

**ПРИМЕНЕНИЕ МЕТОДОВ СИНТЕЗА ОБУЧАЮЩИХ ДАННЫХ
ДЛЯ РАСПОЗНАВАНИЯ ЧАСТИЧНО СКРЫТЫХ ЛИЦ НА ИЗОБРАЖЕНИЯХ**

М. А. ЛЕТЕНКОВ, Р. Н. ЯКОВЛЕВ*, М. В. МАРКИТАНТОВ,
Д. А. РЮМИН, А. А. КАРПОВ

Санкт-Петербургский федеральный исследовательский центр Российской академии наук,
Санкт-Петербург, Россия
*iakovlev.r@mail.ru,

Аннотация. Для решения проблемы автоматического распознавания лиц людей, использующих такие средства индивидуальной защиты, как медицинская маска, предложен и апробирован новый подход, основанный на применении методов генерации синтетических изображений частично скрытых лиц и модели распознавания лиц ArcFace. Предложена стратегия формирования обучающих наборов данных и получен ряд соответствующих моделей распознавания. Проведена серия экспериментов, направленных на оценку качества предсказаний полученного решения, и установлена зависимость между результирующим качеством предсказаний, реализуемых моделями распознавания, и объемом синтетических изображений в обучающих наборах данных. Согласно результатам экспериментальных исследований, нейросетевые модели, дообученные на наборах данных, в которых объем искусственно синтезированных изображений составляет 40—60 %, демонстрируют более высокие значения показателя точности распознавания, выше 87 % по количественной метрике AAc (Averaged Accuracy). Использование предложенного подхода позволяет значительно улучшить качество распознавания частично скрытых лиц по сравнению с базовым подходом.

Ключевые слова: распознавание лиц, нейросетевые модели распознавания, ArcFace, BRAVE-MASKS, генерация синтетических изображений, средства индивидуальной защиты, глубокое обучение

Благодарности: исследование выполнено за счет Российского фонда фундаментальных исследований (проект № 20-04-60529-вирусы), а также частично в рамках ведущей научной школы (грант № НШ-17.2022.1.6)

Ссылка для цитирования: Летенков М. А., Яковлев Р. Н., Маркитантов М. В., Рюмин Д. А., Карпов А. А. Применение методов синтеза обучающих данных для распознавания частично скрытых лиц на изображениях // Изв. вузов. Приборостроение. 2022. Т. 65, № 11. С. 842—850. DOI: 10.17586/0021-3454-2022-65-11-842-850.

**APPLICATION OF TRAINING DATA SYNTHESIS METHODS
FOR RECOGNITION OF PARTIALLY HIDDEN FACES IN IMAGES**

M. A. Letenkov, R. N. Iakovlev*, M. V. Markitantov, D. A. Ryumin, A. A. Karpov

St. Petersburg Federal Research Center of the RAS,
St. Petersburg, Russia
*iakovlev.r@mail.ru,

Abstract. A new approach to solving the problem of automatic face recognition of people using personal protective equipment such as a medical mask has been proposed and tested. This approach is based on the use of methods of generating synthetic images of partially hidden faces and the face recognition model ArcFace. A strategy for training data sets formation is proposed and a number of corresponding recognition models are derived. A series of experiments aimed at assessing the quality of predictions of the obtained solution are carried out, and a relationship between the resulting quality of predictions implemented by recognition models and the volume of synthetic images in training datasets is established. According to the results of experimental studies, neural network models, further trained on datasets with volume of artificially synthesized images of 40-60%, demonstrate values of recognition accuracy above 87% on the AAc quantitative metric (Average Accuracy). Using the proposed approach makes it possible to significantly improve the quality of recognition of partially hidden faces compared to the basic approach.

Keywords: face recognition, neural network recognition models, ArcFace, BRAVE-MASKS, synthetic image generation, personal protective equipment, deep learning

Acknowledgments: the research was carried out at the expense of the Russian Foundation for Basic Research (project N 20-04-60529-viruses), as well as partially within the framework of a leading scientific school (grant N NS-17.2022.1.6).

