



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

**Харчевникова Ангелина Сергеевна**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**Эффективные алгоритмы идентификации лиц на видео на  
основе выбора качественных кадров**

Научный руководитель:  
Доктор технических наук, профессор  
кафедры ИСиТ,  
Савченко Андрей Владимирович

Нижний Новгород, 2020

# Предметная область



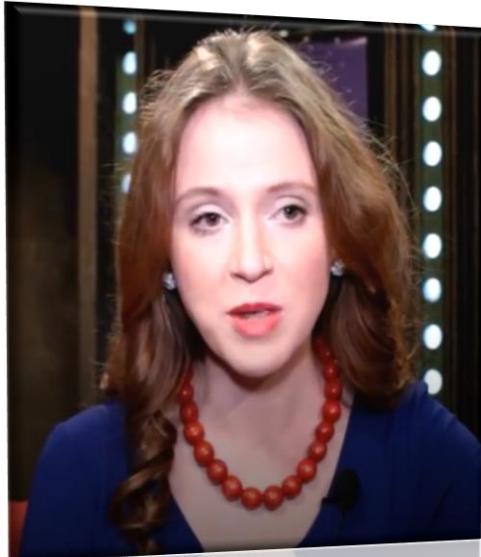
- Сфера безопасности:  
поиск  
злоумышленников или  
пропавших людей

- Face ID: лицо является  
самым надежным  
идентификатором  
человека





# Актуальность



Надежность и точность технологий идентификации непосредственно зависит от качества изображений, приходящих на вход алгоритма

Анализ каждого кадра



Снижение  
эффективности





## Цель и задачи

---

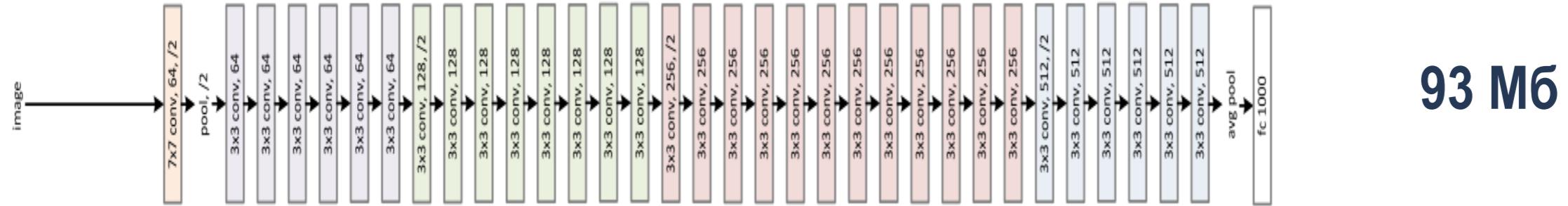
Повышение надежности и скорости работы алгоритмов идентификации лиц на видео с помощью технологий извлечения качественных кадров

- Провести обзор существующих алгоритмов распознавания лиц по изображению с акцентированием внимания на технологии глубокого обучения
- Рассмотреть подходы извлечения ключевых кадров\* по видеопоследовательности
- Провести обзор алгоритмов определения качества изображения лица
- Разработать прототип системы идентификации лиц на основании методов выявления ключевых кадров
- Провести экспериментальные исследования по анализу надежности, производительности и точности работы каждого из алгоритмов

\* Под ключевыми кадрами понимаются видеоизображения достаточного качества, которые в полной мере содержат информацию о видео

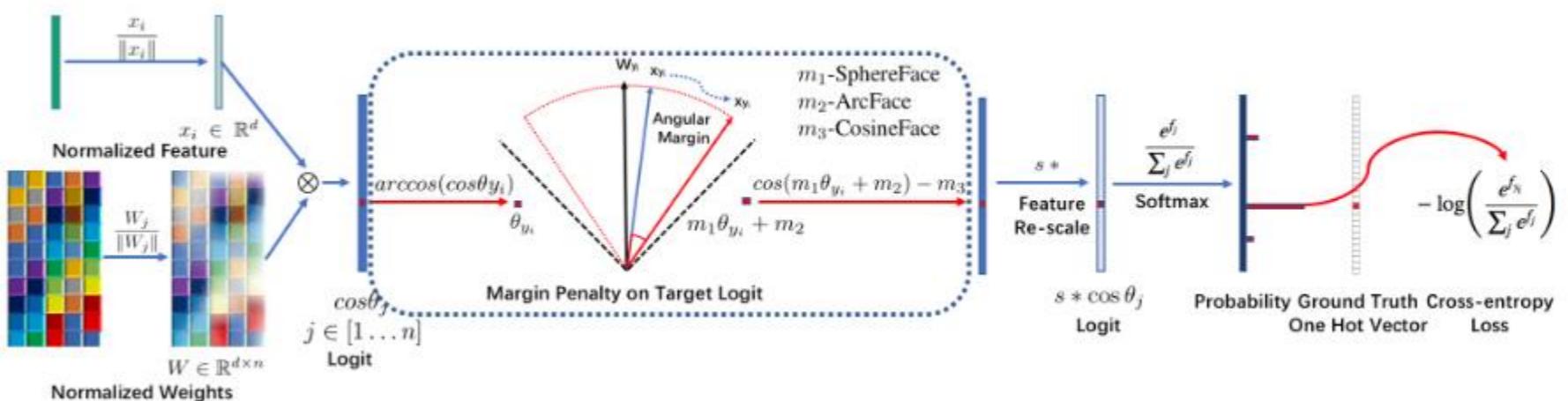


# Сверточные нейронные сети

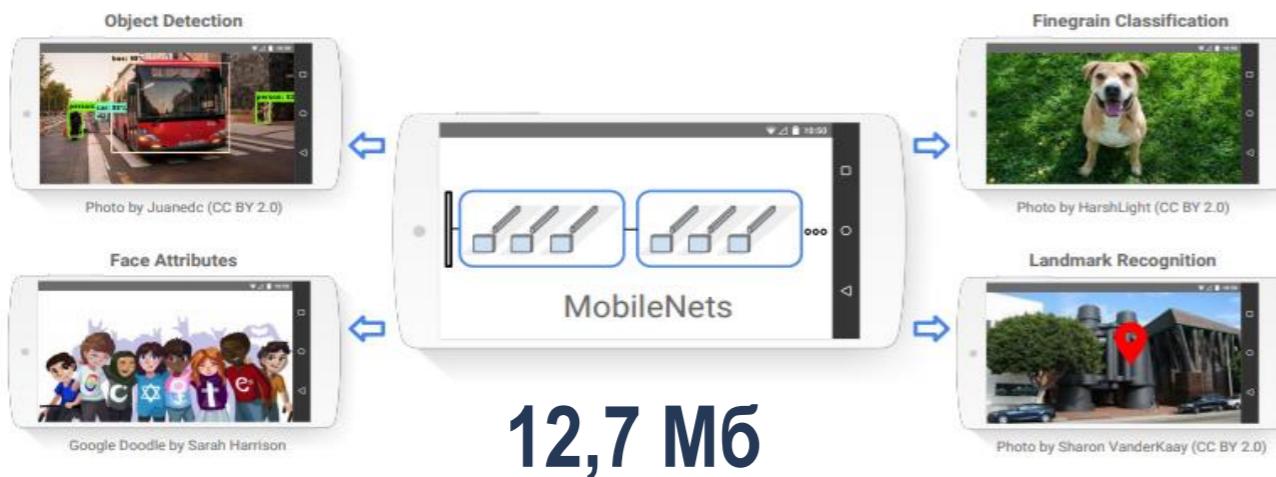


ResNet-50 He, K. Deep residual learning for image recognition, 2016

170 МБ



ARCFace, Jiankang D. ArcFace: Additive Angular Margin Loss for Deep Face Recognition, 2019



12,7 МБ

MobileNet trained on VGGFace2  
Savchenko A.V. Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet, 2019



