Lesion-Aware Reconstruction with Principal Network: Enhancing Pseudo-Label Reliability in Semi-Supervised Clinical Lesion Detection

Shiwan DI[1] 22120036@BJTU.EDU.CN
Jupeng LI[1*] LIJUPENG@BJTU.EDU.CN
Yuxuan YANG[1] 23125011@BJTU.EDU.CN
Qian JIN[1] 1575826445@QQ.COM
Guorui AN[1] 24120022@BJTU.EDU.CN
Jingwen Yang[1] 25120219@BJTU.EDU.CN
Yue WANG[1] 22120143@BJTU.EDU.CN
Yong GUO[1] 7579@BJTU.EDU.CN
[1] *Beijing Jiaotong University, Beijing 100044, China*
Xinyue ZHANG[2] 1810303109@BJMU.EDU.CN
Ruohan MA[2] KQMRH@BJMU.EDU.CN
Gang LI[2] KQGANG@BJMU.EDU.CN
[2] *Peking University School and Hospital of Stomatology, Beijing 100081, China*

Editors: Under Review for MIDL 2026

Abstract

Purpose In lesion detection tasks, labeled medical data are often scarce, limiting the performance of fully supervised models. Teacher-Student frameworks based on semi-supervised learning (SSL) have emerged as effective solutions to leverage unlabeled data. However, the inherent high-confidence bias of teacher networks frequently leads to erroneous pseudo-label propagation, degrading the generalization ability of student networks. To address this critical issue, we propose a novel teacher-principal-student (TPS) framework.

Methods The core innovation lies in introducing a Principal network, which integrates lesion-aware reconstruction to filter low-quality pseudo-labels generated by the teacher network. Specifically, the principal network leverages anatomical prior knowledge and reconstruction consistency constraints to assess the reliability of teacher-generated pseudo-labels, ensuring only high-fidelity pseudo-labeled data are used for student network training. This design fundamentally mitigates the adverse effects of the teacher prediction bias and error propagation.

Results Extensive experiments on jaw lesion detection datasets demonstrate the superiority of our approach. With the same annotation ratio, our SSL network achieves 81.5% mAP@0.5, outperforming mainstream semi-supervised methods by 3% while narrowing the performance gap with fully supervised learning to only 3.3%.

Conclusion Our proposed TPS framework outperforms state-of-the-art semi-supervised approaches in jaw lesion detection task. It not only achieves competitive performance comparable to fully supervised models but also significantly reduces reliance on labeled clinical data, providing a reliable technical solution to promote the translation of clinical lesion detection systems. Our code will be released at https://github.com/dsw847902897/Lesion-Aware-Reconstruction-with-Principal-Network.

Keywords: Semi-supervised learning, Lesion detection, Pseudo-label calibration, Lesion-aware reconstruction, Teacher-Principal-Student framework

* Corresponding author

1. Introduction

Modern medical imaging commonly is utilized in the processes of disease detection and clinical diagnosis (YJ. et al., 2024; A. et al., 2009; M. et al., 2013). However, manual interpretation of these images is subject to the subjective factors and individual experience of the doctors, which can lead to variability in diagnosis (F. et al., 2016; Y. et al., 2024a). To overcome these limitations, deep learning-based computer-aided diagnosis (CAD) systems have been increasingly adopted in clinical practice due to their high accuracy and consistency (XY. et al., 2018; LL. et al., 2019; J. et al., 2016). These advanced systems offer a promising solution by providing more reliable and objective diagnostic support, thereby enhancing the overall quality of medical care. Nonetheless, current mainstream deep learning algorithms require a large amount of labeled data (Y. et al., 2024b; R. et al., 2021), which is costly to obtain in clinical medicine. For locating and diagnosing lesions, the ground truth must be validated by pathological diagnosis and manually annotated and reviewed by multiple experienced clinicians, making the process time-consuming and labor-intensive. Therefore, semi-supervised learning (SSL) for detecting lesions is an effective approach to alleviate the shortage of ground truth and to make better use of unlabeled data.

