國立臺灣師範大學
NATIONAL TAIWAN NORMAL UNIVERSITY

APSIPA ASC 2024

Improving Semi-Supervised Object Detection by ROI-Enhanced Contrastive Learning

Teng-Kuan Huang and Mei-Chen Yeh

Department of Computer Science and Information Engineering

National Taiwan Normal University

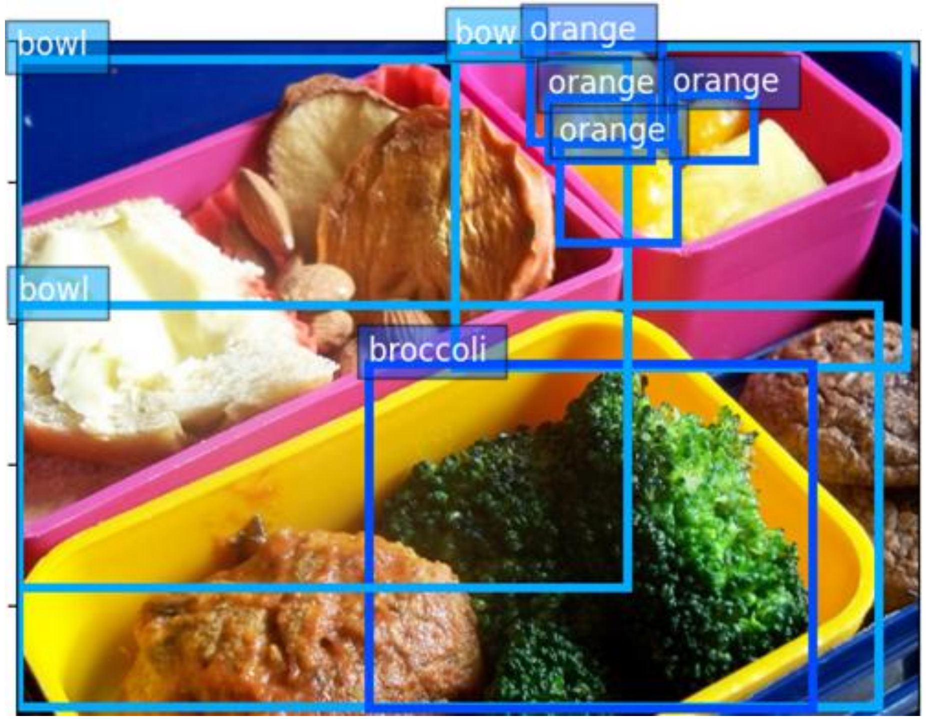
Object Detection

- Detects and locates objects of interest in an image

Semi-Supervised Object Detection

- Utilizes labeled and unlabeled images to train the object detection model

Labeled Images

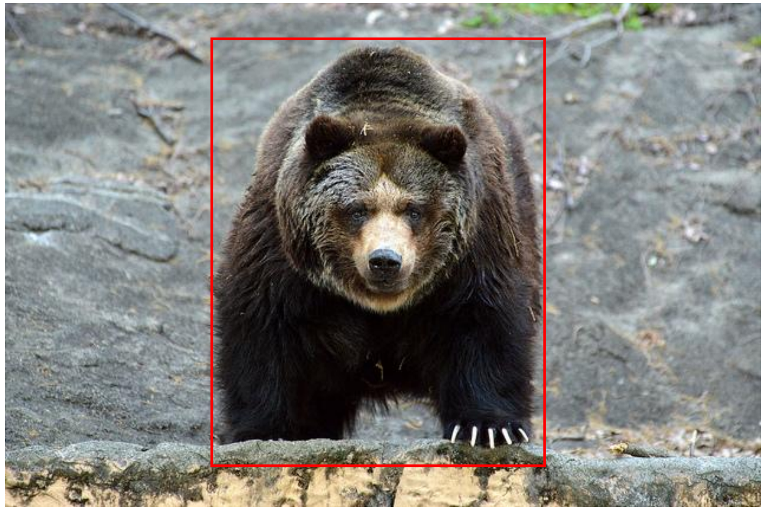


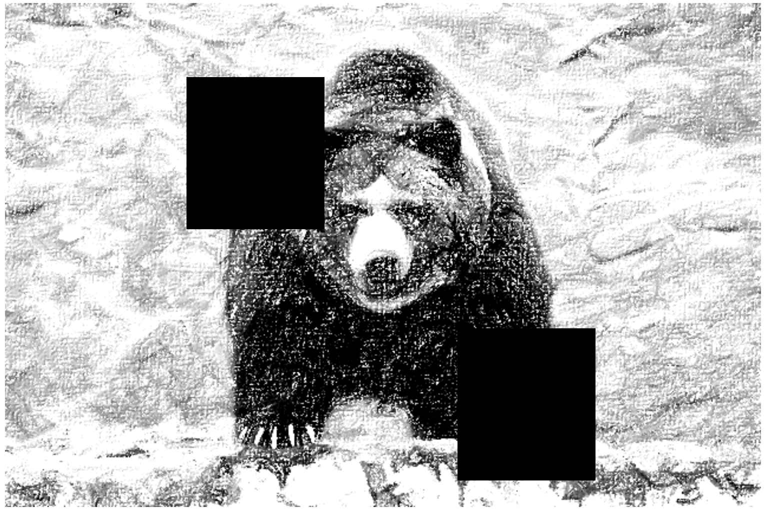
Unlabeled Images

Teacher-Student Mutual Learning

Teacher Model → Generate pseudo-labels on **weakly** augmented Images

