

Искусственный интеллект в промышленности



**Метод обогащения 2D-представлений  
данными из 3D-облаков точек для улучшения  
понимания сцены**

Студент:  
Сергеева Валентина Юрьевна, гр. J4150

Научный руководитель:  
Лаушкина Анастасия Александровна, доцент

Научный консультант:  
Волоха Валерия Дмитриевна, ассистент

# Постановка проблемы

## Проблема:

Современные подходы к анализу инженерных сцен можно разделить на два направления:

- 2D-анализ документов – детекция символов, анализ разметки и структуры на изображения;
- 3D-анализ CAD-моделей – распознавание производственных признаков на основе В-дер, графов или вокселей.

Ни одно из этих направлений не обеспечивает совместного понимания 2D-представлений и 3D-геометрии, что ограничивает применимость методов в реальных промышленных системах восприятия

## Актуальность:

- Рост потребности в автономных CAD-ориентированных системах в промышленности;
- Отсутствие методов, сочетающих результат 2D-детекторов с геометрическими данными 3D-облаков точек.

# Цель и задачи

**Цель:** Разработка и анализ метода обогащения 2D-представлений данными из 3D-облаков точек для повышения качества геометрического понимания промышленных сцен.

**Задачи:**

1. Провести аналитический обзор существующих подходов к 2D- и 3D-анализу инженерных данных;
2. Подготовить синхронизированный датасет на основе ABC-dataset с аннотациями 2D-видов, 3D-облаков и ориентации;
3. Разработать архитектуру YOLOv8\_3DFusion, объединяющую семантическую информацию из 2D-изображения с геометрической информацией из 3D-облака точек;
4. Реализовать и обучить модель для совместной задачи детекции и оценки ориентации;
5. Провести оценку по метрикам mAP@0.5 и средней угловой ошибки.

# Обзор аналогичных исследований ИТМО

Наименование работы	Основной метод/подход	Недостатки/ограничения
<i>Quan, Y. et al. "Self-supervised Graph Neural Network for Mechanical CAD Retrieval."</i>	Предложена графовая нейросеть, которая извлекает геометрические признаки из ВRep-представлений CAD-моделей и использует контрастное обучение для поиска схожих деталей без разметки	Работает исключительно с 3D-данными. Подходит только для поиска похожих CAD-моделей.
<i>Yeo, C. et al. "Machining feature recognition based on deep neural networks to support tight integration with 3D CAD systems"</i>	Предложен специальный feature descriptor, вычисляемый для каждой грани ВRep-модели, как вход в полносвязную нейросеть для классификации типов механических элементов (отверстия, пазы и т.д.)	Не использует 2D-рендеры или изображения. Применяет только полносвязную сеть на признаках, без слияния с визуальной семантикой.
<i>Zhang et al. "M2Doc: A Multi-Modal Fusion Approach for Document Layout Analysis"</i>	Предложен модуль раннего и позднего слияния текстовых (через BERT) и визуальных (через CNN) признаков для улучшения детекции областей в документах	Мультимодальность ограничена текстом и изображением, без учёта 3D-структуры.

Ни один из существующих методов не решает задачу обогащения 2D-признаков геометрическими дескрипторами из 3D-облачков точек.

# Методы и инструменты исследования ИТМО

## Методы:

- 2D-детекция (YOLOv8) – локализация объектов на изображениях
- 3D-обработка облаков точек (PointNet) – извлечение геометрических признаков
- Мультимодальное слияние – объединение 2D и 3D признаков на уровне нейросети
- Синтетическая генерация данных – автоматическое создание изображений, облаков точек и аннотаций

## Данные:

- ABC dataset.

## Инструменты и программный стек:

- DL-фреймворки: PyTorch 2.0, Ultralytics
- Обработка изображений и геометрии: NumPy, OpenCV, trimesh,
- 3D-рендеринг: pyrender
- Визуализация: Matplotlib, PIL, Open3D

# Научная новизна

ІТМО

## Научная новизна:

- Предложена мультимодальная архитектура YOLOv8\_3DFusion, которая обеспечивает эффективную интеграцию 2D-визуальных признаков и 3D-геометрических дескрипторов. В отличие от существующих подходов, где модальности объединяются на уровне изображения или на выходе детектора, предложенный метод реализует двухуровневое слияние признаков.

