

Идентификация лиц: быстро, точно, без смс и регистрации

Андриянов Никита Андреевич,
к.т.н., доцент, Финансовый
университет при Правительстве РФ

О докладчике

Никита Андриянов



*Финансовый университет при
Правительстве РФ*

Доцент, заведующий лабораторией
компьютерного зрения



Мой
ОПЫТ:

В обработке изображений с 2010 года, в глубоком обучении с 2017 года, принимал участие в разработке автоматизированных систем анализа спутниковых снимков, рентгеновских изображений и робота для сбора плодовых растений. В 2023 году руководил НИР по разработке алгоритмов идентификации лиц. Вице-чемпион первого ЧМ по науке (2023)

О себе:

Люблю спорт, особенно теннис)
Пытаюсь познавать мир, перевожу песни и
немного пишу рассказы.

Как сделать систему
распознавания лиц точнее,
быстрее, надежнее?

Демотиватор

ImageNet (1000-классовая база изображений)

Метрика	Значение
Топ-5 ошибка для человека	5.1%
Топ-5 ошибка для нейросети	2.0%

Labeled Faces in the Wild (база изображений лиц в естественных условиях)

Метрика	Значение
Ошибка верификации для человека	2.47%
Ошибка верификации для нейросети	0.17%

Распознавание лиц – НЕ классификация!

Основные задачи компьютерного зрения

Classification



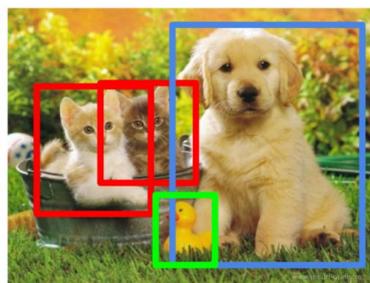
CAT

Classification
+ Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance
Segmentation



