Cross-Domain Semi-Supervised Organ Detection

Nian Li[1] NIAN.LI@TUM.DE
Morteza Ghahremani[1,2] MORTEZA.GHAHREMANI@TUM.DE
Bailiang Jian[1,2] BAILIANG.JIAN@TUM.DE
Pascual Tejero Cervera[1] PASCUAL.TEJERO@@TUM.DE
Benedikt Wiestler[1,2] B.WIESTLER@TUM.DE
Marcus Makowski[1] MARCUS.MAKOWSKI@TUM.DE
Christian Wachinger[1,2] CHRISTIAN.WACHINGER@TUM.DE
[1] *Technical University of Munich (TUM),* [2] *Munich Center for Machine Learning (MCML)*

Editors: Under Review for MIDL 2026

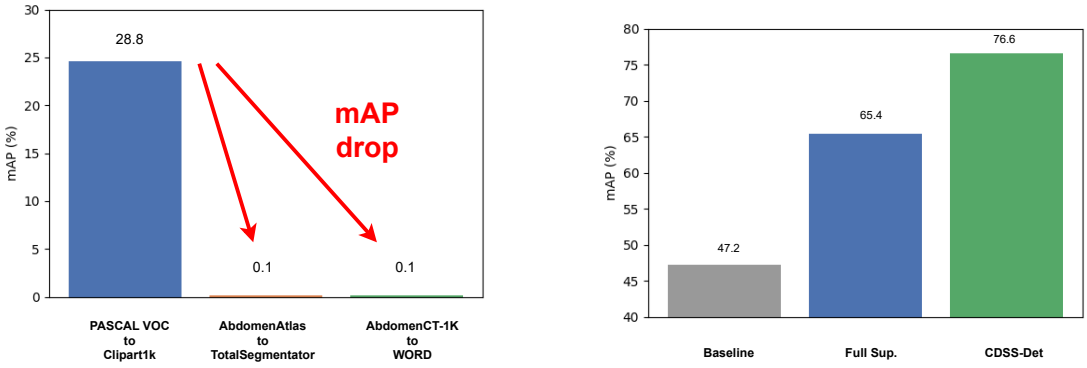
Abstract

Domain adaptation for 3D organ detection in CT imaging is challenging due to variations in scanner types, imaging protocols, and overall acquisition conditions. As supervised detection models require large, annotated datasets from diverse scanners and institutions, semi-supervised approaches have gained attention for their ability to leverage limited unlabeled target data. However, traditional semi-supervised methods typically fail to make effective use of the few labeled target samples and most often do not yield satisfactory results. To address this limitation, we introduce a novel cross-domain semi-supervised detection framework (CDSS-Det) built upon the Transformer-based Organ-DETR model. CDSS-Det synergistically integrates pseudo-labeling, curriculum learning, and domain adaptation to enable effective knowledge transfer from a well-annotated source domain to a target domain with limited labels. Experiments on multi-domain CT datasets demonstrate that incorporating a small number of labeled target samples significantly boosts detection performance over conventional domain adaptation and semi-supervised methods. CDSS-Det consistently achieves higher mean Average Precision (mAP), with notable improvements in detecting small organs, and surpasses a fully supervised model trained solely on the labeled target domain by over 10%. These results underscore the potential of CDSS-Det in efficiently leveraging both labeled and unlabeled target data in cross-domain organ detection, advancing annotation-efficient deep learning models in medical imaging.

Keywords: Organ Detection, Domain Transfer, Domain Adaptation

1. Introduction

Accurate 3D organ detection from CT scans is essential for disease diagnosis, surgical planning, and downstream applications such as segmentation (Ma et al., 2021). Although deep learning-based object detection models have achieved impressive results on well-annotated datasets (Ghahremani et al., 2025), their generalization to new domains is hindered by substantial domain shifts, arising from variations in scanner types, imaging protocols, and patient demographics. As illustrated in Figure 1, this shift is particularly pronounced in medical imaging, where transferring a model between datasets yields a drastic drop in mean Average Precision (mAP). In contrast, object detection generalizes well on natural image datasets (Li et al., 2022).



(a) (b)

Figure 1: (a) Domain gap severity in medical imaging vs. non-medical datasets. (b) Effectiveness of CDSS-Det in cross-domain organ detection on WORD dataset, where CDSS-Det outperforms the baseline and fully supervised model.

