

УДК 004.912

ИССЛЕДОВАНИЕ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ДЛЯ РАСПОЗНОВАНИЯ РУКОПИСНОГО ФРАНКОЯЗЫЧНОГО ТЕКСТА

Скворцова А.Ю. (ИТМО)

Научный руководитель – Михайленко А.Ю. (ИТМО)

Введение. Распознавание рукописных текстов является важной задачей в области компьютерного зрения и обработки естественного языка. С развитием технологий искусственного интеллекта все большее внимание уделяется использованию генеративных сетей для решения этой задачи. Для таких языков, как, например, английский, существует множество решений с высоким качеством моделей. Однако при использовании современных генеративных сетей для распознавания рукописных франкоязычных текстов возникает ряд проблем. Во-первых, французский язык обладает сложной грамматической структурой, что усложняет модель и повышает время обработки текста. Во-вторых, из-за специфики французского письма возникает проблема с распознаванием диакритических знаков и отличиям их от сторонних символов. В-третьих, скорость работы генеративных сетей может быть недостаточной для реального времени, особенно при работе с большим объемом данных. Наконец, требования к вычислительным ресурсам при обучении и использовании генеративных сетей могут быть значительными, что ограничивает их практическое применение.

Основная часть. Для эффективного применения генеративных сетей в задаче распознавания рукописных франкоязычных текстов необходимо провести исследования и оптимизировать методы обучения и работы сетей, чтобы достичь более точных результатов, повысить скорость работы и снизить требования к вычислительным ресурсам. В данном исследовании предлагается рассмотреть существующие архитектуры, разработать базовые модели, применимые к распознаванию англоязычных текстов. После предлагается применить построенные модели к франкоязычным текстам, сравнить результаты. Затем изучить существующие решения и добавить отдельные части к уже разработанным базовым моделям. На основе созданного математического аппарата построить диффузионную модель. Далее подобрать метрики и оценить качество новой модели, сгенерировать тестовые образцы и провести глубокий анализ полученных результатов. Финальным этапом разработки можно считать реконструкцию модели для улучшения показателей скорости и качества распознавания при необходимости.

Выводы. Построенная модель может быть применена для распознавания рукописного текста в отсканированных документах, письмах или рукописных записях. Такая модель может также быть использована для распознавания и анализа рукописных подписей, что может быть полезно в банковском секторе, юридических документах и других областях. Эта модель применима и в искусстве: распознавание рукописных текстов, написанных в различных стилях каллиграфии, может оказаться полезна для художников, дизайнеров и любителей каллиграфии. Отдельно стоит выделить медицинскую сферу: такая модель может распознавать рукописные записи в медицинских картах или отчетах, что поможет в улучшении точности и скорости анализа медицинских данных. Также модель сможет стать основой для разработки генеративных моделей для других языков с письменной структурой, сложнее англоязычной.

Список использованных источников:

1. *Sharon Fogel, Hadar Averbuch-Elor, Sarel Cohen, Shai Mazor and Roe Litman* ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation // arXiv. 2020. V. 1.
2. *Jorge Juan Sueiras Revuelta* Continuous Offline Handwriting Recognition using Deep Learning Models // arXiv. 2021. V. 1.
3. *Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser* Attention Is All You Need // arXiv. 2023. V. 7.