Е. А. Лемберг, В. Н. Веревкин, Ю. А. Плотоненко
Тюменский государственный университет, г. Тюмень
УДК 004.93'1

ИСПОЛЬЗОВАНИЕ СВЕРХТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ПОДОЗРИТЕЛЬНОЙ АКТИВНОСТИ

Аннотация. В статье представлены подходы к решению проблемы обнаружения нарушений во время проведения единого государственного экзамена. Были реализованы YOLO и ResNet архитектуры искусственного интеллекта для выделения областей с учениками и определения подозрительной активности.

Ключевые слова: машинное обучение, модели машинного обучения, компьютерное зрение, средства обработки видеоинформации.

Введение. В современном обществе обеспечение честности процесса проведения государственных экзаменов становится актуальной задачей. Проблему подтверждает статистика Рособрнадзора за 2023 год. По статистике более 1000 человек были уличены в использовании шпаргалок, средств связи и т. д. Из этого количества учащихся с пункта проведения ЕГЭ было удалено 54%. Остальных удалось выявить путем просмотра записей с камер видеонаблюдений уже после проведения ЕГЭ.

В рамках данного проекта предлагается подход к решению этой проблемы с использованием искусственного интеллекта. Проект направлен на помощь наблюдателям Единых государственных экзаменов (ЕГЭ) в выявлении подозрительной активности участников с целью усиления системы контроля, обеспечения честности экзаменационного процесса.

Система будет состоять:

- из модели машинного обучения для обнаружения людей;
- модели машинного обучения для определения подозрительной активности.

Работы на данную тему уже проводились. Tong Lui разрабатывал схожую систему по отслеживания подозрительной активности на офлайн-экзаменах [1]. В его работе был представлен подход с использованием двух моделей. Одна модель сверточной нейронной сети образовывала картинки со всеми людьми бинарно классифицируя каждый отдельный квадрат на наличие на нем человека. Далее вторая сверточная нейронная сеть определяет, есть ли признаки нарушений в данном квадрате. Хотя этот метод показал отличные результаты в ходе проверок на тестовой выборке, 91% точности, он имеет несколько серьезных недостатков:

- Заранее определенные размеры областей, в которых должен быть зафиксирован полный образ ученика, включая руки, ноги и голову, приводят к тому, что для каждой аудитории должен быть размер области, в которую бы помещался полностью ученик. Из-за этого ошибки или особенности телосложения учеников сильно влияют на точность модели.
- Так как классификация проходит для каждой области отдельно, то требуется достаточно много ресурсов для работы. В статье указано, что изображение может быть просмотрено раз в 5-10 секунд.

Также схожую работу проделывала команда, состоящая из четырех разработчиков с темой "AUTOMATIC CHEATING DETECTION IN EXAM HALL" [2]. Работа проводилась с использованием только YOLO-opencv. Достигнутая точность — 86%.

Проблема исследования. Основной проблемой, с которой сталкиваются наблюдатели, является сложность в выявлении подозрительной активности в реальном времени. Среди таких инцидентов могут быть использование запрещенных учебных материалов и электронных приборов.

Цель данного проекта состоит в создании системы которая, используя методы машинного обучения, способна выявлять подозрительные ситуации в ходе проведения государственных экзаменов.

Для достижения этой цели были поставлены следующие задачи:

- Реализовать модели для обнаружения участников ЕГЭ.
- Реализовать модели для обнаружения подозрительной активности участников ЕГЭ.
- Предобработать, разметить и форматировать данные для обучения и тестирования моделей.
- Реализовать связи между моделями.

Материалы и методы. Были получены не размеченные данные (видео) с официально проводимых государственных экзаменов. Данные предоставил Региональный центр обработки информации Тюменской области. Результатом работы над материалом будет размеченный датасет.