# Задача оценки качества кадров

$$\{X(t)\}, X(t) = \|x_{uv}(t)\|, t = \overline{1, T}$$

Входная видеопоследовательность

$$Y_t = 0.2126 \cdot R + 0.7152 \cdot G + 0.0722 \cdot B$$

Яркость (1)

$$b_t = \frac{Y_t}{255 \cdot h \cdot w}$$

$$c_t = \sqrt{\frac{Y_t^2}{h \cdot w} - b_t^2}$$

Контраст (2)

$$\hat{I} = \frac{I_{i,j} - \mu_{i,j}}{\theta_{i,j} + c}$$

BRISQUE (3)

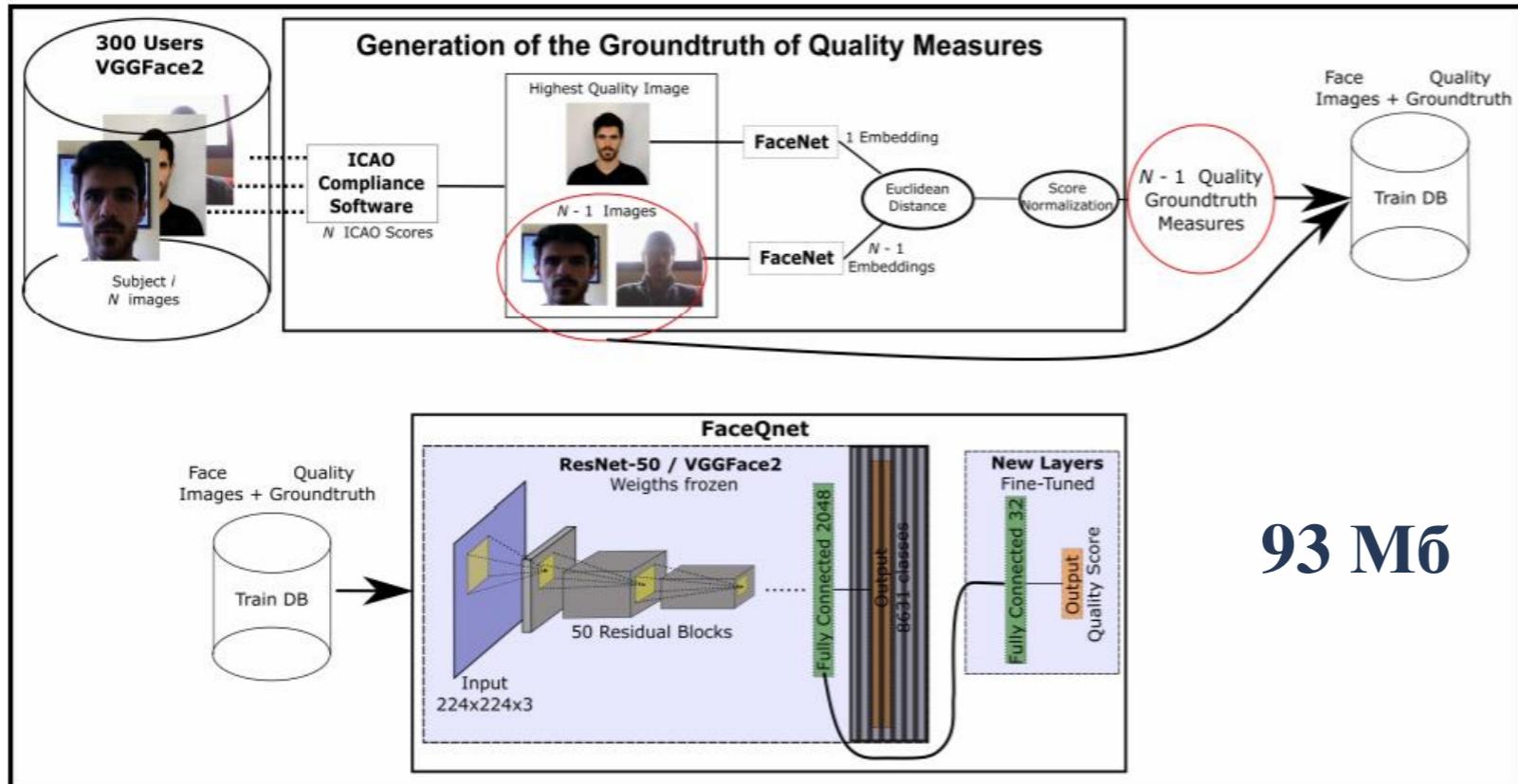
$$Q^*(X(1)) \geq Q^*(X(t)) \geq Q^*(X(T))$$

$$K \in \{1 \dots T\}$$

Последовательность  $K$  качественных кадров



# Оценка качества на основе глубокого обучения

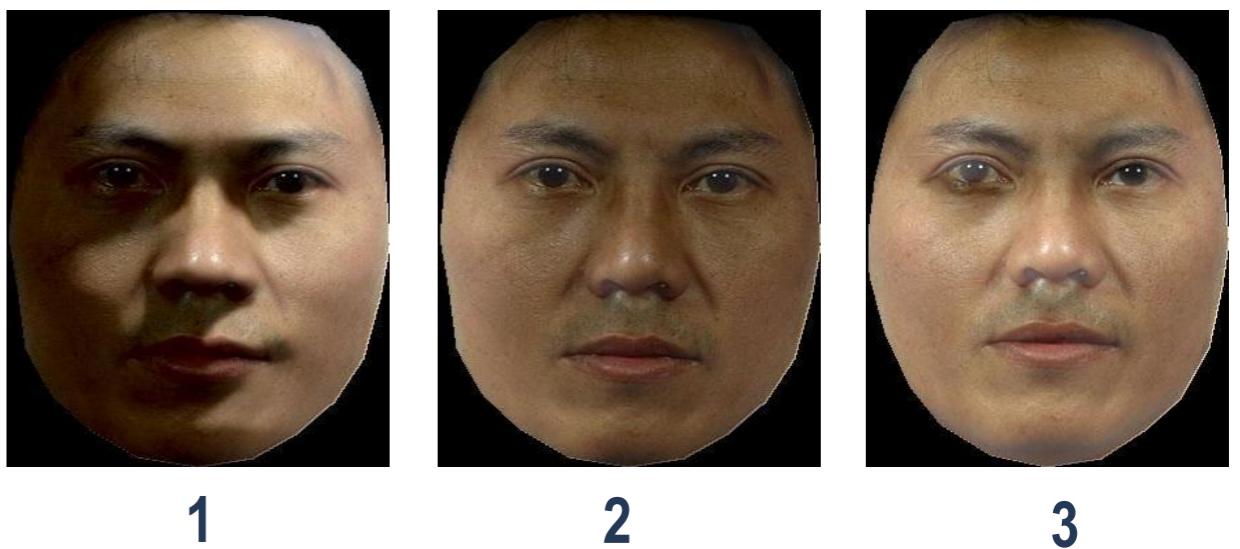


## FaceQNet

Hernandez-Ortega, J. *FaceQnet: Quality Assessment for Face Recognition based on Deep Learning*, 2019

