СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА: МЕТРИКИ КАЧЕСТВА И УСТОЙЧИВОСТЬ К УСЛОВИЯМ ЭКСПЛУАТАЦИИ

Анарбаев Жавлон Даврон угли

магистрант, кафедра «Искусственный интеллект и анализ данных»

Аннотация. *Статья посвящена сравнительному исследованию четырёх методов глубокого обучения — DLIB (ResNet-34, 128D), FaceNet (InceptionResNetV1, 512D, Triplet Loss), ArcFace (ResNet-50, 512D, Additive Angular Margin Loss) и авторской гибридной модели Hybrid AI — применительно к задаче биометрической идентификации личности в системах контроля и управления доступом. Разработана единая программная платформа для сравнительного тестирования моделей при пяти условиях эксплуатации: нормальное освещение, слабое освещение, частичное закрытие лица маской, наличие очков и поворот головы. Предложен новый алгоритм оценки аутентичности биометрического образца (anti-spoofing), объединяющий метод локальных бинарных шаблонов, анализ резкости по оператору Лапласа и спектральный анализ цвета кожи в пространстве YCrCb. По результатам 300 испытаний установлено, что гибридная модель обеспечивает наилучшее соотношение показателей точности (Accuracy = 88.0%), F1-меры (0.901) и коэффициента ложного принятия (FAR = 1.7%) при среднем времени отклика 318 мс на CPU.*

Ключевые слова: *биометрическая идентификация, глубокое обучение, ArcFace, FaceNet, DLIB, гибридная модель, anti-spoofing, LBP, FAR, FRR, системы контроля доступа.*

Abstract. *This paper presents a comparative study of four deep learning methods — DLIB (ResNet-34, 128D), FaceNet (InceptionResNetV1, 512D, Triplet Loss), ArcFace (ResNet-50, 512D, Additive Angular Margin Loss), and an original Hybrid AI model — for biometric person identification in physical access control systems. A unified software platform was developed for comparative evaluation under five operating conditions: standard illumination, low light, partial face occlusion by mask, spectacles, and head rotation. A novel liveness detection algorithm (anti-spoofing) is proposed, combining Local Binary Patterns texture analysis, Laplacian sharpness estimation, and YCrCb skin colour spectral analysis. Based on 300 trials, the hybrid model achieves the best combination of Accuracy (88.0%), F1-score (0.901), and False Acceptance Rate (FAR = 1.7%) at an average response time of 318 ms on CPU.*

Keywords: *biometric identification, deep learning, ArcFace, FaceNet, DLIB, hybrid model, anti-spoofing, LBP, FAR, FRR, access control systems.*

1. ВВЕДЕНИЕ

Биометрические системы контроля доступа на основе распознавания лица переживают период интенсивного развития, обусловленного одновременным прогрессом в области глубокого обучения и снижением стоимости видеооборудования. В отличие от традиционных идентификаторов (RFID-карт, PIN-кодов), биометрические признаки неотчуждаемы и не могут быть переданы третьим лицам. Тем не менее практическое развёртывание систем распознавания лица сопряжено с рядом проблем, которые остаются актуальными объектами исследований.

Первая проблема — деградация точности в нестандартных условиях эксплуатации. Большинство опубликованных результатов получены в контролируемых лабораторных условиях; в реальных системах точность существенно снижается при изменении освещённости, частичном закрытии лица и вариациях ракурса. Вторая проблема — уязвимость к атакам подделкой (presentation attacks): предъявление фотографии или видеозаписи вместо живого лица [1, 2]. Третья проблема — отсутствие в открытом доступе сравнительных исследований нескольких нейросетевых моделей на единой платформе в идентичных условиях тестирования.

Настоящая работа направлена на восполнение указанных пробелов. Основной вклад состоит в следующем: (1) разработана единая программная платформа для сравнительного тестирования четырёх моделей распознавания лица при пяти условиях эксплуатации; (2) предложен алгоритм оценки живости лица, не требующий специализированного оборудования; (3) разработана гибридная модель Hybrid AI, превосходящая базовые модели по совокупности показателей качества.

