Agregar un "asistente" de aprendizaje: mejorando la calidad de pseudo-etiquetas para la detección de objectos semi-supervisada

Adding a teaching "assistant": improving the quality of pseudo-labels for semi-supervised object detection

Fabian Fallas-Moya[1], Amir Sadovnik[2], Quan Zhou[3], Konstantinos Georgiou[4], Hairong Qi[5]

Fecha de recepción: 30 de mayo, 2024
Fecha de aprobación: 19 de agosto, 2024

Fallas-Moya, F; Sadovnik, A; Zhou, Q; Georgiou, K; Qi, H. Agregar un "asistente" de aprendizaje: mejorando la calidad, de pseudo-etiquetas para la detección de objectos semi-supervisada. *Tecnología en Marcha*. Vol. 38, Nº 1. Enero-Marzo, 2025. Pág. 100-111.

https://doi.org/10.18845/tm.v38i2.7166

1 Universidad de Costa Rica
fabian.fallasmoya@ucr.ac.cr
https://orcid.org/0000-0003-0997-2917
2 Oak Ridge National Laboratory, Estados Unidos
sadovnika@ornl.gov
https://orcid.org/0000-0003-0997-2917
3 The University of Tennessee, Estados Unidos
qzhou10@vols.utk.edu
4 The University of Tennessee, Estados Unidos
kgeorgio@utk.edu
5 The University of Tennessee, Estados Unidos
hqi@utk.edu

Keywords

Machine learning; semi-supervised learning; artificial intelligence.

Abstract

This paper focuses on semi-supervised object detection (SS-OD) for its tolerance to small amounts of training samples, which is common in real-world applications. Pseudo-label-based approaches have been the mainstream for SS-OD. In this paper, we first show the impact of accurate pseudo-labeling and the challenge of producing such labels. In contrast to prior research that predominantly focused on refining the main model to enhance localization, this paper introduces a novel strategy, where a standalone "Teaching Assistant" or simply "Assistant" is involved in the popular Teacher/Student paradigm to improve the quality of pseudo-labels. This "Assistant" can be plugged into any existing Teacher/Student-based framework without having to fine-tune the original Teacher/Student model. We exploit two "Assistant" models, both of which center around the non-maximum suppression (NMS) method -- a popular technique used to select only the promising bounding boxes. The first "Assistant" model is referred to as the "pre-NMS" assistant that refines the candidate bounding box scores for a better set of inputs to the NMS process. The second "Assistant" model is referred to as the "post-NMS" assistant which takes advantage of SOTA segmentation models to improve the output from the NMS process. We thoroughly evaluate the performance of pre-NMS vs. post-NMS and the impact of improved pseudo-labels on the OD performance. Experimental results on the COCO dataset demonstrate that post-NMS is better than SOTA methods.

Palabras clave

Aprendizaje máquina; aprendizaje semi supervisado; inteligencia artificial.

Resumen

Este artículo se centra en la detección de objetos semisupervisada (SS-OD) por su tolerancia a pequeñas cantidades de muestras de entrenamiento, lo cual es común en aplicaciones del mundo real. Los enfoques basados en pseudoetiquetas han sido la corriente principal para SS-OD. En este artículo, primero mostramos el impacto del pseudoetiquetado y el desafío de producir dichas etiquetas. En contraste con investigaciones anteriores que se centraron en perfeccionar el modelo principal para mejorar la localización, este artículo presenta una estrategia novedosa, en la que un "Asistente de Profesor" independiente o simplemente un "Asistente" participa en el popular paradigma Profesor/Estudiante para mejorar la calidad de la enseñanza. pseudoetiquetas. Este "Asistente" se puede conectar a cualquier marco existente basado en Profesor/Estudiante sin tener que ajustar el modelo original de Profesor/Estudiante. Explotamos dos modelos "Asistentes", los cuales se centran en el método de supresión no máxima (NMS), una técnica popular utilizada para seleccionar solo los cuadros delimitadores prometedores. El primer modelo "Asistente" se denomina asistente "pre-NMS" y refina las puntuaciones del cuadro delimitador candidato para obtener un mejor conjunto de entradas al proceso NMS. El segundo modelo "Asistente" se conoce como asistente "post-NMS" y aprovecha los modelos de segmentación SOTA para mejorar el resultado del proceso NMS. Evaluamos minuciosamente el rendimiento de pre-NMS frente a post-NMS y el impacto de las pseudoetiquetas mejoradas en el rendimiento de detección de objetos. Los resultados experimentales en el conjunto de datos COCO demuestran que el asistente post-NMS es mejor en comparación con los métodos SOTA.

