# Метод генерации обучающих данных для компьютерной системы обнаружения защитных масок на лицах людей

*Е.В. Рюмина [1], Д.А. Рюмин [1], М.В. Маркитантов [1], А.А. Карпов [1]*
*[1] Санкт-Петербургский Федеральный исследовательский центр РАН (СПб ФИЦ РАН), 199178, Россия, г. Санкт-Петербург, 14-я линия В.О., д. 39*

***Аннотация***

Мониторинг и оценка уровня безопасности отдельных граждан и общества в целом является одной из важнейших проблем современного мира, который вынужден меняться в связи с возникновением коронавируса COVID-19. Для повышения уровня безопасности общества необходимы новые информационные технологии, способные остановить распространение пандемии за счет минимизации угроз новых вспышек и мониторинга соблюдения людьми защитных мер. К таким технологиям относятся, в частности, компьютерные системы для автоматизированного отслеживания наличия защитных масок на лицах людей. Для таких систем предлагается метод генерации обучающих данных, который объединяет такие способы аугментации данных, как Mixup и Insert. Предложенный метод апробируется на двух корпусах – MAsked FAce и Real-World Masked Face Recognition Dataset, для которых достигаются значения невзвешенной средней полноты при обнаружении масок в 98,51 % и 98,50 %. Кроме того, эффективность предложенного метода апробируется на изображениях с имитацией защитных масок на лицах людей и предлагается автоматизированный способ для уменьшения ошибок I и II рода. С помощью предложенного автоматизированного способа удается сократить количество ошибок II рода с 174 до 32 для корпуса Real-World Masked Face Recognition Dataset и с 40 до 14 для изображений с нарисованными защитными масками на реальных лицах людей.

*Ключевые слова*: обнаружение защитных масок, COVID-19, имитация защитных масок, генерация данных, визуальные характеристики, тепловая карта.

*Цитирование*: **Рюмина, Е.В.** Метод генерации обучающих данных для компьютерной системы обнаружения защитных масок на лицах людей / Е.В. Рюмина, Д.А. Рюмин, М.В. Маркитантов, А.А. Карпов // Компьютерная оптика. – 2022. – Т. 46, № 4. – С. 603-611. – DOI: 10.18287/2412-6179-CO-1039.

*Citation*: Ryumina EV, Ryumin DA, Markitantov MV, Karpov AA. A method for generating training data for a protective face mask detection system. Computer Optics 2022; 46(4): 603-611. DOI: 10.18287/2412-6179-CO-1039.

## *Введение*

В последние годы люди стран всего мира вынуждены соблюдать социальную дистанцию и носить средства индивидуальной защиты (СИЗ). Причиной этому является выявление новых штаммов коронавируса COVID-19 и рост количества зараженных людей, что приводит к ужесточению мер по борьбе с распространением пандемии. На сегодняшний день проведены многочисленные научные исследования, показывающие несомненную пользу ношения СИЗ. Так, исследование в работе [1] показало, что ношение маски на лице в общественных местах позволяет уменьшить распространение коронавирусной инфекции за счет снижения количества выбросов инфицированной слюны и респираторных капель от людей с проявлениями COVID-19. Авторы [2] также доказали, что использование многослойной маски и респиратора служит эффективным барьером от передачи инфекционных заболеваний в больнице и в других многолюдных общественных местах. Несмотря на то, что неоднократно была доказана эффективность ношения СИЗ, некоторые люди пренебрегают рекомендациями. В работе [3] анализируется взаимосвязь между возрастом человека, самовосприятием и ношением маски на лице. Опрос показал, что возраст не имеет взаимосвязи с самовосприятием маски на лице. Однако несмотря на то, что пожилые люди больше подвержены тяжелому течению болезни, они соблюдают рекомендации по ношению маски на лице реже молодых [3].

Мониторинг и оценка уровня безопасности общества является одной из важнейших проблем современного мира. Проблема несоблюдения рекомендаций является актуальной, и для борьбы с COVID-19 необходимы новые информационные технологии, способные остановить распространение заражения за счет минимизации угроз новых вспышек и мониторинга соблюдений защитных мер. К таким технологиям относятся цифровые методы автоматизации превентивных мер по борьбе с распространением коронавирусной инфекции путем интеллектуального отслеживания наличия защитных масок на лицах людей (далее обнаружение защитных масок). В настоящее время ведущие зарубежные научные институты и мировые промышленные корпорации проводят ис-

следования и разработки интеллектуальных технологий для решения данной задачи. Технологии, основанные на методах искусственного интеллекта, включая глубокое машинное обучение, позволяют обнаруживать защитные маски по акустическим [4, 5] или визуальным характеристикам людей [6, 7]. Также данные характеристики активно применяются для решения задач, связанных с обнаружением респираторных заболеваний [8, 9].

Обнаружение защитных масок по визуальным характеристикам людей на сегодняшний день является наиболее актуальной задачей. Однако исследователи в своих работах не оценивают эффективность предложенных методов на изображениях с имитацией защитных масок (ИЗМ) на лицах людей, что способно значительно понизить их работоспособность.

### *1. Современные методы обнаружения защитных масок*

Методы обнаружения защитных масок по визуальным характеристикам различаются по двум задачам: 1) реализация детектора обнаружения объектов, объектами являются, например, «лицо в маске» и «лицо без маски», т.е. на вход детектора подаются изображения, его цель – самостоятельно найти на изображениях области с необходимыми объектами; 2) реализация алгоритма машинной классификации по двум классам («лицо в маске» и «лицо без маски»), т.е. на вход алгоритма подаются изображения (содержащие только одно лицо без фоновой составляющей), требуется определить, к каким классам они принадлежат.