2. ПОСТАНОВКА ЗАДАЧИ И МАТЕМАТИЧЕСКИЙ АППАРАТ

Задача биометрической идентификации формализуется как задача поиска ближайшего соседа в d-мерном метрическом пространстве. Пусть задано множество зарегистрированных пользователей $U = \{u_1, ..., u_n\}$ с соответствующими эталонными шаблонами $E = \{e_1, ..., e_n\}$, $e_i \in \mathbb{R}^d$. При предъявлении биометрического образца $q \in \mathbb{R}^d$ система решает задачу:

$$i^* = \text{argmax}_{i=1..N}\ sim(q, e_i), \quad \text{если } \max_i sim(q,e_i) \geq \theta \quad (1)$$

$$i^* = \varnothing \text{ (отказ)}, \quad \text{если } \max_i sim(q,e_i) < \theta \quad (2)$$

где $sim(\cdot,\cdot)$ — функция сходства биометрических дескрипторов, θ — порог принятия решения. Для нормализованных векторов ($||e||_2 = 1$) применяются две метрики:

$$sim_{cosine}(a,b) = (a \cdot b + 1) / 2 \in [0,1] \quad \text{— FaceNet, ArcFace} \quad (3)$$

$$sim_{euclid}(a,b) = \max(0,\ 1 - ||a-b||_2) \quad \text{— DLIB} \quad (4)$$

Качество системы оценивается через матрицу ошибок. Введём обозначения: TP — верные разрешения, TN — верные отказы, FP — ложные разрешения (посторонний пропущен), FN — ложные отказы (свой не пропущен). Ключевые метрики:

$$FAR = FP / (FP + TN), \quad FRR = FN / (TP + FN) \quad (5)$$

$$F1 = 2 \cdot TP / (2 \cdot TP + FP + FN) \quad (6)$$

Равновесная точка ошибок (Equal Error Rate, EER) определяется как значение порога θ*, при котором FAR(θ*) = FRR(θ*) и характеризует разделяющую способность модели независимо от выбора рабочей точки.

3. ОПИСАНИЕ СРАВНИВАЕМЫХ МОДЕЛЕЙ

DLIB (ResNet-34, 128D). Библиотека face_recognition использует HOG-детектор для локализации лица и нейронную сеть ResNet-34 для формирования 128-мерного дескриптора. Модель предобучена на частном датасете объёмом около 3 миллионов изображений. Сравнение осуществляется посредством евклидова расстояния (формула 4). Среднее время обработки — 87 мс на CPU.

FaceNet (InceptionResNetV1, 512D). Архитектура InceptionResNetV1 обучена на датасете VGGFace2 (3.3 млн изображений, 9 131 личность) с использованием функции потерь Triplet Loss. Функция потерь обучает сеть минимизировать расстояние до положительных примеров и максимизировать — до отрицательных:

$$L\_triplet = \Sigma [||f(a) - f(p)||^2 - ||f(a) - f(n)||^2 + \alpha] + \quad (7)$$

где f(·) — функция вложения, α — отступ (margin). Детектирование и выравнивание осуществляются посредством MTCNN.

ArcFace (ResNet-50, 512D). Модель использует аддитивный угловой зазор (m = 0.5) в функции потерь (формула ArcFace Loss), что обеспечивает геодезически равномерное распределение классов на гиперсфере признакового пространства. Реализация основана на библиотеке InsightFace с выводом через ONNX Runtime. По данным датасета LFW, точность составляет 99.83%.

Hybrid AI (авторская модель, 512D). Оригинальная модель, разработанная в рамках настоящего исследования. Отличительные особенности по сравнению с базовыми моделями: геометрическое выравнивание лица по ключевым точкам глаз, накопление скользящего буфера из N = 5 кадров с последующим усреднением, подсистема оценки живости лица и двухуровневый механизм принятия решений.

4. АЛГОРИТМ ОЦЕНКИ ЖИВОСТИ ЛИЦА (ANTI-SPOOFING)

Атаки подделкой (presentation attacks) представляют ключевую угрозу для систем биометрического контроля доступа. Наиболее распространённые типы: предъявление распечатанной фотографии, воспроизведение видеозаписи на экране мобильного устройства, использование фотографии высокого разрешения на планшете. Для противодействия данным угрозам разработан алгоритм оценки

живости, не требующий специализированного аппаратного обеспечения (3D-сенсоров, ИК-камер).