Introduction

Object detection (OD) has been one of the fundamental problems that a robot has to solve before it can set out to explore the world. SOTA OD models are all deep learning-based. Although deep learning has made a revolution in performance in recent years [1], this advancement comes at a notable cost: the hunger for substantial quantities of labeled data for training. The process of obtaining these labeled data is often characterized by its expense in terms of time and availability [2].

In response, the paradigm of semi-supervised (SS) learning has emerged, advocating for the utilization of a small quantity of labeled data in combination with a large number of unlabeled samples. Recent outcomes in this direction have displayed remarkable results [3] [4] [5] in classification tasks. Pseudo-labeling [6] [7] is the widely used SS method, which basically uses the main model to make predictions over the unlabeled data, and these new predictions or pseudo-labels are then used in subsequent training.

The successes of SS for classification have motivated the development of SS models to solve object detection (OD), referred to as SS-OD. However, the nature of OD brings new challenges to SS. This is mainly due to the heavy reliance of SS on the accuracy of pseudo-labels [6]. SS assumes that the pseudo-labels are close in the feature space (smoothness assumption [8]). In OD, it has to predict both the bounding boxes' location and their classes.

One of the most popularly used SS learning models is the teacher/student (TS) model (similar to distillation learning [9] [10] [11]), where the teacher generates pseudo-labels for the student during training. It is interesting to note that in real-life scenarios, in addition to the teacher and students, there is often another important character for each class, the "teaching assistant", who helps explain the teacher's lectures and assignments in a better way such that the students could digest the content better, yielding better learning outcome. Inspired by this analogy, we propose the addition of an "Assistant" model to help improve the quality of the pseudo-labels produced by the teacher model (see Fig. 1).

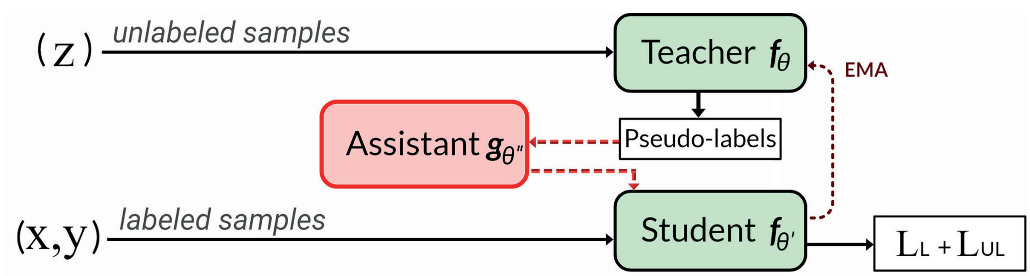

Figura 1. We process the pseudo-labels using an assistant to improve their quality.

Usually, OD generates multiple predictions (or bounding boxes) over the same object, and the non-maximum suppression (NMS) method is often applied to keep only a single box from a set of overlapping predictions (Fig. 2a). However, NMS is based on classification scores, which become unreliable with limited labeled data. To address this issue, we use the assistant model to enhance the classification scores, allowing for more informed decisions. We call this assistant "pre-NMS". It refines candidate scores that represent a higher intersection over union (IoU) with the object (see Fig. 2b).

Our second assistant, "post-NMS", works by correcting the pseudo-labels after NMS. We take advantage of recent advances in image segmentation and use the Segment Anything Model (SAM) [12] to perform this correction (see Fig. 2c). SAM is a visual foundation model (VFM)

[13] that attempts to segment the visual content from an image using visual prompts (e.g., dense points, single points, or boxes) without training. This zero-shot inference is possible because SAM was trained using 11M images that contain nearly 1B masks. There are two main challenges when using SAM: first, it needs visual prompts from the user; second, how to use the segments for the task of bounding box correction. To respond to these challenges, in the “post-NMS” assistant model, we first use a novel approach to create a visual prompt with the teacher’s pseudo-label coordinates. This will create masks from the area inside the pseudo-label and we choose the mask that has the largest area as the prompt. The pseudo-label is then adjusted with the most likely object that is contained.

