

Детектирование и распознавание лиц в системах видеонаблюдения

А. Б. Мудрич, К. В. Ежова

Аннотация—В статье рассматриваются современные подходы для решения задач детектирования и распознавания лиц в биометрических системах видеонаблюдения.

Интеллектуальные системы на основе биометрии с каждым годом получают все большее распространение в различных сферах жизни человека – от систем безопасности до банков и магазинов. В качестве биометрических данных могут выступать различные уникальные биологические идентификаторы человека - отпечатки пальцев, рисунок сетчатки глаза, текстура кожи, почерк и другое. Но наибольшее распространение получили системы на основе распознавания лиц, так как их характеризуют минимальные требования к аппаратной части: достаточно размещения камеры видеонаблюдения и относительная легкость имплементации алгоритма распознавания. В данной работе будут рассмотрены наиболее широко используемые на настоящий момент алгоритмы детектирования и распознавания лиц, а также требования к наборам данных, используемым для обучения моделей глубокого обучения.

В первом разделе приведено понятие лица как идентификатора в биометрической системе распознавания. Во втором разделе – краткое описание наборов данных, используемых для обучения моделей для детекции и распознавания лиц. Третий раздел содержит обзор наиболее распространённых моделей и библиотек детектирования и распознавания лиц. В заключительном разделе представлен пример структуры биометрической системы распознавания лиц.

Ключевые слова — биометрическая система, детектирование лиц, распознавание лиц, система видеонаблюдения, система безопасности.

I. ВВЕДЕНИЕ

Важным элементом любой интеллектуальной системы, в задачи которой входит анализ биометрических данных, является работа с биометрическими данными человека [1]. Эти данные представляют собой идентификатор – признак, по которому можно однозначно определить личность посетителя или пользователя [2]. Для рассматриваемых систем таким идентификатором является лицо человека.

Геометрия лица является уникальным биометрическим идентификатором. В процессе анализа из биометрических данных (например, 2D-снимка) извлекаются значимые признаки (вектор признаков), которые в дальнейшем используются для сравнения

эталона (изображения в базе данных) с детектируемым объектом (кадром с камеры видеонаблюдения). Для лица, как биометрического идентификатора, в зависимости от используемого алгоритма, может быть значимым выделение частей лица (характерных зон) и определение расстояний между этими зонами. В ходе анализа фрагмент с изображением лица, полученный в результате выполнения алгоритма детекции, декомпозируется на основные компоненты (например, глаза, нос, рот и зоны с характерным освещением). Существует отдельное направление исследований, посвященное статистическому анализу значимости отдельных компонентов (зон) лица и построению математических моделей для определения весов (численных коэффициентов) для разных зон лица для последующего применения на этапе декомпозиции и выделения признаков [3].

Также в качестве ключевых признаков могут выступать расстояния между выделенными зонами лица. В таком случае, после декомпозиции для каждой зоны определяются «центр тяжести» - якорная точка элемента и затем между центрами тяжести всех сегментов вычисляются расстояния, используя уравнение расстояния между двумя координатами. В качестве итоговых параметров будет получен набор расстояний [4].



Рис 1. Представление дистанций между выделенными зонами лица [4]

На основе извлеченных признаков формируется карта признаков (вектор признаков), который используется как входные данные для обучения модели или уже непосредственного выполнения задачи идентификации (если речь идет о предобученной модели или алгоритме).

Все изображения, на основе которых выполняется распознавание лиц в биометрической системе, можно условно разделить на две большие группы – плоскостные 2D-изображения и объемные 3D-модели. В

первом случае в качестве эталона используется фотография из базы данных, в качестве изображения для сравнения – кадр из видеопотока с камеры. В случае 3D-модели производится реконструирование трехмерных признаков лица и создание математической объемной модели. Эталонным изображением является модель, полученная в результате 3D-сканирования, или восстановления по нескольким изображениям лица с разных ракурсов (например, с нескольких камер), в качестве изображения для сравнения также может выступать результат сканирования или восстановленная по нескольким изображениям в разных ракурсах 3D-модель.