Далее была проведена работа над полученными данными:

- Проведена выборка для обнаружения записей с подозрительной активностью.
- Осуществлена предобработка данных, включающая в себя: выделения объектов на видео, изменение размеров исходного видео.

Математическая постановка задачи модели обнаружения участников ЕГЭ:

Пусть I — входное изображение размером W×H×C, где W и H — ширина и высота изображения соответственно, а C — количество каналов. Есть конечно множество классов $H = \{H_1, H_2\}$, где H_1 — говорит о наличии человека на изображении, а H_2 — говорит об его отсутствии. Задача обнаружения участников ЕГЭ на изображении состоит в том, чтобы для каждого объекта I (изображение W×H×C) определить его принадлежность к определенному классу из множества H.

Далее входное изображение I разбивается на сетку из S×S ячеек. Каждая ячейка сетки отвечает за предсказание BB прямоугольных рамок и вероятностей принадлежности объекта к каждому из C классов. Для каждой ячейки предсказываются: Bounding boxes с параметрами x, y, w, h, где x и y — координаты верхнего левого угла bounding box'a относительно координат ячейки, а w и h — ширина и высота bounding box'a. Вероятности P(object) и P(classi|object), где P(object) определяет наличие объекта в bounding box'e, а P(classi|object) — вероятность принадлежности объекта к классу i.

Математическое описание обучение модели обнаружения участников ЕГЭ:

Пусть $D = \{(I_1, B_1, C_1), (I_2, B_2, C_2), ..., (I_N, B_N, C_N)\}$ — обучающий набор данных, где I_i — входное изображение, B_i — истинные bounding boxes для объектов на изображении I_i, а C_i — классы объектов. Для каждого изображения I_i модель YOLO вычисляет предсказания bounding boxes и вероятностей классов.

Пусть $B`_i$ — предсказанные bounding boxes, а $C`_i$ — предсказанные вероятности классов. Для каждого изображения I_i вычисляется функция потерь L_i, которая состоит из двух компонент:

- Localization loss: Lloc (I_i, $B`_i$, B_i).
- Confidence loss: Lconf (I_i, $B`_i$, $C`_i$, B_i, C_i).

Обратное распространение ошибки (Backpropagation) в контексте обучения модели YOLO включает вычисление градиентов функции потерь по параметрам модели и обновление этих параметров с использованием метода градиентного спуска.

Вычисление градиентов функции потерь по параметрам модели:

Пусть θ — параметры модели. Для каждого параметра θ_k модели, где j_k = 1, 2, ..., M (где M — общее количество параметров), градиент функции потерь L по этому параметру вычисляется с использованием правила цепочки, как: $\frac{\partial L}{\partial \theta k} = \sum_{i=1}^{N} \frac{\partial Li}{\partial \theta k}$, где N — количество обучающих изображений, а L_i — функция потерь для *i*-го изображения.

Обновление параметров модели. Параметры модели обновляются в направлении, противоположном градиенту функции потерь, с использованием выбранного метода оптимизации. Обновление параметра θj происходит следующим образом: $\theta k = \theta k - \eta \frac{\partial L}{\partial \theta k}$, где η — скорость обучения, определяющая величину шага обновления параметров.

Постановка задачи обнаружения подозрительной активности:

Имеется множество Imgs = {img_1,img_2, ..., img_j}, где img_j — входное изображение размером W×H, где W и H — ширина и высота, состоящее из j объектов, и множество классов SA = {SA_1, SA_2, ..., SA_k}, где SA_k — k-й класс объектов. Задача обнаружения подозрительной активности на изображении (классификации объектов на изображении) состоит в том, чтобы для каждого объекта на изображении определить его принадлежность к определенному классу из множества SA.

Математическое описание используемой модели (Resnet50):

Пусть x — входной сигнал, а F(x) — преобразование, которое должна научиться сеть. Вместо того чтобы напрямую предсказывать преобразование F(x), ResNet50 предсказывает разность между F(x) и входом x, т. е. F(x) — x. Это позволяет обучать сеть на остаточных, или residual, функциях: F(x) = H(x) + x, где H(x) = F(x) — x. Используя эту концепцию, сеть проще изучает остатки изменений вместо попыток напрямую предсказывать оптимальное преобразование.