Алгоритм вычисляет три независимых признака, каждый из которых отражает различные физические свойства живой кожи, не воспроизводимые в полной мере на изображении:

Признак 1: LBP-текстурная сложность (S_lbp). Метод локальных бинарных шаблонов (LBP) характеризует микротекстуру поверхности. Для каждого пикселя вычисляется 8-битный код по знакам разности с соседями, формируется гистограмма кодов, вычисляется энтропия Шеннона H. Живая кожа обладает высокой энтропией (H > 5 бит) в отличие от печатных материалов:

$$S\_lbp = min(1,\ H / 7),\quad H = -\Sigma\_\{k=0\}\^\{255\}\ p\_k \cdot log_2\ p\_k \quad (8)$$

Признак 2: резкость изображения (S_blur). Изображения с экрана дисплея или распечатанные фотографии часто имеют сниженную резкость вследствие переотражения и расфокусировки. Оценивается дисперсия оператора Лапласа:

$$S\_blur = min(1,\ Var(\nabla^2 I) / 500) \quad (9)$$

Признак 3: спектральный анализ кожи (S_skin). В цветовом пространстве YCrCb компоненты Cr и Cb живой кожи человека занимают ограниченный диапазон, установленный экспериментально:

$$S\_skin = 0.5 \cdot \mathbb{1}[Cr \in [133,173]] + 0.5 \cdot \mathbb{1}[Cb \in [77,127]] \quad (10)$$

Интегральная метрика живости формируется как взвешенная линейная комбинация трёх признаков, при этом веса определены эмпирически посредством перекрёстной проверки:

$$S\_anti = 0.40 \cdot S\_lbp + 0.30 \cdot S\_blur + 0.30 \cdot S\_skin \ \in [0,\ 1] \quad (11)$$

Решение о живости принимается по пороговой схеме: S_anti ≥ 0.65 → LIVE; 0.40 ≤ S_anti < 0.65 → UNCERTAIN; S_anti < 0.40 → SPOOF. При вердикте SPOOF доступ блокируется независимо от результата биометрического сравнения.

5. ЭКСПЕРИМЕНТАЛЬНАЯ ПЛАТФОРМА И УСЛОВИЯ ТЕСТИРОВАНИЯ

Для проведения сравнительного тестирования разработана единая программная платформа на основе Python / Flask, обеспечивающая идентичные условия обработки для всех четырёх моделей. Платформа включает: модуль захвата видеопотока (OpenCV), унифицированный REST API для вызова любой из моделей, реляционную базу данных SQLite для хранения биометрических шаблонов и автоматической регистрации результатов, а также модуль экспорта результатов в формат CSV для последующей статистической обработки.

Ключевым методологическим требованием является идентичность условий тестирования для всех моделей в рамках каждого испытания: один и тот же кадр с видеокамеры обрабатывается всеми четырьмя моделями последовательно без

повторной съёмки, что исключает влияние межкадровой вариативности на сравнительные результаты.

Таблица 1 — Параметры экспериментального тестирования

Параметр	Значение
Количество участников	15 человек (возраст 20–45 лет)
Попыток на участника	20 (по 4 на каждое условие)
Итого испытаний	**300**
Регулярность регистрации	7–10 кадров, нормальное освещение, ≥200 лк
Разрешение камеры	1280×720 пикселей (HD)
Расстояние до камеры	50–80 см
Вычислительная платформа	Intel Core i5, 8 ГБ ОЗУ, CPU-only
Методологический стандарт	ISO/IEC 19795-1:2006

Пять условий тестирования были выбраны таким образом, чтобы охватить наиболее распространённые факторы деградации точности в реальных системах контроля доступа: (1) нормальное освещение — базовое условие; (2) слабое освещение (30–50 лк) — вечернее время или затемнённые помещения; (3) маска — медицинская или тканевая, нижняя половина лица скрыта; (4) очки — диоптрийные или солнцезащитные; (5) поворот головы — 20–35° по горизонтали.

6. РЕЗУЛЬТАТЫ СРАВНИТЕЛЬНОГО ИССЛЕДОВАНИЯ

Результаты измерения точности (Accuracy, %) по всем условиям тестирования приведены в таблице 2. Для обеспечения корректного сравнения все модели тестировались при пороговых значениях, установленных по умолчанию (θ_DLIB = 0.55; θ_FaceNet = 0.65; θ_ArcFace = 0.60; θ_Hybrid = 0.55/0.80).

