

# ОТСЛЕЖИВАНИЕ ЛИЦ НА ВИДЕО С ИСПОЛЬЗОВАНИЕМ КЛАССИЧЕСКИХ И НЕЙРОСЕТЕВЫХ ПОДХОДОВ В КОМПЬЮТЕРНОМ ЗРЕНИИ

*Латушкин Н.С.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Сиротко С.И. – к.т.н., доцент*

В докладе приводится описание инструмента позволяющего вести работу с лицами на видео. Также описываются принятые решения и модули необходимые для построения системы.

В современном мире одним из наиболее популярных типов данных являются видеоданные. Поэтому обработка видео является частой и необходимой задачей. Однако обработка такого типа данных требует большого количества времени и ресурсов. Например, если мы хотим обработать 1 минуту видео, в случае отсутствия специальных программных средств для этих целей, мы будем вынуждены обработать целых 1800 кадров (как правило одна секунда видео содержит 30 кадров). Одной из наиболее часто решаемых задач в контексте обработки видео является задача обнаружения и распознавания людей.

Таким образом система для автоматического отслеживания лиц на видео должна содержать в себе следующие логические компоненты: детекция лиц, распознавание лиц, трекер лиц, модуль анализа и объединения информации, а также удобный графический интерфейс для работы с видео в случае, когда нам понадобится исправлять ошибки автоматического аннотирования видео.

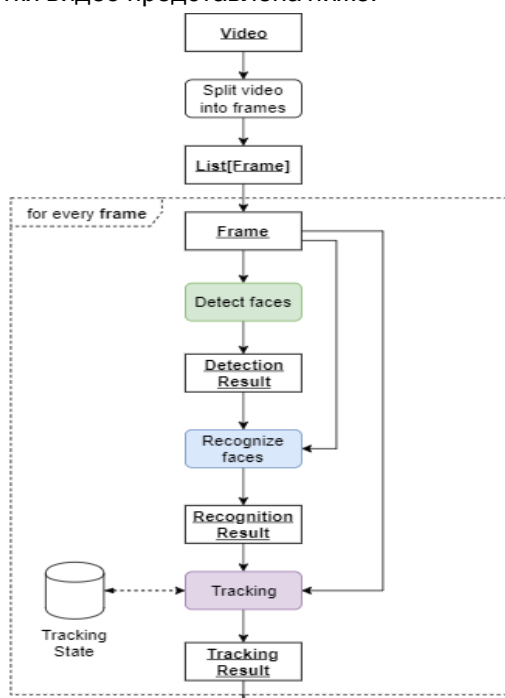
Детектирование лиц необходимо для нахождения всех лиц на кадрах, а также их последующего сопоставления с людьми. В качестве решения для этой задачи будем использовать нейросетевой детектор на основе нейронной сети Retina с использованием в качестве backbone архитектуры

Resnet50. В качестве входа она будет получать, по очереди, каждый кадр из видео. Таким образом, после работы этого модуля мы получим все координаты рамок обрамляющих лица.

Распознавание лиц используется для агрегации рамок, полученных в результате работы модуля детектирования, в коллекции, внутри которых содержатся рамки одного и того же человека. Для этих целей мы будем использовать нейронную сеть по типу архитектуры ArcFace с backbone Resnet50. В качестве входных данных будем использовать вырезанные по координатам рамок детекции изображения. Их также необходимо выровнять и привести к одному размеру. На выходе получим набор дескрипторов, соответствующих каждой из поданных на вход рамок. Данные дескриптора будут обладать следующим свойством: если посчитать метрику косинусной схожести для двух дескрипторов, то у одинаковых лиц она будет высокой, а у разных низкой.

Трекинг лиц используется для устранения минусов работы модулей детекции и распознавания лиц. Иногда модуль детекции лиц может ошибаться и не детектировать все лица на кадре, тогда в случае наличия этого человека на других кадрах видео, с помощью трекинга мы имеем шанс найти этого человека на всех кадрах, на которых он присутствует. В качестве трекера будет использоваться SiamRpn сети, принимающие на вход все ту же рамку детекции.

Общая схема обработки видео представлена ниже:



Общая схема обработки видео

**Список использованных источников:**

1. Examples of visual object tracking Электронный ресурс / Режим доступа: <https://pt.coursera.org/lecture/deep-learning-in-computer-vision/examples-of-visual-object-tracking-methods-NmYR4>
2. Face detection Электронный ресурс / Режим доступа: <https://arxiv.org/abs/1905.01585>
3. Face recognition Электронный ресурс / Режим доступа: <https://arxiv.org/abs/1801.07698>