

DOI: 10.15514/ISPRAS-2022-34(6)-8



Исследование возможности применения нейронных сетей для восстановления изображения лица в системах распознавания

Е.И. Маркин, ORCID: 0000-0003-4848-8757 <evgeniymarkin1@gmail.com>

В.В. Зупарова, ORCID: 0000-0002-7903-083X <zuparova_vv@mail.ru>

А.И. Мартышкин, ORCID: 0000-0002-3358-4394 <mai@penzgtu.ru>

Пензенский государственный технологический университет,
440039, Россия, г. Пенза, проезд Байдукова/ул. Гагарина, д. 1а/11

Аннотация. Идентификация человека на цифровом изображении с помощью компьютерного зрения является важнейшим аспектом этой области. Наличие внешних объектов, таких как медицинские маски, которые закрывают часть лица, может резко снизить точность распознавания и увеличить ошибки от 5% до 50% в зависимости от алгоритма. В данной статье исследуется использование нейронных сетей, в частности генеративной состязательной сети (*GAN*), для решения задачи восстановления изображения лица, закрытого медицинской маской, для повышения точности распознавания лица.

Ключевые слова: компьютерное зрение; нейронные сети, генеративно-состязательные сети, идентификации лиц людей, реконструирование цифрового изображения.

Для цитирования: Маркин Е.И., Зупарова В.В., Мартышкин А.И. Исследование возможности применения нейронных сетей для восстановления изображения лица в системах распознавания. Труды ИСП РАН, том 34, вып. 6, 2022 г., стр. 117-126. DOI: 10.15514/ISPRAS-2022-34(6)-8

Exploring the application of neural networks for facial image reconstruction in recognition systems

E.I. Markin, ORCID: 0000-0003-4848-8757 <evgeniymarkin1@gmail.com>

V.V. Zuparova, ORCID: 0000-0002-7903-083X <zuparova_vv@mail.ru>

A.I. Martyshekin, ORCID: 0000-0002-3358-4394 <mai@penzgtu.ru>

Penza State Technological University,
1a/11, Baidukova Passage/Gagarin st., Penza, 440039, Russia.

Abstract. Identifying a person in a digital image using computer vision is a crucial aspect of this field. The presence of external objects, such as medical masks that cover part of the face, can drastically reduce recognition accuracy and increase errors from 5% to 50%, depending on the algorithm. This paper investigates the use of neural networks, in particular the generative adversarial network (*GAN*), to solve the problem of reconstructing an image of a face covered by a medical mask to improve face recognition accuracy.

Keywords: computer vision; neural networks, generative adversarial networks, human face identification, digital image reconstruction.

For citation: Markin E.I., Zuparova V.V., Martyshekin A.I. Exploring the application of neural networks for facial image reconstruction in recognition systems. Trudy ISP RAN/Proc. ISP RAS, vol. 34, issue 6, 2022. pp. 117-126 (in Russian). DOI: 10.15514/ISPRAS-2022-34(6)-8

1. Введение

Использование передовых технологий и искусственного интеллекта произвело революцию во многих отраслях промышленности, повысив эффективность, точность и производительность. Компьютерное зрение – одна из таких областей, в которой в последние годы наблюдается значительный рост, с разработкой передовых алгоритмов и моделей, позволяющих анализировать и интерпретировать изображения и видео. Идентификация человеческих лиц по фотографическим изображениям – важнейший аспект компьютерного зрения, который находит применение в системах безопасности, контроля доступа, проверки данных и пользователей устройств [1].

Однако наличие посторонних объектов, загромождающих или скрывающих части лица, может сильно повлиять на точность распознавания и повысить коэффициент ошибок на значительную величину [2]. Широкое использование средств индивидуальной защиты во время пандемии коронавируса 2019-*nCoV* привело к повсеместным сбоям в системах биометрической идентификации с масками на лице, что подчеркивает необходимость инновационных решений для решения проблемы идентификации лиц со скрытыми участками [3, 4].