В работе [10] представлен метод обнаружения защитных масок для решения первой задачи. Авторы предлагают усовершенствование детектора обнаружения объектов YOLOv3, добавляя в него механизм внимания с блоком Squeeze and Excitation. Для генерации данных используются аффинные преобразования (горизонтальное отображение, случайная обрезка), регулировка контрастности изображений и метод Mixup [11]. Кроме того, исследовались и другие детекторы обнаружения объектов, результаты экспериментов показали, что предложенный метод обнаружения защитных масок значительно превосходит другие методы по показателю многокатегориального обнаружения объектов (mean Average Precision, mAP), но уступает по скорости. Так, при размере входного изображения 512×512 пикселей значение mAP составило 73,50 %.

В работе [12] авторы используют два детектора обнаружения объектов – YOLOv3 и Faster R-CNN. Как известно [13], одноэтапные детекторы обнаружения объектов с одной нейросетью (например, YOLOv3) превосходят двухэтапные с двумя нейросетями (Faster R-CNN) по скорости, однако уступают по показателю mAP. Поэтому авторы предлагают улучшения для детекторов. Так, для YOLOv3 добавляются 53 слоя к имеющимся 53 слоям, что позволяет извлекать более важную информацию. В то время как при работе с детектором Faster R-CNN авторы упрощают нейросеть определения областей интереса (Regional Proposal Network), рассматривая только 256 областей из более чем 16 тыс. возможных, максимально исключая их перекрытие между собой. Результаты исследований показали, что YOLOv3 уступает Faster R-CNN по показателю mAP на 7 % (55,00 % и 62,00 % соответственно), но значительно превосходит по скорости на 15,55 кадров в секунду (22,22 и 6,67 соответственно).

Методов обнаружения защитных масок для решения первой задачи на сегодняшний день мало. Это связано с недостатком обучающих данных, пригодных для исследований (т.е. изображений с наличием фоновой составляющей) и сложностью предварительной подготовки аннотированных данных, которые необходимо подстраивать под каждый детектор обнаружения объектов индивидуально. К тому же необходимо найти компромисс между двумя показателями (скорость и mAP). В случае применения разработанных методов, например, в контрольно-пропускных пунктах, скорость обработки данных является второстепенной. Однако в многолюдных общественных местах скорость зачастую критически важна. В связи с описанными сложностями, большая часть исследований направлена на решение задачи по разработке надежных алгоритмов машинной классификации (вторая задача).

В работе [6] предлагается комбинация визуальных текстурных признаков, извлекаемых с помощью архитектуры нейросети ResNet-50, и подсчет распределения интенсивности пикселей на изображениях (далее метод RNHist). Данные признаки объединяются и нормализуются до подачи на полносвязную нейросеть для последующей классификации. RNHist продемонстрировал свою эффективность в результате кросс-корпусного анализа на двух тестовых корпусах MAsked FAces (MAFA) [14] и Real-World Masked Face Recognition Dataset (RMFRD) [15], аннотированных на два класса («лицо в маске» и «лицо без маски»). Достигнут прирост для значений средней невзвешенной полноты (Unweighted Average Recall, UAR) на 1 % и составил 98,12 % и 97,68 % соответственно. Однако система плохо справляется с изображениями лиц при ИЗМ (UAR = 45,48 %), когда лицо перекрывается другим объектом (смартфоном, книгой и т.д.).

Еще один комбинированный метод представлен в работе [7], который сочетает методы традиционного и глубокого машинного обучения. Так, ResNet-50 выступает в качестве извлечения текстурных признаков, а метод опорных векторов (Support Vector Machine, SVM) – в качестве классификатора. На проверочных выборках из корпусов RMFRD и Simulated Masked Face Dataset (SMFD) [16] авторы получили точность

распознавания (Accuracy) 99,64 % и 99,49 % соответственно. На тестовом корпусе Labeled Faces in the Wild Simulated Masked Face Dataset (LFW-SMFD) [17] получена точность 100 %, такой результат достигается из-за неестественно (синтетически) наложенных защитных масок на изображения лиц. В своей работе авторы не рассматривают, справляется ли метод с перекрытием лица другими объектами.

Система обнаружения защитных масок SSDMNV2 представлена в работе [18] и включает детектор обнаружения лиц Single Shot Multibox Detector (SSD) и архитектуру нейросети MobileNetv2 для извлечения признаков и классификации. Для исследования авторы собрали свой корпус, в котором аннотированы объекты двух классов – «лицо в маске» и «лицо без маски», и дополнительно применяют аффинные преобразования для генерации новых данных. Так, на проверочной выборке удалось достичь UAR = 92,64 %. С другими работами в этой области можно ознакомиться в обзоре [19].

При анализе методов, разработанных для решения двух задач, можно заметить, что основным показателем является точность (mAP или UAR). Так, при решении второй задачи реализуются более надежные методы обнаружения защитных масок, которые достигают значения UAR выше 90 %, тогда как при решении первой задачи mAP не превышает 75 %. Это связано с тем, что показатель mAP учитывает верные случаи, если область лица найдена правильно и верно классифицирована, при ложной классификации или ложном нахождении лица значение показателя mAP падает.

Таким образом, на сегодняшний день активно разрабатываются методы обнаружения защитных масок. Однако только в одной работе [6] с помощью RNHist выполняется проверка эффективности метода на изображениях лиц при ИЗМ. Поэтому цель текущего исследования заключается в усовершенствовании метода RNHist за счет улучшения процесса генерации обучающих данных, а также в детальном рассмотрении проблемы ИЗМ и предложении метода для уменьшения ошибочно предсказанных классов.

### *2. Исследовательские данные*

При реализации RNHist для обучения и проверки использовался корпус Medical Mask Extended Dataset (MMED). В текущей работе MMED увеличивается за счет изображений из других корпусов: MAFA (обучающая выборка) и Labeled Faces in the Wild (LFW) [20]. Далее объединенный корпус назовем MMED2.

При работе с корпусами MAFA и LFW потребовалось выполнить обнаружение всех областей лиц на изображениях с помощью детектора RetinaFace [21] и ручное аннотирование согласно правилам, предложенным в [6]. Результат нашего ручного аннотирования для корпуса MAFA можно найти в [22]. В табл. 1 представлено распределение изображений по классам в исследуемых корпусах. Где класс 0 – «лицо в маске», класс 1 – «лицо без маски», класс 2 – «некорректно надетая маска», класс 3 – «перекрытие другим объектом». В рамках текущей работы принято решение объединить в MMED2 классы «перекрытие другим объектом» и «лицо без маски». Как показали исследования [6, 10], это объединение необходимо сделать для того, чтобы алгоритм машинной классификации научился не допускать ложные пропуски (уменьшение ошибки II рода) при перекрытии лица иным предметом, отличным от защитной маски. Изображения, принадлежащие к классу «некорректно надетая маска», не использовались в текущей работе.

*Табл. 1. Распределение изображений по классам*

| Корпус | Класс 0 | Класс 1 | Класс 2 | Класс 3 |
|---|---|---|---|---|
| Обучающие корпусы | | | | |
| MMED | 6769 | 6769 | – | – |
| MAFA | 1644 | 23889 | 715 | 3204 |
| LFW | 15054 | – | – | – |
| MMED2 | 26660 | 30647 | – | – |
| Тестовые корпусы | | | | |
| RMFRD | 90468 | 2203 | – | – |
| MAFA | 447 | 3707 | 128 | 653 |

Как можно заметить из табл. 1, распределение изображений в классах не сбалансировано, эта проблема оказывает негативное влияние на эффективность алгоритмов машинной классификации [23]. В рамках текущего исследования для решения этой проблемы используется обратно-пропорциональное взвешивание классов по их частоте. Веса для классов устанавливаются согласно формуле (1):

$$w_i = \frac{N}{n_i \times 2}, \tag{1}$$

где $N$ – количество изображений в корпусе, $n_i$ – количество изображений, принадлежащих классу $i$, $i$ – порядковый номер класса от 0 до $2-1$, 2 – количество классов («лицо в маске», «лицо без маски»).

Для оценивания алгоритма машинной классификации выбраны корпусы MAFA (тестовая выборка) и RMFRD, это необходимо для сравнения значений показателя UAR со значениями, полученными с помощью RNHist. В табл. 1 представлено распределение изображений по классам в тестовых корпусах.

### *3. Предлагаемый метод обнаружения защитных масок*

В нашем методе (RNMask) обнаружения защитных масок используется предварительно обученная ResNet-50, которая продемонстрировала свою эффективность в предыдущем исследовании [6]. Настройка нейросети и классифицирующие слои также соответствуют настройкам, представленным в [6]. Однако, как сказано ранее, в текущем исследовании значительно расширяется объем MMED2, а также уменьшается количество обучающих эпох до 15. Затем к

MMED2 применяются два способа генерации данных: 1) Mixup [11] (далее метод RNMaskMixup); 2) Insert – случайная вставка другого изображения (далее метод RNMaskInsert), а также их комбинация (далее метод RNMaskMixup+Insert). На рис. 1 представлена схема предложенного метода обнаружения защитных масок (RNMaskMixup+Insert).

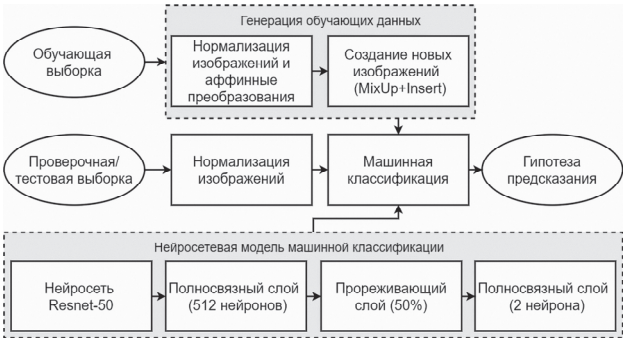


*Рис. 1. Метод обнаружения защитных масок (RNMaskMixup+Insert)*

Из рис. 1 можно заметить, что отдельным блоком выделена генерация обучающих данных, которая поделена на два этапа. Этап нормализации изображений и аффинных преобразований применяется во всех предложенных методах. Этап создания новых изображений используется в методах RNMaskMixup, RNMaskInsert, RNMaskMixup+Insert и отличается в зависимости от способа генерации данных (Mixup и/или Insert). Таким образом, в текущем исследовании предлагается включить дополнительный блок генерации обучающих данных в процессе обучения нейросетевой модели машинной классификации.

### *4. Экспериментальные исследования*

Для проведения экспериментальных исследований MMED2 разделен на 5 равномерных частей с учетом непересекаемости изображений для выполнения перекрестной проверки (Cross-Validation, CV). CV дает более надежную оценку эффективности предлагаемых методов, так как тестирование производится на 5 неидентичных проверочных выборках. Также обязательным этапом в предлагаемых методах обнаружения защитных масок является генерация обучающих данных, которая подробно описана далее.

#### *4.1. Генерация обучающих данных*

Как можно заметить из рис. 1, процесс генерации обучающих данных состоит из двух этапов: 1) нормализация изображений и аффинные преобразования; 2) создание новых изображений.

Для нормализации изображений выполняются следующие действия: 1) канальная нормализация, соответствующая ResNet-50; 2) приведение изображений к единому разрешению 224 ×224 пикселей. Данная нормализация выполняется для изображений из обучающих, проверочных и тестовых выборок исследовательских корпусов. Также в процессе обучения нейросети применяются случайные аффинные преобразования.

Этап создания новых изображений с помощью способов генерации данных Mixup и/или Insert отсутствует в методе RNMask, кроме того, размер партии равен 64 изображениям. Для методов RNMaskMixup, RNMaskInsert размер партии составляет 32 изображения, при этом партия делится на две равные части и выполняется попарное слияния изображений из двух частей. Степень слияния изображений регулируется весовым коэффициентом. Сначала создается новое изображение с большим/меньшим весом изображения из первой части, затем с тем же весом изображения из второй части и так далее. На выходе получается 32 изображения, к которым применены только нормализация и аффинные преобразования, и столько же новых скрещенных изображений после генерации данных Mixup или Insert. Метод RNMaskMixup+Insert схож с методами RNMaskMixup и RNMaskInsert, однако сначала создается новое изображение с помощью способа генерации данных Mixup, затем с помощью Insert. Помимо самих изображений, также изменяются бинарные вектора (One-Hot Vectors) в соответствии с [11]. На рис. 2 представлен пример создания новых изображений, где $W_1$ – вес первого изображения (в %), $W_2$ – вес второго (в %).


*Рис. 2. Пример создания изображений предложенным методом RNMaskMixup+Insert*

Как можно заметить из рис. 2, на каждую пару изображений создаются два новых изображения. Кроме того, на каждой эпохе в одну партию попадают разные изображения, к которым применяются случайные аффинные преобразования, поэтому новые сгенерированные изображения отличны от предыдущих, что позволяет значительно увеличить вариативность изображений.

Для способов генерации данных (Mixup и Insert) устанавливаются следующие ограничения. В Mixup устанавливается случайный весовой коэффициент в интервале [0,3; 0,7]. Так, при коэффициенте 0,3 вес первого изображения составит 30 %, второго – 70 %, подробнее о способе можно ознакомиться в [11]. В

Insert подбираются четыре параметра: ширина, высота, смещение по ширине и высоте вставляемого изображения. Первые два параметра подбираются в интервале значений [80; 150] с шагом 10 пикселей, вторые два – в интервале [0; 140] с тем же шагом. Замена пикселей в первом изображении выполняется с места смещения второго (встраиваемого) изображения. Доля площади вставляемого изображения от размера нормализованного изображения является весовым коэффициентом для изменения бинарного вектора, изменения аналогичны Mixup [11]. Такие параметры позволяют выполнять существенные преобразования в исходных изображениях.

### *4.2. Результаты экспериментов*

В табл. 2 представлены результаты экспериментов, полученные нейросетевой моделью. Для оценки эффективности методов используется показатель UAR. Для тестовой выборки корпуса MAFA значения UAR представлены отдельно для классов 2 и 3 и совокупно для классов 0 и 1. При подсчете значений показателя UAR изображения для классов 2 и 3 учитываются как класс «лицо без маски», т.е. класс 0. Для 5 проверочных выборок при CV представляются усредненные значения показателя UAR и стандартные отклонения (STD). Для проверки эффективности предложенных методов на тестовых корпусах производилось обучение нейросетевой модели на всем обучающем наборе (т.е. без разделения на обучающую и проверочную выборки). Кроме того, в табл. 2 представлены значения разности (Δ) между результатами, достигнутыми с помощью RNHist, и результатами других методов, предложенных в текущем исследовании.

Из табл. 2 видно, что увеличение изображений в MMED2 за счет объединения нескольких корпусов (метод RNMask) дает прирост показателя UAR для корпуса RMFRD на 0,37 % и MAFA (класс 3) на 29,41 %. Это связано с тем, что в MMED2 добавлены сложные изображения лиц при ИЗМ. В свою очередь, RNMaskMixup+Insert достигает лучшее значение показателя UAR при CV и имеет меньшее значение STD, что говорит о том, что предложенный метод показывает более стабильное значение показателя UAR вне зависимости от тестовых данных. Также можно заметить, что данный метод показывает лучшие значения UAR на других исследовательских корпусах (RMFRD, MAFA для классов 0, 1) в сравнении с RNHist. А также для корпуса MAFA (класс 3), т.е. при ИМЗ достигается прирост значения UAR на 40,43 %, что также свидетельствует об эффективности предлагаемого метода RNMaskMixup+Insert. Стоит отметить, что при включении этапа создания новых изображений в процесс обучения нейросети время обучения одной эпохи в среднем увеличилось на 27 %, такое увеличение можно считать незначительным с учетом того, что генерация обучающих данных выполняется «на лету».

Как упоминалось ранее, в предыдущем исследовании [6], предложен метод на основе комбинации визуальных текстурных признаков. В текущей работе этот метод также применен. В табл. 2 (см. MMED2 + RNHist) представлены полученные результаты. Можно заметить, что также достигается прирост значений UAR в сравнении с RNHist. Однако при попытке дополнить методы из табл. 2 (RNMaskMixup, RNMaskInsert и RNMaskMixup+Insert) подсчетом распределения интенсивности пикселей на изображениях нам не удалось улучшить значения UAR. Это связано с тем, что после канальной нормализации изображений и способов генерации данных Mixup и/или Insert распределение интенсивности пикселей на изображении значительно искажается, что не позволяет извлечь надежные информативные признаки.

## *5. Проблема имитации защитных масок*

Надежность методов обнаружения защитных масок на лицах людей значительно ухудшается при перекрытии лица другими предметами, отличными от защитной маски. Поэтому в данном параграфе анализируется работоспособность предложенных методов на случайных изображениях из тестовой выборки корпуса MAFA. Для анализа используются тепловые карты [24]. Построение тепловых карт дает возможность визуализировать то, что оценивает нейросеть, когда делает предсказание в сторону определенного класса. Визуализация позволяет увидеть, какие области лица (области интереса) на изображении нейросеть считает важными для конкретного класса. На рис. 3 представлен результат наложения тепловых карт на изображения, где горячий красный цвет показывает наиболее важные области интереса на изображении, а холодный синий – менее информативные области. В том числе отображены вероятностные прогнозы (в %), где $P_0$, $P_1$ – вероятностные прогнозы для классов 0 и 1.

*Табл. 2. Результаты экспериментов с различными методами генерации обучающих данных (UAR, %)*

| Метод | CV (STD) | RMFRD (Δ) | MAFA (классы 0 и 1) (Δ) | MAFA (класс 2) (Δ) | MAFA (класс 3) (Δ) |
|---|---|---|---|---|---|
| RNHist | – | 97,68 | 98,12 | 37,50 | 45,48 |
| RNMask | 98,23 (± 0,09) | 98,05 (+ 0,37) | 98,00 (– 0,12) | 33,59 (– 3,91) | 74,89 (+ 29,41) |
| RNMaskMixup | 98,35 (± 0,18) | 98,07 (+ 0,39) | 97,35 (– 0,77) | **47,66** (– 10,16) | 82,08 (+ 36,60) |
| RNMaskInsert | 98,41 (± 0,13) | 98,29 (+ 0,61) | 97,70 (– 0,42) | 20,31 (– 17,31) | 75,96 (+ 30,48) |
| RNMaskMixup+Insert | **98,42** (± 0,10) | **98,51** (+ 0,83) | **98,50** (+ 0,38) | 42,31 (+ 4,81) | **85,91** (+ 40,43) |
| MMED2 + RNHist | – | 98,12 (+ 0,44) | 98,16 (+ 0,04) | 43,75 (+6,25) | 70,90 (+ 25,42) |


*Рис. 3. Результат наложения тепловых карт на изображения и вероятностные прогнозы*

Из рис. 3 можно заметить, что в первых двух рядах все предложенные методы из табл. 2 верно распознали классы, представленные на изображениях. Так, все методы акцентируют внимание на область носа, если на изображении представлено «лицо без маски». Когда на лице присутствует маска, то методы RNMask и RNMaskMixup делают акцент на границе между областью без маски и с маской, в то время как два других метода (RNMaskInsert и RNMaskMixup+Insert) на область глаз. В случае с изображениями, принадлежащими к классам 2 и 3, предполагалось, что методы отнесут изображения к классу «лицо без маски». Однако можно заметить, что метод RNMask показывает низкую вероятность принадлежности к классу «лицо без маски», а также областями интереса являются скуловая и подглазная области. Тогда как по другим методам можно заметить, что они извлекают больше полезной информации, учитывая также нижнюю часть лица и даже волосы. Причина акцента на волосах скрывается в том, что в MMED2 имеются изображения, где лицо перекрыто волосами. Так, нейросеть обучилась на том, что обилие волос на изображении в области лица свидетельствует о ИЗМ. Следовательно, при сложных изображениях можно утверждать, что при открытой нижней области лица и присутствии волос на лице использование методов RNMaskInsert и RNMaskMixup+Insert с большой долей вероятности позволит верно отнести изображения к классу «лицо без маски».

Также стоит упомянуть, что в 2021 году зафиксирован случай ИЗМ, при котором на лице человека была нарисована защитная маска. Поэтому для исследований этой проблемы нами собраны из Интернета 53 изображения с нарисованными защитными масками на лицах людей (назовем собранный корпус Painted Face Masks Dataset (PFMD). На рис. 4 продемонстрированы примеры из PFMD.

Далее на изображениях из корпуса PFMD протестированы 4 предложенных метода (из табл. 2). Так, с учетом того, что исследуемые изображения относятся к классу «лицо без маски», максимальное значение UAR = 24,53 % достигается с помощью метода RNMaskMixup+Insert, что является достаточно низким результатом в сравнении с полученным значением 85,91 % для корпуса MAFA (класс 3). Поэтому для повышения эффективности предложенных методов в текущей работе предлагается установить пороговое значение для вероятностных прогнозов. Прогнозная метка класса устанавливается согласно:

$$\hat{y}_j = \begin{cases} y_j, & \text{если} \quad 1-TV \le P_{j0} \le TV; \\ \arg\max(P_j), & \text{иначе}, \end{cases} \tag{2}$$

где $\hat{y}_j$ – прогнозная метка класса для $j$ изображения, $j$ – порядковый номер изображения от 0 до $N-1$, $N$ – количество изображений в корпусе, $y_j$ – верная метка класса (0/1) для $j$ изображения, $TV$ – пороговое значение, $P_{j0}$ – вероятностный прогноз для класса 0 изображения $j$, $\arg\max()$ – функция, возвращающая индекс/метку класса (0/1) с максимальным вероятностным прогнозом, $P_j$ – вектор вероятностных прогнозов для $j$ изображения.

*Рис. 4. Примеры изображений из корпуса PFMD*

Очевидно, что при бинарной классификации пороговое значение $TV$ должно быть установлено в интервале $0{,}5 < TV < 1$. Для рассмотрения всех «неуверенных решений» нейросети с использованием только вероятности класса 0, необходимо установить нижнюю границу для выполнения первого условия (2) в виде $1-TV$. В случае применения предложенных методов (из табл. 2), например, в контрольно-пропускных пунктах, изображения, относящиеся к «неуверенным решениям», необходимо просмотреть вручную проверяющему персоналу (например, оператору). Таким образом, предлагается автоматизированный способ для повышения эффективности предложенных методов за счет анализа изображений, относящихся к «неуверенным решениям» нейросети.

Для исследования использованы корпусы RMFRD, PFMD и изображения из тестовой выборки корпуса MAFA (класс 3). В табл. 3 представлены результаты экспериментов, $TV$ – пороговое значение, $HP$ – количество «неуверенных решений» нейросети (шт.), $ДО$ – количество допущенных ошибок (шт.), $N$ – количество изображений в корпусе, $O_{\text{I}}/O_{\text{II}}$ – количество ошибок I/II рода (шт.), где $O_{\text{I}}$ – ошибочное отнесение изображений к классу «лицо без маски» (ложные срабатывания), $O_{\text{II}}$ – ошибочное отнесение изображений к классу «лицо в маске» (ложные пропуски).

*Табл. 3. Подбор пороговых значений для предсказаний по изображениям из корпусов RMFRD, PFMD и MAFA (класс 3)*

| TV | Метод | HP | ДО | $O_I$ | $O_{II}$ | HP | $O_{II}$ | HP | $O_{II}$ |
|---|---|---|---|---|---|---|---|---|---|
| Корпус | | RMFRD ($N=92671$) | | | | MAFA (класс 3) ($N=653$) | | PFMD ($N=53$) | |
| - | RNMask | - | 238 | 82 | 156 | - | 164 | - | 44 |
| 0,9 | | 626 | 99 | 51 | 48 | 253 | 68 | 14 | 34 |
| 0,8 | | 354 | 131 | 60 | 71 | 175 | 94 | 5 | 42 |
| - | RNMaskMixup | - | 279 | 80 | 199 | - | 117 | - | 46 |
| 0,9 | | 2469 | 82 | 41 | 41 | 236 | 35 | 29 | 20 |
| 0,8 | | 837 | 119 | 49 | 70 | 166 | 53 | 17 | 31 |
| - | RNMaskInsert | - | 245 | 71 | 174 | - | 157 | - | 52 |
| 0,9 | | 1103 | 70 | 38 | 32 | 390 | 39 | 18 | 35 |
| 0,8 | | 595 | 100 | 43 | 57 | 278 | 70 | 10 | 43 |
| | RNMaskMixup+Insert | - | 648 | 51 | 597 | - | 92 | - | 40 |
| 0,9 | | 8325 | 129 | 27 | 102 | 237 | 19 | 36 | 14 |
| 0,8 | | 3437 | 241 | 34 | 207 | 149 | 36 | 27 | 22 |

Из рис. 3 можно заметить, что при установке *TV* больше 0,98 (98 %) в *HP* попадут случаи, которые можно отнести к погрешности нейросети. Поэтому для сравнения эффективности предложенных методов *TV* устанавливается на 0,9 и 0,8, такие значения *TV* вносят существенные различия в ДО.

Результаты из табл. 3 демонстрируют, что чем выше *TV,* тем меньше допускается ошибок, однако необходимо просмотреть больше спорных изображений, на что может потребоваться время, поэтому высокое значение *TV* может быть установлено, например, в системах контрольно-пропускных пунктов. Также можно сказать, что показатель ошибок II рода более значим, чем I рода, поскольку чем меньше «лиц без маски» пропускает система, тем она надежней. Так, для корпуса RMFRD метод RNMaskInsert оказался более надежным, а $O_2$ уменьшилось с 174 до 32 ошибок. Для корпусов MAFA (класс 3) и PFMD меньшие значения $O_2$ достигаются с помощью метода RNMaskMixup+Insert. Следовательно, можно утверждать, что данный метод лучше остальных справляется с проблемой ИЗМ. Таким образом, в данном параграфе предложен способ для уменьшения ошибок I и II рода как для простых изображений лиц с явным наличием/отсутствием защитных масок, так и с ИЗМ.

### *Заключение*

В работе рассмотрены и исследованы современные методы обнаружения защитных масок на лицах людей. Предложены методы генерации обучающих данных, в основе которых лежат способы Mixup и/или Insert. Результаты экспериментов показали, что с помощью одного из предложенных методов генерации данных RNMaskMixup+Insert получены значения UAR 98,51 % и 98,50 %, для тестовых корпусов RMFRD и MAFA (классы 0 и 1), что показывает абсолютный прирост 0,83 % и 0,38 % в сравнении с ранее предложенным нами методом RNHist. Однако исследование изображений лиц с имитацией защитных масок (корпусы MAFA (классы 3) и PFMD) показало значения UAR 85,91 % и 24,53 % соответственно, что значительно меньше по сравнению с изображениями лиц с явным наличием/отсутствием защитных масок (корпусы MAFA (классы 0 и 1) и RMFRD). В связи с этим предлагается автоматизированный способ для уменьшения количества ошибок I и II рода. Так, для корпусов MAFA (классы 3) и PFMD количество ошибок II рода уменьшилось с 92 до 19 и с 40 до 14 соответственно, что говорит об эффективности предложенного автоматизированного способа.

Так как в текущей работе предложены методы генерации обучающих данных для задачи обнаружения защитных масок на лицах людей при заранее локализованных областях лиц, то в последующих исследованиях планируется разработать метод обнаружения защитных масок на лицах людей, который будет решать сразу две задачи, а именно: обнаруживать область лица на изображении с предоставлением ограничительных рамок; классифицировать обнаруженную область лица как «лицо без маски» либо «лицо в маске». В качестве обучающих данных планируется использование корпусов MMED, MAFA, LFW и Biometric Russian Audio-Visual Extended MASKS (BRAVE-MASKS) [25], а тестовых – MAFA и BRAVE-MASKS.

### *Благодарности*

Работа выполнена при поддержке проекта фонда РФФИ № 20-04-60529-вирусы, а также частично в рамках бюджетной темы № 0073-2019-0005.

### *References*

[1] Cheng VC, Wong SC, Chuang VW, So SY, Chen JH, Sridhar S, To KK, Chan JF, Hung IF, Ho PL, Yuen KY. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J Infect 2020; 81(1): 107-114. DOI: 10.1016/j.jinf.2020.04.024.