Furthermore, developing high-performance object detectors requires large-scale labeled datasets, the annotation of which is both resource-intensive and time-consuming. To overcome this challenge, *domain adaptation* has emerged as a promising approach, enabling models trained on a labeled source domain to effectively generalize to a related but distinct target domain (Guan et al., 2021). Despite progress in domain adaptation, most prior works focus on unsupervised domain adaptation (UDA) (Tzeng et al., 2017; Chen et al., 2020). Although no labeled target data is required in UDA methods, they often struggle in medical imaging due to substantial domain shifts caused by variations in image acquisition and patient demographics (Zhang et al., 2020). *Semi-supervised learning* (SSL) has been explored to address the scarcity of labeled data by leveraging unlabeled target data alongside limited labeled samples (Ouali et al., 2021; Bai et al., 2017). However, existing semi-supervised object detection methods in medical imaging often rely solely on labeled source data or unlabeled target data, making it challenging to achieve satisfactory performance due to domain shifts and the lack of direct supervision on the target domain (Jeong et al., 2019; Sohn et al., 2020). Few-shot learning approaches attempt to mitigate this issue by training on a small set of labeled target samples (Wang et al., 2020), but they fail to utilize the large pool of available unlabeled target data, limiting their ability to generalize effectively.

Recent studies have explored domain adaptation and semi-supervised learning for medical image analysis in classification and segmentation tasks. For instance, Yuan et al. (Yuan et al., 2024) utilized pseudo-labeling for COVID-19 detection, demonstrating that incorporating unlabeled target data can enhance domain transfer in medical classification. In the segmentation domain, Cai et al. (Cai et al., 2024) proposed a Class-Aware Mutual Mixup strategy with triple alignments to improve cross-domain semi-supervised segmentation. Basak and Yin (Basak and Yin, 2023) further introduced a consistency-regularized disentangled contrastive learning approach for semi-supervised domain adaptation in medical image segmentation, highlighting the value of combining consistency and contrastive learning in pixel-level tasks. While these works show promising results in classification and segmentation, 3D object detection presents unique challenges such as instance-level localization and class imbalance, which are not addressed by methods designed for pixel-wise or image-level tasks. *Contrastive learning* techniques such as PixPro (Xie et al., 2021a) have

also been introduced to improve feature consistency between domains in self-supervised settings. However, their applicability to volumetric medical detection tasks remains underexplored. Our work focuses on extending semi-supervised cross-domain learning to 3D organ detection, bridging this gap by leveraging both labeled and unlabeled data through curriculum-driven pseudo-labeling.

We address this issue by introducing cross-domain semi-supervised organ detection (CDSS-Det) that effectively leverages both labeled and unlabeled target data. CDSS-Det builds upon the Transformer-based Organ-DETR model (Ghahremani et al., 2025) and integrates pseudo-labeling and curriculum learning strategies to enhance adaptation. Unlike existing domain adaptation and semi-supervised methods, CDSS-Det utilizes a small number of labeled target samples in addition to unlabeled target data, significantly improving performance in challenging small-organ detection scenarios. Experimental results demonstrate that CDSS-Det outperforms existing baselines, achieving superior mean Average Precision and highlighting the importance of integrating both labeled and unlabeled target data for cross-domain organ detection (Figure 1). Our contributions are summarized below.

- To the best of our knowledge, the proposed CDSS-Det is the first cross-domain semi-supervised 3D object detection framework for medical imaging. It jointly leverages labeled source data, limited labeled target data, and abundant unlabeled target data to bridge the domain gap between source and target domains.
- We introduced a curriculum learning strategy for pseudo-labeling, where a dynamic weighting mechanism adjusts the contribution of pseudo-labels during training, leading to more stable optimization and improved accuracy.
- We demonstrated the effectiveness of CDSS-Det through extensive experiments on two cross-domain 3D organ detection benchmarks, showing state-of-the-art results and significant improvements on small organs, which are particularly challenging due to anatomical variability. We also reported ablation studies to analyze the contribution of each component.

2. Methodology

Preliminaries. Organ detection in 3D CT imaging involves localizing anatomical structures using axis-aligned bounding boxes and assigning class labels to detected organs (Shin et al., 2016). Detection performance is evaluated using mAP at different IoU thresholds, mean Average Recall (mAR), precision, and recall. We build upon Organ-DETR (Ghahremani et al., 2025), a Transformer-based 3D object detector designed for medical imaging. It introduces MultiScale Attention (MSA) for handling varying organ sizes and Dense Query Matching (DQM) to improve query-object associations, enhancing detection robustness in CT scans.