В данной работе предлагается нейросетевой подход с использованием глубоких сверточных генеративно-состязательных сетей (*DCGAN*) [5] для восстановления скрытых частей лиц и улучшения качества и детализации изображений для повышения точности распознавания. Использование *DCGAN* для восстановления рисунка лица может значительно повысить точность идентификации личности и снизить коэффициент ошибок, обеспечивая надежное решение проблемы идентификации лиц со скрытыми участками лица.

2. Описание метода решения задачи

По мере развития технологий и разработки новых инноваций, область генерации данных и компьютерного зрения в последние годы претерпела значительные изменения. Одним из наиболее значительных прорывов в этой области стала разработка генеративных состязательных сетей (*GAN*) (рис. 1), которые представляют собой тип нейронных сетей, доказавших свою высокую эффективность в создании фотореалистичных изображений [6].



Рис. 1. Архитектура генеративной состязательной сети

Fig. 1. The architecture of the generative adversarial network

Основная концепция *GAN* основана на теории игр и включает две модели, генератор и критик, которые конкурируют друг с другом, чтобы сделать друг друга сильнее [7-9]. Генератор выдает синтетические образцы, которые должны выглядеть как можно ближе к реальным данным, а критик, также известный как дискриминатор, призван определить, является ли образец настоящим или поддельным. Это соревнование между генератором и

критиком приводит к тому, что обе модели со временем совершенствуются и становятся более эффективными в своих задачах.

Дискриминатор (D) должен с высокой точностью отличать реальные данные от поддельных, в то время как генератор (G) стремится производить образцы, которые D трудно отличить от реальных данных. Чтобы сбалансировать эти две цели, между D и G разыгрывается минимаксная игра, в которой функция потерь оптимизируется следующим образом:

$$\min_G \max_D L(D, G) = E_{x \sim p_r(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \\ = E_{x \sim p_r(x)} [\log D(x)] + E_{x \sim p_g(x)} [\log (1 - D(x))]. \quad (1)$$

$E_{x \sim p_r(x)} [\log D(x)]$ не влияет на G во время обновления градиентного спуска.

Чтобы определить наилучшее значение дискриминатора, нам нужно максимизировать функцию потерь:

$$L(D, G) = \int_x (p_r(x) \log(D(x)) + p_g(x) \log(1 - D(x))) dx. \quad (2)$$

Оптимальное значение $D(x)$ максимизирует $L(G, D)$, его можно представить как:

$$\tilde{x} = D(x), A = p_r(x), B = p_g(x), \quad (3)$$

и оно может быть рассчитано как:

$$f(\tilde{x}) = A \log \tilde{x} + B \log(1 - \tilde{x}), \quad (4) \\ \frac{df(\tilde{x})}{d\tilde{x}} = A \frac{1}{\ln 10} \frac{1}{\tilde{x}} - B \frac{1}{\ln 10} \frac{1}{1 - \tilde{x}} = \frac{1}{\ln 10} \left(A \frac{1}{\tilde{x}} - B \frac{1}{1 - \tilde{x}} \right) = \frac{1}{\ln 10} \frac{A - (A + B)\tilde{x}}{\tilde{x}(1 - \tilde{x})}.$$

Таким образом, установив $\frac{df(\tilde{x})}{d\tilde{x}} = 0$, мы получаем наилучшее значение дискриминатора:

$$D^*(x) = \tilde{x}^* = \frac{A}{A + B} = \frac{p_r(x)}{p_r(x) + p_g(x)} \in [0, 1]. \quad (5)$$

Когда генератор оптимально обучен, вероятности сгенерированных данных (p_g) и реальных данных (p_r) сходятся, и дискриминатор (D^*) должен давать значение $1/2$. Это приводит к функции потерь:

$$L(G, D) = \int_x (p_r(x) \log(D^*(x)) + p_g(x) \log(1 - D^*(x))) dx \\ = \log \frac{1}{2} \int_x p_r(x) dx + \log \frac{1}{2} \int_x p_g(x) dx. \quad (6)$$