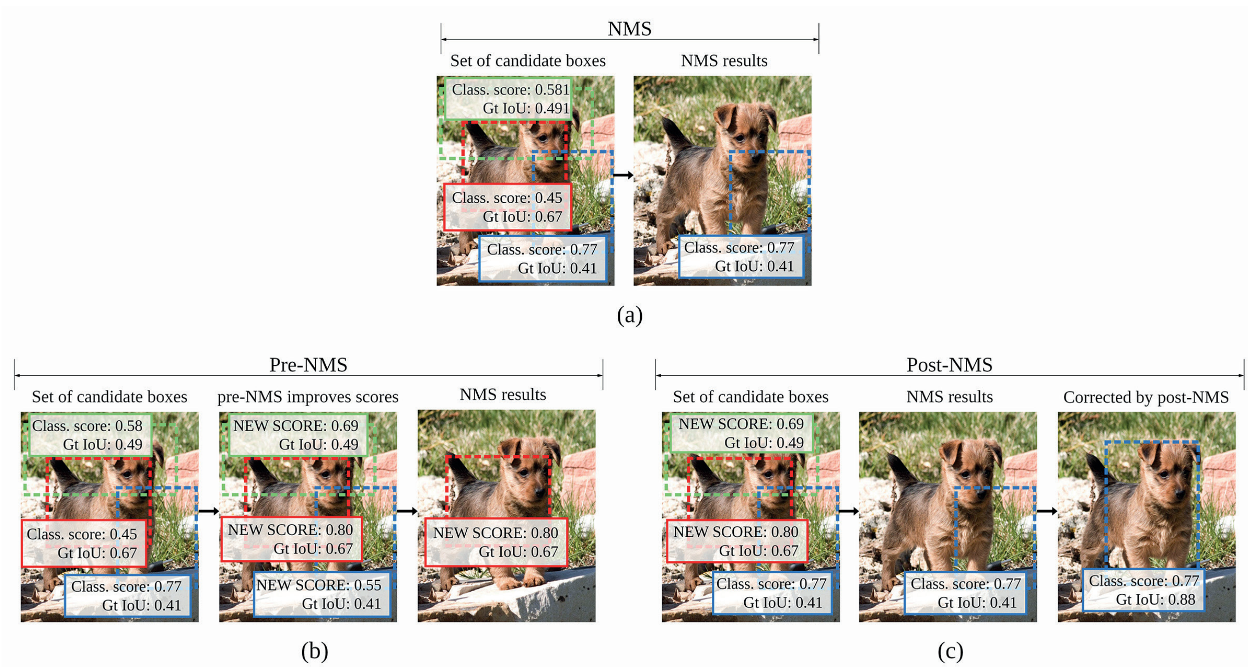

Figura 2. How our two assistants improve the quality of the pseudo-labels around NMS.

Our primary contributions can be summarized as follows.

- Inspired by the important roles that teaching assistants play in a classroom; we expand the TS-SS learning by adding an “assistant” to improve performance.
- We exploit two assistants centered around the NMS that yield improved results.
- We propose to use SAM in the SS-OD pipeline to take advantage of SOTA segmentation model for the purpose of bounding box correction.
- Finally, we thoroughly evaluate the two different assistant models for the purpose of SS-OD and conclude with a superior performance of the post-NMS assistant.

Literature review

The majority of recent semi-supervised classification research adopts the teacher/student framework [9] [10]. SS learning approaches can be further divided into three categories, consistency regularization-based, pseudo-label-based, and hybrid. Consistency regularization [14] [15] assumes that random transformations should not change the predictions given the same input. Pseudo-labeling trains a preliminary model using limited labeled data and then uses this model to assign labels to unlabeled samples. These pseudo-labels are then included in the next-round training. From its first appearance as entropy minimization approach [16] and in deep learning [6], pseudo-labeling has achieved promising performance, including the Meta Pseudo-labels [7], and other hybrid models [4].

Interesting attempts were proposed by Hoffman et al. [17] and Gao et al. [18], where they train using few or no annotated samples. Jeong et al. [19] used consistency regularization to overcome the imbalance of foreground and background. Sohn et al. [20] created a simple framework using pseudo-labeling. Recently, the trend has combined the teacher/student models with exponential moving average (EMA) [10] [21].

Large Language Models (LLMs) have emerged for NLP tasks. Foundation models, pre-trained LLMs on extensive data, enable direct use in downstream tasks without training. SAM [24], a VFM, seamlessly integrates into visual tasks including medical imaging [22], one-shot OD [23], zero-shot segmentation [24], and more [25].

Method

In SS-OD, modern approaches usually build on FixMatch's core concepts [3]: enforcing model consistency through EMA updates between teacher and student models and generating pseudo-labels from the teacher's predictions. Our baseline model, Unbiased Teacher [10], is a recent variant of FixMatch. In the following, before elaborating on the proposed "Assistants", we discuss the importance of having high-quality pseudo-labels.

