



NATIONAL RESEARCH UNIVERSITY

Сверточные нейронные сети в задаче распознавания пола и возраста по видеоизображению

А.С. Харчевникова, А.В. Савченко

Национальный Исследовательский Университет Высшая Школа Экономики, Нижний Новгород



Введение

В коммерческих приложениях контекстной рекламы в целях выявления определенной целевой аудитории применяется автоматическое определение пола и возраста

Надежность существующих решений остается недостаточной для практического применения

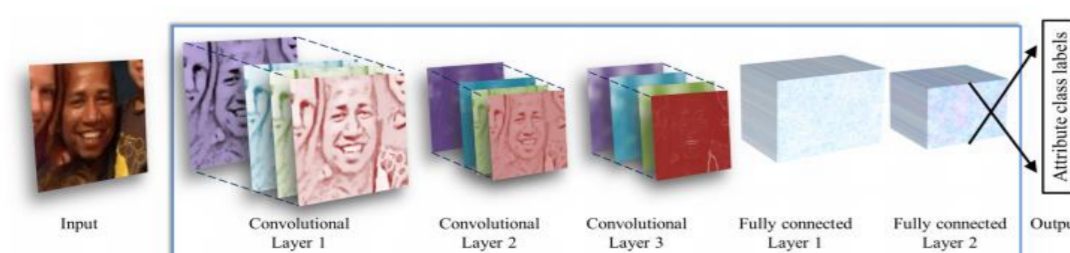
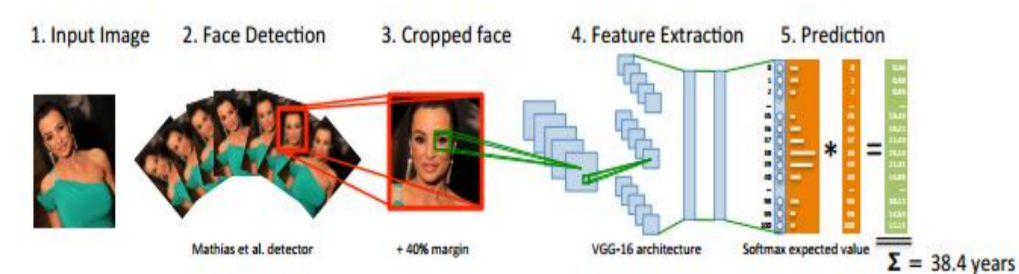


Постановка задачи

В рамках продолжения видеосъемки можно получить более ста кадров классифицируемого объекта в динамике. **Основная идея:** если каждый классификатор допускает определенную ошибку, то общее число ошибок можно снизить с помощью синтеза коллективов решающих правил (КРП)

Задача классификации видеоизображения лица: требуется отнести вновь поступающую на вход последовательность кадров $\{X(t), X(t) = \{x_{uv}(t)\}, t = 1, \overline{T}$ с изображением одного объекта к одному из L классов

В связи с эффективностью осуществления сверточных нейронных сетей (СНС) для задач классификации изображений, в настоящей работе применяется данный подход



Алгоритм

В процессе распознавания на вход СНС подадим RGB матрицу пикселей t -го кадра. Выход обычно получается в слое **Softmax**, который выдает оценку апостериорной вероятности $P(l|X(t))$

