

УДК 004.93:656.25

UDC 004.93:656.25

5.2.2 – Математические и инструментальные методы экономики (экономические науки)

5.2.2 – Mathematical and instrumental methods of economics (economic sciences)

РАЗРАБОТКА ГИБРИДНОЙ СИСТЕМЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА НА ОСНОВЕ YOLOV11 И FACENET

DEVELOPMENT OF A HYBRID FACE-BASED IDENTIFICATION SYSTEM BASED ON YOLOV11 AND FACENET

Павлов Дмитрий Алексеевич

Pavlov Dmitriy Alexeevich

к.ф.-м.н., доцент

Cand.Phys.-Math.Sci., associate professor

РИНЦ SPIN-код, 8822-5089

RSCI SPIN-code, 8822-5089

dp.logic@gmail.com

dp.logic@gmail.com

Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трублина», г. Краснодар, Россия

Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia

Хван Владимир

Hvan Vladimir

Студент

Student

Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трублина», г. Краснодар, Россия

Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia

В работе рассматривается задача идентификации личности по изображению лица в системах реального времени. Предложена гибридная архитектура, объединяющая детектор YOLOv11 с предсказанием пяти ключевых точек и идентификатор FaceNet на основе Inception ResNet v1 с функцией потерь Triplet Loss. Разработан конвейер предобработки, включающий аффинное выравнивание лица по глазам перед подачей в энкодер. Представлены результаты экспериментального исследования на наборах данных LFW, AgeDB и WIDER Face. Достигнута точность идентификации 98,7% на LFW при частоте обработки 12 кадров в секунду на GPU NVIDIA RTX A6000 (48Гб). Проведена оценка экономической эффективности внедрения разработанной системы

This article examines the problem of face image identification in real-time systems. A hybrid architecture is proposed that combines the YOLOv11 detector with five keypoint prediction and a FaceNet identifier based on Inception ResNet v1 with a Triplet Loss loss function. A preprocessing pipeline is developed that includes affine face alignment to eyes before feeding the face to the encoder. Experimental results are presented on the LFW, AgeDB, and WIDER Face datasets. An identification accuracy of 98.7% was achieved on the LFW dataset at a processing frequency of 12 frames per second on a NVIDIA RTX A6000 GPU (48GB). The cost-effectiveness of implementing the developed system is assessed

Ключевые слова: ГЛУБОКОЕ ОБУЧЕНИЕ, КОМПЬЮТЕРНОЕ ЗРЕНИЕ, СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ, РАСПОЗНАВАНИЕ ЛИЦ, ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ, YOLO, FACENET

Keywords: DEEP LEARNING, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORKS, FACE RECOGNITION, PERSONAL IDENTIFICATION, YOLO, FACENET

<http://dx.doi.org/10.21515/1990-4665-221-016>

Введение

В последние годы цифровые сервисы и информационные системы проникают практически во все сферы жизни. Вместе с этим растут

<http://ej.kubagro.ru/2026/07/pdf/16.pdf>

требования к безопасности и надёжности аутентификации пользователей. Старые способы проверки личности – по паролю, пластиковой карте или PIN-коду — уже не дают нужного уровня защиты. Их можно подделать, украсть или передать другому человеку. Кроме того, такие методы требуют активных действий от пользователя: ввести код, приложить карту, запомнить сложную комбинацию.

Биометрическая аутентификация, в частности распознавание по лицу, свободна от многих из этих недостатков. Сканеру не нужно касаться человека – достаточно просто попасть в объектив камеры. Методы на основе глубокого обучения и свёрточных нейросетей, включая архитектуры типа сиамская нейронная сеть [1], позволяют эффективно решать задачу верификации личности [2].

Современные системы распознавания лиц показывают точность, сравнимую с человеческим зрением [3, 4]. Особый интерес представляют архитектуры, способные работать в реальном времени. Среди множества подходов выделяется семейство YOLO (You Only Look Once) [5], обеспечивающее наилучший компромисс между скоростью и точностью детекции, а также метрический подход FaceNet, который использует специальную функцию потерь Triplet Loss и превращает лицо в компактный числовой вектор (эмбеддинг) [3].