For citation: Letenkov M. A., Iakovlev R. N., Markitantov M. V., Ryumin D. A., Karpov A. A. Application of training data synthesis methods for recognition of partially hidden faces in images. *Journal of Instrument Engineering*. 2022. Vol. 65, N 11. P. 842—850 (in Russian). DOI: 10.17586/0021-3454-2022-65-11-842-850.

Введение. В настоящее время существует множество подходов к решению задач детектирования и распознавания лиц на видеопоследовательностях или отдельных изображениях [1—4]. Однако большинство таких методов не универсальны и имеют высокий процент погрешности в условиях частичной окклюзии объекта съемки [5, 6], в том числе в случаях, когда лицо человека, использующего средства индивидуальной защиты (СИЗ), в частности медицинскую маску, частично скрыто. Принимая во внимание высокий уровень распространения коронавирусной инфекции COVID-19, а также установленные требования, связанные с необходимостью ношения СИЗ [7], следует отметить критическую потребность в разработке специализированных систем биометрической идентификации, способных к распознаванию в условиях, когда лицо человека частично скрыто. Ключевым условием разработки соответствующих систем идентификации является подготовка специализированных обучающих наборов данных, включающих изображения лиц людей, использующих маски (Masked Human Faces — МНФ) как средство индивидуальной защиты (далее — МНФ-изображение) [8, 9]. Однако формирование масштабных наборов данных с высокой долей МНФ-изображений в каждом конкретном случае представляет собой крайне трудоемкую и ресурсоемкую задачу [10—12].

Таким образом, настоящее исследование посвящено разработке подхода к распознаванию частично скрытых лиц, основанного на применении методов генерации синтетических (синтезированных) МНФ-изображений при формировании обучающих наборов данных.

Современное состояние исследований. Известные к настоящему времени подходы к распознаванию лиц предусматривают возможность построения биометрических систем идентификации людей по изображениям. При проектировании таких систем наиболее релевантными являются методы, основанные на применении моделей глубокого обучения [3, 6]. Среди современных моделей глубокого обучения, нацеленных на решение задачи идентификации лиц, наибольший интерес представляют Facenet [3], Sphereface [13], Cosface [14], Arcface [4] (далее — ArcFace). Модель глубокого обучения Facenet [3], превзошедшая более распространенные сверточные нейросети и победившая на конкурсе ImageNet 2014, основана на архитектуре GoogleNet [15]. В качестве базовой модели в рамках данной архитектуры используется нейросетевая модель Inception-Resenet-v1. Функция потерь, используемая в рамках модели Facenet, определена поэлементной суммой значений функции потерь triplet loss и функции L2-регуляризации. При оценке качества предсказаний, реализуемых моделью распознавания на наборе данных MegaFace [16], было получено значение метрики Accuracy, равное 70,49 %.

В отличие от модели Facenet, использующей евклидово пространство для размещения признаков векторов с возможностью их классификации путем расчета евклидова расстояния между ними, модель Sphereface [13], представленная в 2017 г., является сверточной нейросетью, включающей 64 слоя, в архитектуре которой задействованы residual-блоки. В качестве функции потерь используется angular softmax (A-Softmax), что позволяет модели классифицировать признаковые векторы путем анализа угловых расстояний между ними. Данный подход позволил превзойти результаты Facenet при оценке модели Sphereface на наборе данных MegaFace [16], где значение метрики Accuracy достигло 75,76 %.

Разработчики модели Cosface [14] предложили отойти от использования стандартных функций потерь, применяемых при решении задачи идентификации лиц. Был сформирован новый вид функции потерь, названный LMCL, получаемой посредством представления softmax loss как cosine loss путем применения L2-нормализации к признаковому и весовому векторам. Такой подход позволяет уменьшить значения радиальных отклонений на последнем этапе процесса идентификации, что положительно влияет на минимизацию внутриклассовых и максимизацию межклассовых расстояний. Cosface является результатом применения разработанной концепции функции потерь к 64-слойной сверточной нейросети Sphreface. Экспериментальные оценки, производимые на наборе данных MegaFace [16], показали достаточно высокий уровень предсказаний, реализуемых моделью Cosface: модели удалось достичь значения Accurasy в 82,72 %.