На основе 6 можно рассчитать дивергенцию Йенсена-Шеннона между двумя распределениями, используя формулу:

$$D_{JS}(p_r \| p_g) = \frac{1}{2} D_{KL}(p_r \| \frac{p_r + p_g}{2}) + \frac{1}{2} D_{KL}(p_g \| \frac{p_r + p_g}{2}) = \\ = \frac{1}{2} \left(\log 2 + \int_x p_r(x) \log \frac{p_r(x)}{p_r + p_g(x)} dx \right) + \frac{1}{2} \left(\log 2 + \int_x p_g(x) \log \frac{p_g(x)}{p_r + p_g(x)} dx \right) \\ = \frac{1}{2} (\log 4 + L(G, D^*)) \quad (7)$$

и формулу

$$L(G, D^*) = 2D_{JS}(p_r \| p_g) - 2 \log 2. \quad (8)$$

Процесс обучения генеративно-сопоставительной сети (GAN) включает использование функции потерь, которая определяет сходство между сгенерированным распределением данных (p_g) и распределением реальной выборки (p_r). Одной из часто используемых функций потерь является дивергенция Дженсена-Шеннона, которая вычисляет разницу между двумя распределениями, когда дискриминатор оптимален. Оптимальный генератор (G^*), который может идеально воспроизвести реальное распределение данных, приводит к минимальным потерям $L(G^*, D^*) = -2 \log 2$ [14, 15].

Существует несколько разновидностей GAN, которые предназначены для различных приложений и контекстов. Например, при полунаблюдаемом обучении дискриминатор обновляется для присвоения реальных меток классам с 1 по K-1 и поддельной метки классу K, при этом генератор пытается обмануть дискриминатор, чтобы присвоить меньшую метку [16].

В данном исследовании для улучшения обучения DCGAN был использован модифицированный метод одностороннего сглаживания меток. Метод сочетает в себе два метода оценки сходства изображений, а именно функцию потерь и индекс структурного сходства (SSIM) [17]. Функция потерь, такая как средняя квадратичная ошибка или средняя абсолютная ошибка, измеряет расхождение между эталонным и сгенерированным изображениями. Между тем, SSIM рассчитывает численное представление сходства между двумя изображениями, используя три параметра: яркость, контраст и структуру. Итоговая оценка варьируется от 0 до 1, где оценка 1 означает, что изображения полностью схожи.

В процессе проектирования генеративно-сопоставительной нейронной сети возникает компромисс между производительностью модели и ее размером. Чем глубже сеть и чем больше в ней фильтров, тем большее количество параметров она будет содержать [18]. Например, слой свертки с размерностью [8, 8, 256] содержит примерно 0,59 миллиона параметров, а слой с размерностью [4, 4, 512] – 2,3 миллиона параметров. Это приводит к увеличению времени обучения. Однако, регулируя глубину и количество фильтров, можно экспериментировать и найти конфигурацию, которая обеспечивает оптимальный баланс между производительностью и размером модели.

После проведения нескольких испытаний была определена следующая оптимальная конфигурация:

- размеры фильтров: [64, 128, 128, 128, 256, 256, 512];
- размер ядра: [7, 7, 7, 7, 3, 3, 3];
- размер входного изображения: [256, 256, 3];
- размер пакета: 12.

3. Обучение нейронной сети

Для получения достаточного количества обучающих данных для генеративной сопоставительной нейронной сети необходимы пары изображений человека с маской и без нее. Однако сбор таких данных вручную может быть затруднен из-за необходимости сопоставления входных и выходных изображений одного и того же человека [19]. Для решения этой проблемы была разработана система, позволяющая генерировать базу данных лиц с масками, используя существующие базы данных изображений лиц.



Рис. 2. Блок-схема алгоритма добавления маски на изображение лица человека
Fig. 2. Block diagram of the algorithm for adding a mask to the human face image

В качестве основной базы данных использовалась база данных *Labeled Faces in the Wild* Массачусетского университета, которая содержит более 5 000 изображений человеческих лиц. Для точного размещения маски на лицах использовался предварительно обученный детектор ориентиров на лицах *dlib*. Этот инструмент не только определяет ориентиры, но и вычисляет наклон лица, что крайне важно для правильного выравнивания изображения маски.

Наклон лица учитывается при выборе наиболее подходящего шаблона маски из набора имеющихся масок. Затем шаблон маски преобразуется на основе шести ключевых точек нижней части лица для достижения наилучшего прилегания. Процесс изображен на блок-схеме, показанный на рис. 2, а пример полученного изображения с маской и без маски показан на рис. 3.



