

УДК 004.852, 004.855.5

РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБЪЕКТОВ НА ВИДЕОИЗОБРАЖЕНИИ

© Додонов А.А., Солдатова О.П.

*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

e-mail: saha.dodonov@gmail.com

Целью разработки интеллектуальной системы является распознавание лиц людей в маске и без нее, изображение которых передается через видеокамеру. Распознавание реализуется с помощью моделей сверточных нейронных сетей. Так как изображение некоторого объекта зависит от множества параметров, определяющих его местоположение по отношению к видеокамере, характеристик камеры, подсветки, статических и динамических характеристик предмета, то снимок или кадр довольно непросто оформить и представить в форме конкретной математической модели. Из этого следует, что способы осуществления компьютерного представления изображения напрямую зависят от поставленных целей работы.

В настоящее время для распознавания видеоизображений и классификации объектов на них используются нейронные сети. Обученные на определенном наборе данных, они представляют собой идеальное решение большого спектра задач. Основной целью данной работы является разработка системы распознавания объектов на видеоизображении при помощи моделей сверточных нейронных сетей.

Для решения поставленной задачи были детально изучены различные архитектуры сверточных нейронных сетей, процесс их обучения методом обратного распространения ошибки с использованием алгоритмов оптимизации, идеи трансферного обучения, методы аугментации данных.

По результатам проведенного анализа для решения данной задачи были выбраны две модели нейронных сетей: ResNet50, представляющая собой сеть, основанную на принципах остаточного обучения, и MobileNetV2 – сеть остаточного обучения для мобильной платформы.

Adam – один из самых эффективных алгоритмов обучения глубоких нейронных сетей, являющийся оптимизатором известного алгоритма стохастического градиентного спуска. Он сочетает в себе идеи алгоритма RMSProp и алгоритма с импульсом или моментом [1]. Вместо того чтобы адаптировать скорость обучения параметров на основе среднего первого момента, как в RMSProp, Adam также использует среднее значение вторых моментов градиентов. В частности, алгоритм вычисляет экспоненциальное скользящее среднее градиента и квадратичный градиент [2].

Идея трансферного обучения строится на том, что знания, накопленные в модели, подготовленной для выполнения одной задачи – например, распознавания цветов на фотографии, – могут быть перенесены на другую модель, чтобы помочь решению другой, родственной задачи.

Существуют различные подходы к трансферному обучению, но один из них – тонкая настройка (finetuning) – находит особенно широкое применение. При таком подходе используется предварительно обученная модель, в которой удаляются или переучиваются последние слои этой модели, чтобы сфокусироваться на новой, похожей задаче [3].

Для того чтобы увеличить размер датасета и тем самым позволить сети обучиться на большем объеме информации, используется техника под названием аугментация данных [4]. Суть сводится к искусственному применению всевозможных преобразований к элементам датасета, например, таких как поворот изображения. Это существенно не меняет внутреннюю структуру данных, но немного изменяет их представление.

Для обучения нейронной сети поставленной задаче, а именно классификации на видеоизображении двух классов объектов, было создано два набора данных. Первый набор содержит изображения лиц людей с отсутствующей маской. Второй набор данных – лица людей, у которых маска на лице присутствует. Оба набора включали по 1500 изображений каждый.

Для оценки качества обученной нейронной была использована метрика качества аккуратность (ассигасу). Она показывает количество правильно определенных данных от общего количества данных.

В результате выполнения поставленной задачи было произведено множество экспериментов, направленных на оценку эффективности использования моделей ResNet50 и MobileNetV2, обученных с нуля и обученных при помощи методов трансферного обучения.

В проведенных экспериментах для оценки качества обучения кроме метрики аккуратности для оценки эффективности исследовались и другие параметры: количество обучаемых параметров, время обучения, потребление памяти для анализа одного кадра, время, затрачиваемое на анализ одного кадра.

Результаты исследований моделей нейронных сетей представлены в следующей таблице.

Таблица – Результаты исследования нейронных сетей

Нейросеть	Параметры	Обучаемые параметры	Размер на диске	Аккуратность	Время обучения
ResNet50	23,850,242	262,530	40,3 МВ	0.9010	0,13 с
MobileNetV2	2,422,210	161,226	36,0 МВ	0,9878	0,05 с

Лучшие результаты показала система, использующая модель MobileNetV2, которая имеет 161 226 обучаемых параметров. Метрика аккуратности на заданном наборе данных составила более 98 %. Для анализа одного кадра данная система заняла 36 МБ оперативной памяти и 0,05 секунды времени.

Библиографический список

1. Николенко С., Кадулин А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.
2. Хабр. Результируем и сравниваем оптимизаторы моделей в глубоком обучении. URL: <https://habr.com/ru/company/skillfactory/blog/525214> (дата обращения: 18.05.2022).
3. Трансферное обучение. URL: <https://academy.yandex.ru/posts/transferное-obuchenie-rochemu-deep-learning-stal-dostupnee> (дата обращения: 18.05.2022).
4. Википедия. Data augmentation URL: https://en.wikipedia.org/wiki/Data_augmentation (дата обращения: 18.05.2022).