



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Вычислительно эффективные алгоритмы распознавания изображений на основе последовательного анализа нейросетевых признаков

Соколова Анастасия Дмитриевна

adsokolova@hse.ru

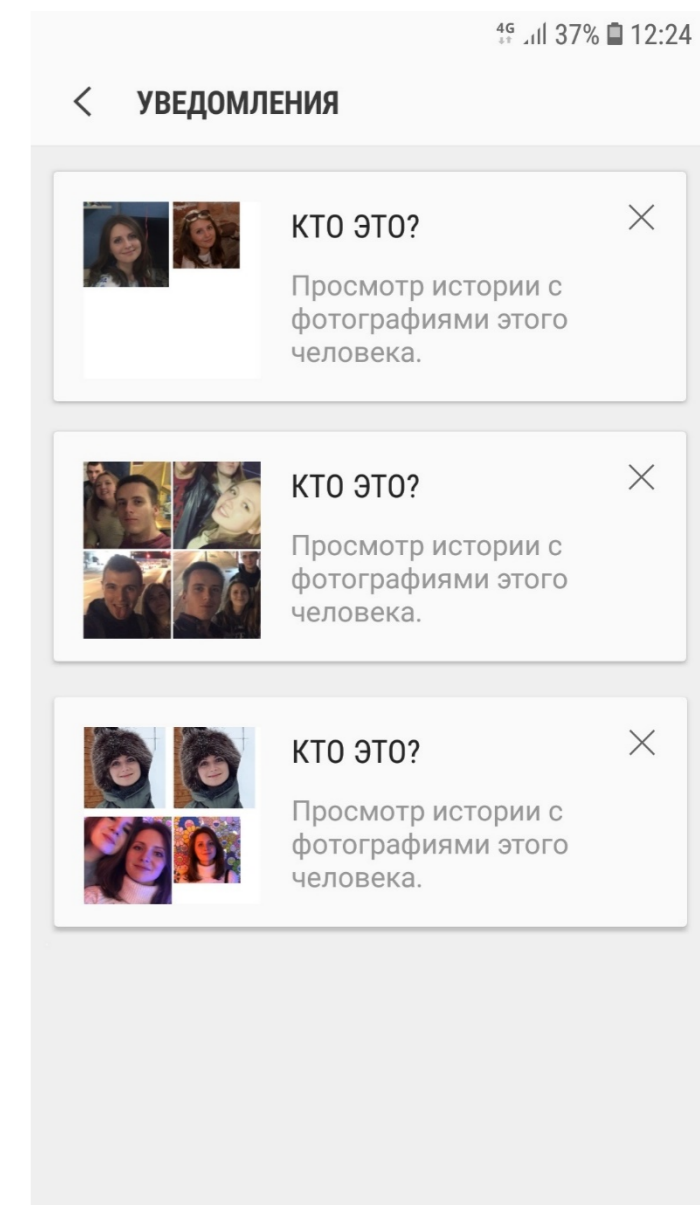
Савченко Андрей Владимирович

avsavchenko@hse.ru



АКТУАЛЬНОСТЬ

- Системы навигации по галереи изображений
- Группировка видеоданных для ведения статистики
- Распознавание объектов в сфере общественной безопасности
- И т.д.





ПОСТАНОВКА ЗАДАЧИ

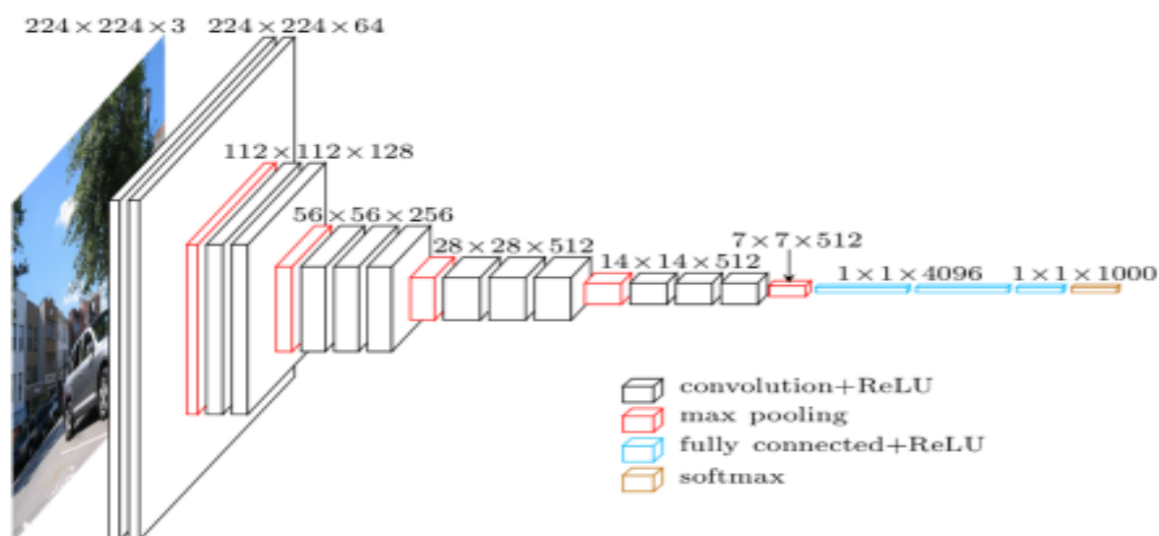
- Поступающему на вход изображению нужно поставить в соответствие один из $C > 1$ заранее точно неопределенных классов
- Классы задаются с помощью обучающего множества из $R \geq C$ эталонных изображений с известной меткой класса $c(r) \in \{1, \dots, C\}$
- Для каждого человека достаточно мало эталонных изображений - принятие решения может осуществляться с помощью метода ближайшего соседа (k -NN)
- Размерность вектора признаков D (256 - 4096). Сложность линейная $O(RD)$





ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ ИЗОБРАЖЕНИЯ

- Lightened CNN (version C)
1x256 вектор
- FaceNet
1x512 вектор
- VggFace2
1x2048 вектор
- VggFace
1x4096 вектор





АЛГОРИТМ КЛАССИФИКАЦИИ

- Анализ главных компонент векторов признаков эталонных изображений
- Входной вектор признаков x преобразуется в вектор главных компонент $\tilde{x} = [\tilde{x}_1, \dots, \tilde{x}_D]$
- На каждом l -м уровне иерархии ($l \in \{1, \dots, L\}$) сопоставляются только d_l главных компонент
- Количество компонент на каждом уровне иерархии d_l определяется, исходя из фиксированной доли объясненной дисперсии $\sigma_l^2 \in (0; 1]$ обучающего множества, при этом $\sigma_1^2 < \sigma_2^2 < \dots < \sigma_L^2 = 1$



АЛГОРИТМ КЛАССИФИКАЦИИ

- Расстояние между признаками с помощью уровней иерархии признаков

$$\rho(\tilde{x}^{(l)}, \tilde{x}^{(l)}) = \rho(\tilde{x}^{(l-1)}, \tilde{x}^{(l-1)}) + \sum_{d=d_{(l-1)}+1}^{d_l} \rho(\tilde{x}_d, \tilde{x}_{r;d})$$

- Определение принадлежности к классу

$$c_l^* = \operatorname{argmin}_{c \in C_l} \rho_c(\tilde{x}^{(l)})$$

- Множество кандидатов

$$C_{l+1} = \left\{ c \in C_l \mid \frac{\rho_c(\tilde{x}^{(l)})}{\rho_{c_l^*}(\tilde{x}^{(l)})} \leq \delta \right\}$$



