JUNE 18-22, 2023
CVPR VANCOUVER, CANADA
MultiEarth Workshop 2023

Debiased Learning from Naturally Imbalanced xView Data

Chun-Hsiao Yeh[1], Xudong Wang[1], Stella X. Yu[1,2], Charles Hill[3], Zackery Steck[3], Scott Kangas[3], Aaron A. Reite[4]

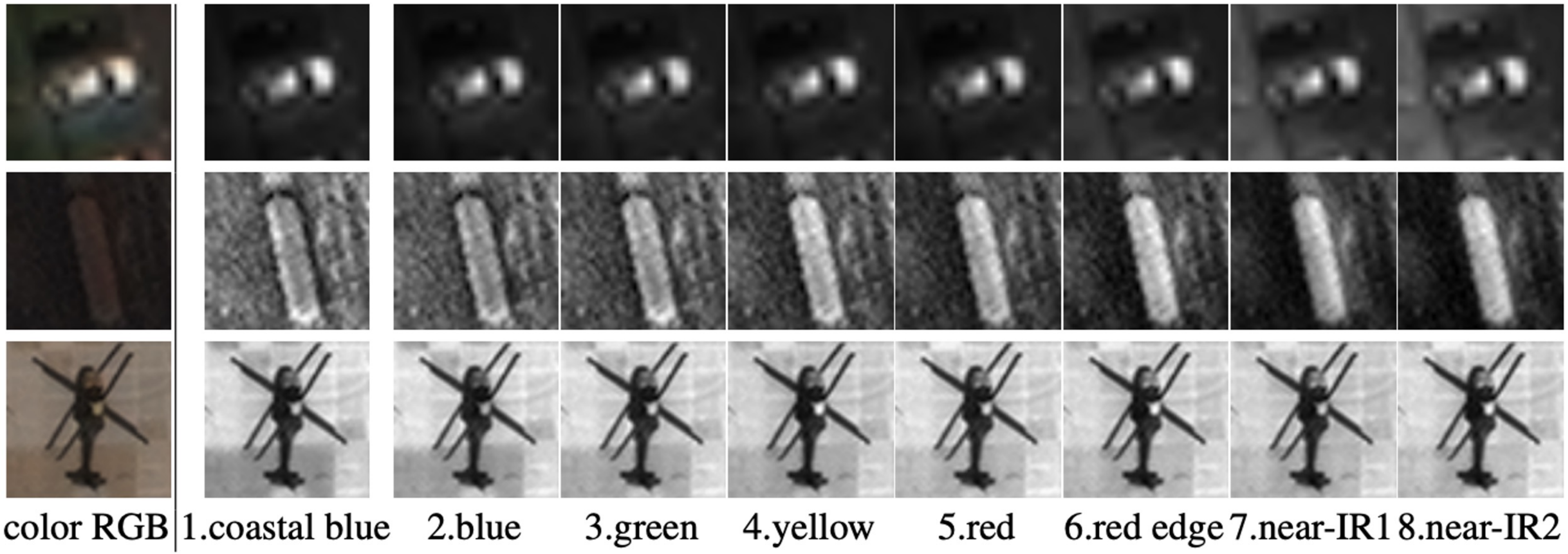
[1]UC Berkeley [2]University of Michigan [3]Etegent Technologies, Ltd [4]NGA

06/19/2023

This work was supported, in part, by the National Geospatial Intelligence Agency / Etegent Technologies Ltd. contract HM047620C0063.