На настоящий момент 2D-распознавание лиц получило наибольшее распространение, так как существуют готовые базы данных эталонов и готовая техническая инфраструктура в виде камер видеонаблюдения. В свое время 3D-распознавание требует более сложных алгоритмов обработки и восстановления объемного изображения, а также специальных сканеров или несколько камер, размещенных специальным образом. Но при этом оно обладает меньшим значением коэффициентов ошибок FAR (False Acceptance Rate) и FRR (False Rejection Rate) [5] и более устойчиво к спуфинговым атакам.

II. НАБОРЫ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ

Для работы любого алгоритма распознавания и детекции важным является этап извлечения признаков и формирования входного вектора для обработки выбранной моделью. Извлечение признаков и определение их значимости происходит на этапе обучения (тренировки модели) на наборе данных. Поэтому качество всей модели будет определяться не только выбранным перечнем и значением параметров, влияющих на ее структуру (например, глубина дерева решений, количество слоев в нейронной сети), но и данными, используемыми на этапе обучения.

Для обучения моделей распознавания лиц используются общедоступные наборы данных. Перечень всех открытых для использования в исследовательских целях датасетов можно найти в специальной базе данных «Datasets for face recognition», которая содержит название, описание набора данных и его лицензии [6].

Условно все существующие наборы, используемые для обучения моделей для детекции и распознавания лиц, можно разделить на несколько больших групп:

1) Наборы, содержащие фотографии и видео с размеченными лицами. Например, Flickr-Faces-HQ (FFHQ) [7], The Yale Face Database [8] и другие.



Рис 2. Примеры изображений набора The Yale Face Database [8]

2) Крупномасштабные наборы, состоящие из изображений с разной модальностью (видимое, ближнее инфракрасное, тепловое, компьютеризированный эскиз,

LYTRO, записанное видео, 3D-изображения и др.). К данному типу относится известный набор данных Tufts Face Dataset [9].



Рис 3. Примеры изображений набора Tufts Face Dataset [9]

3) Наборы из реальных и сгенерированных изображений, целью которых является борьба с фальсификацией (спуфингом) биометрических данных (чаще всего используются для обучения состязательных нейронных сетей GAN) Примером такого датасета является Real and Fake Face Detection [10].



Рис 4. Примеры изображений набора Real and Fake Face Detection [10]

На результат распознавания значительно влияют характеристики физической среды, в которой получено изображение (уровень освещенности, разрешение камеры и другое), и наличие шумов и помех на самом изображении (кадре) – например, очки, медицинская маска. Необходимо учитывать эти факторы, снижающие качество распознавания, при выборе данных для обучения модели. Набор данных должен максимально охватывать все возможные случаи изображений лица, также желательно использование для обучения модели нескольких наборов данных.

Под специфические задачи интеллектуального анализа биометрических данных могут использоваться специальные наборы данных. Например, для повышения устойчивости к атакам на видеосистемы можно использовать датасеты, содержащие реальные и фейковые изображения лиц.

III. ПОДХОДЫ К ДЕТЕКТИРОВАНИЮ И РАСПОЗНАВАНИЮ ЛИЦ

В процессе интеллектуального анализа биометрических данных из видеопотока можно выделить два последовательных этапа выделения и обработки объекта (лица человека). Это детектирование – выделение фрагмента с лицом на кадре, захваченном из непрерывного потока видео и непосредственно распознавание – сопоставление выделенной как лицо области с эталонным изображением и расчет меры соответствия (близости) двух изображений.

В качестве меры близости (вычисляемой дистанции) эталонного изображения и полученного кадра могут выступать различные математические меры расстояния двух величин. Например, косинусное расстояние, евклидово расстояние и другие. Как правило, конкретная мера расстояния задается в параметрах модели.

А. Детектирование лиц

Процесс обработки видеопотока в биометрической системе проходит последовательно – сначала выделяются отдельные кадры из общего потока данных, затем для каждого кадра в сформированном массиве выделяются фрагменты, содержащие целевой объект – лицо человека. Формируется новый массив из фрагментов, которые затем подаются на вход модели распознавания.

Задача детектирования относится к этапу выделения фрагментов с лицами в массиве кадров и успешно решается базовыми алгоритмами библиотек OpenCV и Dlib. Рассмотрим несколько классических алгоритмов детектирования, реализованных в данных библиотеках.