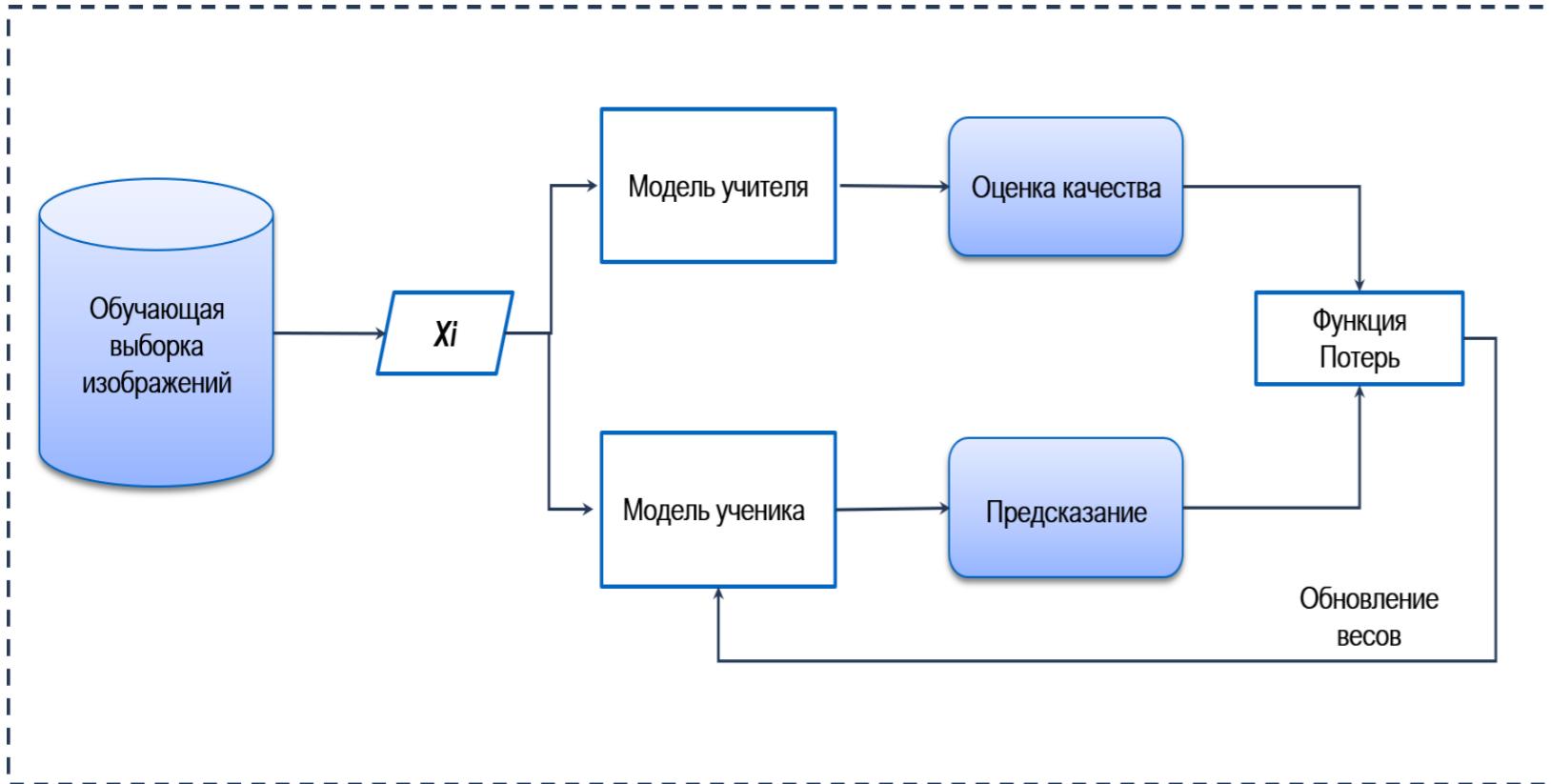
## FIIQA: Face Image Illumination Quality Dataset

Zhang, L. *Illumination quality assessment for face images: A benchmark and a convolutional neural networks based model*, 2017

