Developing an Image Classifier Using Synthetic Data from CAD Models

Alessandro Morato

Professional Certificate in Machine Learning and Artificial Intelligence

"Vision is the act of knowing what is where by looking" - Aristotle

Overall, computer vision plays a crucial role in a wide range of applications. Its ability to analyse and interpret visual data enables machines to perceive and understand the world around them, leading to numerous practical benefits and advancements in technology.

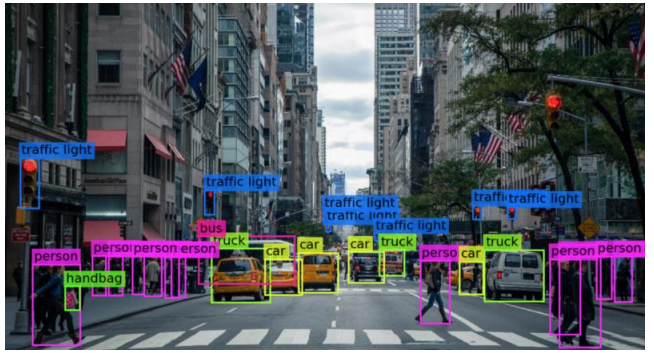
Computer vision is essential in many fields and applications for several reasons

Perception and Understanding: Computer vision enables machines to perceive and understand the visual world, similar to how humans do. Recognize objects, scenes, and patterns, which is crucial for tasks such as object detection, image classification, and scene understanding.

Automation: This includes quality control in manufacturing, sorting and categorizing items in logistics, and monitoring processes in various industries. By automating these tasks, computer vision improves efficiency, reduces errors, and frees up human resources for more complex tasks.

Augmented Reality (AR) and Virtual Reality (VR): Fundamental to AR and VR technologies, which overlay digital information onto the real world or create immersive virtual environments.

Medical Imaging: Tasks such as MRI and CT image analysis, cancer detection, and diagnostic support. Algorithms can analyse medical images to detect abnormalities, assist radiologists in interpretation, and aid in treatment planning.



YOLO

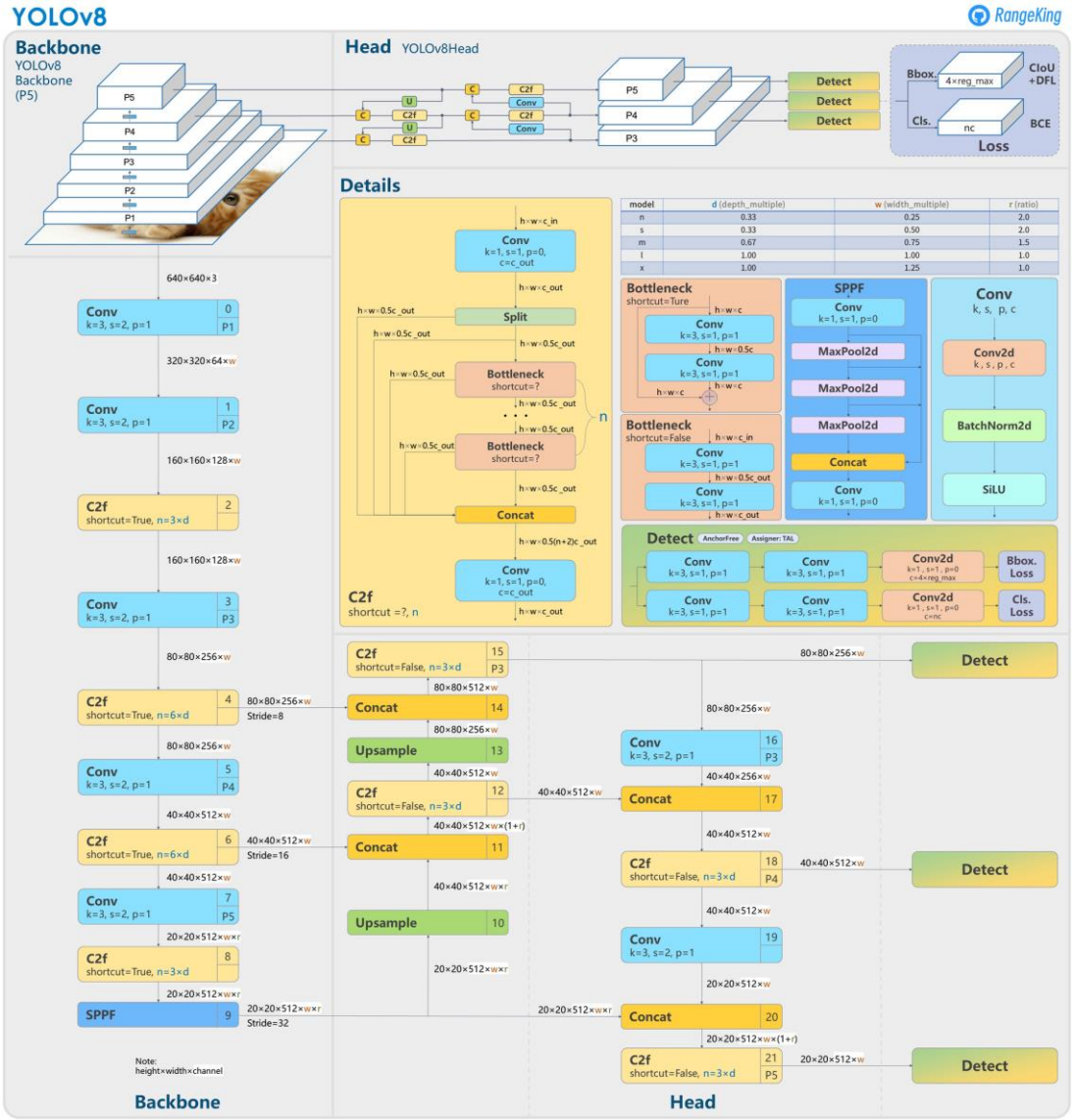
(You Only Look Once)