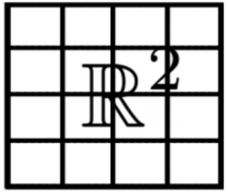
Our Task: Remote Sensing Data Classification

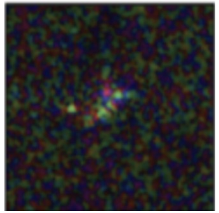
8-Band Images (xView Dataset)



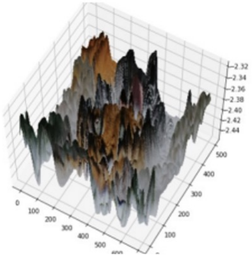
An Image is a Function from Domain to Co-Domain

Domain: Pixel Locations (x,y)

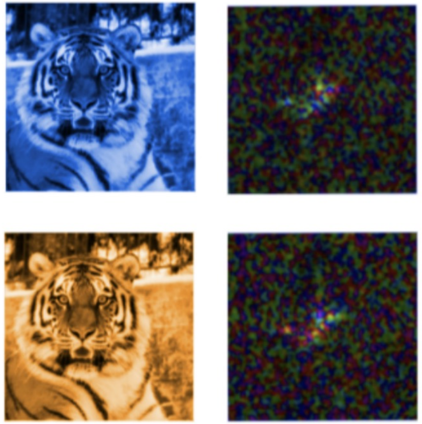




Co-Domain: Pixel Values (RGB)

