УДК 004.421
DOI 10.17513/snt.40149

ДЕТЕКТИРОВАНИЕ ЗАЩИТНЫХ МАСОК В ПРОМЫШЛЕННОСТИ В ПОТОКОВОМ ВИДЕО НА МОДЕЛИ MOBILENET V3 SSD

Эвиев В.А., Лиджи-Гаряев В.В., Бадрудинова А.Н., Гермашева Ю.С., Абушинов О.А.

ФГБОУ ВО «Калмыцкий государственный университет имени Б.Б. Городовикова», Элиста, e-mail: goryaeff@mail.ru

В условиях пандемии важность готовности к чрезвычайным ситуациям стала очевидной, в частности необходимость ношения масок для снижения распространения инфекционных заболеваний. Цель данного исследования – разработка эффективного метода автоматического распознавания защитных масок для обеспечения соблюдения норм безопасности в таких отраслях, как строительство и химическое производство. Набор данных был создан на основе общедоступного набора изображений со строительными защитными масками и без них Construction Mask Dataset. Особое внимание уделяется интеграции технологий глубокого обучения в процессы мониторинга соблюдения мер безопасности. Это позволит значительно повысить уровень защиты работников и улучшить общий контроль за соблюдением стандартов. В исследовании были проведены эксперименты по разработке и усовершенствованию моделей глубокого обучения для обнаружения масок. Модель была реализована на платформе графического процессора (graphics processing unit) с использованием TensorFlow и библиотеки глубоких нейронных сетей. Для извлечения признаков использованы заранее обученные базовые модели, применяемые для выполнения задач классификации высокого качества. В качестве базовой модели для детектора Single-Shot Multi Box была выбрана сеть MobileNet V3 с весами, предобученными на наборе ImageNet. На тестовом наборе достигнута точность около 99%, а прирост скорости классификации на мобильных устройствах составил от 1,2 до 1,5 раз по сравнению с MobileNet V2. Результаты показывают, что предлагаемая модельная схема OpenCV + Deep Neural Networks + MobileNet с предварительно обученной конволюционной нейронной сетью является эффективным решением для обнаружения лиц в медицинских масках. Потери при обучении и валидации стремятся к нулю, что подтверждает точность и эффективность алгоритма для наборов данных более 5000 изображений.

Ключевые слова: нейросеть, MobileNet, компьютерное зрение, распознавание защитных масок, Single-Shot MultiBox

DETECTION OF PROTECTIVE MASKS IN INDUSTRY IN SWEAT VIDEO ON THE MOBILENET V3 SSD MODE

Eviev V.A., Lidzhi-Garyaev V.V., Badrudinova A.N., Germasheva Yu.S., Abushinov O.A.

Kalmyk State University named after B.B. Gorodovikov, Elista, e-mail: goryaeff@mail.ru

In the context of public health emergencies, the importance of preparedness has become increasingly evident, particularly the need for wearing protective gear. This study aims to develop an effective method for automatic recognition of construction masks to ensure compliance with safety standards in industries such as construction and chemical manufacturing. A dataset was created using a publicly available collection of images, both with and without construction masks, which is referred to as the “Construction Mask Dataset”. The study focuses on integrating deep learning technologies into safety monitoring processes to enhance worker protection and improve overall oversight of compliance with standards. Experiments were conducted on the development and improvement of deep learning models for mask detection, using a graphics processing unit (GPU) with TensorFlow and deep neural network libraries. Pre-trained base models were utilized for feature extraction to perform high-quality classification tasks. The MobileNet V3 network was chosen as the base model for the Single-Shot MultiBox Detector, achieving an accuracy of approximately 99% on the test set, with a classification speed increase on mobile devices ranging from 1.2x to 1.5x compared to MobileNet V2. The results indicate that the proposed model scheme of OpenCV + Deep Neural Networks + MobileNet is an effective solution for detecting faces in medical masks. The losses during training and validation tend towards zero, confirming the accuracy and effectiveness of the algorithm for datasets exceeding 5000 images.

Keywords: neural networks, MobileNet V3, computer vision, mask detection, Single-Shot Multi Box

Введение

Глубокое обучение продемонстрировало огромный потенциал в различных реальных приложениях, включая обнаружение объектов. Эта технология привела к многообещающим результатам в обнаружении объектов на изображениях, особенно в области строительства и безопасности трудах [1]. Обнаружение защитных строительных масок по-прежнему имеет решающее значение для обеспечения безопасности труда на строительных площадках, соблюдения гигиенических норм в строительных учреждениях и поддержания безопасности

в различных отраслях промышленности. Кроме того, достижения в области моделей глубокого обучения для обнаружения масок расширяют область компьютерного зрения, предлагая потенциальные возможности для обнаружения других видов средств индивидуальной защиты (СИЗ).

Целью исследования является анализ эффективности аннотирования и локализации объектов защитных строительных масок на лице работника в видеопотоке, на основе метрик точности для двух реализаций архитектуры MobileNet. Для решения поставленной цели необходимо решить следующие задачи:

Для достижения цели потребуется решить следующие задачи:

− выбор актуального набора данных для машинного обучения;

− предварительная обработка данных;

− обучение модели на основе аннотированного набора данных;

− определение оценочных показателей для измерения точности модели и расчет этих показателей для обеих реализаций архитектуры MobileNet;

− сравнение производительности этих реализаций архитектуры MobileNet с использованием оценочных показателей.

Материалы и методы исследования

Для обнаружения масок была использована сверточная нейронная сеть (CNN) с архитектурой MobileNet [2]. Скрипт был разработан с использованием библиотек Python, Tensorflow/Keras и OpenCV [3, 4]. Работа алгоритма обнаружения в потоковом видео основана на захвате кадров в виде входных изображений с заданными границами объектов. Метод прогнозирования объектов на изображении основан на известных режимах свертки. Для каждого пикселя на данном изображении оценивается набор ограничивающих рамок (обычно 4) разных размеров и соотношений сторон. Кроме того, для каждого пикселя вычисляется степень достоверности для всех возможных объектов, включая метку «Маски нет». Этот процесс повторяется для нескольких карт объектов. Для извлечения карт объектов используются предварительно обученные методы (базовые модели). Эти методы используются для решения задач классификации с высокой точностью. В качестве базовой модели для Single-Shot Multi Box (SSD) была использована сеть MobilNet версии 3. А, соответственно, ImageNet – это база данных изображений, которая была предварительно обработана на сотнях тысяч изображений, что отлично подходит для классификации изображений [5, 6]. Во время обучения предполагаемые границы сравниваются с фактическими. При обратном распространении параметры корректируются в соответствии с требованиями. Перед слоем классификации в модели MobilNet V3 добавляются слои объектов. Размер этих слоев постепенно уменьшается [7]. Каждое пространственное пространство объектов имеет ядро, которое выдает результаты, показывающие, существует объект или нет. Так же определяются размеры ограничивающей рамки. Из-за небольшого размера фильтров, применяемых к изображению, существует множество весовых параметров, которые в конечном итоге могут привести к повышению производительности. Последняя версия OpenCV включает модуль Deep Neural Network (DNN), который содержит предварительно обученную нейронную сеть (kCNN) для распознавания лиц [8, 9].

Набор данных был создан на основе общедоступного набора данных изображений со строительными защитными масками и без них Construction Mask Dataset [10, 11]. Для эксперимента было отобрано 8020 изображений, в том числе 4408 изображений с масками и 3612 изображений без масок. Эти изображения были использованы для обучения и тестирования модели с использованием мультисенсорного потока и современных методов распознавания объектов в CNN [12]. Набор данных был предварительно обработан с помощью функции preprocess_input в MobileNet V2. Базовая сеть была модифицирована путем добавления слоев: среднего объединяющего слоя, выравнивающего слоя, плотного слоя (128 единиц, активация ReLU), слоя с отсевом половины и последнего плотного слоя (2 единицы, активация сигмовидной кишки) [13, 14]. Общий процесс проиллюстрирован ниже.

Проект распознавания лиц по маскам разделен на две части. Обучение модели с помощью свертки или любой предварительно обученной модели для обнаружения масок на изображениях, что решает проблему бинарной классификации. Распознавание лиц на видео или изображениях и прогнозирование ношения масок с помощью обученной модели. Применялось обучение переносу с использованием предварительно обученных моделей MobileNet. Сеть использует разделяемые по глубине свертки, ее основной структурный блок показан на рис. 1. Этот блок включает в себя три сверточных слоя, последние два из которых являются глубинными свертками, фильтрующими входные данные, за которыми следует слой поточечной свертки размером

1x1. В отличие от V1, V3 уменьшает количество каналов в поточечной свертке 1x1, известной как проекционный слой, уменьшая размерность. Например, слой по глубине может обрабатывать тензор со 144 каналами, которые проекционный слой уменьшает до 24 (рис. 2). Этот уровень также называется уровнем узких мест, поскольку он уменьшает объем данных, проходящих через сеть. Свертка по глубине применяет свои фильтры к тензору. Наконец, проекционный слой преобразует 144 отфильтрованных канала в меньшее число, например в 24. Входные и выходные данные блока представляют собой тензоры низкой размерности, в то время как фильтрация выполняется по тензору высокой размерности. Остаточные соединения, аналогичные ResNet, помогают управлять градиентным потоком в сети, при этом каждый слой имеет пакетную нормализацию и активацию ReLU6. Однако выходные данные проекционного слоя не имеют функции активации, что приводит к получению данных малой размерности.

F-Measure объединяет точность и отзывчивость в единую метрику, отражающую оба свойства. В качестве альтернативы точности классификации обычно используют показатели прецизионности и отзыва (табл. 1).

Для анализа регрессионной модели на соответствие набору данных используется среднеквадратичная ошибка (RMSE):

$$RMSE = \sqrt{\sum \frac{\left(P_i - O_i\right)^2}{n}},$$

где P_i – это прогнозируемое значение в наборе данных с n фото, O_i – наблюдаемое значение для i-го наблюдения в наборе данных.

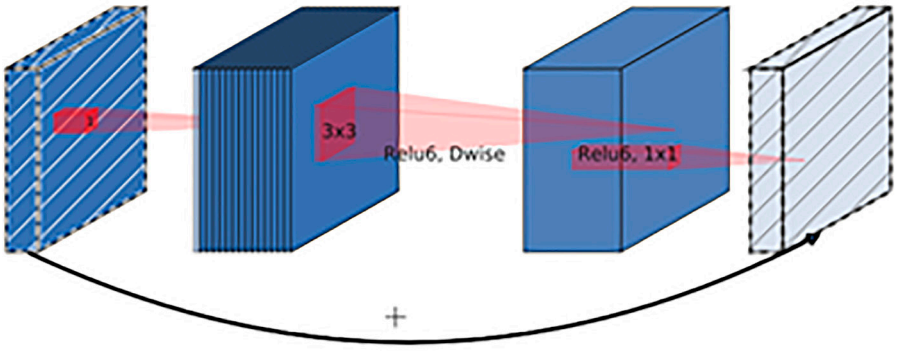

Рис. 1. Слои и свертки MobileNet V2

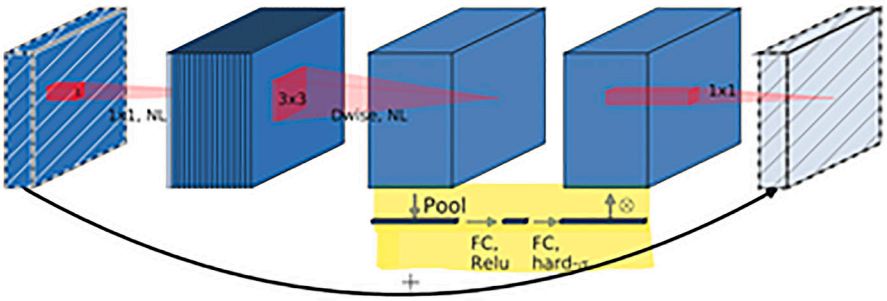

Рис. 2. Слои и свертки MobileNet V3

Таблица 1

Формулы оценки

	Accuracy	Precision	Recall	F1
Формула	$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$\frac{TP}{TP+\frac{1}{2}(FP+FN)}$
Python sklearn.metrics	accuracy_score	precision_score	recall_score	f1_score

Результаты исследования и их обсуждение

Архитектура MobileNet V2 представлена инвертированными остаточными блоками (IRB), которые заменяют традиционные остаточные блоки более эффективной структурой. MobileNet V3 еще больше повысил эффективность благодаря модифицированной архитектуре инвертированного остаточного блока (MIRB), которая добавляет блок сжатия и возбуждения (SE) для повторной калибровки откликов функций в зависимости от канала (табл. 2).

Помимо извлечения объектов модель может быть расширена для обнаружения объектов и сегментации. Простой мерой оценки модели является точность классификации (общее количество правильных прогнозов, деленное на общее количество прогнозов), но она не подходит для задач несбалансированной классификации, подобных текущей. Чтобы оценить производительность классификатора, для измерения производительности моделей классификации, целью которых является предсказание категориальной метки для каждого входного экземпляра, необходимо изучить матрицы производительности (табл. 3).

Загрузка весов модели и создание объекта MobileNet V3 выполняется с использованием tf.keras.applications.mobilenet_v3. В результате тест показал высокий результат:

93s 465ms/step – loss: 0.0085 – accuracy: 0.9958 – lr: 1.0000e-04

Таблица 2

Сети MobileNet V2 и MobileNet V3

Архитектура MobileNet V2				Архитектура MobileNet V3			
Тип / Этап	Входной размер	c	s	Тип / Этап	Входной размер	c	s
Conv d2	224 × 224 × 3	32	2	Conv d2	224 × 224 × 224 × 3	16	2
Bneck	112 × 112 × 32	16	1	Bneck, 3 × 3	112 × 112 × 16	16	1
Bneck	112 × 112 × 16	24	2	Bneck, 3 × 3	112 × 112 × 16	64	2
Bneck	56 × 56 × 24	32	2	…	…		
Bneck	28 × 28 × 32	64	2	Bneck, 5 × 5	14 × 14 × 112	160	2
Bneck	14 × 14 × 64	96	1	Bneck, 5 × 5	7 × 7 × 160	160	1
Bneck	14 × 14 × 96	160	2	Bneck, 5 × 5	7 × 7 × 160	160	1
Bneck	7 × 7 × 160	320	1	Conv d2, 1 × 1	7 × 7 × 160	960	1
Conv d2	7 × 7 × 320	1280	1	Conv / s1	7 × 7 × 960	-	1
Avg Pool 7 × 7	7 × 7 × 1280	-	-	Conv d2, 1 × 1, NBN	1 × 1 × 960	1280	1
Conv d2 1 × 1	1 × 1 × 1280	k		Conv d2, 1 × 1, NBN	1 × 1 × 1280	k	1

Примечание: S определяет, на сколько пикселей сдвигается окно свертки (фильтр), c – коэффициент, указывает, во сколько раз увеличивается количество каналов в первом слое блока.

Таблица 3

Матрица путаницы

	Позитивный прогноз	Отрицательный прогноз
Положительный класс	Истинно положительный (TP)	Ложноотрицательный (FN)
Отрицательный класс	Ложноположительный (FP)	Истинно отрицательный (TN)

Таблица 4

Сравнение оптимизаторов SGM и ADAM

Оптимизатор	Эпохи	Итерации	Прошедшее время	Мини-пакет (RMSE)	Валидация (RMSE)	Потери мини-пакетов	Потери на тесте
SGM	25	100	0:16:03	0,88	0,87	0,8025	0,7917
	50	200	0:44:21	0,73	0,82	0,5420	0,6484
Adam	25	100	0:23:55	0,59	0,69	0,3561	0,4956
	50	200	0:46:01	0,52	0/70	0,2656	0,4936

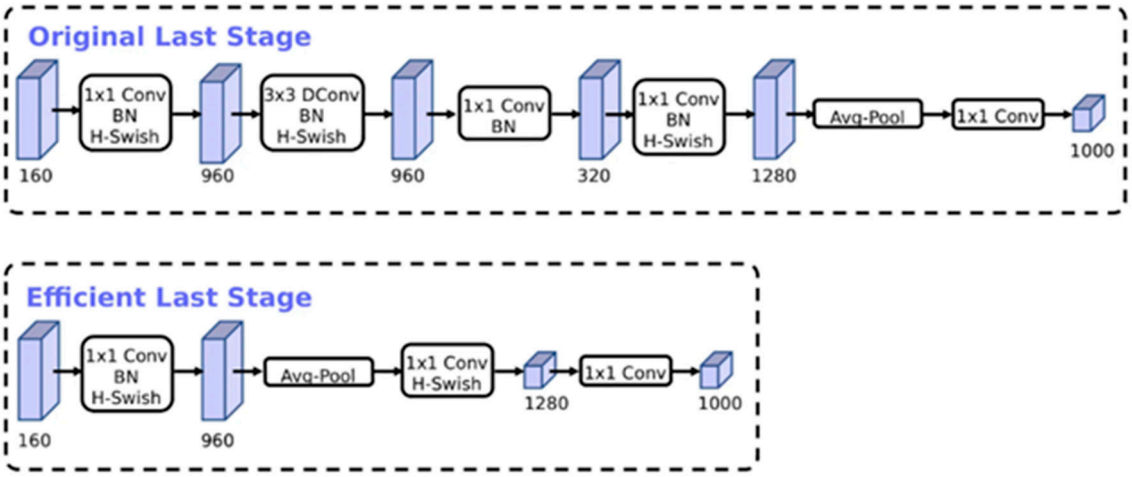

Рис. 3. Нейросеть MobileNet V3

Набор данных разделен на 80% обучающих, 10% валидационных и 10% тестовых изображений. Начальная скорость обучения составляет 0,001, количество эпох – 50. В качестве методов оптимизации использованы ADAM и градиентный спуск (SGM) для сравнения производительности (табл. 4).

Проверка модели в реальном времени проводилась на изображениях двух классов (маска и отсутствие маски). Результаты показали, что SGM достигает меньших затрат времени на обучение по сравнению с ADAM. Первый сверточный слой модели включает слой BN и слой активации h-переключателя. Средняя часть содержит сверточные слои (MB) с узкими местами и SE-структуры сжатия и возбуждения, что уменьшает общий объем вычислений. MobileNet V3 использует больше параметров из-за модуля SE, однако это компенсируется улучшенной точностью и скоростью. SE и h-swish слегка замедляют работу сети, добавляя некоторые задержки, однако это приемлемо ввиду повышения точности. H-swish используются на более глубоких уровнях, где тензоры меньше, что снижает задержку. Использование стандартного SSDLite-MobileNet V2 расширяет последнюю свертку 1x1 с глубины 320 до 1280, что неактуально для текущей задачи с двумя классами (рис. 3).

Сравнение метрик эффективности архитектур MobileNet V2 и V3

Для оценки эффективности архитектур MobileNet V2 и V3 в распознавании медицинских масок использованы данные, показывающие улучшения в производительности MobileNet V3 (табл. 5).

Использование HardSwish и модуля SE улучшает точность и полноту, общий F1-Score также высок. MobileNet V3 демонстрирует значительное увеличение скорости распознавания благодаря новым разделяемым по глубине сверткам и уменьшению числа параметров.

На устройстве Google Pixel MobileNet V3 обеспечивает 1,5-кратное ускорение задач классификации изображений по сравнению с MobileNet V2. На Samsung Galaxy S10 MobileNet V3 ускоряет задачи обнаружения объектов в 2,5 раза.

Таблицы путаницы для MobileNet V3 показывают точность ~99% на тестовом наборе данных.

Таблица 5

Метрики производительности моделей

Модель	Категории	precision	recall	f1-score	Support
MobileNet V2	С маской	0,96	0,98	0,98	882
	Без маски	0,98	0,99	0,99	722
MobileNet V3	С маской	0,97	0,99	0,98	882
	Без маски	0,99	0,99	0,99	722

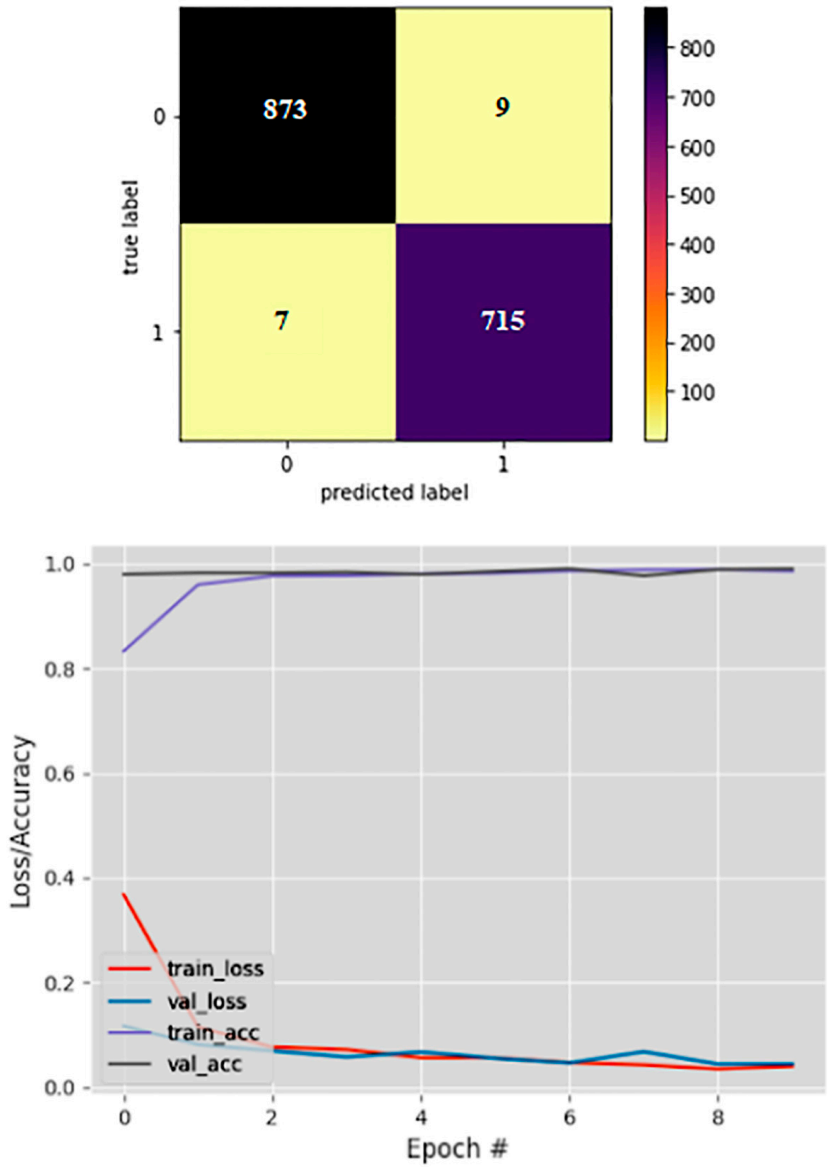

Рис. 4. Матрица путаницы и график модели потерь

Если точность высока, а потери невелики, то модель допускает небольшие ошибки только в некоторых данных (рис. 4), но при этом наблюдаются признаки переобучения: потери при проверке ниже, чем при обучении.

Для обнаружения защитных строительных масок в потоковом видео была использована архитектура OpenCV DNN вместе с предварительно обученной моделью MobileNet SSD V3 на основе каскадного классификатора для получения высокопроизводительных результатов. Оценку f1 используют в тех случаях, когда нужно иметь как хорошую точность, так и хорошую отзывчивость. Это указывает на то, что использование оценки f1 существенно, когда есть заинтересованность в получении наибольшего количества истинных положительных результатов, и при этом мы хотим быть более уверенными, что текущий прогноз верен.

Заключение

Целью исследования был анализ эффективности по распознаванию защитных строительных масок на лице работника в потоковом видео. В процессе обучения и проверки детектора в контролируемом состоянии был использован набор данных на основе общедоступного набора данных с лицами в масках и без них. Кроме того, в экспериментах были изучены такие показатели производительности, как средняя оценка частоты про-

махов по логарифму. На тестовом наборе получена точность ~99 %, при этом можно наблюдать, что есть небольшие признаки переобучения, при этом оптимизатор Adam показал потери при проверке ниже (0,3113), чем у SGM (0,3847). При этом полученные результаты показывают, что предлагаемая модельная схема OpenCV DNN + MobileNet SSD V3 с предварительно обученной конволюционной нейронной сетью (kCNN) для обнаружения лица в защитной строительной маске является эффективной моделью для обнаружения лица в защитной строительной маске. MobileNet V3 представляет собой значительное улучшение по сравнению с MobileNet V2, предлагая лучшую точность и скорость благодаря новым архитектурным решениям и оптимизациям. Эти улучшения делают MobileNet V3 предпочтительным выбором для мобильных и встроенных приложений, требующих высокой производительности и эффективности. Объединение в единую пространственную пирамиду (LR-ASP) позволяет достичь новых результатов в области мобильной классификации, обнаружения и сегментации. MobileNet V3-Large на 3,2 % точнее в классификации ImageNet при одновременном снижении задержки на 20 % по сравнению с MobileNet V2 и работает на 25 % быстрее, чем MobileNet V2 R-ASPP, при аналогичной точности сегментации.

В целом задачи исследования решены, полученные результаты подтверждают эффективность предлагаемой модели для обнаружения защитных строительных масок в потоковом видео, и можно надеяться, что она будет полезна для различных приложений в области строительства и безопасности труда.

Список литературы

1. Adegun A.A., Viriri S., Tapamo J.R. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis // Journal of Big Data. 2023. Vol. 10, Is. 1. P. 93–97.

2. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. DOI: 10.1109/CVPR.2016.90.

3. Глубокое обучение с подкреплением: теория и практика на языке Python // Системный администратор. 2021. № 12 (229). С. 72–94.

4. Моделирование нейронных сетей в пакетах Keras и Tensorflow: методические указания к лабораторным работам / Сост. С.М. Наместников. Ульяновск: УлГТУ, 2021. С. 1–14.

5. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, Is. 6. P. 1137–1149.

6. Rusakovsky J. et al. Imagenet large scale visual recognition challenge // International Journal of Computer Vision. 2015. Vol. 115, Is. 3. P. 211–252.

7. Sinha D. Thin MobileNet: An Enhanced MobileNet Architecture // IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference. 2019. P. 0280–0285.

8. Русанова Е.Г. Обзор методов классификации эмоций человека для задач распознавания эмоций // Политехнический молодежный журнал. 2022. № 8. С. 8–10.

9. Горяев В.М., Басангова Е.О. Исследование производительности различных моделей машинного обучения при неинвазивном измерении артериального давления на основе сигналов PPG и ЭКГ // Вестник Башкирского университета. 2023. № 1. С. 36–44.

10. Construction site Safety Image Dataset. [Электронный ресурс]. URL: https://agie.ai/datasetsdetails/construction-safety-object-detection (дата обращения: 21.07.2024).

11. Goryaev V.M., Basangova E.O. et al. Forecasting steppe fires using remote sensing data of time series // IOP Conference Series: MSE. Vol. 1047, Is. 1. 2021. P. 12092–12098.

12. Goryaev V.M. et al. Development of a statistical forecast model to improve accuracy based on statistical analysis of weather historical data for the Kalmyk region // IOP Conference Series: Earth and Environmental Science. 2019. Vol. 350. P. 012058–012064.

13. Dumoulin J., Houshmand P., Jain V., Verhelst M. Enabling Efficient Hardware Acceleration of Hybrid Vision Transformer (ViT) Networks at the Edge // IEEE International Symposium on Circuits and Systems (ISCAS). 2024. P. 1–5.

14. Goryaev V.M., Uchurova E.O., Basangova E.O., Bembitov D.B., Miloshenko A.P. Analysis of digital filters for preprocessing biomedical signals from ECG apparatus // AIP Conference (MIST: Aerospace-IV). AIP pablishing, 2023. Vol. 2700. P. 050037–050043. DOI: 10.1063/5.0125057.