Student Model → Learn on **strongly** augmented images and pseudo-labels



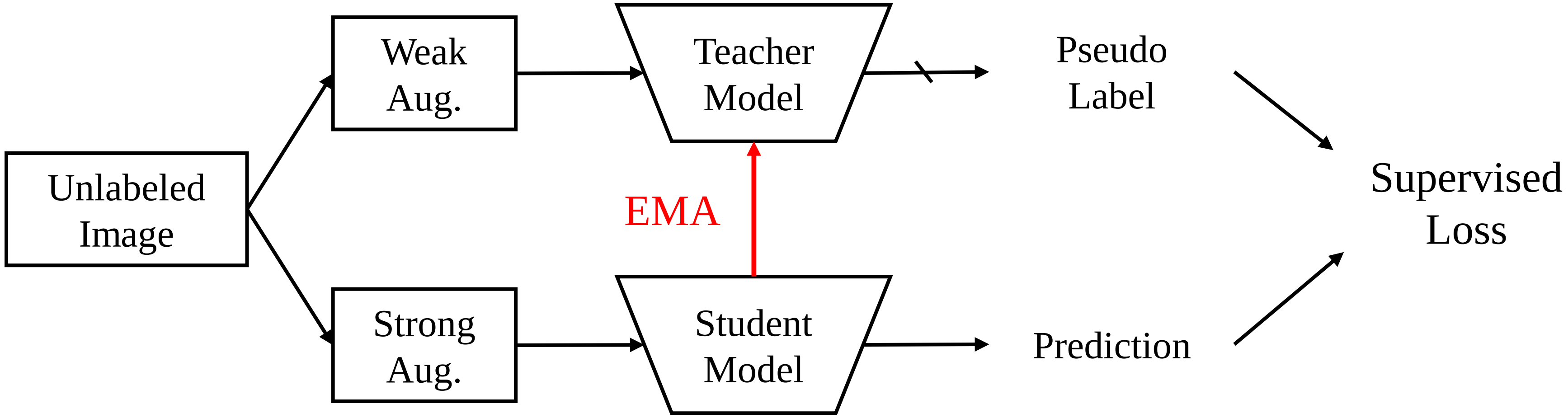


- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the knowledge in a neural network," arXiv:1503.02531v1, 2015.
- Kihyuk Sohn *et al.*, "A simple semi-supervised learning framework for object detection," arXiv:2005.04757v2, 2020.

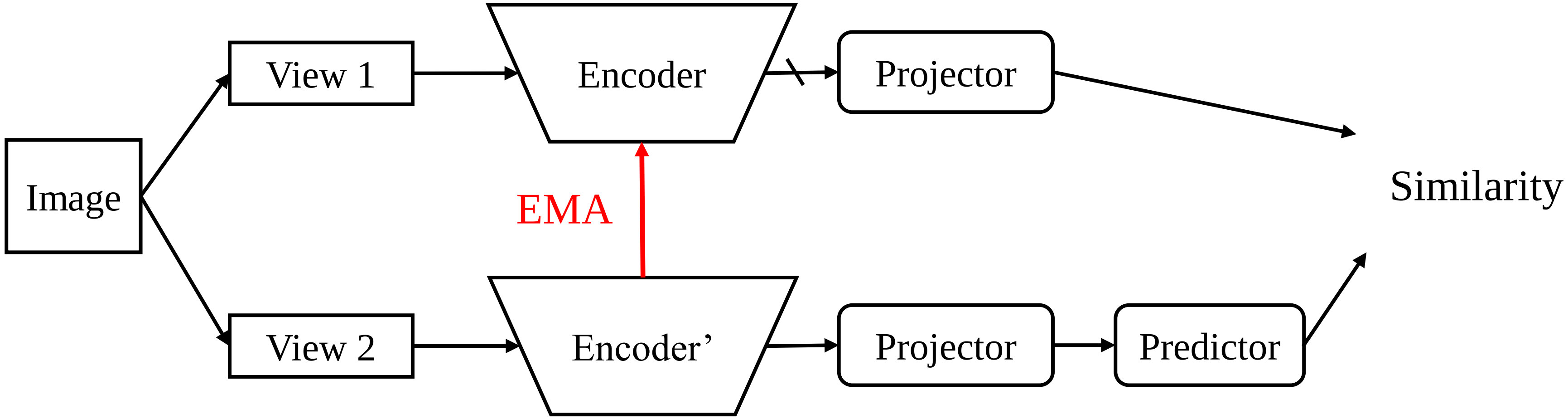
Motivations

- Pseudo-labels generated by the teacher model inevitably introduce label noise. However, deep neural networks have shown the ability to effectively memorize arbitrary noisy labels during training.
- We need a new strategy to utilize unlabeled data.