Версия YOLOv11 поддерживает детекцию ключевых точек лица, что позволяет совместить поиск лица и его выравнивание в одной модели [6]. Это открывает возможность построить целостную, практичную систему, где не нужно собирать отдельные модули из разных библиотек [7]. Для обучения идентификатора используются большие размеченные наборы данных, такие как VGGFace2 [8].

Объект исследования: Процесс идентификации человека по фотографии или видеокадру внутри информационных систем.

Предмет исследования: Алгоритмы и методы глубокого обучения, которые отвечают за поиск лица на изображении, его выравнивание и последующее распознавание, а также способы объединения этих этапов в единую работающую программу.

Цель работы: Изучить существующие методы глубокого обучения для задачи распознавания лиц и на их основе разработать программную систему, которая использует гибридную архитектуру: детектор YOLOv11 (для поиска лица и его ключечных точек) и идентификатор FaceNet на базе Inception ResNet v1 с функцией потерь Triplet Loss. Система должна показывать высокую точность и работать в реальном времени.

Экспериментальное исследование проводилось на общедоступных наборах: LFW [9] (Labeled Faces in the Wild), AgeDB [10] для оценки устойчивости к возрастным изменениям и WIDER Face [11] для оценки качества детекции. Сравнительный анализ включал такие детекторы, как MTCNN [12] и RetinaFace [13]. Для эффективного поиска по базе эталонных эмбеддингов использовалась библиотека FAISS [14].

Материалы и методы исследования

Разработанная система имеет клиент-серверную топологию и состоит из трёх основных компонентов: модуля ввода (веб-камера, файл, видеопоток), модуля обработки (конвейер глубокого обучения) и модуля вывода (API, веб-интерфейс). На рисунке 1 представлена компонентная диаграмма системы.



Рисунок 1 - Компонентная диаграмма системы идентификации личности

Конвейер обработки включает следующие этапы: детекция лица и предсказание пяти ключевых точек (YOLOv11) → аффинное выравнивание по глазам → нормализация → извлечение эмбединга или представления, который отображает сложные многомерные данные в векторное пространство числовых векторов меньшей размерности 128 (FaceNet) → поиск ближайшего соседа (FAISS/ pgvector).

Для детекции лиц выбран YOLOv11 [6] в конфигурации medium по следующим причинам: оптимальный баланс скорости и точности, поддержка задачи определения ключевых точек (keypoint detection) из границ объекта, наличие экосистемы Ultralytics [7] и возможность выравнивания лица без дополнительных моделей. На рисунке 2 представлена диаграмма активности модуля детекции.

