Effects of adding unlabeled training data through pseudo-labeling

Reducing labeling efforts for deep learned object detectors

Author: Žygimantas Liutkus
(Liutkus@student.tudelft.nl)

Responsible professor: Dr. Jan van Gemert

Supervisor: Dr. Osman Kayhan

1. Introduction

Object detectors require lots of training data for best performance, and this requires a lot of labeling effort.

Semi-Supervised Object Detection (SSL) [1]:

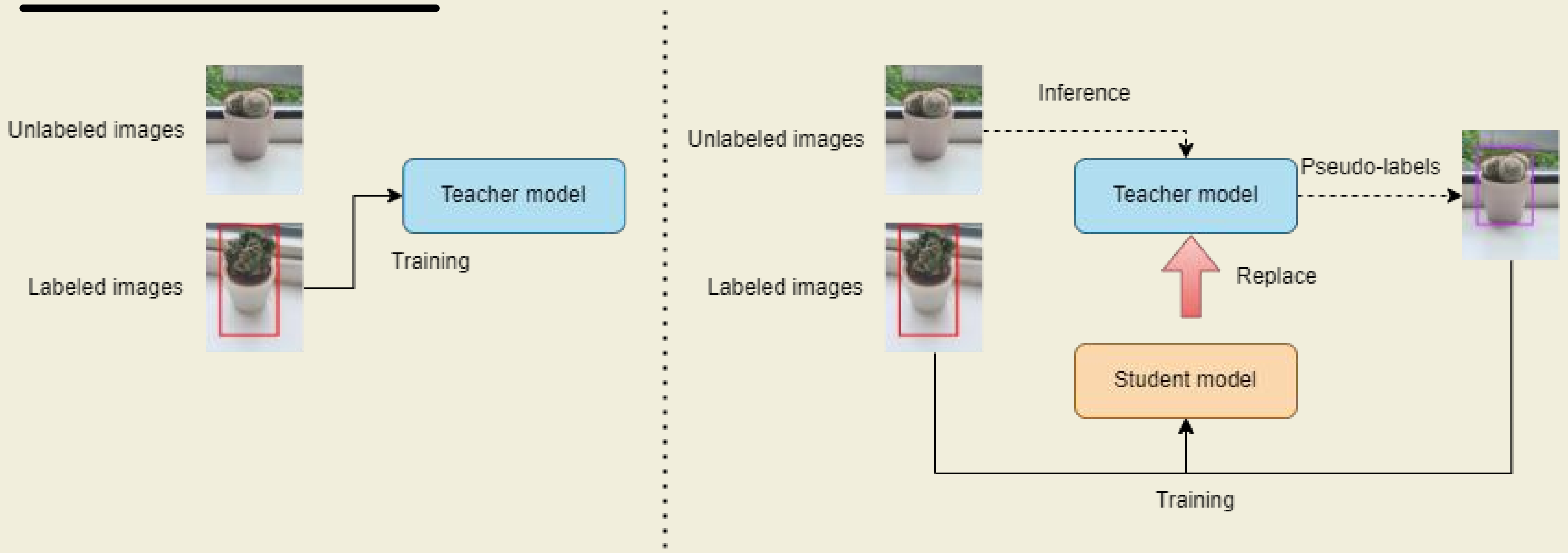
- ***Pseudo-labeling***
- Consistency training

In related research, individual effects of ***pseudo-labeling*** are not investigated [1].

Research goal:

- Exploring the individual effects of pseudo-labeling with the robust YOLOv8 object detector [2].

2.1 Method: Base



Teacher model is trained *supervisedly* on the labeled data. The **teacher** then **produces pseudo-labels** for unlabeled data and the **student** model **is trained on the combined labeled and pseudo-labeled data**. Afterwards, the **student replaces the teacher** and the second step is repeated iteratively.

2.2 Method: Proposed Improvements

Dynamic thresholds [3]:

- **Problem:** Underrepresented classes suffer from biased predictions.
- **Solution:** pseudo-label threshold applied class-wise based on class ratio in training set.