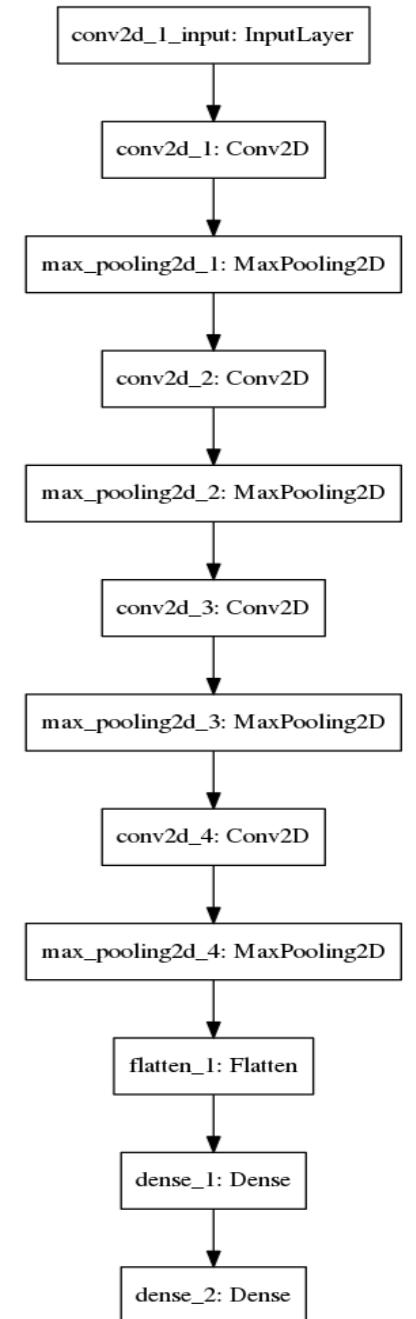




# Перенос знаний сети



**FaceQNet mobile** 12.94 Мб  
**FaceQNet light** 4.4 Мб

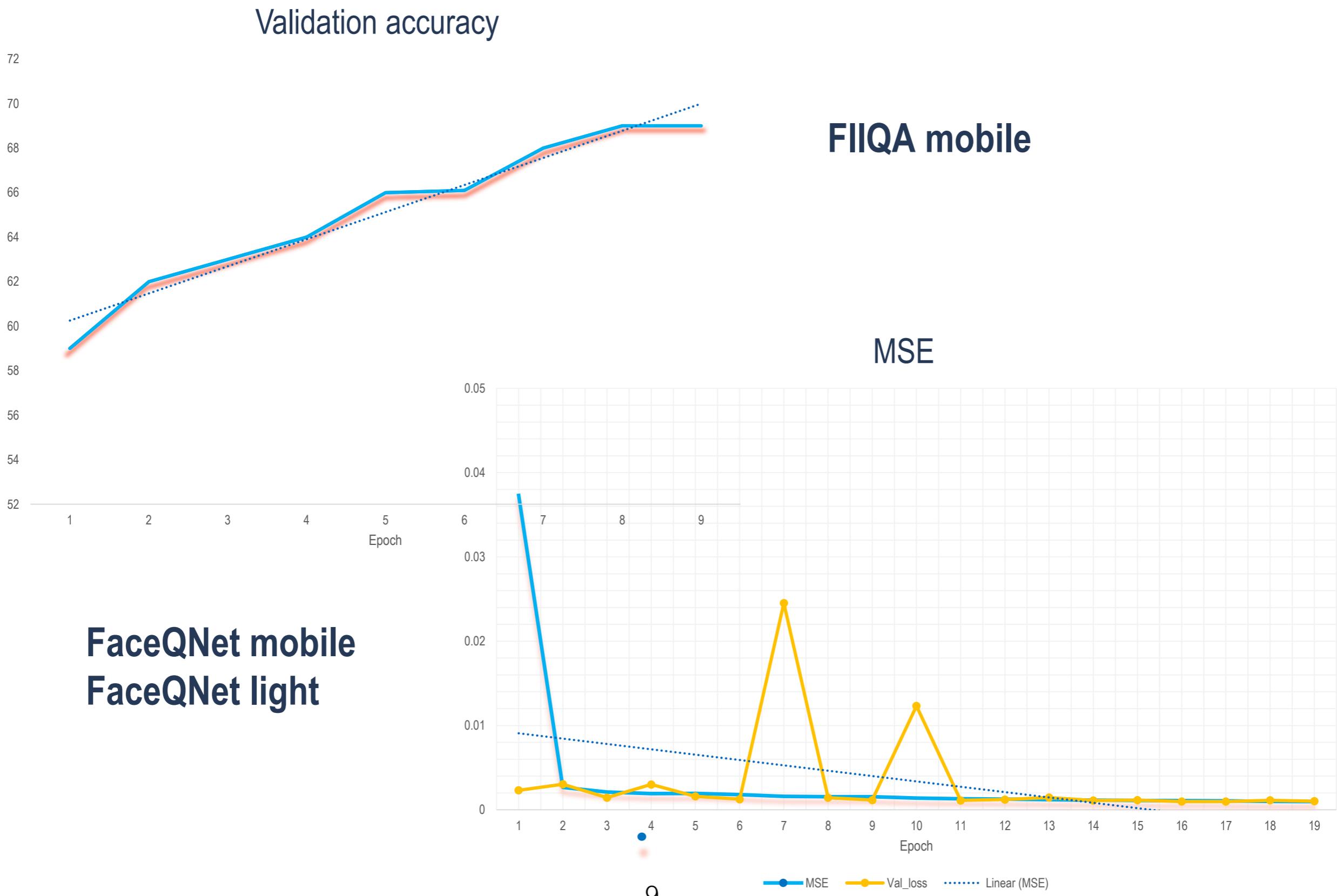


**FaceQNet** = 0.595  
**FaceQNet mobile** = 0.601  
**FaceQNet light** = 0.384

**FaceQNet** = 0.352  
**FaceQNet mobile** = 0.391  
**FaceQNet light** = 0.636

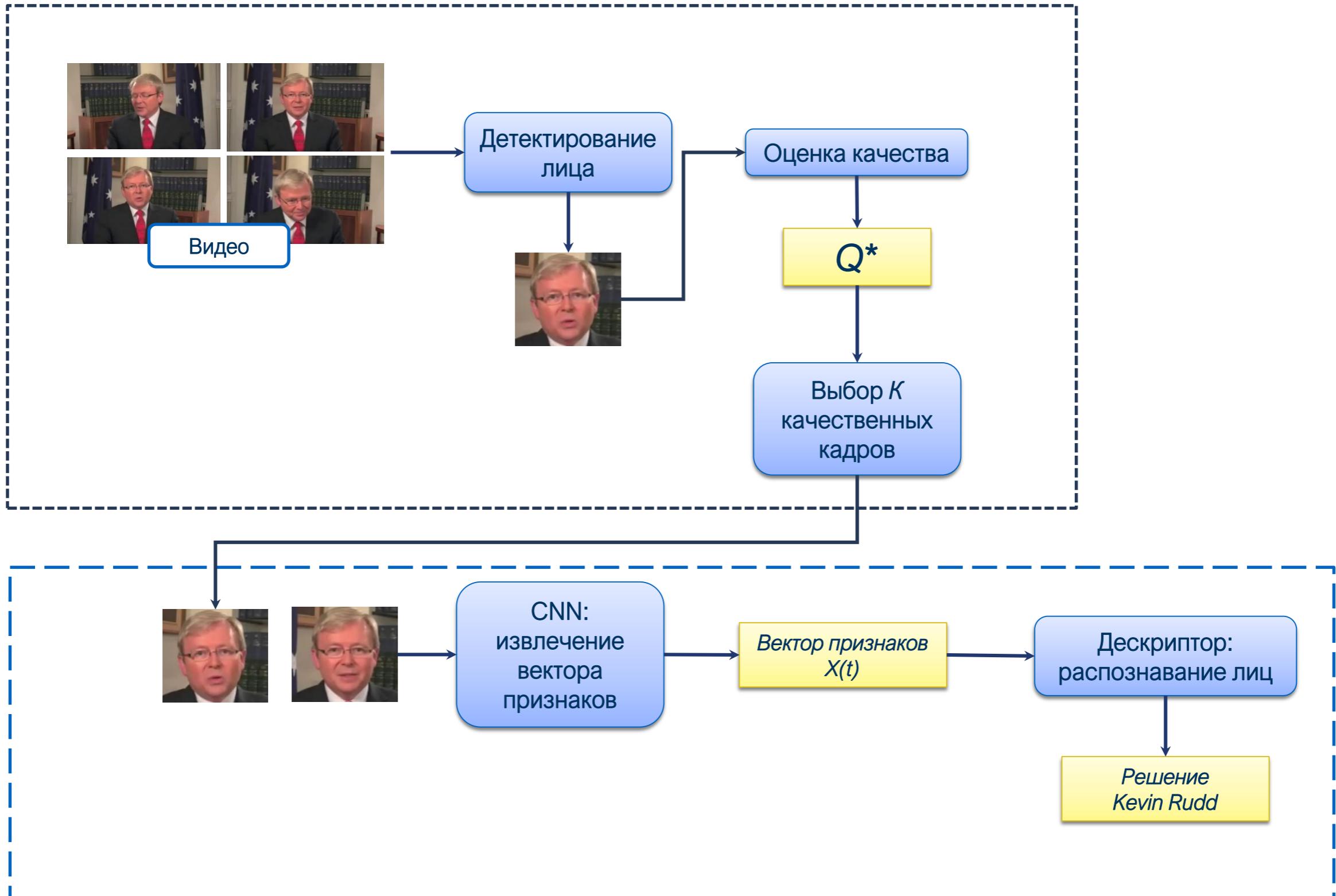


# Обучение моделей оценки качества





# Структурная схема системы





# Графический интерфейс системы

The screenshot displays the graphical interface of a Face ID system. It includes:

- A main window titled "Face ID" showing a video feed of a person with a blue bounding box around their face. The result "angelina" is displayed below the video.
- An open file dialog titled "Select Video" showing a file named "face.MP4" selected from the path "C:\Users\akharche\Desktop\ui\videos".
- A secondary window titled "Select Video" with the title "Collect Frames:" containing a text input field with the value "100", a checked checkbox "Load video", a "Select video" button, and a large blue "START" button.
- A preview grid labeled "Key frames:" showing a 3x2 grid of thumbnail images of the same person's face.



# Результаты экспериментов. Эффективность

TensorFlow Keras

Модель СНС	Время обработки одного кадра, мс.	
	GPU	CPU
ResNet-50	9.186	48.917
MobileNet-vgg2	6.574	20.246
InsightFace	15.735	93.407

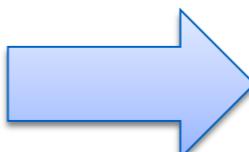
$$T = \sum_{i=1}^N t_i$$



# Эффективность

Метод оценки качества	Время обработки одного кадра, мс.	
	GPU	CPU
Luminance (1)	-	0.027
Contrast (2)	-	0.071
Brisque (3)	-	285.123
FaceQNet	12.039	47.897
FaceQNet mobile	6.351	22.088
FaceQNet light	3.111	14.213
FIIQA mobile	6.642	21.409