However, most mainstream SSL approaches rely on the generation and utilization of pseudo-labels (Z. et al., 2024b; X. et al., 2021). These methods have a significant drawback: errors in pseudo-labels can be amplified in subsequent operations, resulting in a large bias in the final network output. This issue arises from that the incorrect pseudo-labels are not effectively filtered out, especially when the network predicts an incorrect result with high confidence. Particularly in jaw lesion detection tasks, during the early stages of lesion development or at lesion boundary regions where pathological features are indistinct, the teacher network tends to generate incorrect predictions with high confidence values. These high-confidence errors are often not adequately screened, leading to successive optimization based on incorrect results during subsequent training of the student network. As a result, the performance of these networks becomes significantly limited. Over time, this accumulation of errors hinders the ability of network to learn accurately, ultimately compromising the reliability and effectiveness of the model.

To address the issue of inadequate filtering of incorrect predictions in SSL, this paper proposes a teacher-principal-student (TPS) framework. Based on the classic teacher-student (TS) model, we introduce one principal network that employs lesion-aware reconstruction to evaluate and filter the predictions of teacher network. Principal network objectively evaluates the quality of the predictions from teacher model and filters out poorly predicted results, ensuring that the pseudo-label data used for training student network is not influenced by the potential biases of teacher network. Specifically, the principal network conducts image reconstruction for the lesion areas localized by the teacher network, reconstructing these pathological regions into healthy tissue images. These altered images, along with the original images, are input into principal network for scoring based on contrastive learning. This model effectively measures the feature space disparity between the image pairs; a larger disparity indicates more precise location of lesion features, signifying more accurate predictions by teacher network. By filtering out erroneous predictions, our approach significantly improves diagnostic accuracy and reliability. Our main contributions are as follows:

- Propose an innovative principal network for robust pseudo-label quality assessment in semi-supervised lesion detection, with integrating lesion-aware reconstruction to enhancing reliability of pseudo-labels from teacher network, fundamentally mitigating prediction bias-induced adverse impacts on student network training.
- Develop one contextual attention-guided feature destruction algorithm, which adaptively disrupts lesion-related features while preserving normal anatomical integrity, laying a reliable foundation for pseudo-label quality evaluation.
- Our proposed TPS framework outperforms state-of-the-art semi-supervised approaches in clinical lesion detection tasks, achieves competitive performance comparable to fully supervised models and significantly reduces reliance on labeled clinical data.

2. Related Work

Wang used a double-uncertainty method for semi-supervised segmentation, enhancing accuracy in medical imaging with a teacher-student model (Y. et al., 2020). Huang compared semi- and self-supervised methods, finding MixMatch performs best under realistic conditions(Z. et al., 2024a). Studies (X. et al., 2023) have proposed a SSL framework for DME classification from OCT images, enhancing accuracy with self-correction and unlabeled data. Peng used a semi-supervised framework with adaptive threshold pseudo-labeling and contrastive loss to improve medical image classification(Z. et al., 2023). Qayyum(A. et al., 2023) proposed a semi-supervised method for dental caries detection using self-training to enhance accuracy with limited labeled data. Studies (Y., 2023) have shown that a semi-supervised deep learning method can be effective for tumor pathology image analysis. Wang introduced FocalMix, a SSL framework that improves 3D medical image detection by combining focal loss with MixUp augmentation to leverage unlabeled data(D. et al., 2020). Similarly, Zhou proposed SSMD, a semi-supervised medical image detection framework, improving detection with adaptive consistency and heterogeneous perturbation(HY. et al., 2021). Although some studies have begun to optimize the selection of pseudo-labels in semi-supervised training, the results remain unsatisfactory. Filipiak (D. et al., 2024) filtered noisy pseudo-labels by setting a single confidence threshold . Nozarian designed category-specific foreground IoU thresholds to classify the IoU relationship between pseudo-labels and student candidate boxes (F. et al., 2023). The aforementioned methods filter pseudo-labels by setting thresholds, but threshold setting is subjective and lacks specificity, failing to fundamentally eliminate the impact of teacher network errors on the student network. The Principal model proposed in this paper effectively controls the quality of the teacher network's pseudo-labels, thereby mitigating this issue.