РАСПОЗНАВАНИЕ ВИДЕО

- Нахождение среднего вектора

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x(t)$$

- Нахождение медоида

$$x^* = \operatorname{argmin}_{x(t)} \sum_{t'=1}^T \rho(x(t), x(t'))$$

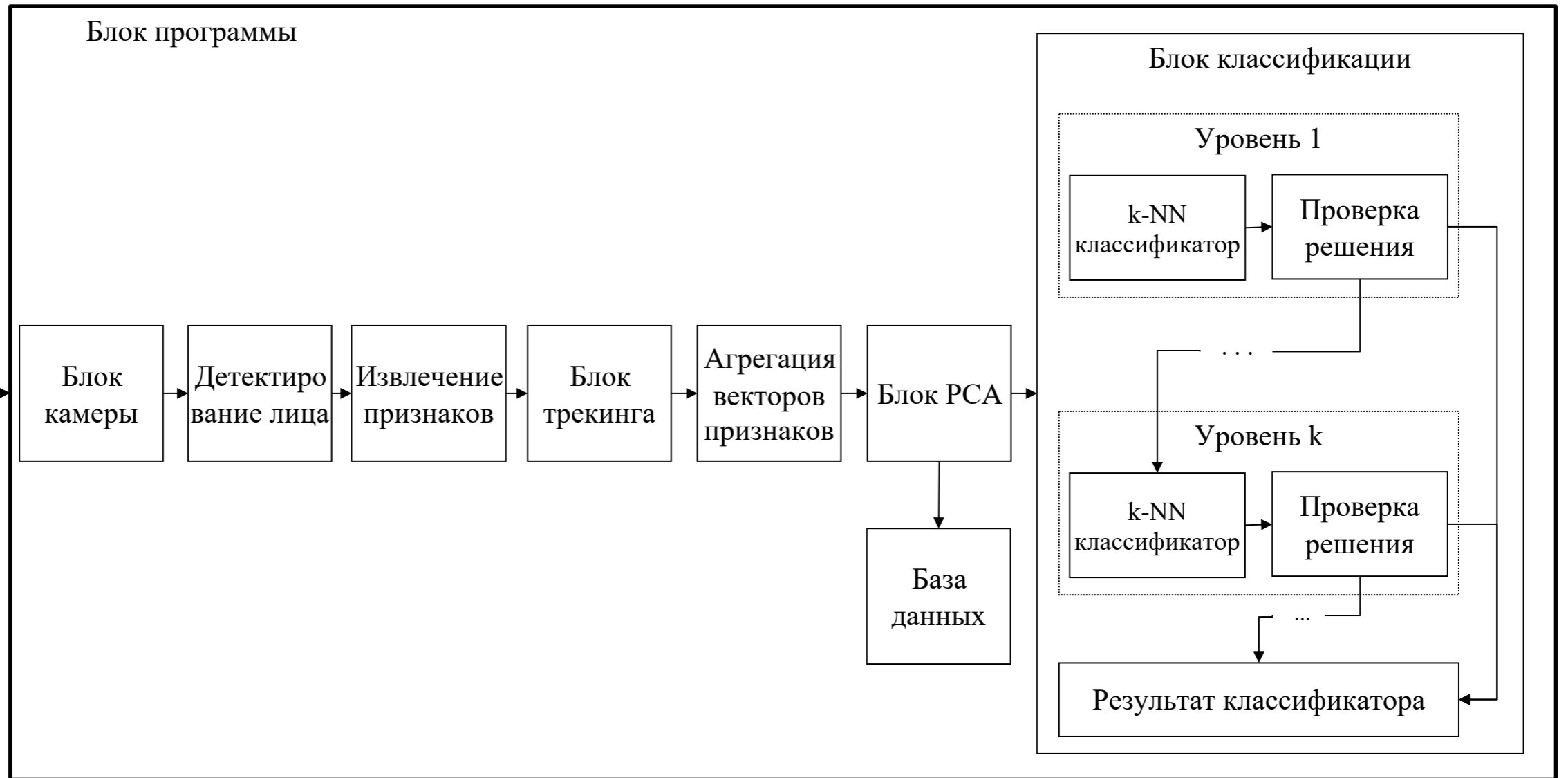
- Нахождение медианы

$$x' = [x'_1, \dots, x'_D]$$

где x'_i – медиана i -ой компоненты всех векторов $\{x(t), t = 1, 2, \dots, T\}$

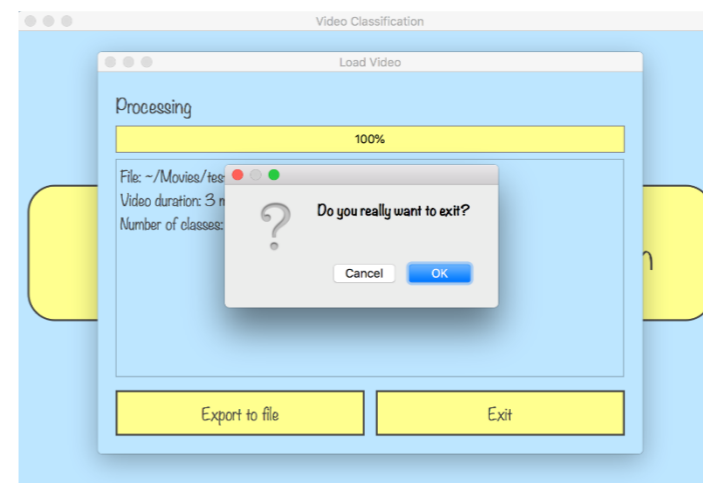
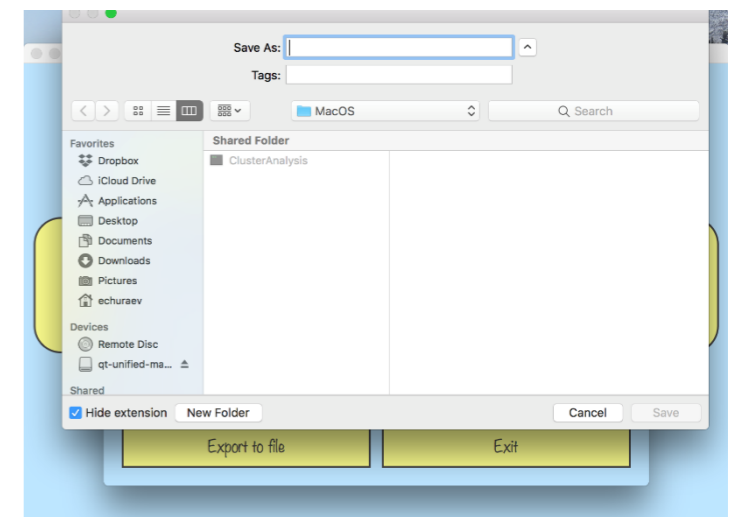
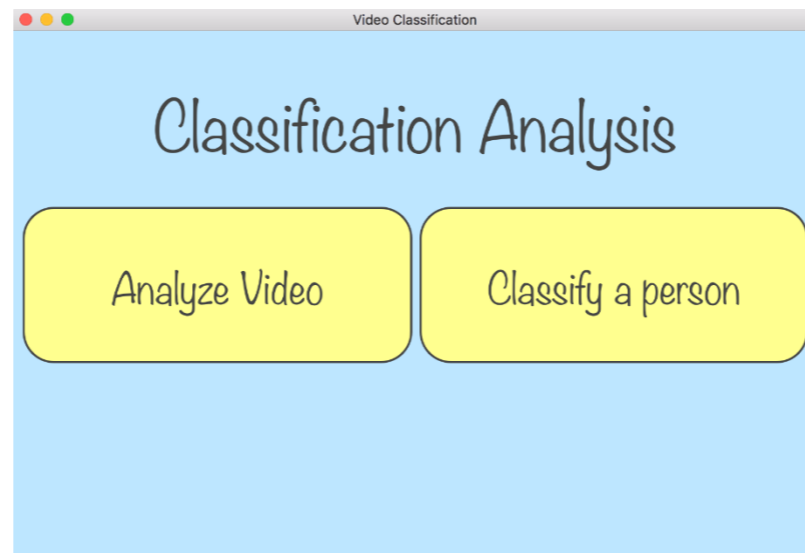
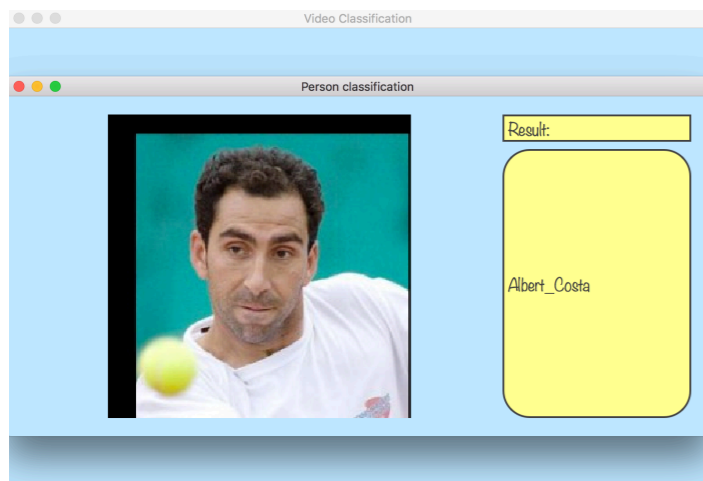
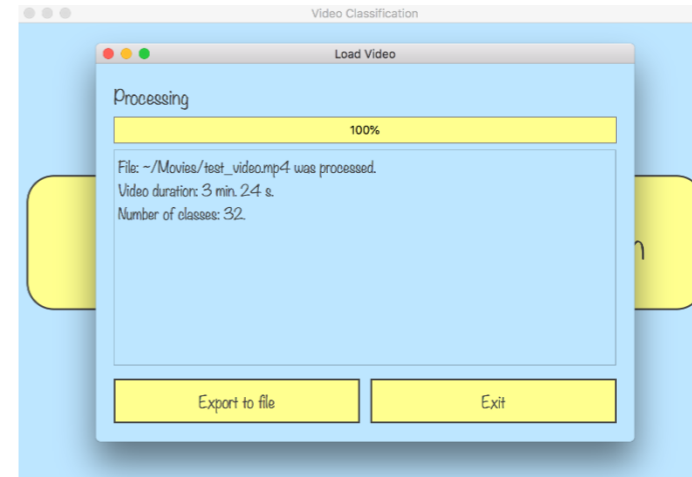
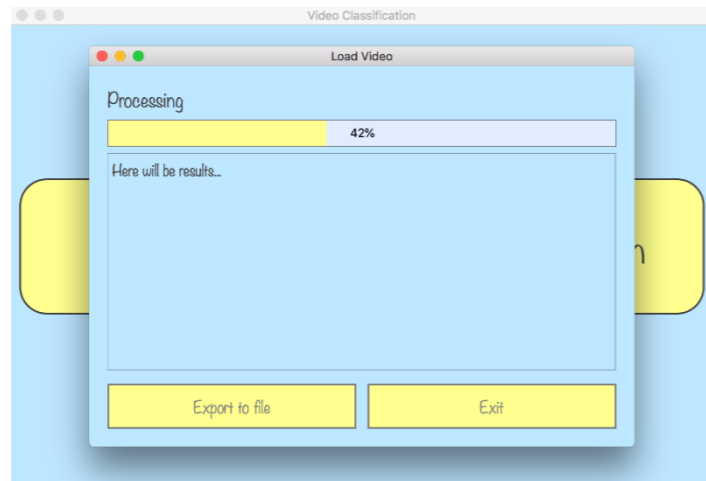