Semi-Supervised Learning



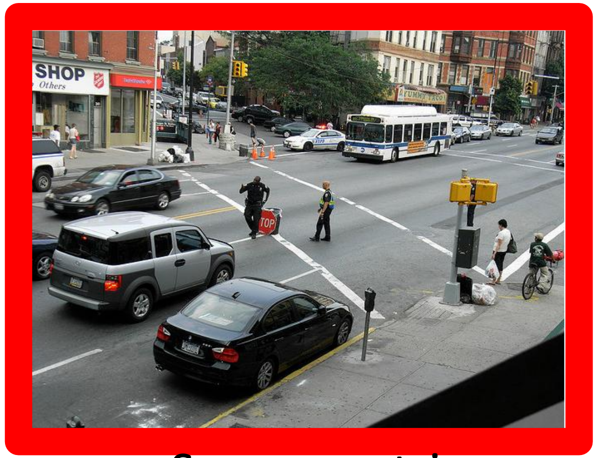
Self-Supervised Learning



Motivations and Challenges

- Self-supervised learning diminishes the reliance on annotation.
- We incorporate this strategy to foster feature learning independent of (pseudo) labels in the context of semi-supervised object detection.
- Images are scene-centric, rather than object-centric.

Object-centric

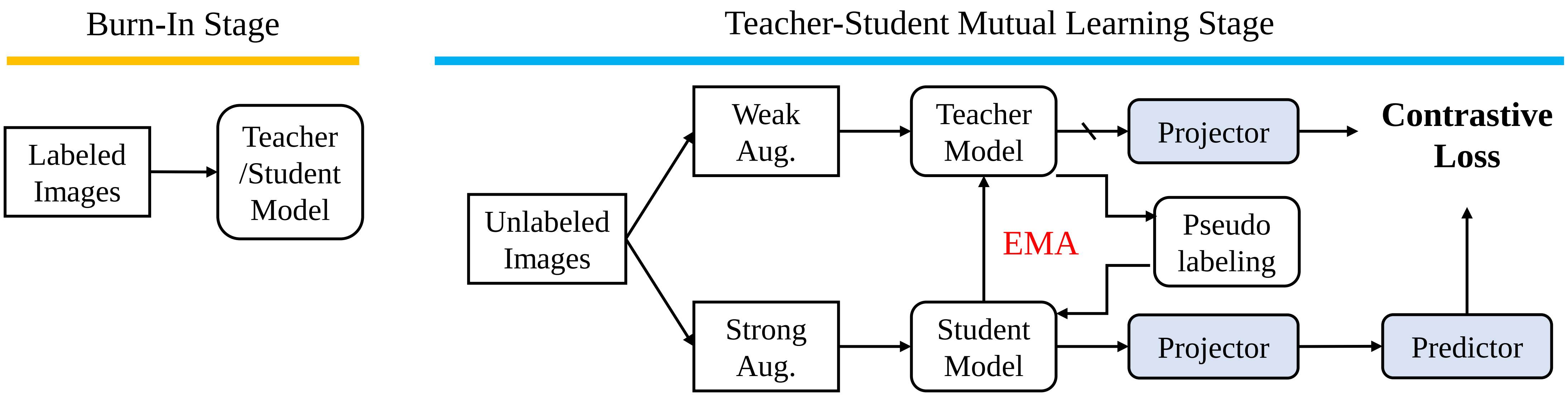


Scene-centric

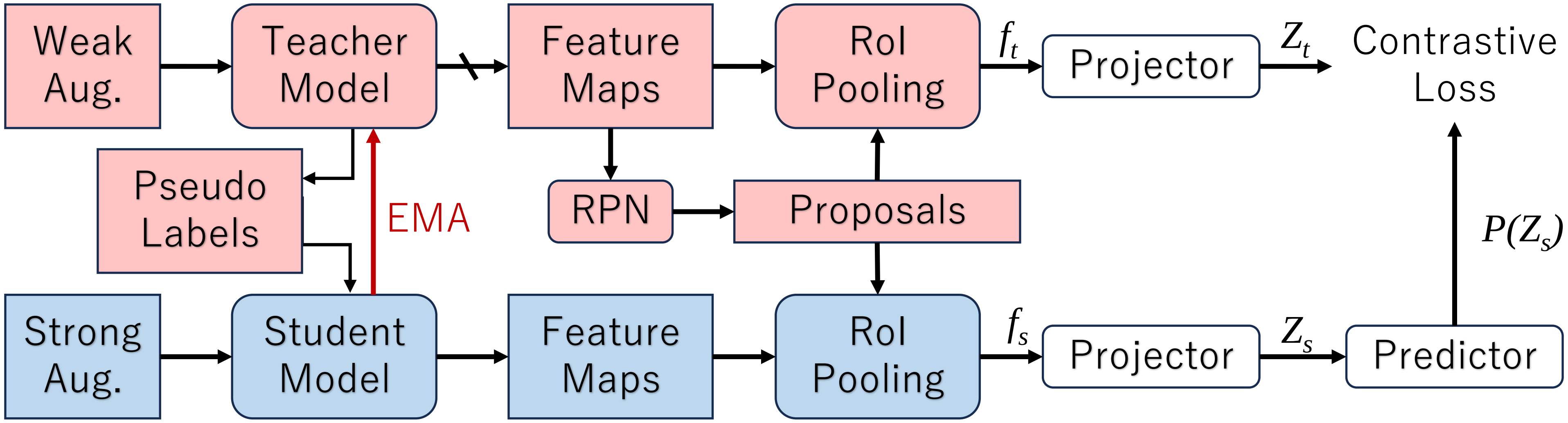
Contributions

- We present a contrastive learning approach to improve semi-supervised object detection, performing consistency regularization not only by aligning the box predictions to pseudo boxes but also by considering feature-level representations.
- To address the challenge of object detection, the contrastive loss is computed at the box level, rather than on the entire image. Furthermore, the loss computation is spatially aware.
- Through experiments, we demonstrate that contrastive learning on RoI features can enhance the model's ability to gain additional information from unlabeled data.

