WEAKLY-SUPERVISED LUNG NODULE SEGMENTATION IN CT WITH ONE-CLICK ANNOTATION

Tianyi Qu, Dichao Liu†, Ze Jin, Yuqiao Yang, Kenji Suzuki

Department of Information and Communications Engineering, Institute of Science Tokyo, Japan

ABSTRACT

Deep Learning (DL) has become a predominant choice for medical image segmentation. However, the requirement for a large volume of precisely-annotated manual segmentation data poses a significant barrier. This study proposes a weakly-supervised approach to mitigating manual annotation labor in lung nodule segmentation in CT. Unlike conventional segmentation models that rely on manual pixel-wise annotations, our approach iteratively refines pseudo labels for 3D semantic segmentation generated from "one-click" input through function maps derived from an Explainable AI (XAI) method. The iterative retraining process using pseudo labels resulted in the performance (Dice Score = 78.83%, HD_{95} = 2.178) comparable to the "gold-standard" fully-supervised model trained with manual pixel-wise annotations (Dice Score = 79.26%, HD_{95} = 2.335). Additionally, we compared our approach with state-of-the-art weakly supervised methods. Experimental results demonstrated that our model achieved better performance while significantly reducing both parameter count and inference time. These findings highlight the effectiveness of our method for pulmonary nodule segmentation in 3D CT volumes.

Index Terms— Weakly supervised learning, lung nodule segmentation, explainable AI (XAI), function maps

1. INTRODUCTION

Semantic segmentation is a prominent topic of image processing and computer vision, with applications extending across diverse fields such as scene understanding, robotic perception, and medical image analysis. The primary objective of 3D semantic segmentation is to assign an object class label to each voxel in an image. Presently, Deep learning (DL) architectures such as DeepLab V3+ [1] and nnUNet [2] have significantly advanced this field.

Despite their remarkable performances, DL-based segmentation models typically necessitate large-scale, pixel-wise annotated datasets. While annotation acquisition is relatively manageable in natural image processing, acquiring such detailed labels with high precision and accuracy is exceptionally challenging. This challenge is particularly evident in computer-aided diagnosis (CAD), such as medical image segmentation tasks. Annotating medical images for segmentation requires expert knowledge, which substantially escalates the cost of data labeling. Furthermore, inter-observer variability introduces subjectivity, affecting annotation consistency.

Reducing annotation burden while maintaining high segmentation accuracy is crucial for clinical deployment. Weakly-supervised learning has therefore emerged as a promising alternative. Researchers strive to introduce various strategies for weak supervision, such as sparse annotation [3]-[4], which depends on a limited number of accurate annotations of 2D slices, and active learning [5]-[6], which allows the model to be trained with only a small set of representative samples alongside a large pool of unlabeled data.

In this study, we propose a 3D weakly-supervised lung nodule segmentation framework based on massive-training artificial neural networks (MTANNs) with iteratively refining pseudo labels. Instead of using precisely annotated manual segmentation data, we initialize training with coarsely labeled images obtained from a "one-click" annotation. Throughout the iterative process, function maps generated by an Explainable AI (XAI) technique [7] serve as refined supervision signals. Our approach eliminates the need for labor-intensive manual labeling, requiring only a "one-click" indication to achieve effective segmentation while maintaining an extremely low parameter count and inference time. The proposed method eliminates the need for the cumbersome and labor-intensive process of manual annotation in a 3D lung nodule segmentation task.

2. METHODS AND MATERIALS

2.1. Lung CT database with nodules

The Lung Nodule Analysis 2016 (LUNA16) dataset was used in this study, which contains 888 thoracic CT scans with 1,186 nodule candidates and numerous non-nodule candidates. Pre-processing involved cropping the volumes of interest (VOIs) around nodule candidates. The dimension of the VOI image was set to 49×49×49. Furthermore, an auto-generated lung mask was applied to the VOI to eliminate the regions outside the lung.

2.2. Massive-training artificial neural network (MTANN) deep learning

†Corresponding author.

In the field of image processing, Suzuki et al. proposed supervised nonlinear filters and edge enhancers based on neural networks (NNs), referred to as neural filters and neural edge enhancers [8], respectively. Suzuki et al. extended neural filters and edge enhancers to invent an MTANN deep learning model to accommodate pattern recognition and classification tasks [9].

The input of a 3D MTANN is voxel values in a 3D patch extracted from the original input image, and the output is a single voxel value, which corresponds to the center voxel of the input patch and is located at the same position in the teaching images. Teaching images are the desired output images. A large number of patches are acquired prior to training. Employing this patch-based training methodology allows the MTANN to be trained super-efficiently, even with a limited number of input images [10].