YOLO is a family of computer vision models

Initially introduced (in 2016) as the first object detection model that combined bounding box prediction and object classification into a single end to end differentiable network. From v5 models are written in the PyTorch framework.

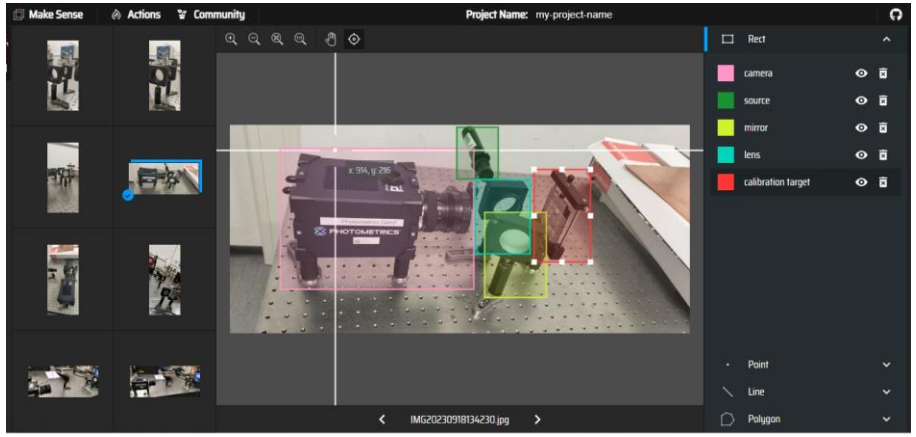
In treating the detection task as a single shot regression approach for identifying bounding boxes, YOLO models are often very fast and very small – often making them faster to train and easier to deploy, especially to edge devices.

YOLOv8 augments images during training online. One of those augmentations is called **mosaic augmentation**. This involves stitching four images together, forcing the model to learn objects in new locations, in partial occlusion, and against different surrounding pixels.



Large amount of images properly annotated

To train an object classifier, we need to supervise its learning with bounding box annotations. If the objects and environment are uncommon, the data need to be labelled MANUALLY.



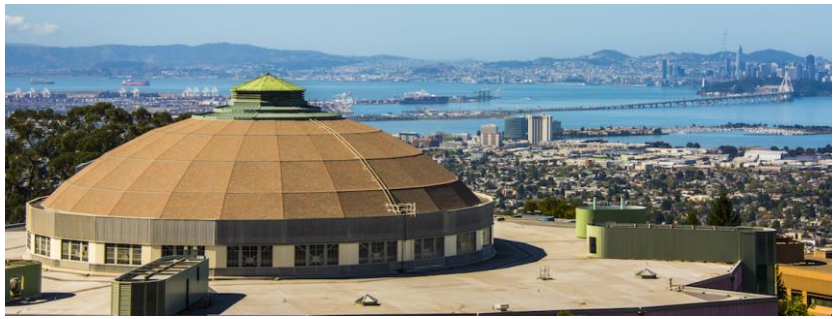
Limited Data Availability: Collecting a large and diverse dataset for object detection can be challenging, especially for custom or specialized domains. Limited data can lead to overfitting, where the model fails to generalize well to unseen data.

Annotation Quality: Annotating objects in images with bounding boxes or segmentation masks requires manual effort and expertise. Poorly annotated data can negatively impact model training and performance.

Question: Is it possible to develop an effective image recognizer using a collection of 'synthetic' images produced by CAD software?