CAT, DOG, DUCK

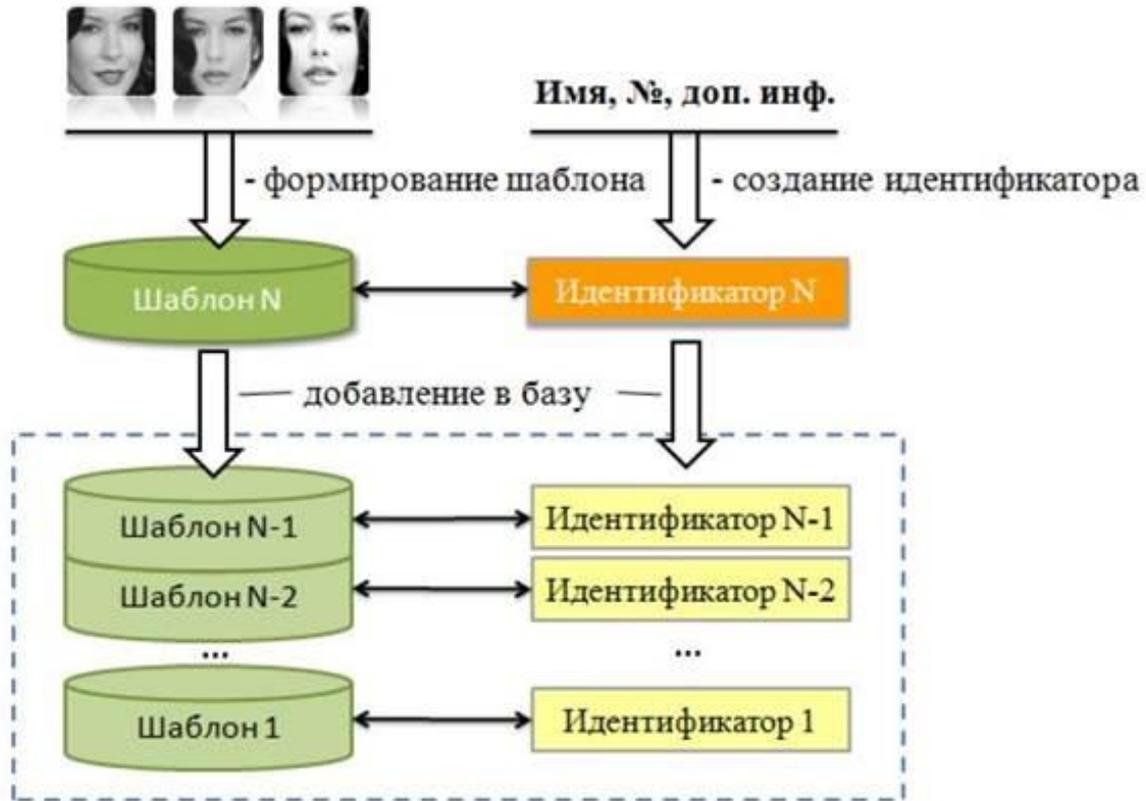
Single object

Multiple objects

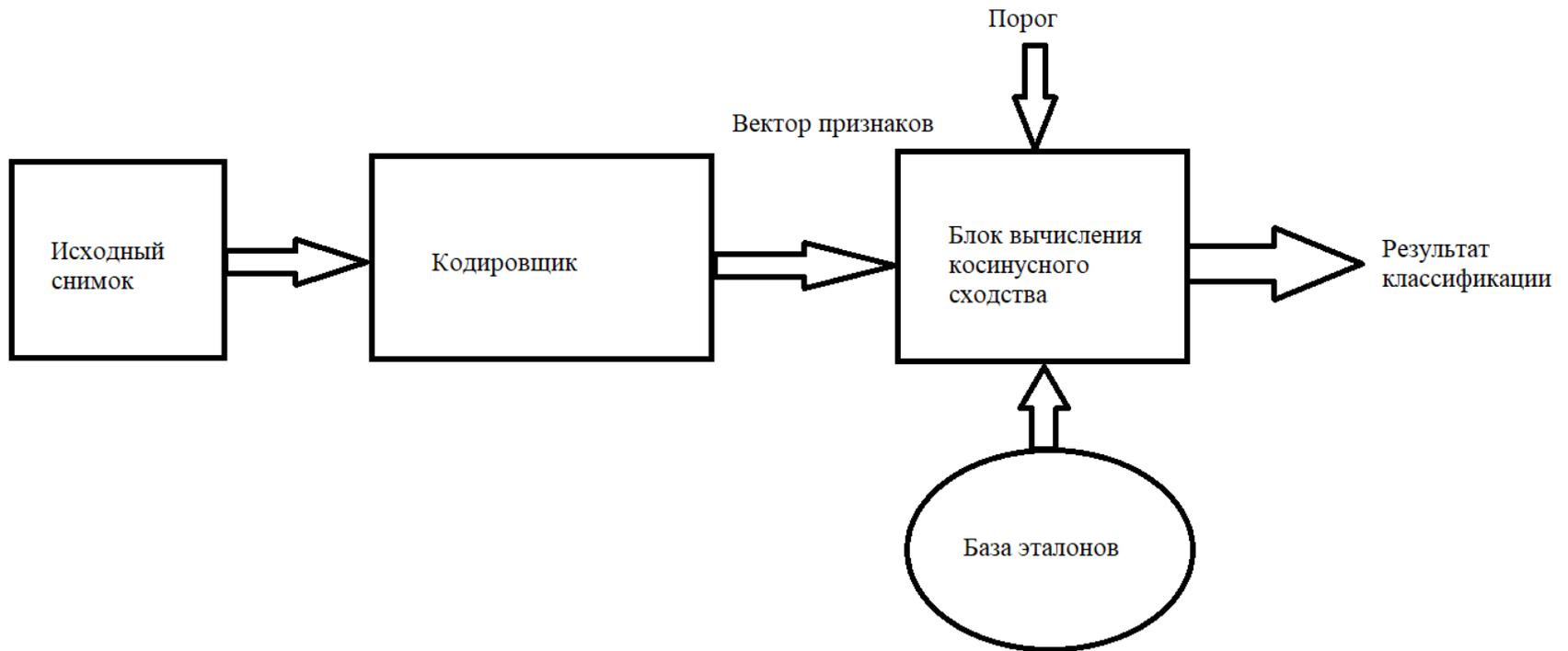
Традиционный классификатор всегда дает один из N известных в рамках обучения ответов