2.3. Explainable AI (XAI) approaches with MTANN

Our weakly supervised approach is inspired by the investigation of an XAI approach [7] for a fully-supervised MTANN in distinguishing between lung nodules and non-nodule structures. Fig. 1 illustrates an overall schematic diagram of the XAI approach for the MTANN, which contains the following steps: 1) training a large MTANN model, 2) pruning the model with a sensitivity-based structure-based optimization, and 3) visualizing function maps and clustering them by an unsupervised hierarchy clustering algorithm.

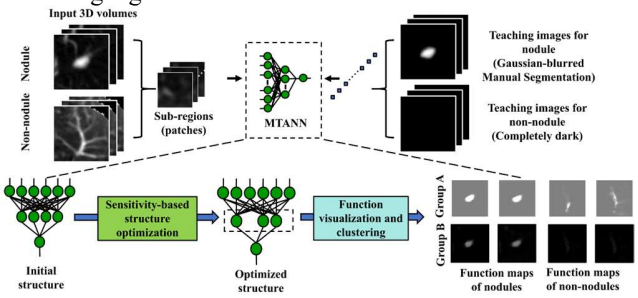


Fig. 1. Overall schematic diagram of the XAI approach for a fully-supervised MTANN model to provide function maps in classification between lung nodules and non-nodules in CT.

The function maps for the retained hidden units are categorized via unsupervised hierarchical clustering, leveraging a composite loss function that integrates mean square error (MSE) and structural similarity (SSIM). The units that are clustered together into different groups (e.g., Group A and B in Fig. 1) exhibit comparable functional characteristics. A grayscale-coding scheme is utilized in the function maps: lighter region highlights areas accentuated by the hidden unit, medium gray indicates unchanged areas, and dark shade marks suppression areas, as shown in Fig. 1. In our experiment, in the case of a nodule image, Group A was enhancing the nodule region, or namely lesion enhancement while not influencing other regions. On the other hand, Group B was suppressing most non-nodule regions while keeping nodule regions unchanged, namely non-lesion suppression, which resulted in a gray region in the function map of Group B.

2.4. Weakly-supervised training with iterative strategy

To address the challenge of heavy reliance on precise annotations for training, we propose a weakly-supervised methodology that does not require manual segmentation.

Motivation: In a fully-supervised MTANN, when precise manual lung nodule segmentation masks were employed as teaching images, one group of the function maps in the model demonstrated the ability to highlight the nodule region accurately (Group A in Fig. 1). To investigate how the function maps adapt under weak supervision, we replaced the manual segmentation teaching images with alternative teaching images that were independent of manual annotations. Most lung nodules exhibit approximately spherical or ellipsoidal shapes, with pixel intensity gradually decreasing from the nodule center toward the periphery. To simulate this characteristic, we substituted the manual segmentation teaching images with a "one-click" generated Gaussian distribution image (σ via LUNA16 stats) near nodule center to mimic the shape and size of a nodule region. Remarkably, function maps from the weakly supervised model closely resembled those from the fully supervised model. Compared to Gaussian distribution teaching images, certain function maps that highlighted the nodule region provided more precise information about the nodule, leading to reduced errors. Consequently, we utilized these informative function maps as refined instructional images to retrain the model, thereby improving its performance iteratively.

Fig. 2 depicts the framework of our proposed weakly-supervised training strategy for the segmentation task of lung nodules with MTANN. Leveraging the Explainable AI (XAI) capabilities of the MTANN, the original rough Gaussian distribution teaching image, as an initial pseudo label, crafted based on basic prior knowledge, is refined through the application of effective function maps generated by the previous iteration of the model.

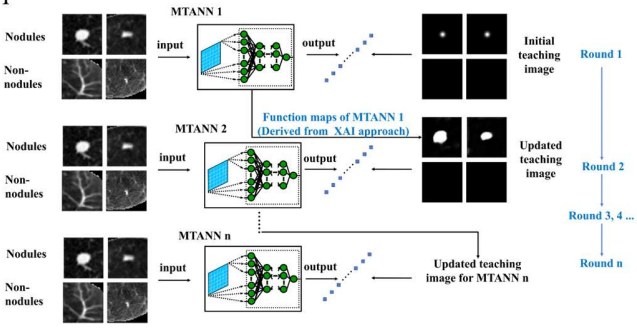


Fig. 2. Comprehensive framework of weakly-supervised training strategy for 3D semantic segmentation of lung nodules in CT with MTANN: iterative training from an initial inexact label from "one-click" input with a teaching-image

refinement strategy coupled with function map generation by our XAI approach.