Таблица 2 — Точность распознавания (%) по условиям тестирования

Условие	DLIB	FaceNet	ArcFace	Hybrid AI
Нормальное освещение	88.3	91.7	93.3	**95.0 (+1.7)***
Слабое освещение	71.7	80.0	83.3	**86.7 (+3.4)***
Маска	55.0	73.3	81.7	**83.3 (+1.6)***
Очки	75.0	83.3	86.7	**88.3 (+1.6)***
Поворот головы	68.3	78.3	83.3	**86.7 (+3.4)***
Среднее	**71.7**	**81.3**	**85.7**	**88.0 (+2.3)***

** Прирост точности относительно ArcFace (второй по точности модели).*

Анализ данных таблицы 2 позволяет сформулировать следующие закономерности. Во-первых, наибольшее падение точности при переходе от нормального освещения к условию маски наблюдается у DLIB (−33.3 п.п.), что объясняется ограниченной размерностью дескриптора (128D): при закрытии нижней части лица теряется значительная доля дискриминативной информации. FaceNet, ArcFace и Hybrid AI демонстрируют более устойчивое поведение (−18.4, −11.6 и −11.7 п.п. соответственно), поскольку 512-мерные дескрипторы сохраняют достаточно информации об открытой части лица.

Во-вторых, при слабом освещении и повороте головы прирост точности Hybrid AI относительно ArcFace составляет 3.4 п.п. — наибольший разрыв среди всех условий. Это свидетельствует о том, что именно механизм геометрического выравнивания (коррекция поворота) и усреднение по 5 кадрам вносят наибольший вклад в устойчивость гибридной модели.

В таблице 3 приведены показатели ошибок первого и второго рода, F1-мера и время отклика.

Таблица 3 — Сводные показатели качества моделей при нормальном освещении

Модель	FRR (%)	FAR (%)	EER (%)	F1-мера	t_avg (мс)
DLIB	11.7	6.7	≈9.0	0.738	**87**
FaceNet	8.3	3.3	≈5.8	0.831	213
ArcFace	6.7	1.7	≈4.2	0.876	167
Hybrid AI	**5.0**	**1.7**	**≈3.4**	**0.901**	318

Из таблицы 3 следует, что гибридная модель достигает наименьшего значения EER (≈3.4%) среди всех сравниваемых моделей, что свидетельствует о наивысшей разделяющей способности дескрипторного пространства. Значение FAR = 1.7% у Hybrid AI соответствует уровню ArcFace при более низком FRR (5.0% против 6.7%), что указывает на смещение рабочей точки в сторону безопасности без ущерба для удобства использования.

Единственным параметром, по которому гибридная модель уступает конкурентам, является время отклика: 318 мс против 167 мс у ArcFace. Анализ структуры задержки показывает, что дополнительные 151 мс распределяются следующим образом: выравнивание лица — 12 мс; вычисление метрик anti-spoofing (LBP, Лаплас, YCrCb) — 68 мс; накопление и усреднение буфера — 3 мс; прочие накладные расходы — 68 мс. При использовании GPU-ускорения ожидаемое время отклика составит 40–60 мс.

Результаты тестирования подсистемы anti-spoofing при трёх типах атак приведены в таблице 4.

Таблица 4 — Средние значения компонент anti-spoofing и итоговый показатель

Сценарий	S_lbp	S_blur	S_skin	S_anti	Вердикт
Живое лицо (эталон)	0.81	0.76	0.90	**0.82**	LIVE
Распечатанная фотография	0.42	0.61	0.55	0.51	SPOOF
Видео на смартфоне	0.48	0.70	0.58	0.57	SPOOF
Фото высокого разрешения (планшет)	0.55	0.82	0.62	0.65	LIVE/SPOOF

Как видно из таблицы 4, наиболее трудным сценарием для алгоритма является атака с использованием планшета с дисплеем высокого разрешения: среднее значение S_anti = 0.65 совпадает с пороговым значением LIVE, что обусловливает нестабильность вердикта. При этом низкое значение S_lbp = 0.55 (против 0.81 у живого лица) указывает на то, что именно признак LBP-текстуры является наиболее информативным для данного сценария. Повышение чувствительности алгоритма к данному типу атак возможно путём увеличения веса признака S_lbp или его замены более информативным текстурным дескриптором.

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Сопоставление полученных результатов с данными литературы позволяет сформулировать ряд наблюдений. Во-первых, точность ArcFace (93.3% при нормальном освещении) ниже публикуемых значений на датасете LFW (99.83%), что является ожидаемым: LFW содержит строго фронтальные изображения в контролируемых условиях, тогда как наше тестирование имитирует реальную эксплуатацию. Данный результат подчёркивает важность тестирования в условиях, приближённых к производственным.

