

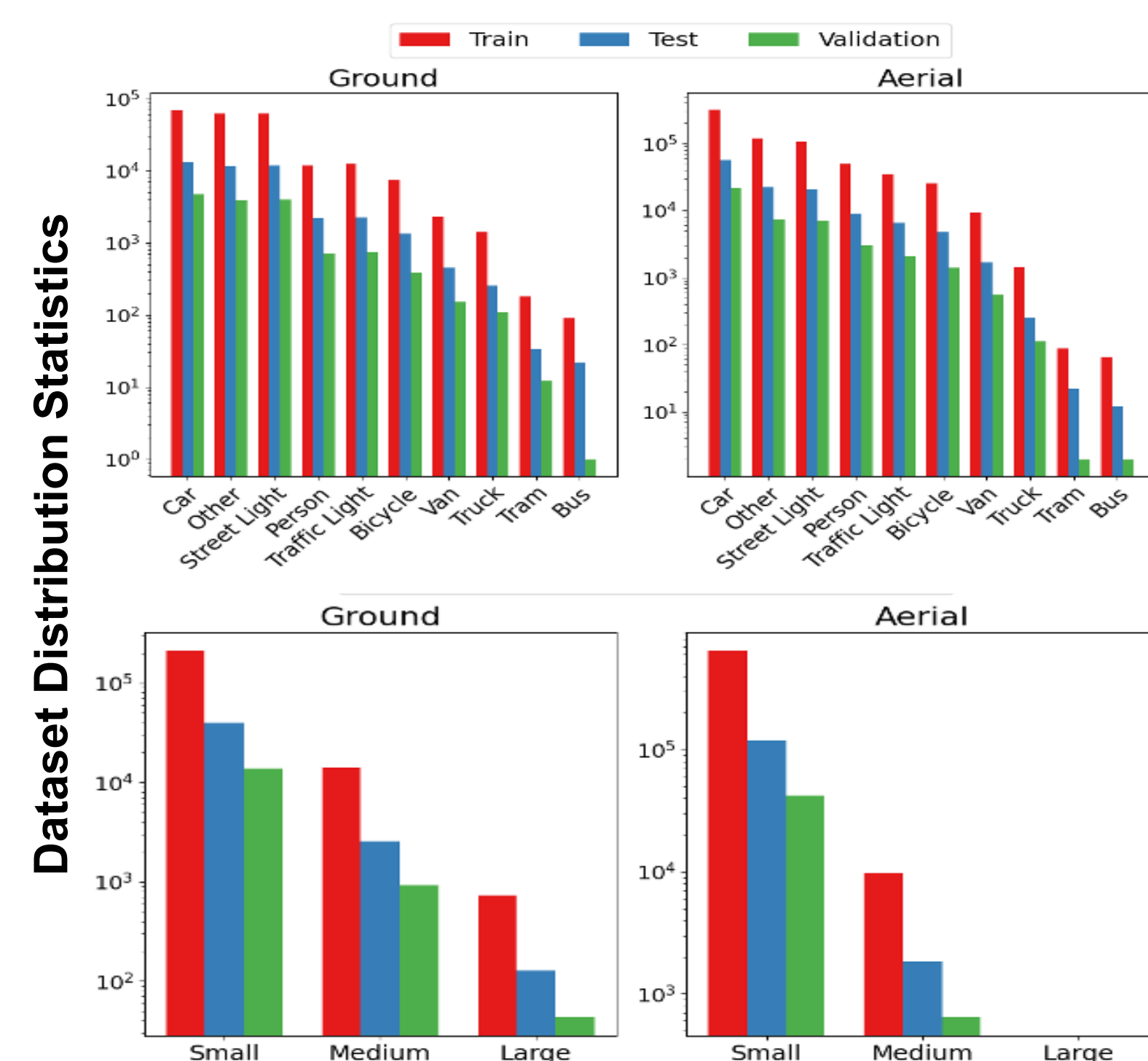
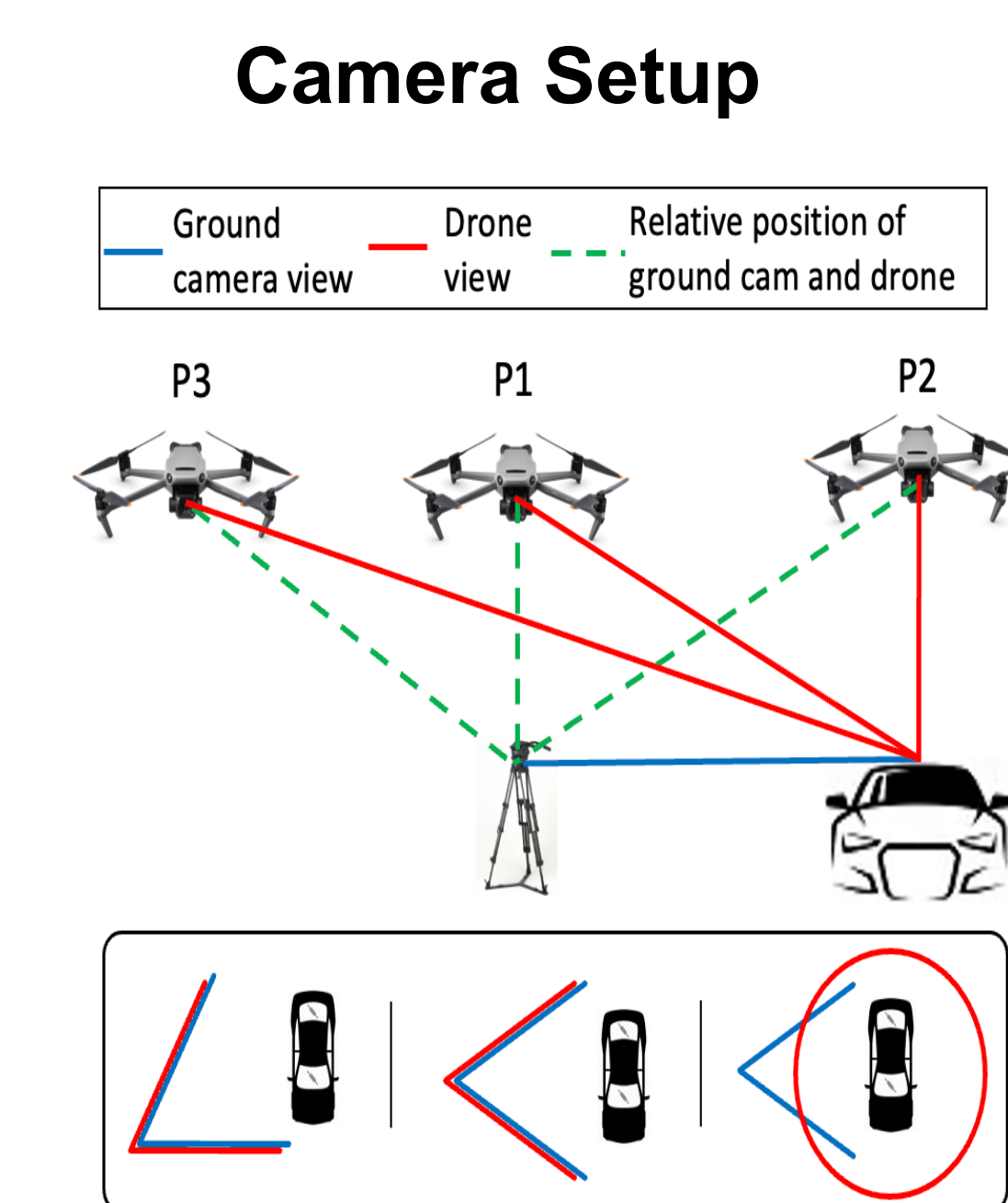
Multiview Aerial Visual Recognition (MAVREC):

