Object Detection Storage System for Workplace Safety

Carl Zhang, Eric Jia, Freya Zhang, Jiamin Wang, Ruoqi Yang, Xinshu Zhao, Nhi Le, Arko Barman

Introduction

- TechnipFMC delivers services in energy operations.
- A crucial part is continuous video monitoring to ensure safety in real time.
- Current video monitoring relies on processes that are time-consuming and onerous.

Objectives

- Identify and train a suitable object detection model
- Build a queryable detected object storage system
- Allow users to efficiently retrieve specific video segments

Workflow

Data Acquisition → Data Preprocessing → Modeling & Evaluation → Inference → Database & Retrival

Data: YouTube-VIS

Source

- 3477 short real-world YouTube clips with rich object motion and appearance changes
- Wide coverage of scenes: person, sports, animals, vehicles, etc.

- Each video provides images with bounding boxes for all annotated objects

Processing

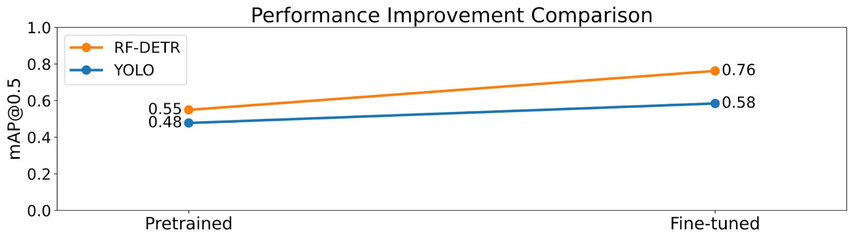
- Selected 18 target categories
- Convert to image-level annotations format
- Subsample uniformly to increase temporal diversity
- Create video-level train/validation/test splits with balanced class distribution

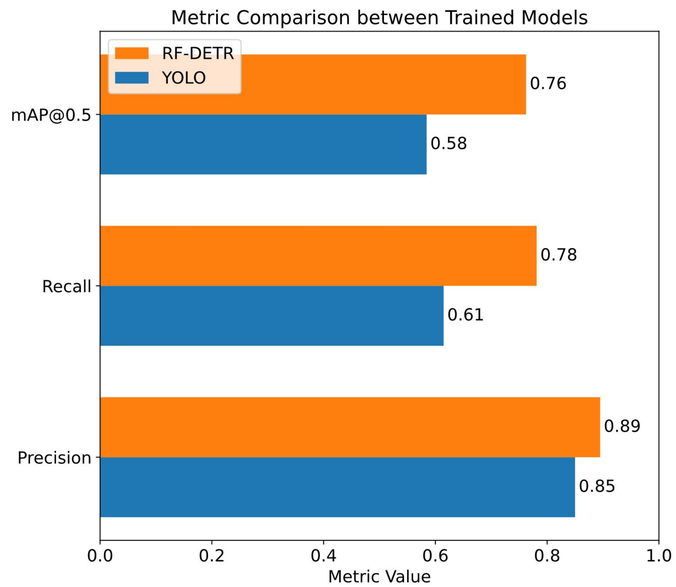
Modeling

- Evaluated RF-DETR(Medium) and YOLO(v12Medium) as candidates for real-time safety monitoring
- Started with pretrained checkpoints and fine-tuned on our training/validation subset
- Ran several training runs per model with varied configurations and compared them using mAP@0.5 and latency to identify the most suitable model

RF-DETR vs. YOLO

- The fine-tuned RF-DETR model achieves superior detection accuracy (mAP@0.5=0.762), significantly outperforming the fine-tuned YOLO (mAP@0.5=0.584).
- However, YOLO is still much faster, at 21.29 ms per image, less than half of RF-DETR's 55.96 ms.
- All evaluations ran on an NVIDIA RTX 2060 GPU.



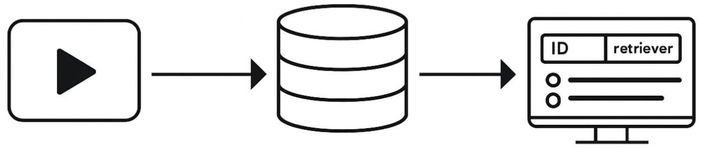


Retrieval System

- Stores all model-generated detections in PostgreSQL database, recording the object type, time range, and key metadata for fast lookup
- Lets users instantly retrieve an intuitive and interactive visualization of results

User Experience: Front-End Panel

- Enter a video ID and object ID
- See the video and its detection intervals, with direct jumps to any video segment



Discussion

- We trained and tested RF-DETR and YOLO models using the public YouTube-Vis dataset
- RF-DETR performs better, with more accurate detections while still being relatively fast
- We designed and built a queryable database system that integrates model output with a flexible front-end

Future Work

- Train models on internal TechnipFMC videos
- Support large-scale batch video ingestion and real-time detection streaming for faster workflows
- Extend to detect actions and to track objects
- Refine the user interface for a more user-friendly experience