В контексте крупномасштабного распознавания лиц следует выделить исследование [4], посвященное нейросетевой модели распознавания ArcFace. В работе авторы решают задачу оптимизации функции потерь. Для достижения большей разделимости классов в евклидовом пространстве предложена функция Additive Angular Margin Loss. В качестве базовой модели использовалась сверточная нейросеть Sphreface. Модель ArcFace по сравнению с моделями Sphreface и Cosface обладает постоянным линейным угловым запасом (constant linear angular margin), что позволяет существенно повысить разделимость классов в признаковом пространстве. При тестировании модели ArcFace на наборе данных MegaFace [16] удалось достичь значения метрики Accurasy в 98,35 %, что является лучшим результатом среди обзриваемых моделей. Таким образом, исходя из особенностей рассмотренных моделей глубокого обучения при разработке подхода к распознаванию частично скрытых лиц наиболее предпочтительным представляется использование модели глубокого обучения ArcFace.

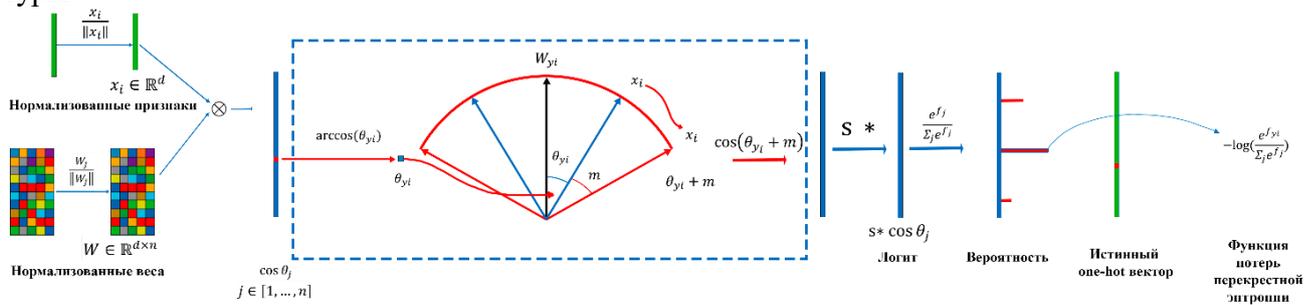
Для реализации предлагаемого подхода необходимо также определить наиболее релевантные методы генерации синтетических МНФ-изображений, которые могут быть использованы для формирования обучающих наборов данных и последующего тестирования полученного решения. В рамках предыдущего этапа работ коллективом авторов настоящей статьи проведен детальный анализ современного состояния исследований по данной проблематике и предложен собственный метод генерации синтетических изображений лиц людей в защитных масках, основанный на комбинированном использовании средств трехмерного моделирования объектов и нейросетевого метода детектирования трехмерных лицевых ориентиров на изображениях (3D-FAN) [17]. В рамках проведенного эксперимента разработанный метод продемонстрировал высокое и стабильное качество: доля корректно сгенерированных изображений лиц в масках по результатам применения разработанного метода к тестовому набору данных оказалась равной 95,9 %.

Таким образом, в рамках настоящего исследования в качестве метода генерации синтетических МНФ-изображений предлагается использовать авторский метод, представленный в работе [17].

Описание разработанного подхода. В соответствии с результатами проведенного анализа связанных методов и подходов для распознавания частично скрытых лиц в качестве исходной выбрана нейросетевая модель глубокого обучения ArcFace [4]. Обобщенный вид процедуры обучения нейросетевой архитектуры ArcFace представлен на рисунке [9]. На первом этапе на базе нормализованного набора признаков x_i и нормализованного набора весов W_j для каждого класса формируется функция $\cos\theta_j$ (Logit, логит) в виде $W_j^T x_i$. Далее вычисляются значения $\arccos\theta_{yi}$ и определяются углы между вектором признаков x_i и векторами истинных весов W_{yi} . Затем к целевым углам θ_{yi} добавляются компоненты углового смещения m , после чего определяются величины $\cos(\theta_{yi} + m)$, которые в дальнейшем проходят процедуру масштабирования по параметру s . Далее логиты передаются в функцию softmax, после чего полученные значения используются при расчете функции потерь перекрестной энтропии (cross-

entropy loss). Использование предложенного подхода приводит к тому, что связанные признаки распределяются по гиперсфере с радиусом s , а аддитивная составляющая углового смещения m при этом одновременно способствует снижению внутриклассовых и увеличению межклассовых расстояний.

