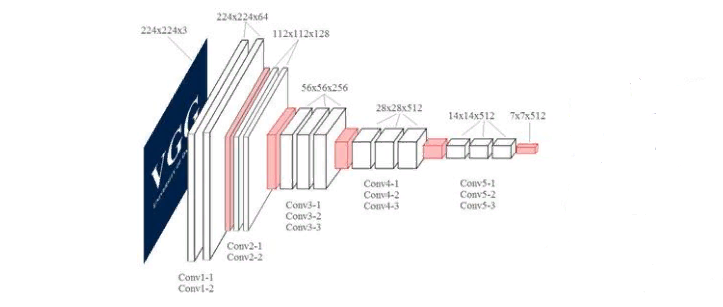


Рисунок 6. Измененная модель на основе архитектуры VGG19

Далее будем выбирать те слои нейронной сети с которых будет высчитываться ошибка по стилю и по контенту () conv1, conv2, conv3, conv4, conv5. Слой для потери по контенту – conv4.

Следующая группа гиперпараметров – это те которые будут задействованы непосредственно во время обучения.

Параметры *,* которые отвечают завклад ошибки по стилю и по контенту будут равны 1000000 и 100. Так как исходный холст инициализируется контент изображением, то параметр отвечающий за потерю по контенту будет принимать довольно малое значение, так как его вклад в итоговую ошибку не существенен. Количество итераций обучения num\_steps полагаем равным 250. В качестве метода оптимизации целевой функции предлагается использовать алгоритм LBFGS(Алгоритм – Бройдена – Флетчера – Гольдфарба - Шанно).

## Результаты работы алгоритма

Реализуем алгоритм с описанным набором гиперпараметров, который представляет наилучший его результат. Рассмотрим работу алгоритма с входными данными в виде следующих двух изображений.

****

Рисунок 7. Выходные контент-изображение и изображение – стиль.

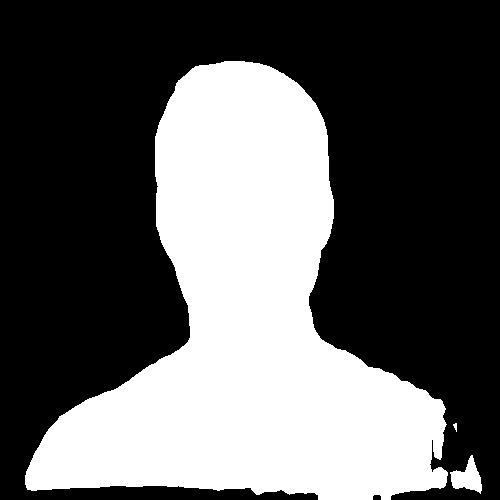


Рисунок 8. Сегментация мужчины, перенос стиля на фон

Время работы алгоритма - 52.6 с.

**Заключение**

В ходе данной работы определена постановка задачи нейросетевого переноса художественного стиля в изображениях. Описана методика решения задачи, а так же дана математическая интерпретация существенных для этой задачи понятий таких как стиль изображения, схожесть по стилю. Был предложен алгоритм решающий как задачу семантической сегментации изображения так и задачу переноса стиля. В качестве конечной конфигурации алгоритма были выбраны модели ResNet101,Vgg19, метод оптимизации – Алгоритма – Бройдена – Флетчера – Гольдфарба – Шанно, число итераций обучения –300.

Время работы алгоритма по двум входным изображениям в среднем составляет 52 с.

**Литература**

1. Курс продвинутого потока Deep learning School 1 часть.
2. URL: <https://towardsdatascience.com/introduction-to-neural-style-transfer-with-tensorflow-99915a5d624a> intoduction to Neural Style Transfer with TensorFlow.
3. Neural Network in Market Segmentation: A Descriptive study of Market Segmentation using Artificial Neural Networks Paperback – June 2, 2017
4. NEURAL NETWORKS A Comprehensive Foundation Second Edition. Simon Haykin
5. A Neural Algorithm of Artistic Style URL: <https://arxiv.org/pdf/1508.06576.pdf>