Problem Definition. We address *cross-domain semi-supervised* organ detection, where a model is trained using a labeled source dataset $D_s = \{(X_i^s, Y_i^s)\}$, a small set of labeled target samples $D_t = \{(X_i^t, Y_i^t)\}$, and a larger set of unlabeled target scans $U_t = \{X_i^t\}$. The primary challenge is the domain gap between D_s and D_t, which can degrade detection performance

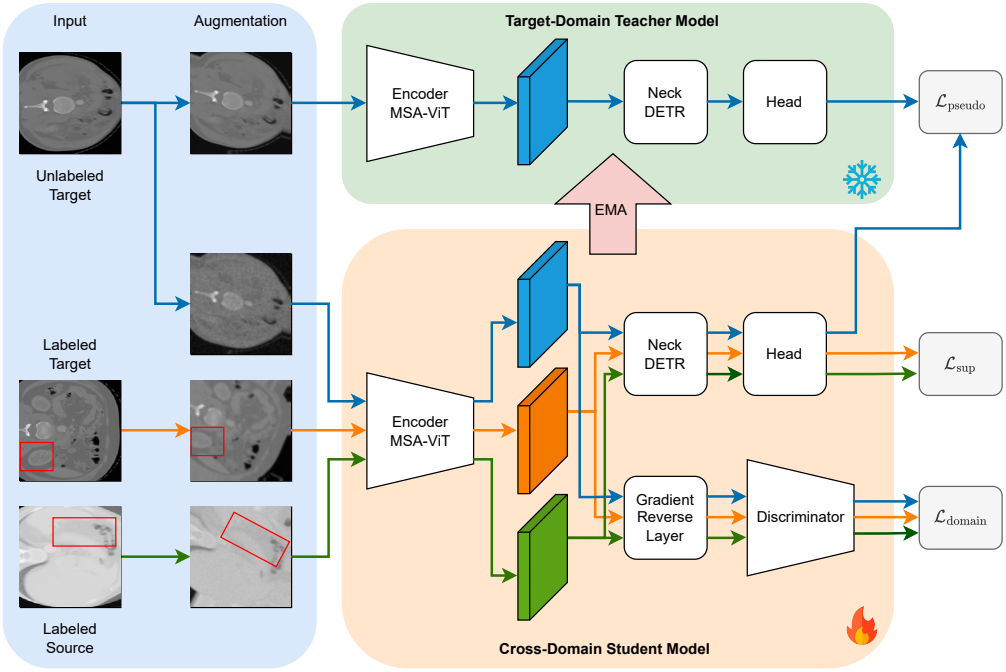


Figure 2: Overview of the CDSS-Det framework. Both student and teacher branches are based on Organ-DETR, which incorporates Multi-Scale Attention (MSA) and Dense Query Matching (DQM) to enhance 3D organ detection. The teacher model processes unlabeled target data to generate pseudo-labels, while the student model is trained using labeled source data, labeled target data, and unlabeled target data. Supervised loss is computed from labeled predictions, domain loss is obtained via a discriminator with a gradient reversal layer, and pseudo loss is calculated between student predictions and pseudo labels. The teacher is updated via Exponential Moving Average (EMA) of the student weights.

when models trained on D_s are directly applied to D_t, see Figure 1. Annotating large-scale target domain data is costly and time-consuming, making fully supervised adaptation impractical. Semi-supervised learning offers a solution by leveraging unlabeled target data through pseudo-labeling, self-training, and teacher-student learning (Ouali et al., 2021).

2.1. CDSS-Det Framework for Cross-Domain Semi-Supervised Learning

Figure 2 provides an overview of the CDSS-Det framework. Our approach builds upon the strengths of semi-supervised learning in medical imaging, where teacher-student frameworks have proven effective at refining pseudo-labels and enhancing model generalization (Xie et al., 2020). However, existing methods such as Adaptive Teacher (Li et al., 2022) rely exclusively on pseudo-labels derived from unlabeled target data. This reliance is problematic for cross-domain adaptation, as the inherent domain gap between source and target datasets can lead to the propagation of erroneous pseudo-labels. In contrast, CDSS-Det explicitly leverages a small amount of labeled target data in conjunction with unlabeled samples, while incorporating replay-based learning, adversarial domain adaptation, and curriculum learning to improve robustness.

Mitigating Noisy Pseudo-Labels: To harness additional target samples via pseudo-labeling without degrading performance due to noise, we first apply confidence-based filtering and spatial refinement. A pseudo-label is retained only if its classification confidence exceeds a threshold τ:

$$\hat{y}_i^t = \arg\max p(y_i|X_i^t), \quad \text{if} \quad p(y_i|X_i^t) > \tau, \tag{1}$$

where X_i^t is the i-th unlabeled target sample, y_i represents the ground-truth class label, and $\hat{y}_i^t$ is the assigned pseudo-label. To eliminate redundant detections, we further apply IoU-based Non-Maximum Suppression (NMS). Notably, since pseudo-labels are selected based on classification confidence, their bounding box predictions may remain unreliable. Instead of using regression loss for pseudo-labels, we apply only cross-entropy classification loss calculated by predicted class of the unlabeled sample y_i^t and its pseudo-label $\hat{y}_i^t$:

$$\mathcal{L}_{\text{pseudo}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathcal{L}_{\text{CE}}(y_i^t, \hat{y}_i^t). \tag{2}$$