Компоненты Renset:

- Основная архитектура: ResNet50 состоит из нескольких блоков, каждый из которых содержит несколько слоев связей. Внутри каждого блока есть два типа связей: прямые (direct connections) и остаточные связи (residual connections).
- Слои: каждый слой в ResNet50 — это полносвязный слой (fully connected layer), который принимает на вход вектор и выдает вектор. В каждом слое используется нелинейный активационный функционал, такой как ReLU (Rectified Linear Unit).
- Остаточные связи: в каждом блоке есть остаточные связи, которые позволяют сети «пропустить» информацию через слои. Остаточные связи добавляются к результату прямой связи, чтобы уменьшить сложность обучения и улучшить производительность сети.
- Уплотнительные слои: в каждом блоке есть уплотнительные слои, которые уменьшают размерность вектора на входе. Это делается путем применения оператора свертки (convolution) с определенным количеством фильтров и размером ядра.
- Слой классификации: в конце сети есть слой классификации, который принимает на вход вектор и выдает вероятности принадлежности к каждому классу.

Особенности ResNet:

Соединения быстрого доступа (shortcut connections) пропускают один или несколько слоев и выполняют сопоставление. Их выходы добавляются к выходам (stacked layers). Используя ResNet, можно решить множество проблем, таких как:

– ResNet-ы относительно легко оптимизировать: «простые» сети (которые просто складывают слои) показывают большую ошибку обучения, когда глубина увеличивается.

– ResNet позволяет относительно легко увеличить точность благодаря увеличению глубины, чего с другими сетями добиться сложнее.

– ResNet имеют меньше фильтров и меньшую сложность (по сравнению с другими сетями, например — VGG). ResNet обеспечивает производительность примерно в 3,6 млрд операций, что по сравнению VGG быстрее в 1,7 раз (примерно 2,1 млрд операций). Слои имеют одинаковое количество фильтров для одного и того же размера (map size). Количество фильтров удваивалось, если map size уменьшался в двое. Это нужно, чтобы сохранить временную сложность для каждого слоя.

Результаты. В результате были изучены подходы к решению проблемы обнаружения нарушений во время проведения единого государственного экзамена. Учтены минусы альтернативных подходов и на их основе были выбраны оптимальные методы. Реализованы YOLO и ResNet архитектуры искусственного интеллекта для выделения областей с учениками и определения подозрительной активности.

Благодаря такому выбору архитектур моделей удалось обеспечить быстродействие работы программного обеспечения и добиться того, что ученики определяются в почти любом положении и с любого угла.

Проведено тестирование работы моделей на различных входных данных в ручном и автоматизированном режиме.

Для модели, предназначенной для обнаружения участников на пунктах проведения протестированы следующие варианты входных данных.

– Изображение, поступившее в модель, не содержит людей.

– Изображение, поступившее в модели, содержит переполненный класс детей (рис. 1).

– Изображение, поступившее в модель содержит количество и расстановку участников единого государственного экзамена согласно правилам пунктов проведения (см. рис. 2).

Рис. 1. Класс с большим количеством людей

Рис. 2. Класс с нормальным количеством людей

По итогам тестирования были получены результаты, представленные в табл. 1.

Таблица 1

	Количество изображений для тестирования	*Количество изображений, на которых были верно определены все люди*
Пустой класс	50	50
Переполненный класс	50	12
Нормально заполненный класс	1000	873

Для модели определения человека была получена функция потерь, которая составила 29,3%. Аналогичный подход использовался для тестирования работы второй модели (табл. 2). Входными данными являлись 3 вида изображений.