Много классов



Метрические системы

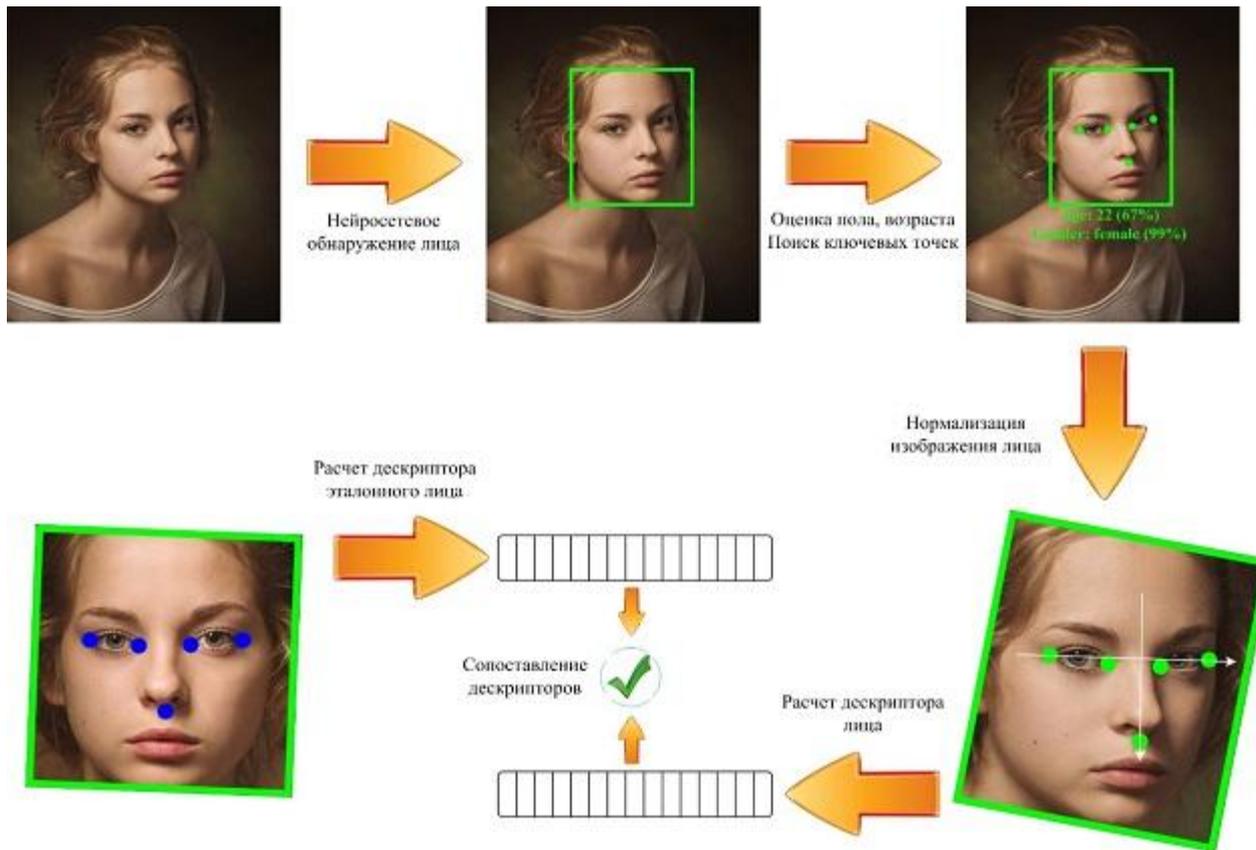


Обнаружение лица

Метод Виолы-Джонса

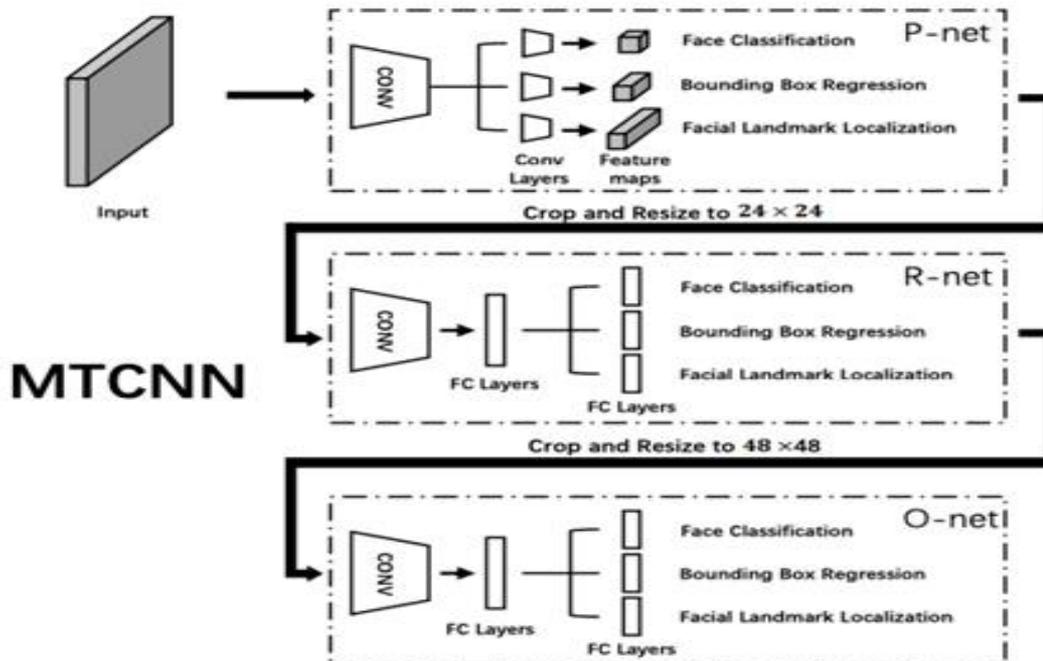


Нормализация лиц



Мультизадачные сети

Мультизадачная каскадная сверточная нейронная сеть MTCNN:



P-Net

для генерации гипотез

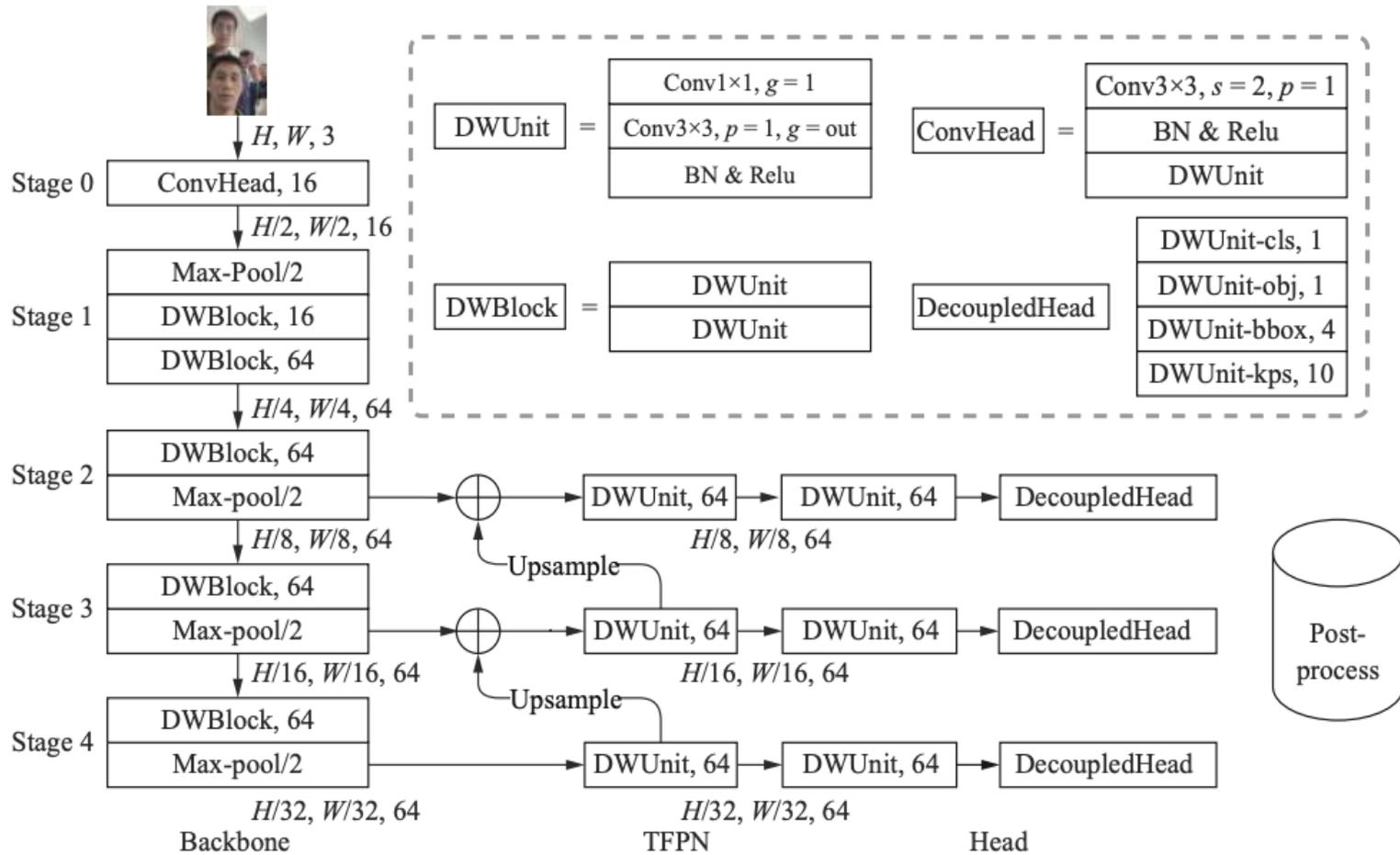
R-Net

для отбрасывания ложных гипотез

O-Net

для корректировки
ограничительных рамок
и классификации лица

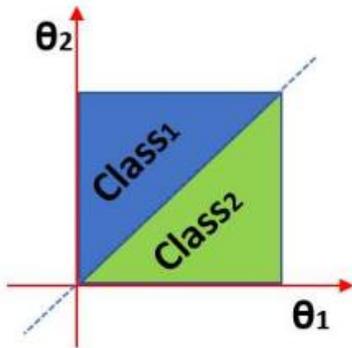
YuNet



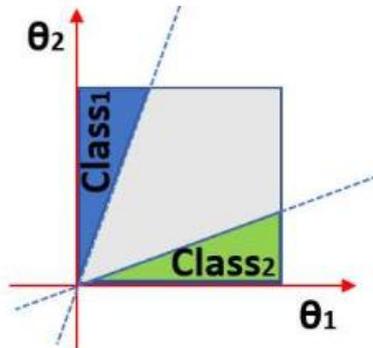
ArcFace потери

$$\mathcal{L}_{\text{CosFace}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp\{s(\cos(\theta_{y_i,i}) - m)\}}{\exp\{s(\cos(\theta_{y_i,i}) - m)\} + \sum_{j \neq y_i} \exp\{s \cos(\theta_{j,i})\}}$$

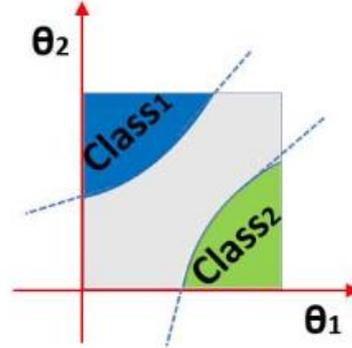
$$\mathcal{L}_{\text{ArcFace}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp\{s \cos(\theta_{y_i,i} + m)\}}{\exp\{s \cos(\theta_{y_i,i} + m)\} + \sum_{j \neq y_i} \exp\{s \cos(\theta_{j,i})\}}$$