Reducing Confirmation Bias in Pseudo-Labeling: A major drawback of pseudo-labeling is confirmation bias, where incorrect pseudo-labels reinforce prediction errors. To mitigate this, we introduce a curriculum learning approach that dynamically adjusts the pseudo-label weight based on the student's classification loss:

$$\lambda_{\text{pseudo}}(t) = \begin{cases} \min(\lambda_{\text{pseudo}}(t-1) + \Delta, \lambda_{\max}), & \text{if } \mathcal{L}_{\text{cls}} < \delta \\ \max(\lambda_{\text{pseudo}}(t-1) - \Delta, \lambda_{\min}), & \text{if } \mathcal{L}_{\text{cls}} \geq \delta. \end{cases} \tag{3}$$

If the student classification loss on labeled data is lower than a threshold δ, the pseudo-label loss coefficient increases by Δ, up to a maximum value $\lambda_{\max}$. Conversely, if the classification loss remains high, the coefficient is decreased by Δ, but not below a minimum value $\lambda_{\min}$. This ensures that the influence of pseudo-labels grows only when the model is confident and remains bounded to avoid over-reliance on noisy labels.

Preventing Forgetting with Replay Strategy: Cross-domain adaptation can lead to catastrophic forgetting of source domain knowledge. To prevent this, we adopt a replay strategy inspired by continual learning (Rolnick et al., 2019). Unlike prior cross-domain methods that either randomly select source samples or use the entire source dataset during training, we explicitly identify and replay the *hardest* labeled source samples, determined based on low detection confidence. These hard examples are more informative for adaptation and are replayed alongside target data to reinforce discriminative feature learning. This targeted replay improves transferability while mitigating forgetting.

Bridging Domain Gap with Adversarial Adaptation: The pronounced domain shift between source and target datasets is addressed through adversarial domain adaptation. A domain discriminator is attached to the backbone features extracted by the student model and trained to differentiate between source and target domains. Simultaneously, the student model is optimized to produce domain-invariant features via a gradient reversal layer (GRL) (Zhang et al., 2020), which inverts the gradients of the domain loss. This adversarial interplay effectively aligns the feature distributions across domains, enhancing detection performance.

Teacher-Student Learning with EMA: In our framework, the teacher model is maintained as an Exponential Moving Average (EMA) of the student model's parameters. Rather than receiving direct gradient updates, the teacher is updated as:

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s, \tag{4}$$

where θ_t and θ_s denote the teacher and student model parameters, respectively, and α is the EMA decay factor. This prevents unstable pseudo-labeling updates and provides a more consistent training signal for the student.

Student Loss Criterion: The overall training loss for the student model integrates three key components: supervised loss on labeled data, pseudo-label loss on unlabeled data, and domain loss for alignment:

$$\mathcal{L}_{\text{student}} = \mathcal{L}_{\text{sup}} + \lambda_{\text{pseudo}}(t)\mathcal{L}_{\text{pseudo}} + \lambda_{\text{domain}}\mathcal{L}_{\text{domain}}. \tag{5}$$

The supervised loss $\mathcal{L}_{\text{sup}}$ is computed on both labeled source and labeled target data:

$$\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{sup}}^{\text{src}} + \mathcal{L}_{\text{sup}}^{\text{tgt}}, \tag{6}$$

where each term includes a combination of classification and localization losses:

$$\mathcal{L}_{\text{sup}}^{(\cdot)} = \mathcal{L}_{\text{cls}}^{(\cdot)} + \mathcal{L}_{\text{bbox}}^{(\cdot)} + \mathcal{L}_{\text{giou}}^{(\cdot)} + \mathcal{L}_{\text{seg}}^{(\cdot)}. \tag{7}$$

Here, $\mathcal{L}_{\text{cls}}^{(\cdot)}$ is classification loss for organ category prediction; $\mathcal{L}_{\text{bbox}}^{(\cdot)}$ is L1 loss for bounding box regression; $\mathcal{L}_{\text{giou}}^{(\cdot)}$ is Generalized IoU loss to improve spatial accuracy; and $\mathcal{L}_{\text{seg}}^{(\cdot)}$ is an optional segmentation loss (if segmentation maps are available), composed of cross-entropy and Dice losses:

$$\mathcal{L}_{\text{seg}}^{(\cdot)} = \mathcal{L}_{\text{ce}}^{(\cdot)} + \mathcal{L}_{\text{dice}}^{(\cdot)}. \tag{8}$$

The dynamic weight $\lambda_{\text{pseudo}}(t)$ controls the influence of pseudo-labels based on the model's confidence, following a curriculum strategy. Meanwhile, λ_{domain} scales the domain loss, which helps minimize domain shift through adversarial learning. Together, these components enable CDSS-Det to effectively integrate both labeled and unlabeled target data for cross-domain 3D organ detection.