ПРЕДЛОЖЕННЫЙ ПОДХОД





ПРОТОТИП СИСТЕМЫ





НАБОРЫ ДАННЫХ

- LFW (Labeled Faces in the Wild)
 - 1680 людей
 - 13000 изображений
 - 1-10 изображений на человека
- YTF (YouTube Faces)
 - 1595 людей
 - 3425 видео
 - 48-6070 изображений на человека
- IJB-C (IARPA Janus Benchmark-C)
 - 31334 отдельных изображений
 - 3531 людей (~6 снимков)
 - 11779 видео
 - 117542 кадров



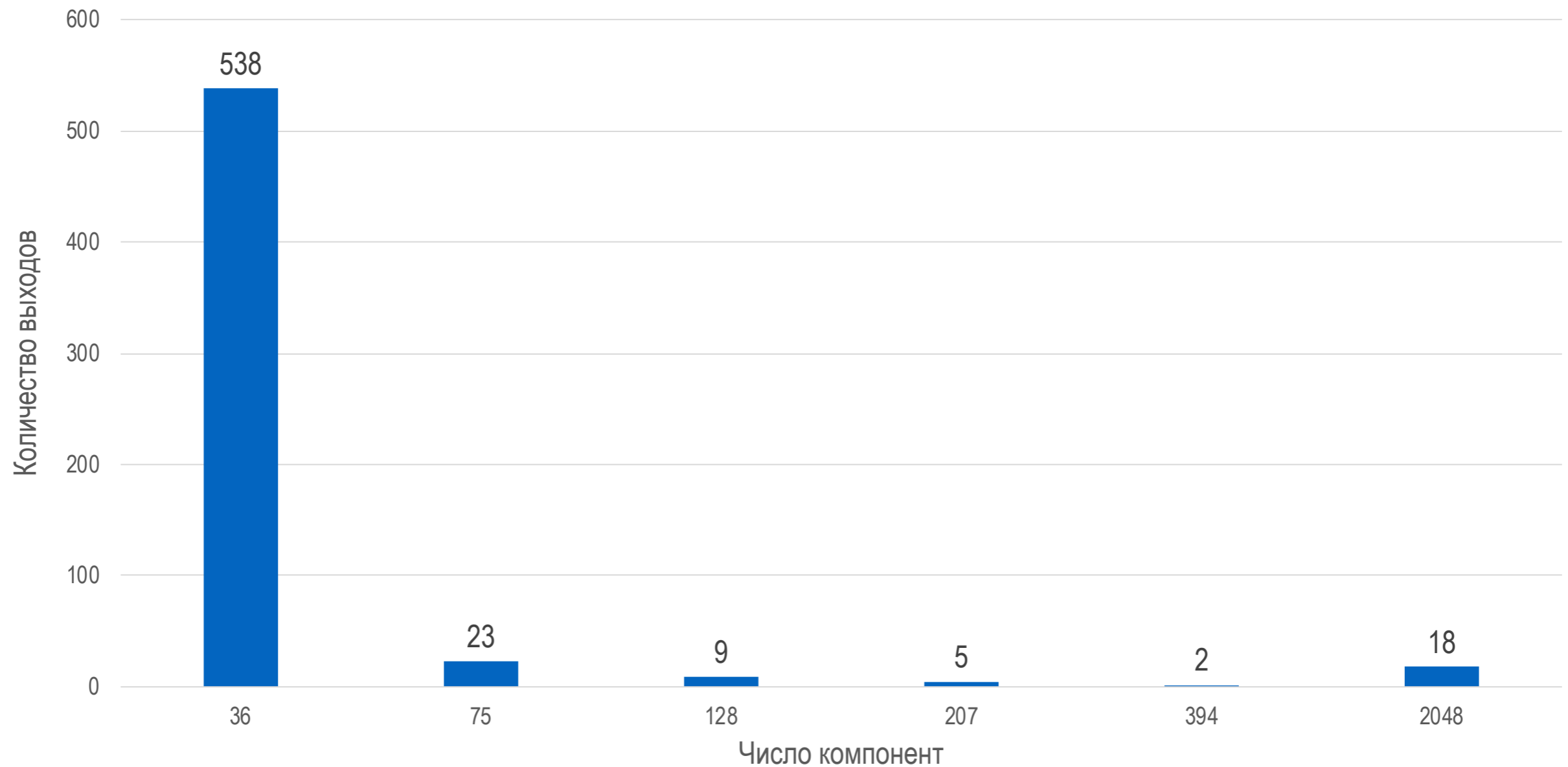


РЕЗУЛЬТАТЫ РАСПОЗНАВАНИЯ ЛИЦ K-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ LFW)