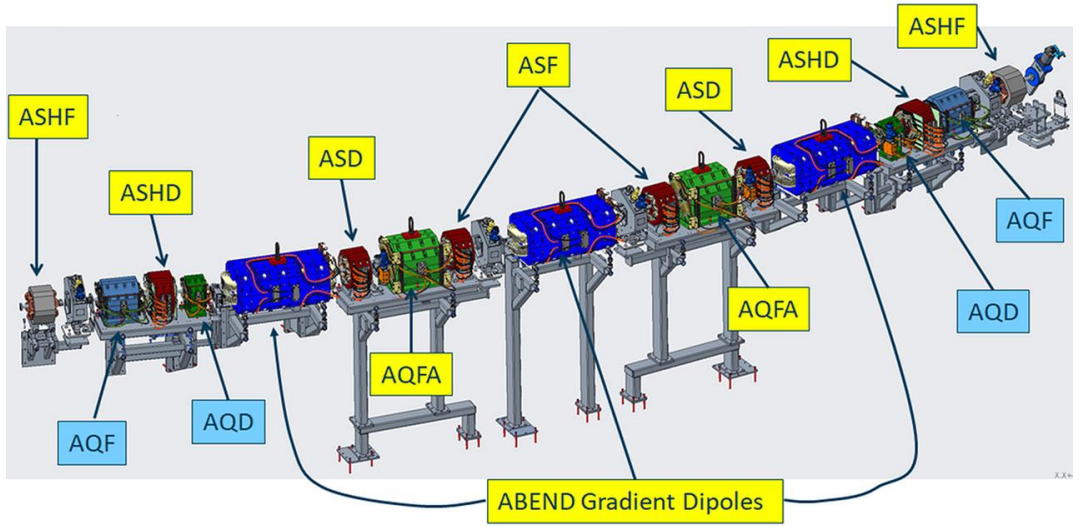
The Application

ALS – Advance Light Source Particle Accelerator (Lawrence Berkeley Lab, Berkeley, CA)



Computer vision to recognize every single part of the facility

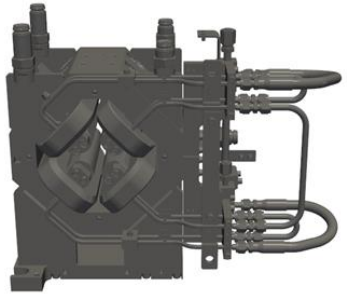
Let's start with a small part: a family of magnets (Accumulator Ring)



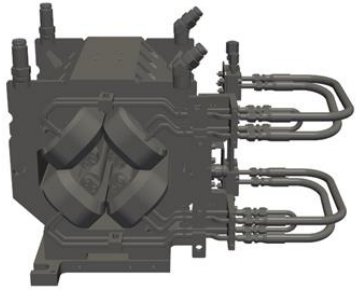
CAD Models

CAD models can be used to produce large amount of images (same object different view angle) and allow for automated labelling.

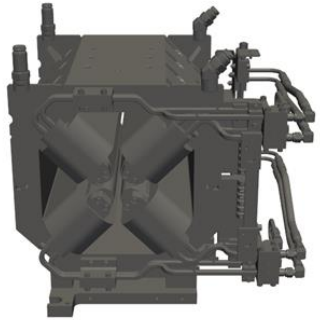
When model trained in combination with real-world images it improved its prediction accuracy.