3. Experiments

Datasets.We evaluate our method on two cross-domain 3D organ detection settings. The first setting, AbdomenAtlas → TotalSegmentator, uses AbdomenAtlas, a large-scale, multi-center dataset with annotations for multiple abdominal organs (Qu et al., 2023). We use AbdomenAtlas 1.0 with 5,195 scans, where 3,524 scans were used as the training set to pre-train our model. To the best of our knowledge, this is the first study utilizing AbdomenAtlas for cross-domain, semi-supervised organ detection. The scans in this dataset include both healthy and diseased organs such as tumors and fatty liver. Axis-aligned bounding boxes are extracted from segmentation maps and used as detection labels. Scans are normalized using the 0.5 and 99.5 percentiles of non-background voxels, clipped to the [0, 1] range. Augmentations (applied with 50% probability) include random intensity scaling/shifting (up to 10%), rotation (±5°), translation (up to 10%), and zooming (±10%).

In the target dataset, TotalSegmentator (Wasserthal et al., 2022), the 113 scans are split into 8 labeled target scans, 105 unlabeled target scans, 21 validation scans, and 29 test scans, creating a challenging adaptation scenario with limited supervision. 8 common organs from these two datasets are selected for our detection task. The TotalSegmentator dataset also includes both healthy and pathological cases.

The second setting, AbdomenCT-1K → WORD, involves AbdomenCT-1K, a dataset of 1,112 high-resolution 3D CT scans from five sources, covering the liver, left kidney, right kidney, spleen, and pancreas (Ma et al., 2021). These scans exhibit variability in slice thickness and pixel spacing, making them suitable for cross-domain adaptation. 732 samples in the training set are used to pre-train the model. The scans contain both healthy and diseased organs, including cancer and tumors. The same normalization and augmentation techniques as above are applied to improve robustness.

The target dataset, WORD, contains 150 CT scans acquired from a single medical center with high-resolution imaging and multiple organ annotations (Miao et al., 2021). We use 31 labeled target scans, 75 unlabeled target scans, 14 validation scans, and 29 test scans. The domain shift between these datasets presents significant challenges for adaptation and pseudo-labeling. 5 common organs from these two datasets are selected for our detection task. Similar to other datasets, bounding boxes are derived from segmentation masks.

Training and Evaluation Setup. Pseudo-labels are filtered with a confidence threshold of 0.8 and refined via NMS (IoU = 0.5). The teacher model is updated using EMA (decay = 0.9996). Training uses AdamW (weight decay 1×10^{-4}) with an initial learning rate of 2×10^{-4}, decaying by 0.1 every 500 epochs, for a total of 2,500 epochs. Each iteration processes one labeled source, one labeled target, and one unlabeled target sample. The supervised loss includes classification, bounding-box regression, and optional segmentation, with weights: cls = 2, bbox = 5, giou = 2, segce = 2, segdice = 2. The pseudo-label weight $\lambda_{\text{pseudo}}(t)$ adjusts by ± 0.1 based on a classification-loss threshold of 0.01, and is clipped to $[0, 2]$. Domain adaptation loss is added with weight 0.2. A replay strategy selects as many hard source samples as labeled target samples. All experiments are conducted on an NVIDIA A100 GPU (80 GB).

To evaluate CDSS-Det, we compare it against multiple baselines, including a baseline model trained solely on labeled target data, a pre-trained variant initialized with a source-trained model, and a fully supervised (Full Sup.) model trained with full target annotations. Additionally, we conduct an ablation study to assess the contribution of different components within CDSS-Det.

Recent cross-domain semi-supervised methods in classification (Yuan et al., 2024) and segmentation (Cai et al., 2024; Basak and Yin, 2023) have shown encouraging results, but they lack publicly available source code and implementation details, making direct comparisons infeasible. Furthermore, methods like (Basak and Yin, 2023), which focus on 2D segmentation, are difficult to adapt to 3D object detection due to substantial differences in task formulation and model architecture. As an alternative, we include PixPro (Xie et al., 2021b), a pixel-level contrastive learning-based consistency method originally developed for 2D natural images. In our setup, PixPro is used for self-supervised feature consistency in the 3D medical domain, with a loss coefficient of 0.01. Note that all reported results are obtained using the student model during inference.

Experimental Results. Table 1 summarizes the detection performance of different training strategies on the WORD and TotalSegmentator datasets. CDSS-Det consistently achieves the highest detection performance across different IoU thresholds and organ sizes, demonstrating its effectiveness in leveraging labeled source, labeled target, and unlabeled target data for cross-domain semi-supervised organ detection. The baseline model, trained only on labeled target data, achieves the lowest performance on both datasets, highlighting the

Table 1: Detection performance on the WORD and TotalSegmentator datasets under different training strategies. The baseline is trained solely on labeled target data. The Pre-trained model is initialized with a source-trained model to improve generalization. The Full Sup. assumes access to all target data with full annotations. We report mAP grouped by organ size for small (S), medium (M), and large (L) organs.