3. Methods

3.1. Overview

Our proposed TPS model structure is illustrated in Figure 1. Teacher and student networks function as target detection networks, specifically for identifying and locating lesions. Teacher network generates predictions that serve as pseudo-labels for the fully supervised training of student network. For Student, two loss functions are adopted:

Classification Loss ($\mathcal{L}_{cls}$):

$$\mathcal{L}_{cls} = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{1}$$

where α_t denotes the class-balanced weight, p_t is the predicted probability of the true class, and γ is the focusing parameter to downweight easy negative samples.

Bounding Box Regression Loss ($\mathcal{L}_{reg}$):

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}0.5x_i^2 & |x_i| \le 1\\ |x_i| - 0.5 & |x_i| > 1\end{cases} \tag{2}$$

where N is the number of positive samples, and $x_i = t_i - \hat{t}_i$ represents the normalized difference between the ground-truth bounding box parameters (t_i: center coordinates, $\hat{t}_i$: the predicted parameters).

Principal network is innovatively introduced to evaluate the accuracy of these pseudo-labels to ensure quality of predictions of the teacher network. Inaccurate predictions receive lower quality scores and are filtered out as low-quality pseudo-label data. This process ensures that the data used to train student network is of high quality, which is crucial for improving the accuracy and reliability of the detection results.

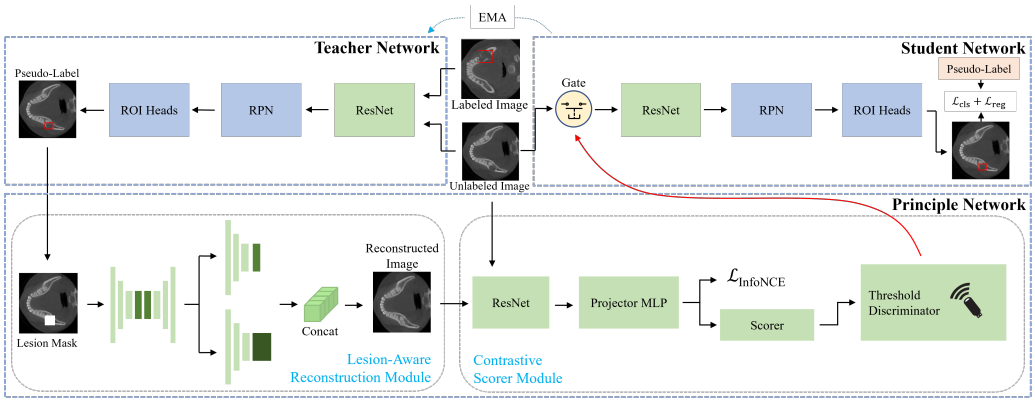

Figure 1: The Structure of TPS architecture. The pre-trained Teacher Network generates pseudo-labels, and the Student Network adopts these labels based on the reliability assessment results of the Principal Network (via lesion-aware reconstruction and contrastive scoring). L_{cls} is a classification loss and L_{reg} is a bounding box regression loss.

Principal network, which is the core innovation of this paper, comprises two main components: lesion-aware reconstruction module based on Contextual Attention (CAtt) (J. et al., 2018) and scoring module based on contrastive learning. The lesion-aware reconstruction module is tasked with disrupting the features of the lesion areas identified by teacher network and reconstruct the region into a healthy image. This change helps in evaluating how accurately teacher network has predicted the lesions. The scoring module then compares

the reconstructed image with the original image, using this comparison to assign a quality score to the predictions from teacher network. By scoring the pseudo-labels, principal network ensures that only high-quality labels are used for training student network. This mechanism significantly enhances the performance of the student network by reducing the impact of teacher network mistake.