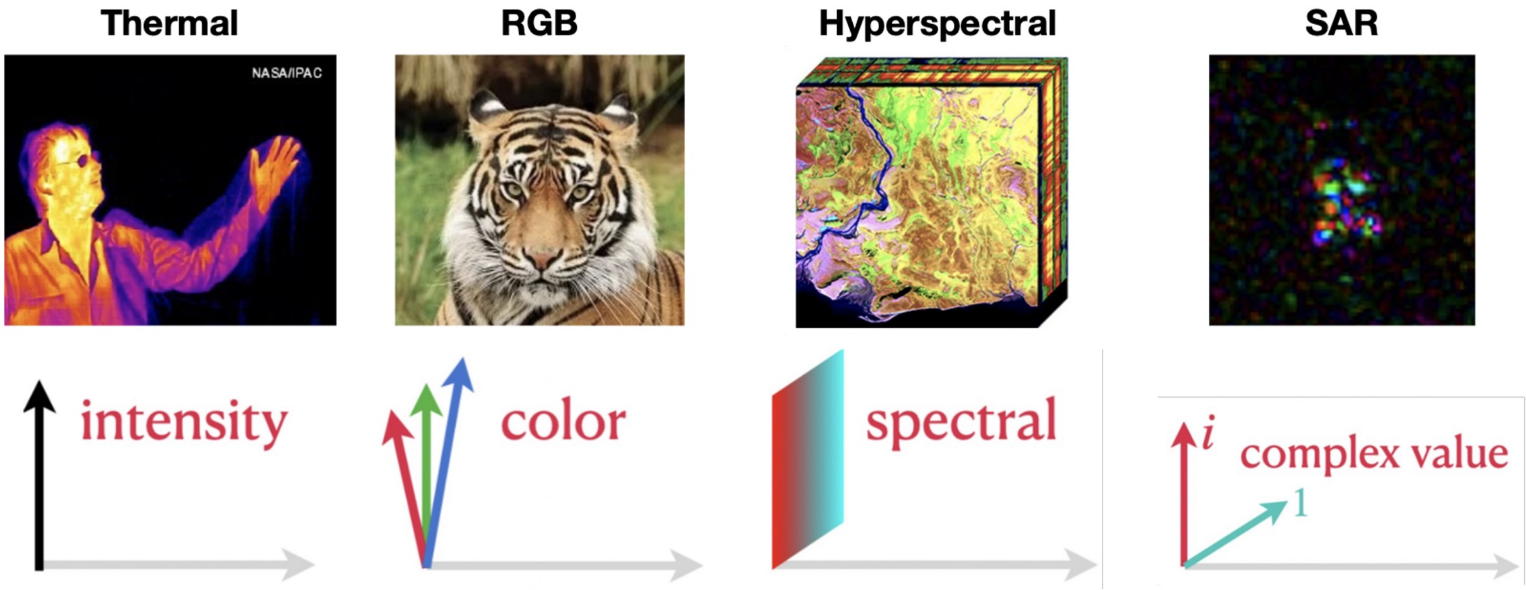


scaling   rotation



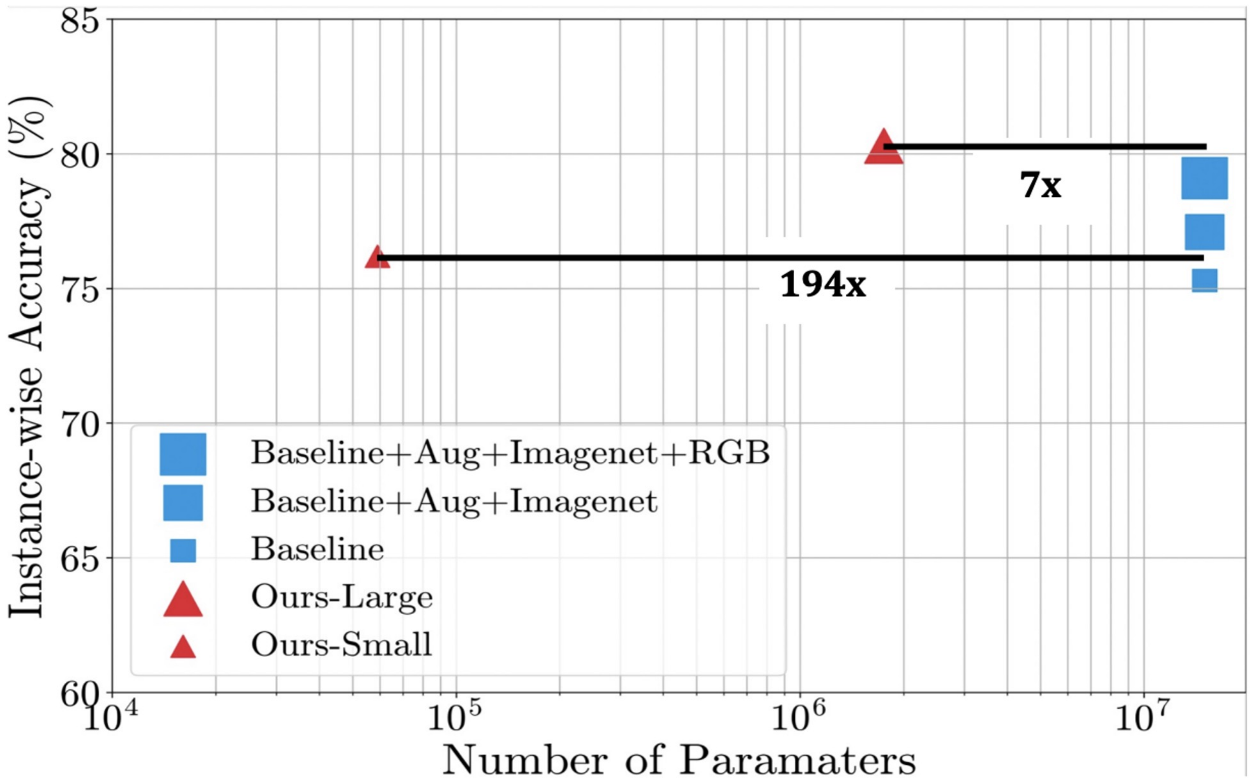
color   C-scaling

Co-Domain is Related to Diversity of Image Types



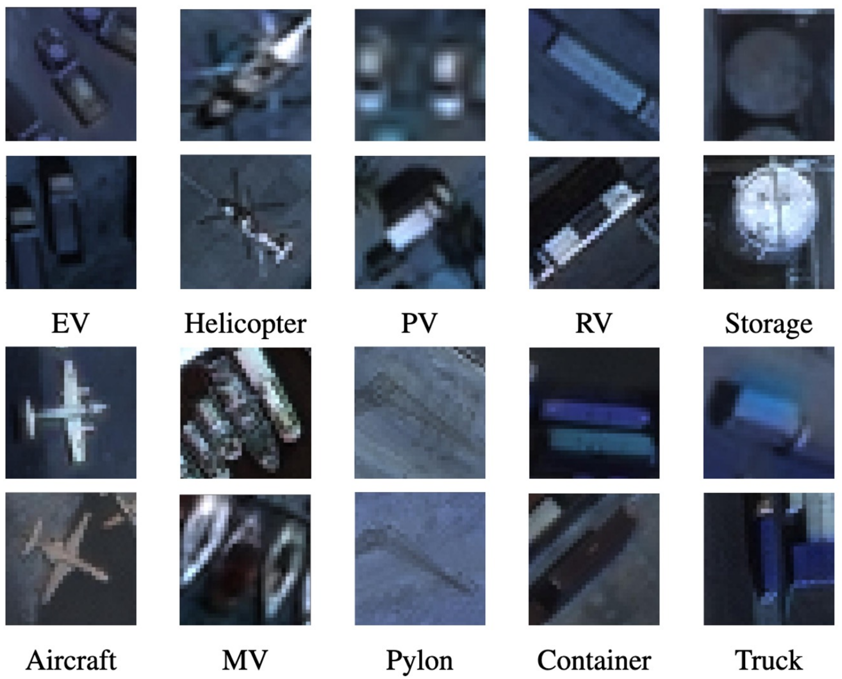