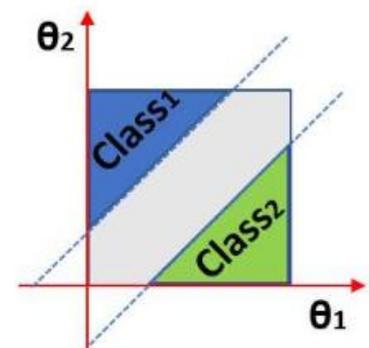
Softmax



SphereFace



CosFace



ArcFace

Датасет LFW



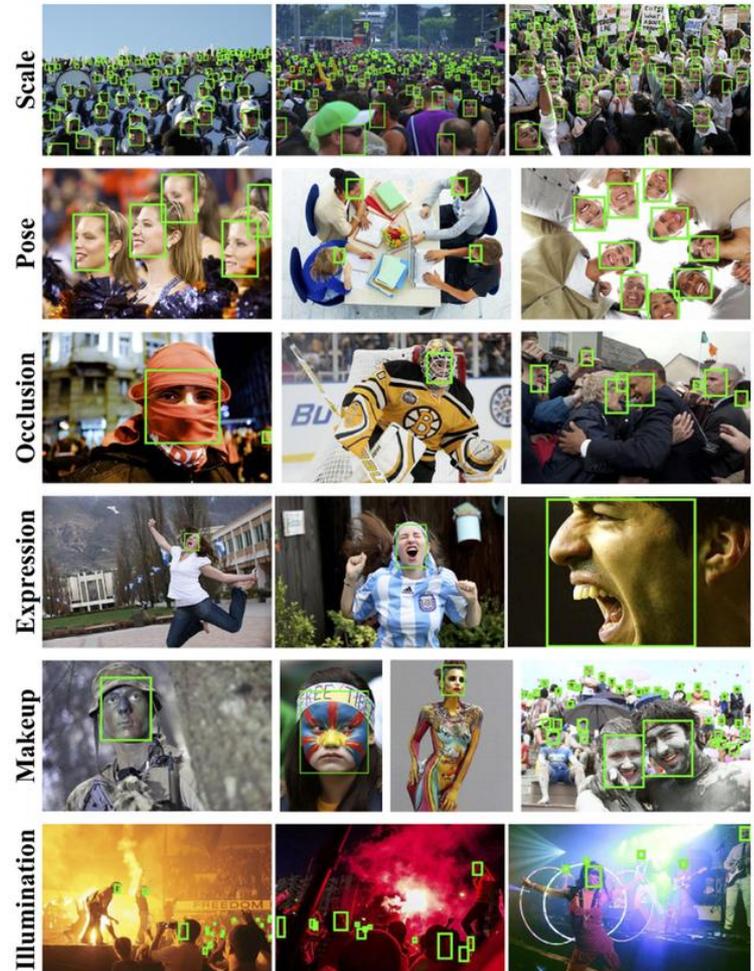
Для задачи распознавания лиц хорошо подходит датасет Labelled Faces in the Wild (LFW). Чтобы создать данный датасет использовался метод Виолы-Джонса. Итого получилось собрать 13 233 изображения, содержащих информацию о 5 749 людях. При этом в нем всего 1680 человек, для которых имеется не менее двух фотографий.

Ссылка:

<https://www.kaggle.com/datasets/jessicali9530/lfw-dataset>

Датасет WIDER FACE

Он был разработан для решения задачи обнаружения лиц в реальных условиях с широким разнообразием масштабов, точек обзора и фоновых условий. Набор данных WIDER Face содержит более 32 000 изображений, в которых представлены более 40 000 размеченных лиц. Ссылка: <https://www.kaggle.com/datasets/mksaad/wider-face-a-face-detection-benchmark>



Меры для сравнения

Performance: число кадров в секунду

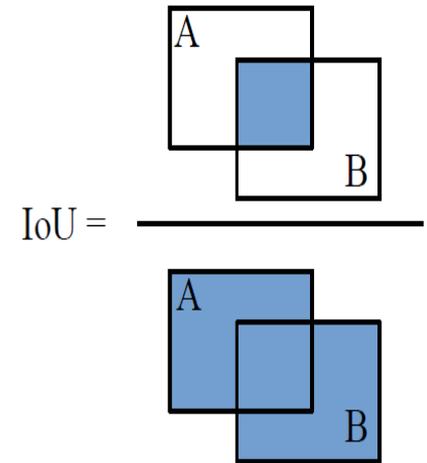
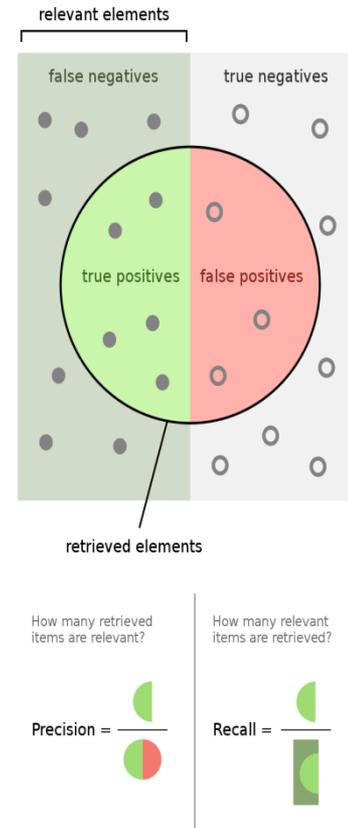
Accuracy: доля верных распознаваний

