

АНАЛИЗ ПРИМЕНИМОСТИ КАРТ ВИЗУАЛЬНОГО ВНИМАНИЯ, ПОСТРОЕННЫХ С ПОМОЩЬЮ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ, К ЗАДАЧЕ СЖАТИЯ ВИДЕО

Долганов Станислав Викторович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: sdolganov@graphics.cs.msu.ru

Зрительная система человека устроена таким образом [1], что только малая часть видимой области захватывает большое количество внимания и требует высокого разрешения, в то время как остальная часть на периферии получает небольшое количество внимания и не требует высокого разрешения. Распределение внимания человека принято моделировать с помощью карт визуального внимания. Наиболее перспективной задачей с их использованием в данное время является сжатие видео.

Нейронные сети глубокого обучения преуспели во многих задачах, как, например, классификация изображений, распознавание движений и распознавание объектов. В области предсказания карт визуального внимания, подобные алгоритмы также научились автоматически извлекать пространственные признаки для оценки карт. В статье [2] предлагается использовать двухкомпонентную архитектуру, позволяющую напрямую предсказать попиксельную карту визуального внимания. Такой подход дает возможность изучать более сложное распределение внимания человека.

В данной работе анализируется использование карт визуального внимания, полученных с помощью рекуррентных нейронных сетей, для задачи сжатия видео. В качестве алгоритма предсказания карт использовался метод, описанный в [2], а для сжатия видео — модифицированная версия кодека x264, предложенная в статье [3], которая устанавливает для областей с высоким значением внимания более низкие значения параметра квантизации. Метод также позволяет контролировать процент битрейта выделяемый областям повышенного интереса.

Для сравнения визуального качества видео, сжатого с использованием карт визуального внимания, с видео, сжатого общепринятым способом, была написана и использована платформа [4], которая позволяет проводить экспертную оценку методов обработки медиа контента. В исследовании приняло участие 583 респондента (474 успешно). Каждому было предложено из 12 пар видео выбрать бо-

лее предпочтительное с точки зрения качества. В сравнении все пары состояли из видео, сжатого с помощью стандартного пресета x264 с битрейтом 1500 Кб/с, и видео, сжатого с учетом карты визуального внимания с битрейтами 1000, 1100, 1200, 1300, 1400, 1500 Кб/с. В сравнении использовались 18 сцен из датасета SAVAM [3].

На основании результатов экспертной оценки, при использовании для сжатия карт визуального внимания, полученных с помощью глубокой нейронной сети [2], можно в среднем получить выигрыш в битрейте около 11%, т.е. видео размером на 11% меньше будет иметь такое же визуального качество. Так как видео составляет большую часть трафика, передаваемого через Интернет, такая экономия позволит существенно сократить нагрузку на сети.

Иллюстрации

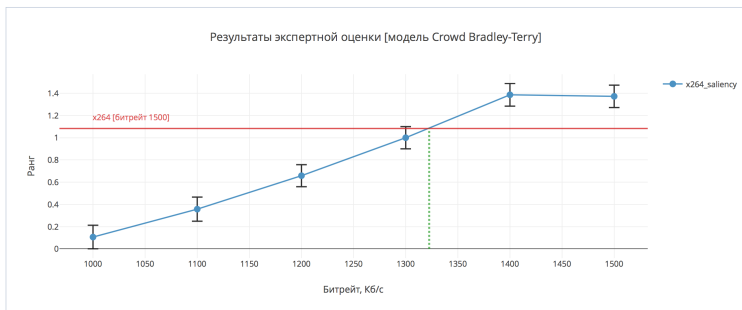


График результатов сравнения. Для ранжирования методов на основании результатов попарного сравнения использовался метод Crowd Bradley-Terry [5]. Красная линия отвечает методу x264, 1500 Кб/с. Как можно заметить, пересечение двух линий проходит в точке 1325 Кб/с, что соответствует 11% выигрыша.

Литература

1. Matin E., Saccadic suppression: a review and an analysis // In Psychological bulletin, volume 81, number 12, 1974, P. 899.
2. Jiang L., Xu M., and Wang Z., Predicting Video Saliency with Object-to-Motion CNN and Two-layer Convolutional LSTM // In arXiv preprint arXiv:1709.06316, 2017.
3. Lyudvichenko V., Erofeev M., Gitman Y., and Vatolin D., A semiautomatic saliency model and its application to video

compression // In IEEE International Conference on Image Processing, 2017.

4. Multimedia subjective quality evaluation platform:
<http://www.subjectify.us>
5. Chen X., Bennett P., Collins-Thompson K., and Horvitz E., Pairwise ranking aggregation in a crowdsourced setting // In Proceedings of the sixth ACM international conference on Web search and data mining, 2013, P. 193–202.