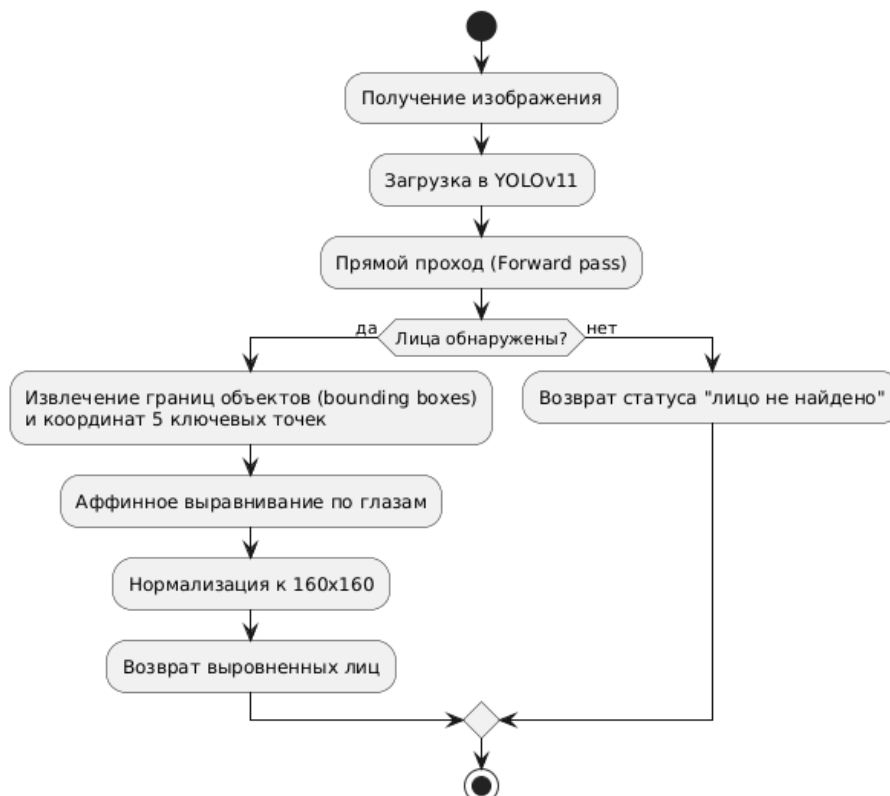


Рисунок 2 - Диаграмма потоков данных системы компьютерного зрения

Аффинное выравнивание выполняется по левому и правому глазу. Целевые координаты глаз в выходном изображении (160×160) выбраны как (0.35×160, 0.35×160) для левого и (0.65×160, 0.35×160) для правого, что обеспечивает симметричное размещение лица.

Для извлечения эмбеддингов используется Inception ResNet v1 с функцией потерь Triplet Loss (1).

$$L = \sum_{i=1}^N \max \left(\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha, 0 \right) \quad (1)$$

где N - количество триплетов (набор из трёх изображений лиц, используемый для обучения нейронной сети метрическому пространству, состоящем из компонент: x_i^a - эталонное изображение человека; x_i^p - положительное изображение человека; x_i^n - отрицательное изображение человека) в обучающем батче. $f(x_i^a)$, $f(x_i^p)$, $f(x_i^n)$ – соответственно эмбединги изображения человека эталонного, положительного и отрицательного.

$\|f(x_i^a) - f(x_i^p)\|_2^2$ – квадрат евклидова расстояние между эмбедингами эталонного и положительного изображения (должно быть малым). $\|f(x_i^a) - f(x_i^n)\|_2^2$ - квадрат евклидова расстояние между эмбедингами эталонного и отрицательного изображения (должно быть большим). α - Пороговое значение, определяющее минимальное требуемое различие между расстояниями (в работе $\alpha=0,2$).

Выбор Inception ResNet v1 обусловлен сочетанием Residual и Inception модулей, оптимальным размером (~25 млн параметров) и доступностью предобученных весов на VGGFace2 [8]. Размер эмбединга принят равным 128.

Для поиска ближайшего соседа используется библиотека FAISS [14]. При размере базы <1000 человек применяется плоский индекс IndexFlatL2, при большем размере – IndexIVFFlat. Порог принятия решения калибруется по точке равенства ошибок (EER) на валидационной выборке. На рисунке 3 представлена диаграмма последовательности идентификации.