Примером наиболее раннего алгоритма является каскадный классификатор Хаара. Каскадный классификатор можно отнести к группе ансамблевых алгоритмов бустинга. В основе работы каскада Хаара лежит метод Виолы-Джонса, позволяющих обнаруживать объекты на изображениях в реальном времени.

Позапно алгоритм Виолы-Джонса проходит следующие шаги: вейвлет-преобразование Хаара, получение интегрального изображения, обучение классификатора AdaBoost и использование каскадных классификаторов [11]. На этапе вейвлет-преобразования Хаара происходит извлечение признаков с помощью функций Хаара: находится первая функция как разница интенсивностей темной и светлой областей фрагмента лица, затем найденный прямоугольник применяется в качестве сверточного ядра ко всему изображению. Для осуществления вейвлет-преобразования можно применять несколько разных типов прямоугольников в качестве функций.

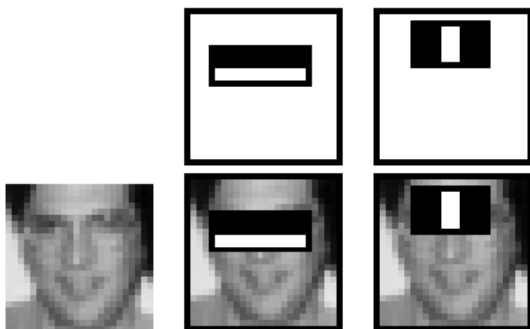


Рис 5. Примеры выбора признаков с помощью преобразования Хаара [11]

Выбранные функции применяются к изображению с использованием классификатора AdaBoost. Далее вычисляется прямоугольная сумма на основе четырех значений используя интегральное изображение. На завершающем этапе для определения областей изображения, содержащих лицо, используются

каскадные классификаторы – деревья решений, которые последовательно исключают окна, не содержащие пиксели с лицом. Данный алгоритм реализован в библиотеке OpenCV.

Вторым популярным алгоритмом детектирования объектов является Гистограмма направленных градиентов (Histogram of Oriented Gradients (HOG)). Алгоритм заключается в формировании вектора признаков, представляющих из себя распределения направленных градиентов. Градиенты позволяют определять граничные (переходные) области лица, содержащие различные зоны. Затем сформированный вектор передается на вход алгоритма классификации (например, SVM) [12].

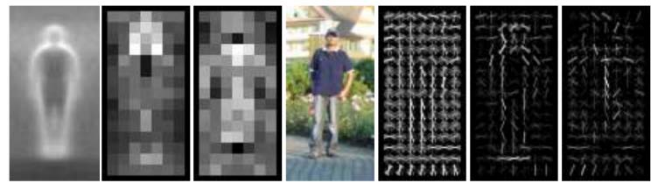


Рис 6. Гистограммы направленных градиентов [12]

Примером использования алгоритма глубокого обучения является использование сверточной нейронной сети (CNN). В процессе обработки происходит свертка исходного изображения с помощью набора заданных фильтров. На выходе мы получаем значение бинарной классификации – 0 или 1, что соответствует отсутствию или наличию лица на входном изображении.

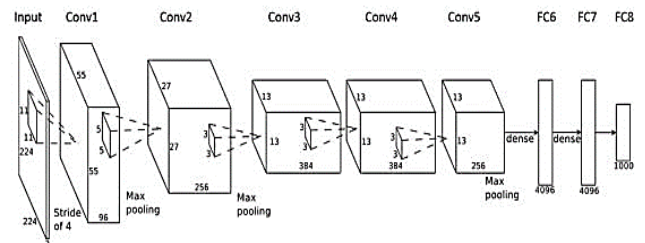


Рис 7. Архитектура CNN [13]

Алгоритмы Каскад Хаара и HOG показывают более быстрые результаты вычислений, при этом CNN выдает более точные результаты классификации. Таким образом, выбор конкретного алгоритма зависит от требований к точности и скорости вычислений в ходе детектирования объектов.

В. Модели распознавания лиц

Последним этапом в работе алгоритма биометрической системы является распознавание – сопоставление полученного фрагмента с лицом с эталонным изображением.

Для решения задач распознавания уже существует большое количество как классических, так и нейросетевых алгоритмов. Рассмотрим несколько примеров наиболее современных из них.

