

"Распознавание текста на основе скелетного представления линий с шириной и сверточных сетей"

Н. А. Рябов¹, И. А. Рейер², В. И. Петров¹

¹Московский Физико-Технический Институт (национальный исследовательский университет)

²Федеральный исследовательский центр «Информатика и управление» РАН

В текущий момент в области машинного зрения все популярнее становятся “глубокие” сети, способные замечать самые разные закономерности. Но проблема такого подхода заключается в необходимости больших вычислительных затрат не только для обучения, но и для применения. В данной работе предлагается метод распознавания текста, основанный на графовом представлении линий с учетом их ширины и обработке с помощью сверточных графовых сетей. Представление графа занимает меньше памяти и более содержательно, чем картинка, поэтому ожидается, что графовые сети будут сильно оптимальнее по потребляемым ресурсам, чем стандартные сверточные на картинках.

Для работы мы выбрали датасет MNIST, содержащий одноканальные изображения цифр и в качестве базовой модели выбрали LeNet5, содержащую около 60000 обучаемых параметров.

В ходе работы мы сначала бинаризовали изображения – сделали, чтобы значение каждого пикселя было либо 0, либо 1. Далее мы воспользовались алгоритмом скелетизации, описанном в статье [1]: сначала каждая “компонента связности” изображения аппроксимируется многоугольником. Далее в этот многоугольник вписываются окружности максимально возможного радиуса и центры этих окружностей образуют начальные вершины графа. После этого проводится удаление шумовых вершин. Таким образом мы получили скелетное представление символа: граф, у которого признаки вершины: координаты и радиус окружности.

Далее я реализовали архитектуру нейронной сети, предложенную в статье [2] и сравнил метрики. Архитектура сети состоит из семи спектральных графовых сверток, на выход идут признаки добавленной в начале фиктивной вершины. Также я сравнил архитектуры с разными свертками: пространственными и спектральными: пространственная агрегирует признаки соседних вершин, а спектральная использует матрицы, описывающие общую структуру графа. При этом обе этих свертки не изменяют количество вершин и матрицу смежности, поэтому я попробовал разные подходы унификации выхода сетей и сравнил метрики: добавление фиктивной вершины, выход которой будет считаться выходом сети, различные пулинги по всем вершинам, добавление фиктивных вершин в граф до достижения определенного их числа и т.д.

При этом в нашей работе было важно, чтобы размер сети не превышал размера базовой модели, ведь задачей поставлена оптимизация архитектуры и мы хотели либо показать лучшее или равное качество при меньшем размере модели.

Литература

[1] Mestetskiy L., Semenov A. Binary Image Skeleton - Continuous Approach // Proceedings 3rd International Conference on Computer Vision Theory and Applications, VISAPP 2008. P. 251-258

[2] Lomov, N. and Arseev, S.: NEURAL NETWORKS FOR SHAPE RECOGNITION BY MEDIAL REPRESENTATION, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W12, 137–142,
<https://doi.org/10.5194/isprs-archives-XLII-2-W12-137-2019>, 2019.