Среди существующих реализаций архитектуры ArcFace наиболее популярной является имплементация, выполненная с использованием программных библиотек PyTorch и TensorFlow. В контексте настоящего исследования задействован открыто распространяемый программный пакет InsightFace [18], включающий предобученную реализацию данной архитектуры.



Одним из ключевых условий реализации предлагаемого подхода является отсутствие потребности в наличии обучающего набора данных, включающего МНФ-изображения. Таким образом, в рамках настоящего исследования была предложена следующая стратегия формирования обучающего набора данных: 1) разделение исходного набора данных на блоки A ($n, \%$) и B ($100 - n, \%$) от изначального размера набора данных (независимо по каждому классу); 2) формирование набора синтетических данных (B') за счет применения к блоку B метода генерации синтетических МНФ-изображений [17]; 3) формирование результирующего набора данных для обучения путем объединения блоков A и B' .

Предлагаемая стратегия формирования обучающего набора данных, необходимого для дообучения нейросетевой модели ArcFace, предполагает подготовку обучающего набора данных на основе некоторого исходного набора. Соответствующий исходный набор данных по умолчанию не включает в себя экземпляров данных, представляющих лица в масках, и должен обладать достаточной репрезентативностью и отсутствием зашумленных данных. Согласно предложенной стратегии, исходный набор данных разделяется на блоки A и B , разделение осуществляется независимо по каждому классу k ($k \in \{1 \dots K\}$), представленному в исходном наборе данных.

По результатам разделения блок B обрабатывается методом генерации синтетических МНФ-изображений, в результате формируется блок B' . Размер и структура данных в блоке B' определяются заданными параметрами метода генерации, такими как набор вариаций СИЗ, набор текстур СИЗ, фактор репликации. По умолчанию размер результирующего блока B' будет эквивалентен размеру исходного блока B .

На заключительном шаге осуществляется формирование результирующего набора данных для обучения путем объединения блоков A и B' . В случае единичного значения фактора репликации полученный набор данных будет обладать идентичной структурой распределения данных по классам по сравнению с исходным набором данных.

В рамках настоящего исследования принятая стратегия дообучения для модели ArcFace является достаточно распространенной и предполагает разбиение набора данных на обучающую и валидационную части. Дообучение происходит исключительно на обучающей части набора данных с применением стратегии кросс-валидации. Реализация модели, продемонстрировавшая по результатам применения кросс-валидации наиболее высокое качество идентификации, проходит финальную проверку на валидационной части набора данных.

Предложенная стратегия позволит минимизировать риски переобучения модели в процессе ее дообучения.

Далее перейдем к оценке реализации предложенного подхода.

Эксперименты. В качестве исходного набора данных для обучения использовалось подмножество корпуса аудиовизуальных русскоязычных данных (BRAVE-MASKS — Biometric Russian Audio-Visual Extended MASKS corpus) [19], включающее 26 информантов, использующих защитные маски, — BRAVE-MASKS-CL. Каждое из изображений, используемых при формировании BRAVE-MASKS-CL, содержит один объект (лицо), ориентированный под произвольными углами относительно осей координат Ox , Oy и Oz соответственно. Изображения, входящие в состав BRAVE-MASKS-CL, отображают лица без масок, отснятые в двух вариантах отдаления от фиксирующей камеры: ближнее и дальнее расположение. Общая характеристика исходного набора данных BRAVE-MASKS-CL представлена в табл. 1.