Can Multi-view Improve Aerial Visual Perception?

Aritra Dutta, Srijan Das, Jacob Nielsen, Rajatsubhra Chakraborty, and Mubarak Shah

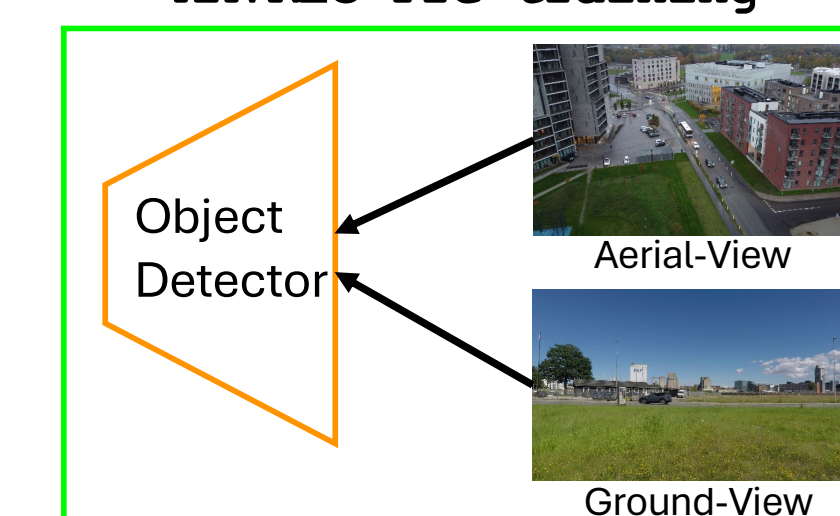


 0.5 M Frames
  2.5 hours of 2.7K resolution video
  10 Object Classes
  12 diverse scenes
  1.1 M Bounding Boxes



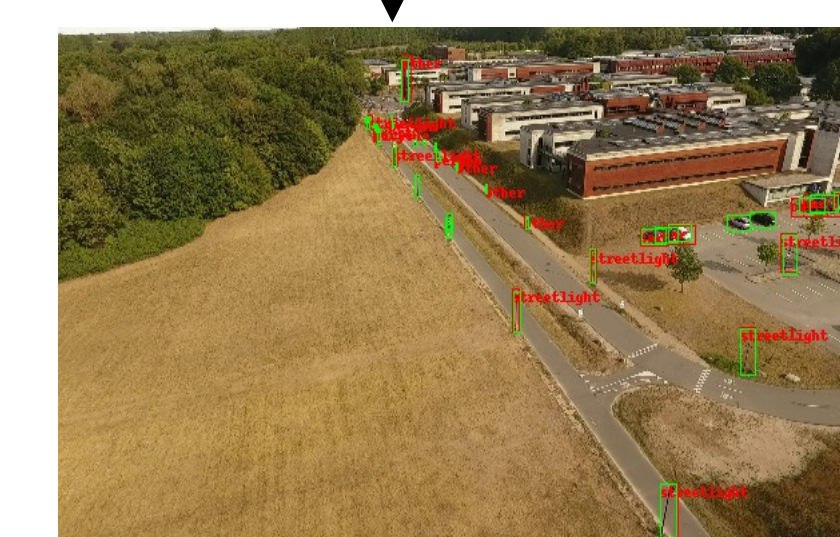
YES

European Landscape
MAVREC Pre-training



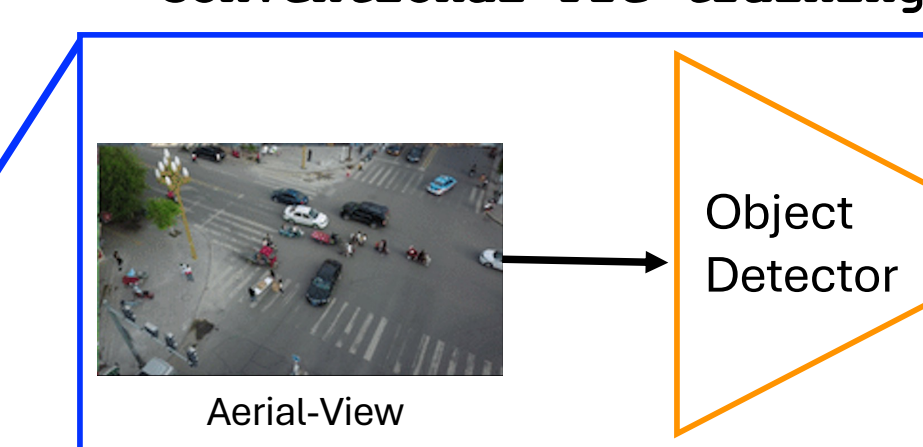
★★★★

Inference



Introducing MAVREC — the largest ground and aerial synchronized dataset

Other Regions
Conventional Pre-training



Inference