The iterative process persists until outputs on training cases stabilize (paired *t*-test on prediction maps, $p > 0.05$ between consecutive iterations), which indicates the completion of the refinement process. At this stage, the inexact label becomes very similar to manual segmentation. Non-nodule input images are also included in training to preserve the function maps that effectively highlight the nodule region within the classification task.

3. EXPERIMENT AND RESULTS

A total of 100 nodule and 100 non-nodule 3D volumes were selected from distinct patients of the LUNA16 dataset for training and another 100 nodule images with "gold-standard" manual segmentation were selected for testing. For each image, 100 patches were extracted, resulting in a total training set of 20,000 patches. The training of the MTANN employed the mean absolute error (MAE) as the loss function, with a focus on minimizing the discrepancy between the predicted output and the actual result value.

The original trained large MTANN was designed with 343 input units, matching a 3D kernel of size 7×7×7 and 50 hidden units. In the pruning step with the XAI approach [7] , the mean absolute error (MAE) was monitored through sensitivity-based structure optimization to evaluate whether the compact model could perform equivalently to the original large model on the training images. A notable turning point was detected when the number of hidden units was reduced to fewer than eight. We selected a compact model configuration comprising eight hidden units for subsequent training, capitalizing on its effective functional mapping during the iterative training process.

From the second round of iterative training onwards, we employed the effective function maps derived from the previously trained MTANN model as updated teaching images to further train new MTANN models. This iterative process persisted, continually refining the model by incorporating these refined function maps as new teaching images until there was no further segmentation improvement.

In this study, we carried out experiments and observed that performance peaked during the third iteration. To scientifically validate convergence, we manually proceeded with two additional iterations beyond the peak. For ease of reference, these weakly supervised models (Ws-MTANN) have been named according to their iteration index. Additionally, a fully supervised model (Fs-MTANN) was trained with manual annotations for comparison. Quantitative evaluation was performed on entire 3D volumes, while segmentation results were visualized as 2D central slices.

Fig. 3 compares updated teaching images during the initial third iterations. As the iteration continues, the teaching images refined and became much closer in size and shape compared with actual nodules in input images. Alternative refinement strategies (previous model's output, CAM-based methods) failed to achieve satisfying iterative results.

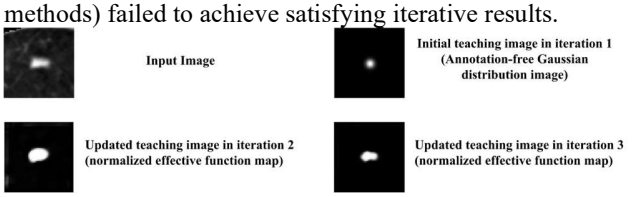


Fig. 3. Comparison between input image, initial and updating teaching images in the initial third iterations.

As shown in Table 1, the weakly-supervised model achieved a significant improvement of 7.28% in Dice score ($p < 0.01$) by the end of the iterative process, closely approaching the fully supervised model's Dice score without statistical significance ($p > 0.05$). The marginal but consistent performance decline observed in Ws-MTANN 4 and Ws-MTANN 5 further validated our stopping criterion.

Table 1. Quantitative comparison of weakly-supervised models (Ws-MTANNs), starting from "one-click" annotation, generated in the iterative process with the "gold-standard" fully-supervised model (Fs-MTANN).

Model name	Dice score (%) ↑	HD_{95} ↓
Ws-MTANN 1	71.55	2.987
Ws-MTANN 2	75.30	2.435
Ws-MTANN 3	**78.83**	**2.178**
Ws-MTANN 4	78.57	2.196
Ws-MTANN 5	78.04	2.213
Fs-MTANN	**79.26**	**2.335**

Fig. 4 showcases the segmentation comparison of the models in the initial third iterations. The segmentation gradually improved during the refined teaching images in an iterative training process. The model Ws-MTANN 3, after three iterations, had the best segmentation performance.

Fig. 5 represents the segmentation comparison of the weakly supervised model Ws-MTANN 3 (yellow curve), the fully supervised model Fs-MTANN (green curve), and the ground truth (red curve). The weakly supervised model demonstrated segmentation results that were not only very close to the ground truth but, in some instances, even surpassed the performance of the fully supervised model. This is particularly notable because the model Ws-MTANN 3 was trained without extensive manual segmentation, which is typically a high-cost and labor-intensive process.

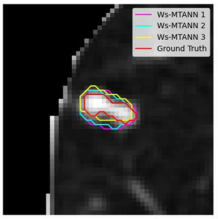


Fig. 4. Progress of segmentation results of the models during the initial third iterations (purple, blue, yellow curves) and the ground truth (red curve).

