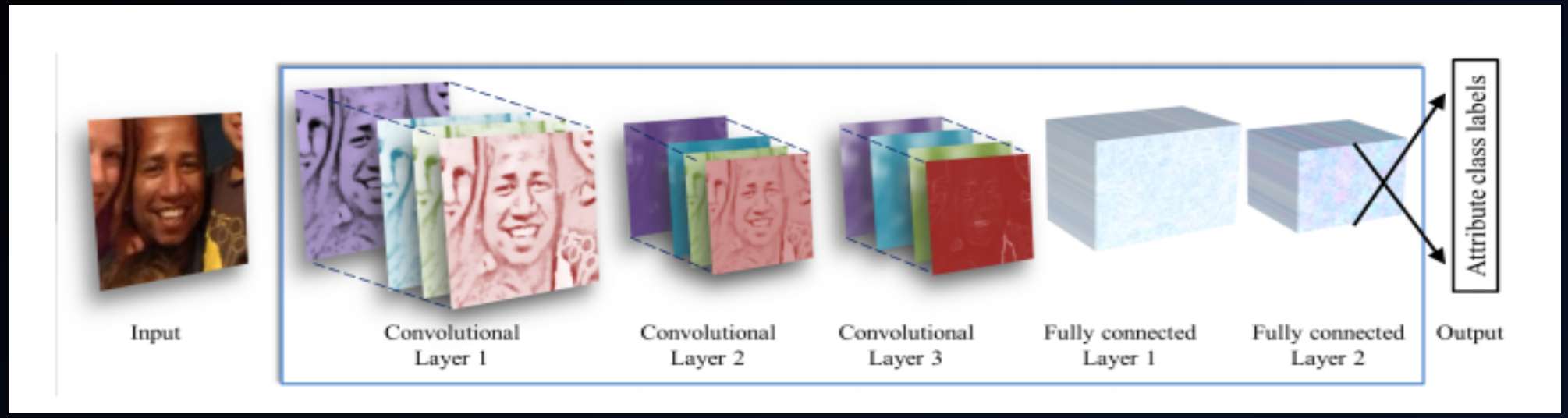
Распознавание пола и возраста по видеоизображению лица на основе сверточных нейронных сетей

РАБОТУ ВЫПОЛНИЛА:
ХАРЧЕВНИКОВА АНГЕЛИНА
СТУДЕНТ 3 КУРСА
БАКАЛАВРИАТА ФАКУЛЬТЕТА
ИНФОРМАТИКИ МАТЕМАТИКИ
И КОМПЬЮТЕРНЫХ НАУК
НИУ ВШЭ

НИЖНИЙ НОВГОРОД 2017

- В данной работе задача распознавания пола и возраста на видеоизображении рассматривается как задача выбора наиболее надежного решения из нескольких на основе синтеза коллективов решающих правил (КРП, комитетов классификаторов)

Строение модели



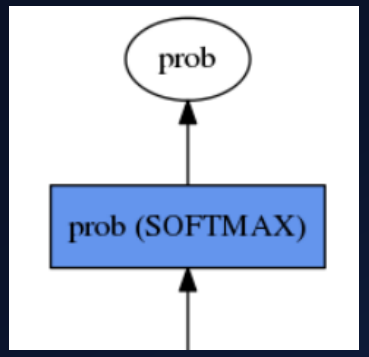
Входной объект

Вероятности

Выход нейросетевой модели обычно получается в слое Softmax, который выдает вероятности принадлежности объекта к каждому классу

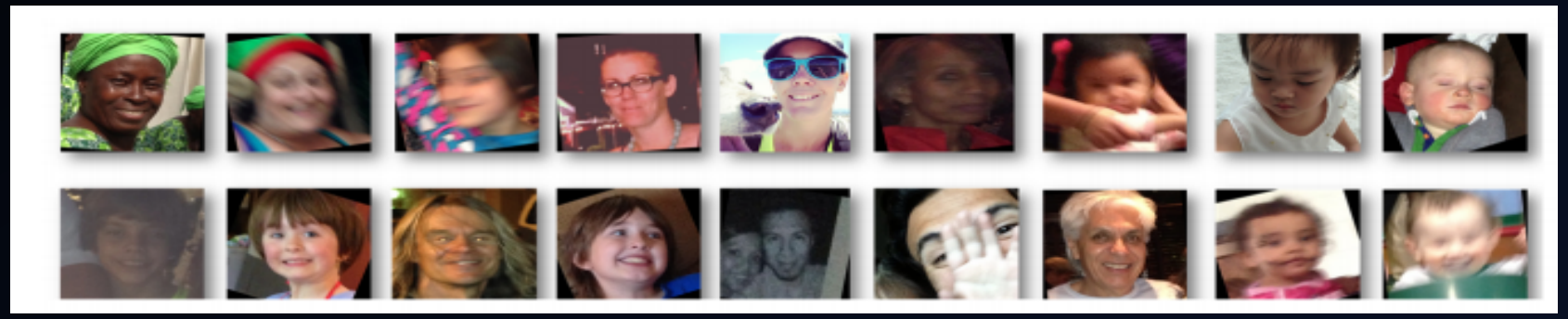
$$P(X(t)|l) = \text{softmax } z_l(t) = \frac{\exp(z_l(t))}{\sum_{j=1}^{L} \exp(z_j(t))}, l = 1, 2, \ldots, L,$$

$$l^*(t) = \underset{l=1,2,\ldots,L}{\text{argmax}} \; P(X(t)|l).$$



Результаты тестирования моделей

- Тестирование проводилось с помощью dataset Audience collection (Audience benchmark)
- Коллекция содержит 26000 нефильтрованных изображений, полученных с камер смартфонов