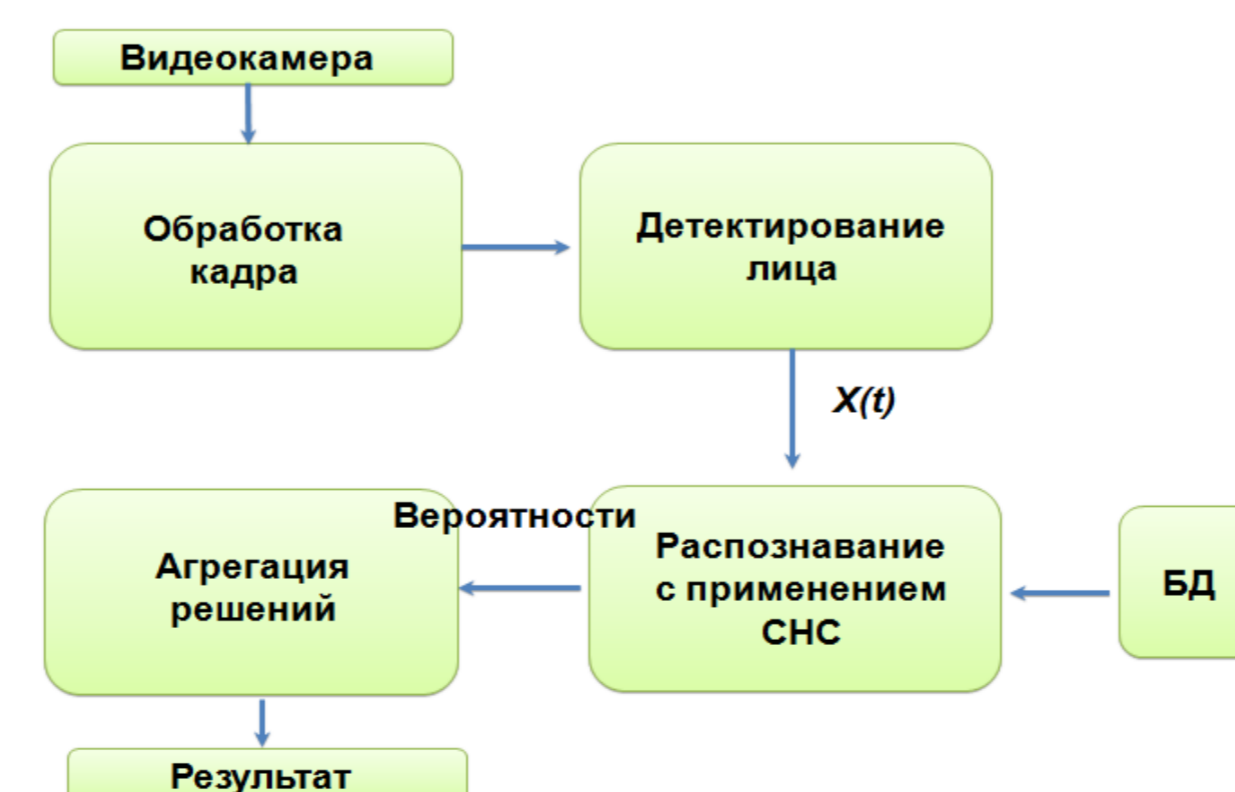
$$P(l|X(t)) = \text{softmax } z_l(t) = \frac{\exp(z_l(t))}{\sum_{j=1}^L \exp(z_j(t))}, l = 1, 2, \dots,$$

Методы агрегации

- Простое голосование
$$l^* = \underset{l = 1, \overline{L}}{\text{argmax}} \sum_{t=1}^T \delta(l^*(t) - l)$$
- Среднее арифметическое (правило суммы)
$$l^* = \underset{l = 1, \overline{L}}{\text{argmax}} \frac{1}{T} \sum_{t=1}^T P(l|X(t))$$
- Среднее геометрическое (правило произведения)
$$l^* = \underset{l = 1, \overline{L}}{\text{argmax}} \prod_{t=1}^T P(l|X(t)) = \underset{l = 1, \overline{L}}{\text{argmax}} \sum_{t=1}^T \log P(l|X(t))$$
- Оценка математического ожидания

$$l^* = \sum_{l=1}^L P(l|X(t)) \cdot l$$

Схема алгоритма



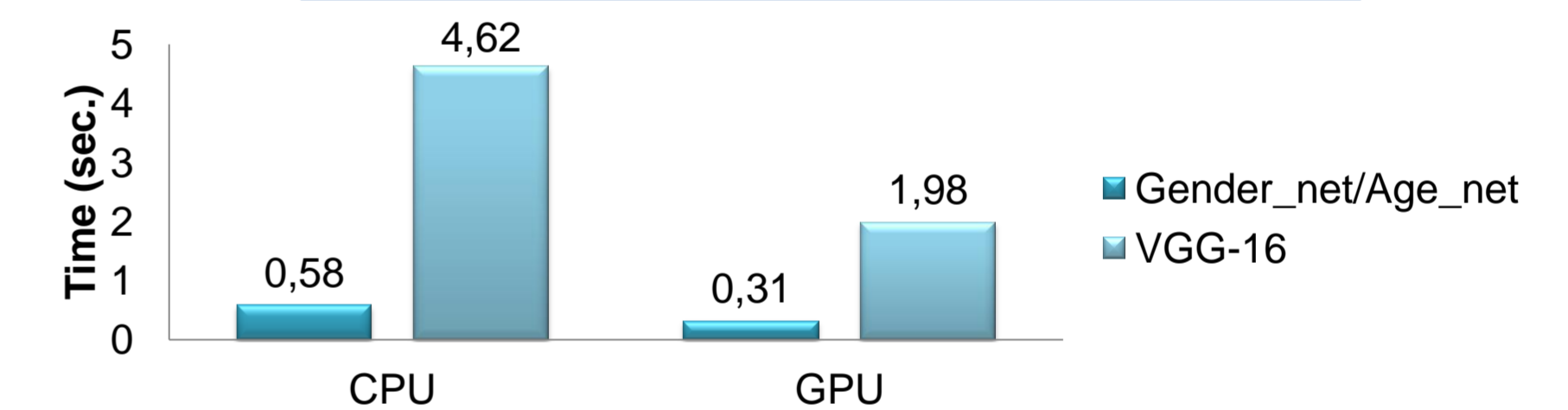
Экспериментальные результаты

Наборы данных: IARPA Janus Benchmark A (JJB-A), Indian Movie и EURECOM Kinect, которые состоят из видео кадров с известной информацией пола и возраста