The pre-NMS Assistant

The pre-NMS assistant model aims to measure pseudo-label quality by constructing a model that, when given an image with predicted bounding boxes from the teacher model, predicts a new objectness score. This score signifies the degree to which the object is present.

IoU-Augment. To train the pre-NMS assistant, we generate image crops with varied IoU based on the ground truth given using labeled data (Fig. 3a). From each image's bounding boxes, we select random points as centroids. New samples, with dimensions scaled proportionally (between 20% and 100%) of the original box, are thus created. This process produces multiple random boxes from a single box, forming new data samples along with their corresponding labels denoting the intersection percentage. We refer to this step as IoU augmentation.

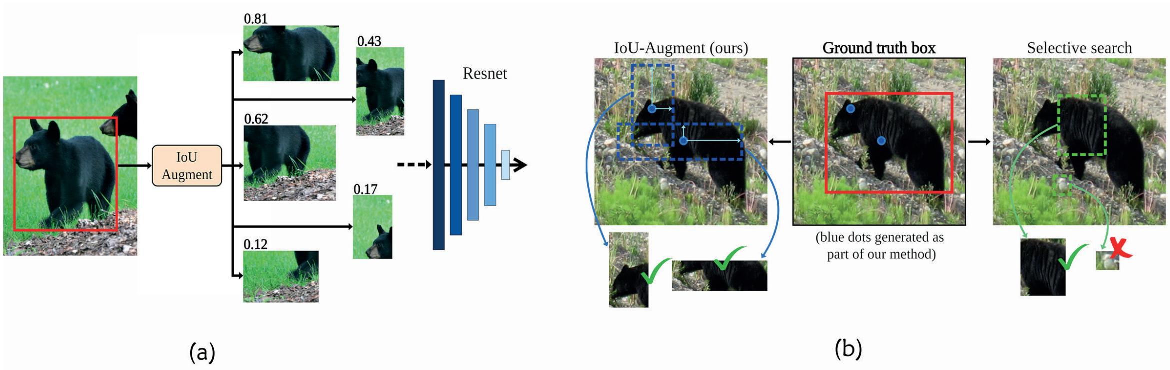

Figura 3. (a) Pre-NMS assistant training process. (b) Our method vs selective search.

Comparison with selective search. An alternative method for generating random samples is selective search [26]. However, it comes with two key limitations. Firstly, selective search can yield samples with irrelevant image sections, as depicted in Fig. 3b-right where a plant is chosen instead of the intended bear. In contrast, our IoU-Augment method guarantees the object's presence within a specific bounding box percentage, ensuring relevance. Secondly, selective search lacks control over sample and label distribution, potentially leading to dataset imbalance (e.g., creating more samples with an IoU < 50%). In comparison, our method manages the number of samples in different IoU ranges, promoting diverse data.

Implementation details. Resnet-50 is used as the backbone for pre-NMS. The training set is first augmented using the IoU-Augment method, with the mean absolute error (MAE) as the loss function. To address overfitting that arises from generating numerous samples from the same image distribution, we use strong data augmentation strategies. The model convergence is fast, typically within 10,000 iterations.

The post-NMS Assistant

The post-NMS assistant model incorporates SAM to distill information about the object's location (see Fig. 4) using the NMS prediction result directly. SAM is intended for segmentation tasks rather than OD and does not undergo training in this context. Initial attempts [22] [25] [27] reveal that while SAM is powerful, it requires fine-tuning or tailored visual prompting for effective domain adaptation. Fine-tuning is relevant mainly for considerably different domains like medical [22], multi-spectral [28], or 3D objects [27].

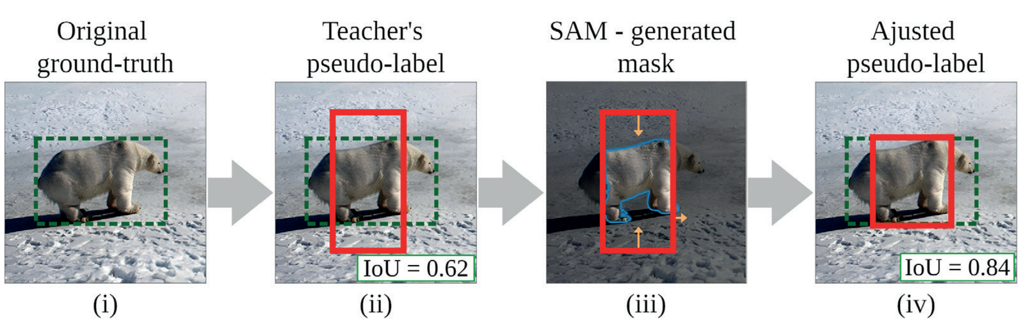

Figura 4. Post-NMS assistant uses the pseudo-labels as visual prompts to feed SAM.

