ПРЕДОБРАБОТКА ДАННЫХ ДЛЯ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ YOLO В ШАХТАХ

Петровский В.В.
Томский политехнический университет, ИШИТР, аспирант гр. А3-39,
e-mail: vvp32@tpu.ru

Аннотация

В статье рассматриваются методы предобработки изображений и видео, предназначенных для обучения нейронных сетей, применяемых для распознавания и оценки позы человека в условиях подземной шахты. Проведён анализ влияния параметров аугментации на точность моделей YOLOv8n-pose и YOLO11n-pose. Исследование ориентировано на улучшение устойчивости моделей к условиям плохого освещения, запыленности и ограниченности данных.

Ключевые слова: YOLO, шахта, нейросеть, распознавание позы, аугментация, предобработка данных.

Введение

Системы автоматического распознавания объектов и оценки позы человека находят всё более широкое применение в промышленной сфере, в том числе в условиях подземных шахт. Мониторинг состояния и поведения шахтёров с помощью компьютерного зрения позволяет повысить уровень промышленной безопасности.

Работа нейронных сетей в шахтных условиях осложняется особенностями среды: плохим и неравномерным освещением, запыленностью, высоким уровнем шума на изображениях и ограниченным количеством размеченных данных. В таких условиях особенно важно качество и объём тренировочного датасета, а также методы его предобработки.

Цель исследования – проанализировать влияние параметров аугментации изображений (цветовой тон, насыщенность, яркость, угол поворота изображения и т. д.) на точность нейросетевых моделей YOLOv8n-pose и YOLO11n-pose при решении задачи детекции и оценки позы человека в подземной шахте. Работа также направлена на выбор наилучших параметров предобработки данных с учетом ограниченных вычислительных ресурсов, характерных для портативных устройств, используемых в горнодобывающей промышленности.

Методы предобработки и аугментации данных

Для распознавания и оценки позы человека в условиях подземной шахты рассматривались следующие методы:

- Традиционные методы аугментации. Они включают геометрические преобразования (повороты, масштабирование, сдвиги, отражения), изменения цветовой гаммы (яркость, контрастность, насыщенность), а также добавление шума [1]. Например, увеличение яркости может помочь имитировать условия более светлой шахты, а добавление шума может сделать модель более устойчивой к зашумленным изображениям (пылевая завеса).
- Аугментации, основанные на GAN (Generative Adversarial Networks). GAN могут быть использованы для генерации синтетических изображений шахтных сцен с людьми [2] а также для генеративно-обученных моделей на больших датасетах (например, ImageNet, COCO) и их дальнейшая дообучка на шахтных данных [3].
- Методы борьбы с запыленностью и плохим освещением. Существуют специализированные методы улучшения качества изображений, направленные на устранение запыленности и коррекцию освещения. Например, использование алгоритмов гистограммного выравнивания или фильтров медианной фильтрации [4].

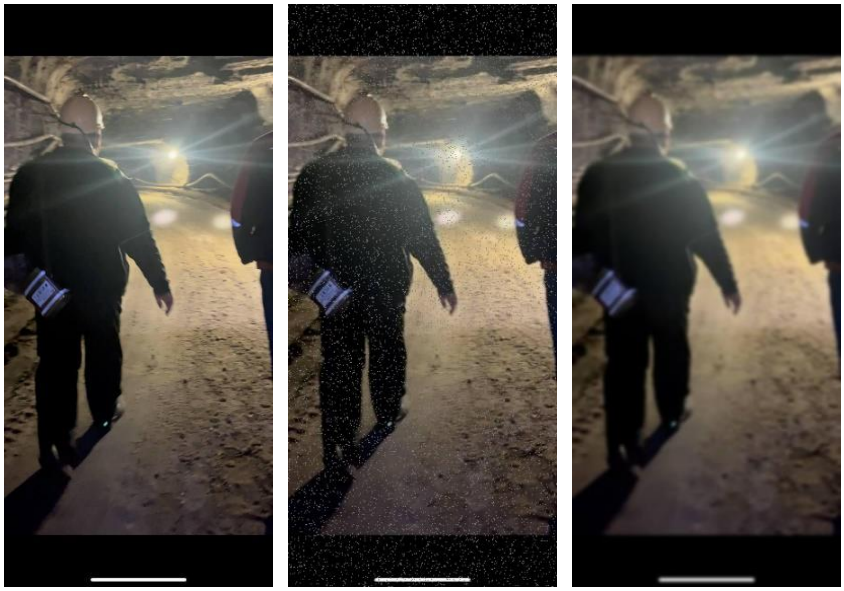
• Mixup и CutMix. Эти методы аугментации основаны на смешивании изображений и меток из различных классов [5, 6]. Это позволяет улучшить обобщающую способность модели и сделать ее более устойчивой к зашумленным данным.

