

ИСПОЛЬЗОВАНИЕ ДВУХУРОВНЕВЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ПРИЗНАКОВ ВИДЕО

Гайбуллаев Абдулла Зиядуллаевич

Студент

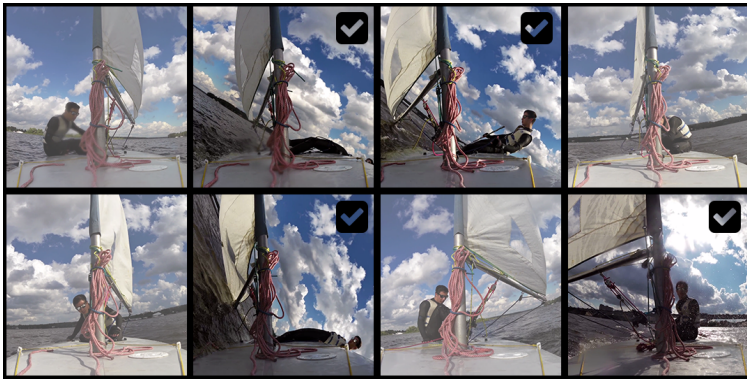
Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: abdulla.gaibullaev@graphics.cs.msu.ru

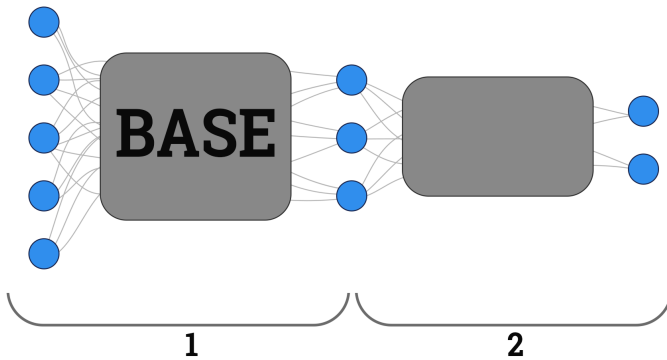
В настоящее время широкую популярность приобрели экшен-камеры, на которые записываются, например, трансляции спортивных мероприятий. Обычно продолжительность необработанных видео измеряется часами, тогда как интерес может представлять лишь небольшая их часть. Например, средний прыжок с парашютом длится около 10 минут, а само видео может составлять более часа (подготовка, подъем, завершение). В работе рассматривается задача автоматического поиска интересных моментов в таких видеопоследовательностях.

Помимо выделения интересных моментов, может возникнуть множество других задач, таких как анализ соревнований, поиск интересных моментов, и т. д. Эти задачи решаются с помощью глубоких нейронных сетей, однако высокоуровневые признаки видео очень тяжело поддаются обучению. При этом точность результатов оказывается не выше 70%. Главной проблемой для обучения таких признаков является огромный размер входных данных. Для решения данной проблемы используются многоуровневые нейронные сети.

Иллюстрации



Пример интересных моментов в видео, снятых на парусных гонках.



Двухуровневая нейронная сеть.

Рассмотрим двухуровневые сети. Назовем первую часть этой структуры *базовой нейронной сетью*. Также предположим, что эта часть может быть обучена отдельно. Обычно первый уровень представляет из себя сложную и хорошо обученную сеть без нескольких последних слоев. Например, AlexNet или Inception. Но, как уже было сказано выше, такие модели выдают очень низкую точность.

В данной работе предлагается в качестве базовой нейронной сети использовать узкоспециализированную сеть, т. е. сеть, обученную на выборке определенного вида спорта. Применение такого подхода позволяет повысить точность за счет сужения области применения.

Литература

1. Yao T., Mei T., Rui Y. Highlight detection with pairwise deep ranking for first-person video summarization. – 2016.
2. Naik J. DeepVideo: Video Summarization using Temporal Sequence Modelling.
3. Xu B. et al. Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization //IEEE Transactions on Affective Computing. – 2016.
4. Potapov D. et al. Category-specific video summarization //European conference on computer vision. – Springer, Cham, 2014. – С. 540-555.