

Разработка и анализ методов обработки сверхвысокоразмерных изображений для семантического анализа с применением нейронных сетей

Пименов А. В. (ИТМО)

Научный руководитель - Ефимова Валерия Александровна (ИТМО)

Введение. На текущий момент существуют актуальные задачи в области анализа векторных изображений: поиск схожих изображений, обнаружение плагиата, генерация изображений и другие подобные задачи, которые уже решены при работе с растровыми изображениями. Одной из проблем является неограниченность векторных изображений в количестве объектов на изображении. Это накладывает существенные ограничения в плане анализа и генерации новых изображений [1]. Мы хотим предложить решение которые поможет обрабатывать векторные изображения неограниченной сложности.

Основная часть. Наиболее простым методов является расширение входной последовательности. К сожалению, архитектура трансформера такова что она ограничена квадратичной сложностью что накладывает существенные ограничения на технические возможности. Кроме того, векторные изображения могут содержать неограниченное количество объектов что дает возможность получить изображений невозможное для обработки даже при достаточно большой входной последовательности.

Возможное решение основано на исследованиях, проведенных на трансформерах для решения задачи анализа длинных последовательностей токенов [2]. В его основе лежит идея работы над последовательностью в виде сегментированной рекуррентной памятью. Механизм реализован следующим образом:

1. Токенизация данных
2. Разбитие данных на сегменты
3. Добавление специальных токенов указывающих на начало и конец сегмента

Таким образом модель трансформера без изменений в архитектуре получает возможность управлять этими сегментами подобно памяти для выявления наиболее важных признаков из длинных последовательностей токенов. Кроме того, внедрение этого решения не оказывает существенного влияния на требования к техническим ресурсам. При тестировании базового трансформера с расширенной входной последовательностью и трансформера с сегментированной рекуррентной памятью было выявлено несущественное отличие в ходе обработки последовательности в 1 миллион токенов [3].

Вывод. Предложенное решение способно решить проблему, связанную с неограниченностью в количестве объектов векторных изображений. Решение этой проблемы дает возможность генерировать изображения без ограничений на количество объектов, а также анализировать самые сложные векторные изображения.

Список использованных источников:

1. Wu, R., Su, W., Ma, K., & Liao, J. (2023b). IconShop: Text-Guided Vector Icon Synthesis with Autoregressive Transformers. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2304.14400>
2. Bulatov, A., Kuratov, Y., & Burtsev, M. (2022). Recurrent Memory Transformer. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2207.06881>
3. Bulatov, A., Kuratov, Y., & Burtsev, M. (2023b). Scaling Transformer to 1M tokens and beyond with RMT. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2304.11062>