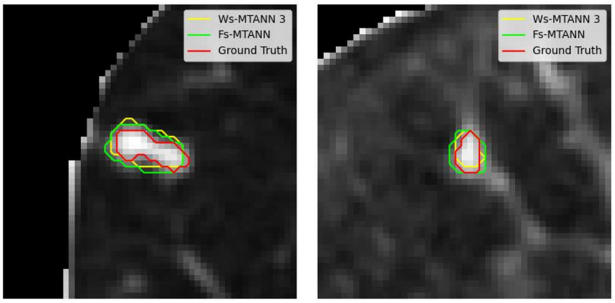


Fig. 5. Comparison of segmentation results of weakly (yellow curve), fully supervised (green curve) models, and the ground truth (red curve).

Table 2 compares our method with recent weakly-supervised segmentation approaches on the selected dataset, evaluating them in terms of Dice score, parameter count (million), and inference time (ms) per image. Fig. 6 visualizes the segmentation results of these models. Our proposed weakly-supervised approach not only achieved a segmentation contour that better matched the ground truth but also accomplished this with an extremely lightweight parameter set and rapid inference speed, demonstrating a strong alignment with the demands of clinical applications.

Table 2. Comparison with recent weakly-supervised segmentation algorithms.

Model name	Dice score (%) ↑	Parameters ↓	Time ↓
S2C [11]	66.07	106.121	10.692
CTI [12]	75.39	24.705	15.746
Ws-MTANN	**78.83**	**0.003**	**0.526**

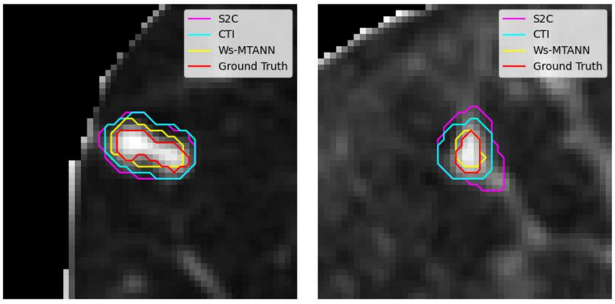


Fig. 6. Segmentation result visualization of different weakly-supervised algorithms.

4. CONCLUSIONS

In this study, we proposed a clinically viable weakly-supervised approach to practically eliminate manual annotation in semantic segmentation of lung nodules in 3D CT volumes. We started training an initial MTANN model with an inexact Gaussian teaching image obtained from "one-click" input on the nodule. The iterative process gradually improved imperfect teaching images with effective function maps obtained from the previous model by using our XAI approach. The results indicated that our weakly-supervised model not only demonstrated the performance statistically on par with the "gold-standard" fully-supervised model, but also possessed a lightweight architecture and rapid inference, thus obviating the need for labor-intensive annotations through manual segmentation in clinical scenarios.

5. ACKNOWLEDGEMENT

This work was supported in part by a NEDO Program Grant.

6. COMPLIANCE WITH ETHICAL STANDARDS

Ethical approval was not required because of the retrospective use of the open-source data.

7. REFERENCES

[1] L.-C. Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *ECCV*, 2018, pp. 833–851.
[2] F. Isensee et al., "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation," *arXiv preprint arXiv:1809.10486*, 2018.
[3] H. Cai et al., "3D Medical Image Segmentation with Sparse Annotation via Cross-Teaching Between 3D and 2D Networks," in *MICCAI*, 2023, pp. 614–624.
[4] F. Gao et al., "Segmentation only uses sparse annotations: Unified weakly and semi-supervised learning in medical images," *Medical Image Analysis*, vol. 80, p. 102515, 2022.
[5] F. Bai et al., "Discrepancy-Based Active Learning for Weakly Supervised Bleeding Segmentation in Wireless Capsule Endoscopy Images," in *MICCAI*, 2022, pp. 24–34.
[6] H. Liu et al., "COLosSAL: A Benchmark for Cold-Start Active Learning for 3D Medical Image Segmentation," in *MICCAI*, 2023, pp. 25–34.
[7] Z. Jin et al., "Explaining Massive-Training Artificial Neural Networks in Medical Image Analysis Task Through Visualizing Functions Within the Models," in *MICCAI*, 2023, pp. 713–722.
[8] K. Suzuki et al., "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 25, no. 12, pp. 1582–1596, 2003.
[9] K. Suzuki et al., "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Medical Physics*, vol. 30, no. 7, pp. 1602–1617, 2003.
[10] N. Tajbakhsh and K. Suzuki, "Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs," *Pattern Recognition*, vol. 63, pp. 476–486, 2017.
[11] H. Kweon et al., "From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation," in *CVPR*, 2024, pp. 19499–19509.
[12] S. -H. Yoon et al., "Class Tokens Infusion for Weakly Supervised Semantic Segmentation," in *CVPR*, 2024, pp. 3595–3605.