Singhal, Utkarsh, Yifei Xing, and Stella X. Yu. "**Co-domain symmetry for complex-valued deep learning**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

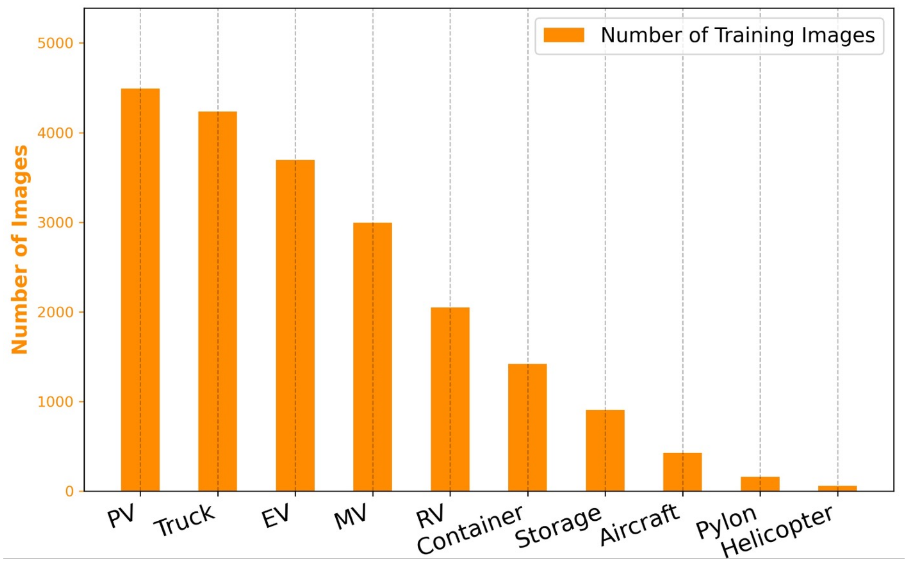
xView Results: Simpler and Better Ultra-lean Models