При апробации предложенного подхода в целях проведения сравнительного анализа полученных результатов была сформирована эталонная модель распознавания. В качестве набора данных для обучения эталонной модели также использовалось подмножество визуального корпуса BRAVE-MASKS, включающее те же 26 информантов — BRAVE-MASKS-M. Все экземпляры данных, входящие в состав набора, представляют собой МНФ-изображения, прочие параметры BRAVE-MASKS-M в целом аналогичны BRAVE-MASKS-CL. Общая характеристика набора данных BRAVE-MASKS-M также приведена в табл. 1.

Таблица 1

Характеристика	BRAVE-MASKS-CL	BRAVE-MASKS-M
Общее количество изображений, шт.	3900	3900
Разрешение изображений, пиксел	1920×1080	1920×1080
Количество информантов	26	26
Среднее количество изображений на информанта, шт.	150	150
Число представленных типов СИЗ, шт.	—	4

В рамках настоящего эксперимента было апробировано множество реализаций стратегии формирования обучающего набора данных с различными параметрами как разбиения исходного набора данных на блоки, так и метода генерации синтетических изображений. Апробированные реализации данной стратегии с указанием значений представленных выше параметров отражены в табл. 2. По результатам применения указанных ранее реализаций стратегии формирования обучающих наборов данных к исходному BRAVE-MASKS-CL получено 14 наборов данных разного объема, различающихся долей МНФ-изображений, входящих в состав соответствующих наборов.

Таблица 2

Набор данных	n , %	Фактор репликации	Число вариаций СИЗ	Число вариаций текстур СИЗ
BRAVE-MASKS-CL-100	100	—	—	—
BRAVE-MASKS-CL-75/1	75	1	3	4
BRAVE-MASKS-CL-60/1	60	1		
BRAVE-MASKS-CL-60/2	60	2		
BRAVE-MASKS-CL-50/1	50	1		
BRAVE-MASKS-CL-50/2	50	2		
BRAVE-MASKS-CL-50/3	50	3		
BRAVE-MASKS-CL-40/1	40	1		
BRAVE-MASKS-CL-25/1	25	1		
BRAVE-MASKS-CL-25/2	25	2		
BRAVE-MASKS-CL-0/2	0	2		

В ходе экспериментов дообучение нейросетевой модели ArcFace проводилось на всех полученных наборах данных. Процесс дообучения осуществлялся независимо согласно описанной ранее стратегии дообучения с применением кросс-валидации, а также выделением

валидационной выборки. Число блоков кросс-валидации было задано равным 5, объем валидационной выборки составлял 20 %. Разбиение проводилось таким образом, чтобы все блоки содержали экземпляры данных каждого из классов с отсутствием множественных включений одних и тех же единиц данных в различные подмножества изображений. Аналогичная стратегия была использована и для дообучения эталонной модели идентификации, где в качестве обучающего набора данных выступал набор BRAVE-MASKS-M-80, сформированный путем исключения 80 % изображений по каждому классу из BRAVE-MASKS-M. Оставшаяся часть набора данных, обозначенная как BRAVE-MASKS-M-20, выполняла функцию контрольной выборки при определении результирующей оценки качества идентификации по МНФ-изображениям.

По итогам выполнения предложенной стратегии дообучения на каждом из обучающих наборов данных, представленных в табл. 2, определена дообученная реализация модели ArcFace, продемонстрировавшая наиболее высокое качество идентификации на соответствующих валидационных выборках, — модели (1)—(11). Значения метрики ААс (Averaged Accuracy), полученные данными моделями на валидационных выборках, отражены в табл. 3. На заключительном этапе эксперимента модели (1)—(11) были применены к набору данных BRAVE-MASKS-M-20 — в табл. 3 представлены полученные ими значения показателей ААс, АР (Averaged Precision) и АR (Averaged Recall), а также дополнительно приведены аналогичные показатели, полученные эталонной моделью идентификации.