SAM offers two options for visual prompting: user-defined prompts, where users mark points or boxes on images (similar to active learning [1]), and dense sampling through point grids. However, dense sampling is not well suited for SS-OD. Rather than relying on user-generated prompts, we utilize the pseudo-label created by the teacher model as the visual prompt that will be used for SAM. This generates segmentation just inside the pseudo-label. From all the resulting segments we keep the one with the largest area and recover the coordinates of the box that the segment generates. This would be the correction applied to the pseudo-labels. See Fig. 4 for an illustration of this process.

Our pipeline

Fig. 5 shows the main components of our pipeline and how these components interact.

- Burn-in stage: for pre-NMS, during the burn-in stage, we train with labeled data to obtain pre-trained weights for the semi-supervised framework [10] [20]. These weights initialize both the student and the teacher models.
- Mutual learning: our assistants cooperate with the teacher model during pseudo-label generation. The teacher, a Faster R-CNN model, employs the ROI head for which NMS relies on potentially inaccurate classification scores, particularly in initial iterations. Mutual learning integrates our assistants with the teacher and student. It is worth revisiting that pre-NMS enhances NMS decisions through new (more accurate) scores, while post-NMS refines pseudo-labels.

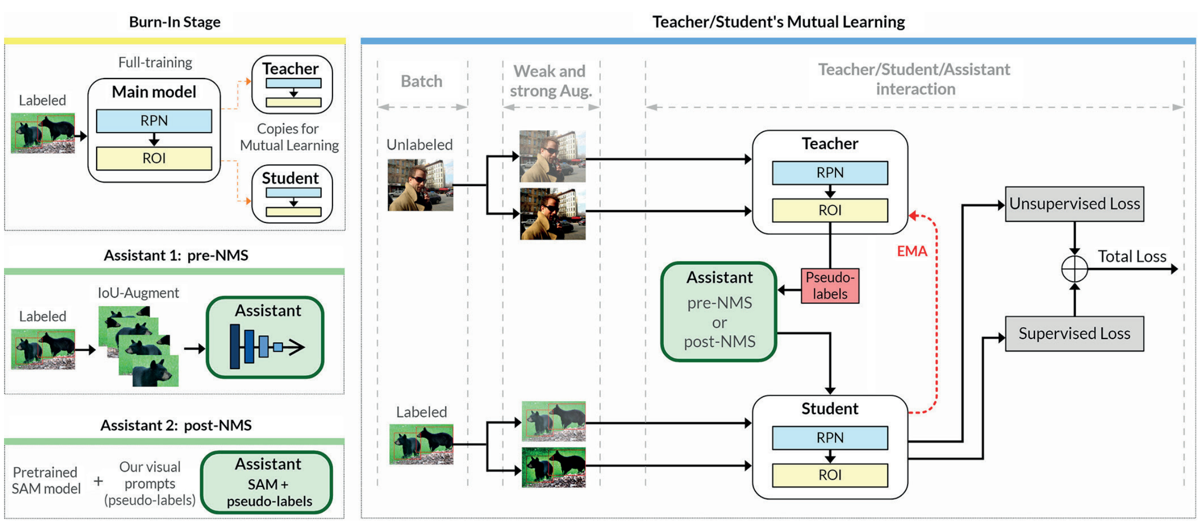

Figura 5. Our main pipeline showcasing how our assistants interact with the main model.

Fig. 5 also shows that the student model is updated using a compound loss for the labeled and unlabeled data, given by $D_L = \{x_i^l, y_i^l\}_{i=1}^{N_l}$ and $\mathrm{D_{UL}} = \{\mathrm{x_i^u}\}_{\mathrm{i=1}}^{\mathrm{N_u}}$, respectively. Since we are using Faster R-CNN [18], the labeled data have four losses:

$$\mathcal{L}_L = \sum_i \mathcal{L}_{cls}^{rpn}(x_i^l, y_i^l) + \mathcal{L}_{reg}^{rpn}(x_i^l, y_i^l)\backslash + \mathcal{L}_{cls}^{roi}(x_i^l, y_i^l) + \mathcal{L}_{reg}^{roi}(x_i^l, y_i^l)$$

