Применение машинного обучения в задачах CV

на примере объекта сложной геометрической формы

Руководитель проекта

Титова Александра Михайловна

st076009@student.spbu.ru

Формулировка задачи

Формальная:

Задача определения дефектов сборки промышленного оборудования

В терминах CV:

Комплексная задача детекции объектов на изображении, сегментации изображений, классификации

Задачи промышленной дефектоскопии

Severstal: Steel Defect Detection (2019) [kaggle.com]

PHYGITALISM и IT-центр МАИ [habr.com/ru/company/itmai]

Проблемы

1. Ограниченный объём исходных данных
2. Сложная структура данных, специфика предметной области
3. Уникальность задачи
4. Большое число регулируемых параметров

Команда

Титова Александра [PM]

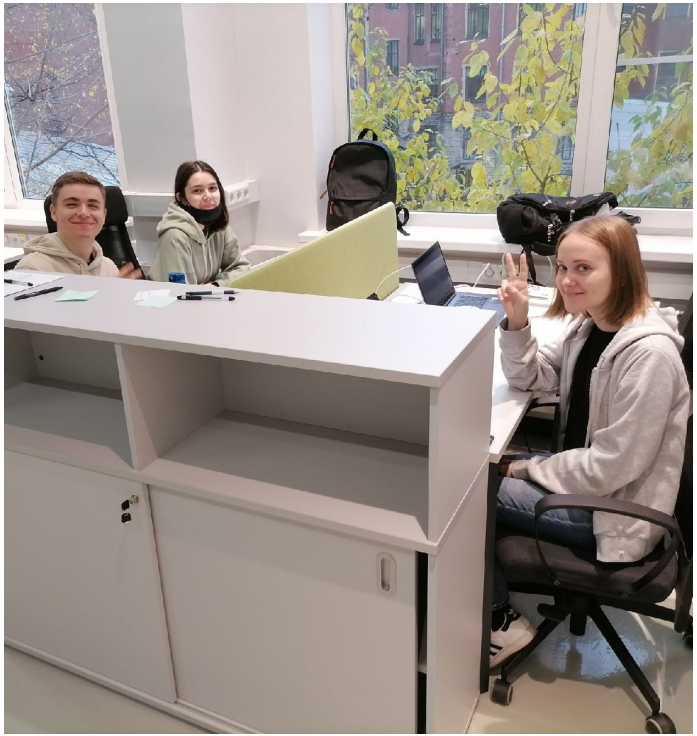
Галягина Ольга [ML]

Горбачева Полина [Data]

Коновалов Илья [Data]

Киреев Сергей [ML]

Кравцов Тимур [F]

