

Конкатенация нейронных сетей в системе биометрической аутентификации пользователей компьютерной информационной системы

М.Б. Гузаиров, А.С. Исмаилова, Н.Д. Лушников

Уфимский университет науки и технологий, Уфа

Аннотация: В данной статье описана конкатенированная архитектура нейронной сети при распознавании личности по изображению лица и по голосу в ходе обучения. Для обучения нейронных сетей в качестве входных векторов применяются извлеченные признаки изображения лица и акустического сигнала. Описан сравнительный анализ рассчитанных показателей эффективности применения различных методов распознавания пользователей компьютерной информационной системы на основе нейронных сетей.

Ключевые слова: биометрическая аутентификация, голос, датасет, изображение лица, компьютерная информационная система, конкатенация, нейронная сеть.

Введение

Инновационные разработки и технологии, которые позволяют повысить точность процедур идентификации и аутентификации, интенсивно прогрессируют. К числу наиболее эффективных инструментов идентификации и аутентификации следует отнести архитектуры нейронных сетей [1,2]. В работе в процессе обучения при распознавании пользователей компьютерных информационных систем используются такие подходы, как трансферное обучение и конкатенация (слияние) нейронных сетей [3]. Ранее данные подходы были исследованы и описаны в трудах Белова Ю.С., Нассера Н.М. и Солеймани С. [4].

Целью работы является повышение точности процессов распознавания пользователей по извлеченным признакам изображения лица и голоса посредством применения архитектур нейронных сетей на основе трансферного обучения и конкатенации (слияния). Результатом исследования является разработанная система биометрической аутентификации пользователей компьютерной информационной системы. В подсистеме биометрической системы распознавания пользователей по изображению лица применяется сверточная нейронная сеть с применением обучения на

авторском наборе данных (датасет) DataSet 600, а при распознавании по голосу – архитектура нейронной сети Wav2vec с набором данных VoxCeleb. Выбор приведенных датасетов обусловлен наибольшей долей корректно распознанных образов согласно полученным сравнительным показателям проведенных экспериментов с различными архитектурами нейронных сетей на разных наборах обучающих выборок. На основе вышеперечисленных архитектур была разработана и реализована конкатенированная нейронная сеть системы биометрической аутентификации пользователей компьютерной информационной системы.

1. Распознавание пользователей по изображению лица

Процесс распознавания пользователей по изображению лица в режиме онлайн является неотъемлемой частью подсистемы программного обеспечения биометрической аутентификации. Для обучения нейронной сети необходим входной вектор, который состоит из основных биометрических признаков (доминантные точки лица), к которым относятся локальные бинарные шаблоны и гистограммы направленных градиентов [5].

Локальные бинарные шаблоны вычисляются по следующей формуле [5]:

$$LBP_{Y,R} = \sum_{p=0}^{Y-1} s(g_p - g_c) 2^p$$

где $s(x)$ – пороговая функция, g_p – значение интенсивности p -ого пикселя, g_c – значение интенсивности центрального пикселя, p – номер пикселя, Y – окрестность с пикселями, R – радиус, $p = 0, \dots, Y - 1$.

Контур, состоящий из точек изображения лица, имеет приближенное значение величины градиента, которое возможно рассчитать на основе полученных приближенных значений производных [5]:

$$G = \sqrt{G_x^2 + G_y^2}$$

где G – величина градиента, G_x и G_y – координаты изображения.

При формировании входного вектора необходимо применить морфологические преобразования на основе фильтрации. Гистограмма направленных градиентов и ее дескриптор является конкатенированным вектором, состоящим из элементов нормализованных гистограмм ячеек [6].

2. Распознавание пользователей по голосу

Для извлечения биометрических признаков по голосу необходимо разбить входной акустический сигнал на временные окна с шагом смещения. Для каждого полученного кадра применяются следующие преобразования:

— Предварительная фильтрация с конечной импульсной характеристикой (КИХ-фильтр) [6]:

$$y_t = x_t - b \times x_{t-1}$$

где y_t – акустический сигнал после фильтрации, x_t – входной акустический сигнал, t – количество кадров, b – коэффициент фильтрации.

— Дискретное преобразование Фурье (ДПФ) [6]:

$$F_k = \sum_{t=0}^{T-1} w_t \times y_t e^{\frac{-2\pi i}{T} kt}$$

где T – отсчеты в кадре, w_t – весовая оконная функция, k – индекс частоты.