Первым является специализированный фреймворк fViT framework на базе алгоритмов визуальных трансформеров (visual transformers). Модель состоит из облегченной нейронной сети и визуального

трансформера. Нейронная сеть отвечает за выделение основных координат лица (нос, глаза и другие), а визуальный трансформер анализирует выделенные участки [14].

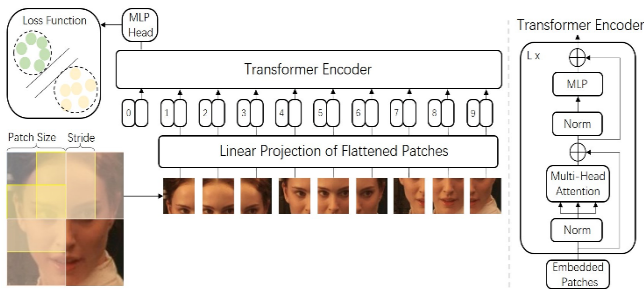


Рис 8. Архитектура fViT framework [15]

Также активно развивается направление использования для идентификации личности данных разных модальностей, что повышает точность распознавания и повышает устойчивость системы к атакам. В качестве входных данных могут выступать как биометрические идентификаторы одной природы (например, изображения лица разной модальности), так и разные по природе данные (например, изображение лица и запись голоса). Обычно в таких системах разные типы данных обрабатываются отдельно, а затем результат суммируется в соответствии с заданным интегральным алгоритмом и делается общий вывод о совпадении эталона и полученных данных. Например, ниже представлена простая модель из двух сверточных нейронных сетей, обрабатывающих входные изображения разной модальности [16].

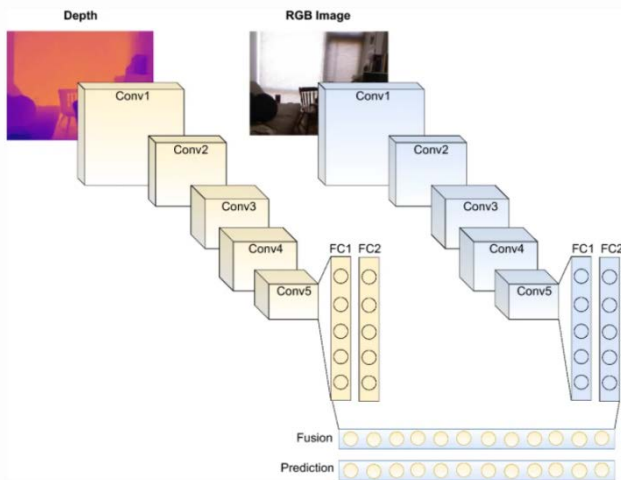


Рис 9. Пример бимодальной CNN [16]

ArcFace является примером модели обучения подобия (similarity learning). В качестве результата вычисляется расстояние между изображениями. В качестве метрики расстояния используется косинусное расстояние. Вместо Softmax Loss в качестве функции потерь используется Angular Margin Loss [16].

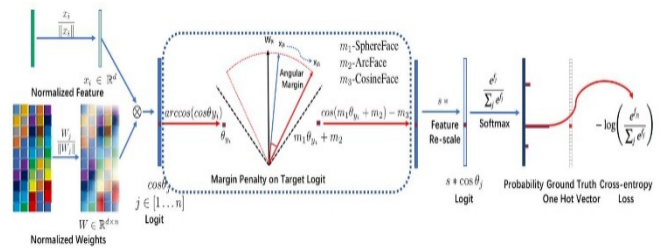


Рис 10. Архитектура ArcFace [17]

Выше были приведены примеры наиболее современных подходов к построению моделей распознавания лиц. При этом необходимо учитывать, что в этой области по-прежнему активно используются ставшие уже классическими сверточные (CNN) и рекуррентные нейронные сети (RNN). Для задач обработки видеопотока как правило они используются совместно.

IV. ПОСТРОЕНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ РАСПОЗНАВАНИЯ ЛИЦ

Любая интеллектуальная система распознавания лиц, несмотря на огромное количество различных алгоритмов детектирования и распознавания, будет содержать в себе базовые функциональные элементы, среди которых:

- 1) Входные данные в виде кадров видеопотока.
- 2) Классификатор для детектирования фрагментов, содержащих лица.
- 3) Блок распознавания.
- 4) Выходные данные: результат бинарной классификации (0 или 1) или значение метрики близости эталонного и входного изображения.

