В решении использованы открытая библиотека transformers и модели с платформы HuggingFace.

Реализовано 2 подхода к генерации обложки:

1. Пользователь загружает видео в сервис. Видео переводится в картинки. Берется одна картинка из каждой секунды и к ней генерируется описание (модель Salesforce/blip-image-captioning-large) для детекции объектов в видео. Все описания объединяются и суммаризируются (модель facebook/bart-large-cnn). Это первая часть итогового промпта. Если пользователь добавил текстовое описание к видео, то по нему делается саммари (модель IlyaGusev/rubert\_telegram\_headlines), переводится на английский (модель Helsinki-NLP/opus-mt-ru-en) – это вторая часть итогового промпта. Если пользователь указал теги, которые хочет видеть на обложке, они переводятся на английски и добавляются к итоговому промпту. Итоговый промпт подается на вход генеративной сети (модель animelover/novelai-diffusion), генерируется картинка в стиле аниме. При повторном запросе на генерацию, будет создаваться новая картинка.
2. Полностью повторяет пайплайн получения первой части промпта из первого подхода. Далее рассчитывается удаленность каждого из описаний к кадрам из видео от итогового промпта, и выбирается кадр, который с минимальным расстоянием. При повторном запросе выберется та же самая картинка.

Ограничения, с которыми мы столкнулись:

1. Чем дольше видео, тем дольше оно обрабатывается. Видео до 1 минуты обрабатывается до 1 минуты, видео около 15 минут обрабатывается около 5 минут, и т.д.
2. Первый подход не очень хорошо работает на мультиках

Точки роста:

1. Добавить функционал по выбору желаемого стиля изображения: реализм, импрессионизм и т.д.
2. Добавить функционал по генерации аватарки из загруженного фото пользователя, с использованием переноса стиля, выбранного пользователем.

Результат:

|  |  |  |
| --- | --- | --- |
| Кадр из видео | 1 подход (сгенерированное изображение) | 2 подход (выбранное изображение) |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |