

6. Li N. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction / N. Li, P. C. Loizou. — Text : direct // J. Acoust. Soc. Am. — 2008. — Vol. 123, № 3. — P. 1673-1682.
7. Brungart D. S. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation / D. S. Brungart, P. S. Chang, B. D. Simpson, D. Wang. — Text : direct // J. Acoust. Soc. Am. — 2006. — Vol. 120, № 6. — P. 4007-4018.
8. Zermini A. Deep Learning for Speech Separation: PhD thesis / A. Zermini. — Text : direct // University of Surrey, faculty of engineering, physical sciences, Centre for Vision. — Speech : Signal Processing (CVSSP), South East of England, UK, 2020.
9. Xia S. Using Optimal Ratio Mask as Training Target for Supervised Speech Separation / S. Xia, H. Li, X. Zhang. — Text : direct // 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). — Kuala Lumpur, 2017. — P. 163-166.
10. Allen J. B. Image method for efficiently simulating small-room acoustics / J. B. Allen, D. A. Berkley. — Text : direct // The Journal of the Acoustical Society of America. — 1998. — Vol. 65, № 4. — P. 943-950.

М. А. УСТЕЛЕМОВ, М. С. ВОРОБЬЕВА

Тюменский государственный университет, г. Тюмень

УДК 004.94

ИССЛЕДОВАНИЕ ПОДХОДОВ К КЛАССИФИКАЦИИ НАЛИЧИЯ СРЕДСТВ ИНДИВИДУАЛЬНОЙ ЗАЩИТЫ ПО КАДРУ ЛИЦА НА ОСНОВЕ АНАЛИЗА ВИДЕОПОРЯДА МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В статье отражено исследование методов машинного обучения, которые могут служить компонентами автоматизированной системы выявления лиц, не использующих средства индивидуальной защиты.

Ключевые слова: fine-tuning, transfer learning, data augmentation, видеопоряд, классификация, машинное зрение, трекинг объектов, набор данных, определение лица, средства индивидуальной защиты.

Введение. Работодатель, в рамках организации контроля трудовых условий, обязан обеспечивать ношение и правильное применение средств индивидуальной и коллективной защиты на территории организации [1]. Обеспечение контроля применения средств индивидуальной защиты (далее — СИЗ) зачастую выполняется сотрудниками службы безопасности, совмещающими данную функцию с другими своими обязанностями по обеспечению контроля правопорядка.

Проводимый ручной контроль ношения посетителями СИЗ приводит к увеличению нагрузки на сотрудников службы безопасности и возможности невыявления нарушения. Использование методов машинного обучения для анализа видеоряда с камер наблюдения позволит автоматизированно фиксировать факты нарушения.

Проблема исследования. Существующие подходы к определению наличия СИЗ на видеокадре лица используют наборы данных с изображениями, условия подготовки которых не соответствуют входным данным решаемой задачи: качество, размер и ракурс съемки изображений лиц [2, 3]. Для проведения исследования подходов к автоматизированному контролю наличия СИЗ методами машинного обучения потребуется подготовить набор изображений лиц, соответствующих условиям проводимого контроля.

Для подготовки набора данных следует использовать инструменты отслеживания объектов и определения лица, которые в дальнейшем могут применяться для проведения автоматизированного контроля. На основе подготовленного набора данных требуется разработать классификатор наличия СИЗ на видеокадре лица.

Цель исследования — разработать классификатор наличия СИЗ на видеокадре лица на основе анализа видеоряда методами машинного обучения.

Материалы и методы. Входными данными для проведения исследования является видеоряд, получаемый с камер видеонаблюдения, установленных на контрольно-пропускных пунктах институтов ТюмГУ. Формат видеоряда — MJPEG-видеопоток с частотой обновления 14 FPS, размерность изображения — 1280×720 пикселей (720 p).

Областью интереса данных в используемом видеоряде будет являться проход через турникеты, размерность области — 400×400 пикселей.

Результатом классификации наличия СИЗ на кадре лица является:

- изображение лица нарушителя режима ношения СИЗ на кадре: формат — JPEG, минимальный размер — 128×128 пикселей, каждый факт нарушения соответствует одному изображению лица нарушителя;
- вероятность принадлежности к классу для возможных дальнейших доработок модели классификации;
- время обнаружения лица относительно времени получения входных данных.

На рис. 1 приведен конвейер данных при проведении анализа видеоряда, выделены следующие компоненты:

- компонент получения данных: чтение следующего кадра, изменение размеров и компоновки субпикселей;
- компонент детектирования лица на изображении: выделение объектов на изображении, выделение лиц, соответствующих пороговым значениям, при помощи подавления немаксимумов;
- компонент сопровождения объектов: обновление результатов детектирования на основе границ лица, удаление из отслеживания объектов с необновленной детекцией;
- компонент хранения данных по лицам: хранение во внутреннем буфере данных по каждому идентификатору лица, при удалении объекта из отслеживания — выбор из буфера элемента с наибольшим размером изображения и наилучшей оценкой детекции лица;
- компонент определения наличия маски: изменение размеров и компоновки субпикселей детектированного лица, отнесение изображения лица к классу наличия или отсутствия СИЗ;
- компонент хранения и просмотра данных о нарушителях: сохранение результатов для лиц с классом отсутствия СИЗ.

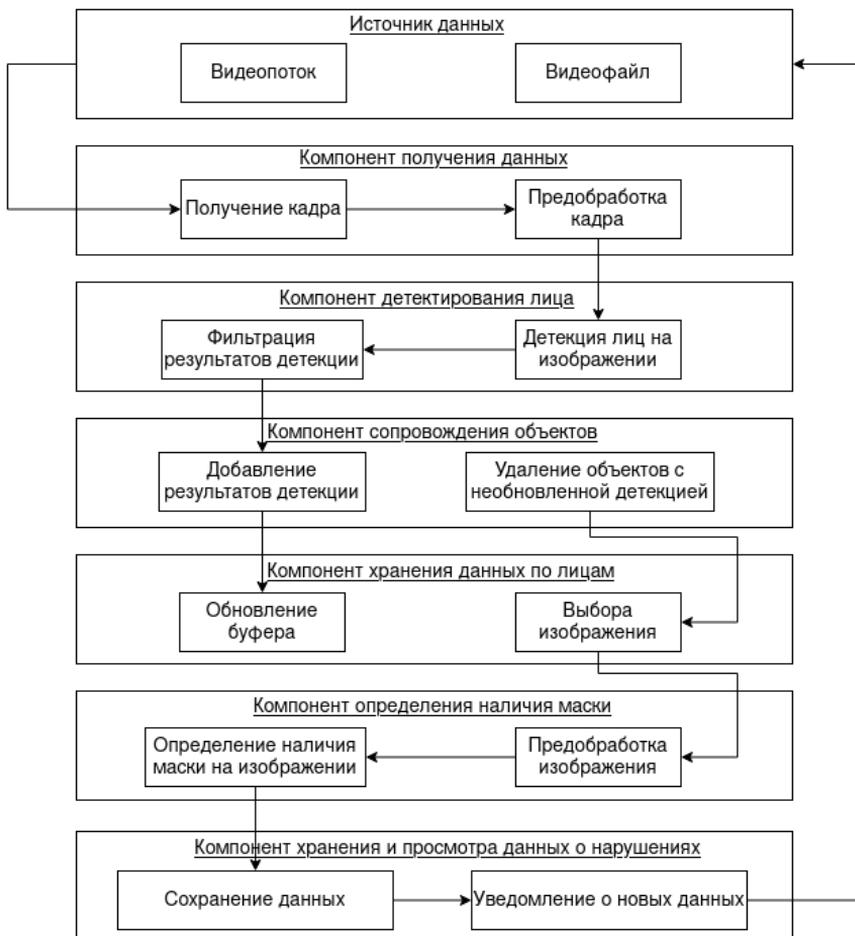


Рис. 1. Конвейер данных при проведении анализа видеоряда для классификации наличия СИЗ на кадре лица

Сопровождение движущегося объекта (Object Tracking, «трекинг») — задача компьютерного зрения, заключающаяся в определении одного и того же объекта на каждом кадре видеоряда.

Преимущества использования алгоритма отслеживания объекта:

- классификаторы детектируемых объектов могут запускаться несколько раз на одном объекте, и результирующее решение будет приниматься по самой точной оценке;
- возможность сбора статистики по точному количеству объектов;
- бóльшая устойчивость к ошибочным детекциям объекта (объекты с количеством детекций меньше порогового значения могут быть отброшены).

В качестве инструмента отслеживания объекта на видеопотоке был выбран алгоритм SORT. Авторы A. Bewley, Z. Ge, L Ott [et al.] отмечают, что данный алгоритм имеет лучшие значения точности и скорости работы в совокупности по сравнению с другими существующими алгоритмами [4].

Входными данными для решения задачи бинарной классификации наличия СИЗ у человека являются результаты детектирования лиц на кадре. Существует два подхода к детектированию лица, совместимых с определением трекинга объекта:

1) вычисление координат границ лица на основе координат ключевых точек тела — использование детектора позы человека (Pose Detector);

2) получение координат границ лица как результат работы детектора лица (Face Detector).

Основные требования к работе инструмента детектирования лица — это скорость работы, приближенная к частоте получения кадров источника данных, и минимизация прерываний трекинга лиц, что определяется возможностью модели детектировать лицо при различных углах поворота.

В качестве детектора позы человека была выбрана модель LightweightOpenPose (tiny VGG) [5], детектора лица — YOLOv5s библиотеки YOLO5Face [6].

Были проведены замеры скорости работы моделей HyperPose и YOLOv5s на CPU Intel(R) Xeon(R) @ 2.20GHz и GPU Tesla K80. Размеры изображений:

- 346×342 пикселя для HyperPose,
- 320×320 пикселей для YOLOv5s.

Результаты замеров скорости и сравнение моделей по их основным характеристикам приведены в табл. 1, 2. Полученные расчеты являются средними значениями за 100 операций.

Исходя из проведенного анализа, для разработки компонента детектирования лица была выбрана модель YOLOv5s.

При подготовке набора данных для разработки компонента определения наличия маски (см. рис. 1) были определены два класса для разметки объектов датасета: «mask» и «no_mask».

Таблица 1

Сравнительный анализ скорости работы HyperPose и YOLOv5s

<i>Тип замера скорости работы</i>	<i>HyperPose CPU скорость, мс / FPS</i>	<i>HyperPose GPU скорость, мс / FPS</i>	<i>YOLOv5s CPU скорость, мс / FPS</i>	<i>YOLOv5s GPU скорость, мс / FPS</i>
Один проход модели	1383,48 / 0,72	11,99 / 83,4	103,06 / 9,7	29,03 / 34,45
Постобработка	218,58 / 4,58	125,82 / 7,95	Не требуется	Не требуется
Полная обработка изображения	1627,44 / 0,61	156,18 / 6,4	108,81 / 9,19	34,31 / 29,12

Таблица 2

Сравнительный анализ HyperPose и YOLO5Face

<i>Характеристика</i>	<i>HyperPose</i>	<i>YOLO5Face</i>
<i>1</i>	<i>2</i>	<i>3</i>
Вычисление координат лица	На основе определения 17 ключевых точек тела человека	Как результат работы модели
Устойчивость детектирования при положении лица не анфас	Возможность детектирования полностью отвернутого от камеры лица. Чтобы определить отвернуто ли лицо, можно использовать оценку точности детектирования точек лица	Возможность детектирования лица при повороте в профиль. Чтобы определить повернуто ли лицо в профиль, можно использовать оценку точности детектирования лица и размер его границ

1	2	3
Устойчивость детектирования лица при его близости к объективу камеры	Невозможность вычисления границ лица при отсутствии детектирования ключевых точек тела, не относящихся к лицу	Устойчивое детектирование лица при его попадании в кадр
Скорость работы на GPU, FPS	6,4	29,12

Правила присвоения класса:

- прикрыты нос и рот человека — маска надета (класс «mask»),
- прикрыт только рот — маска надета (класс «mask»),
- не прикрыты рот и нос — маска не надета (класс «no_mask»),
- прикрыт только нос — маска не надета (класс «no_mask»).