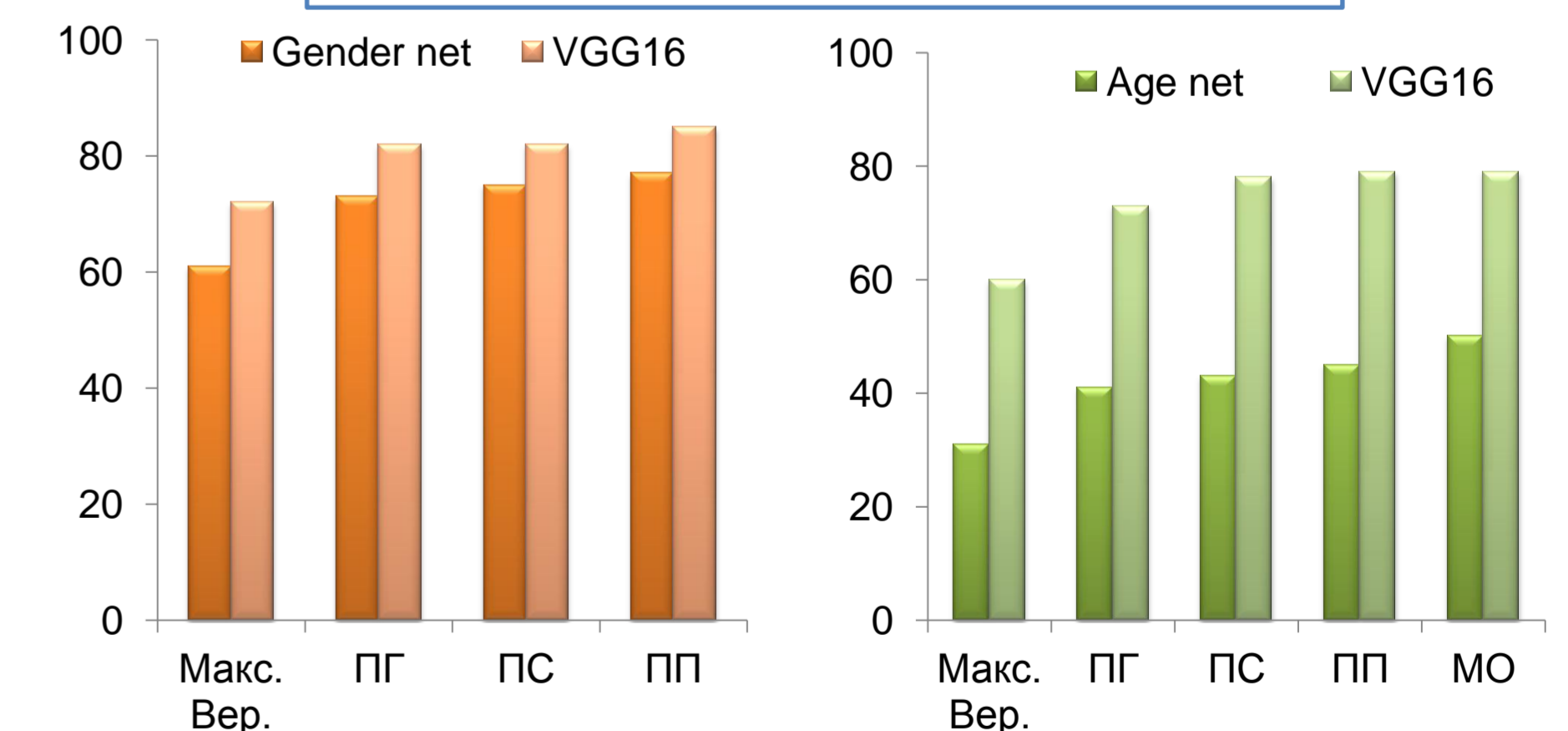


Сравниваются СНС модели: Age_net и Gender_net и глубокая VGG-16, обученная для распознавания пола и возраста

Среднее время распознавания



Точность



Заключение

Проведенное экспериментальное исследование продемонстрировало увеличение точности распознавания при реализации КРП в сравнении с традиционным подходом принятия решения. Оценка среднего геометрического (правило произведения) с нормализацией входных данных оказалась более точной в задаче классификации по полу. В то же время наиболее точный результат распознавания возраста достигается с применением алгоритма оценки математического ожидания