Dataset	Method	mAP ↑			mAR ↑			mAP ↑ by size		
		Total	75%	50%	Total	75%	50%	S	M	L
WORD	Baseline	47.2	44.7	96.5	54.0	57.9	97.2	26.2	50.4	58.7
	Pre-trained	72.9	85.6	97.5	77.7	89.0	**98.6**	54.0	77.2	79.0
	Full Sup.	65.4	78.1	**98.1**	71.1	87.1	**98.6**	40.4	71.4	72.6
	CDSS-Det	**76.6**	**88.8**	97.4	**79.9**	**91.7**	97.9	**58.4**	**80.7**	**82.7**
TotalSeg	Baseline	13.7	2.1	50.7	21.0	8.2	62.5	6.0	13.9	21.0
	Pre-trained	62.5	68.1	**94.5**	68.2	75.5	95.9	32.1	76.6	64.7
	Full Sup.	52.7	62.7	88.6	59.2	69.9	91.8	20.5	64.4	61.6
	CDSS-Det	**70.1**	**82.2**	94.1	**75.2**	**85.8**	**96.3**	**42.2**	**81.7**	**74.5**

challenges of learning from limited labeled data in the target domain. Pre-training on the source dataset significantly improves performance, confirming the importance of transferring knowledge from a larger labeled dataset. Remarkably, the Full Sup. model, despite full supervision, falls short of CDSS-Det, indicating that refined pseudo-labeling can effectively supplement sparse annotations and improve generalization. Figure 3 presents a qualitative comparison of organ detection results between Full Sup. and CDSS-Det.

CDSS-Det provides notable gains, particularly for small organs, which are traditionally difficult to detect due to their anatomical variability and limited representation in training data. As shown in Figure 4, CDSS-Det significantly outperforms the Full Sup. model across all organ sizes, with the largest improvements observed on small organs in both WORD and TotalSegmentator datasets. These results demonstrate that pseudo-label refinement and dynamic weighting are especially effective in addressing the challenges of detecting anatomically variable and underrepresented structures.

Recent works in medical domain adaptation and semi-supervised learning, such as Yuan et al. (Yuan et al., 2024), Cai et al. (Cai et al., 2024), and Basak et al. (Basak and Yin, 2023), have explored related ideas in the context of classification and segmentation. However, these methods focus on image-level classification (Yuan et al., 2024) or pixel-wise segmentation (Cai et al., 2024; Basak and Yin, 2023), and do not address the instance-level challenges inherent in 3D object detection. In addition, they lack open-source implementations and sufficient details for reproducibility, and methods designed for 2D segmentation tasks are not straightforward to adapt to volumetric 3D detection problems. As a result, we include PixPro (Xie et al., 2021a) as a representative self-supervised learning baseline in our ablation study to evaluate the potential of pixel-level consistency in 3D detection tasks.

Table 2 reports results for the ablation study that evaluated the contribution of individual components within CDSS-Det. The replay strategy improves feature stability by retaining harder samples from the source domain. Domain adaptation provides an additional performance boost by reducing feature discrepancies between source and target distribu-

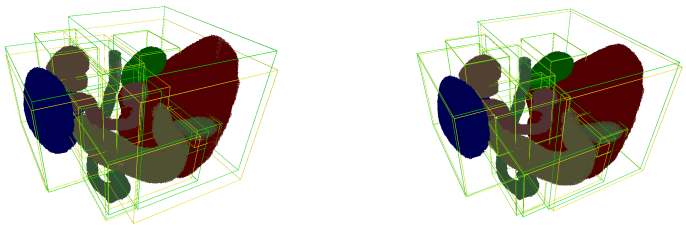
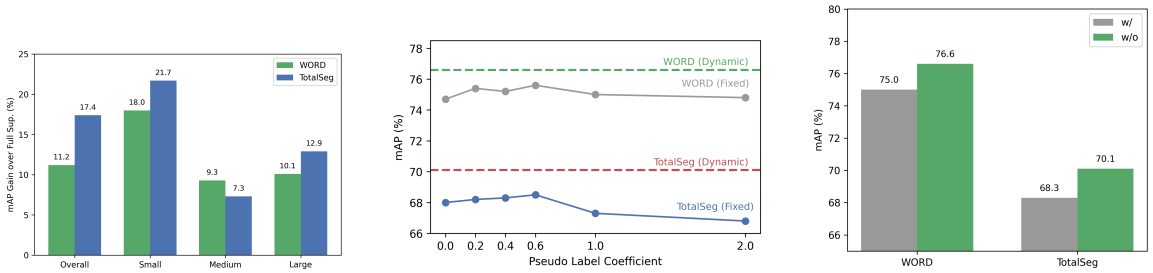


Figure 3: Visualization of organ detection results on WORD dataset. The left shows results from Full Sup. and the right presents results from CDSS-Det. Ground truth bounding boxes are in green, and predicted bounding boxes are in yellow.



