

# Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Литовченко Л.А.

Московский физико-технический институт (Государственный университет)  
litovchenko.la@phystech.edu

**Аннотация** Данная работа посвящена исследованию применения скелетного представления растрового изображения в задаче классификации печатных символов. Сравнивается 3 подхода к решению задачи: классический с использованием свёрточных нейросетей небольшой глубины, с построением скелета изображения и кодировкой его в вектор, и их комбинация.

*Ключевые слова: сверточные нейронные сети, распознавание текста, скелетное представление, graph embedding*

## 1 Введение

Распознавание растровых изображений уже давно успело стать классической задачей компьютерного зрения. Сюда же относится и распознавание изображений с символами. Каждый год публикуются научные работы, улучшающие и без того высокое качество классификации (1) (2). Традиционным подходом к проектированию архитектур для решения таких задач является использование свёрточных слоев (3) (4). Обычно, входом таких алгоритмов является растровое изображение, однако, существуют подходы, которые производят построение некоторой графовой структуры путём скелетизации изображения с последующим преобразованием в вектор и обучении на нём.

Скелетизация представляет из себя выделение графа изображенного объекта, форма которого может приближённо описать изначальный объект. В общем случае это можно представить как заполнения внутренностей объектов кругами, центры которых будут являться вершинами скелетного графа и будут соединяться ребрами графа. В литературе можно найти различные статьи, посвященные вопросам скелетизации: (5) описывает генерацию рукописного текста при помощи толстых линий; (6) сравнивает формы изображений при помощи скелетизации; (7) описывает альтернативные подходы к построению графа.

Получаемый после скелетизации граф объекта не позволяет непосредственно использовать его в существующих архитектурах нейронных сетей. В связи с этим необходимо произвести векторизацию графа с помощью его эмбединга.

Для проведения экспериментов использовался успешный стандарт классическим датасет MNIST (8). Он состоит из рукописных цифр, и большое количество работ, в том числе и современных, используют данную выборку для валидации и анализа архитектур (9) (10) (11).

## Литература

- [1] Zou, Xianli Fast Convergent Capsule Network with Applications in MNIST—Advances in Neural Networks—ISNN 2018—Springer International Publishing—pp. 3–10
- [2] Palvanov A., Im Cho Y Comparisons of deep learning algorithms for MNIST in real-time environment—International Journal of Fuzzy Logic and Intelligent Systems—2018. – Т. 18. – №. 2. – С. 126-134.
- [3] LeCun Y. et al. Convolutional networks for images, speech, and time series //The handbook of brain theory and neural networks. – 1995. – Т. 3361. – №. 10. – С. 1995.
- [4] Ciresan D. C. et al. Convolutional neural network committees for handwritten character classification //Document Analysis and Recognition (ICDAR), 2011 International Conference on. – IEEE, 2011. – С. 1135-1139.
- [5] Клименко С. В., Местецкий Л. М., Семенов А. Б. Моделирование рукописного шрифта с помощью жирных линий //Труды. – 2006. – Т. 16.
- [6] Кушнир О. и др. Сравнение формы бинарных растровых изображений на основе скелетизации //Машинное обучение и анализ данных. – 2012. – Т. 1. – №. 3. – С. 255-263.
- [7] Масалович А., Местецкий Л. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений //Труды Международной конференции «Графикон», Новосибирск.–2006.–4 с.
- [8] LeCun Y., Cortes C., Burges C. J. MNIST handwritten digit database // Available: <http://yann.lecun.com/exdb/mnist>. – 2010. – Т. 2.
- [9] Zhu D. et al. Negative Log Likelihood Ratio Loss for Deep Neural Network Classification //arXiv preprint arXiv:1804.10690. – 2018.
- [10] Nair P., Doshi R., Keselj S. Pushing the limits of capsule networks //Technical note. – 2018.
- [11] Hsieh P. C., Chen C. P. Multi-task Learning on MNIST Image Datasets. – 2018.