

## Превращаем горизонтальные видео в вертикальные

Как правило, кино снимают в таком соотношении сторон, в котором ширина кадра больше высоты, потому что глаза человека расположены горизонтально по отношению друг к другу. Будем называть такие видео горизонтальными.

Мы в Иви показываем кино на разных устройствах, в том числе и мобильных телефонах. А телефоном удобнее пользоваться, когда держишь его вертикально, то есть ширина экрана меньше высоты. Если показать горизонтальное видео на вертикальном экране, останется много неиспользуемого места, и масштаб будет сильно меньше, чем мог бы быть.

Чтобы смотреть видео на телефонах было более приятно, мы хотим научиться автоматически превращать горизонтальные видео в вертикальные.

В этой задаче, вам нужно создать модель или алгоритм, который вычисляет координату кропа, который превращает горизонтальное видео в вертикальное с соотношением сторон 9:16.

Участникам будут даны следующие данные:

- 2849 коротких видео из фильмов и мультфильмов, у которых ширина больше высоты
- списки ID актёров по каждому фильму, из которых вырезаны фрагменты. ID будут расположены в порядке значимости актёра в фильме. Сначала — самые важные, в конце — менее важные.
- каждому из видео будет соответствовать файл с метаданными, среди которых
  - координаты боксов лиц, имеющихся в видео
  - номера треков лиц
  - для каждого бокса лица — скор, который показывает, разговаривает ли персонаж, которому принадлежит это лицо. Если скор больше 0, значит разговаривает.
  - ID актёра, которому принадлежит лицо, если удалось распознать или -1, если не удалось

Для 10 файлов будет предоставлена разметка, то есть правильные координаты кропов от редакции Иви.

Ещё некоторое количество размеченных видео останутся у команды Иви в качестве приватного тестового сета. Он будет использоваться при оценке решений участников хакатона.

Все остальные видео будут предоставлены без разметки. Если у участников возникнет желание, они могут доразметить их вручную.

Датасет: <https://disk.yandex.ru/d/Opcyyoos3mPzyg>

Результат: Код на языке Python, который принимает на вход пути к видео-файлу и его метаданным, а на выходе генерирует json-файл со списком.

Порядковые номера в списке соответствуют номерам кадров во входном видео, начиная с 0.

Каждый элемент списка — это координата кропа для соответствующего кадра.

Высота кропа будет всегда равна высоте исходного видео, а ширина должна быть такая, чтобы кроп имел соотношение сторон 9:16.

Координата кропа — это расстояние от левого края кадра исходного видео до левого края кропа.

Код лучше оформлять в виде репозитория на github.

В репозитории обязательно должен быть файл README.md с инструкцией по запуску кода.

Победители будут определяться командой компьютерного зрения Иви и редакцией контента Иви по двум критериям:

- субъективное качество кропов, получаемых моделью с точки зрения команды Иви
- метрика Mean Squared Error между предсказаниями модели и координатами на приватном тестовом сете