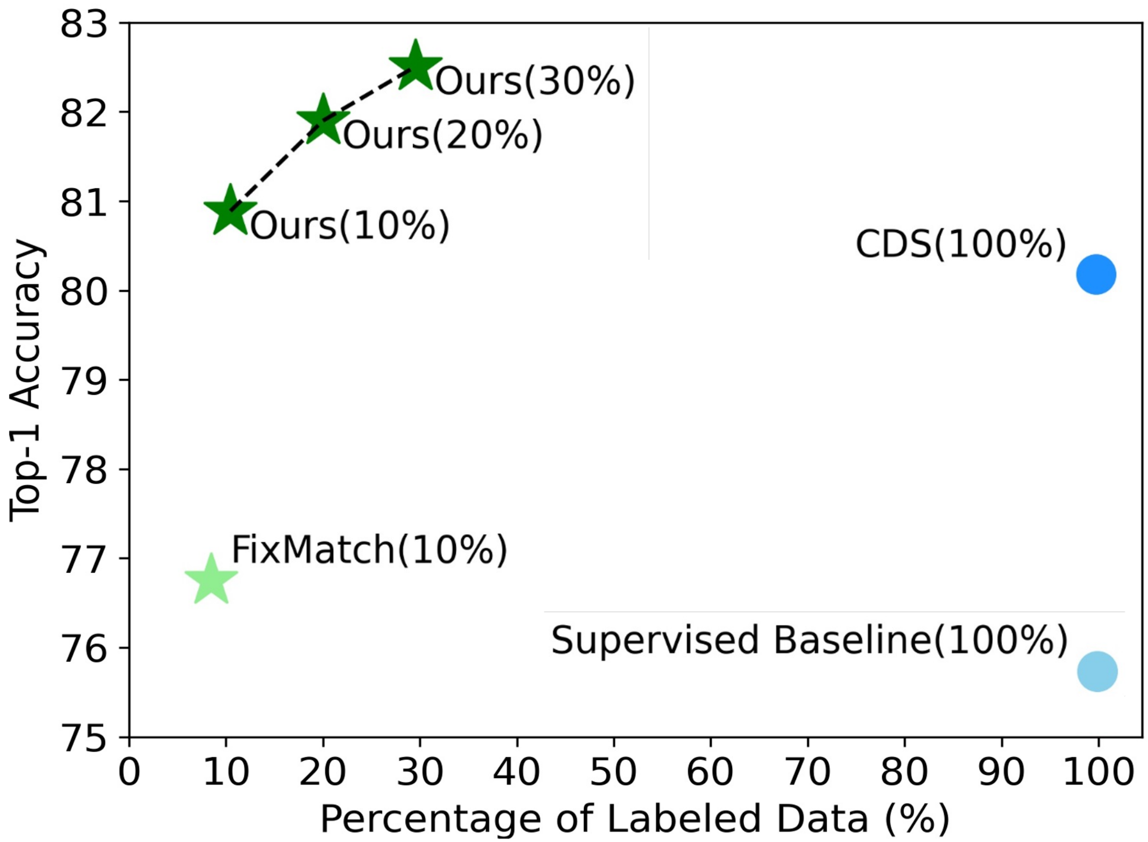
Our Task: Remote Sensing Data Classification