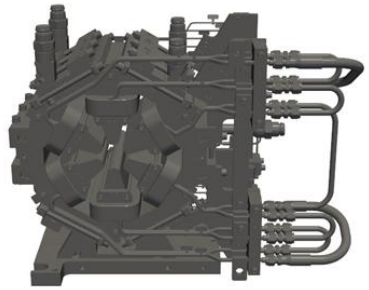
QD



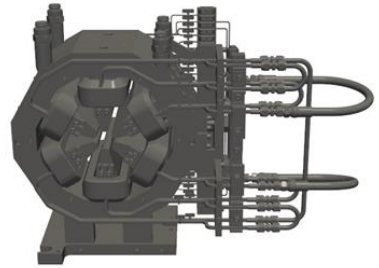
QF



QFA

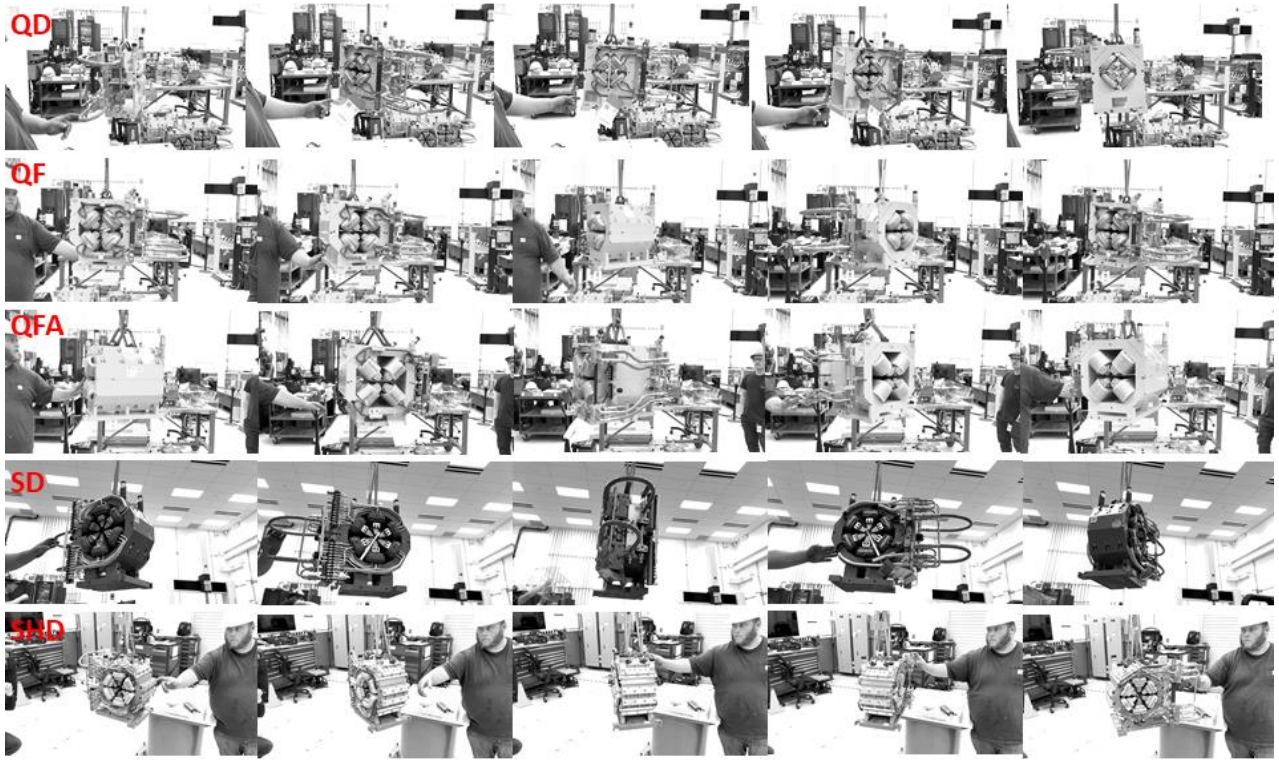


SHD



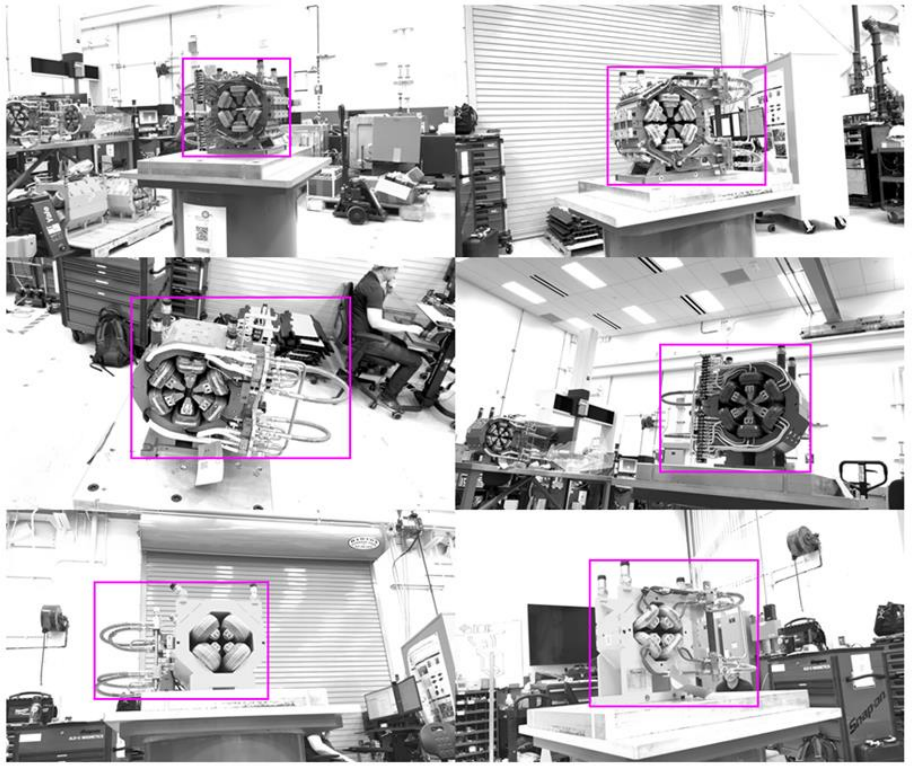
SD

Real-World Images Photo Shooting



Acquisition of Training dataset