3.2. Lesion-Aware reconstruction

Firstly, based on the lesion areas predicted by teacher network, a mask corresponding to the original image is generated. Ideally, this mask will cover all the lesion areas. The resulting masked image will be used as the input for the subsequent CAtt inpainting module. The purpose of this module is to fill the masked area based on contextual features, ensuring that contextual features of this filled area are consistent with non-lesion areas, thereby achieving our goal of disrupting features of lesion areas.

For lesion-aware reconstruction, the method adopts a generative image inpainting with contextual attention proposed in the study (J. et al., 2018). Specifically, the masked image first passes through coarse network to generate coarse result. This network is composed of series of dilated convolution layers, which increase the receptive field by inserting holes into the convolution kernels. Convolution layers are not effective at borrowing features from distant spatial locations, so contextual attention layers are introduced to identify where to obtain feature information from the background and to generate the missing area. First, convolution is used to compute the matching score of foreground patches with background patches (as convolutional filters). Then, softmax layer is applied to compare and obtain attention score for each pixel. Finally, foreground patches are reconstructed with background patches by performing deconvolution on the attention score. The contextual attention layer is differentiable and fully convolutional. This network structure effectively utilizes the contextual features of the background area to reconstruct the masked area, ensuring that the final image retains only the contextual features of the background area.

For the predictions of teacher network, if the lesion area is accurately located, the extracted contextual features after mask operation will only contain features of background area and not lesion features. Thus, reconstructed image will also not contain lesion features, only features of non-lesion area. If lesion area is not accurately located, the mask will not completely cover lesion area, and extracted contextual features will include both non-lesion area features and lesion area features. The final reconstructed image will not completely eliminate lesion features. In other words, the image reconstructed based on contextual features can restore lesion area without introducing new features. If the predictions from teacher network are ideal, the final reconstructed image will not contain any lesion features, achieving the goal of disrupting lesion features. Otherwise, the final reconstructed image will still contain lesion features, indicating incomplete reconstruction of lesion features.

Through lesion-aware reconstruction based on Catt module, we obtain original image and the lesion-aware reconstructed image, which form one pair of samples used as input for the subsequent scoring module. As shown in Figure 1, precisely reconstructed image no longer exhibit the typical appearance of lesions. Instead, they resemble the texture of the surrounding healthy tissue. This approach ensures the destruction of lesion features without introducing unrelated features, laying a solid foundation for subsequent quality scoring.

3.3. Contrastive Scorer

The scoring module is designed to evaluate the accuracy of the Teacher network's predictions. Its core principle states: when the Teacher network's predictions are accurate, the lesion-aware reconstruction network can completely reconstruct the lesion region features to lesion-aware, while minimizing damage to non-lesion region features. This change of original image and reconstructed image is quantified by the mapping distance in feature space - that is, the accuracy of lesion-aware reconstruction is positively correlated with the prediction accuracy of the Teacher network. Consequently, the evaluation metrics of lesion-aware reconstruction can serve as detection result assessment indicators, exhibiting properties similar to the Intersection over Union (IoU) metric.

To achieve this objective and obtain quantifiable metrics for lesion-aware reconstruction accuracy, this section designs a contrastive learning-based scoring network. The core concept involves bringing the representations of samples containing lesions closer together in the embedding space, while pushing them away from representations of non-lesion samples. The contrastive loss function used for model training is $\mathcal{L}_{\text{InfoNCE}}$ (Noise Contrastive Estimation):

$$\mathcal{L}_{InfoNCE} = -\log\left(\frac{\exp\left(\text{sim}(z_i, z_j)/\tau\right)}{\sum_{k=1}^{N}\exp\left(\text{sim}(z_i, z_k)/\tau\right)}\right) \tag{3}$$

where z_i and z_j are the representations of the positive sample pair, z_k is the representation of the negative sample pair, τ is the temperature parameter, and sim is the similarity function.

