

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

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

¹Московский Физико-Технический Институт (национальный исследовательский университет)

²Федеральный исследовательский центр «Информатика и управление» РАН

С ростом популярности глубоких нейросетевых архитектур в задачах компьютерного зрения всё более актуальной становится проблема их высокой ресурсоёмкости. Особенно остро она проявляется при применении таких моделей в условиях ограниченных вычислительных мощностей. В данной работе исследуется альтернативный способ обработки изображений, основанный на переходе от растрового к графовому представлению, что может позволить существенно сократить объем памяти и ускорить инференс.

Для исследования использовался датасет MNIST, содержащий изображения рукописных цифр. После бинаризации изображения были преобразованы в графовую форму с помощью алгоритма, описанного в статье [1]. Его суть заключается в аппроксимации компонент связности многоугольниками и построении вписанных окружностей наибольшего радиуса. Центры этих окружностей образуют вершины графа, а признаки включают координаты и радиус окружности. Далее производится регуляризация с удалением излишних вершин.

На построенных графах обучалась модель, основанная на архитектуре из статьи [2], включающей семь спектральных графовых свёрточных слоёв. Особое внимание было уделено методам агрегации признаков: использовались фиктивные вершины, глобальные пулинги, а также добавление дополнительных вершин для выравнивания графовой структуры. Помимо спектральных, рассматривались и пространственные свертки, агрегирующие информацию от соседей в пределах локальной окрестности.

В качестве базовой модели для сравнения использовалась LeNet-5 с $\approx 60\,000$ обучаемыми параметрами. Ключевым критерием оценки было достижение более высокого качества классификации при том же числе параметров, либо сохранение качества 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.