– Пустое изображение без человека.
– Изображение с человеком во время нарушения.
– Изображение нескольких людей, которые были обнаружены первой моделью как один человек (см. рис. 3).

Таблица 2

	Количество изображений для тестирования	*Количество изображений, на которых были верно классифицировано поведение*
Пустое изображение	50	50
Изображение с нарушителем	15	11
Изображение с ошибкой определения человека	10	3

Рис. 3. Изображение с несколькими людьми в одной выделенной области

Произведены работы над выбором моделью обнаружения участников на пункте проведения (YOLO). Рассмотрены и протестированы все существующие версии 8 и 9 моделей. В качестве примеров для сравнения приведены результаты работы моделей yolov8m и yolov9e. Сравнивая результаты по метрике confused matrix (рис. 4-5) можно сделать выводы о том, что yolov8m показывает лучшие результаты. Модель yolov9e некорректно определяет наблюдателей, путая их с фоном, а также не определяет учеников.

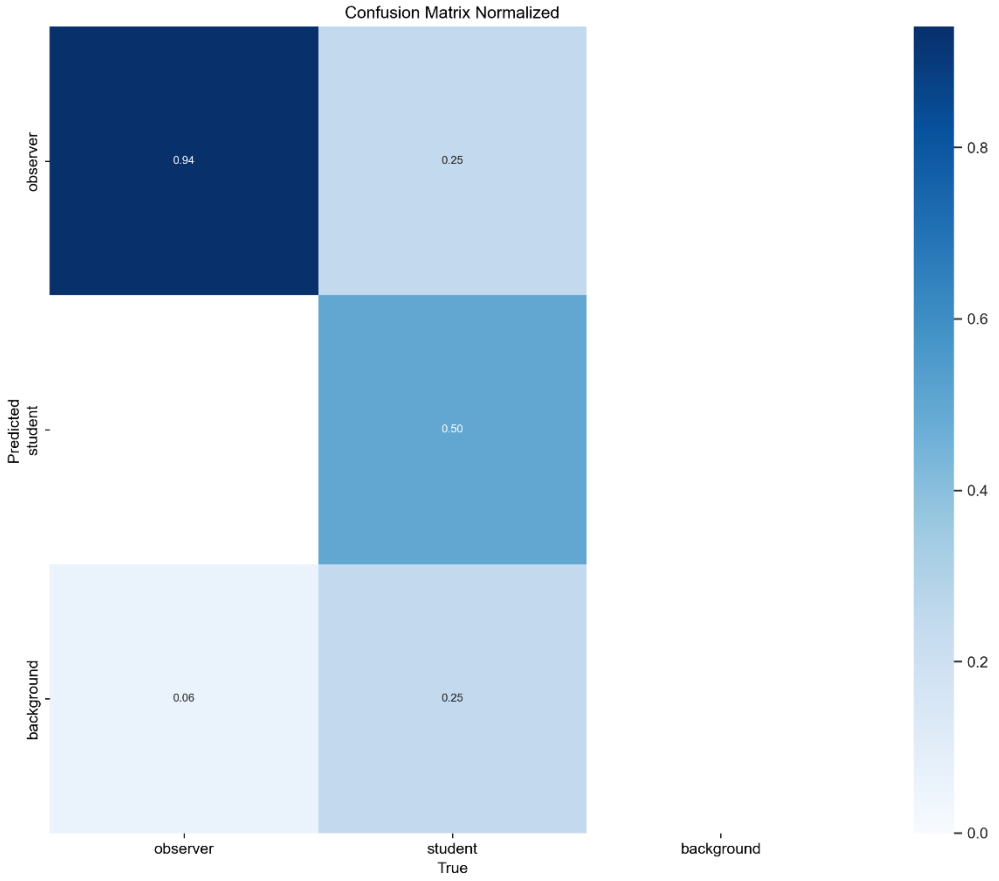

Рис. 4. Yolov8m

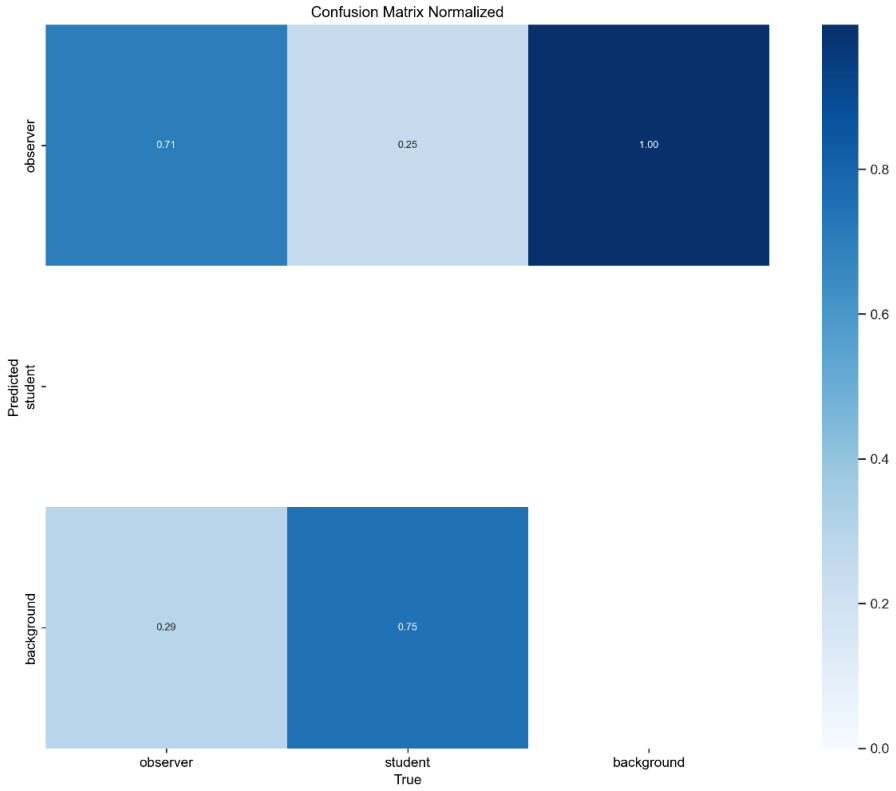

Рис. 5. Yolov9e

Заключение. Выполнена программная реализация комплексной системы обнаружения нарушителей правил проведения единого государственного экзамена для наблюдателей регионального центра обработки информации.