$$T = \sum_{i=1}^N t_i$$

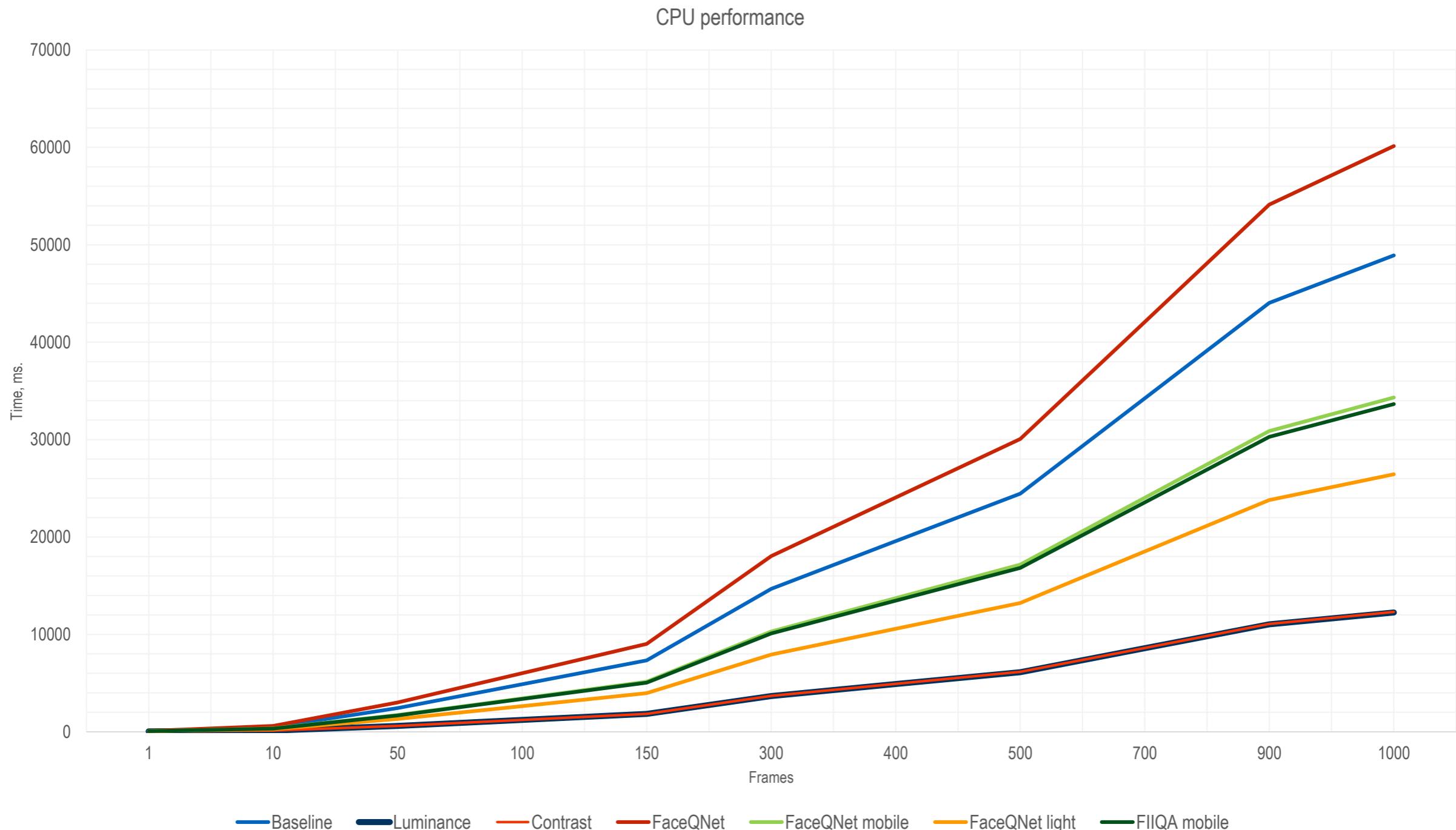


$$T^* = \sum_{i=1}^N t'_i + \sum_{i=1}^K t''_i$$



# Эффективность. CPU

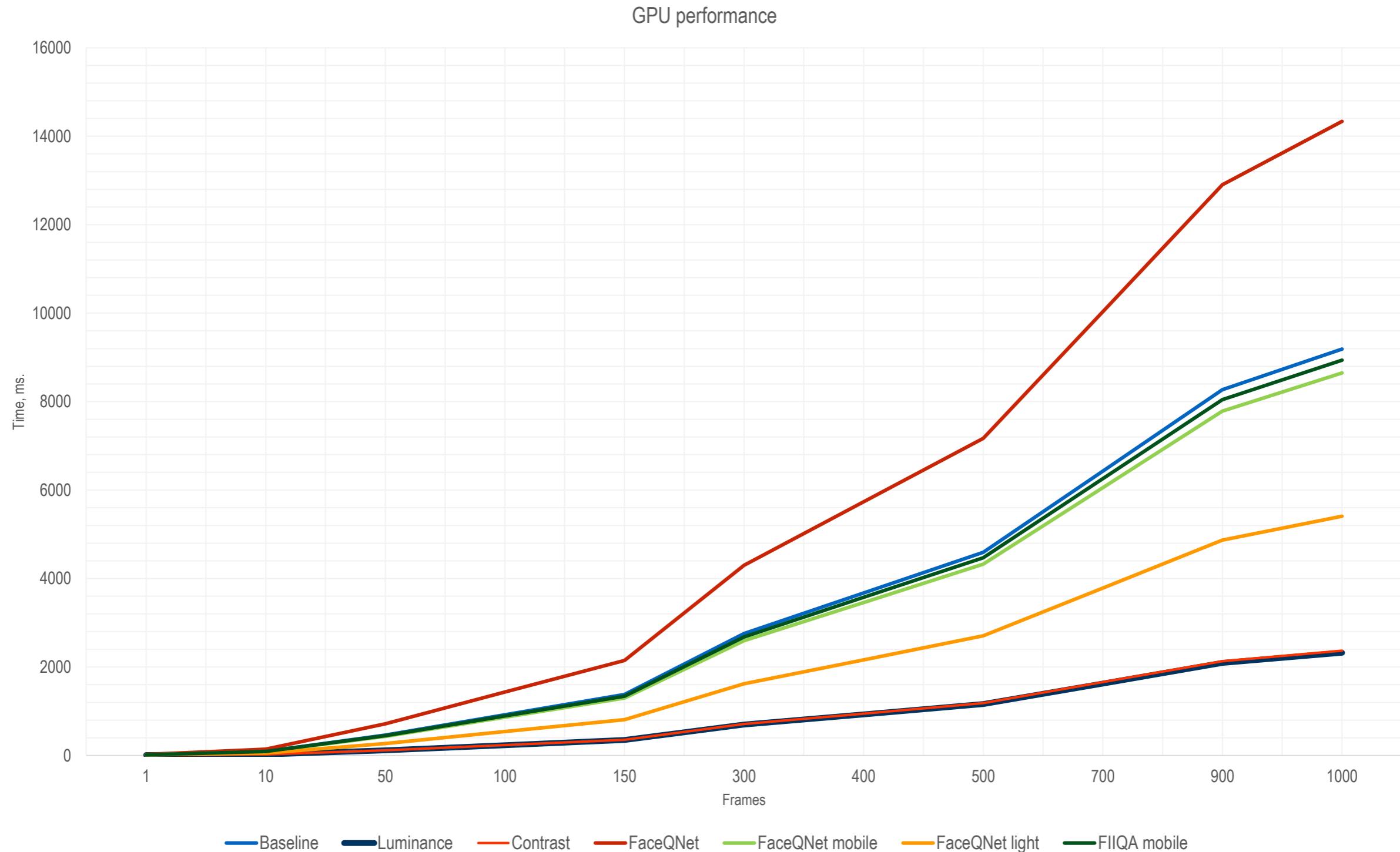
## ResNet-50





# Эффективность. GPU

## ResNet-50





# Наборы данных

- IJB-C

3530 уникальных лиц

21956 статических изображений

457512 видео кадров



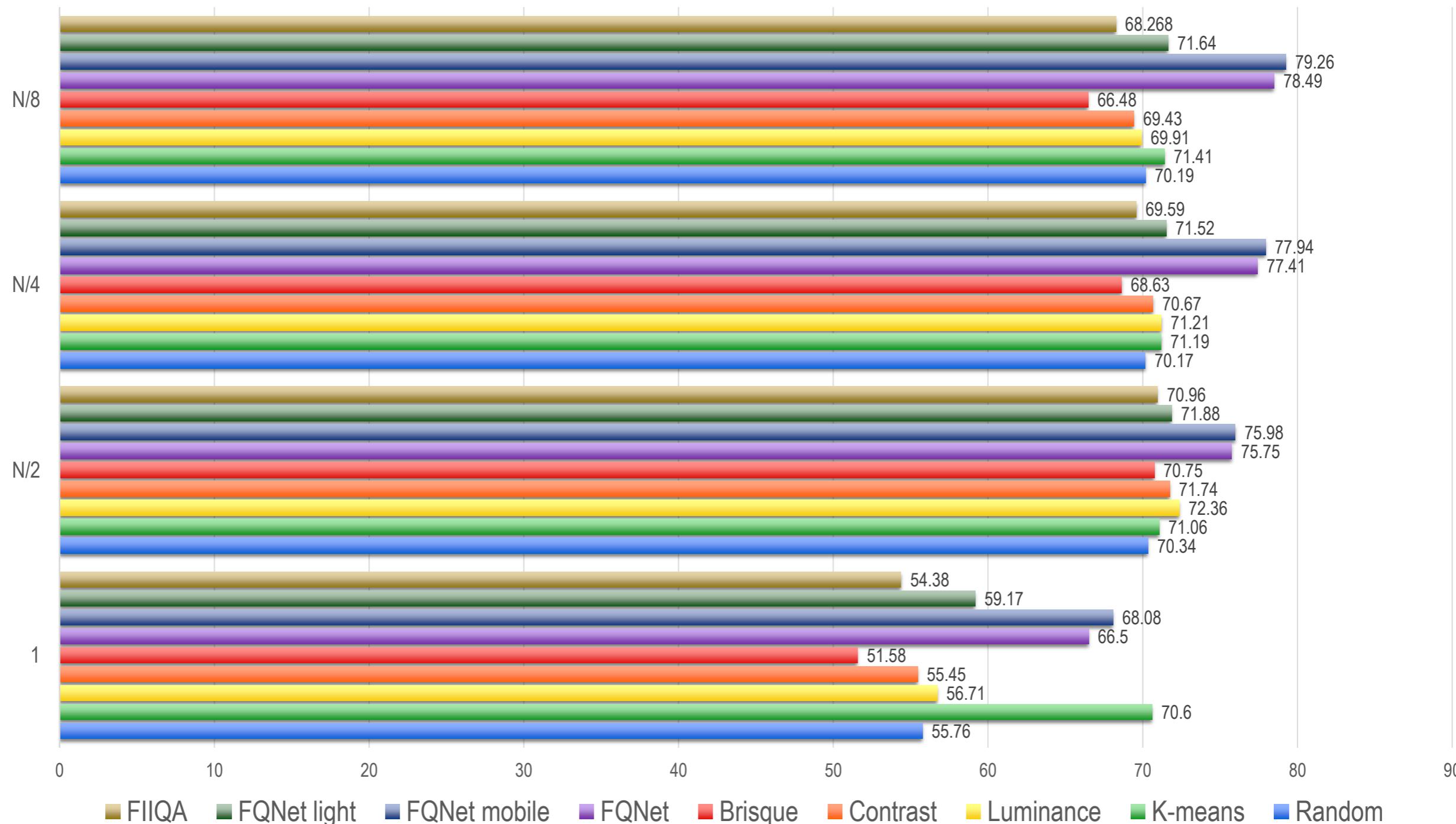


# Результаты экспериментов. IJB-C

ResNet-50

Baseline\* = 70.89%

Точность, %



\* Baseline – традиционный подход распознавания по каждому кадру

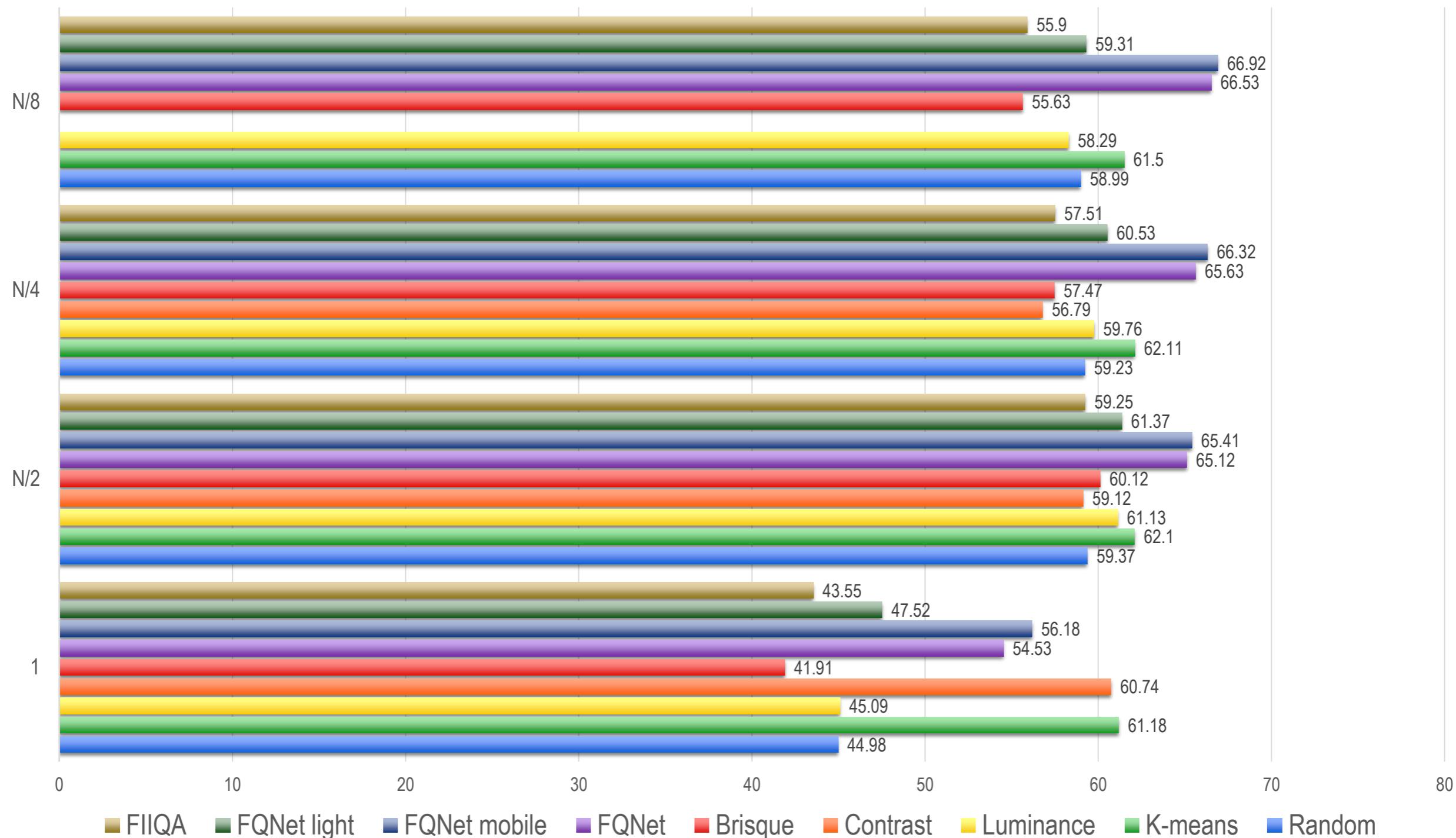


# Результаты экспериментов. IJB-C

MobileNet-vgg2

Baseline = 61.29%

Точность, %

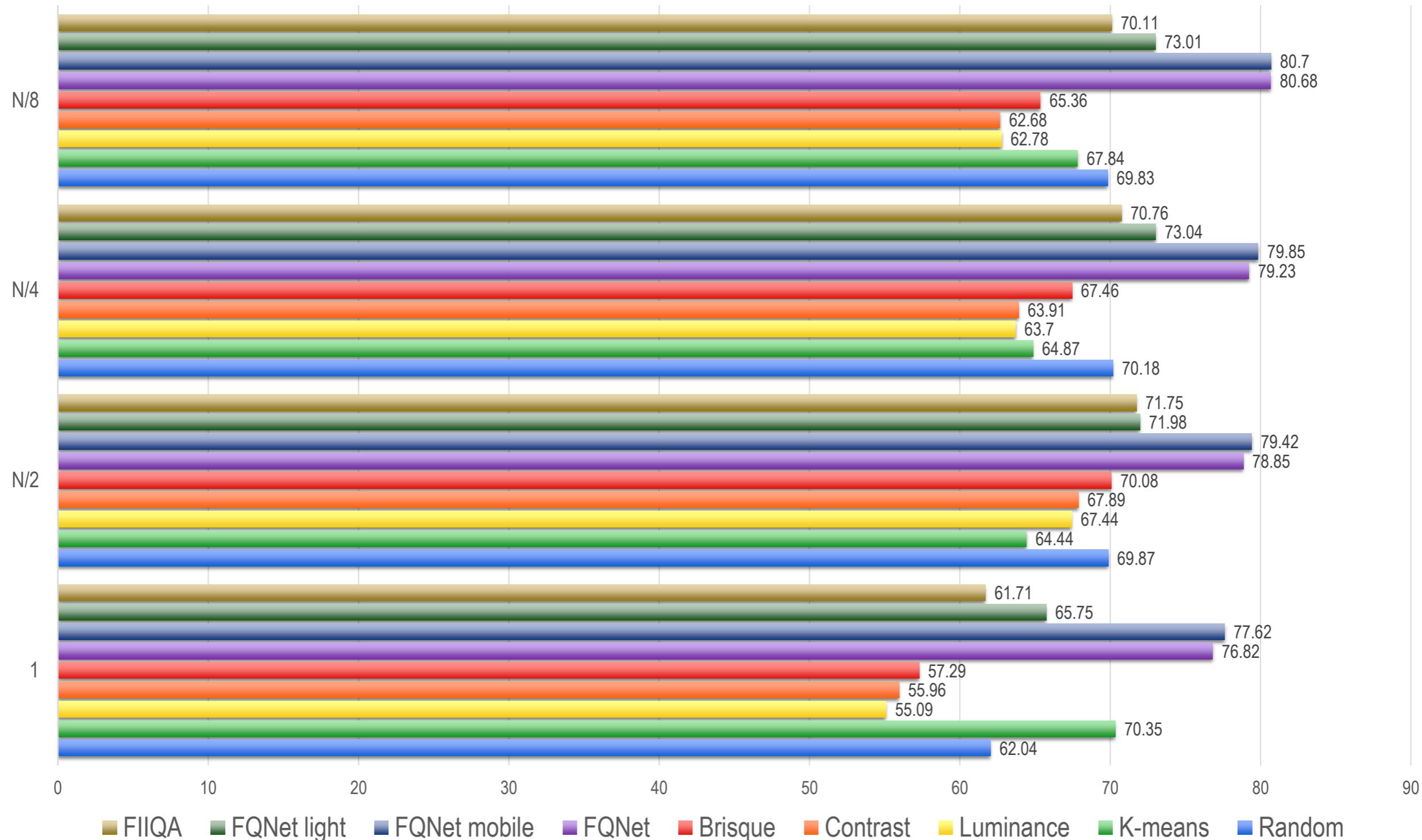