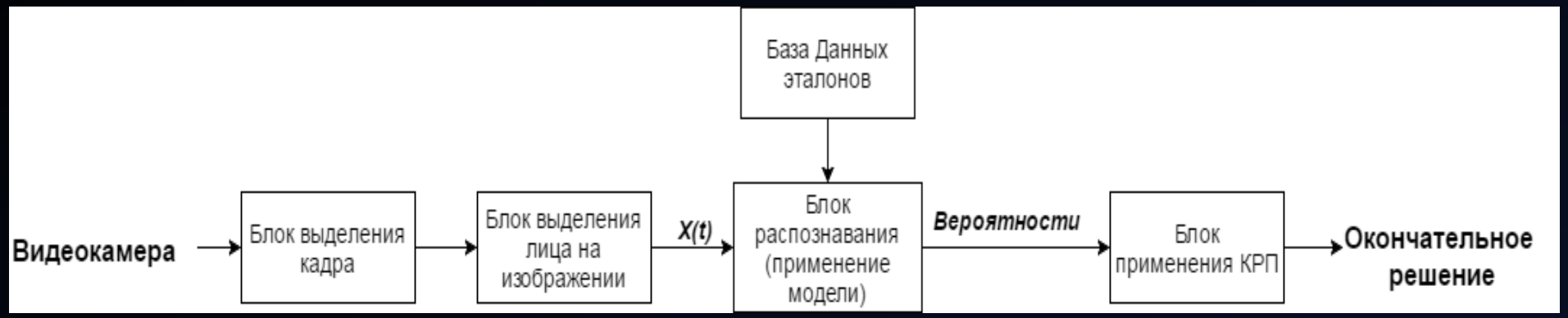
	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

Table 1. **The AdienceFaces benchmark.** Breakdown of the AdienceFaces benchmark into the different Age and Gender classes.

- Точность: распознавание пола – 87%

  распознавание возраста – 85%

Предлагаемый алгоритм



- Задача распознавания видеоизображения состоит в том, чтобы отнести вновь поступающую (на вход) последовательность кадров $\{X(t)\},\ X(t) = \left\| x_{uv}(t) \right\|,\ t = \overline{1,T}$ с изображением одного объекта к одному из *L* классов
- Распознавание пола: L = 2
- Распознавание возраста: L = 8

Применение КРП

- Простое голосование

$$l^* = \arg\max_{l=\overline{1,L}} \sum_{t=1}^{T} \delta(l^*(t) - l).$$

- Метод усреднения вероятностей

1. $$l^* = \operatorname{argmax}_{l=\overline{1,L}} \frac{1}{T} \sum_{t=1}^{T} P\left(X(t) \middle| l\right)$$

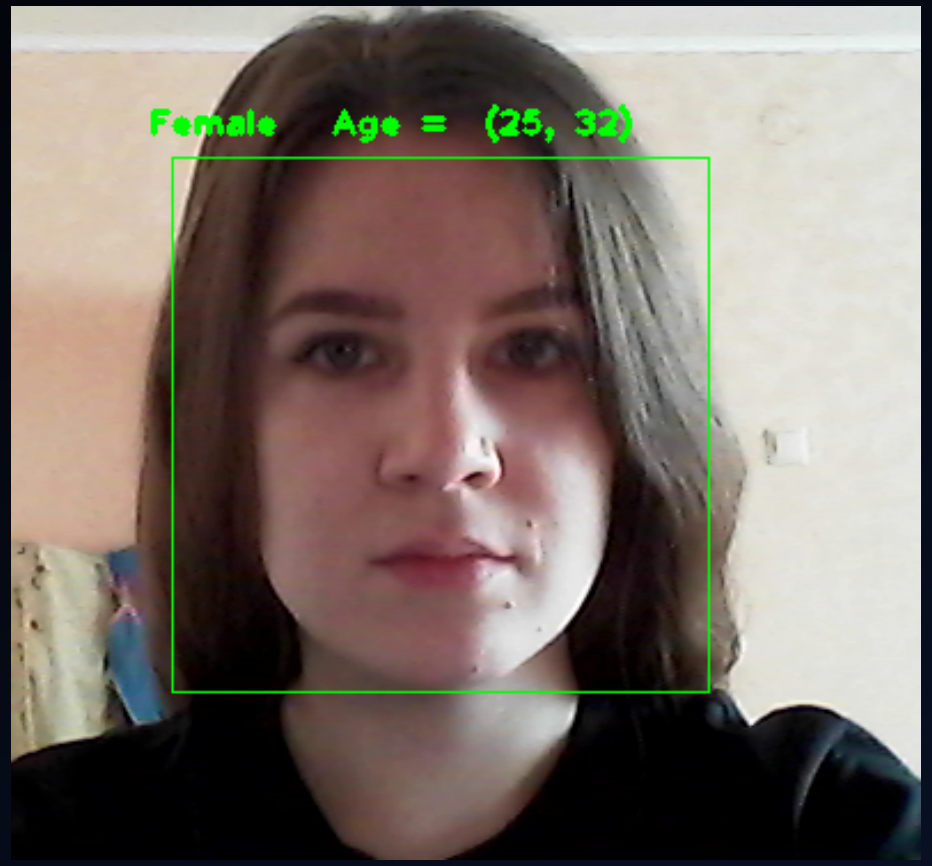
2. $$l^* = \operatorname{argmax}_{l=\overline{1,L}} \prod_{t=1}^{T} P\left(X(t) \middle| l\right) = \operatorname{argmax}_{l=\overline{1,L}} \sum_{t=1}^{T} \log P\left(X(t) \middle| l\right)$$