SAMPLE FRAMES FROM BOTH VIEWS



INTRODUCTION AND MOTIVATION

There exists no *synchronized ground-aerial multiview datasets*

Literature lacks *datasets originating from diverse geographies*

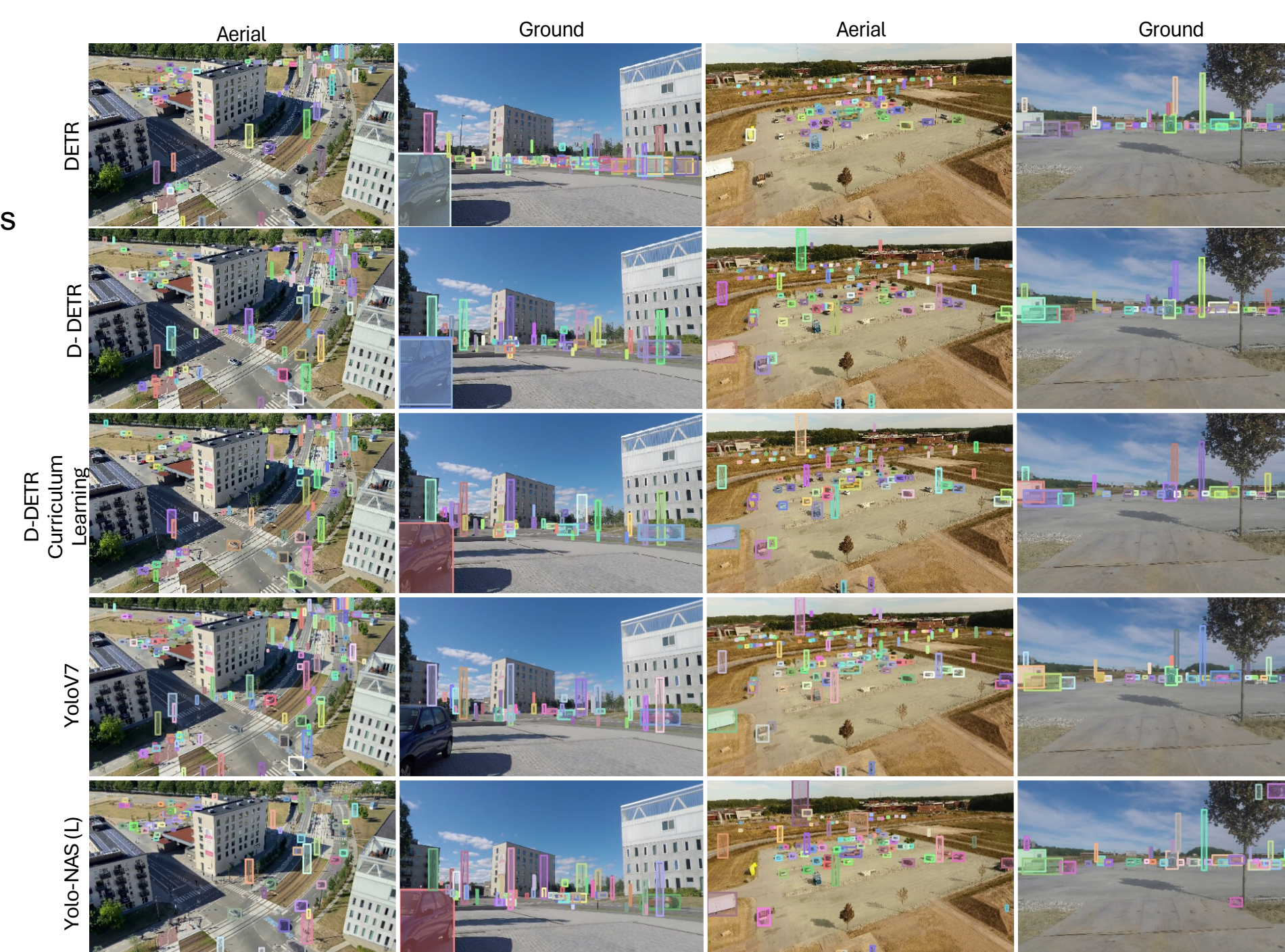
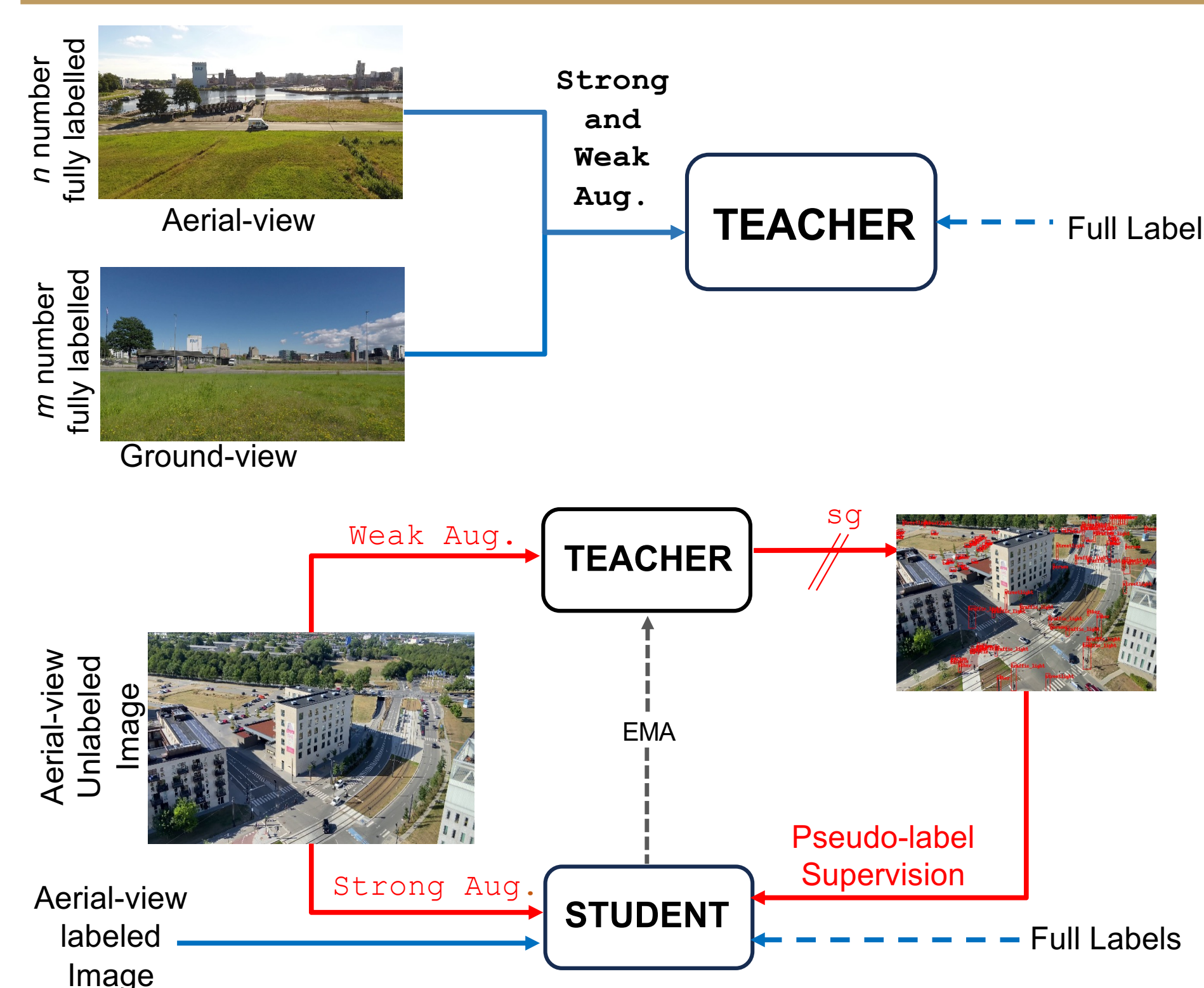
Existing datasets *are low resolution & small-scale*

Data curation ignores effects of multiple *factors such as solar zenith angle, latitude, longitude, etc.*

UNIQUE CHALLENGES IN MAVREC

- ✓ **Diverse object density** presents challenges in training object detectors compared to uniformly dense scenes.
- ✓ MAVREC's aerial perspective includes **small objects** that are prone to **misdetetection** by detection algorithms.

CURRICULUM LEARNING BASED SSOD



Qualitative Inference Results on MAVREC

BENCHMARKING MAVREC

Training Protocol	Pre-training	Test Set			
		AP	AP ₅₀	AP _S	AP _M
Trained from scratch	×	10.3	25.0	10.1	29.4
Grounding-DINO	O365, GoldG, Cap4M	20.4	40.9	18.6	32.5
FT on MAVREC Aerial view	Visdrone (Zhu et al., TPAMI 2022)	20.9	41.9	20.6	43.8
FT on MAVREC Aerial view	MS-COCO (Tsung et al., CVPR 2014)	33.2	61.9	31.5	51.0
FT on MAVREC Aerial view	MAVREC Ground view	44.8	71.5	42.9	72.4