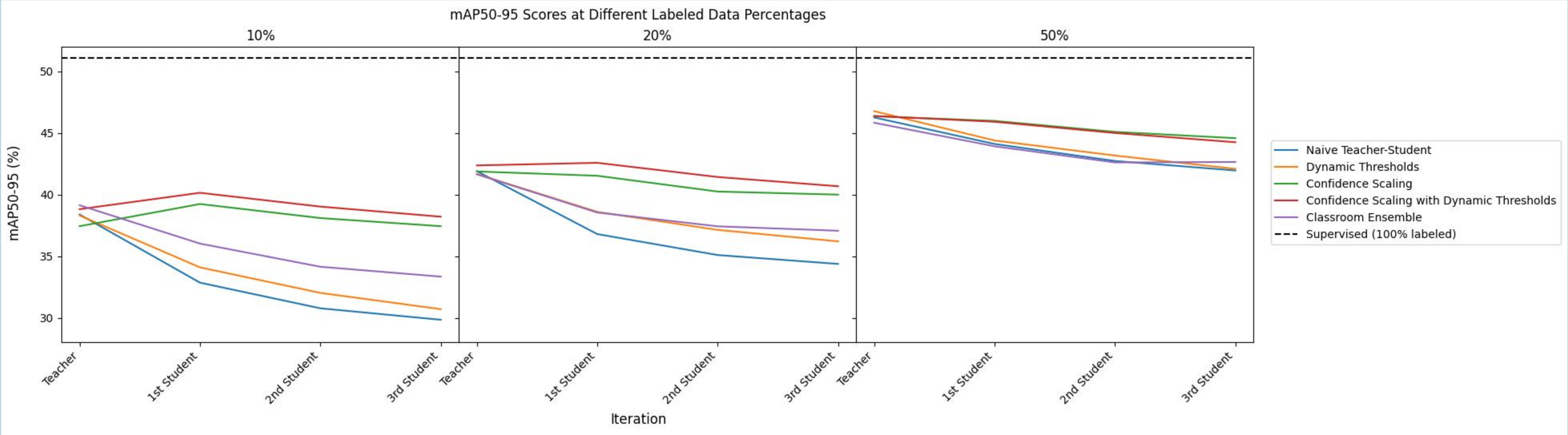
Confidence scaling:

- **Problem:** low confidence predictions can still have valuable information.
- **Solution:** apply scaling according to how distinctively one class is predicted above others.

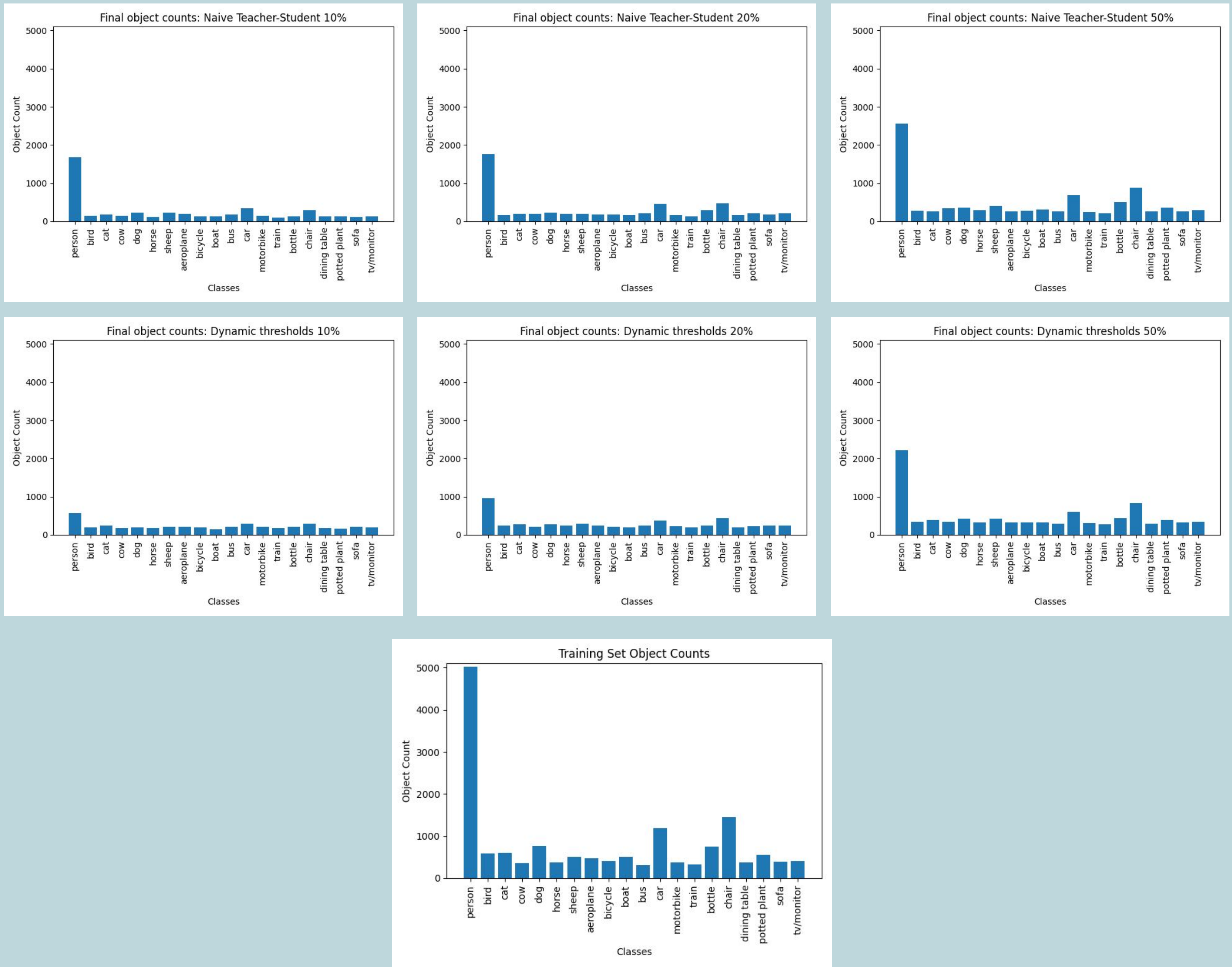
Classroom Ensemble:

- Test if model ensembling benefits transfer to pseudo-labeling.