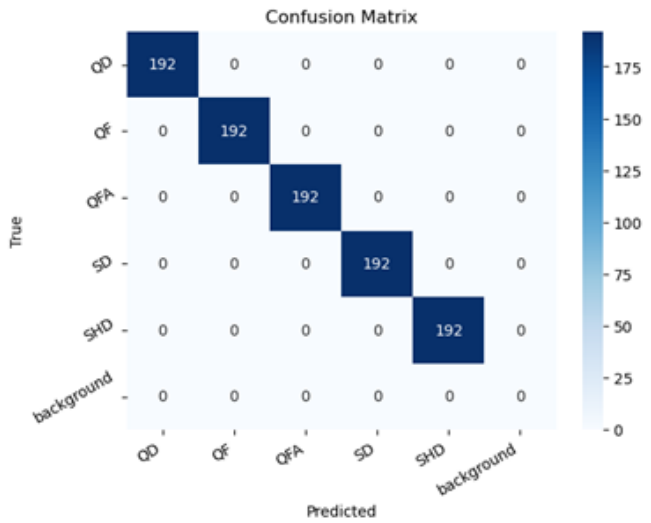
Acquisition of Validation dataset

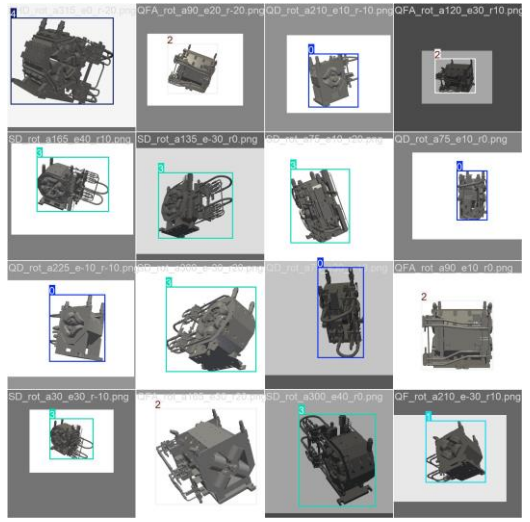


Detect
(full CAD train)

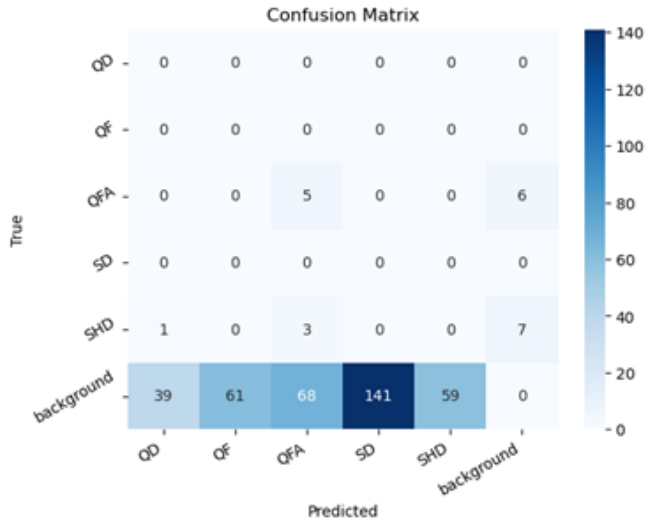
POOR performance in real-world. Almost incapable to recognize any object.