Training based on contrastive learning enables the network to effectively learn feature representations of lesions, thereby capturing the differences in feature space before and after image reconstruction. This reflects effectiveness of the reconstruction. To quantify this capability of distinguishing feature differences, linear layer and sigmoid operation is added to final layer of the contrastive network, converting the output into score that indicates the performance of reconstruction. Score closer to 0 suggests that the input image slice contains almost no lesion features, meaning lesion-aware reconstruction was highly effective. Conversely, score closer to 1 indicates that input image slice contains prominent lesion features.

Additionally, as discussed earlier, this output score can also reflect the performance of the detection network, exhibiting IoU-like properties. Leveraging this characteristic, the scoring network can effectively evaluate predictions of teacher network. In the absence of labeled data to compute IoU, this scoring metric provides an objective assessment of prediction quality. Consequently, it allows for selection of high-quality predictions as training samples for student network while mitigating impact of teacher network's bias in mandibular lesion detection.

4. Experiments and Results

4.1. Dataset

We validated effectiveness of principal network using a jaw lesion dataset. The dataset in this study comes from the Radiology Department of Peking University School and has received approval from the ethics review board (Approval Number: 2023-City Nature-43).

Pathological and radiological diagnoses were used as ground truth. The dataset was annotated and verified by one team comprising three radiologists with over five years of experience, one radiologist with over ten years of experience, and one senior radiologist. In total, we have collected three common types of jaw lesions: cystic lesions, solid lesions, and mixed lesions, some example images are shown in Figure 2.

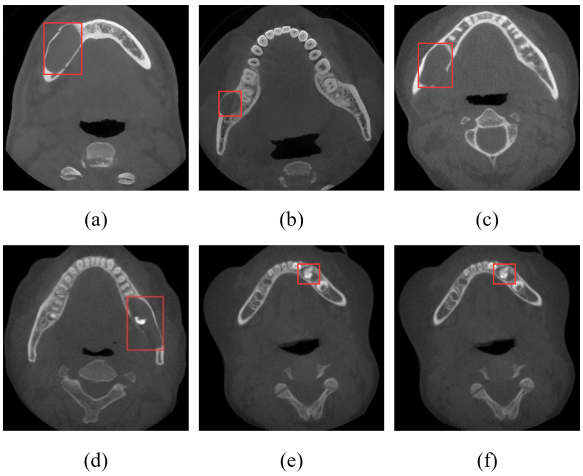

Figure 2: Examples of CBCT images, with lesion areas marked in red boxes, (a)-(d) represent cystic lesions, (e) represents a solid lesion, and (f) is a mixed lesion.

4.2. Experiments and Analysis of lesion-aware reconstruction Network

Healthy mandibular image slices were used as the training set for the lesion-aware reconstruction network to facilitate better learning of imaging features of healthy mandibles. Irregular masks were randomly generated for training purposes. During training process, total variation loss is used for both the coarse and fine generation stages.

$$\mathcal{L}_{TV} = \sum_{i,j} \left(|y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}| \right) \tag{4}$$

where y represents the normalized pixel value.

Additionally, we conducted tests on the trained lesion-aware reconstruction network to evaluate its performance in reconstructing lesion region features. Coarse reconstructed results and final reconstructed slices are shown in Figure 3. As can be observed, the lesion-aware reconstruction model based on contextual attention successfully removes the original lesion features while maximally preserving non-lesion features. By leveraging the characteristic information from healthy background mandibular images, the network generates new mandibular image slices that effectively eliminate pathological patterns.