Метрика	Классификатор	VGGFace	LCNN	VGGFace2	FaceNet
Точность (%)	k-NN, все признаки	96.31	97.48	98.66	98.15
	k-NN, фиксированное число компонент	94.10	96.21	96.95	97.36
	Последовательный k-NN, фиксированное число компонент	95.97	96.97	98.32	98.15
	Последовательный k-NN, число компонент по объясненной дисперсии	95.80	96.81	97.98	97.94
Время (мс)	k-NN, все признаки	50.39	4.88	34.78	9.37
	k-NN, фиксированное число компонент	2.53	2.50	2.53	2.52
	Последовательный k-NN, фиксированное число компонент	2.92	2.54	2.84	2.69
	Последовательный k-NN, число компонент по объясненной дисперсии	3.20	2.23	2.81	1.87

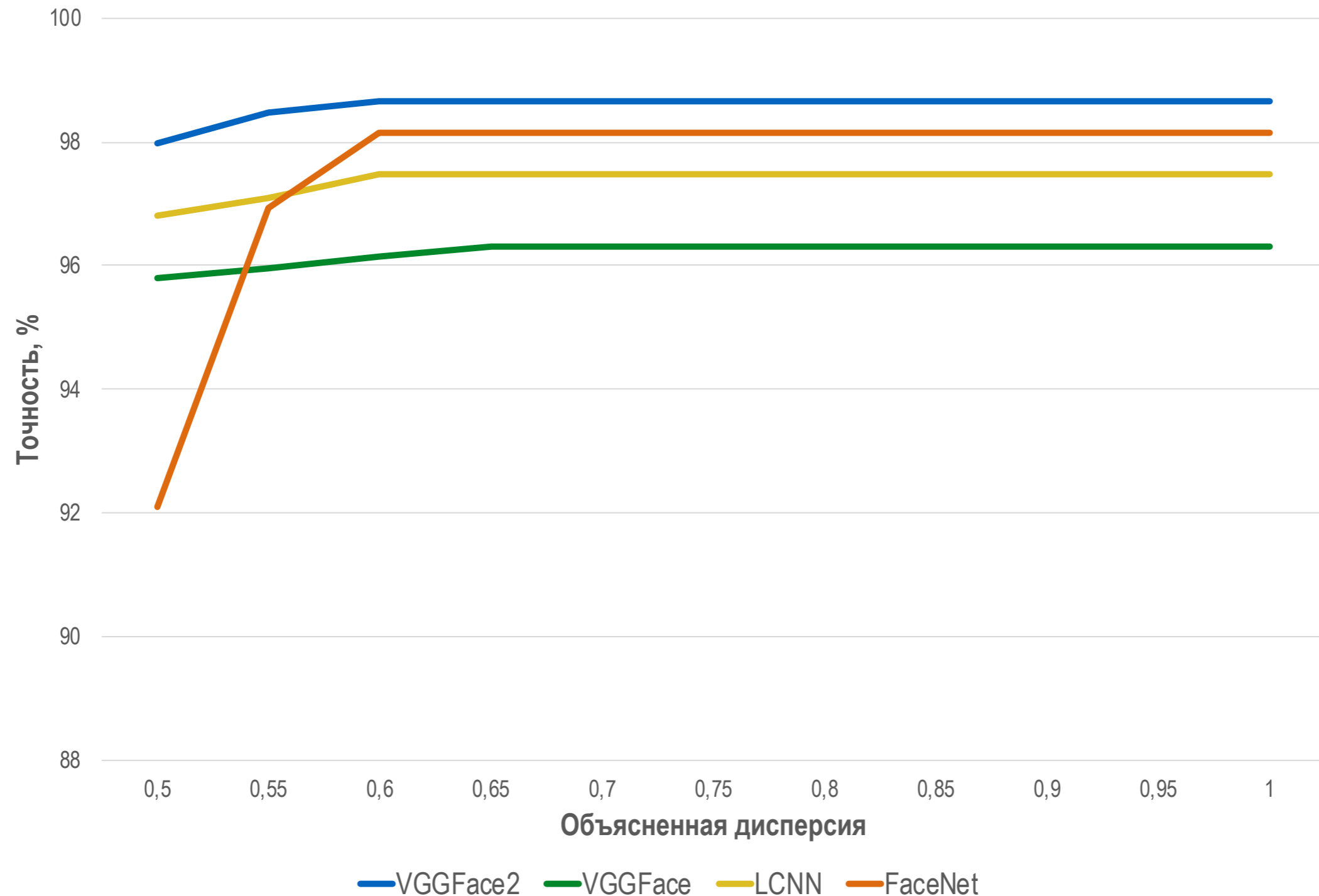


РЕЗУЛЬТАТ ВЫХОДОВ НА ОПРЕДЕЛЕННОМ ЧИСЛЕ КОМПОНЕНТ (НАБОР ДАННЫХ LFW)