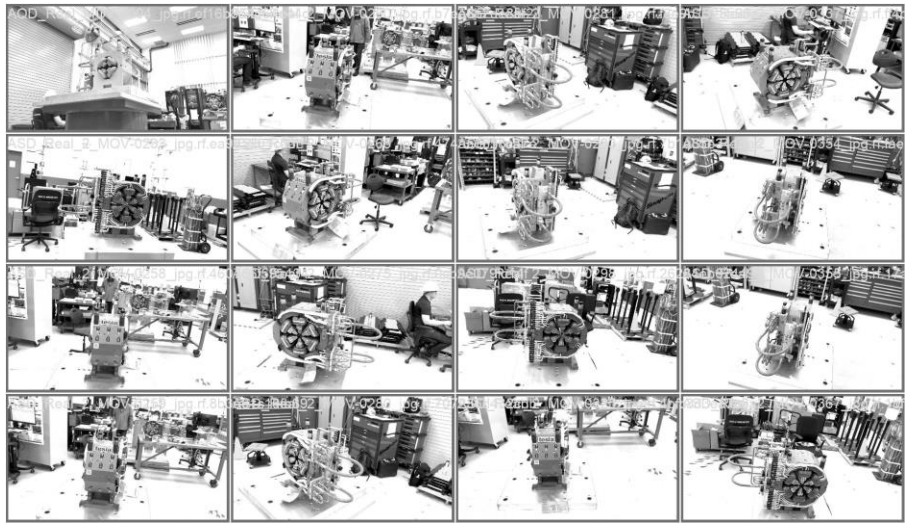
CAD validation dataset





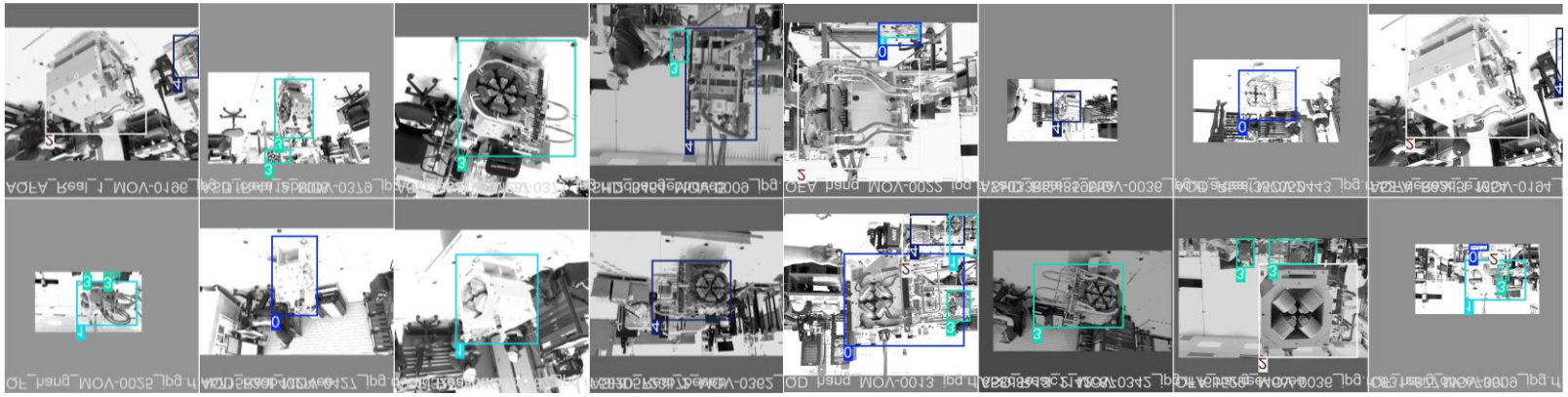
Real-World validation dataset





Berkeley Engineerin

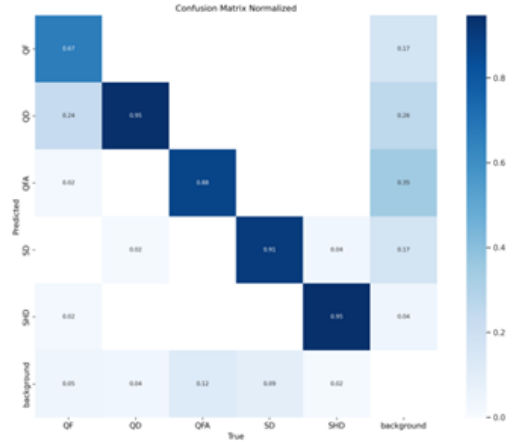
Detect
(full real train)

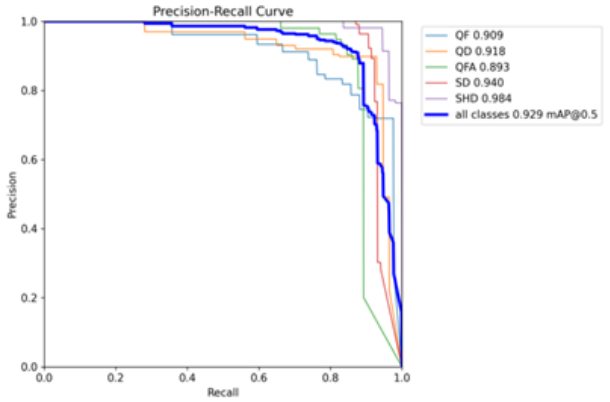


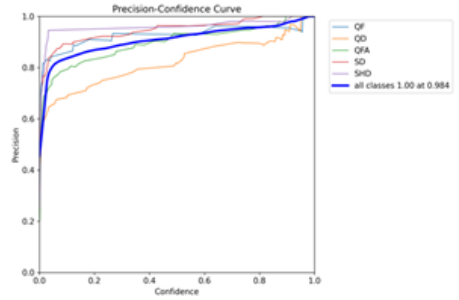
GOOD performance in real-world. Promising…

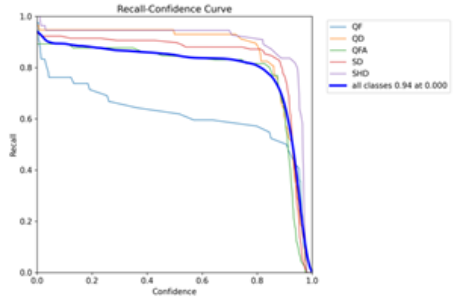
> 600 real-world images

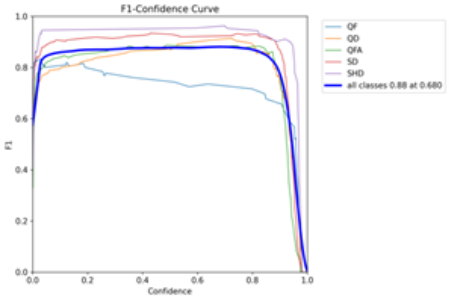
Can we get something similar with much less real-world images in the train dataset?