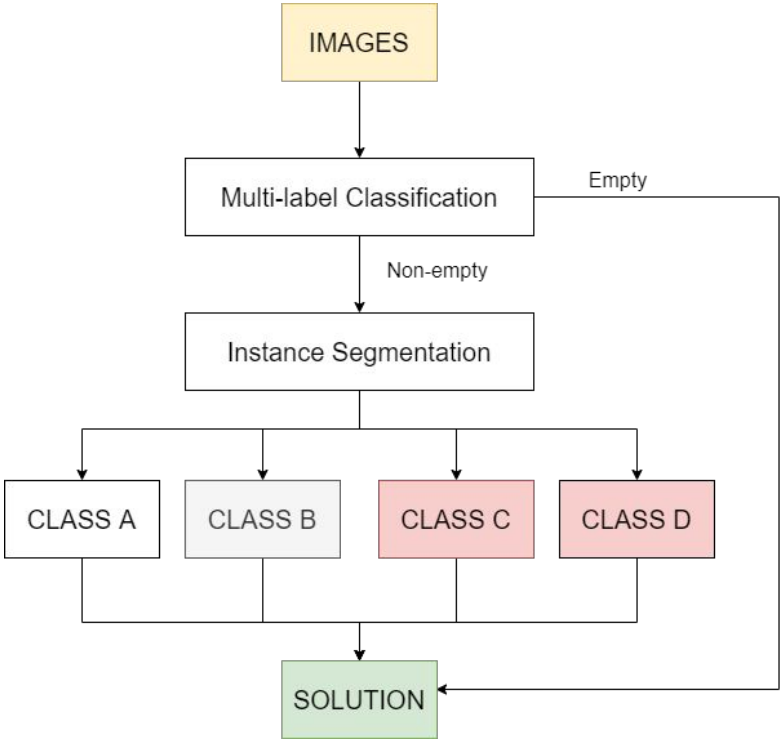


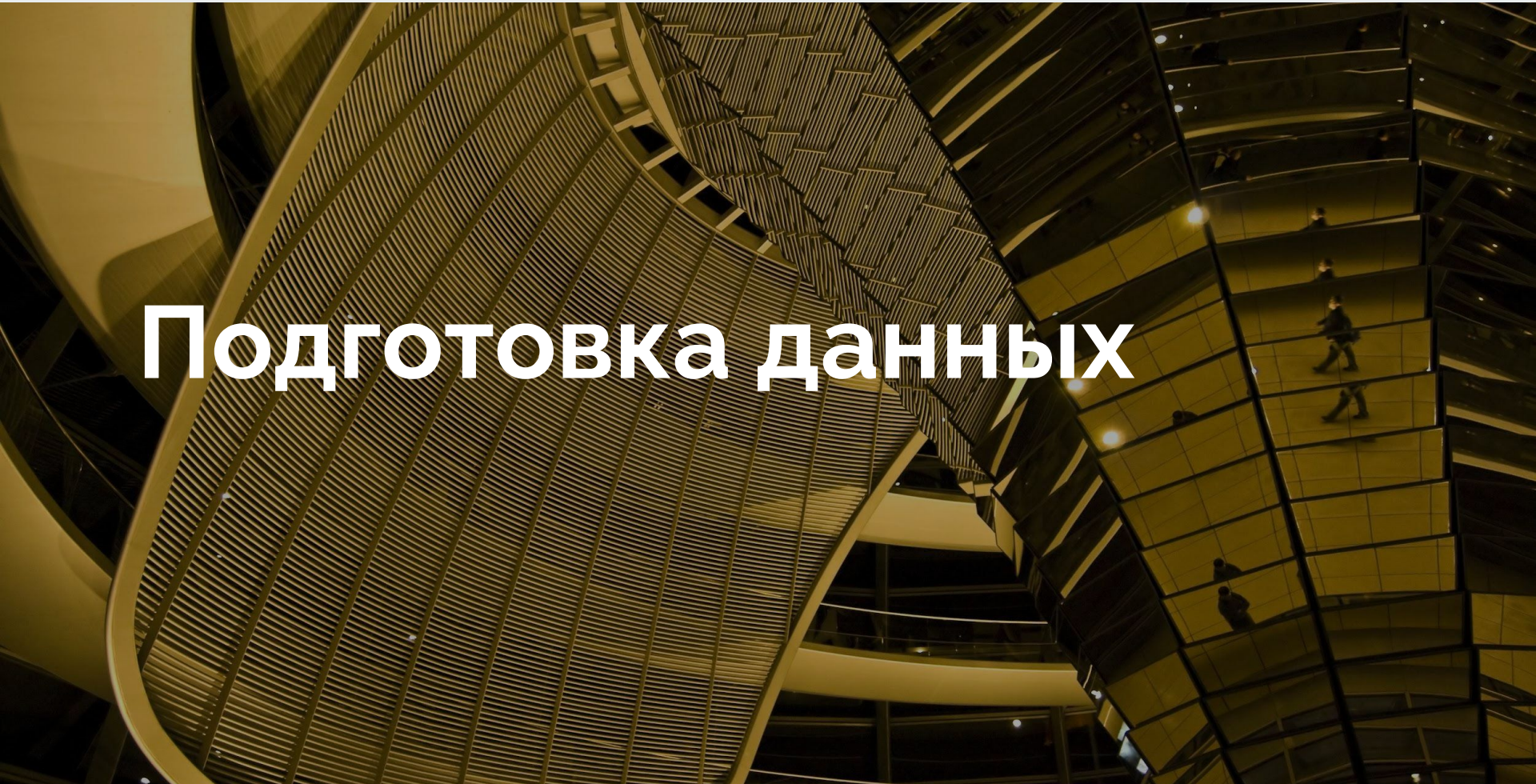
Предложенное решение

Convolutional neural networks (CNN)

Основные инструменты разработки:

- **Python (Detectron2, PyTorch, TensorFlow, Keras)**
- **Google Colab**





Подготовка данных

Разметка данных (160 изображений)

Форматы:

- COCO (.JSON)
- .png

Инструменты:

- CVAT
- LabelME -> labelme2coco

Аугментация данных (160 -> 800 изображений)

Инструменты:

X Augmentor
X albumentations
V **imgaug**

Методы:

· Отражения (вертикальные, горизонтальные)