Таблица 3

Модель	Обучающий набор данных	Валидационные выборки	Набор данных BRAVE-MASKS-M-20		
		ААс	ААс	АР	АР
(1)	BRAVE-MASKS-CL-100	0,989	0,795	0,796	0,773
(2)	BRAVE-MASKS-CL-75/1	0,993	0,835	0,847	0,764
(3)	BRAVE-MASKS-CL-60/1	0,987	0,855	0,869	0,767
(4)	BRAVE-MASKS-CL-60/2	0,992	0,872	0,893	0,793
(5)	BRAVE-MASKS-CL-50/1	0,995	0,868	0,863	0,789
(6)	BRAVE-MASKS-CL-50/2	0,986	0,874	0,899	0,775
(7)	BRAVE-MASKS-CL-50/3	0,984	0,881	0,909	0,798
(8)	BRAVE-MASKS-CL-40/1	0,996	0,875	0,899	0,781
(9)	BRAVE-MASKS-CL-25/1	0,996	0,853	0,868	0,757
(10)	BRAVE-MASKS-CL-25/2	0,991	0,833	0,847	0,762
(11)	BRAVE-MASKS-CL-0/2	0,99	0,813	0,827	0,750
Эталон	BRAVE-MASKS-M-80	0,983	0,957	0,974	0,892

На основе полученных результатов можно выявить определенную зависимость между долей синтетических МНФ-изображений в обучающем наборе данных и результирующим качеством предсказаний, реализуемых полученной моделью распознавания лиц. Данная зависимость условно носит параболический характер: первоначально, по мере увеличения в составе обучающей выборки доли синтетических МНФ-изображений (модели (1)—(7)), качество предсказаний достаточно устойчиво возрастает, при этом для моделей, обученных на выборках с более высоким фактором репликации, характерна более высокая точность распознавания. Однако при последующем увеличении доли синтетических МНФ-изображений в составе обучающих выборок до 60 % и более (без учета фактора репликации) зависимость между качеством предсказаний и долей синтетических изображений становится обратной, более того, фактор репликации также начинает вносить отрицательный вклад в конечное качество предсказаний. Таким образом, по результатам эксперимента наилучшее качество распознавания было продемонстрировано моделями (4), (6), (7) и (8), в данном случае метрика ААс на наборе данных BRAVE-MASKS-M-20 достигла значений 0,872, 0,874, 0,881 и 0,875 соответственно.

Важно отметить, что результирующее качество распознавания для полученных моделей так и не достигло уровня эталонной модели, где метрика ААс на наборе данных BRAVE-MASKS-M-20 достигла значения 0,957. Тем не менее качество распознавания для модели (7), продемонстрировавшей в рамках эксперимента наиболее высокое качество предсказаний среди разработанных моделей, оказалось по показателю ААс на 8,6 % выше по сравнению с моделью (1), где в состав обучающей выборки не были включены синтетические МНФ-изображения.

Заключение. На основе авторского метода генерации синтетических изображений лиц людей в защитных масках и модели распознавания лиц ArcFace разработан и успешно апробирован подход к распознаванию лиц по изображениям в условиях, когда лицо человека частично скрыто. Предложенный подход продемонстрировал высокие показатели качества распознавания лиц (ААс=0,881). По сравнению с базовым подходом, где в состав обучающей выборки не были включены синтетические МНФ-изображения, прирост точности распознавания по показателю ААс составил 8,6 %, а по AP — 11,3 %. Полученные результаты свидетельствуют о практической значимости предложенного подхода и позволяют сделать вывод о целесообразности дообучения нейросетевых моделей распознавания лиц с использованием смешанных наборов данных.