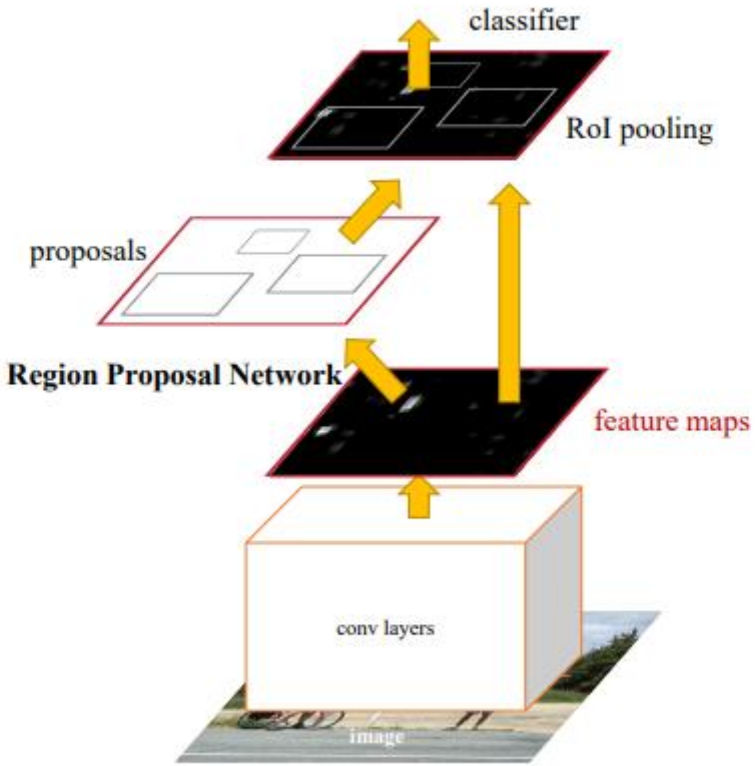
Method

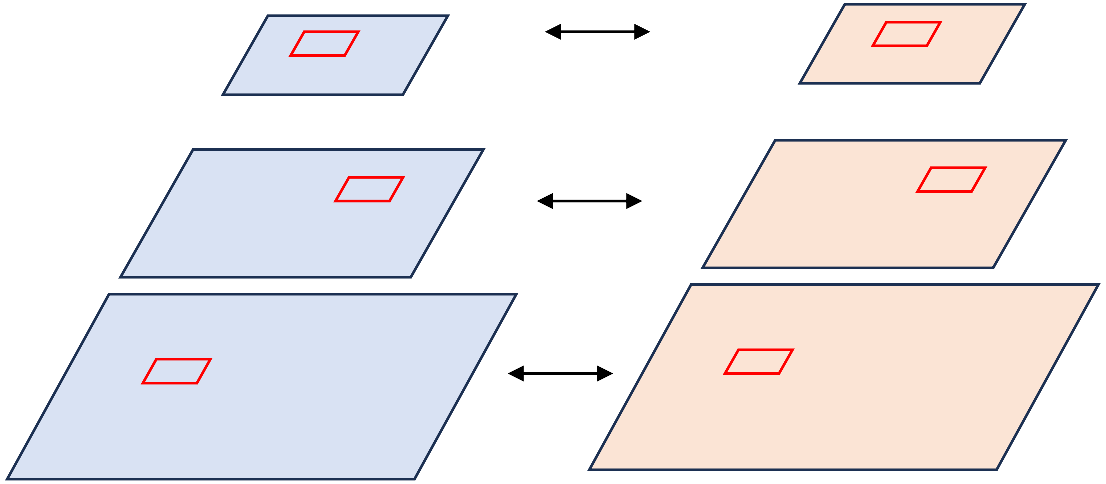

Burn-In Stage
Labeled Images
Teacher /Student Model
Teacher-Student Mutual Learning Stage
Unlabeled Images
Weak Aug.
Strong Aug.
Teacher Model
Student Model
EMA
Projector
Pseudo labeling
Projector
Predictor
Contrastive Loss