Аналогично можно унифицировать процесс распознавания и разделить его на этапы: определения лица, выравнивание контура лица, определение базовых параметров и сравнение параметров эталонного и входного изображения [18].



Рис 11. Общая структура распознавания лиц [18]

При проектировании интеллектуальной системы распознавания лиц важно учитывать необходимость реализации всех функциональных блоков и возможность их совмещения при использовании конкретных технологий для каждого функционального блока. Также важным параметром при разработке системы данного класса выступает ее быстродействие, поэтому уже на этапе проектирования недостаточно просто разрабатывать программную реализацию модели, но также необходимо учитывать аппаратные возможности как самой системы видеонаблюдения, так и серверной части для обработки данных и вычислений.

V. ЗАКЛЮЧЕНИЕ

Среди биометрических систем наибольшее развитие в настоящий момент получили именно системы на основе распознавания лиц. Любая интеллектуальная система распознавания лиц содержит в себе базовые функциональные модули детекции и распознавания, а в качестве выходных данных выдает результат бинарной классификации или значение метрики близости.

При этом значительное влияние на результат распознавания оказывает не только архитектура выбранных моделей для функциональных блоков детекции и распознавания, но и качество данных, использованных для обучения модели. Модели распознавания чувствительны к внешним условиям среды, в которых получено изображение и к наличию шума и помех на самом входном изображении. Поэтому важно использование наиболее полного датасета (нескольких наборов данных) при обучении модели.

Другим важным параметром при проектировании интеллектуальной системы распознавания лиц является необходимость согласования работы всех блоков и учет быстродействия, так как предполагается работа системы в режиме реального времени. Именно это делает важным аппаратную реализацию системы.

БИБЛИОГРАФИЯ

- [1] Мудрич А.Б., Ежова К.В. Анализ подходов к разработке системы контроля и управления доступом предприятия. *International Journal of Open Information Technologies*. 2023. Т. 11. № 3. С. 95-99.
- [2] Ворона В. А., Тихонов В. А. Системы контроля и управления доступом. М.: Горячая линия–Телеком. – 2013. – 272 с.
- [3] Калинин М. А. Исследование наиболее информативных анатомических признаков лица человека и их формализация // *Вестник СибАДИ*. 2010. - №16. (<https://cyberleninka.ru/article/n/issledovanie-naibolee-informativnyh-anatomicheskikh-priznakov-litsa-cheloveka-i-ih-formalizatsiya>).
- [4] Widodo С.Е., Adi K. Face geometry as a biometric-based identification system. *Journal of Physics: Conference Series*, 2020. – Vol. 1524.
- [5] Якубов Н. Лицевая биометрия в системах контроля и управления доступом и не только. *Современные технологии автоматизации*, 2020. - №3. – С. 12–16.
- [6] Datasets for face recognition // [interstellarengine.com](https://interstellarengine.com/ai/dataset-face-recognition.html) (<https://interstellarengine.com/ai/dataset-face-recognition.html>).
- [7] Flickr-Faces-HQ (FFHQ) (<https://github.com/NVLabs/ffhq-dataset>).
- [8] Yale Face Database (<http://vision.ucsd.edu/content/yale-face-database>).
- [9] Tufts Face Database (<https://www.kaggle.com/datasets/kpvisionlab/tufts-face-database>).
- [10] Real and Fake Face Detection (<https://www.kaggle.com/ciplab/real-and-fake-face-detection>).
- [11] Viola P., Jones M. J. Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001. - Vol. 1. – P. 511-518.
- [12] Dalal N., Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2001.
- [13] Fabien M. A full guide to face detection (<https://maelfabien.github.io/tutorials/face-detection/#c-compute-the-hog>).
- [14] Sun Zh., Tzimiropoulos G. Part-based Face Recognition with Vision Transformers, 2022 (https://www.researchgate.net/publication/365943272_Part-based_Face_Recognition_with_Vision_Transformers).
- [15] Face-Transformer (<https://github.com/zhongyy/Face-Transformer>).
- [16] Bayouhdh Kh, Knani R., Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, 2022. - №38. – P. 2939–2970.
- [17] Deng D., Guo J., Yang J., Xue N., Kotsia I. and Zafeiriou St. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *JOURNAL OF LATEX CLASS FILES*, 2015. - VOL. 14, №. 8.
- [18] Эмомов М. И. Анализ современных подходов распознавания и сопоставления лиц для систем биометрического контроля. *Вестник магистратуры*, 2019. - № 1-2(88). – С. 41-46.