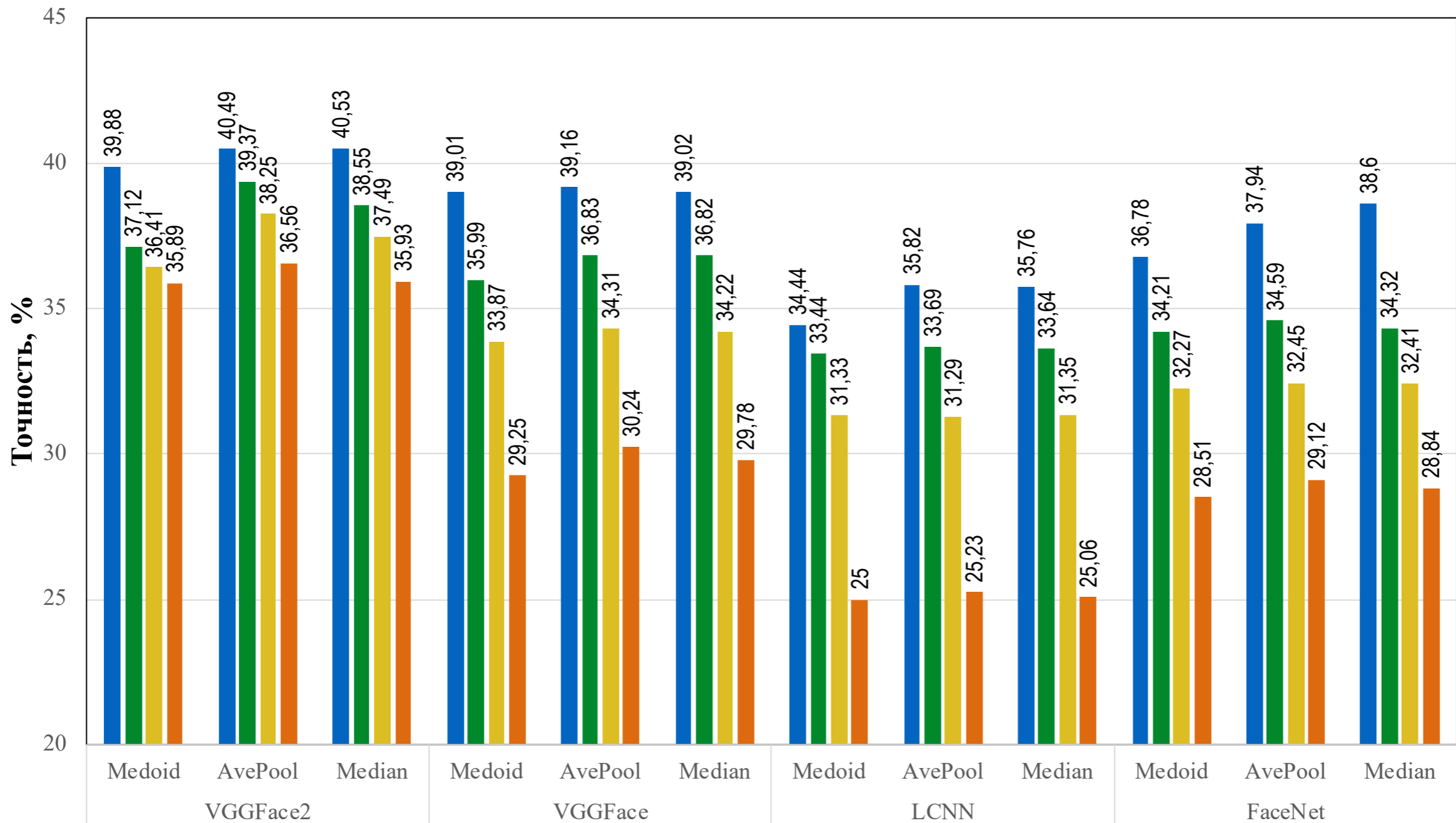


ЗАВИСИМОСТЬ ОТ ПРОЦЕНТА ОБЪЯСНЕННОЙ ДИСПЕРСИИ (НАБОР ДАННЫХ LFW)





ТОЧНОСТЬ K-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ YTF)



■ k-NN, все признаки

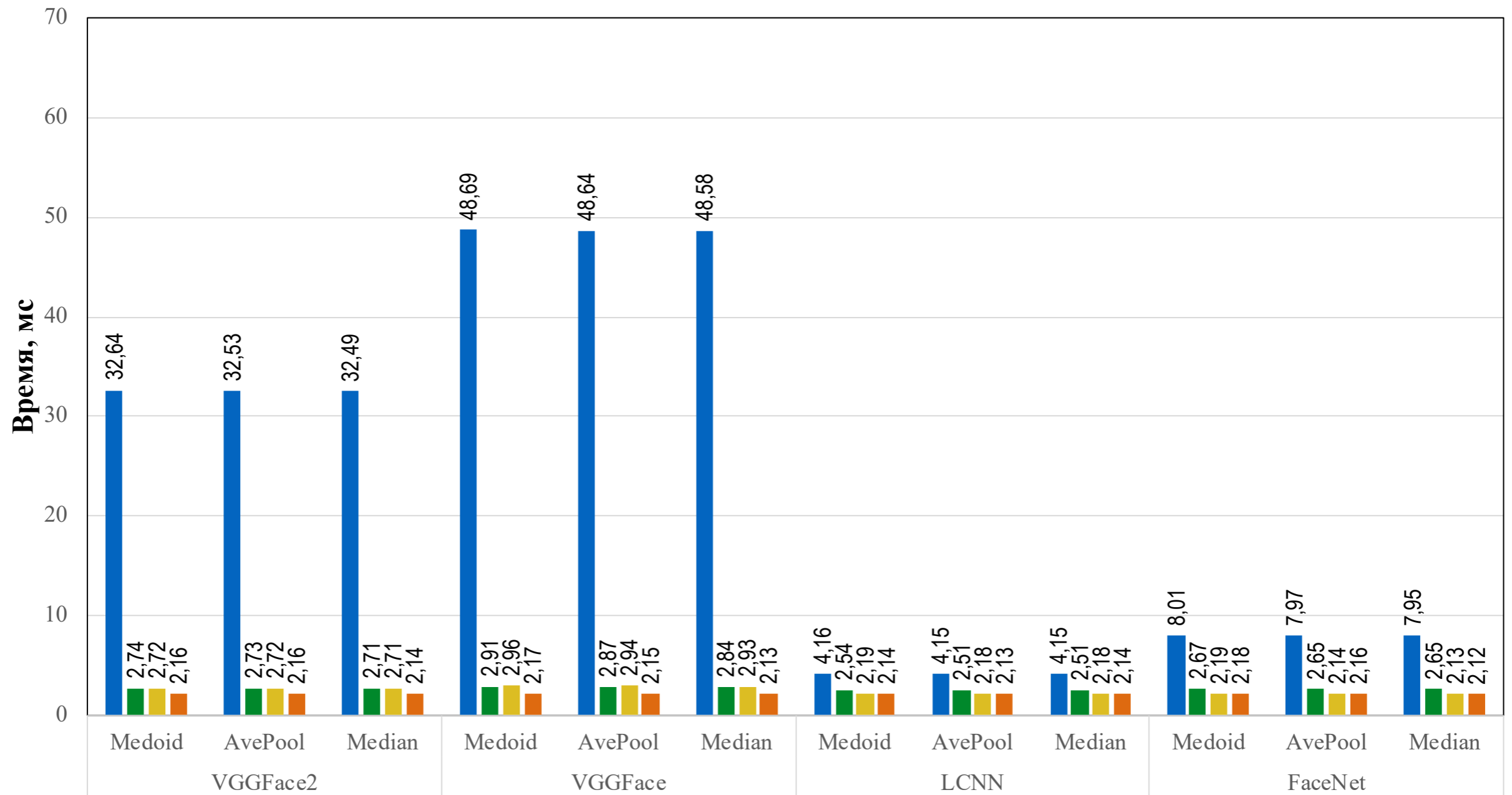
■ Последовательный k-NN, фиксированное число компонент