В иных случаях считается, что у человека отсутствует СИЗ.

Для подготовки датасета был обработан видеопоток с камеры наблюдения общей длительностью 9:29:23. Количество обработанных кадров: 34715. Реальное обработанное время: 18:54:30. Датасет был сбалансирован путем случайного выбора элементов из класса с бóльшим числом объектов до количества элементов класса с меньшим количеством объектов.

Размер итогового датасета — 2100 объектов, из них 1050 соответствует классу «mask», 1050 — классу «no_mask».

Результаты. Для решения задачи классификации был использован подход BigTransfer (BiT) [7]: State-of-the-art Transfer, ставящий своей целью получения более точных результатов классификации для новых датасетов даже с небольшим количеством данных по классу. Для разработки была выбрана ViT-R50x1 модель как наиболее точная и производительная в соответствии с характеристиками в [8].

Для fine-tuning модели ViT-M-R50x1 использовался размер входного слоя (размер изображения) 128×128 пикс. Последний Dense-слой модели заменен на Dense-слой с размерностью 2 по числу классов решаемой задачи.

Параметры обучения:

- количество шагов — 500 : 50 эпох по 10 шагов в эпохе;
- batch size: 512 изображений;
- размер обучающей выборки увеличен для соответствия его размерности количеству шагов путем повторения;
- размер тестирующей выборки увеличен в 5 раз путем повторения;
- оптимизатор — Adam со стандартным начальным значением learning rate — 0,001;
- функция потерь — SparseCategoricalCrossentropy для работы с числовыми метками классов.

Разделение датасета на обучающую и тестовую выборки проводилось случайным образом: доля обучающей выборки от исходных данных — 90%, доля тестовой — 10%.

Для повышения устойчивости модели к освещению и увеличению разнообразия обучающей выборки к изображениям применялись элементы data augmentation.

Результаты обучения модели:

- итоговая accuracy на обучающей выборке: 0,9641;
- итоговая accuracy на валидирующей выборке: 0,9652;
- итоговая precision: 0,9583;
- итоговая recall: 0,9787;
- время обучения составило 16 минут на GPU Tesla T4.

Графики метрик модели по ходу обучения представлены на рис. 2. По графикам видно, что происходит последовательное уменьшение значения функции потерь и увеличение точности, к завершающей эпохе выходящее к итоговому граничному значению.

Модель показала высокий результат точности с учетом отличающейся выборки на каждом шаге из-за выполняемой предобработки. Скачки значений точности для обучающей выборки объясняются числом шагов в эпохе и размером batch size, однако последовательное увеличение точности модели говорит об отсутствии переобучения.