RoI-Enhanced Contrastive Learning



RoI-Enhanced Contrastive Learning



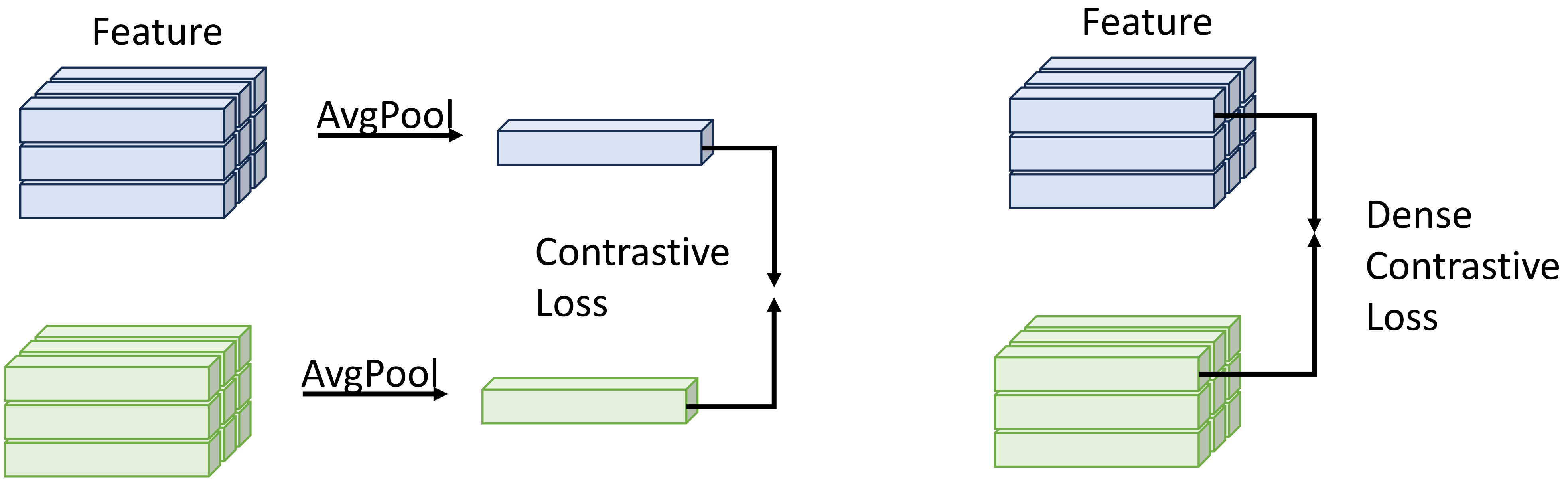


Weakly-augmented feature maps

Strongly-augmented feature maps

Shaoqing Ren *et al.*, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, 2015.

Dense Contrastive Learning



Xinlong Wang *et al.*, "Dense contrastive learning for self-supervised visual pre-training," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

Experimental Settings

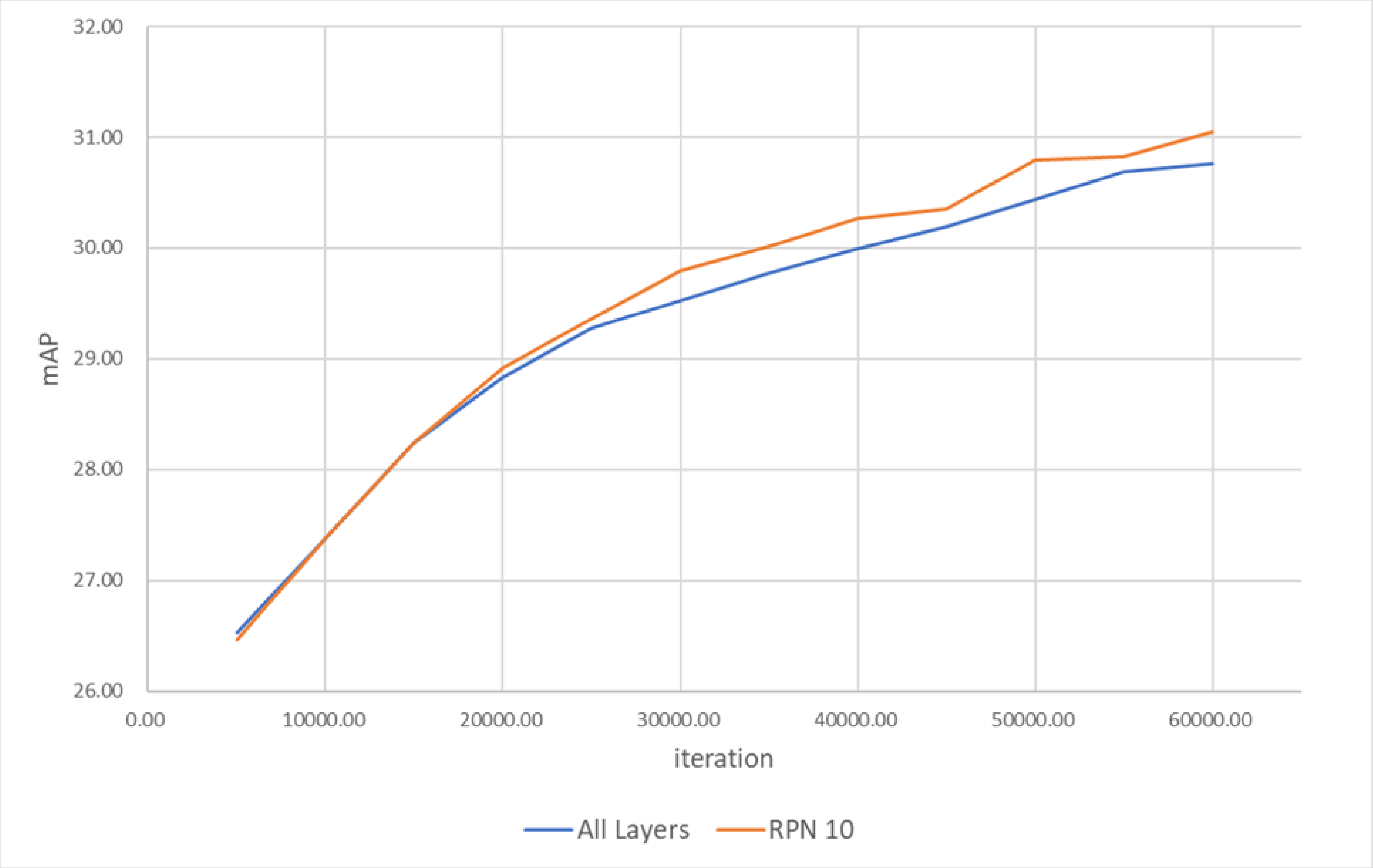
- Dataset: COCO-Standard
    - Training: 118,000 images (train2017)
    - Test: 5,000 images (val2017)
- Evaluation protocol
    - Labeled Images: 1%, 5%, 10% training set
    - Metric: mAP
- Implementation details
    - Backbone: Faster-RCNN with ResNet50-FPN
    - Burn-in stage: 30k iterations
    - Batch size: 16 (smaller than other works)

