

Необходимо проанализировать содержание видео.

Подготовьте 10 секундный видеоролик. Снимите себя или друга. Еще лучше двух человек сразу. Выполнить любые не сложные движения, на любом фоне, в любой обстановке. Если на видео 2 человека, лучше будет если они выполняют разные действия.

Можно взять любой приличный кусок ролика с любой соцсети или любой ролик из датасета [Kinetics 400](#).

### **Задание 1.**

С помощью модели, полученной на практическом занятии по обучению модели анализа изображения (можно использовать предь обученную отсюда), напишите скрипт проходящий через каждый кадр видео.

У вас получится много предсказаний (на каждый кадр), соберите самые часто встречающиеся существительные и глаголы со всех кадров.

Похоже ли описываемые моделью действия и обстановка на те, что изображены на видео? В README.md отразите что модель определила, что нет. Предложите гипотезу, почему модель не все описала?

### **Задание 2.**

С помощью фреймворка SlowFast примените модель slowfast net к исходному видео. У вас должно получиться видео с выделенными объектами и подписанными действиями. Опишите точно ли модель распознала действия, совершаемые на видео.

Изучите описание Slowfast net. В чем сильные стороны данной модели?

Предложите гипотезу как объединить image capturing и video recognition чтобы получить модель, описывающую видео?

Чем усложняется задача video capturing по сравнению с image capturing?

### **Требования:**

- Репозиторий `github/gitla`
- скрипты `.py` или `jupyter notebook` с инференсом моделей
- Поместите в README ваши выводы и ответы на вопросы
- Поместите в README визуализацию результатов работы моделей (несколько кадров с описанием от модели `image capturing`, гифку с отрезком видео с визуализацией работы `slow fast`)