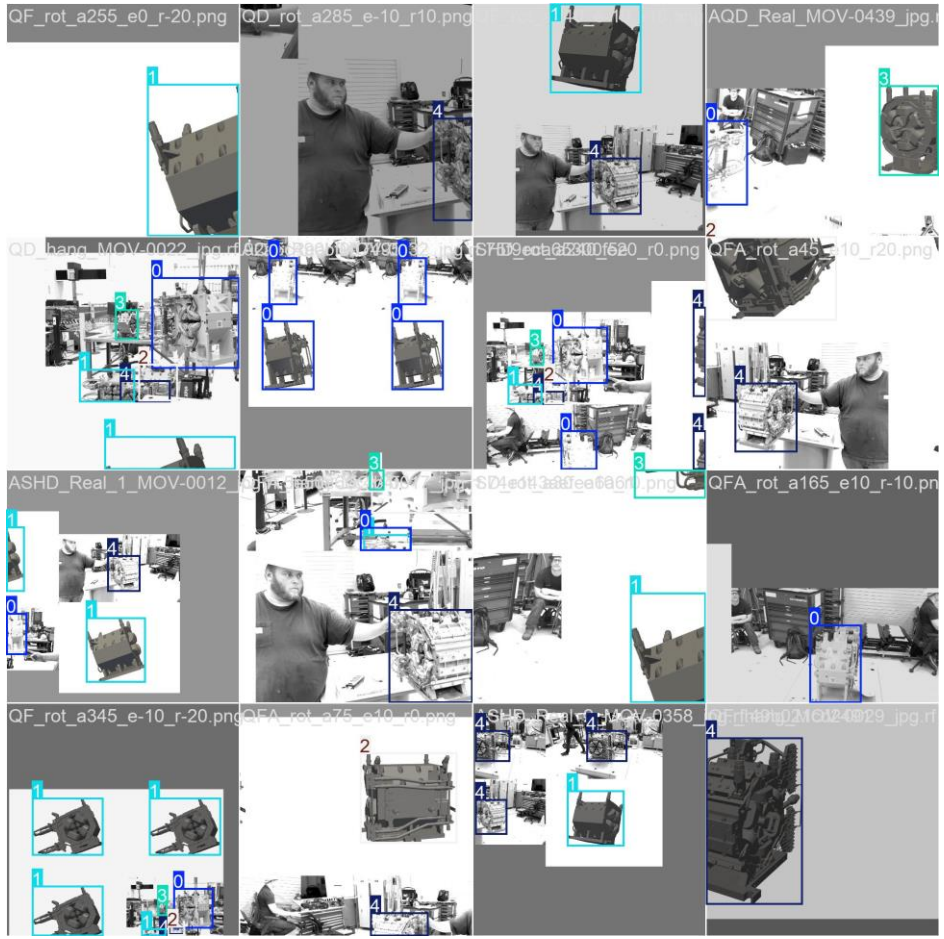






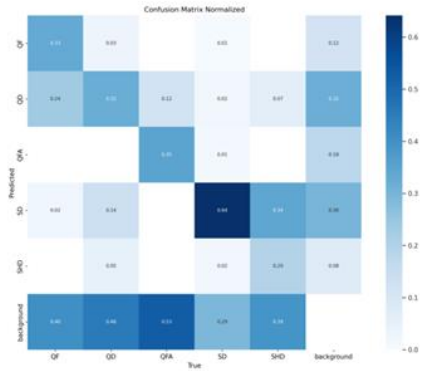


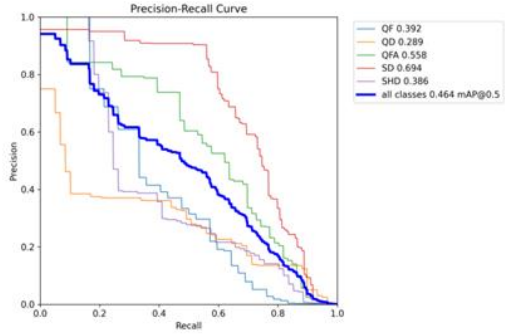
Detect (mixed train)

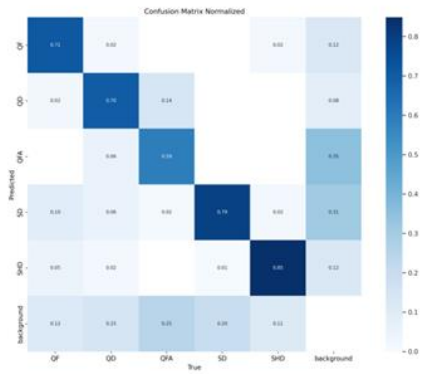


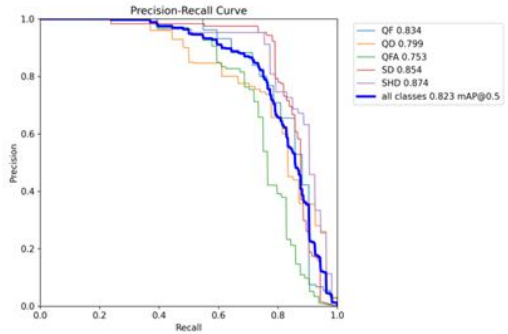
Starting with only 30 real-world train images (5 per class + 5 background), then adding 150 CAD images
NICE IMPROVEMENT!

CAD!

