Д. М. Бобырев [1], А. О. Румянцев [1], Д. И. Волобуев [1], Ю.А.Егоров [1,2]

[1] Тюменский государственный университет, г. Тюмень

[2] Научно-технический университет «Сириус», г.Сочи

УДК 532.546.2

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ АЛГОРИТМОВ И МЕТОДОВ РАСПОЗНАВАНИЯ ОБРАЗОВ ДЛЯ ПОДСЧЁТА КОЛИЧЕСТВА ЛЮДЕЙ В АУДИТОРНОМ ПОМЕЩЕНИИ

Аннотация. В статье представлен процесс разработки системы поиска количества людей на видео с помощью нейронных сетей.

Ключевые слова: нейронные сети, обнаружение людей на видео.

Введение

На сегодняшний день одной из основных задач высших учебных заведений является повышение качества учебного процесса. Для этого совершенствуется учебный процесс, внедряются инновационные методы преподавания, корректируются учебные программы в соответствии современным тенденциям, но остается нетронутым немаловажный фактор, как посещаемость занятий студентами. Широкое распространение цифровых технологий заставляет пересматривать традиционные инструменты учета и мониторинга посещаемости студентами аудиторных занятий и вести поиск новых подходов к решению этой проблемы. Имеющиеся инструменты мониторинга, такие как ведения журнала посещаемости лаборантами на этажах, ведение журналов посещаемости старостами группы, СКУД, не обладают достаточной точностью, актуальностью и зависимы от человеческого фактора. Для решения этой проблемы нами было предложено создать систему мониторинга учета

количества людей на аудиторных занятиях с использованием камер, установленных в аудиториях.

Подсчёт людей на видео

Для решения задачи поиска количества людей на видео был предложен алгоритм, изображенный на рисунке 1.

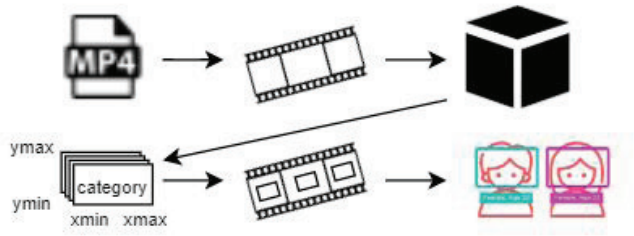

Рис. 1. Визуализация алгоритма

На вход алгоритму подаётся видеофайл, который затем разбивается на кадры и каждый кадр подаётся на вход детектору людей. Далее детектор людей выдаёт для каждого кадра одну или несколько ограничительных рамок, каждая из которых состоит из четырёх координат [xmin, xmax, ymin, ymax] и показывает, что в данной области находится человек. Затем, все полученные рамки накладываются на соответствующие кадры и на выходе получается видеофайл, на котором находятся выделенные люди, а также файл, в котором отражена метаинформация о видео и количество людей на кадрах.

Разбиение видео на кадры

Первым шагом в разработке данного алгоритма является разбиение видеофайла на кадры, а затем последовательная передача каждого кадра на вход детектору объектов.

Для решения задачи разбития видеофайла на кадры была использована библиотека под названием OpenCV - кроссплатформенная библиотека, ориентированная на приложения, связанные с компьютерным зрением. Данная библиотека предоставляет ряд методов, частью которых является эффективная обработка видеопотоков.

Обнаружение людей в кадре

В настоящее время нейронные сети распознают объекты гораздо лучше людей [2] и используются во многих проектах в качестве инструмента для обнаружения различных объектов как в статических изображениях, так и в видео потоках [3, 4, 5], поэтому, в качестве инструмента для создания детектора объектов, было решено использовать нейронные сети.

Создание точных моделей машинного обучения, способных локализовать и идентифицировать несколько объектов на одном изображении является основной проблемой в компьютерном зрении. В качестве основы для детектора объектов было решено использовать Google Object Detection API в основе которого лежит tensorflow - библиотека с открытым исходным кодом, предоставляющая множество инструментов для работы с нейронными сетями.

Tensorflow предлагает для использования множество заранее обученных моделей, подготовленных на наборе данных COCO. Данные модели могут использоваться “из коробки”, а также они могут быть полезны в качестве инициализации для обучения моделей новым классам.

Топология нейросети

В качестве заранее обученной модели использовалась архитектура Faster R-CNN [1], которая поставляется в tensorflow.

Архитектура данной модели состоит из трех частей (рис. 2):

1. Свёрточные слои - в данных слоях обучаются фильтры для извлечения соответствующих характеристик изображения.
2. Region Proposel Network (RPN) - небольшая нейронная сеть, скользящая по последней карте характеристик слоёв свертки и предсказывающая ограничительную рамку и наличие объекта на каждой выделенной характеристике.
3. На данном этапе используется полностью подключенная нейронная сеть, которая, в качестве входа, использует области, предложенные RPN слоем и на выходе, предсказывает класс объекта и ограничивающие блоки.