Precision: доля верных распознаваний среди класса (по прогнозам)

Recall: доля верных распознаваний среди класса (по реальным меткам)

mAP: средняя точность по всем классам под кривой PR (в детекторе)

IoU: мера совпадения ограничивающих рамок при детекции



SOTA

FaceNet (PyTorch) достигает высокой точности на наборе LFW (99.65%) и имеет относительно высокую скорость обработки* (примерно 14.5 FPS).

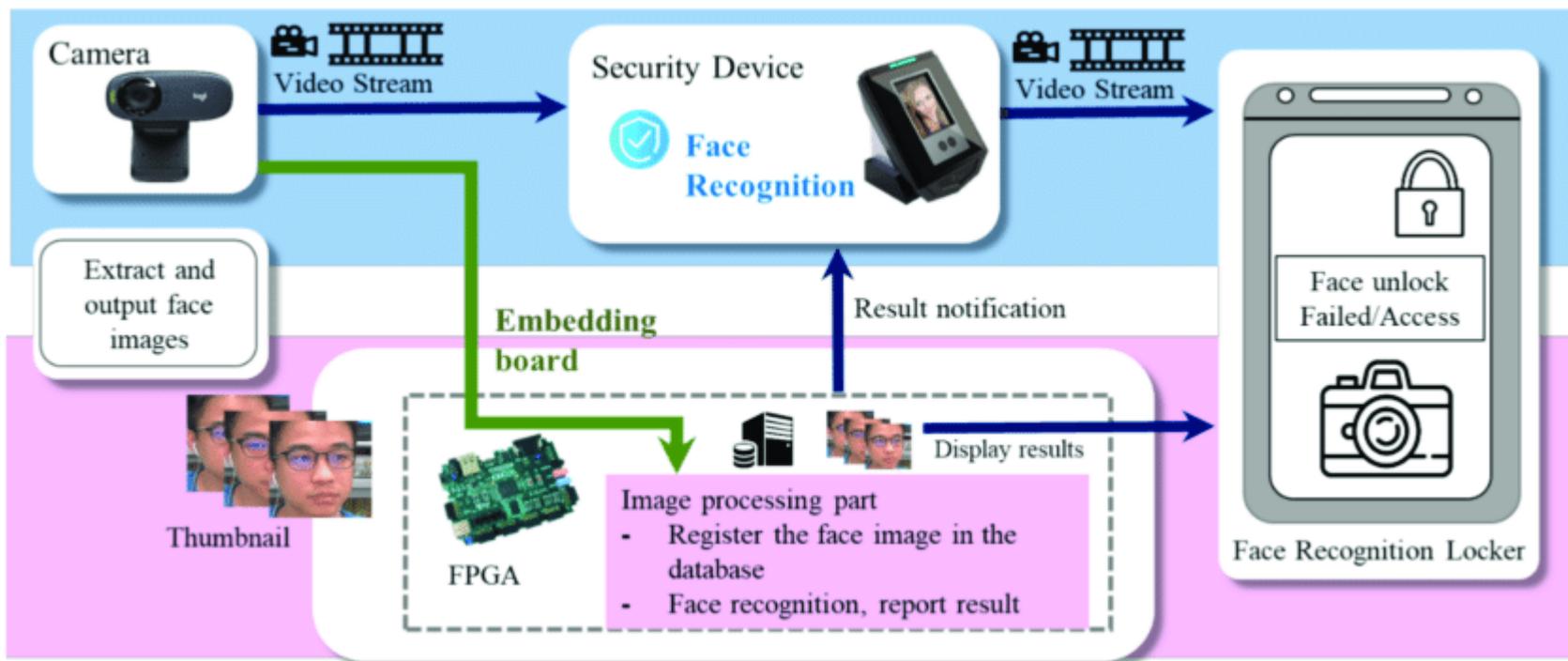
OpenFace (TensorFlow) имеет немного меньшую точность на наборе LFW (93.80%), но все равно обеспечивает приемлемую скорость обработки (примерно 6.8 FPS).

GhostFaceNet (TensorFlow) обладает высокой точностью на наборе LFW (99.68%), но скорость обработки немного ниже (примерно 4.4 FPS).

VGG-Face (TensorFlow) показывает хорошую точность на наборе LFW (98.78%), но обладает самой низкой скоростью обработки (примерно 1.38 FPS).

ArcFace (TensorFlow) также демонстрирует высокую точность на наборе LFW (99.41%), но его скорость обработки немного ниже (примерно 2.9 FPS).