· Смещения

· Масштабирования

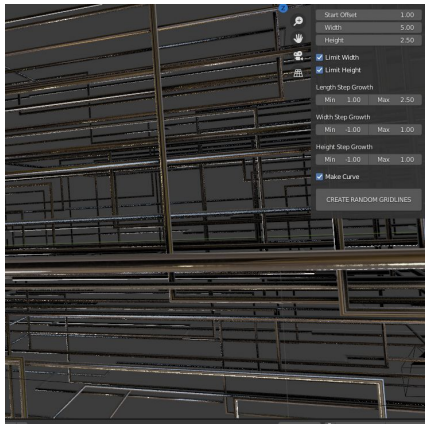
· Изменение яркости

· Изменение контрастности

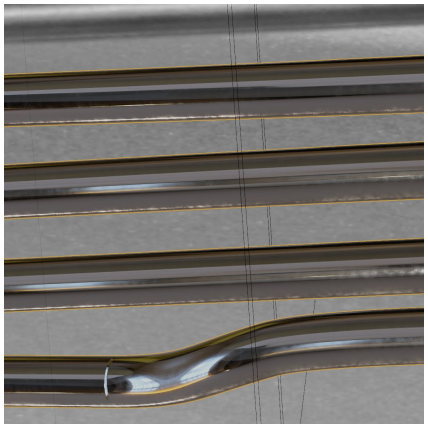
· Наложение шумов

· Размытие

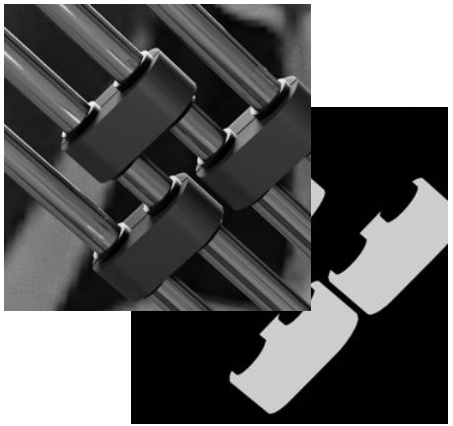
Генерация синтетических данных (800 -> 1000+ изображений)



моделирование структуры



моделирование материала



рендеринг изображений и автоматическая генерация разметки


Выбор ML-модели

Рассмотренные модели

MASK R-CNN (github.com/matterport/Mask_RCNN)
https://proxy.library.spbu.ru:2068/science/article/pii/S0263224121003158

UNET (github.com/zhixuhao/unet)
https://proxy.library.spbu.ru:2281/document/9198370

DEEPLAB (github.com/tensorflow/models/tree/master/research/deeplab)
https://proxy.library.spbu.ru:2068/science/article/pii/S0169260721002844

YOLOv5 (github.com/ultralytics/yolov5)
https://proxy.library.spbu.ru:2068/science/article/pii/S0927025621002512