Рис. 3. Пример сгенерированных изображений лиц в маске
Fig.3. Example of generated images of faces in the mask

4. Практические результаты

Нейронная сеть обучалась в течение 50 эпох, используя 500 случайно выбранных изображений лиц с масками из собранной базы данных в каждой эпохе. Ход обучения отслеживался с помощью функции потерь и индекса структурного сходства (SSIM).

Выбор подходящей функции потерь имеет решающее значение для эффективности модели. Для решения этой задачи была использована комбинация двух функций потерь - среднеквадратичной ошибки (MSE) и средней абсолютной ошибки (MAE). Эти функции оценивают разницу в пикселях между изображением, полученным с помощью модели, и истинным изображением из набора данных до применения маски.

Чтобы модель могла создать реалистичное изображение лица с маской, необходимо, чтобы она понимала особенности, скрытые маской, такие как рот и нос, и, возможно, даже эмоции, отображаемые на видимых частях лица, таких как глаза. Комбинация функций потерь

позволяет модели научиться обобщать любое лицо, а не только те, которые есть в наборе данных для обучения, даже если она не может идеально соответствовать каждому пикселю. SSIM, оценивает сходство между двумя изображениями. Он выдает числовой балл (от 0 до 1) на основе взвешенной комбинации яркости, контраста и структуры, а не просто подсчитывает количество совпадающих пикселей.

В табл. 1 представлена статистика производительности нейронной сети на этапе обучения для реконструкции лиц. В таблице показано значение функции потерь и точность сети во время обучения и тестирования.

Табл. 1. Статистика производительности нейронной сети
Table 1. Neural network performance statistics

№ эпохи	Loss (обучение)	SSIM (обучение)	Loss (тестовый)	SSIM (тестовый)
1	0,3956	0,7597	0,3686	0,7255
2	0,2877	0,8067	0,2792	0,7157
3	0,1968	0,8084	0,1957	0,7566
4	0,1570	0,8252	0,1760	0,7677
5	0,1542	0,8415	0,1794	0,8104
...
46	0,0625	0,9304	0,0727	0,9496
47	0,0628	0,9243	0,0764	0,9215
48	0,0605	0,9221	0,0738	0,8960
49	0,0600	0,9064	0,0721	0,9117
50	0,0606	0,9308	0,0717	0,9546

На рис. 4-5 представлены результаты обучения нейронной сети.

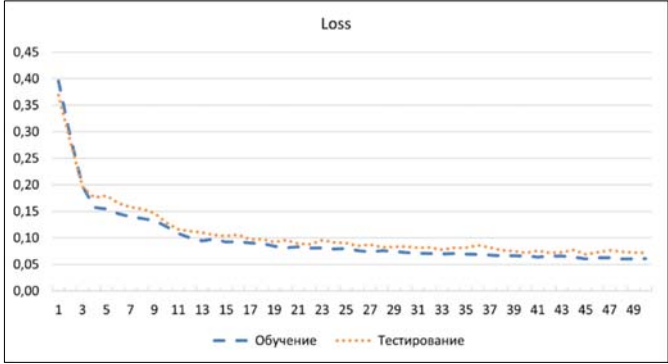


Рис. 4. График изменения функции потерь во время обучения
Fig. 4. Graph of the change in the loss function during training

На рис. 6 представлены результаты, полученные в процессе обучения. На рисунке представлены три изображения: оригинальное изображение слева, изображение с маской в центре и восстановленные скрытые части лица справа.

Как показывают представленные данные, предложенная нейронная сеть демонстрирует замечательные результаты на тестовых наборах данных. Сеть демонстрирует способность к обобщению и, обладает достаточным уровнем распознавания эмоций, чтобы генерировать

улыбающиеся или грустные лица. Тем не менее, все еще есть возможности для дальнейшего совершенствования.