where *rpn* stands for the region proposal network (RPN) model, *roi* for the region of interest (ROI), *cls* for the classification, and *reg* for the regression boxes. Also, we have the unlabeled loss defined by:

$$\mathcal{L}_{UL} = \sum_{i} \mathcal{L}_{cls}^{rpn}(x_i^u, \widehat{y_i^u}) + 0 \times \mathcal{L}_{reg}^{rpn}(x_i^u, \widehat{y_i^u})$$

where $\hat{y}$ stands for the pseudo-label. Notice that we do not use the ROI loss, disabled by being multiplied by zero (following [10]), because we know the location of the pseudo-labels have inaccuracies, and even though these inaccuracies are lowered by the assistants, the corrections made by them are not integrated into the graph of the derivations and cannot contribute to the loss function. In the end, we have a total loss:

$$L_{total} = L_L + L_{UL}$$

Finally, the teacher model is updated (in every iteration) using EMA [15] with $\theta_{teacher} = \alpha \times \theta_{teacher} + (1 - \alpha) \times \theta_{student}$, where α is the smoothing factor that we set to **α= 0.9996** following [10] which has shown the best performance.

Experiments

Datasets. We evaluate our framework using the COCO dataset under various conditions: (1) Single class: focusing on one COCO class, (2) 10-classes: using a subset of 10 classes, and (3) COCO-full: using the entire COCO dataset.

Implementation Details. We use Unbiased Teacher [10] as our baseline. The pre-NMS assistant is trained using MAE and data augmentation with strong augmentation policies (flip, color, and slight rotation).

Metrics. We use both mean average precision (mAP) and mean-IoU (mIoU) measured with ground truth for pseudo-label quality.

The pre-NMS Assistant: Training

Table I illustrates the required labeled samples for effective regressor training. “No overlap” denotes training the assistant with images containing objects absent in the evaluation set. “Overlap” signifies that the train and evaluation sets may share class objects. The column “Labeled samples” shows the number of training samples.

During experiments with various sizes of labeled data sets, we observed that augmenting the size of the labeled data set does not consistently improve performance. Table I demonstrates our attempt to enhance performance by enlarging the labeled dataset; however, we found that the best performance is achieved using only the samples labeled with 240. This outcome arises because our IoU-Augment generates numerous new training samples per image. Despite the small 240 sample count, the effective training sample size is significantly larger, for instance, with an average of 8 objects per COCO image [8], and n=10 augmentations per bounding box, this results in 240 ×10 ×8 = 19200 samples. Moreover, the objective function converges fast, typically around 10K iterations, due to its straightforward single-regression nature, and the focus on objectness within boxes.

Table I. Comparing different numbers (from 120 to 4800) of labeled samples during training.

		Loss (MAE)	
Train/Eval relation	Labeled samples	train	eval
No-overlap	120	0.047	0.105
	240	0.076	0.094
	480	0.102	0.101
	960	0.086	0.102
	1200	0.081	0.103
	2400	0.079	0.105
	4800	0.105	0.101
No-overlap with finetuning	240	0.023	0.095
Overlap	240	0.032	0.102

“No-overlap” outperforms the other settings, suggesting training efficacy with objects distinct from the evaluation set. “No-overlap with fine-tuning” initially trains on nonoverlapping classes and fine-tunes with overlapping class data, yet the gain is marginal. Crucially, this table underscores the assistant’s ability to predict pseudo-label IoU accurately, irrespective of the class of the object, which is an important insight.

Table II. Different post-NMS configurations.

Method	SAM model	Thresh	mAP
Regular	-	0.7	41.01
Aug. method	-	0.7	40.67
Super-pseudo	-	0.7	30.71
Super-pseudo + aug. method	-	0.7	30.83
SAM - NMS	h	0.7	34.75
SAM post-NMS	b	0.7	34.61
	l	0.7	42.13
	h	0.7	55.25
	h	0.5	47.33
	h	0.6	53.62
SAM alone	h	-	31.05

The post-NMS Assistant: How to Use SAM

“Post-NMS” required thorough experimentation to adapt SAM. Table II presents our varied experiments for post-NMS configurations. Identifying the optimal setup posed challenges. Initial data augmentation trials (first 4 rows) without SAM showed negligible impact. Experimenting with the model size of SAM (base, large, huge) highlighted “huge” as the most effective. The column “Thresh” specifies the pseudo-label score threshold, that is, the minimum score a pseudo-label needs to have to be considered valid; 0.7 proved to be the most successful. Notably, SAM alone (last row) cannot accomplish OD autonomously.