Face detection and recognition in video surveillance systems

A. B. Mudrich, K.V.Ezhova

Abstract—The article discusses modern approaches to research the tasks of face detection and recognition in biometric video surveillance systems.

Intelligent systems based on biometrics are becoming more widespread every year in various spheres of human life – from security systems to banks and shops. Various unique biological identifiers of a person can act as biometric data - fingerprints, retinal pattern, skin texture, handwriting, and more. But the most widespread systems are based on facial recognition. Such systems are characterized by minimal hardware requirements: it is enough to place a video surveillance camera and the ease of implementation of the recognition algorithm. The most common algorithms for detecting and recognizing faces and the requirements for data sets used for training models will be considered in this paper.

The first section describes the concept of a face as an identifier in a biometric recognition system. The second section describes different data sets used to train models for face detection and recognition. The third section contains a description of the basic common models and libraries of face detection and recognition. The final section provides an example of the structure of a biometric facial recognition system.

Keywords—biometric system, face detection, face recognition, video surveillance system, security system.

REFERENCES

- [1] Mudrich A. B., Ezhova K. V. Analysis of the approaches to the development control access systems. International Journal of Open Information Technologies. 2023. Vol. 11. No. 3. pp. 95-99.
- [2] Vorona V. A., Tikhonov V. A. Access control and management systems. M.: Goryachaya liniya–Telecom. – 2013. – 272 pp.
- [3] Kalinin M. A. Investigation of the most informative anatomical features of a human face and their formalization // Vestnik SibADI. 2010. - №16. (<https://cyberleninka.ru/article/n/issledovanie-naibolee-informativnyh-anatomicheskikh-priznakov-litsa-cheloveka-i-ih-formalizatsiya>).
- [4] Widodo C.E., Adi K. Face geometry as a biometric-based identification system. Journal of Physics: Conference Series, 2020. – Vol. 1524.
- [5] Yakubov N. Facial biometrics in access control and management systems and not only. Modern automation technologies, 2020. – Vol. 3. – P. 12–16.
- [6] Datasets for face recognition // interstellarengine.com (<https://interstellarengine.com/ai/dataset-face-recognition.html>).
- [7] Flickr-Faces-HQ (FFHQ) (<https://github.com/NVLabs/ffhq-dataset>).
- [8] Yale Face Database (<http://vision.ucsd.edu/content/yale-face-database>).
- [9] Tufts Face Database (<https://www.kaggle.com/datasets/kpvisionlab/tufts-face-database>).
- [10] Real and Fake Face Detection (<https://www.kaggle.com/ciplab/real-and-fake-face-detection>).
- [11] Viola P., Jones M. J. Rapid Object Detection using a Boosted Cascade of Simple Features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. - Vol. 1. – P. 511-518.
- [12] Dalal N., Triggs B. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2001.
- [13] Fabien M. A full guide to face detection (<https://maelfabien.github.io/tutorials/face-detection/#c-compute-the-hog>).
- [14] Sun Zh., Tzimiropoulos G. Part-based Face Recognition with Vision Transformers, 2022 (https://www.researchgate.net/publication/365943272_Part-based_Face_Recognition_with_Vision_Transformers).
- [15] Face-Transformer (<https://github.com/zhongyy/Face-Transformer>).
- [16] Bayouhd Kh, Knani R., Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer, 2022. - Vol. 38. – P. 2939–2970.
- [17] Deng D., Guo J., Yang J., Xue N., Kotsia I. and Zafeiriou St. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. JOURNAL OF LATEX CLASS FILES, 2015. - VOL. 14, №. 8.
- [18] Emomov M. I. Analysis of modern approaches to face recognition and matching for biometric control systems. Vestnik Magistratury, 2019. – Vol. 1-2(88). – P. 41-46.