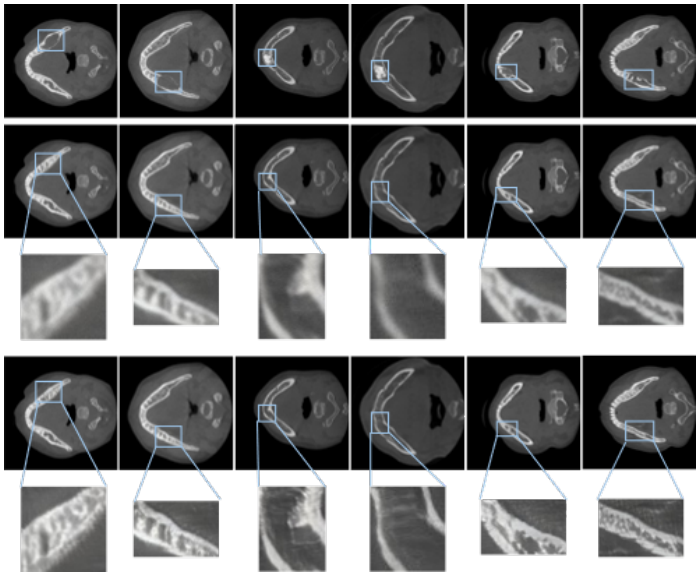

Figure 3: Feature-destroyed image slices: first row: original lesion images; second and third rows: coarse results and fine results from lesion-aware reconstruction; fourth and fifth rows: final lesion-aware reconstructed slices refined through contextual attention.

To better visualize these lesion features, this study focuses on examining distinction between lesion and reconstructed features in the feature space. For this purpose, we take pre-reconstruction and post-reconstruction images as paired samples and feed them into the feature extraction layer of the target detection network. Then employ four dimensionality reduction techniques to map the high-dimensional features into lower-dimensional spaces (2D and 3D) for visualization: 1) Principal Component Analysis (PCA): linear dimensionality reduction method that projects high-dimensional data onto lower-dimensional space by identifying directions of maximum variance (principal components). 2) t-Distributed Stochastic Neighbor Embedding (t-SNE): nonlinear approach that preserves local similarities between data points when mapping to lower dimensions. 3) Uniform Manifold Approximation and Projection (UMAP): manifold-based nonlinear technique that approximates high-dimensional manifold structure before projecting it into lower-dimensional space. 4) Spectral Embedding (SE): graph-theory-based method that performs dimensionality reduction through eigen decomposition of the graph Laplacian matrix. The final visualization results are shown in Figure 4. The analysis reveals that for the same lesion slice, the pre- and post-reconstruction images exhibit measurable separation in high-dimensional feature space - this separation distance corresponds to lesion features. Notably, more complete reconstruction of lesion features leads to greater separation distances, with the processed images retaining only non-lesion characteristics.

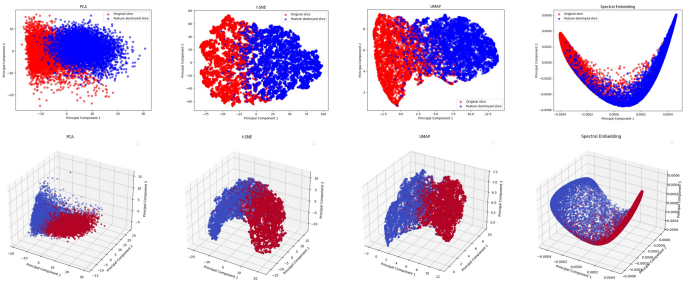

Figure 4: The first row shows 2D visualization images, and the second row shows 3D visualization images. Red represents the high-dimensional features of original images, blue represents the high-dimensional features of reconstruction images.

4.3. SSL Performance Validation and Analysis

Table 1: Performance with amount of labeled data unchanged and unlabeled data increasing

Metric	1:0 (Fully Supervised)	1:1	1:2	1:3	1:4
P	0.793	0.805	0.814	0.823	0.828
R	0.664	0.683	0.697	0.705	0.717
F1-score	0.723	0.739	0.751	0.759	0.769
mAP@0.5	0.694	0.725	0.742	0.758	0.767
mAP@[.5: .95]	0.506	0.539	0.578	0.604	0.627

This study conducted series of comparative experiments, with results presented as follows: Table 1 shows performance of student network when increasing amount of unlabeled data while keeping that of labeled data fixed. Experimental results demonstrate that, under a constant labeled data volume, student network's performance exhibits a continuous improvement trend as more unlabeled training samples are incorporated into the training set. Table 2 presents final performance of student network when progressively reducing proportion of labeled data while maintaining total annotation quantity (i.e., redistributing labels across fewer samples with higher density). These results indicate that even with reduced label ratios, student network maintains competitive performance as long as total annotation budget remains unchanged.