Experimental Result

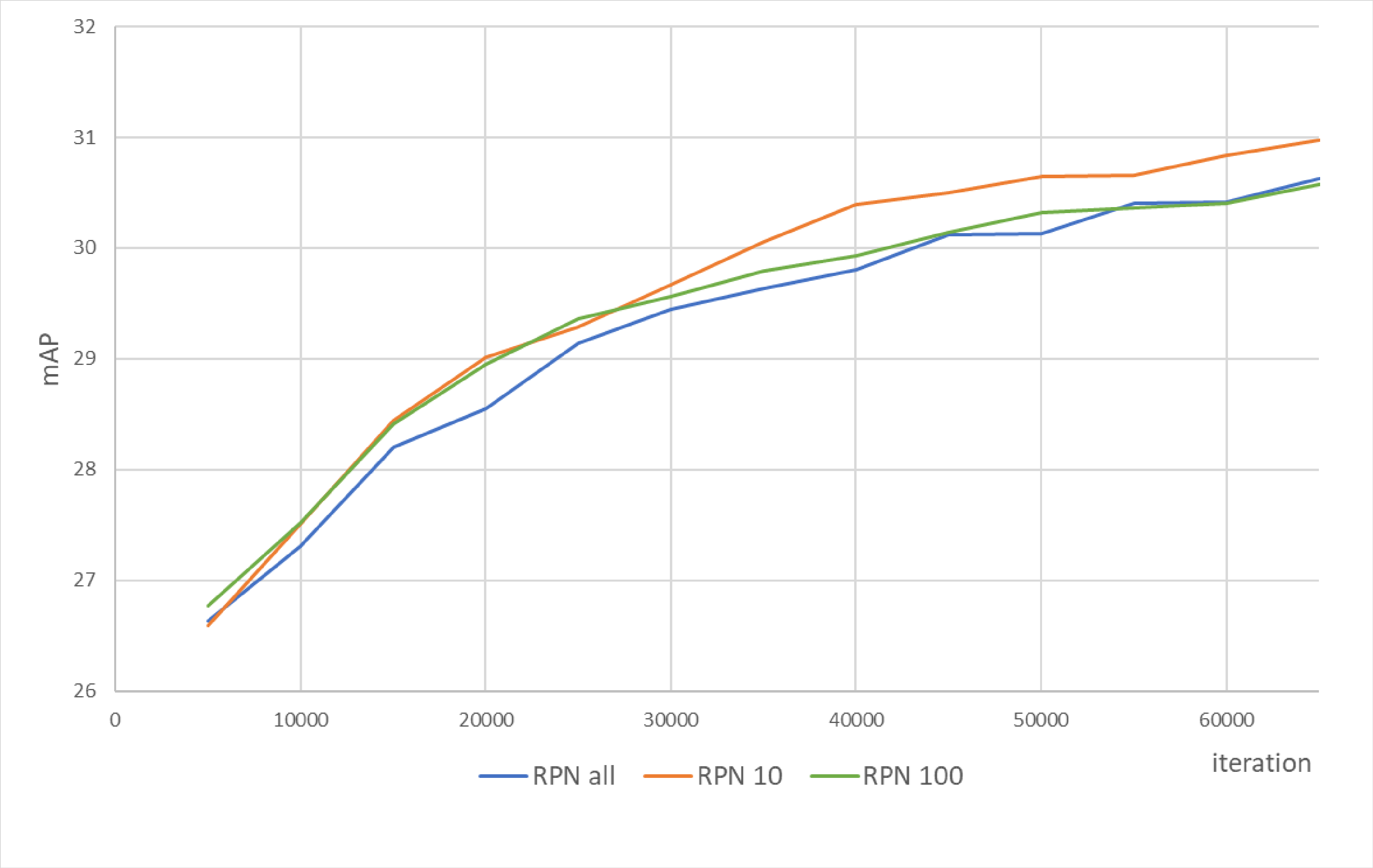
Method	1%	5%	10%
Supervised	9.05	18.47	23.86
Unbiased Teacher	20.19	28.20	31.46
Ours	**20.78**	**28.73**	**31.77**