Конечные устройства

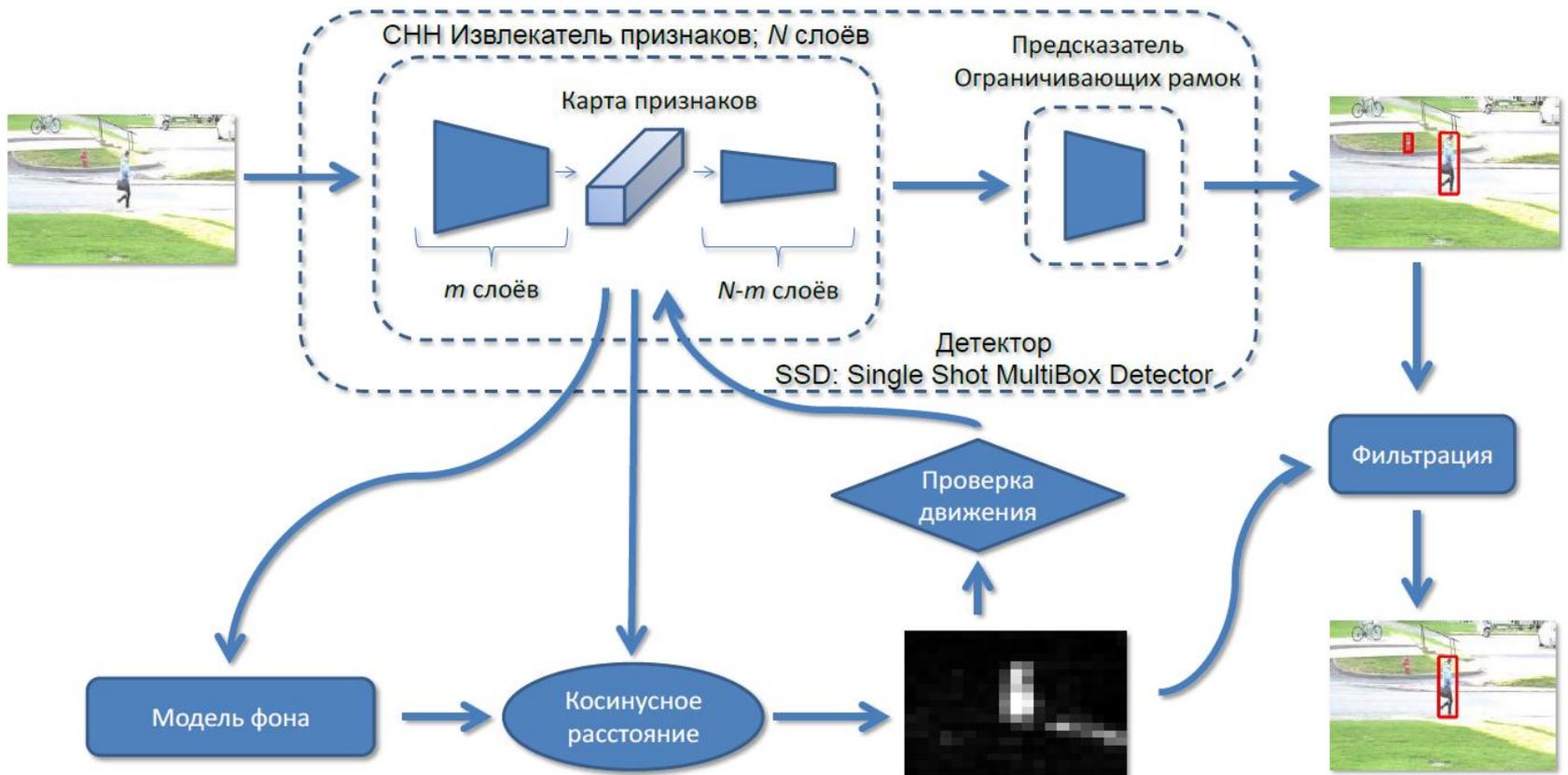


Оптимизация

- 1) Разработка нового алгоритма
- 2) Нейроархитектурный поиск
- 3) Прунинг
- 4) Квантование
- 5) Дистилляция
- 6) OpenVINO и т.п.
- 7) Железо мощнее



Amphibian Detector



Сравнение моделей

Тестовые снимки



Очки (318x159)



Маска (1280x800)



Ночь (183x275)

Сравнение моделей

Тестовые снимки WIDER Face



Выборная кампания (1024x683)

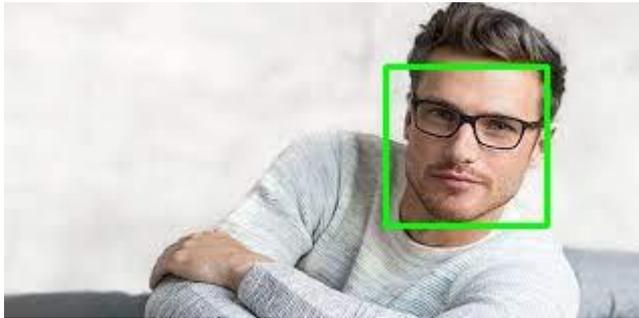
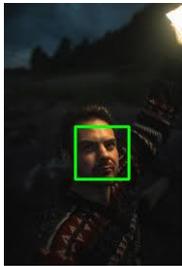