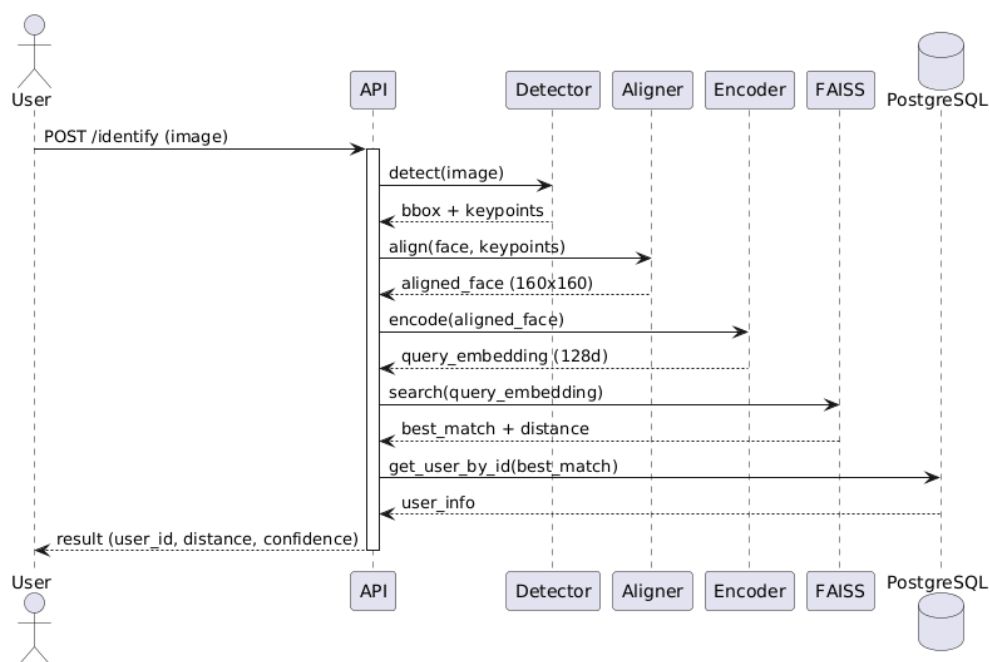


Рисунок 3 – Диаграмма последовательности процесса идентификации

Сводные таблицы используемых представлены в таблицах 1-8.

Таблица 1 - Параметры запроса (Request)

Параметр	Тип	Обязательность	По умолчанию	Описание
image	file (JPEG/PNG)	Да	—	Изображение для идентификации
threshold	float (0.5–2.0)	Нет	1.05	Порог принятия решения
top_k	int (1–10)	Нет	1	Количество возвращаемых кандидатов
return_embeddings	bool	Нет	false	Вернуть эмбединг в ответе

Таблица 2 - Параметры детектора YOLOv11

Параметр	Значение	Описание
conf_threshold	0.5	Минимальная уверенность детекции
iou_threshold	0.45	Порог для Non-Maximum Suppression
max_detections	10	Максимальное количество лиц в кадре
imgsz	640	Размер входного изображения (пиксели)
classes	[0]	Только класс "face"
device	"cuda" / "cpu"	Вычислительное устройство

Таблица 3 - Выходные данные детектора (DetectedFace)

Поле	Тип	Пример	Описание
bbox	Tuple[int,int,int,int]	(120, 85, 210, 190)	Координаты рамки (x1,y1,x2,y2)
confidence	float	0.87	Уверенность детекции (0–1)
class_id	int	0	Идентификатор класса (лицо)
keypoints	Dict[str, Tuple]	{"left_eye": (145, 112)}	5 ключевых точек лица

Таблица 4 - Параметры выравнивания (FaceAligner)

Параметр	Значение	Описание
target_size	160×160	Выходной размер выровненного лица
target_left_eye	(56, 56)	Целевая позиция левого глаза
target_right_eye	(104, 56)	Целевая позиция правого глаза
interpolation	INTER_LINEAR	Метод интерполяции OpenCV
border_mode	BORDER_CONSTANT	Режим границ (чёрный фон)

Таблица 5 - Выходные данные энкодера (FaceNet)

Поле	Тип	Пример	Описание
embedding	List[float]	[0.023, -0.145, ...]	128-мерный эмбединг
shape	Tuple[int]	(128,)	Размерность эмбединга
norm	float	1.0	L2-норма (после нормализации)
dtype	str	"float32"	Тип данных

Таблица 6 - Результаты поиска FAISS

Поле	Тип	Пример	Описание
distances	List[float]	[0.782, 1.234, ...]	L2-расстояния до эталонов
indices	List[int]	[42, 157, 89, ...]	Идентификаторы эмбедингов в базе
k	int	5	Количество возвращённых соседей

Таблица 7 - Информация о пользователе (из БД)

Поле	Тип	Пример	Описание
user_id	string	"usr_42"	Уникальный идентификатор
full_name	string	"Иванов Иван Иванович"	Полное имя
role	string	"employee"	Роль пользователя
department	string	"IT"	Отдел/подразделение
registered_at	datetime	"2026-01-15T10:00:00Z"	Дата регистрации
is_active	bool	true	Активен ли пользователь

Таблица 8 - Ответ API (Response)

Поле	Тип	Пример	Описание
status	string	"recognized" / "unknown"	Результат идентификации
user_id	string	"usr_42"	ID пользователя (если узнан)
full_name	string	"Иванов Иван"	Имя (если узнан)
distance	float	0.782	Расстояние до ближайшего эталона
threshold	float	1.05	Использованный порог
confidence	float	0.94	Уверенность (преобразованное расстояние)
processing_time_ms	int	87	Время обработки в миллисекундах
timestamp	string (ISO 8601)	"2026-05-29T10:30:00Z"	Время запроса

Для расчета параметра уверенности (confidence) используется формула:

$$confidence = \max(0, \min(1, 1 - \frac{dist}{threshold}))$$

где *dist* - евклидово расстояние до ближайшего эталона. *threshold* - порог принятия решения (по умолчанию 1.05).