• Алгоритмы выравнивания гистограммы. Эти алгоритмы улучшают контрастность изображения путем перераспределения значений пикселей [7]. Это особенно полезно в условиях низкой освещенности, характерных для шахт. Адаптивное выравнивание гистограммы (CLAHE) может быть особенно эффективным для локального улучшения контрастности.

Экспериментальная часть

Датасет был собран автором на основании видео и снимков, полученных при производственном мониторинге в шахтах. Было размечено 500 снимков в разрешении 298×640 пикселей с помощью платформы Label Studio [10]. Разметка включала прямоугольные рамки, охватывающие фигуру человека, и его позу в кадре. Все программные операции проводятся на персональном компьютере с CPU Intel Core i5-12400F OEM, GPU NVIDIA GeForce RTX 4060, 16 ГБ ОЗУ, 1 ТБ ПЗУ.

Для увеличения размеченного датасета создаются новые снимки на основе исходных, к которым применяются серый шум (пылевая завеса при ярком освещении) и гауссовское размытие (загрязнение камер), что представлено на рис. 1.



а б в

Рис. 1. Предобработка снимка из датасета:
а – исходный снимок, б – снимок с серым шумом, в – снимок с гауссовским размытием

После увеличения датасета до 1500 снимков, случайным образом разделяем снимки: 1200 на обучающую и 300 на валидационную выборки. Применяем аугментацию для обучения моделей YOLOv8-pose nano и YOLOv11-pose nano. Параметры, применяемые в эксперименте:

- Цветовой тон.
- Насыщенность.
- Яркость.
- Угол поворота изображения.
- Перемещение по горизонтали и вертикали на долю размера изображения.
- Коэффициент масштабируемости изображения.
- Сдвиг изображения на заданный коэффициент.
- Поворот изображения по центральной вертикальной оси.

Оптимизатор и функция потерь применяются как в официальной обученной модели YOLO. Обучение проводиться на 200 эпохах. Далее проводится обучение моделей. В таблице 1 находятся параметры аугментации, применяемые к каждому из наборов.

Таблица 1. Настраиваемые параметры аугментации

Параметры	цветовой тон	насыщенность	яркость	угол поворота	перемещение	масштабируемость	сдвиг	поворот по оси
Параметры, установленные в обученных официальных моделях YOLO	0,015	0,700	0,400	0,000	0,100	0,500	0,000	0,500
Настраиваемые параметры №1 (без аугментации)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Настраиваемые параметры №2	0,015	0,150	0,150	15,00	0,150	0,150	15,00	0,150
Настраиваемые параметры №3	0,030	0,300	0,300	30,00	0,300	0,300	30,00	0,300
Настраиваемые параметры №4	0,045	0,450	0,450	45,00	0,450	0,450	45,00	0,450
Настраиваемые параметры №5	0,060	0,600	0,600	60,00	0,600	0,600	60,00	0,600

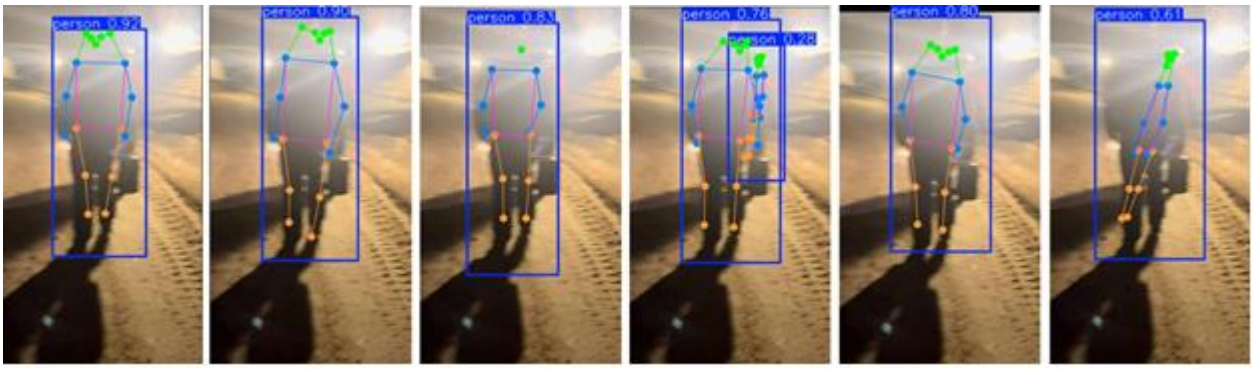
Результаты валидации обученных моделей представлены в таблице 2. Таблица содержит метрики предсказаний: Recall, mAP50, mAP50-95 по двум задачам – детекция рамки и оценка позы.