# Ход решения

## **1. Анализ предметной области.**

- Изучение задач распознавания CAD-объектов в промышленных системах восприятия.
- Формулировка задачи: разработка механизма эффективного слияния 2D-визуальных и 3D-геометрических признаков для единой архитектуры детекции.

## **2. Подготовка и генерация данных.**

- База – ABC Dataset (масштабные CAD-модели) Анализ ограничений современных 2D-детекторов.
- Генерация синхронизированных пар: 2D-рендеры (с вариацией ракурса и освещения) + 3D-облака точек.

## **3. Предобработка и валидация данных.**

- Нормализация геометрии и масштаба моделей.
- Валидация корректности проекции 3D → 2D.
- Разделение на обучающую, валидационную и тестовую выборки .

# Ход решения

## 4. Построение 2D baseline.

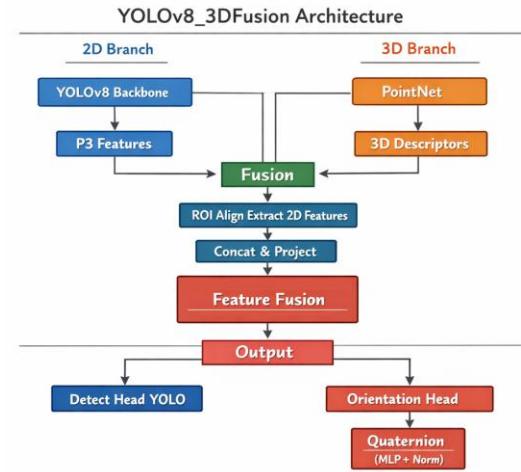
- Обучение YOLOv8 для задачи детекции CAD-компонентов.

## 5. Разработка модуля мультимодального слияния.

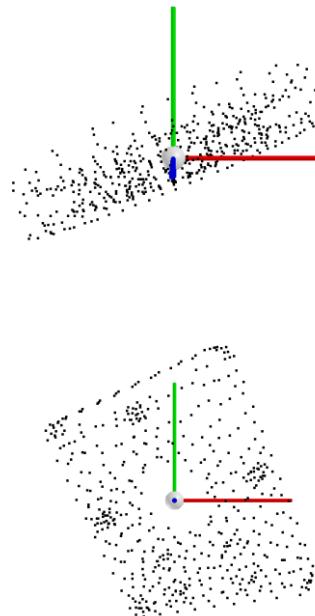
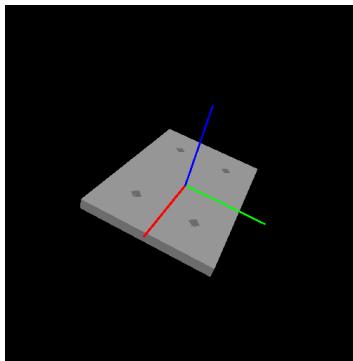
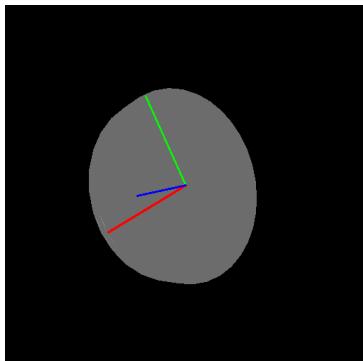
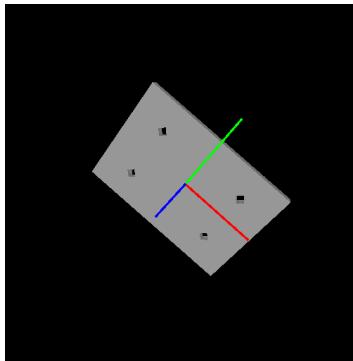
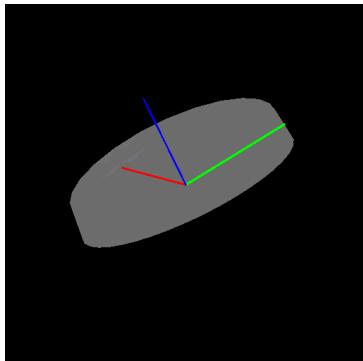
- Извлечение глобальных 3D-признаков с помощью PointNet.
- Выделение 2D-признаков в ROI с помощью ROI Align.
- Объединение 2D и 3D признаков и встраивание в сеть.
- Совместное обучение задачи детекции и оценки 3D-ориентации.