3. Results

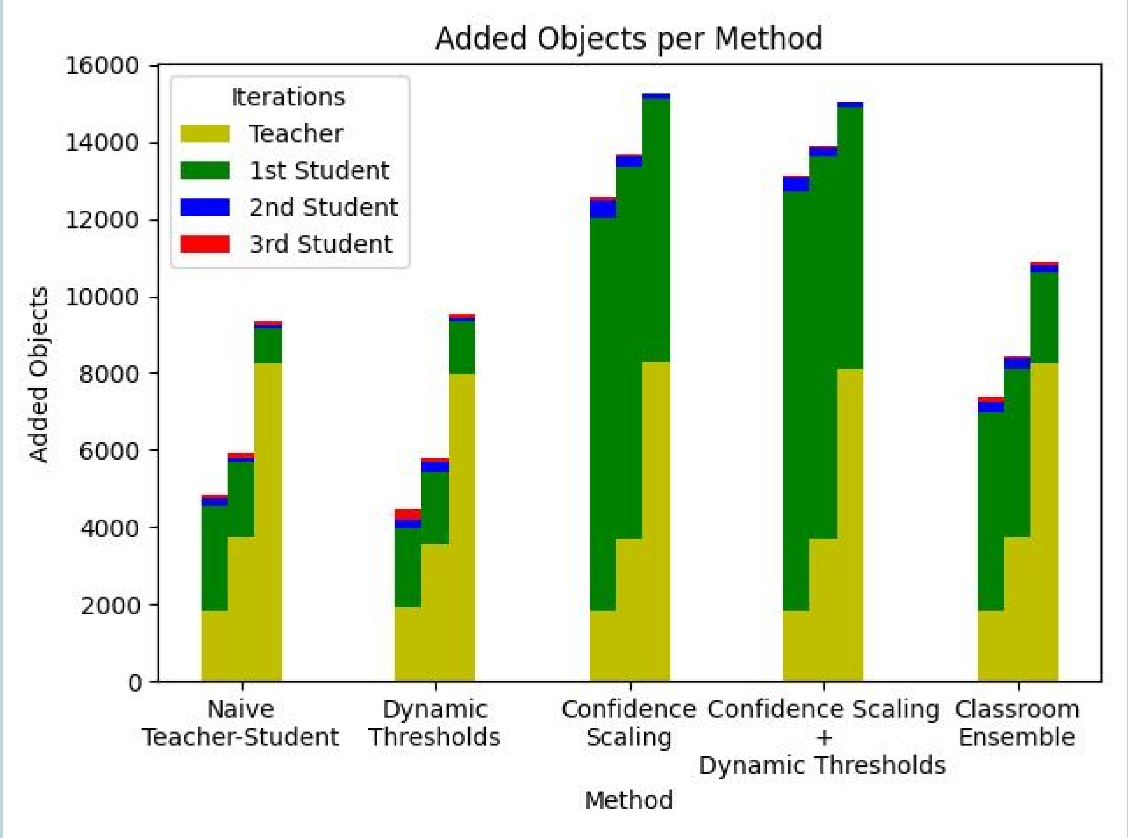


mAP results of all methods at different percentage labeled data splits. A continuous **decreasing trend** is observed in all methods without confidence scaling at lower labeled data percentages. With confidence scaling, the *mAP* decreases after the first iteration.



The class distributions at the end of the naive approach (top) and dynamic threshold (middle) experiments, compared to the whole training set's class distribution (bottom).

Dynamic thresholds produce pseudo-labels more uniformly, without decreasing their total added number.



The 3 bars per method represent 10%, 20% and 50% labeled data splits in that order.

- Confidence scaling significantly increases the number of added pseudo-labels.
- Classroom Ensemble adds more pseudo-labels than the naive and dynamic threshold methods.

4. Limitations

- A likely overfitting issue is present due to overtraining. Epochs could be balanced with amounts of training data.
- Proposed methods do not allow pseudo-label regeneration, leading to missing information.
- Dynamic thresholds are inefficient when there are underrepresented classes in unlabeled data.
- Confidence scaling likely introduces noise to the training set due not considering localization and classification separately.

5. Conclusions and future work

- Experiments showed decreasing performance trends, possibly due to overfitting.
- All proposed improvements achieved individual success relative to the naive basis.
- Results could not be directly compared to related works due to differing datasets and absence of consistency training.

In the future:

- Optimize proposed function hyperparameters.
- Ensemble using different techniques and models.
- Evaluate methods using the same datasets as related works to allow comparisons.

References:

[1] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection, 2020.

[2] Ultralytics YOLOv8 Docs. https://docs.ultralytics.com/. Accessed: 2024-06-24.

[3] Hengduo Li, Zuxuan Wu, Abhinav Shrivastava, and Larry S. Davis. Rethinking pseudo labels for semi-supervised object detection, 2021.