xView Dataset



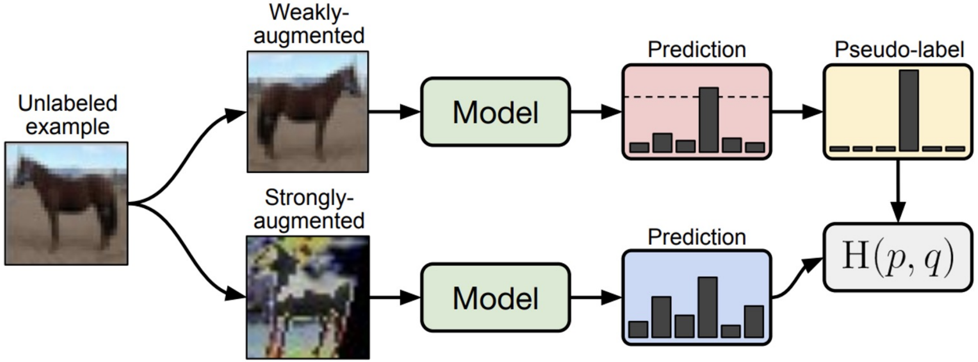


Our Task: Remote Sensing Data

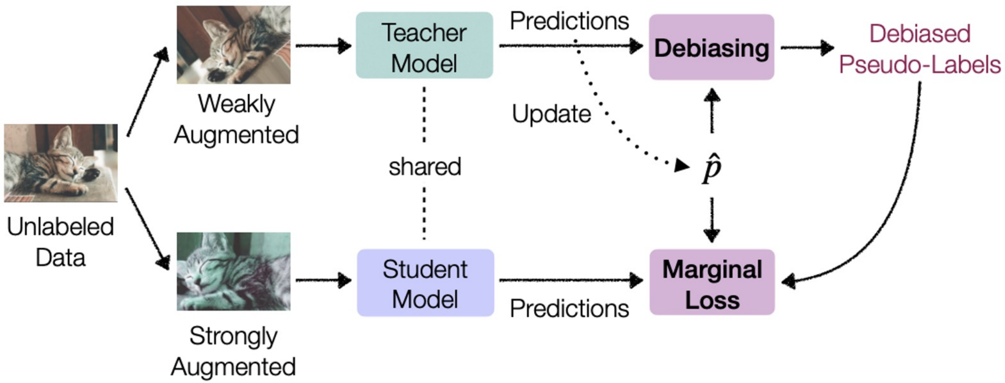


Related Semi-Supervised Learning Methods

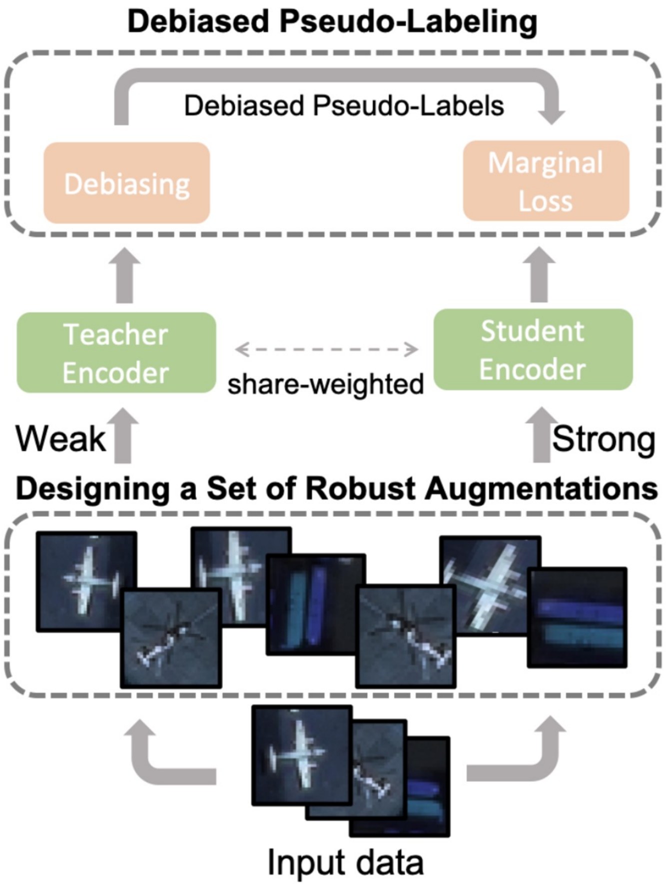
FixMatch (NeurIPS'20)



DebiasedPL (CVPR'22)