Построены и обучены 2 модели глубоких нейронных сетей:

- модель детекции и выделения участников на пункте проведения ЕГЭ;
- модель обнаружения подозрительной активности на основе ResNet.

Произведены тесты по выявлению ошибок в компонентах системы. Полученные ошибки рассмотрены командой и приняты меры по их исправлению.

СПИСОК ЛИТЕРАТУРЫ

1. Tong Liu. Ai proctoring for offline examinations with 2-longitudinal-stream / high school of shandong province- weifang, china. — URL: https://www.sciencedirect.com/science/article/pii/S2666920X22000704 (дата обращения 20.03.2024). — Текст: электронный.
2. Sushmita Mishra, Roopikha, Roshini, Rithika. AUTOMATIC CHEATING DETECTION IN EXAM HALL. Department of Computer Science & Engineering, Rajalakshmi Engineering College, Chennai, India. — URL: https://www.researchgate.net/publication/375790733 (дата обращения: 20.03.2024). — Текст: электронный.
3. A comparison of QoS parameters of WebRTC videoconference with conference bridge placed in private and public cloud. Авторы Robert R. Chodorek, Grzegorz Rzym, Agnieszka Chodorek, Krzysztof Wajda. Опубликован на IEEE 26th international conference on enabling technologies, Web of Science, 2017, Poznan, Poland. — 91 с.
4. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.