Экспериментальное исследование проводилось на наборах данных LFW [9], AgeDB [10] для оценки идентификации и WIDER Face [11] – для детекции. Оценка производительности выполнялась на GPU NVIDIA A6000 (48 ГБ).

Результаты и обсуждение

Качество детекции лиц представлено в таблице 9.

Таблица 9 - Результаты детекции на WIDER Face

Метрика	Значение
mAP50	96,2%
mAP50-95	87,5%
Recall (IoU=0.5)	94,8%
Precision (IoU=0.5)	95,3%

Полученные результаты сопоставимы со специализированными детекторами (RetinaFace [17]) и превосходят классический MTCNN [12] по скорости при близкой точности.

На рисунке 4 представлена ROC-кривая идентификации на наборе LFW.

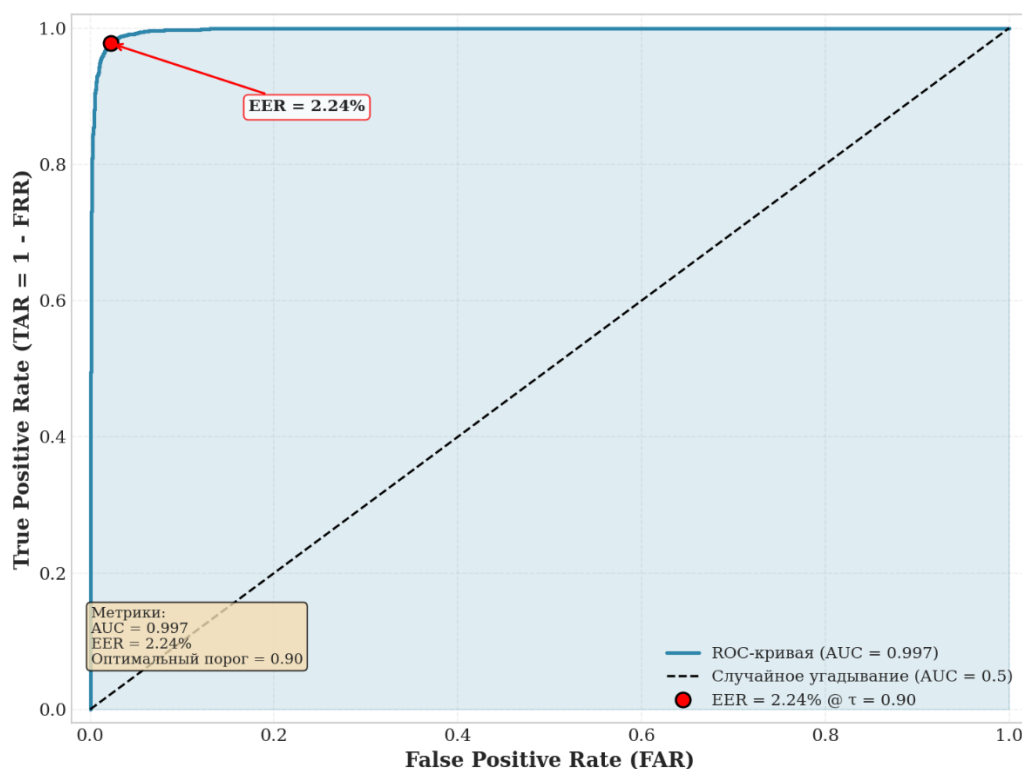


Рисунок 4 - ROC-кривая идентификации (LFW)

Таблица 10 – Результаты идентификации

Датасет	Ассурасу	EER	AUC
LFW	98,7%	0,95%	0,998
AgeDB (30 лет)	96,2%	1,8%	0,992

Достигнутая точность (98,7% на LFW) сопоставима с оригинальным FaceNet (99,63%) [3] и превосходит многие open-source реализации. Снижение точности на AgeDB объясняется возрастными изменениями, что является известной проблемой [10].