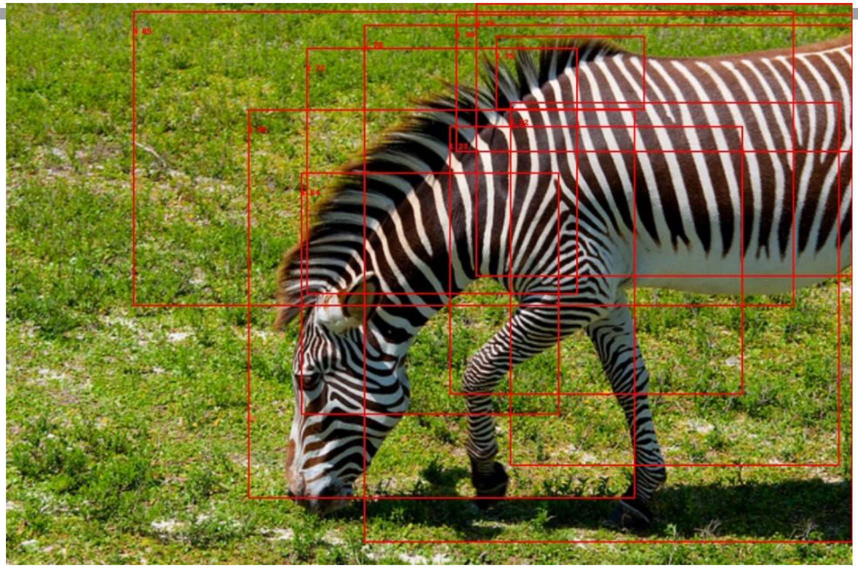
Y. -C. Liu *et al.*, “Unbiased teacher for semi-supervised object detection,” arXiv:2102.09480, 2021.

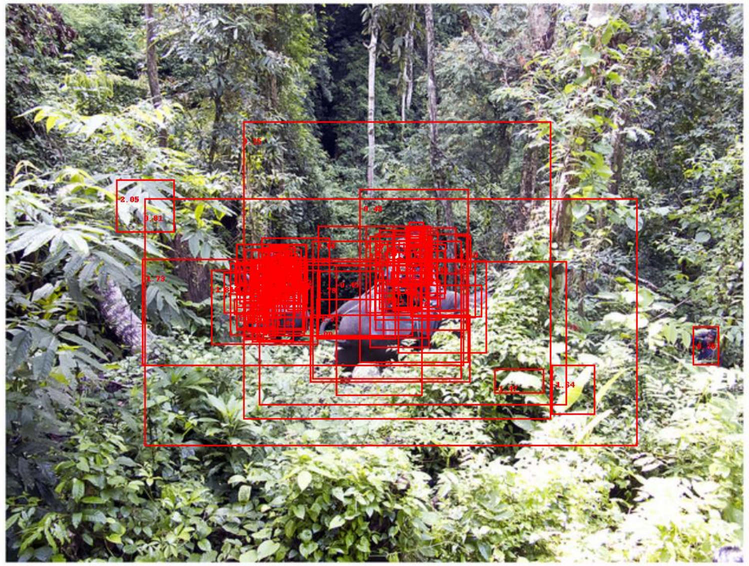
Effect of RoI-Enhanced Contrastive Learning