Таблица 2. Валидационные данные обученных моделей YOLOv8n-pose и YOLO11n-pose

Параметры	Модель	Box				Pose			
		Predict	Recall	mAP50	mAP50-95	Predict	Recall	mAP50	mAP50-95
Параметры, установленные в обученных официальных моделях YOLO	YOLOv8n-pose	0,958	0,882	0,950	0,626	0,848	0,692	0,723	0,300
	YOLO11n-pose	0,952	0,885	0,943	0,605	0,944	0,731	0,760	0,289
Настраиваемые параметры №1 (без аугментации)	YOLOv8n-pose	0,987	0,883	0,904	0,584	0,807	0,654	0,674	0,253
	YOLO11n-pose	0,958	0,872	0,903	0,590	0,743	0,577	0,557	0,243
Настраиваемые параметры №2	YOLOv8n-pose	0,975	0,880	0,962	0,545	0,808	0,654	0,608	0,180
	YOLO11n-pose	0,889	0,921	0,938	0,584	0,794	0,615	0,560	0,197
Настраиваемые параметры №3	YOLOv8n-pose	0,967	0,769	0,891	0,48	0,775	0,531	0,518	0,165
	YOLO11n-pose	0,957	0,885	0,925	0,541	0,722	0,538	0,569	0,176
Настраиваемые параметры №4	YOLOv8n-pose	0,952	0,846	0,910	0,467	0,459	0,385	0,299	0,114
	YOLO11n-pose	0,989	0,808	0,904	0,447	0,518	0,423	0,327	0,102
Настраиваемые параметры №5	YOLOv8n-pose	0,921	0,731	0,857	0,357	0,444	0,231	0,241	0,056
	YOLO11n-pose	0,954	0,769	0,888	0,350	0,346	0,269	0,154	0,005

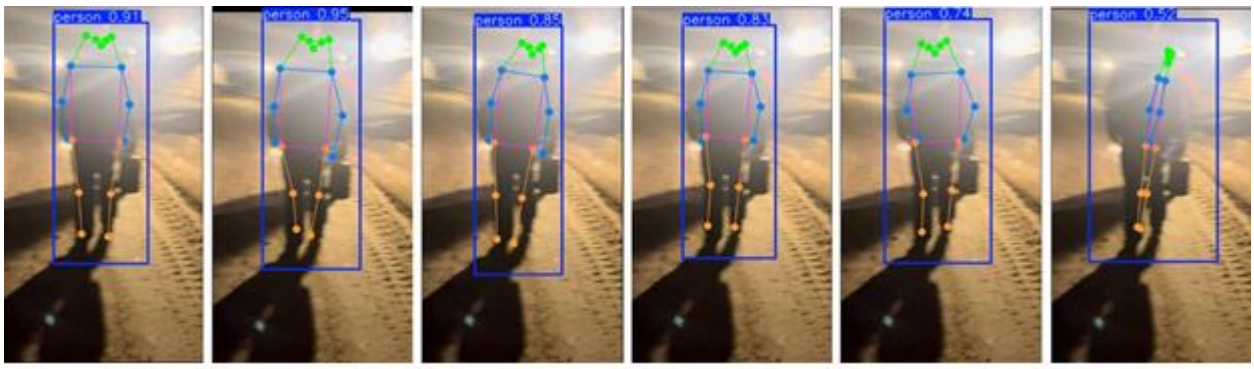
По результатам валидации видно, что после настраиваемых параметров №3 метрики точностией mAP50 и mAP50-95 рамки выделения человека (прямоугольник, предсказанный моделью вокруг фигуры шахтёра, box) и его позы (pose) сильно упали.