The Assistants: pre-NMS vs. post-NMS

Table III presents results for our SS-OD methods, labeled proportions, and datasets. Pre-NMS is trained with labeled data. The table displays the mAP (50:95) achieved during training and mIoU at 50% overlap, indicating mIoU between pseudo-labels and ground truth when pseudo-labels overlap 50% or more. We contrast our approach with the Unbiased Teacher (baseline) and standard supervised training (no semi-supervised step) solely using labeled proportions.

The key finding of the table is that our assistants work better for small datasets. Single-class shows the best mAP results with our assistants. With 10 classes, the improvement is significant with post-NMS, but the contribution of pre-NMS is small. In the case of COCO-full, the improvement is less compared to the previous ones; nonetheless, we still observe that the quality improvement of the pseudo-labels (mIoU) is higher with our assistants.

The noteworthy point is the consistent enhancement of pseudo-label quality, as indicated by mIoU, underscoring the need for pseudo-label fine-tuning in mutual learning. Pre-NMS elevates mIoU by , reflecting its ability to refine pseudo-labels. However, its sole reliance on labeled samples limits further corrections to pseudo-labels. Importantly, pre-NMS operates exclusively with the teacher's NMS, making its impact particularly meaningful due to its minimal intervention in training. Conversely, post-NMS consistently emerges as the strongest performer across experiments, offering compelling evidence for the necessity of corrections to pseudo-labels to enhance results.

Table III. Results with experiments with three different percentages (0.5%, 1%, and 5%) of labeled data using $mAP_{(50:95)}$ and the mean intersection over union (mIoU) at 50%.

		0.5%		1%		5%	
Dataset	Method	mAP	mIoU	mAP	mIoU	mAP	mIoU
Single-class	Supervised	15.67	-	29.51	-	46.19	-
	Unbiased	20.11(+5.68)	78.1	40.09(+10.51)	79.0	56.48(+10.29)	85.8
	pre-NMS	21.48(+5.81)	82.8	41.01(+11.5)	81.1	57.83(+11.64)	85.3
	post-NMS	**31.57(+15.9)**	**93.7**	**55.25(+25.74)**	**92.7**	**62.13(+15.94)**	**92.8**
10-classes	Supervised	12.78	-	21.52	-	28.4	-
	Unbiased	22.38(+9.61)	74.8	31.27(+9.75)	79.5	41.06(+12.66)	85.2
	pre-NMS	22.15(+9.37)	76.3	31.31(+9.79)	81.8	40.58(+12.18)	84.8
	post-NMS	**29.01(+15.86)**	**91.6**	**34.77(+13.25)**	**90.2**	**41.42(+12.89)**	**93.05**
COCO-full	Supervised	5.42	-	8.05	-	12.53	-
	Unbiased	10.09(+4.67)	68.5	17.74(+9.69)	73.3	21.29(+8.76)	84.1
	pre-NMS	9.14(+3.72)	70.0	18.03(+9.98)	75.3	20.45(+7.92)	82.8
	post-NMS	**10.41(+4.99)**	**88.0**	**18.53(+10.48)**	**89.2**	**22.64(+10.11)**	**87.7**

Conclusions and Discussions

We revealed the challenge of achieving precise localization in pseudo-labels, especially in OD's regression. This is due to the localization accuracy not meeting expectations in SS.

We explored using assistant models to help the teacher in pseudo-label creation, particularly before (pre) and after (post) NMS is performed. Our findings showed significant enhancements in three datasets with minimal labeled data ($1%, 0.5%, and 5%).

There is still room for improvement in bridging the gap between SS-OD and fully supervised methods. Our current assistant models enhance pseudo-label quality, but to excel on the COCO-full dataset with more classes, a more robust approach is needed.

Referencias

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015, number:7553 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nature14539

[2] J. E. Van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” Machine learning, vol. 109, no. 2, pp. 373–440, 2020.

[3] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence,” in Advances in Neural Information Processing Systems, vol. 33. Curran Associates, Inc., 2020, pp. 596–608.

[4] A. Kurakin, C. Raffel, D. Berthelot, E. D. Cubuk, H. Zhang, K. Sohn, and N. Carlini, “Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring,” in ICLR, 2020. [Online]. Available: https://openreview.net/pdf?id=HklkeR4KPB

[5] S. Laine and T. Aila, “Temporal Ensembling for Semi-Supervised Learning,” July 2022. [Online]. Available: https://openreview.net/forum?id=BJ6oOfqge

[6] D.-h. Lee, “Pseudo-Label: The Simple Semi-Supervised Learning Method for Deep Neural Networks,” in International Conference on Machine Learning (ICML). ACM, Aug. 2013.