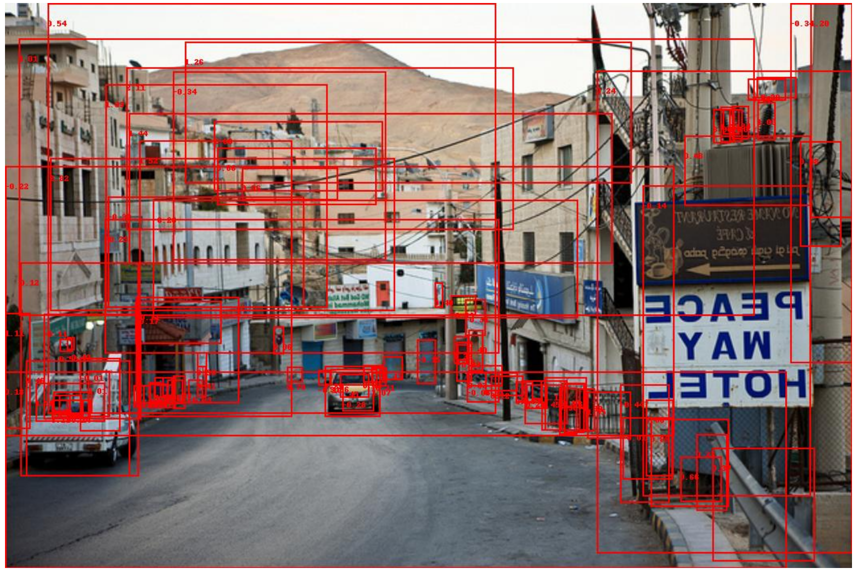


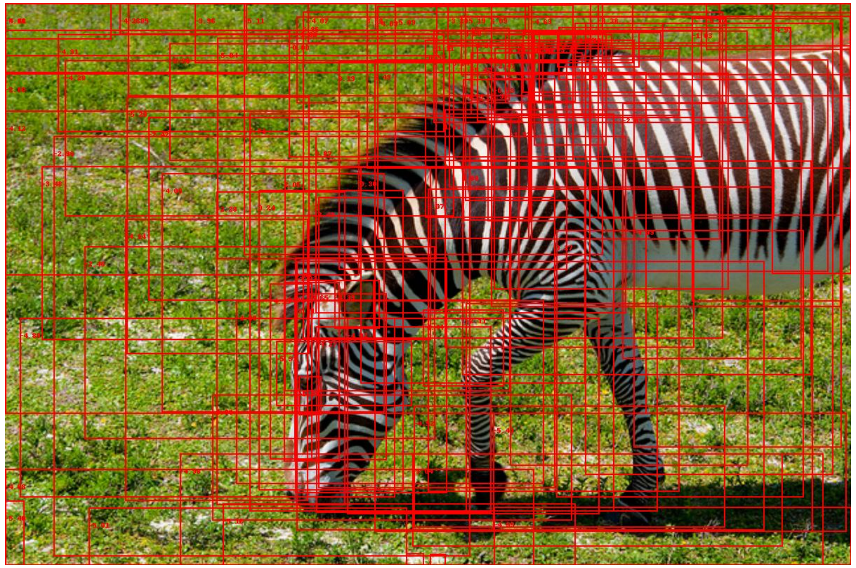
Number of RoI Proposals



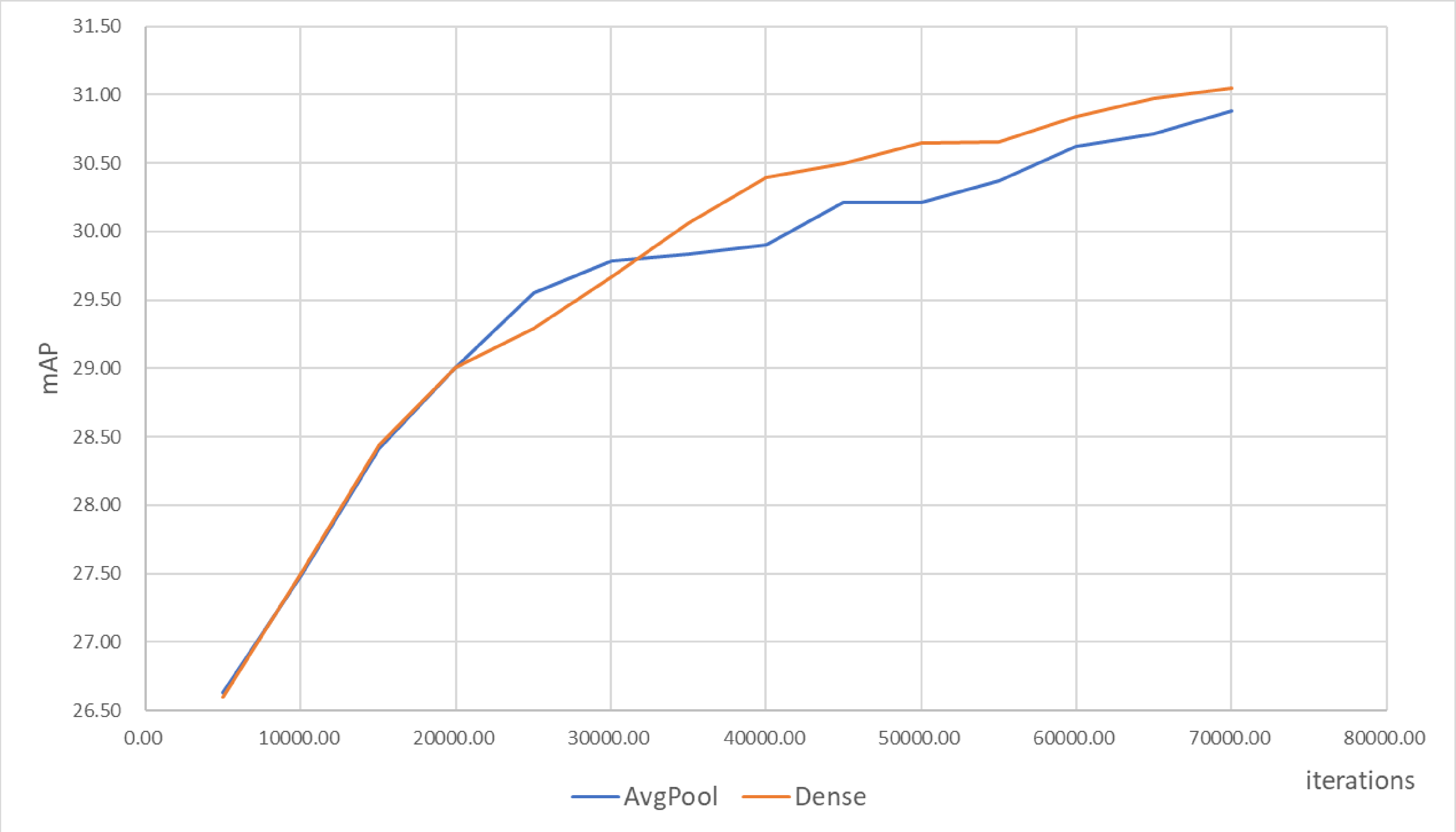








Conventional Loss vs. Dense Loss



Conclusion

- We present a contrastive learning approach to enhance semi-supervised object detection.
    - Leveraging the candidate boxes selected by the Region Proposal Network (RPN) to facilitate RoI-based contrastive learning
    - Incorporating pixel-level comparisons to enable spatial-aware loss calculation
- We will validate the proposed plug-and-play method on alternative detection frameworks beyond Faster-RCNN.

Questions?

More Information:

https://web.ntnu.edu.tw/~myeh/