

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Тушин К.А.

Московский физико-технический институт (Государственный университет)
tushin.ka@phystech.edu

Аннотация В работе рассматриваются подходы решения задачи распознавания символов. Один из методов использует сверточные сети для классификации изображений. Другой метод заключается в анализе графовых структур с помощью скелетного представления, полученных по изображению. Так же приведены сравнения точности этих подходов и архитектур на датасетах MNIST и Chars74K.

Ключевые слова: сверточные нейронные сети, CNN, распознавание символов, скелетное представление, Graph embedding

1 Введение

Распознавание символов это классическая задача компьютерного зрения. Основным подходом в таких задачах это использование сверточных слоев в нейросетях (1) (2). В таких нейросетях на вход подается изображение а на выходе получаем вероятность принадлежности изображения к каждому классу. Существует другой подход, при котором растровое изображение переводится в векторное представление путем построения скелета символа, а потом подается на вход модели для предсказания к какому классу принадлежит изображение.

Скелетизация представляет собой процесс заполнения внутренностей символов кругами, центры которых - вершины графа, соединяются с ребрами графа. Было проведено много много исследований на эту тему. В работе (3) обсуждается моделирование рукописного текста с помощью жирных линий. В работе (4) проводится сравнение формы бинарных растровых изображений на основе скелетизации. Однако, существует и альтернативные методы, как, например описанный в работе (5). В этой работе главной задачей было распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений.

Для оценки качества работы алгоритма использовалась метрика accuracy на датасете, MNIST и датасете Chars74k.

Литература

- [1] 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.
- [2] 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.
- [3] Клименко С. В., Местецкий Л. М., Семенов А. Б. Моделирование рукописного шрифта с помощью жирных линий //Труды. – 2006. – Т. 16.
- [4] Кушнир О. и др. Сравнение формы бинарных растровых изображений на основе скелетизации //Машинное обучение и анализ данных. – 2012. – Т. 1. – №. 3. – С. 255-263.
- [5] Масалович А., Местецкий Л. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений //Труды Международной конференции «Графикон», Новосибирск. – 2006. – 4 с.
- [6] LeCun Y., Cortes C., Burges C. J. MNIST handwritten digit database // Available: <http://yann.lecun.com/exdb/mnist>. – 2010. – Т. 2.
- [7] Zhu D. et al. Negative Log Likelihood Ratio Loss for Deep Neural Network Classification //arXiv preprint arXiv:1804.10690. – 2018.
- [8] Nair P., Doshi R., Keselj S. Pushing the limits of capsule networks //Technical note. – 2018.
- [9] Hsieh P. C., Chen C. P. Multi-task Learning on MNIST Image Datasets. – 2018.