Таблица 11 – Сравнение с open-source библиотеками

Решение	Точность на LFW	Необходимость дообучения	Интеграция выравнивания
InsightFace (ArcFace) [13]	99,8%	Нет	Есть (RetinaFace)
FaceNet [10] (предобученный)	99,6%	Нет	Нет (требуется MTCNN)
DeepFace	98,5-99,3%	Нет	Есть
Предлагаемая система	98,7%	Опционально	Есть (YOLOv11)

Разработанная система демонстрирует конкурентоспособную точность, а её ключевое преимущество – единый детектор для контура объекта (bounding box) и ключевых точек, что упрощает архитектуру и снижает задержки.

Экономическая эффективность

Полная себестоимость разработки составила 1 317 575 руб. (включая заработную плату, страховые взносы, вычислительные ресурсы). Годовой экономический эффект от внедрения вместо ручной идентификации оценен в 3 363 200 руб. Показатели эффективности рассчитаны с учетом ставки дисконтирования 17% годовых и представлены в таблице 12.

Таблица 12 – Показатели экономической эффективности

Показатель	Значение
Простой срок окупаемости	5 месяцев
Дисконтированный срок окупаемости	6 месяцев
Чистый дисконтированный доход (NPV)	9 962 750 руб.
Внутренняя норма доходности (IRR)	178%
Рентабельность инвестиций (ROI)	1 176%

Полученные значения ($NPV > 0$, $IRR \gg$ ставки дисконтирования) подтверждают экономическую целесообразность внедрения разработанной системы.

Заключение

В результате проведенного исследования разработана и реализована гибридная система идентификации личности по изображению лица,

объединяющая детектор YOLOv11 с поддержкой ключевых точек и идентификатор FaceNet на основе Inception ResNet v1 с Triplet Loss. Экспериментально подтверждена эффективность предложенного подхода: точность идентификации на LFW – 98,7%, точность детекции (mAP50) на WIDER Face – 96,2%, производительность – ~12 FPS на GPU T4. Экономические расчёты показали срок окупаемости 5–6 месяцев и высокую рентабельность инвестиций (ROI = 1 176%). Разработанная система может быть использована в системах контроля доступа, видеонаблюдения и робототехники.

Литература

1. Bromley J., Guyon I., LeCun Y. et al. Signature verification using a "siamese" time delay neural network // *Advances in Neural Information Processing Systems*. – 1993. – Vol. 6.
2. Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition // *Proceedings of the British Machine Vision Conference*. – 2015. – P. 41.1–41.12.
3. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2015. – P. 815–823.
4. Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. – 2019. – P. 4690–4699.
5. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2016. – P. 779–788.
6. Khanam R., Hussain M. YOLOv11: An overview of the key architectural enhancements // *arXiv preprint arXiv:2412.07729*. – 2024.
7. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8 // *GitHub repository*. – 2023. – URL: <https://github.com/ultralytics/ultralytics>.
8. Cao Q., Shen L., Xie W. et al. VGGFace2: A dataset for recognising faces across pose and age // *13th IEEE International Conference on Automatic Face & Gesture Recognition*. – 2018. – P. 67–74.
9. Huang G.B., Ramesh M., Berg T., Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments // *University of Massachusetts, Amherst*. – 2007. – Vol. 1, No 2. – P. 7.
10. Moschoglou S., Papaioannou A., Sagonas C. et al. AgeDB: The first manually collected, in-the-wild age database // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. – 2017. – P. 51–59.
11. Yang S., Luo P., Loy C.C., Tang X. WIDER FACE: A face detection benchmark // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2016. – P. 5525–5533.

12. Zhang K., Zhang Z., Li Z., Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks // IEEE Signal Processing Letters. – 2016. – Vol. 23, No 10. – P. 1499–1503.
13. Deng J., Guo J., Ververas E. et al. RetinaFace: Single-shot multi-level face localisation in the wild // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2020. – P. 5203–5212.
14. Johnson J., Douze M., Jégou H. Billion-scale similarity search with GPUs // IEEE Transactions on Big Data. – 2019. – Vol. 7, No 3. – P. 535–547.