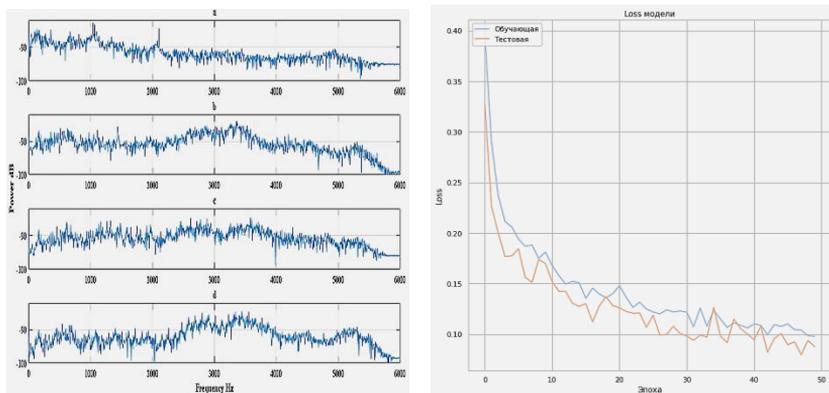


Рис. 2. Метрики модели по ходу обучения

Заключение. Определен конвейер данных для проведения анализа видеоряда при классификации наличия СИЗ на кадре лица, рассмотрены методы сопровождения объектов и детектирования лица, подготовлен набор данных и на основе fine-tuning подхода разработан классификатор с итоговой ассурасу на валидирующей выборке 0,9652.

СПИСОК ЛИТЕРАТУРЫ

1. Российская Федерация. Законы. Трудовой кодекс Российской Федерации : текст с изменениями и дополнениями на 1 марта 2022 г. : [принят Государственной думой 21 декабря 2001 г.: одобрен Советом Федерации 26 декабря 2001 г.]. — Текст : электронный // КонсультантПлюс : [сайт]. — URL: http://www.consultant.ru/document/cons_doc_LAW_34683/ (дата обращения: 26.05.2022).
2. Real-World-Masked-Face-Dataset. — Text : electronic // Github : [website]. — URL: <https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset> (date of the application: 26.05.2022).
3. Face Mask Detection Dataset. — Text : electronic // Kaggle : [website]. — URL: <https://www.kaggle.com/datasets/andrewmvd/face-mask-detection> (date of the application: 26.05.2022).
4. Bewley A. Simple online and realtime tracking / A. Bewley, Z. Ge, L. Ott [et al.]. — Text : electronic // Proc. of IEEE International Conference on Image Processing (ICIP). — 2016. — 5 p. — URL: <https://ieeexplore.ieee.org/document/7533003> (date of the application: 26.05.2022).

5. Guo Y. Fast and Flexible Human Pose Estimation with HyperPose / Y. Guo, J. Liu, G. Li [et al.]. — Text : electronic // MM '21: Proc. of the 29th ACM International Conference on Multimedia. — 2021. — P. 3763-3766. — URL: <https://dl.acm.org/doi/10.1145/3474085.3478325> (date of the application: 26.05.2022).
6. Qi D. YOLO5Face: Why Reinventing a Face Detector / D. Qi, W. Tan, Q. Yao, J. Liu. — Text : electronic // arXiv : [website]. — 2022. — URL: <https://arxiv.org/abs/2105.12931> (date of the application: 26.05.2022).
7. Kolesnikov A. Big Transfer (BiT): General Visual Representation Learning / A. Kolesnikov, L. Beyer, X. Zhai [et al.]. — Text : electronic // arXiv : [website]. — 2019. — URL: <https://arxiv.org/abs/1912.11370> (date of the application: 26.05.2022).
8. cnn-benchmarks. — Text : electronic // Github : [website]. — URL: <https://github.com/jcjohnson/cnn-benchmarks> (date of the application: 26.05.2022).

Д. А. ЗАХАРОВ, И. Г. ЗАХАРОВА

Тюменский государственный университет, г. Тюмень

УДК 004.93

РАЗРАБОТКА СИСТЕМЫ МОНИТОРИНГА СОСТОЯНИЯ ПОКРЫТИЯ ПРОЕЗЖЕЙ ЧАСТИ

***Аннотация.** В статье представлены описание методов обнаружения решения дефектов дорожного покрытия и их программная реализация в рамках системы мониторинга состояния покрытия проезжей части. В качестве основного подхода к решению выбраны создание и применение моделей машинного обучения на базе искусственных нейронных сетей.*

***Ключевые слова:** обнаружение ям, система мониторинга, машинное обучение, нейронные сети, YOLO.*

Введение. С каждым годом растет количество автомобилей и пропорционально этому увеличивается и количество аварий. При этом во многих случаях непосредственной причиной могут послужить серьезные дефекты дорожного покрытия, в частности ямы различных размеров и конфигурации.