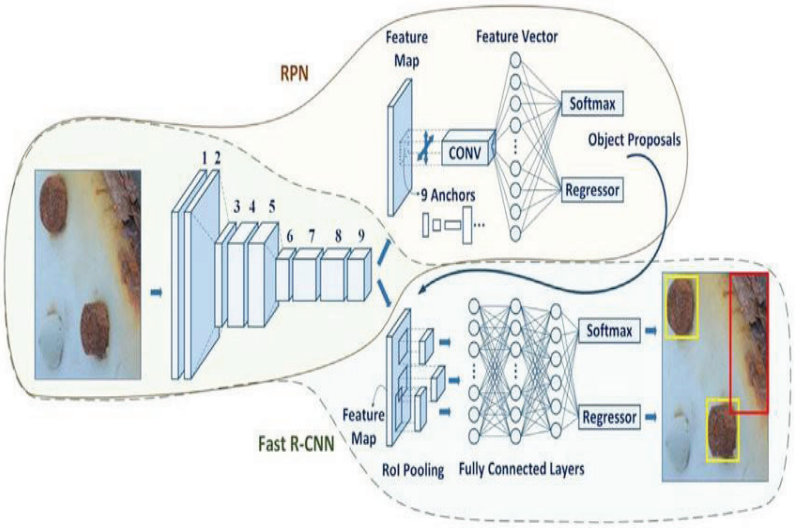

Рис. 2. Архитектура Faster R-CNN

Набор данных

Исходными данными для тестирования предложенного подхода послужили видеозаписи проведения ЕГЭ. Данный набор данных соответствовал следующим характеристикам:

Характеристика	Значение
Расширения файлов	mp4
Количество кадров в секунду	15-20
Длительность одного видео	60 минут
Разрешение	320x240 пикселей
Количество аудиторий	3
Камер в аудитории	2

После изучения данного набора данных стало понятно, что, помимо учеников, на кадре могут присутствовать еще учителя и независимые наблюдатели, сидящие на задних рядах. Так же на различных видео различалось освещение, были блики и люди были плохо различимы даже на глаз.

Рис. 3. Пример исходного набора данных.

Вычислительный эксперимент

Так как заранее обученная модель уже умела распознавать класс “Человек”, было проведено тестирование на предоставленном наборе данных (рис. 4).

Рис. 4. Распознавание без переобучения модели.

Из рисунка 4 видно, что заранее обученная модель не распознает несколько типов людей на видео:

- находящихся за задними партами;
- наблюдатели, которые сидят в конце аудитории;
- наклонившиеся к столу;

Для решения данных проблем было решено дополнительно обучить модель. Были взяты 3 случайных кадра из 30 видео и вручную размечены для проведения дополнительного обучения. Затем было проведено дополнительное обучение модели, результат можно увидеть на рисунке 5.

Рис. 5. Распознавание с дополнительным обучением модели.

Ниже приведена таблица сравнения до и после дополнительного обучения модели в сравнении с визуальной оценкой. В качестве основы для подсчёта количества людей в конкретной секунде бралось среднее по всем кадрам в данной секунде.

Таблица 1.

время на видео	визуальная оценка	до обучения	после обучения
00:00	14	6	13

00:15	14	9	11
00:30	1	0	1
01:00	0	0	0
01:15	14	4	12
01:30	7	4	7
02:00	5	4	6
02:15	10	8	10
02:30	9	8	9
03:00	10	6	10
03:15	3	1	3
03:18	3	2	3

Из таблицы видно, что процент совпадения с визуальной оценкой у модели до дополнительного обучения гораздо ниже, чем у модели после обучения. Это связано с тем, что набор данных, на котором заранее обучалась модель значительно отличался от данного нам, в нём отсутствовали картинки с таким малым разрешением, а также он не был нацелен на задачу распознавания людей в аудитории. После дополнительного обучения процент совпадения с визуальной оценкой значительно вырос.

Заключение

В рамках данной работы был разработан прототип системы распознавания людей в аудиторных помещениях, посредством анализа видеозаписей.

В начале была протестирована заранее обученная, на наборе данных COCO, модель и выяснилось, что существует необходимость в дополнительном обучении данной модели из-за низкого качества распознавания в аудиторном помещении при малом разрешении камеры.

Затем была сформирована репрезентативная выборка и модель была дополнительно обучена и протестирована. Результатом дополнительного обучения модели стало более точное распознавание людей в аудитории.

Дальнейшее исследование предполагает развитие системы с помощью интеграции с расписанием, а также отслеживание концентрации внимания студентами и построение отчетов по полученным данным, которые можно использовать для повышения качества образования.

Благодарности

Статья подготовлена в рамках разработки образовательного кейса для НТУ Сириус при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028.

СПИСОК ЛИТЕРАТУРЫ

1. Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – С. 91-99.
2. Nguyen A., Yosinski J., Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 427-436.
3. True N. Vacant parking space detection in static images //University of California, San Diego. – 2007. – Т. 17. – С. 659-662.

4. Tschentscher M., Neuhausen M. Video-based parking space detection //Proceedings of the Forum Bauinformatik. – 2012. – C. 159-166.
5. Amato G. et al. Deep learning for decentralized parking lot occupancy detection //Expert Systems with Applications. – 2017. – T. 72. – C. 327-334.