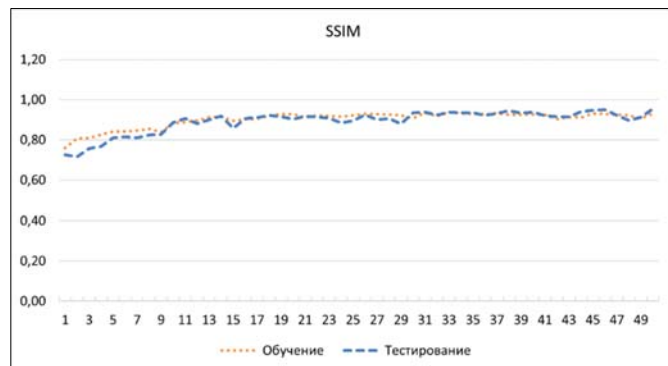


Рис. 5. График изменения функции схожести изображений во время обучения
Fig. 5. The graph of the change in the image similarity function during training

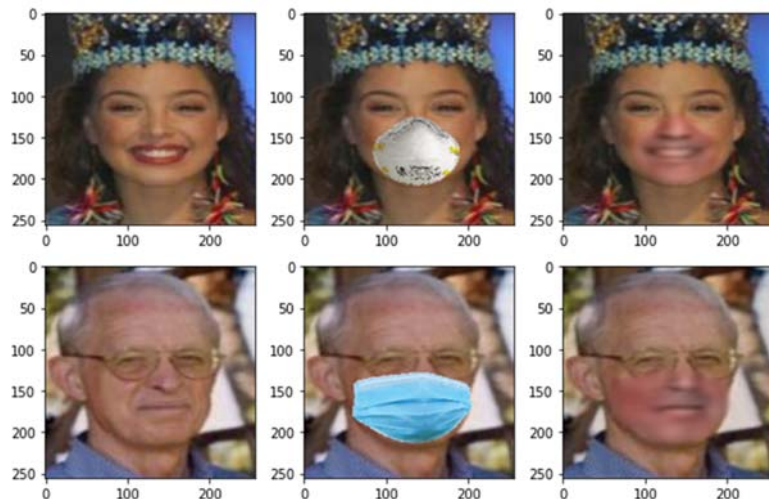


Рис. 6. Результаты восстановления изображения лица человека обученной нейронной сетью
Fig. 6. Results of reconstruction of a human face image by a trained neural network

5. Заключение

В заключение следует отметить, что результаты обучения *DCGAN* для восстановления скрытых участков лица на двух оцениваемых параметрах показало высокие результаты. Использование функции потерь, основанной на разнице пикселей, и метрики сходства изображений, предложенной Вангом, позволило получить значение точности 6,06% при обучении и 7,17% на тестовом наборе данных для функции потерь и 93,08% при обучении и 95,46% на тестовом наборе данных для метрики *SSIM*. Эти результаты показывают, что реконструированные изображения скрытых областей лица очень похожи по деталям и очертаниям на оригинальные изображения до применения масок, что делает их полезными для идентификации лиц в масках

Данная исследовательская работа подчеркивает важность разработки инновационных решений для решения проблем, возникающих при идентификации лиц со скрытыми областями, и потенциал использования глубоких сверточных генеративно-состязательных сетей для восстановления скрытых частей лиц и повышения точности распознавания. Результаты данного исследования имеют значительный вклад в области компьютерного зрения и его различных применений, подчеркивая необходимость постоянных инвестиций в исследования и разработки для продвижения передового опыта в этой области.

