

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Александр Бойко

boyko.am@phystech.edu

В данной статье рассматривается комбинация трех подходов к классификации символов латинского алфавита: стандартных методов обучения без учителя, методов глубокого обучения нейронных сетей и предлагаемого нами метода графового описания символов с помощью скелетного представления толстых линий, оптимального для последующей обработки сверточной нейронной сетью. Проведен сравнительный анализ архитектур, работающих непосредственно с растровыми изображениями, и архитектур, работающих с графовыми представлениями символов на датасете Chars74K, достигнут значительный выигрыш в точности классификации за счет предложенного способа генерации графов.

Ключевые слова: *Сверточная нейронная сеть, CNN, скелет бинарного образа, скелетный граф, распознавание текста, классификация символов, толстая линия, растровое изображение*

1 Введение.

Задача распознавания символов на изображении, содержащем текст - одна из классических задач машинного обучения. Можно выделить два основных подхода к представлению изображений в данной задаче: дискретные и непрерывные. Так, дискретными называются методы, оперирующие с растровыми (т.е. представленными в виде последовательности пикселей) представлениями текста на изображениях. Существующие реализации алгоритма распознавания символов на растровых изображениях используют сверточные нейронные сети (например, [1], [2], [3]) и показывают хорошие результаты на тестах. Тем не менее, есть причины полагать, что для данной задачи лучше подходит непрерывное представление символов, то есть, представление символов с использованием фигур и форм. Мотивацией для использования непрерывных методов является их родство человеческому восприятию (человеческий глаз не распознает отдельные пиксели, склеивая их в единую форму). Кроме того,

непрерывные методы помогают бороться с искажениями текста на обрабатываемых изображениях, что является серьезной проблемой данной задачи ([4]). Различные непрерывные методы представления символов, а так же развернутая мотивация их использования приведены в книге [5]. В качестве непрерывного метода представления символов в данной работе используются так называемые алгоритмы скелетного представления символов. Алгоритмы описывают процесс заполнения внутренностей символов кругами, центры которых принимаются за вершины графа представления. Работа [6] посвящена нахождению оптимальных метрик в пространстве скелетных объектов для классификации символов и показывает, какие признаки можно выделить из скелетного представления. В нашей статье предлагается алгоритм обработки скелетного графа сверточной нейронной сетью для классификации символов по их скелетным графам. Эта сеть состоит из последовательных операций свёртки и уплотнения. В операции свёртки по отдельности рассматривается каждая небольшая часть описания изображения и в ней выделяются характерные паттерны. Операции уплотнения состоит в уменьшении числа признаков путём замены нескольких частей описания изображения на одну часть, аккумулирующую информацию о найденных паттернах. Основной задачей данной работы является конструирование оптимального алгоритма классификации на датасете Char74K и его сравнение с базовыми алгоритмами, использующими дискретное представление.

2 Постановка задачи.

Пусть дано множество A , состоящее из пар (y, I_{mn}) , где I_{mn} - растровая картинка размера $m \times n$ пикселей с буквой латинского алфавита, закодированная в оттенках серого, а y - метка класса Y (номер буквы в алфавите).

Зададим пространство неориентированных графов скелетного представления символа G . Считаем, что существует множество преобразований F такое, что $\forall g \in F \implies g : I_{mn} \rightarrow G$. Обозначим за Φ в общем случае бесконечномерное пространство признаков графов скелетных представлений. Пусть $\phi \subset \Phi$ и ϕ - конечномерное подпространство Φ .

Определим линейное векторное пространство V и будем считать, что существуют сюръективные преобразования $\gamma : G \rightarrow V$. Обозначим B, C за множества функций, действующих из пространства G в пространства ϕ

и V соответственно.

Эти функции, таким образом, переводят графы скелетных представлений в пространство графов пониженной размерности и в векторное пространство. Обозначим за D, E множества функций, действующих из ϕ, V в Y соответственно.

Обозначим за $L(\hat{y}, y)$ функцию потерь кросс-энтропии. Задачей настоящего исследования является оценка функций $\hat{f} : I_{mn} \Rightarrow Y, \hat{\chi} \in D, \hat{\kappa} \in E$, отображающих из пространств признаков в пространство классов Y . Кроме того, целью исследования является поиск функций $h \in F, t \in B$ из условия минимизации функционала $\arg \min_{h \in F, t \in B} L(\hat{\chi}(t(h)), \hat{f})$, а так же поиск функций $h \in F, s \in C$ из условия минимизации функционала $\arg \min_{h \in F, s \in C} L(\hat{\kappa}(s(h)), \hat{f})$.

3 Список литературы

1. Simard Patrice Y, Steinkraus Dave, Platt John C. Best practices for convolutional neural networks applied to visual document analysis // null / IEEE. — 2003. — P. 958.
2. Jaderberg, M., Simonyan, K., Vedaldi, A. et al. Int J Comput Vis (2016) 116: 1. <https://doi.org/10.1007/s11263-015-0823-z>
3. Qiang Guo, Jun Lei, Dan Tu, Guohui Li, "Reading numbers in natural scene images with convolutional neural networks Security Pattern Analysis and Cybernetics (SPAC) 2014 International Conference on, pp. 48-53, 2014.
4. Масалович Антон, Местецкий Леонид. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений // Труды Международной конференции «Графикон», Новосибирск.—2006.—4
5. “Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры” (Л.Местецкий) - ФИЗМАТЛИТ, 2009
6. Кушнир О et al. Сравнение формы бинарных растровых изображений на основе скелетизации // Машинное обучение и анализ данных. — 2012. — Vol. 1, no. 3. — P. 255–263.