— Весовая оконная функция (окно Хэмминга и окно Ханна) применяется с целью уменьшения краевых эффектов, возникающих в результате разбиения сигнала на кадры [7]:

$$w_t^{hamm} = 0.54 - 0.46 \cos\left(\frac{2\pi t}{T-1}\right)$$

$$w_t^{hann} = 0.5 \left(1 - \cos\left(\frac{2\pi t}{T-1}\right)\right)$$

— Дискретное косинусное преобразование для значений энергий фильтров E_s [7]:

$$C_l = \sum_{s=0}^{M-1} E_s \cos \left(l \left(s + \frac{1}{2} \right) \frac{\pi}{M} \right)$$

где C_l – коэффициент под номером l , M – количество фильтров.

3. Конкатенированная нейронная сеть системы биометрической аутентификации пользователей компьютерной информационной системы

Для обучения нейронных сетей и принятия решения авторами была разработана конкатенированная нейронная сеть, которая представлена на рис. 1.

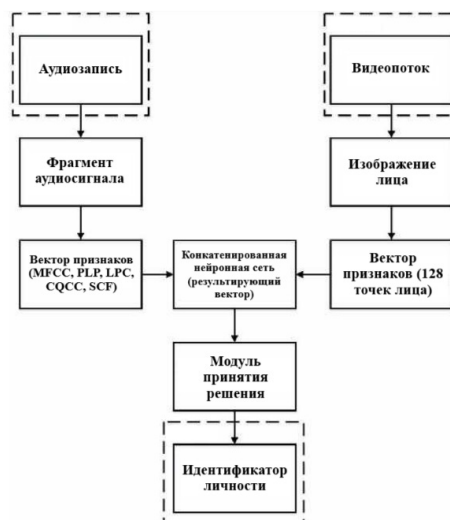


Рис. 1. – Схема биометрической системы аутентификации пользователей компьютерной информационной системы

В результате анализа архитектур сверточных нейронных сетей биометрической аутентификации пользователей компьютерной информационной системы создаются два вектора биометрических образов

одинаковой размерности. Конкатенационная нейронная сеть представляет собой объединение векторов в результирующий вектор размера 512 (рис. 2). При данном способе конкатенации биометрических образов влияние каждой подсистемы на итоговый результат распознавания является равнозначным [8].

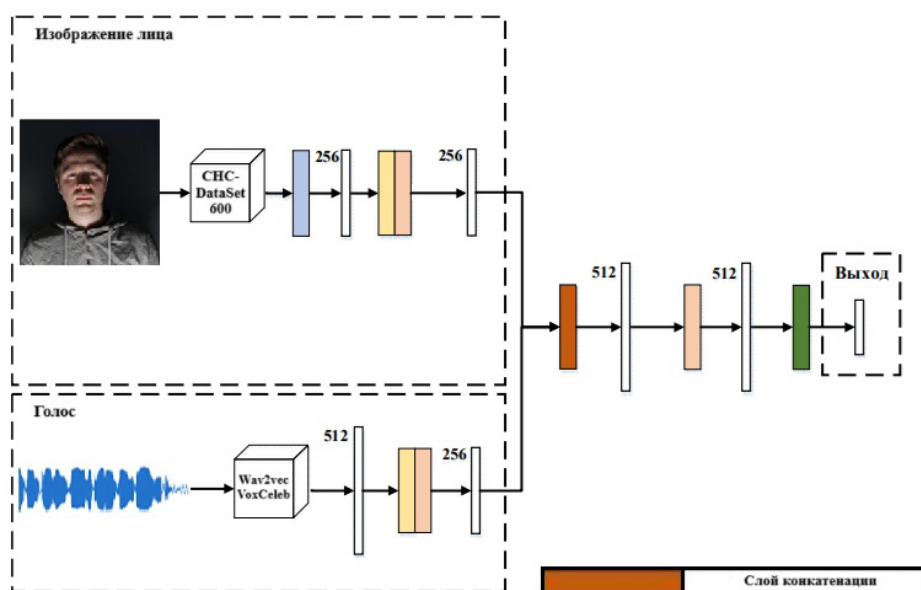


Рис. 2. – Конкатенированная нейронная сеть системы биометрической аутентификации пользователей компьютерной информационной системы

4. Показатели биометрической системы распознавания пользователей по извлекаемым признакам

Оценка эффективности и сравнение различных подходов при распознавании пользователей по биометрическим признакам рассчитаны на основе показателей ошибок первого и второго рода, которые приведены в таблицах № 1-3. Ошибка первого рода вычисляется по формуле [9]:

$$FRR = \frac{\alpha}{N},$$

где FRR – значение ошибки первого рода,

a – показатели некорректного распознавания авторизованных пользователей системы,

N – общее количество проведенных экспериментов.

