

ИССЛЕДОВАНИЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ YOLOV8 И F-RCNN ДЛЯ ЗАДАЧИ РАСПОЗНОВАНИЕ ОБРАЗОВ (СПЕЦ ОДЕЖДЫ НА ЛЮДЯХ)

Трудовая безопасность на производстве является ключевым аспектом. Это двухсторонние отношения, как производства, так и работника. Предприятие обеспечивает безопасные условия труда, соответствующие нормативным требованиям и стандартам охраны труда, посредством внедрения современных технологий управления рисками, регулярного мониторинга производственной среды и обучения персонала методам безопасного выполнения работ.

Для того что бы отслеживать процесс выполнения правил безопасности, можно использовать специально обученного человека или машину. Человек ограничен в количестве информационных потоков, которые он способен обработать одновременно. Кроме того, на его продуктивность могут повлиять такие человеческие факторы, как усталость, уровень заинтересованности (мотивация) и недостаточная концентрация на выполняемой задаче. В отличии от человека, машина, не имеет данных недостатков. Она может автоматизированно оповещать об нарушении, вносить данные об нарушении в базу данных.

Ядром такой системы будет нейросетевая модель для решения задачи детекции объектов или, что ещё лучше, задачи сегментации экземпляров. Для решения подобных задач используются как двухэтапные подходы Faster RCNN, Mask RCNN [4], так и одноэтапные – YOLO (You Only Look Once), UNet, SSD (Single Shot Multibox Detector) и др.

Для исследования качества работы, производительности и требования к вычислительным ресурсам рассмотрим модели YOLOv8m и Fasten RCNN с измененной классификационной головой с использованием FPN (Feature Pyramid Network).

YOLO состоит из 4 основных частей. Backbone (основная сеть) используется для нахождения признаков из изображения. Neck (шея) используется для агрегации признаков с различных уровней пирамиды. Часто используется FPN (Feature Pyramid Network) или PAN (Path Aggregation Network). Эти слои помогают объединять признаки из глубоких и поверхностных уровней, что улучшает способность сети обнаруживать объекты разных размеров. Head (Голова) необходим для предсказания ограничивающих рамок для объектов, используя якорные

точки (anchor points) и их смещения. Также предсказываются вероятности принадлежности к определенным классам. В качестве входных данных используется изображение или видеопоток, который проходит через сверточные слои, извлекающие иерархические признаки [3].

Задача Fasten R-CNN быстрое детектирование объектов на изображении или видео. На рис. 1 изображена топология предлагаемой сети.

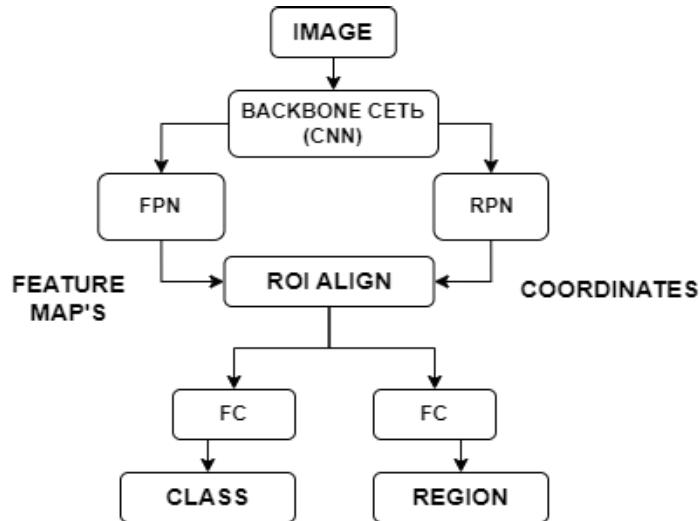


Рисунок 1 – Архитектура FRCNN

Модель решает проблему производительности RCNN, где отдельные региональные предложения классифицируются напрямую на уровне изображения, вместо раздельной обработки [2]. Структура модели состоит из модулей FPN, RPN, ROI Align и классификационной «головы», которая в свою очередь выступает для создания карт признаков в Fasten RCNN [1]. Inception V3 был модифицирован для работы с Fasten RCNN + FPN.

Из классификатора был удален выходной слой, данный классификатор был интегрирован в классификационную «голову». Так же что бы сформировать пирамиду признаков используется FPN сеть. Для данной сети были выбраны 5 сверточных слоев из последнего блока в модели Inception V3.

Для обучения нейросети использовалось следующее оборудование: AMD Ryzen 5 4500 (4.4 GHz), RTX 3060 12 GB, 16 GB ОЗУ. Использовались Python 3.10.6, CUDA 12.1, Torch 2.1.0. Набор данных включал 500 изображений, размер батча – 5, learning rate – 0.01.

Каждая модель обучалась 150 эпох, классифицировались 5 классов: каска, спецовка, фиксирующая лямка, положение человека (лицом или спиной) [5].