# Результаты экспериментов. IJB-C

**InsightFace      Baseline = 69.83%**

Точность, %

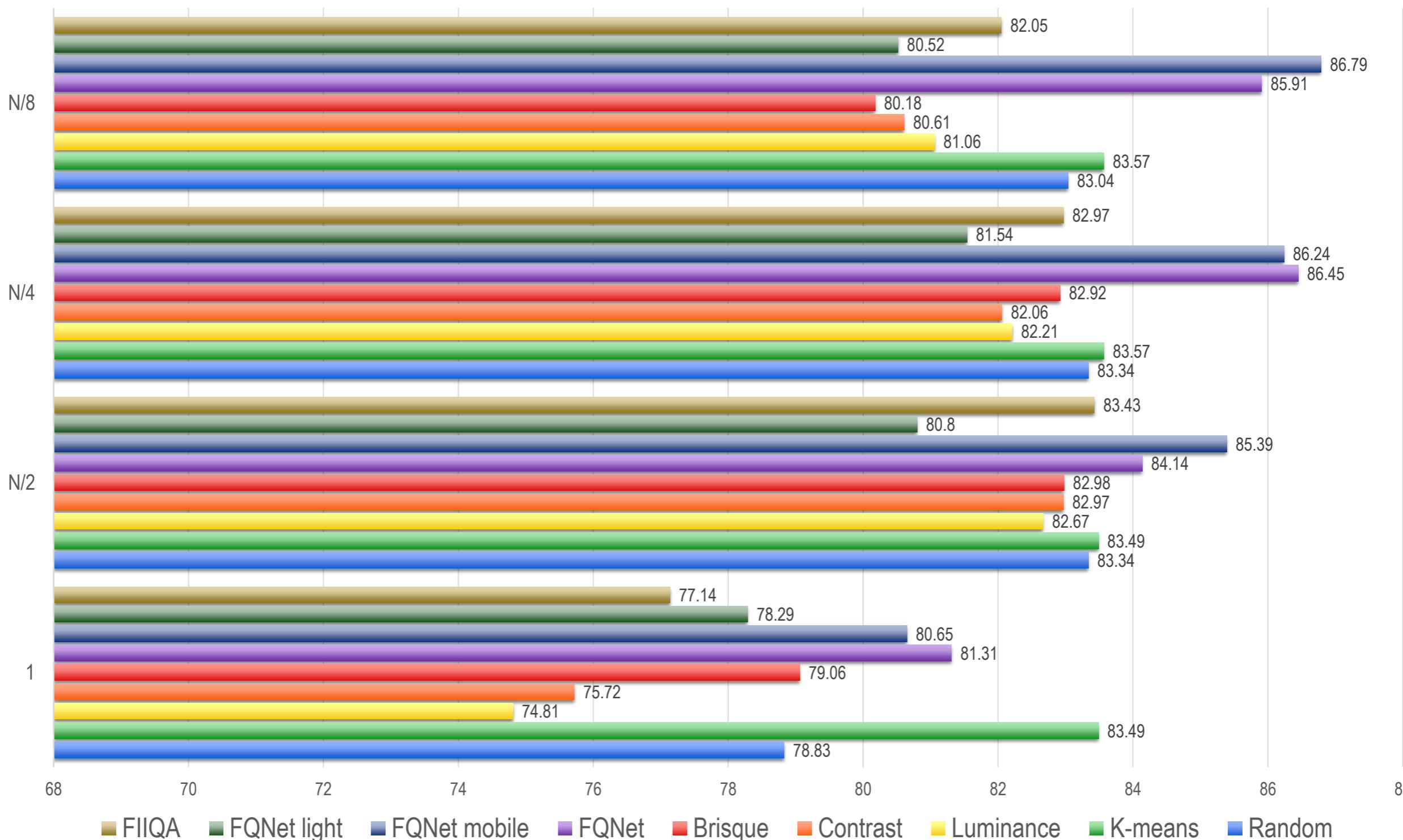




# Результаты экспериментов. YouTube Faces

ResNet-50      Baseline = 83.5%

Точность, %



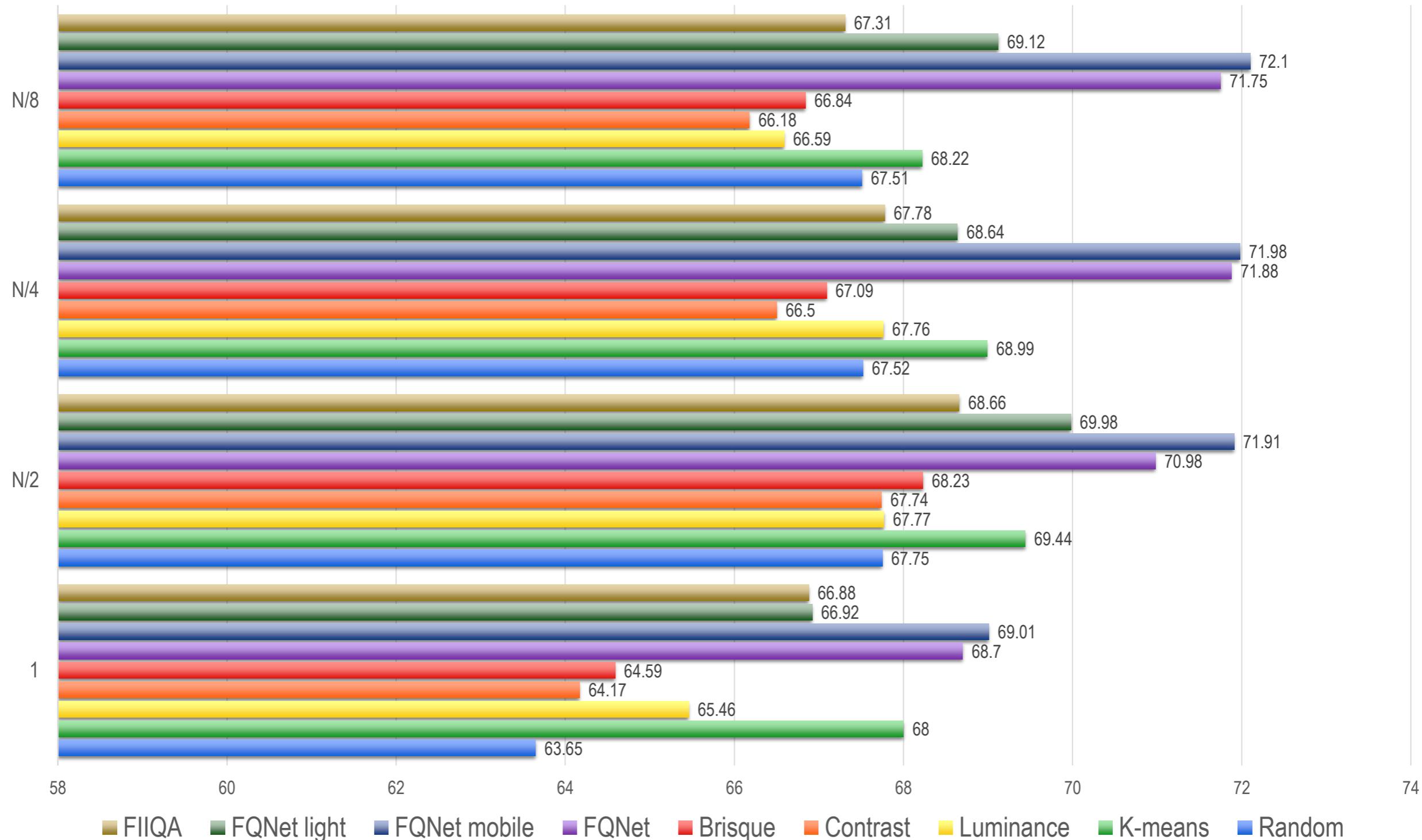


# Результаты экспериментов. YouTube Faces

MobileNet-vgg2

Baseline = 68.33%

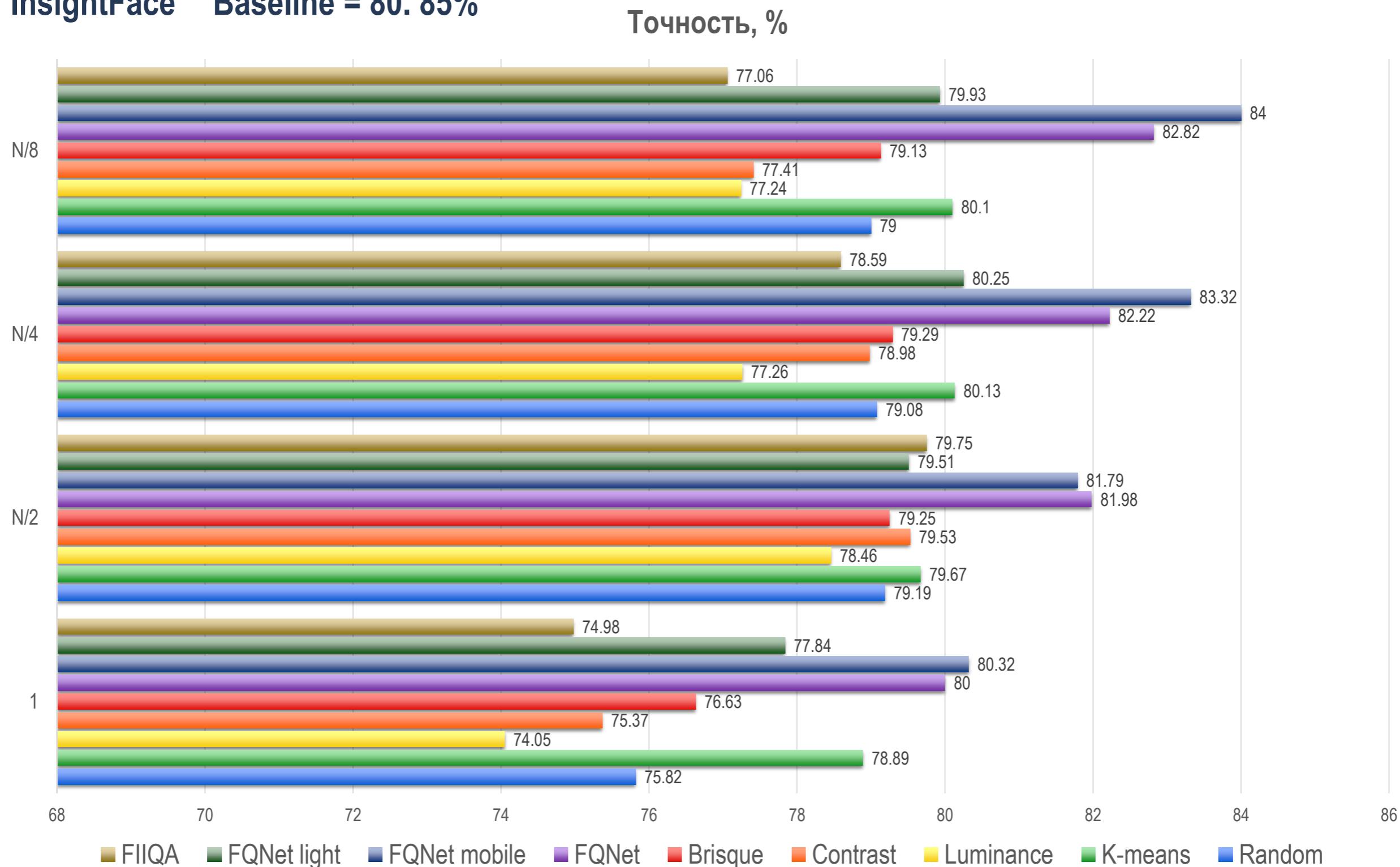
Точность, %





# Результаты экспериментов. YouTube Faces

**InsightFace    Baseline = 80. 85%**





# Результаты

---

- Проведен обзор алгоритмов распознавания лиц по видеопоследовательности. Показано преимущество реализации архитектур сверточных нейронных сетей