Список литературы / References

- [1] Goldstein A.J., Harmon L.D., Lesk A.B. Identification of human faces. Proceedings of the IEEE, vol. 59, issue 5, 1971, pp. 748–760.
- [2] Ngan M.L., Grother P.J., Hanaoka K.K. Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms. NIST Interagency/Internal Report (NISTIR) no. 8331, 2020, 83 p.
- [3] Блейхут Р.Э. Быстрые алгоритмы цифровой обработки сигналов. М., Мир, 1989 г., 448 стр. / Blahut R.E. Fast Algorithms for Digital Signal Processing. Addison-Wesley, 1985, 455 p.
- [4] Оппенгейм А., Шафер Р. Цифровая обработка сигналов. М., Техносфера, 2019 г., 1048 стр. / Oppenheim A.V., Schaffer R.W. Discrete-Time Signal Processing. Pearson, 2009, 1144 p.
- [5] Бершадская Е.Г., Маркин Е.И., Мартышкин А.И. Методы идентификации личности по изображению лица. XXI век итоги прошлого и проблемы настоящего плюс, том 9, вып. 1, стр. 49–53 / Bershadskaya E.G., Markin E.I., Martyshekin A.I. Methods for personal image identification. XXI Century: Resumes of the Past and Challenges of the Present plus, vol. 9, issue 1, pp. 49–53 (in Russian).
- [6] Choi J., Han B. MCL-GAN: Generative Adversarial Networks with Multiple Specialized Discriminators. arXiv2107.07260, 2021, 20 p.
- [7] Adler J., Lunz S. Banach wasserstein GAN. In Proc. of the 32nd International Conference on Neural Information Processing Systems, 2018, pp. 6755–6764.
- [8] Arjovsky M., Bottou L. Towards principled methods for training generative adversarial networks. arXiv1701.04862, 2017, 17 p.
- [9] Маркин Е.И., Мартышкин А.И., Зупарова В.В. Анализ возможностей нейронных сетей для генерации фотореалистичных изображений. Современные информационные технологии, вып. 33, 2021, стр. 30–34 / Маркин Е.И., Мартышкин А.И., Zuparova V.V. Analysis of the capabilities of neural networks to generate photorealistic images. Modern Information Technology, vol. 33, 2021, pp. 30–34 (in Russian).
- [10] Chi N.N.K. Active shape models their training and application. Bachelor' Thesis. International University HCMC, Vietnam, 2011.
- [11] Gulrajani I., Ahmed F. et al. Improved training of wasserstein GANs. In Proc. of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 5769–5779.
- [12] Salimans T., Goodfellow I. et al. Improved techniques for training GANs. In Proc. of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2234–2242.
- [13] Huszar F. How (not) to train your generative model: Scheduled sampling, likelihood, adversary? arXiv1511.05101, 2015, 9p.
- [14] Zhang W., Liu Y. et al. RankSRGAN: Super Resolution Generative Adversarial Networks with Learning to Rank. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, 2022, pp. 7149–7166.
- [15] Zhou E., Fan H. et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In Proc. of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 386–391.
- [16] Wei J., Liu M. et al. DuelGAN: A Duel Between Two Discriminators Stabilizes the GAN Training. Lecture Notes in Computer Science, vol. 13683, 2022, pp. 290–317.
- [17] Wang Z., Bovik A.C. et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, vol. 13, issue 4, 2004, pp. 600–612.
- [18] Yamashita R., Nishio M. et al. Convolutional neural networks: an overview and application in radiology. Insights into Imaging, vol. 9, issue 4, 2018, pp. 611–629.
- [19] Heo J. Three-dimensional generic elastic models for two-dimensional pose synthesis and face recognition Proquest, Umi Dissertation Publishing, 2011, 154 p.

Информация об авторах / Information about authors

Евгений Игоревич МАРКИН – кандидат технических наук, ассистент кафедры «Программирование». Сфера научных интересов: распознавание образов, обработка цифровых изображений, системы компьютерного зрения, машинное обучение и нейронные сети.

Evgeny Igorevich MARKIN – Candidate of Technical Sciences, Assistant of the Department "Programming". Research interests: pattern recognition, digital image processing, computer vision systems, machine learning and neural networks.

Валентина Владимировна ЗУПАРОВА – аспирант кафедры «Программирование». Сфера научных интересов: обработка данных, распознавание образов, распределенные вычислительные системы.

Valentina Vladimirovna ZUPAROVA – Postgraduate Student of the Department of Programming. Research interests: data processing, pattern recognition, distributed computing systems.

Алексей Иванович МАРТЫШКИН – кандидат технических наук, доцент, заведующий кафедрой «Программирование». Сфера научных интересов: обработка данных, моделирование вычислительных систем, исследование высокопроизводительных систем, распределенные вычислительные системы.

Alexey Ivanovich MARTYSHKIN – Candidate of Technical Sciences, Associate Professor, Head of the Department "Programming". Research interests: data processing, modeling of computing systems, research of high-performance systems, distributed computing systems.