Проводится тест обученных моделей YOLOv8n-pose и YOLO11n-pose на видеозаписях из шахты. Рассматривались кадры, в которых результат предсказания значительно различается между конфигурациями моделей. На рисунках 2 и 3 изображены точности предсказания моделей от примененных параметров аугментации.



а б в г д е

Рис 2. Тест обученной модели YOLOv8n-pose при разных аугментационных параметрах:
а – параметры, установленные в обученных официальных моделях YOLO,
б – настраиваемые параметры № 1 (без аугментации),
в – настраиваемые параметры № 2,
г – настраиваемые параметры № 3,
д – настраиваемые параметры № 4,
е – настраиваемые параметры № 5



а б в г д е

Рис 3. Тест обученной модели YOLO11n-pose при разных аугментационных параметрах:
а – параметры, установленные в обученных официальных моделях YOLO,
б – настраиваемые параметры № 1 (без аугментации),
в – настраиваемые параметры № 2,
г – настраиваемые параметры № 3,
д – настраиваемые параметры № 4,
е – настраиваемые параметры № 5

По результатам тестирования видно, что новая версия YOLO11n-pose работает точнее, чем YOLOv8n-pose. Применение аугментационных параметров до 30 % изменения входных данных («настраиваемые параметры №3») приводит к гибкой работе моделей, при этом точность предсказания уменьшается не существенно. После 30 % происходит сильный спад

точности предсказания, модели работают хуже, и выдают ложные срабатывания. Результаты обучения с аугментационными параметрами как в официальных моделей, и обучения без аугментационных параметров, имеют самые высокие показатели как по определению рамки, так и позы человека.

Учитывая сложные условия эксплуатации в шахтах, модели обучались в конфигурации nano – с упором на минимальное потребление ресурсов. Такие модели могут быть реализованы на одноплатных устройствах (например, Jetson Nano, Raspberry Pi с TPU, Odyssey и т. д.) с поддержкой запуска в условиях ограниченного энергопотребления, охлаждения, запыленности и вибраций.

Заключение

Проведён анализ влияния параметров аугментации изображений на точность моделей YOLO в шахтных условиях. Оптимальными оказались конфигурации, где изменения входных данных не превышают 30%. Повышение параметров аугментации приводят к потере точности. Для повышения точности распознавания и гибкости моделей параметры аугментации следуют выставлять опытным путем. Также имеет смысл завершать обучение при неизменяемости показателей точности, а не по заданному количеству эпох.

Данное исследование может помочь в дальнейшем изучении по улучшению качества обучающих моделей. Дальнейшие работы направлены на расширение датасета и внедрение моделей в реальное производство.

Список использованных источников

1. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning //Journal of big data. – 2019. – Т. 6. – №. 1. – С. 1-48.
2. Goodfellow I.J. et al. Generative adversarial nets //Advances in neural information processing systems. – 2014. – Т. 27.
3. Yosinski J. et al. How transferable are features in deep neural networks? //Advances in neural information processing systems. – 2014. – Т. 27.
4. Espinosa A.R., McIntosh D., Albu A.B. An efficient approach for underwater image improvement: Deblurring, dehazing, and color correction //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. – 2023. – С. 206-215.
5. Zhang H. et al. mixup: Beyond empirical risk minimization // arXiv preprint arXiv:1710.09412. – 2017.
6. Yun S. et al. Cutmix: Regularization strategy to train strong classifiers with localizable features // Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – С. 6023-6032.
7. Yeuseyenka I. Detection and selection of moving objects in video images based on impulse and recurrent neural networks // Journal of Data Analysis and Information Processing. – 2022. – Т. 10. – №. 2. – С. 127-141.
8. Explore Ultralytics YOLOv8. [Электронный ресурс]. – URL: docs.ultralytics.com/models/yolov8/ (дата обращения: 28.03.2025).
9. Ultralytics YOLO11. [Электронный ресурс]. – URL: docs.ultralytics.com/ru/models/yolo11/ (дата обращения: 28.03.2025).
10. Open-Source Data Labeling Platform. Label Studio: сайт. [Электронный ресурс]. – URL: labelstud.io/ (дата обращения: 28.03.2025).