Ошибка второго рода вычисляется по формуле [10]:

$$FAR = \frac{b}{N},$$

где FAR – значение ошибки второго рода,

b – показатели некорректного распознавания неавторизованных пользователей системы,

N – общее количество проведенных экспериментов.

Вычисления показателей эффективности распознавания пользователей компьютерной информационной системы по аудиозаписям (на основе ошибок первого и второго рода) отражены в таблице 1.

Таблица 1

Сопоставительные данные по эффективности методов распознавания пользователей компьютерной информационной системы по голосу

Краткое описание метода	Ошибка первого рода (FRR, %)	Ошибка второго рода (FAR, %)
Многослойный персептрон на основе извлеченных акустических признаков MFCC, LPC и PLP	1,2	3
Многослойный персептрон на основе извлеченных акустических признаков MFCC, LPC, PLP, CQCC и SCF	0,7	0,3
Сверточная нейронная сеть Wav2vec на наборе данных VoxCeleb	0,11	<0,1

Согласно приведенным результатам в таблице 1 наиболее оптимальной является архитектура сверточной нейронной сети Wav2vec на наборе данных VoxCeleb при распознавании пользователей компьютерной информационной системы по аудиозаписям.

Показатели эффективности распознавания пользователей компьютерной информационной системы по аудиозаписям приведены в таблице 2.

Таблица 2

Сопоставительные данные по эффективности методов распознавания пользователей компьютерной информационной системы по изображению лица

Краткое описание метода	Ошибка первого рода (FRR, %)	Ошибка второго рода (FAR, %)
Многослойный персептрон	3,3	8
Рекуррентная искусственная нейронная сеть	2,37	2,37
Сверточная нейронная сеть на наборе данных Yalefaces	2,42	2,42
Сверточная нейронная сеть на наборе данных CelebFaces	10,5	10,5
Скрытые марковские процессы, ИНС	2	2
Сверточная нейронная сеть на наборе данных DataSet 600	1,16	0,17
Сверточная нейронная сеть с предобработкой изображений (на основе фильтра Калмана) на наборе данных DataSet 600	0,83	<0,1

Опираясь на полученные показатели ошибок первого и второго рода в таблице 2, сверточная нейронная сеть с набором данных DataSet 600 на основе предобработки изображений посредством фильтра Калмана является

наиболее эффективной архитектурой при распознавании пользователей компьютерной информационной системы по голосу.

В таблице 3 приведены показатели ошибок первого и второго рода конкатенированной архитектуры нейронной сети с применением сверточной нейронной сети на основе предобработки изображений (на основе фильтра Калмана) на наборе данных DataSet 600 и сверточной нейронной сети Wav2vec на наборе данных VoxCeleb.

Таблица 3

Сопоставительные данные по эффективности методов распознавания пользователей компьютерной информационной системы

Краткое описание метода	Ошибка первого рода (FRR, %)	Ошибка второго рода (FAR, %)
Конкатенация сверточной нейронной сети с предобработкой изображений (на основе фильтра Калмана) на наборе данных DataSet 600 и сверточной нейронной сети Wav2vec на наборе данных VoxCeleb	0,01	<0,1

Как показывает приведенный результат в таблице 3, предложенная авторами архитектура конкатенированной нейронной сети является оптимальной в задаче распознавания личности и по лицу, и по голосу.

Результаты

Проведены исследования и составлена архитектура конкатенированной нейронной сети. Для повышения точности распознавания пользователей компьютерной информационной системы предложен метод слияния сверточных нейронных сетей модулей распознавания личности биометрической аутентификации пользователей. Проведены экспериментальные работы системы биометрической аутентификации пользователей при распознавании личности по изображению лица и по голосу. В ходе исследования было выявлено, что конкатенированная

нейронная сеть является оптимальной архитектурой для принятия решения на основе нескольких биометрических подсистем (критериев). Авторами помимо показателей эффективности методов распознавания пользователей компьютерной информационной системы получен результат точности обучения (ассурасу) разработанной архитектуры, который составляет 97,33%; показатели потерь (loss) равны 3,17% при обучении в 100 эпох.