■ Последовательный k-NN, число компонент по объясненной дисперсии

■ k-NN, 64 компоненты



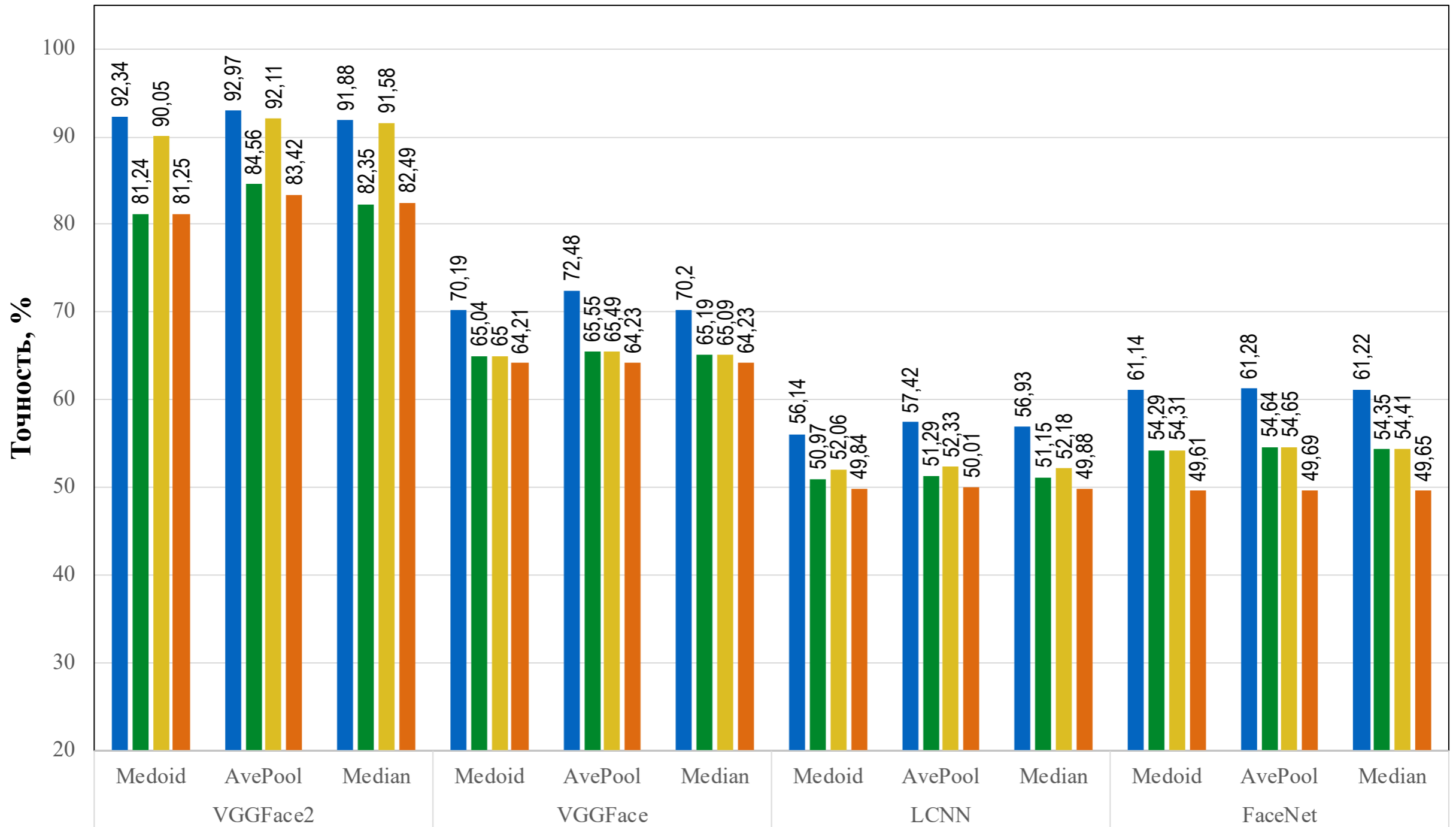
ВРЕМЯ РАБОТЫ К-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ YTF)