СПИСОК ЛИТЕРАТУРЫ

1. 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, N 10. P. 1499—1503. DOI: 10.1109/LSP.2016.2603342.
2. Zhang F., Fan X., Ai G., Song J., Qin Y., Wu J. Accurate face detection for high performance // arXiv preprint arXiv:1905.01585. 2019. P. 1—9.
3. Schroff F., Kalenichenko D., Philbin J. Facenet: A unified embedding for face recognition and clustering // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. P. 815—823. DOI: 10.1109/CVPR.2015.7298682.
4. Deng J., Guo J., Xue N., Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition // Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. P. 4690—4699.
5. He Y., Xu D., Wu L., Jian M., Xiang S., Pan C. LFFD: A light and fast face detector for edge devices // arXiv preprint arXiv:1904.10633. 2019. P. 1—10. DOI: 10.48550/arXiv.1904.10633.
6. Parkhi O. M., Vedaldi A., Zisserman A. Deep face recognition // British Machine Vision Conf.: Proc. 2015. P. 1—12. DOI: 10.5244/C.29.41.
7. Rab S., Javaid M., Haleem A., Vaishya R. Face masks are new normal after COVID-19 pandemic // Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020. Vol. 14, N 6. P. 1617—1619.
8. Martínez-Díaz Y., Méndez-Vázquez H., Luevano L. S., Nicolás-Díaz M., Chang L., González-Mendoza M. Towards Accurate and Lightweight Masked Face Recognition: an Experimental Evaluation // IEEE Access. 2021. Vol. 10. P. 7341—7353.
9. Anwar A., Raychowdhury A. Masked face recognition for secure authentication // arXiv preprint arXiv:2008.11104. 2020. P. 1—8.
10. Cao Q., Shen L., Xie W., Parkhi O. M., Zisserman A. Vggface2: A dataset for recognising faces across pose and age // 13th IEEE Intern. Conf. on Automatic Face & Gesture Recognition (FG 2018). 2018. P. 67—74.
11. Guo Y., Zhang L., Hu Y., He X., Gao, J. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition // European Conf. on Computer Vision. Cham: Springer, 2016. P. 87—102.
12. Wang Z., Wang G., Huang B., Xiong Z., Hong Q., Wu H., Liang J. Masked face recognition dataset and application // arXiv preprint arXiv:2003.09093. 2020. P. 1—3.
13. Liu W., Wen Y., Yu Z., Li M., Raj B., Song L. SpheroFace: Deep hypersphere embedding for face recognition // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. P. 212—220.

14. Wang H., Wang Y., Zhou Z., Ji X., Gong D., Zhou J., Liu W. Cosface: Large margin cosine loss for deep face recognition // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. P. 5265—5274.
15. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. Going deeper with convolutions // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. P. 1—9.
16. Kemelmacher-Shlizerman I. et al. The megaface benchmark: 1 million faces for recognition at scale // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. P. 4873—4882.
17. Letenkov M. A., Iakovlev R. N., Markitantov M. V., Ryumin D. A., Saveliev A. I., Karpov A. A. Method for Generating Synthetic Images of Masked Human Faces // Scientific Visualization. 2022. Vol. 14, N 2. P. 1—17. DOI: 10.26583/sv.14.2.01.
18. InsightFace: 2D and 3D Face Analysis Project [Электронный ресурс]: <<https://github.com/deepinsight/insightface>> 08.07.2022.
19. Markitantov M., Ryumina E., Ryumin D., Karpov A. Biometric Russian Audio-Visual Extended MASKS (BRAVE-MASKS) Corpus: Multimodal Mask Type Recognition Task // Proc. of ISCA Intern. Conf. INTERSPEECH-2022. Korea. 2022.

Сведения об авторах

- | | |
|--------------------------------------|--|
| Максим Андреевич Летенков | — СПбФИЦ РАН, СПИИРАН, лаборатория технологий больших данных социкиберфизических систем; мл. научный сотрудник; E-mail: letenkovmaksim@yandex.ru |
| Роман Никитич Яковлев | — СПбФИЦ РАН, СПИИРАН, лаборатория технологий больших данных социкиберфизических систем; мл. научный сотрудник; E-mail: iakovlev.r@mail.ru |
| Максим Викторович Маркитантов | — СПбФИЦ РАН, СПИИРАН, лаборатория речевых и многомодальных интерфейсов; мл. научный сотрудник; E-mail: m.markitantov@yandex.ru |
| Дмитрий Александрович Рюмин | — канд. техн. наук; СПбФИЦ РАН, СПИИРАН, лаборатория речевых и многомодальных интерфейсов; ст. научный сотрудник; E-mail: ryumin.d@iias.spb.su |
| Алексей Анатольевич Карпов | — д-р техн. наук, доцент; СПбФИЦ РАН, СПИИРАН, лаборатория речевых и многомодальных интерфейсов; гл. научный сотрудник; E-mail: karpov@iias.spb.su |

Поступила в редакцию 18.07.2022; одобрена после рецензирования 26.07.2022; принята к публикации 30.09.2022.