[2] Wang J, Pan L, Tang S, Ji JS, Shi X. Mask use during COVID-19: A risk adjusted strategy. Environ Pollut 2020; 266(1): 115099. DOI: 10.1016/j.envpol.2020.115099.

[3] Howard MC. The relations between age, face mask perceptions and face mask wearing. J Public Health (Oxf) 2021: fdab018. DOI: 10.1093/pubmed/fdab018.

[4] Markitantov M, Dresvyanskiy D, Mamontov D, Kaya H, Minker W, Karpov A. Ensembling end-to-end deep models

for computational paralinguistics tasks: ComParE 2020 mask and breathing sub-challenges. Proc Interspeech 2020: 2072-2076. DOI: 10.21437/Interspeech.2020-2666.
[5] Montacié C, Caraty M. Phonetic, frame clustering and intelligibility analyses for the INTERSPEECH 2020 ComParE challeng. Proc Interspeech 2020: 2062-2066. DOI: 10.21437/Interspeech.2020-2243.
[6] Ryumina E, Ryumin D, Ivanko D, Karpov A. A novel method for protective face mask detection using convolutional neural networks and image histograms. Int Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences 2021; XLIV-2/W1-2021: 177-182. DOI: 10.5194/isprs-archives-XLIV-2-W1-2021-177-2021.
[7] Loey M, Manogaran G, Taha MHN, Khalifa NEM. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement 2021; 167: 108288. DOI: 10.1016/j.measurement.2020.108288.
[8] Deshpande G, Schuller BW. Audio, speech, language, & signal processing for COVID-19: A comprehensive overview. arXiv Preprint 2020. Source: ⟨https://arxiv.org/abs/2011.14445⟩.
[9] Efremtsev VG, Efremtsev NG, Teterin EP, Teterin PE, Bazavluk ES. Chest X-ray image classification for viral pneumonia and Covid-19 using neural networks. Computer Optics 2021; 45(1): 149-153. DOI: 10.18287/2412-6179-CO-765.
[10] Jiang X, Gao T, Zhu Z, Zhao Y. Real-time face mask detection method based on YOLOv3. Electronics 2021; 10(7): 837. DOI: 10.3390/electronics10070837.
[11] Zhang H, Cissé M, Dauphin Y, Lopez-Paz D. Mixup: Beyond empirical risk minimization. Proc. International Conference on Learning Representations (ICLR) 2018; 1-13.
[12] Singh S, Ahuja U, Kumar M, Kumar K, Sachdeva M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimed Tools Appl 2021; 80(13): 19753-19768. DOI: 10.1007/s11042-021-10711-8.
[13] Vizilter YV, Gorbatsevich VS, Moiseenko AS. Single-shot face and landmarks detector. Computer Optics 2020; 44(4): 589-595. DOI: 10.18287/2412-6179-CO-674.
[14] Ge S, Li J, Ye Q, Luo Z. Detecting masked faces in the wild with LLE-CNNs. Proc IEEE Conf on Computer Vision and Pattern Recognition 2017: 2682-2690. DOI: 10.1109/CVPR.2017.53.
[15] Wang Z, Wang G, Huang B, Xiong Z, Hong Q, Wu H, Yi P, Jiang K, Wang N, Pei Y, Chen H, Miao Y, Huang Z, Liang J. Masked face recognition dataset and application. arXiv Preprint 2020. Source: ⟨https://arxiv.org/abs/2003.09093⟩.
[16] The simulated masked face dataset. Source: ⟨https://github.com/prajnasb/observations/⟩.
[17] The labeled faces in the wild simulated masked face dataset. Source: ⟨https://www.kaggle.com/muhammeddalkran/lfw-simulated-masked-face-dataset/⟩.
[18] Nagrath P, Jain R, Madan A, Arora R, Kataria P, Hemanth J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustain Cities Soc 2021; 66: 102692. DOI: 10.1016/j.scs.2020.102692.
[19] Dvoynikova AA, Markitantov MV, Ryumina EV, Ryumin DA, Karpov AA. Analytical review of audiovisual systems for determining personal protective equipment on a person's face [In Russian]. Informatics and Automation 2021; 20(5): 1116-1152. DOI: 10.15622/ia.2021.20.5.
[20] Learned-Miller E, Huang GB, RoyChowdhury A, Li H, Hua G. Labeled faces in the wild: A survey. In Book: Kawulok M, Celebi E, Smolka B, eds. Advances in face detection and facial image analysis. New York: Springer; 2016: 189-248. DOI: 10.1007/978-3-319-25958-1_8.
[21] Deng J, Guo J, Ververas E, Kotsia I, Zafeiriou S. RetinaFace: Single-shot multi-level face localisation in the wild. Proc IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2020: 5203-5212. DOI: 10.1109/CVPR42600.2020.00525.
[22] The annotation for MAsked FAce. Source: ⟨https://github.com/ElenaRyumina/AnnotationMAFA/⟩.
[23] Ryumina EV, Karpov AA. Comparative analysis of methods for imbalance elimination of emotion classes in video data of facial expressions [In Russian]. Scientific and Technical Journal of Information Technologies, Mechanics and Optics 2020; 20(5:129): 683-691. DOI: 10.17586/2226-1494-2020-20-5-683-691.
[24] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. IEEE Int Conf on Computer Vision 2017: 618-626. DOI: 10.1109/ICCV.2017.74.
[25] Markitantov MV, Ryumin DA, Ryumina EV, Karpov AA. Corpus of audiovisual Russian-language data of people in protective masks (BRAVE-MASKS – Biometric Russian Audio-Visual Extended MASKS corpus) [In Russian]. Database state registration certificate N2021621094 of May 26, 2021.