(a) (b) (c)

Figure 4: (a) mAP gain of CDSS-Det over the Full Supervised model across different organ sizes (Small, Medium, Large) on the WORD and TotalSegmentator datasets. (b) Ablation study on the pseudo-label loss coefficient strategies, and (c) Ablation study on the weak-strong augmentation impact.

tions, but its impact remains relatively small compared to pseudo-labeling. Self-supervised feature consistency with PixPro does not yield substantial improvements, indicating that contrastive learning may be less effective for volumetric medical imaging. Overall, the highest gains come from pseudo-labeling and dynamic weighting, which allow the model to gradually incorporate pseudo-labels without introducing excessive noise. Note that CDSS-Det corresponds to the last row in the table.

To investigate the impact of curriculum learning, we compare CDSS-Det using fixed pseudo-labeling loss coefficients against our dynamic curriculum strategy. As shown in 3, using fixed coefficients leads to unstable performance, with mAP values fluctuating between 74.7 and 75.6 in WORD dataset and fluctuating between 66.8 and 68.5 in TotalSegmentator dataset depending on the coefficient. Notably, large coefficients such as 1.0 and 2.0 result in degraded performance in both two datasets due to training divergence. In contrast, our curriculum-based dynamic weighting strategy achieves a significantly higher mAP of 76.6 in WORD dataset and 70.1 in TotalSegmentator dataset, demonstrating its ability to balance supervision and mitigate overfitting or label noise during training. This highlights that dynamic pseudo loss weighting is crucial for stable and effective semi-supervised learning in medical detection scenarios.

To investigate the effect of weak-strong augmentation in the context of medical image detection, we compare CDSS-Det's performance with and without this strategy. As shown in Figure 3, applying weak-strong augmentation decreases the mAP from 76.6 to 75.0 in WORD dataset and decreases from 70.1 to 68.3 in TotalSegmentator dataset, contrary to

Table 2: Ablation study on the WORD and TotalSegmentator datasets. Configuration settings include Replay (R), Domain Alignment (D), PixPro (P), Pseudo-Labeling (PL), and Dynamic Pseudo-Labeling (Dyn).

	Configuration					mAP↑			mAR↑			mAP ↑ by size		
	R	D	P	PL	Dyn	Total	75%	50%	Total	75%	50%	S	M	L
WORD	✓					74.5	87.7	96.1	79.7	**92.4**	97.9	52.4	78.7	**84.2**
	✓	✓				74.7	86.2	**97.5**	79.6	91.0	**98.6**	52.7	79.6	82.0
	✓	✓	✓			74.8	87.7	96.4	79.5	91.7	97.9	55.3	79.5	80.2
	✓	✓		✓		75.6	88.2	97.4	79.5	91.7	97.9	56.6	80.0	81.3
	✓	✓		✓	✓	**76.6**	**88.8**	97.4	**79.9**	91.7	97.9	**58.4**	**80.7**	82.7
TotalSeg	✓					67.4	78.8	92.7	73.2	84.3	95.1	40.3	78.5	72.3
	✓	✓				68.0	78.8	94.2	73.8	83.8	95.9	39.3	79.9	73.0
	✓	✓	✓			68.1	78.7	93.9	73.1	83.4	95.1	39.9	80.2	72.3
	✓	✓		✓		68.5	80.1	**94.2**	74.2	85.1	96.0	41.7	78.9	74.3
	✓	✓		✓	✓	**70.1**	**82.2**	94.1	**75.2**	**85.8**	**96.3**	**42.2**	**81.7**	**74.5**

trends observed in natural image domains. For instance, in Adaptive Teacher (Li et al., 2022), weak-strong augmentation significantly boosts detection performance across various cross-domain scenarios. However, in our experiments in 3D CT data, this technique appears to hinder performance. One possible explanation is that aggressive augmentation may distort subtle anatomical structures and degrade the quality of pseudo labels, especially when the model already generates high-quality predictions under weak augmentation alone. This highlights a critical difference between medical and natural image domains, where preserving spatial fidelity is often more important than encouraging invariance through strong perturbations.

Overall, our results demonstrate that CDSS-Det consistently outperforms a range of strategies, from models trained solely on labeled target data and those leveraging pre-training, to fully supervised approaches. By integrating refined pseudo-labeling and dynamic weighting with both labeled and unlabeled data, CDSS-Det effectively mitigates domain shifts.