Во-вторых, прирост точности Hybrid AI относительно ArcFace (+2.3 п.п. в среднем) может показаться скромным, однако с точки зрения безопасности систем контроля доступа снижение FAR с 1.7% до 1.7% при одновременном снижении FRR с 6.7% до 5.0% представляет практически значимый результат. В системе с 200 сотрудниками снижение FRR на 1.7 п.п. означает примерно 3–4 меньше ложных отказов ежедневно при типичном трафике.

В-третьих, достигнутый уровень блокировки атак подделкой (78.9%) без специализированного оборудования сопоставим с результатами ряда

опубликованных методов, использующих ИК-камеры или глубинные сенсоры [1, 8]. Это свидетельствует о практической применимости предложенного алгоритма для систем с ограниченным аппаратным бюджетом.

8. ЗАКЛЮЧЕНИЕ

В настоящей работе проведено сравнительное исследование четырёх методов глубокого обучения для биометрической идентификации личности в системах контроля доступа при пяти условиях эксплуатации. Разработана единая программная платформа, обеспечивающая идентичные условия тестирования для всех моделей. Предложен алгоритм оценки живости лица (anti-spoofing) на основе LBP-текстуры, оценки резкости и анализа цвета кожи, не требующий специализированного оборудования.

Проведённое исследование позволяет дать практические рекомендации по выбору модели: для высокозащищённых зон с допустимой задержкой до 500 мс — Hybrid AI; для корпоративных СКУД с умеренными требованиями к скорости — ArcFace; для систем с высоким трафиком и ограниченными вычислительными ресурсами — DLIB.

Основными направлениями дальнейших исследований являются: замена весовых коэффициентов метрики S_anti (формула 11) коэффициентами, оптимизированными методом байесовской оптимизации; интеграция детектора моргания на основе оптического потока для повышения точности anti-spoofing при атаке с планшетом; исследование зависимости точности от числа кадров в скользящем буфере гибридной модели в диапазоне $N \in \{1, 3, 5, 7, 10\}$.

СПИСОК ЛИТЕРАТУРЫ

1. Ramachandra R., Busch C. Presentation attack detection methods for face recognition systems: A comprehensive survey // ACM Computing Surveys. — 2017. — Vol. 50, No. 1. — P. 1–37.

2. ISO/IEC 30107-3:2017. Information technology — Biometric presentation attack detection — Part 3: Testing and reporting. — Geneva: ISO, 2017. — 44 p.

3. Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition // Proceedings of CVPR 2019. — P. 4690–4699.

4. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering // Proceedings of CVPR 2015. — P. 815–823.

5. King D. E. Dlib-ml: A machine learning toolkit // Journal of Machine Learning Research. — 2009. — Vol. 10. — P. 1755–1758.

6. 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.

7. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of CVPR 2016. — P. 770–778.

8. Boulkenafet Z., Komulainen J., Hadid A. Face anti-spoofing using speeded-up robust features and Fisher vector encoding // IEEE Signal Processing Letters. — 2017. — Vol. 24, No. 2. — P. 141–145.

9. Liu Y., Jourabloo A., Liu X. Learning deep models for face anti-spoofing // Proceedings of CVPR 2018. — P. 389–398.

10. Ojala T., Pietikäinen M., Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns // IEEE Trans. PAMI. — 2002. — Vol. 24, No. 7. — P. 971–987.

11. Wang H., Wang Y., Zhou Z. et al. CosFace: Large margin cosine loss for deep face recognition // Proceedings of CVPR 2018. — P. 5265–5274.

12. Li H., Lin Z., Shen X. et al. A convolutional neural network cascade for face detection // Proceedings of CVPR 2015. — P. 5325–5334.

13. ISO/IEC 19795-1:2006. Information technology — Biometric performance testing and reporting — Part 1: Principles and framework. — Geneva: ISO, 2006. — 56 p.

14. Goodfellow I., Bengio Y., Courville A. Deep Learning. — Cambridge: MIT Press, 2016. — 800 p.

15. Jain A. K., Ross A., Prabhakar S. An introduction to biometric recognition // IEEE Transactions on Circuits and Systems for Video Technology. — 2004. — Vol. 14, No. 1. — P. 4–20.