### *Сведения об авторах*

**Рюмина Елена Витальевна**, 1991 года рождения, в 2021 году окончила Национальный исследовательский университет ИТМО по специальности 09.04.02 «Информационные системы и технологии», работает младшим научным сотрудником в лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН). Область научных интересов: аффективные вычисления, цифровая обработка изображений, распознавание визуальных сигналов, автоматическое распознавание паралингвистических явлений, машинное обучение, нейронные сети, биометрические системы, человеко-машинные интерфейсы. E-mail: *ryumina.e@iias.spb.su* .

**Рюмин Дмитрий Александрович**, 1991 года рождения, в 2016 году окончил факультет информационных технологий и программирования, а в 2020 году защитил кандидатскую диссертацию на тему «Модели и методы автоматического распознавания элементов русского жестового языка для человеко-машинного взаимодействия» в Университете ИТМО. Старший научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН). Область науч-

ных интересов: цифровая обработка изображений, распознавание образов, автоматическое распознавание визуальной речи, многомодальные интерфейсы, машинное обучение, нейронные сети, биометрия, человеко-машинные интерфейсы. E-mail: *ryumin.d@iias.spb.su* .

**Маркитантов Максим Викторович**, 1995 года рождения, в 2019 году окончил Национальный исследовательский университет ИТМО по специальности 09.04.04 «Программная инженерия», работает младшим научным сотрудником в лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН). Область научных интересов: искусственный интеллект, машинное обучение, речевые технологии, компьютерная паралингвистика, распознавание характеристик диктора, распознавание пола и возраста диктора, обнаружение защитных масок по аудиоинформации. E-mail: *m.markitantov@yandex.ru* .

**Карпов Алексей Анатольевич**, 1978 года рождения, доктор технических наук (2013), доцент по специальности 05.13.11 (2012). Работает главным научным сотрудником (руководителем лаборатории) речевых и многомодальных интерфейсов Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН). Область научных интересов: речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика и другие. E-mail: *karpov@iias.spb.su* .

*ГРНТИ: 28.23.15*
*Поступила в редакцию 3 сентября 2021 г. Окончательный вариант – 27 октября 2021 г.*

# A method for generating training data for a protective face mask detection system

*E.V. Ryumina [1], D.A. Ryumin [1], M.V. Markitantov [1], A.A. Karpov [1]*
*[1] St. Petersburg Federal Research Center of the RAS (SPC RAS), 199178, St. Petersburg, Russia, 14th Line V.O. 39*

***Abstract***

Monitoring and evaluation of the safety level of individuals is one of the most important problems of the modern world, which was forced to change due to the emergence of the COVID-19 virus. To increase the safety level of individuals, new information technologies are needed that can stop the spread of infection by minimizing the threat of outbreaks and monitor compliance with recommended measures. These technologies, in particular, include intelligent tracking systems of the presence of protective face masks. For these systems, this article proposes a new method for generating training data that combines data augmentation techniques, such as Mixup and Insert. The proposed method is tested on two datasets, namely, the MAsked FAce dataset and the Real-World Masked Face Recognition Dataset. For these datasets, values of the unweighted average recalls of 98.51% and 98.50% are obtained. In addition, the effectiveness of the proposed method is tested on images with face mask imitation on people's faces, and an automated technique is proposed for reducing type I and II errors. Using the proposed automated technique, it is possible to reduce the number of type II errors from 174 to 32 for the Real-World Masked Face Recognition Dataset, and from 40 to 14 for images with painted protective face masks.

*Keywords*: protective face mask detection, COVID-19, protective face mask imitation, data augmentation, visual features, heatmap.

*Citation*: Ryumina EV, Ryumin DA, Markitantov MV, Karpov AA. A method for generating training data for a protective face mask detection system. Computer Optics 2022; 46(4): 603-611. DOI: 10.18287/2412-6179-CO-1039.

*Acknowledgements*: This work was supported by the Russian Foundation for Basic Research № 20-04-60529.

***Authors' information***

**Elena Vitalievna Ryumina** (b. 1991) graduated from ITMO University in 2021, majoring in Information Systems and Technologies. Currently she works as the junior researcher at the St. Petersburg Federal Research Center of the RAS (SPC RAS) in the Speech and Multimodal Interfaces Laboratory. Research interests are affective computing, digital image processing, visual signal recognition, automatic recognition of paralinguistic phenomena, machine learning, neural networks, biometric systems, human-machine interfaces. E-mail: *ryumina.e@iias.spb.su* .

**Dmitry Alexandrovich Ryumin** (b. 1991) graduated from Information Technologies and Programming faculty in 2016. He defended Ph.D. thesis on Models and Methods for Automatic Recognition of Russian Sign Language Elements for Human-Machine Interaction in ITMO University in 2020. He is a senior researcher of the Speech and Multimodal Interfaces Laboratory of the St. Petersburg Federal Research Center of the RAS (SPC RAS). Research interests are digital image processing, pattern recognition, automatic visual speech recognition, multimodal interfaces, machine learning, neural networks, biometrics, human-machine interfaces. E-mail: *ryumin.d@iias.spb.su* .

**Maxim Viktorovich Markitantov** (b. 1995) graduated from ITMO University in 2019, majoring in Software Engineering. Currently he works as the junior researcher at the St. Petersburg Federal Innovation Center of the RAS (SPC RAS). Research interests are artificial intelligence, machine learning, speech technologies, computational paralinguistics, recognition of the speaker's characteristics, speaker's age and gender recognition, detection of protective masks by audio information. E-mail: *m.markitantov@yandex.ru* .

**Alexey Anatolievich Karpov** (b. 1978) is Doctor of Technical Sciences (2013), Associate Professor (2012). Currently he works as the chief researcher and head of the Speech and Multimodal Interfaces Laboratory at the St. Petersburg Federal Research Center of the RAS (SPC RAS). Research interests are speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics. E-mail: *karpov@iias.spb.su* .

*Received September 3, 2021. The final version – October 27, 2021.*