4. Conclusion

We demonstrated that CDSS-Det effectively leverages labeled and unlabeled data for cross-domain 3D organ detection, surpassing both pre-trained and fully supervised models. Pseudo-label refinement contributes the most to performance gains, while replay and domain adaptation further enhance generalization. These findings highlight the potential of semi-supervised learning not only to reduce annotation efforts and enhance detection robustness but also to address the domain shifts inherent in medical imaging. By facilitating reliable domain transfer, our approach takes a crucial step toward translating organ detection approaches into clinical practice. Future work will explore extending this framework to additional imaging modalities and anatomical structures to further validate its effectiveness.

References

Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hirofumi Suzuki, Martin Rajchl, and Paul M. Matthews. Semi-supervised learning for network-based cardiac mr image segmentation. In *MICCAI*, pages 253–260. Springer, 2017.

Hritam Basak and Zhaozheng Yin. Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2023*, pages 260–270. Springer, 2023. doi: 10.1007/978-3-031-43901-8_25.

Zhuotong Cai, Dongze Han, Yizhe Zhang, Yuting Ding, and Lequan Lin. Class-aware mutual mixup with triple alignments for semi-supervised cross-domain segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 402–412. Springer, 2024.

Yunjey Chen, Jaehyun Kim, Jiashi Wang, and Quanquan Zhang. Revisiting unsupervised domain adaptation for medical imaging. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 441–450. Springer, 2020.

Morteza Ghahremani, Benjamin Raphael Ernhofer, Jiajun Wang, Marcus Makowski, and Christian Wachinger. Organ-DETR: 3d organ detection transfomer with multiscale attention and dense query matching. *IEEE Transactions on Medical Imaging*, 2025. doi: 10.1109/TMI.2025.3543581.

Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, and L. Zheng. Domain adaptation for medical image analysis: A survey. *IEEE Transactions on Medical Imaging*, 40(11):2975–2993, 2021. doi: 10.1109/TMI.2021.3106867.

Jongwon Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak. Consistency-based semi-supervised learning for object detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 10758–10767, 2019.

Yu-Jhe Li, Xiaoliang Dai, Chih-Yao Ma, Yen-Cheng Liu, Kan Chen, Bichen Wu, Zijian He, Kris Kitani, and Peter Vajda. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7581–7590, 2022.

J. Ma, Y. Zhang, S. Gu, P. Nie, Y. Zhu, S. Yang, and Y. Sun. Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6692–6708, 2021. doi: 10.1109/TPAMI.2021.3054826.

S. Miao, Z. Wang, L. Pan, and R. Liao. Whole abdominal organ segmentation with deep convolutional neural networks. *Medical Physics*, 48(4):2038–2050, 2021. doi: 10.1002/mp.14720.

Yassine Ouali, Céline Hudelot, and Mohamed Tami. Semi-supervised deep learning taxonomy and survey for medical imaging applications. *Medical Image Analysis*, 71:102057, 2021.

Chongyu Qu, Tiezheng Zhang, Hualin Qiao, Jie Liu, Yucheng Tang, Alan L Yuille, and Zongwei Zhou. Abdomenatlas-8k: Annotating 8,000 ct volumes for multi-organ segmentation in three weeks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, G. Wayne, and J. Rae. Experience replay for continual learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019. URL https://arxiv.org/abs/1811.11682.

H-C Shin, Matthew Orton, David J Collins, S Doran, and MO Leach. Organ detection using deep learning. In *Medical image recognition, segmentation and parsing*, pages 123–153. Elsevier, 2016.

Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Caiming Lee, and Thomas Pfister. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*, 2020.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7167–7176. IEEE, 2017.

Yue Wang, Quanming Yao, Bohan Ni, Yu Wang, Tongliang Wang, and Dacheng Tao. Few-shot learning: A survey. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.

J. Wasserthal, M. Meyer, H. Breit, J. Cyriac, S. Yang, and M. Segeroth. Totalsegmentator: Robust segmentation of 104 anatomical structures in ct images. *arXiv preprint*, 2022. doi: 10.48550/arXiv.2208.05868.

Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le. Self-training with noisy student improves imagenet classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10687–10698, 2020. doi: 10.1109/CVPR42600.2020.01070.

Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu. Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16684–16693, 2021a. doi: 10.1109/CVPR46437.2021.01641.

Zhenda Xie, Yutong Lin, Zheng Zhang, Yue Cao, Stephen Lin, and Han Hu. Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16684–16693, 2021b.

Runtian Yuan, Heng Hu, Ziqi Yang, Mingxuan Xu, Yiyu Zhang, Maosheng Xu, and Yefeng Zheng. Domain adaptation using pseudo labels for covid-19 detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3977–3985, 2024.

Y. Zhang, L. Yang, W. Zheng, Z. Lin, Q. Tian, and D. Zhang. Generalizing deep networks for medical image segmentation. *Medical Image Analysis*, 65:101761, 2020. doi: 10.1016/j.media.2020.101761.