Table 2: Performance with total amount of data unchanged and proportion of unlabeled data increasing

Metric	1:0	1:1	1:2	1:3	1:4
P	0.887	0.873	0.866	0.842	0.828
R	0.811	0.765	0.741	0.730	0.717
F1-score	0.847	0.815	0.799	0.782	0.769
mAP@0.5	0.848	0.815	0.791	0.776	0.767
mAP@[.5:.95]	0.706	0.671	0.655	0.634	0.627

To further verify the effectiveness of the proposed SSL strategy in mandibular lesion detection, this study compared the performance of different semi-supervised strategies on the same dataset using the mAP@0.5 metric (as shown in Table 3). The experimental results demonstrate that compared to other methods, the semi-supervised training strategy proposed in this paper shows significant advantages under conditions of high unlabeled data ratios. Given that clinical medical mandibular data is characterized by scarce labeled samples but abundant unlabeled data, this finding fully confirms the clinical application value of the proposed algorithm.

Table 3: Performance comparison of different semi-supervised strategies in different ratios of quantities of labeled and unlabeled data

Method	1:4	1:3	1:2	1:1
STAC	0.693	0.715	0.743	0.758
Unbiased Teacher	0.725	0.741	0.752	0.764
Soft Teacher	0.712	0.733	0.756	0.784
Ours	0.767	0.776	0.791	0.815

5. Conclusion

Our TPS framework offers robust solution for semi-supervised lesions detection by addressing the critical issue of high-confidence erroneous pseudo-labels. Principal network, through its lesion-aware reconstruction and scoring mechanisms, ensures the quality of pseudo-labels, leading to more accurate and reliable detection outcomes. This model enhances the overall performance of SSL in medical image analysis, providing a valuable tool for clinical diagnosis. Future work may focus on refining the lesion-aware reconstruction and exploring its applicability to other medical imaging domains.

References

FLORES A., RYSAVY S., ENCISO R., et al. Non-invasive differential diagnosis of dental periapical lesions in cone-beam ct. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 566–569, 2009. doi: 10.1109/isbi.2009.5193110.

QAYYUM A., TAHIR A., BUTT MA., et al. Dental caries detection using a semi-supervised learning approach. *Scientific Reports*, 13(1):749, 2023. doi: 10.1038/s41598-023-27808-9.

FILIPIAK D., ZAPAŁA A., TEMPCZYK P., et al. Polite teacher: Semi-supervised instance segmentation with mutual learning and pseudo-label thresholding. *IEEE Access*, 12: 37744–37756, 2024. doi: 10.1109/access.2024.3374073.

WANG D., ZHANG Y., ZHANG K., et al. Focalmix: Semi-supervised learning for 3d medical image detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3951–3960, 2020. doi: 10.1109/cvpr42600.2020.00401.

ABDOLALI F., ZOROOFI RA., OTAKE Y., et al. Automated classification of maxillofacial cysts in cone beam ct images using contourlet transformation and spherical harmonics. *Computers in Biology and Medicine*, 72:108–119, 2016. doi: 10.1016/j.compbiomed.2016.03.014.

NOZARIAN F., AGARWAL S., REZAEIANARAN F., et al. Reliable student: Addressing noise in semi-supervised 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4981–4990, 2023. doi: 10.1109/cvprw59228.2023.00526. Accessed: 2025-12-03.

ZHOU HY., WANG C., LI H., et al. Ssmd: Semi-supervised medical image detection with adaptive consistency and heterogeneous perturbation. *Medical Image Analysis*, 72: 102117, 2021. doi: 10.1016/j.media.2021.102117.