REFERENCES

1. Zhang K., Zhang Z., Li Z., Qiao Y. *IEEE Signal Processing Letters*, 2016, no. 10(23), pp. 1499–1503, DOI: 10.1109/LSP.2016.2603342.
2. Zhang F., Fan X., Ai G., Song J., Qin Y., Wu J. *arXiv preprint arXiv:1905.01585*, 2019, pp. 1–9.
3. Schroff F., Kalenichenko D., Philbin J. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823, DOI: 10.1109/CVPR.2015.7298682.
4. Deng J., Guo J., Xue N., Zafeiriou S. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.
5. He Y., Xu D., Wu L., Jian M., Xiang S., Pan C. *arXiv preprint arXiv:1904.10633*, 2019, pp. 1–10, DOI: 10.48550/arXiv.1904.10633.
6. Parkhi O. M., Vedaldi A., Zisserman A. *Deep face recognition*, 2015, pp. 1–12. DOI: 10.5244/C.29.41.
7. Rab S., Javaid M., Haleem A., Vaishya R. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 2020, no. 6(14), pp. 1617–1619.
8. Martínez-Díaz Y., Méndez-Vázquez H., Luevano L. S., Nicolás-Díaz M., Chang L., González-Mendoza M. *IEEE Access.*, 2021, vol. 10, pp. 7341–7353.
9. Anwar A., Raychowdhury A. *arXiv preprint arXiv:2008.11104*, 2020, pp. 1–8.
10. Cao Q., Shen L., Xie W., Parkhi O.M., Zisserman A. *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, 2018, pp. 67–74.
11. Guo Y., Zhang L., Hu Y., He X., Gao J. *European conference on computer vision*, Springer, Cham, 2016, pp. 87–102.
12. Wang Z., Wang G., Huang B., Xiong Z., Hong Q., Wu H., Liang J. *arXiv preprint arXiv:2003.09093*, 2020, pp. 1–3.
13. Liu W., Wen Y., Yu Z., Li M., Raj B., Song L. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 212–220.
14. Wang H., Wang Y., Zhou Z., Ji X., Gong D., Zhou J., Liu W. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5265–5274.

15. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
16. Kemelmacher-Shlizerman I. et al. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4873–4882.
17. Letenkov M.A., Iakovlev R.N., Markitantov M.V., Ryumin D.A., Saveliev A.I., Karpov A.A. *Scientific Visualization*, 2022, no. 2(14), pp. 1–17, DOI: 10.26583/sv.14.2.01.
18. *InsightFace: 2D and 3D Face Analysis Project*, <https://github.com/deepinsight/insightface>.
19. Markitantov M., Ryumina E., Ryumin D., Karpov A. *Proceedings of ISCA International Conference INTERSPEECH-2022*, Korea, 2022.

Data on authors

Maxim A. Letenkov	—	St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Laboratory of Big Data Technologies in Socio-Cyberphysical Systems; Junior Researcher; E-mail: letenkovmaksim@yandex.ru
Roman N. Iakovlev	—	St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Laboratory of Big Data Technologies in Socio-Cyberphysical Systems; Junior Researcher; E-mail: iakovlev.r@mail.ru
Maxim V. Markitantov	—	St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Speech and Multimodal Interfaces Laboratory; Junior Researcher; E-mail: m.markitantov@yandex.ru
Dmitry A. Ryumin	—	PhD; St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Speech and Multimodal Interfaces Laboratory; Senior Researcher; E-mail: ryumin.d@iias.spb.su
Alexey A. Karpov	—	Dr. Sci., Associate Professor; St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Speech and Multimodal Interfaces Laboratory; Chief Researcher; E-mail: karpov@iias.spb.su

Received 18.07.2022; approved after reviewing 26.07.2022; accepted for publication 30.09.2022.