## 6. Оценка и анализ слияния.

- Сравнение методов: 2D-only, 3D-only, 2D+3D.
- Анализ результатов.



# Описание датасета



## Датасет: ABC-dataset

4750 CAD-моделей с детальной геометрией.

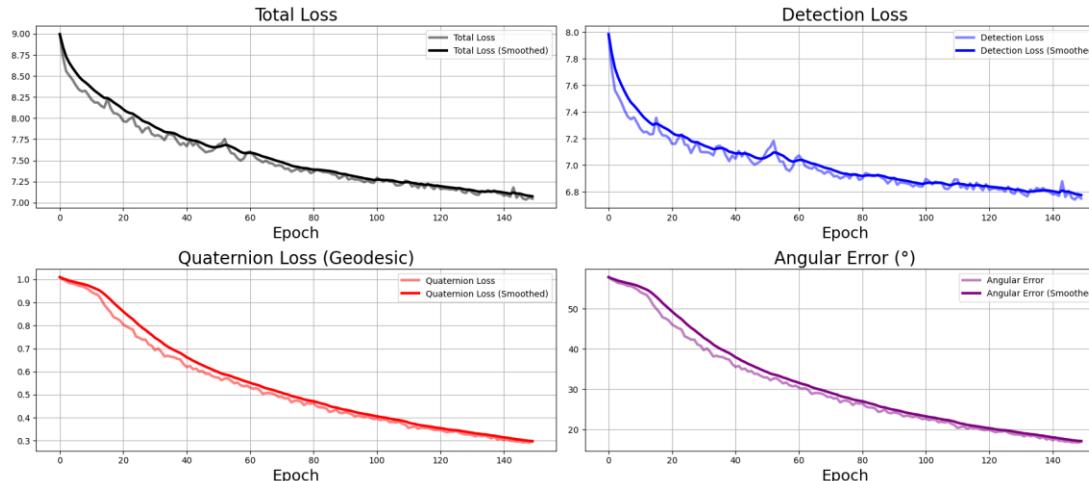
57 000 синхронизированных сцен.

Каждая сцена содержит:

- RGB-изображение,
- 3D-облако точек,
- ориентация в виде кватерниона,
- bounding box объекта.

# Основные результаты

Метод	Тип данных	map0.5	Angle error
YOLOv8 (baseline)	RGB	0.92	~57.9°
3D-only (ICP + GT bbox)	dense point cloud	-	~22–25°
Предложенный метод (YOLOv8_3DFusion)	RGB + sparse point cloud	0.89	17.5°



Ссылка на  
репозиторий



# Основные результаты

## Результаты:

1. Проведён аналитический обзор современных подходов к 2D- и 3D-анализу инженерных данных;
2. Подготовлен синхронизированный датасет на основе ABC-dataset;
3. Разработана архитектура YOLOv8\_3DFusion, объединяющая 2D детекцию и 3D дескрипторы;
4. Реализована и обучена модель на 100 эпохах. Достигнуты следующие результаты: mAP@0.5 = 0.89 (снижение на 3% по сравнению с baseline Yolo). Угловая ошибка = 17.5 ° (снижение с 57.9° ).

## Выводы:

- Модель стабильно сходится. Угловая ошибка снижается с 57.9° до 17.5°;
- Интеграция 3D-геометрии в 2D-детектор (YOLOv8\_3DFusion) улучшает геометрическое понимание сцены. Но требуется расширение датасета реальными данными;
- Оставшаяся ошибка (~ 17.5 °) объяснима симметрией объектов.

# Ожидаемые результаты

ІТМО

- Повышение надёжности классификации типов механических элементов в реальных промышленных сценариях;
- Внедрение методов обогащения с применением графовых нейронных сетей;
- Подготовка и подача публикации.

# Заключение

## В рамках первого семестра:

- Сформулирована цель и задачи;
- Сгенерирован базовый датасет из 3D моделей синхронизированных с облаками точек;
- Обучена базовая модель, позволяющая успешно обогащать 2D признаки признаками, извлеченными из облака точек.

В дальнейшем планируется модификация методов обогащения признаков с применением графовых нейронных сетей для создания GNN-детектора CAD-признаков, который улучшает распознавание признаков в условиях неполных, зашумлённых или частично скрытых CAD-сцен.

**Спасибо  
за внимание!**

**iTMO** *more than a*  
**UNIVERSITY**