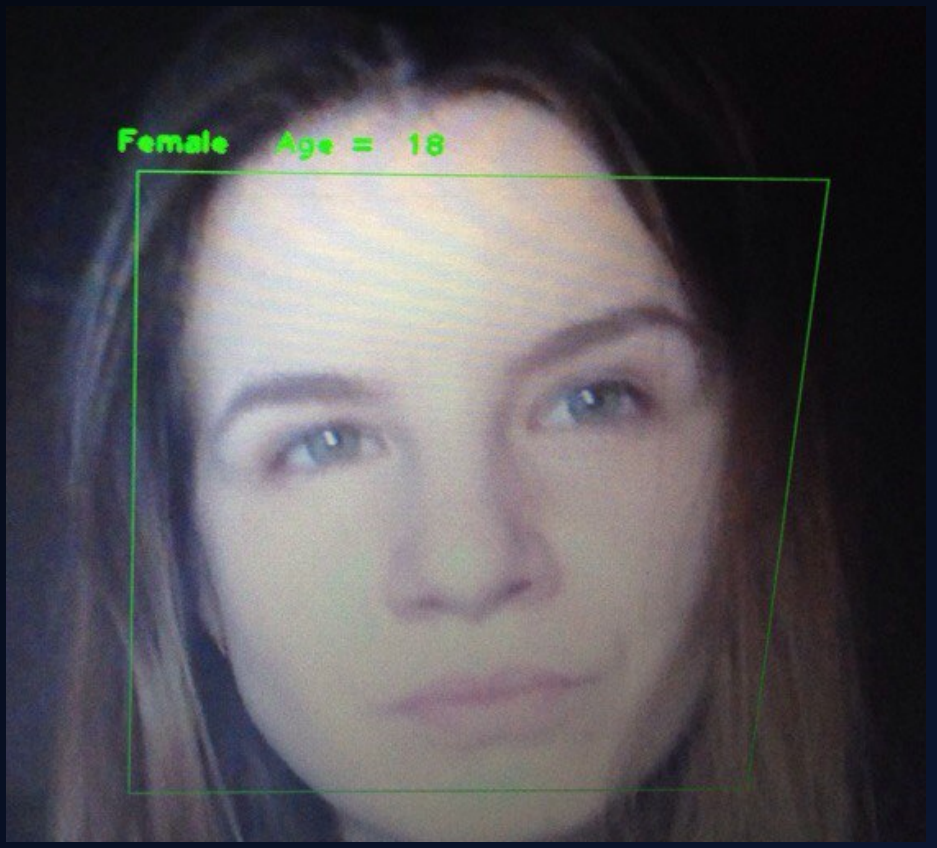
- Математическое ожидание для возраста

Реализация

- ПК Intel Core i5-3210M CPU, 64-разрядная операционная система Windows 7
- Opencv
- Opencv DNN module
- Gender_net caffemodel
- Age_net caffemodel

Результат эксперимента





Тестирование. База Данных IJB-A

Распознавание возраста: количество видео - 1375

Простое голосование = 0.29 (398)

Среднее Арифметическое = 0.28 (385)

Среднее геометрическое = 0.27 (372)

Тестирование. База Данных Indian Movie

Распознавание пола: количество видео - 332

Простое голосование = 0.156 (52)

Среднее Арифметическое = 0.158 (53)

Среднее геометрическое = 0.158 (53)

Тестирование. База Данных Indian Movie

Распознавание возраста: количество видео - 332

Простое голосование = 0.45 (149)

Математическое ожидание = 0.42 (147)

Тестирование. База Данных Kinect

Распознавание пола: количество видео - 100

Простое голосование = 0.33

Среднее Арифметическое = 0.32

Среднее геометрическое = 0.31

План работы

- Проведение сравнительного анализа способов агрегации решений, полученных при распознавании каждого видеокадра
- Экспериментальное исследование их точности и быстродействия
- Проведение хорошего эксперимента
- Применение более сложных способов агрегации решений

Спасибо за внимание!