- Рассмотрены существующие подходы извлечения ключевых кадров
- Предложена общая схема алгоритма идентификации лиц по видео, включающая дополнительный шаг поиска качественных кадров для повышения стабильности работы системы
- Обучен ряд легких СНС архитектур методом дистилляции знаний в целях повышения эффективности выполнения идентификации лиц
- Проведены экспериментальные исследования по анализу производительности и точности разработанного подхода с помощью наборов данных *IJB-C*, *YouTube Faces*. Показано, что наиболее эффективным методом оценки качества является применение СНС модели *FaceQnet mobile*, с помощью которой удалось достичь повышения точности распознавания на 10% моделей *ResNet-50*, *InsightFace*



# Результаты

---

Исследование было проведено в рамках научно-учебной группы «Анализ мультимедийных данных пользователей мобильных устройств»

- В будущем планируется рассмотреть еще больше методов оценки качества кадров, а также произвести сравнение подходов с технологией сверточных нейронных сетей на основе внимания. Дополнительно планируется учитывать полученные веса качества изображения в процессе агрегации векторов-признаков изображений в рамках видеопоследовательности. Стоит отметить возможность реализации предложенной системы в качестве решения Face ID для современных мобильных устройств



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

СПАСИБО ЗА ВНИМАНИЕ!