

УДК 004.93

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ СЖАТЫХ ВИДЕОРАДОВ

Николай Я.В. (Национальный исследовательский университет ИТМО)

Научный руководитель – д.т.н., профессор Тропченко А.Ю.

(Национальный исследовательский университет ИТМО)

В данной работе предложен подход для классификации видеорядов на основе векторов движения, извлекаемых из сжатого в формате MPEG видео. Реализованная на языке Python модель показала увеличение скорости обработки по сравнению с базовой моделью, применяемой к несжатым кадрам.

Введение. С каждым годом увеличивается количество хранимых цифровых материалов, в том числе и видео, которые должны быть каким-либо образом проаннотированы для ускорения получения наиболее релевантных для пользователя данных. Несмотря на то, что видеоряды обычно хранятся в сжатом виде, используемые в таких системах алгоритмы применяются для сырых кадров, что увеличивает требуемый для работы объем памяти, а так же время обработки из-за необходимости выполнять декодирование кадров. В последние годы для задач классификации видео российские и зарубежные ученые применяют нейронные сети, что позволяет существенно увеличить точность предсказания класса видеоряда.

Основная часть. Как показал обзор существующих решений, использование оптического потока как входных данных для нейронной сети улучшает качество распознавания образов в видео. Вектора движения, которые создаются на этапе компенсации движения при сжатии видео используются для той же цели, что и оптический поток – определить новое местоположение пикселей на следующем или предыдущем кадре. Следовательно, представляется целесообразной разработка модели, которая бы использовала вектора движения для классификации видео. В предлагаемом подходе вектора движения извлекаются из сжатого видео с помощью библиотеки ffmpeg и передаются на вход нейронной сети. Каждый вектор обрабатывается сначала сверточным блоком, который извлекает признаки, далее блоки одного кадра агрегируются в блоке LSTM. Кадры, которые были сжаты без использования обрабатываются другим LSTM-блоком. Выходы покадровых LSTM-блоков передаются в LSTM-слой для учета временных изменений, а затем на классифицирующий Softmax-слой.

Выводы. Предложенная архитектура нейронной сети была реализована на языке Python. Сравнение с базовой моделью, применяемой к несжатым кадрам, показало в среднем ускорение скорости обработки одного видеофрагмента. Предложенная в данной работе модель может быть использована в системах хранения сжатых видео для аннотирования.

Николай Я.В. (автор)

Подпись

Тропченко А.Ю. (научный руководитель)

Подпись