Литература

1. Петров А.И., Белов Ю.С. Трансферное обучение модели VGG-16 // Региональная научно-техническая конференция, 2023, Т.2, С.57-61.
2. Титова А.А., Козина А.В., Белов Ю.С. Модель пограничных блоков для распознавания объектов // Научное обозрение. Технические науки, 2021, № 3, С. 61-65.
3. Soleymani S., Dabouei A., Kazemi H., Dawson J., Nasrabadi N.M. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification // 24th International Conference on Pattern Recognition (ICPR). – 2018. – pp. 3469-3476.
4. Soleymani S., Torfi A., Dawson J., Nasrabadi N.M. Generalized bilinear deep convolutional neural networks for multimodal biometric identification // 25th IEEE International Conference on Image Processing. – 2018. – pp. 763-767.
5. Todisco M., Delgado H., Evans N. A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients // Odyssey 2016: The Speaker and Language Recognition Workshop. Bilbao: ISCA SIG, 2016. pp. 283–290.
6. Исмаилова А.С., Лушников Н.Д. Комплексная биометрическая аутентификация пользователей информационной системы с применением нейронных сетей // Инженерный вестник Дона, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8961.

7. Щербань И.В., Доброходский В.В., Ефименко А.А. Online–программа аутентификации, основанная на оконном преобразовании Фурье речевых фраз пользователя // Символ науки, 2016, №1. URL: cyberleninka.ru/article/n/online-programma-autentifikatsii-osnovannaya-na-okonnom-preobrazovanii-furie-rechevyh-fraz-polzovatelya.
8. Девицына С.Н., Елецкая Т.А., Балабанова Т.Н., Гахова Н.Н. Разработка интеллектуальной системы биометрической идентификации пользователя // Научные ведомости. Серия: Экономика. Информатика, 2019, № 1, С.148-160.
9. Осовский С. Нейронные сети для обработки информации. М.: Телеком, 2017. 448 с.
10. Lawrence R. Rabiner, Ronald W. Schafer. Theory and Applications of Digital Speech Processing. Prentice Hall, 2010. 1056 p.

References

1. Petrov A.I., Belov Yu.S. Regional'naya nauchno-tekhnicheskaya konferentsiya, 2023, T.2, pp.57-61.
2. Titova A.A., Kozina A.V., Belov Yu.S. Nauchnoye obozreniye. Tekhnicheskkiye nauki, 2021, № 3, pp. 61-65.
3. Soleymani S., Dabouei A., Kazemi H., Dawson J., Nasrabadi N.M. 24th International Conference on Pattern Recognition (ICPR), 2018. pp. 3469-3476.
4. Soleymani S., Torfi A., Dawson J., Nasrabadi N.M. 25th IEEE International Conference on Image Processing, 2018. pp. 763-767.
5. Todisco M., Delgado H., Evans N. Odyssey 2016: The Speaker and Language Recognition Workshop. Bilbao: ISCA SIG, 2016. pp. 283–290.
6. Ismagilova A.S., Lushnikov N.D. Inzhenernyj vestnik Dona, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8961.
7. Shcherban' I.V., Dobrokhodskiy V.V., Yefimenko A.A. Simvol nauki, 2016, №1 URL: cyberleninka.ru/article/n/online-programma-autentifikatsii-osnovannaya-na-okonnom-preobrazovanii-furie-rechevyh-fraz-polzovatelya.



8. Devitsyna S.N., Yeletskaia T.A., Balabanova T.N., Gakhova N.N. Nauchnyye vedomosti. Seriya: Ekonomika. Informatika, 2019, № 1, pp. 148-160.
9. Osovskiy S. Neyronnyye seti dlya obrabotki informatsii. M.: Telekom, 2017. 448 p.
10. Lawrence R. Rabiner, Ronald W. Schafer. Theory and Applications of Digital Speech Processing. Prentice Hall, 2010. 1056 p.

Дата поступления: 7.03.2025

Дата публикации: 25.04.2025