[7] H. Pham, Z. Dai, Q. Xie, and Q. V. Le, “Meta pseudo labels,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 11 557–11

[8] O. Chapelle, B. Sch ̈olkopf, and A. Zien, Eds., Semi-supervised learning, ser. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2006.

[9] L. Wang and K.-J. Yoon, “Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 6, pp. 3048–3068, 2022.

[10] Y.-C. Liu, and C.-Y. Ma, “Unbiased teacher for semi-supervised object detection,” in International Conference on Learning Representations, 2021.

[11] H. Li, Z. Wu, A. Shrivastava, and L. S. Davis, “Rethinking Pseudo Labels for Semi-supervised Object Detection,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 2, pp. 1314–1322, June 2022, number: 2.

[12] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Doll ́ar, and R. Girshick, “Segment Anything,” Apr. 2023, arXiv:2304.02643 [cs]. Available: http://arxiv.org/abs/2304.02643

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” June 2021, arXiv:2010.11929 [cs]. [Online]. Available: http://arxiv.org/abs/2010.11929

[14] P. Bachman, O. Alsharif, and D. Precup, “Learning with Pseudo-Ensembles,” in Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc., 2014.

[15] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” in Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc., 2017.

[16] Y. Grandvalet and Y. Bengio, “Semi-supervised Learning by Entropy Minimization,” in Advances in Neural Information Processing Systems, vol. 17. MIT Press, 2004.

[17] J. Hoffman, S. Guadarrama, E. S. Tzeng, R. Hu, J. Donahue, R. Girshick, T. Darrell, and K. Saenko, “LSDA: Large Scale Detection through Adaptation,” in Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc., 2014.

[18] J. Gao, J. Wang, S. Dai, L.-J. Li, and R. Nevatia, “Note-rcnn: Noise tolerant ensemble rcnn for semi-supervised object detection,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.

[19] J. Jeong, S. Lee, J. Kim, and N. Kwak, “Consistency-based Semi-supervised Learning,” in Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., 2019.

[20] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister, "A Simple Semi-Supervised Learning Framework for Object Detection," Tech. Rep., May 2020, publication Title: arXiv e-prints ADS Bibcode: 2020arXiv200504757S Type: article. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv200504757S

[21] P. Mi, J. Lin, Y. Zhou, Y. Shen, G. Luo, X. Sun, L. Cao, R. Fu, Q. Xu, and R. Ji, "Active teacher for semi-supervised object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 14 482–14 491.

[22] K. Zhang and D. Liu, "Customized Segment Anything Model for Medical Image Segmentation," Apr. 2023, publication Title: arXiv ADS Bibcode: 2023arXiv230413785Z. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230413785Z

[23] R. Zhang, Z. Jiang, Z. Guo, S. Yan, J. Pan, H. Dong, P. Gao, and H. Li, "Personalize Segment Anything Model," May 2023, publication Title: arXiv e-prints ADS Bibcode: 023arXiv230503048Z. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230503048Z

[24] S. Roy, T. Wald, G. Koehler, M. R. Rokuss, N. Disch, J. Holzschuh, D. Zimmerer, and K. H. Maier-Hein, "SAM.MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model," Apr. 2023, publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230405396R: https://ui.adsabs.harvard.edu/abs/2023arXiv230405396R

[25] W. Ji, J. Li, Q. Bi, T. Liu, W. Li, and L. Cheng, "Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-world Applications," Apr. 2023, publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230405750J. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230405750J

[26] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," International Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, Sept. 2013. [Online]. Available: https://doi.org/10.1007/s11263-013-0620-5

[27] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, H. Chi, X. Hu, D.P. Fan, F. Dong, and D. Ni, "Segment Anything Model for Medical Images?" Apr. 2023, publication Title: arXiv e- prints ADS Bibcode: 2023arXiv230414660H. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230414660H

[28] J. Dyson, A. Mancini, E. Frontoni, and P. Zingaretti, "Deep Learning for Soil and Crop Segmentation from Remotely Sensed Data," Remote Sensing, vol. 11, no. 16, p. 1859, Jan. 2019, number: 16 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/11/16/1859

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores aquí firmantes declaramos que no se utilizó ninguna herramienta de IA para la conceptualización, traducción o redacción de este artículo.