LIU J., ZHANG SS., ZHENG SG., et al. Oral health status and oral health care model in china. *Chinese Journal of Dental Research*, 19(4):207–215, 2016. doi: 10.3290/j.cjdr.a37145.

YU J., LIN Z., YANG J., et al. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018. doi: 10.1109/cvpr.2018.00577. Accessed: 2025-12-09.

DA LL., XIE YF., and RONG S. Statistical analysis of current oral health care and dental education resources in china. *Chinese Journal of Dental Research*, 22(1):37–43, 2019. doi: 10.3290/j.cjdr.a41773.

SHWEEL M., AMER MIK., and EL-SHAMANHORY AF. A comparative study of cone-beam ct and multidetector ct in the preoperative assessment of odontogenic cysts and tumors. *The Egyptian Journal of Radiology and Nuclear Medicine*, 44(1):23–32, 2013. doi: 10.1016/j.ejrnm.2012.12.002.

SINGH R., BHARTI V., PUROHIT V., et al. Metamed: Few-shot medical image classification using gradient-based meta-learning. *Pattern Recognition*, 120:108111, 2021. doi: 10.1016/j.patcog.2021.108111.

WANG X., CHEN H., XIANG H., et al. Deep virtual adversarial self-training with consistency regularization for semi-supervised medical image classification. *Medical Image Analysis*, 70:102010, 2021. doi: 10.1016/j.media.2021.102010.

WANG X., TANG F., CHEN H., et al. Deep semi-supervised multiple instance learning with self-correction for dme classification from oct images. *Medical Image Analysis*, 83: 102673, 2023. doi: 10.1016/j.media.2022.102673.

SUN XY., CHAO Y., WANG XZ., et al. Report of the national investigation of resources for oral health in china. *Chinese Journal of Dental Research*, 21(4):285–297, 2018. doi: 10.3290/j.cjdr.a41087.

GUO Y., GUO J., LI Y., et al. Rapid detection of non-normal teeth on dental x-ray images using improved mask r-cnn with attention mechanism. *International Journal of Computer Assisted Radiology and Surgery*, 19(4):779–790, 2024a. doi: 10.1007/s11548-023-03047-1.

JIANG Y. Research on tumor histopathological image analysis method based on deep semi-supervised learning. *Frontiers in Oncology*, 12, 2023. doi: 10.3389/fonc.2022.1044026.

WANG Y., ZHANG Y., TIAN J., et al. *Double-Uncertainty Weighted Method for Semi-Supervised Learning*, volume 23, pages 542–551. Springer International Publishing, 2020. doi: 10.1007/978-3-030-59710-8_53.

WANG Y., CHEN HJ., LI JP., et al. Adaptive cross-view feature mining for jaw lesions detection and recognition in cbct image. In *2024 IEEE EMBS International Conference on Biomedical and Health Informatics*, pages 1–7, 2024b. doi: 10.1109/bhi62660.2024.10913593.

SHI YJ., LI JP., WANG Y., et al. Deep learning in the diagnosis for cystic lesions of the jaws: A review of recent progress. *Dentomaxillofacial Radiology*, 53(5):271–280, 2024. doi: 10.1093/dmfr/twae022.

HUANG Z., JIANG R., AERON S., et al. Systematic comparison of semi-supervised and self-supervised learning for medical image classification. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22282–22293, 2024a. doi: 10.1109/cvpr52733.2024.02103.

PENG Z., TIAN S., YU L., et al. Semi-supervised medical image classification with adaptive threshold pseudo-labeling and unreliable sample contrastive loss. *Biomedical Signal Processing and Control*, 79:104142, 2023. doi: 10.1016/j.bspc.2022.104142.

REN Z., KONG X., ZHANG Y., et al. Ukssl: Underlying knowledge based semi-supervised learning for medical image classification. *IEEE Open Journal of Engineering in Medicine and Biology*, 5:459–466, 2024b. doi: 10.1109/ojemb.2023.3305190.