Our Semi-Supervised Learning Framework



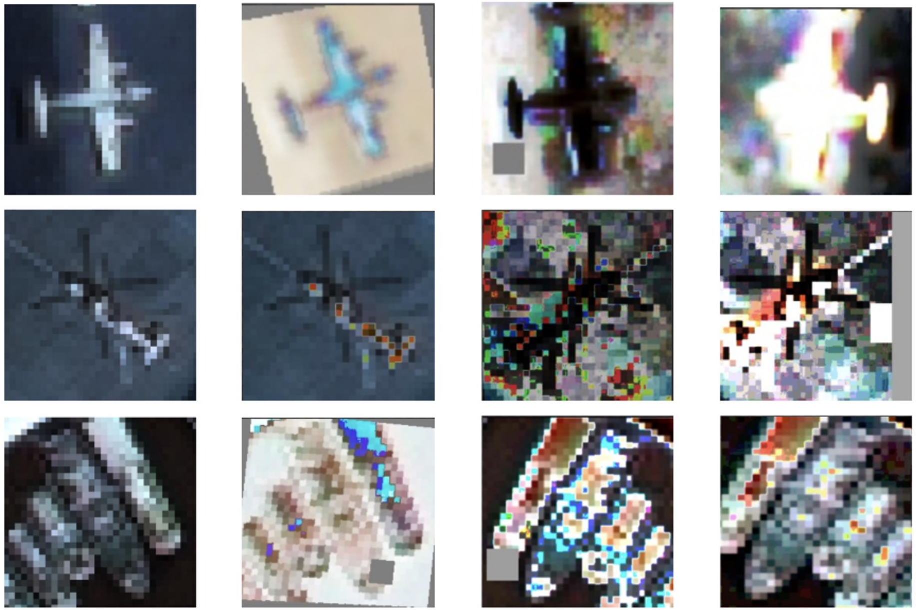
Contribution #1: Robust Augmentations

Design a set of robust strong and weak augmentations suitable for remote sensing data.

Contribution #2: Debiased Pseudo-Labeling

Leverage DebiasedPL to mitigate the bias in pseudo-labeling.

RandAugment Appears to be too Complicated for Remote Sensing Data



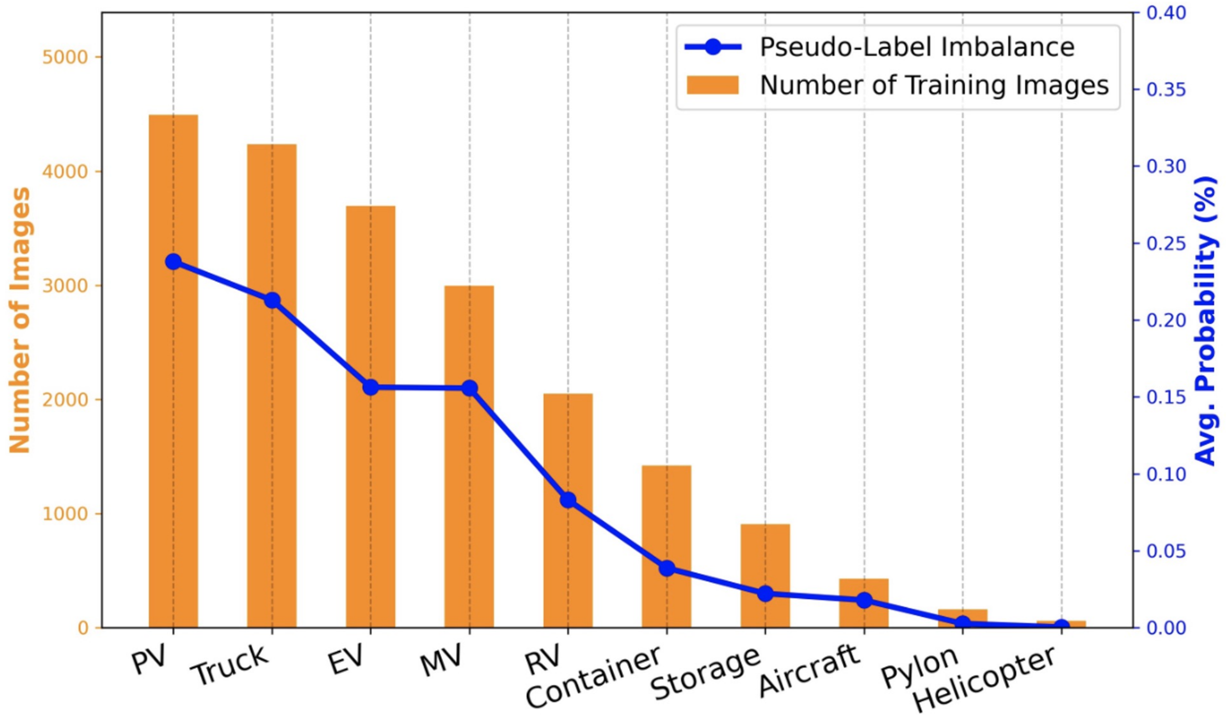
Original | RandAugment ($m = 6$) | RandAugment ($m = 8$) | RandAugment ($m = 10$)