■ k-NN, все признаки ■ Последовательный k-NN, фиксированное число компонент
■ Последовательный k-NN, число компонент по объясненной дисперсии ■ k-NN, 64 компоненты



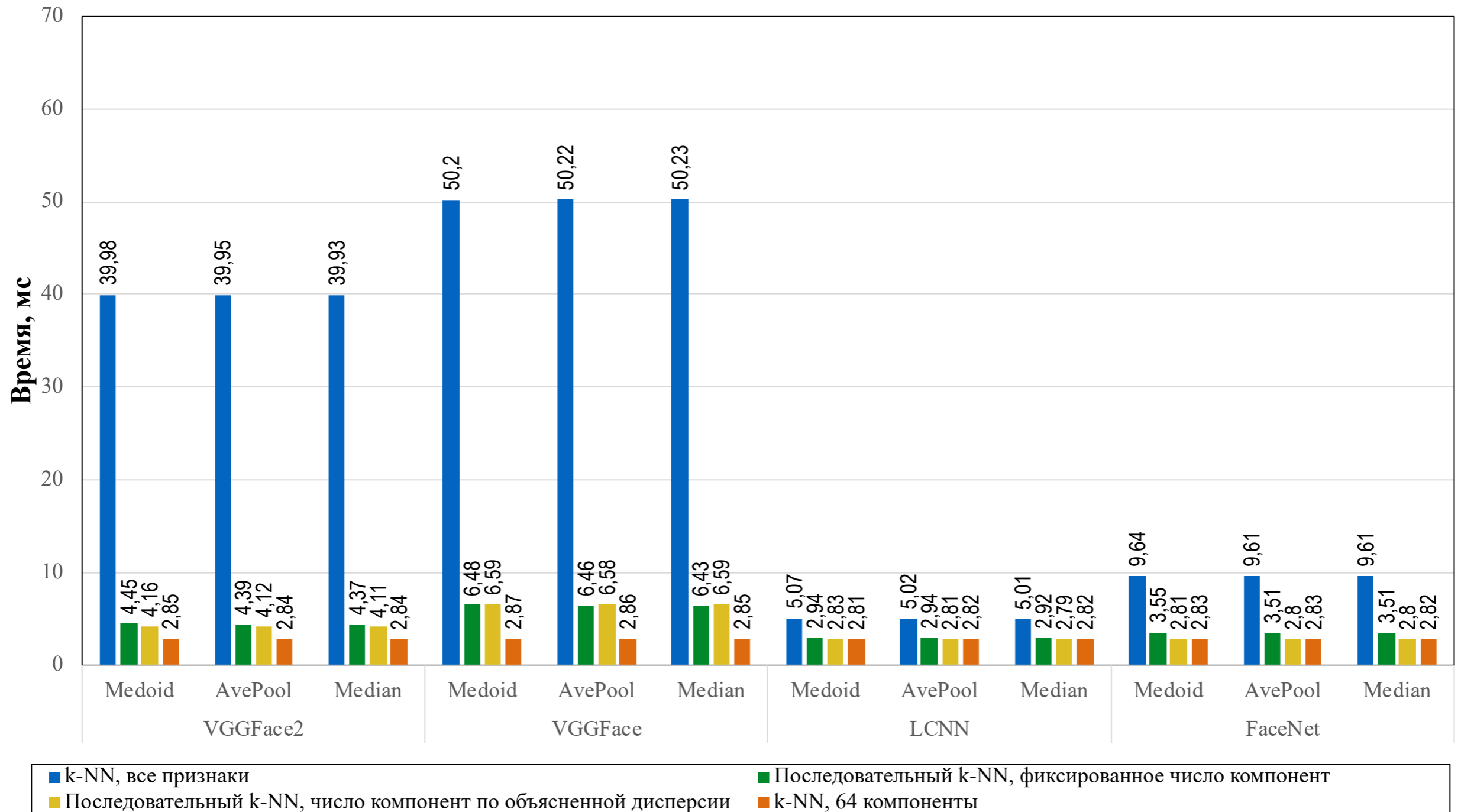
ТОЧНОСТЬ K-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ LFW-YTF)



■ k-NN, все признаки ■ Последовательный k-NN, фиксированное число компонент
■ Последовательный k-NN, число компонент по объясненной дисперсии ■ k-NN, 64 компоненты



ВРЕМЯ РАБОТЫ К-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ LFW-YTF)





ПОКАЗАТЕЛИ ТОЧНОСТИ K-NN КЛАССИФИКАТОРА (НАБОР ДАННЫХ IJB-C)

Метрика	Классификатор	VGGFace2	FaceNet
Точность (%)	k-NN, все признаки	87.24	53.08
	k-NN, фиксированное число компонент	82.53	49.73
	Последовательный k-NN, фиксированное число компонент	85.50	51.29
	Последовательный k-NN, число компонент по объясненной дисперсии	86.07	52.36
Время (мс)	k-NN, все признаки	42.17	16.63
	k-NN, фиксированное число компонент	3.12	3.11
	Последовательный k-NN, фиксированное число компонент	3.64	3.22
	Последовательный k-NN, число компонент по объясненной дисперсии	3.42	3.09



ЗАКЛЮЧЕНИЕ

- Была предложена модификация алгоритма k-NN классификации на основе последовательного анализа нейросетевых признаков
- Время принятия решения предложенного алгоритма в 10 раз меньше, чем время традиционного подхода поиска ближайшего соседа
- Экспериментальным исследованием было показано, что наилучшей точностью характеризуется использование сети VGGFace2 и алгоритма поиска среднего вектора однородного сегмента нормализованных признаков



СПАСИБО ЗА ВНИМАНИЕ!



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