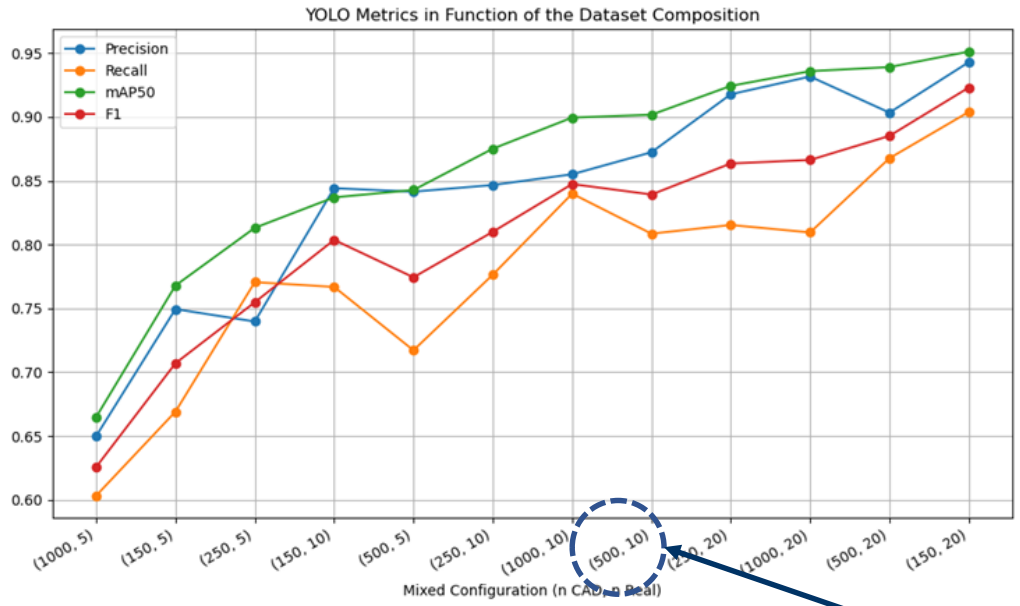








Finding the Right Combination

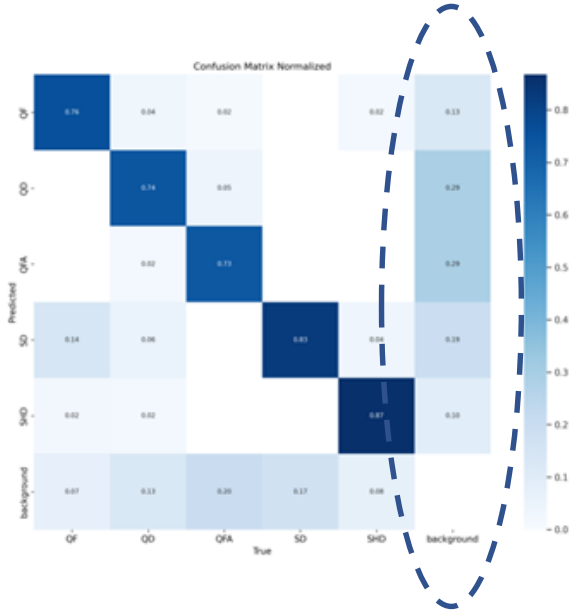


Performance metrics for different combinations of CAD and real images

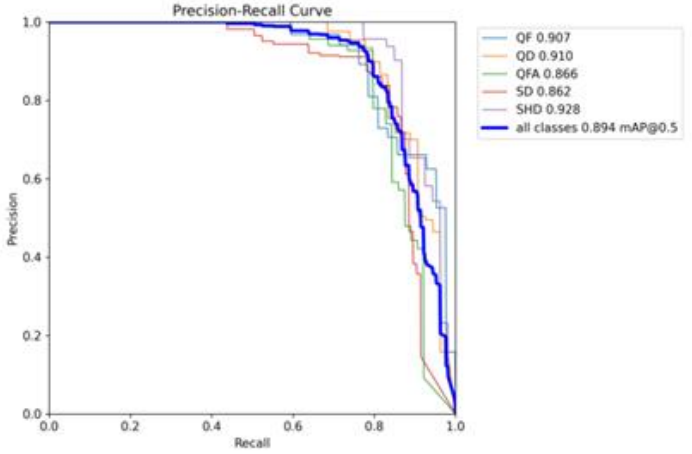
Model	Precision	Recall	mAP50	F1
Full Real	0.946722	0.835058	0.929050	0.887391
Full Real - Reduced	0.545860	0.409720	0.464200	0.468092
Mixed - Simple	0.813192	0.725323	0.822760	0.766749
Mixed - Optimized	0.908239	0.792265	0.892283	0.846297

Mixed Model Selected:
A good blend of real and CAD images

500 CAD (100 per class)
+
60 Real (10 per class + 10 background



WARNING
Still room for improvement: a not negligible number of not-existent objects identified



Next Steps and Recommendations

- Further develop model tuning (explore YOLO tuning functionalities and expand the search for optimal parameters)
- Study the effect of the selected real images: Are some images more effective than others?
- How can the number of false positives (i.e. detection of non-existent objects) be reduced? Is this related to the photo shooting process?
- Analysis performed in greyscale: Would color impact the overall results? And how?

- **Deploy the best model, ideally in a real-time computer vision tool.**

Is it possible to develop an effective image recognizer using a collection of 'synthetic' images produced by CAD software?

YES! CAD images have proven to be a very powerful tool for easily developing an image recognition tool.
However, they cannot be used alone and will always require a minimum number of real-world images to allow the model to correlate the artificial dataset with the real world.

Thank you!