УДК 004.932

Применение глубокой нейронной сети трансформера для задачи повторной идентификации транспортных средств

Саитов И.А. (Университет ИТМО)
Научный руководитель – доцент, к.ф-м.н, Фильченков А.А.
(Университет ИТМО)

Введение. Хотя глубокие нейронные сети трансформера доказали свою эффективность в области обработки естественного языка, их применение в компьютерном зрении ещё слабо изучено. При обработке изображений механизм внимания (ключевой компонент архитектуры трансформера) применяется как правило в дополнении к свёрточной нейронной сети либо же заменяя некоторые её части. Одним из главных ограничений [1] использования архитектуры трансформера является необходимый объём тренировочных данных при обучении. В частности, для задачи повторной идентификации транспортных средств существует проблема малого количества примеров для каждого объекта. Поэтому в работе представляется использование полноценного трансформера для задачи повторной идентификации транспортных средств. Рассматривается эффективность предобучения модели на более крупном датасете схожей доменности. Разработанная модель сравнивается относительно свёрточной нейронной сети.

Основная часть. Модели с механизмом внутреннего внимания, в частности архитектура глубокой нейронной сети трансформера, стали предпочтительной моделью в обработке естественного языка (англ. NLP). Преобладающий подход заключается в предварительном обучении на большом текстовом корпусе, а затем тонкой настройке на меньшем наборе данных для конкретной задачи. Благодаря вычислительной эффективности и масштабируемости трансформеров стало возможным обучать модели беспрецедентного размера с более чем 100 миллиардами параметров. С увеличением количества моделей и наборов данных по-прежнему нет признаков насыщения производительности.

Однако в области компьютерного зрения доминирующую позицию занимают модели глубоких свёрточных нейронных сетей (англ. CNN). Благодаря успехам в обработке естественного языка появились архитектуры, включающие комбинацию CNN и механизма внимания. И хотя они имели теоретическую эффективность во многих задачах классические модели всё ещё показывали свою эффективность.

На примере классической задачи распознавания изображений полноценная архитектура трансформера [1] показала свою эффективность. Как уже было сказано для успешного применения данной модели необходим большой объём тренировочных данных. Создание в последние годы таких датасетов, как VehicleID и VeRi-776, для задачи повторного распознавания транспортных средств позволяет применять архитектуры глубокой нейронной сети трансформера. Одной из моделей продемонстрировавшей применение этой идеи является [3]. Авторы продемонстрировали преимущества методологии в плане устойчивости обученных представлений объектов и эффективности использования вычислительных мощностей.

В данной работе исследуется применение такой модели глубокой нейронной сети трансформера для задачи повторной идентификации транспортных средств. В качестве датасетов для обучения использовались упомянутые VehicleID и VeRi-776, имеющие 229567 с 776 классами и 49357 с 26328 классами изображений. Для чистоты экспериментов в качестве исходной модели извлечения фич изображений выбрана ResNet50. Для сравнения использовалась мощная реализация свёрточной сети с характерными и эффективными для данной задачи методиками.

Выводы. Выполнено исследование по использованию чистой архитектуры трансформера с применением технологии переноса знаний для задачи повторной идентификации транспортных средств. И приведено сравнение относительно архитектуры свёрточной нейронной сети.

Список использованных источников:

1. Dosovitskiy A. et al. An image is worth 16x16 words: Transformers for image recognition at scale // arXiv preprint arXiv:2010.11929. – 2020.
2. Luo H. et al. A strong baseline and batch normalization neck for deep person re-identification // IEEE Transactions on Multimedia. – 2019. – Т. 22. – №. 10. – С. 2597-2609.
3. Zheng Z. et al. VehicleNet: Learning robust visual representation for vehicle re-identification // IEEE Transactions on Multimedia. – 2020. – Т. 23. – С. 2683-2693.