Таблица 1 – метрики объектов для FRCNN

Метрика	Общее	Малые объекты	Средние объекты	Большие объекты
mAP	0.9384	0.7174	0.9860	0.9807
mAR(top 1)	0.6559	0.8351	0.9995	0.9978
mAR (top 10)	0.9719	-	-	-
mAR (top 100)	0.9725	-	-	-

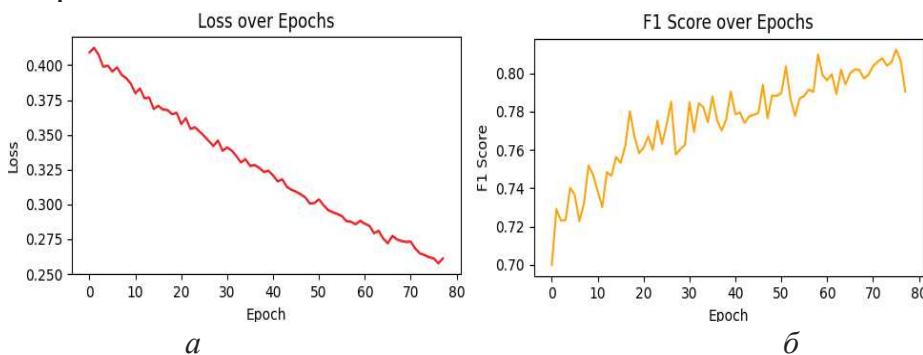
После обучения YOLO8 и FRCNN проявляются следующие результаты. Обе модели отлично определяют ограничивающие рамки для больших и средних объектов, но FRCNN имеет сложность в обнаружении мелких деталей таких как лямки касок и небольшие кусочки касок. В таблицах 1 и 2 изображена метрика mAP и mAR для FRCNN, в таблице 2 изображены метрики для YOLO8 соответственно.

Согласно таблице 2, модели наименее точно распознают каски и их лямки. Причина этого кроется в особенностях исходного набора данных: на изображениях присутствуют каски и их лямки размером от 5% до 30% от стандартного размера каски и лямки. Более того, в общем объеме классифицируемых объектов лямки касок представлены наименьшим количеством примеров.

Таблица 2 – метрики классов для модели FRCNN и YOLO8

Класс	FRCNN mAP50	FRCNN mAR (top 100)	YOLO8 mAP50	YOLO8 mAR (top 100)
Каска	0.9881	0.9964	0.995	0.998
Ремешок каски	0.7774	0.8754	0.956	0.914
Спецодежда	0.9855	0.9968	0.995	0.996
Люди спиной	0.9767	0.9959	0.995	0.998
Люди лицом	0.9644	0.9979	0.995	0.994

Как можно увидеть из таблицы 2, что YOLO8 лучше распознает как большие объекты, так и малые, как например лямка каски или часть каски из-под капюшона. Большие и средние объекты распознаются на должном, высоком уровне, для двух моделей. На рис. 2 изображен график с метрикой F1 score и loss для FRCNN.



a – график функции потерь, *б* – график метрики F1 score
Рисунок 2 – Loss и F1 score для модели FRCNN

Характерной особенностью является дублирование уже помеченных областей для FRCNN, что подтверждается низким показателем mAR. Высокий mAP указывает на корректное определение и классификацию ограничивающих рамок [2]. F1-score модели остается на приемлемом уровне (0.8), но меняется крайне медленно. Для сравнения, модель YOLO8 показывает существенно более высокий F1-score – 0.95.

На основании представленных данных можно сделать следующие выводы. Faster R-CNN в текущей конфигурации демонстрирует не-плохие результаты, однако имеет существенные ограничения при обнаружении мелких деталей и страдает от периодических ложных срабатываний, что делает её неподходящей для задачи детектирования спецодежды рабочего персонала. Рекомендации по оптимизации:

- пересмотр гиперпараметров модели;
- дополнение FPN низкоуровневыми слоями для более эффективного извлечения признаков.

По предварительным тестам, более предпочтительной является модель YOLOv8m. Она не имеет вышеперечисленных недостатков в отличии от FRCNN.

ЛИТЕРАТУРА

1. Жерон О. Прикладное машинное обучение с помощью ScikitLearn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем / Жерон О. – СПб.: Диалектика, 2018. – 688 с.
2. Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение / Гудфеллоу Я., Бенджио И., Курвиль А. – М.: ДМК Пресс, 2018. – 652 с.
3. Kurachka K.S., Luchshava T.V., Panarin K.A. Localization of human percentages on X-ray images with use of Darknet YOLO / Kurachka K.S., Luchshava T.V., Panarin K.A.// Doklady BGUIR. – 2018. – Vol. 113, No. 3. – P. 32-38.
4. Kurochka K. S., Panarin K. A. An algorithm of segmentation of a human spine X-ray image with the help of Mask R-CNN neural network for the purpose of vertebrae localization // 2021 56th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST). – IEEE, 2021. – C. 55-58.
5. Компьютерное зрение: распознавание одежды на фотографии с помощью мобильного приложения [Электронный ресурс]. Режим доступа: <https://habr.com/ru/companies/anetika/articles/241343/> (дата обращения: 12.01.2025).