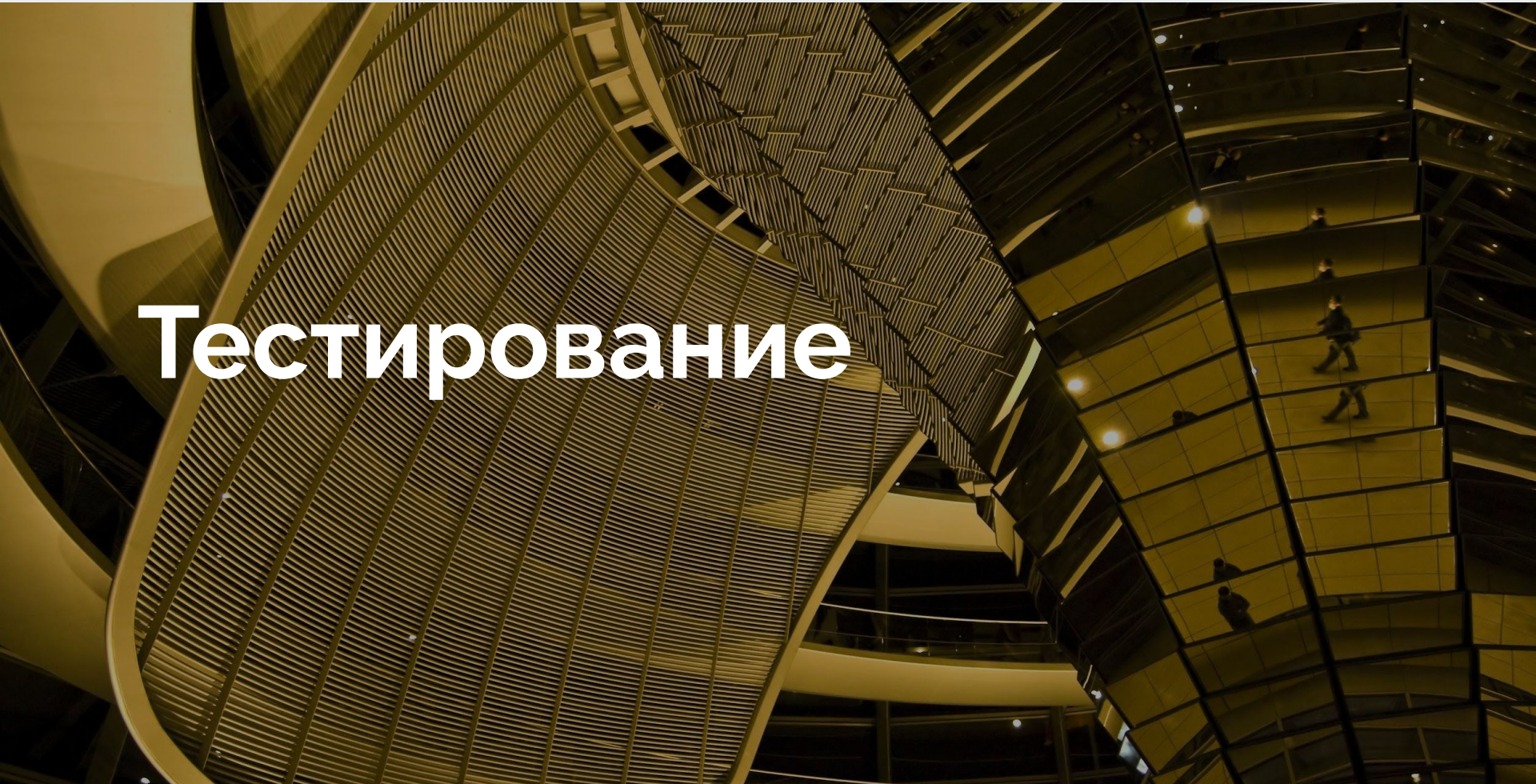
PSPNet (github.com/hszhao/PSPNet)
https://proxy.library.spbu.ru:2068/science/article/pii/S0169260720317302

Panoptic FPN (github.com/facebookresearch/detectron2)
https://proxy.library.spbu.ru:2281/document/8954091

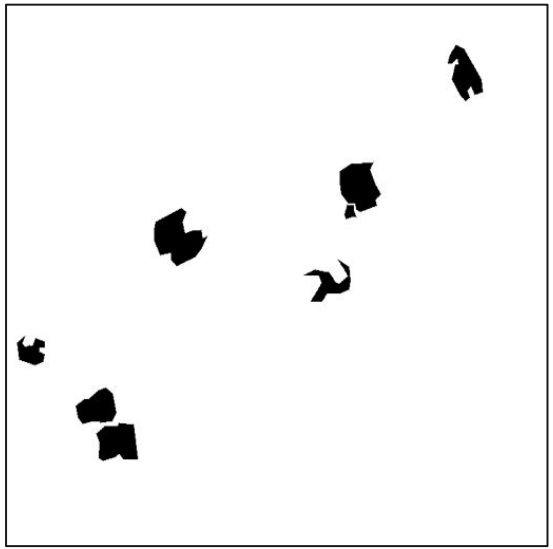
и другие

Parameters	Mask R-CNN	UNET	DEEPLAB
created by	facebook ai research	Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany	Liang-Chieh Chen, George Papandreou, Senior Member, IEEE, Iasonas Kokkinos, Member, IEEE, Kevin Murphy, and Alan L. Yuille, Fellow, IEEE
used for	semantic segmentation, instance segmentation	image segmentation (mostly biomedical)	semantic segmentation
input	MS COCO auto download	rgb images with binary masks	TFRecord format
output	predicted class, box offset, binary mask for each Region of Interest	segmentation map (every pixel is assigned an integer value that represents the segmentation class)	segmentation map
stages	The first stage scans the image and generates proposals (areas likely to contain an object). The second stage classifies the proposals and generates bounding boxes and masks	Trained end-to-end. A U-shaped architecture consists of a specific encoder-decoder scheme: the encoder reduces the spatial dimensions in every layer and increases the channels. On the other hand, the decoder increases the spatial dims while reducing the channels. In the end, the spatial dims are restored to make a prediction for each pixel in the input image.	end-to-end training
metrics	mean average precision, recall, f1-score	IoU, accuracy, precision, recall, f1-score, specificity, Dice score	accuracy, precision, recall, f1-score, Dice score, mean IoU
advantages	efficiency, performance (can run at about 5 fps, 200ms per frame on a GPU), simplicity, flexibility (can be extended on other tasks), pre-trained weights for MS COCO	it is trainable with a small dataset, is computationally efficient, has very good performance on different biomedical segmentation applications	accurate (combining ideas from deep convolutional neural networks and fully-connected conditional random fields)
libraries	python3 keras and tensorflow	Tensorflow, Keras	Tensorflow