Лыжники (1024x683)

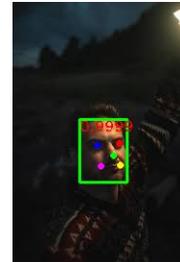
Результаты

Обнаружения по тестовым кадрам

VJ



YuNet



Результаты

Обнаружения по тестовым кадрам WIDER Face

VJ



5 верных детекций, 3 ложных

YuNet



15 верных детекций, не более 1 ложной

Производительность

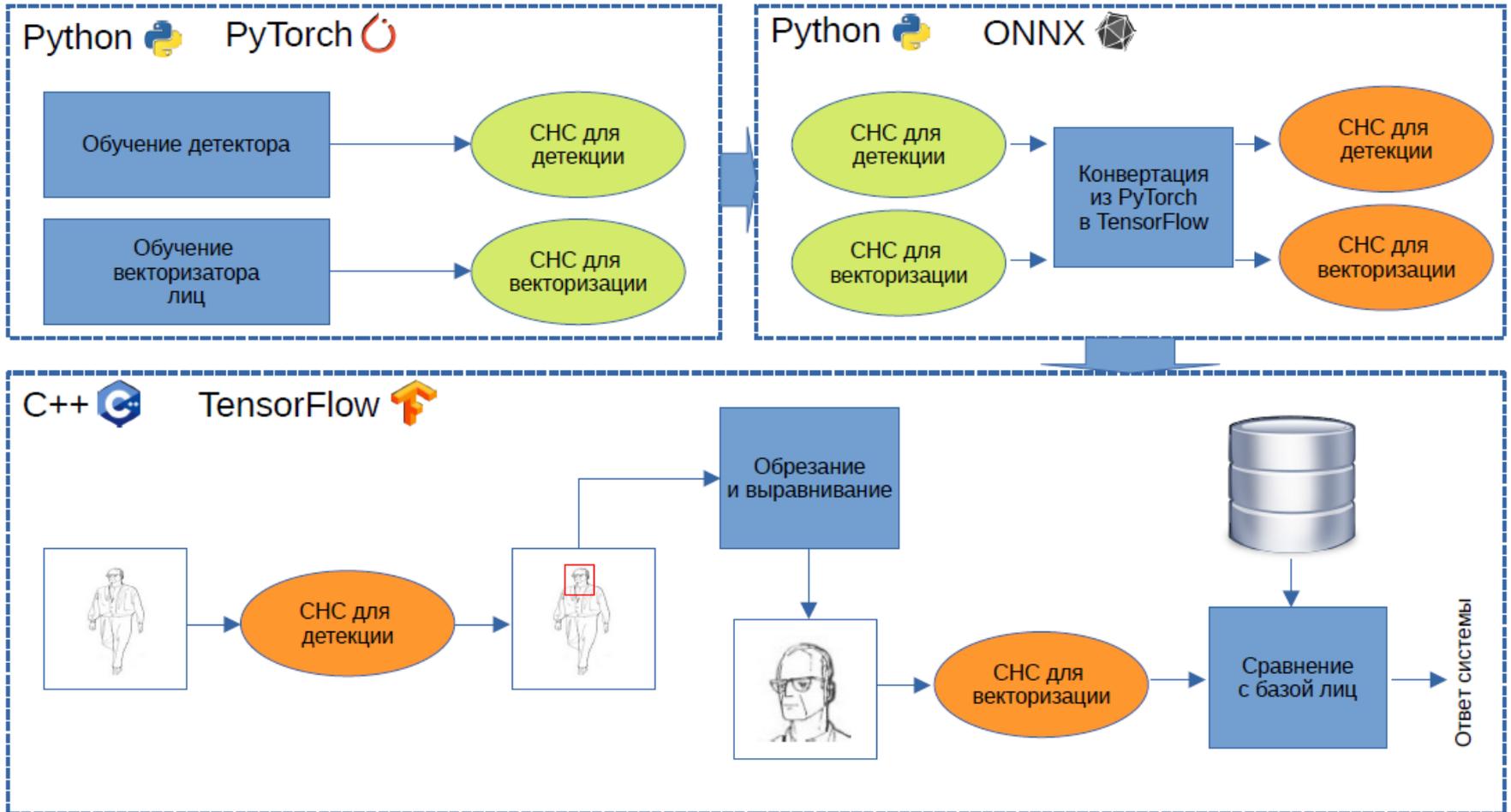
Тестовые изображения

Изображение	Среднее время исполнения VJ, мс	Среднее время исполнения YuNet, мс
Очки	38,6	12,3
Маска	361	156
Ночь	17,2	11,2

Изображения Wider Face

Изображение	Среднее время исполнения VJ, мс	Среднее время исполнения YuNet, мс
Выборная кампания	660	113
Лыжники	640	116

Перенос во встраиваемые системы



Спасибо за внимание!

Никита Андриянов

Email: nikita-and-nov@mail.ru

TG: [@nikita_and_nov](https://t.me/@nikita_and_nov)

VK: [@nikandriyanov](https://vk.com/nikandriyanov)