Our New Findings: Data Augmentations for Remote Sensing Data

Augmentations	Used	Reason
RandAugment [8]	×	The use of RandAugment with high intensity of transformations may result in overfitting to the training set, which can be detrimental for generalization on remote sensing data.
Rotation	✓	random rotation (± 10 degrees) can simulate variations in remote sensing imaging angles.
Scaling	✓	scaling(0.8, 1.2) can simulate variations in ground sampling distances, which is important for generalization on remote sensing data.
Horizontal Flip	✓	horizontal flipping simulates mirror-reflected scenes in remote sensing data

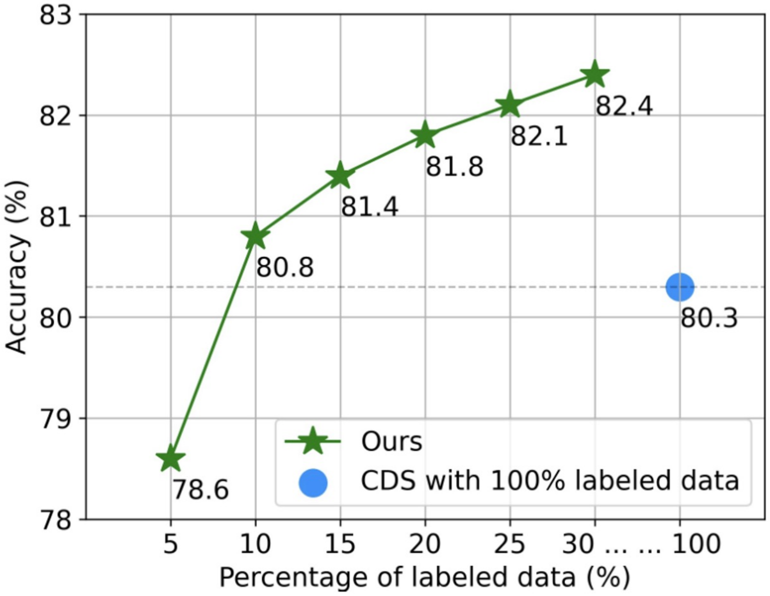
Weak augmentation						
None	✓	×	×	✓	×	×
Random Resize Cropping	×	✓	✓	×	✓	✓
Horizontal flipping	×	✓	✓	×	✓	✓
Strong augmentation						
ResizeCropping + Horizontal flipping	✓	✓	✓	✓	✓	✓
Rotation(± 10 degrees)	×	×	×	✓	✓	✓
Scaling(0.8, 1.2)	×	×	×	✓	×	✓
RandAugment(m=10)	✓	✓	×	×	×	×
RandAugment(m=5)	×	×	✓	×	×	×
Top-1 accuracy (%)	69.8	75.6	76.3	79.4	79.7	**80.8**

Two Sources of Imbalance (Training Label & Pseudo Label)



Quantitative Results

Increasing the amount of labeled data



Class-wise accuracy