Parameters	YOLOv5	TensorMask	Panoptic FPN	PSPNet
created by	Glenn Jocher	Xinlei Chen, Ross Girshick, Kaiming He, Piotr Dollár	Alexander Kirillov, Ross Girshick, Kaiming He, Piotr Dollár (Facebook AI Research)	Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia (The Chinese Univeristy of Hong Kong, Sensetime Group Limited)
used for	object detection	semantic segmentation, instance segmentation	semantic segmentation, instance segmentation	semantic segmentation
input	.txt files, each file is responsible for one image and each line - for one object. There is an auto download for COCO datasets	list[dict], each dict is responsible for one image; there is a ready transformation function for COCO format	COCO Panoptic annotations format	
output	tensors (cx, cy, w, h, conf, classes)	structured high-dimensional tensors to represent image content: (V, U, H, W), where (H, W) is responsible for object position and (V, U) - for relative mask position		segmentation map
stages	one-stage model	dense sliding window method, multi-class classification in parallel to mask prediction	It starts with an FPN backbone, widely used in object detection, for extracting rich multi-scale features. Then a region-based branch on top of FPN is used for instance segmentation. In parallel, a lightweight denseprediction branch is added on top of the same FPN features for semantic segmentation.	Given an input image, firstly, CNN is used to get the feature map of the last convolutional layer, then a pyramid parsing module is applied to harvest different sub-region representations, followed by upsampling and concatenation layers to form the final feature representation, which carries both local and global context information. Finally, the representation is fed into a convolution layer to get the final per-pixel prediction
metrics	Mean Average Precision (mAP)	mAP, IoU, sensitivity	Panoptic segmentation metrics: panoptic quality; Single-task metrics: mIoU, AP	sensitivity, accuracy, specificity, average precision, average recall, mean IoU, f1-score, dice coefficient
advantages	speed, accuracy, is trainable on small datasets	Each mask is itself a 2D spatial mask, which is benefitial for large objects' masks. Reperesenting masks as structured geometric sub-tensors enables the definition of novel operations and network architectures (including coordinate transformation, up-/downscaling, and use of scale pyramids)	Outperforms separate networks on a joint task of panoptic segmentation (semantic+instance segmentation)	
libraries	PyTorch	PyTorch	PyTorch, TensorFlow	PyTorch


Тестирование

Результаты тестирования: Unet



Точная маска

Предсказанная маска

Результаты тестирования: Mask-RCNN

```
[11/21 20:37:10 d2.evaluation.coco_evaluation]: Evaluation results for bbox:
|   AP   |  AP50  |  AP75  |  APs  |  APm   |  APl   |
|:------:|:------:|:------:|:-----:|:------:|:------:|
| 41.006 | 73.090 | 44.345 |  nan  | 22.472 | 42.478 |
[11/21 20:37:10 d2.evaluation.coco_evaluation]: Some metrics cannot be computed and is shown as NaN.
[11/21 20:37:10 d2.evaluation.coco_evaluation]: Per-category bbox AP:
| category   | AP     | category     | AP     | category   | AP     |
|:-----------|:-------|:-------------|:-------|:-----------|:-------|
| clamp      | 47.612 | clamp_metal  | 41.593 | clamp_bb   | 33.813 |
```

Результаты

Требования заказчика удовлетворены:

- представлены результаты работы с данными
- обучена и протестирована модель

Размечено изображений

160

Увеличен объём датасета (раз)

x15

Протестировано моделей

2

Результаты

Повышены квалификации всех участников команды

- Пройдены онлайн-курсы по ML и разметке данных
- Получен и проанализирован опыт командной работы, коммуникации и управления проектом
- Проведено обширное исследование в рамках задачи

Спасибо!