SinusNet: Label-Free Segmentation of Maxillary Sinus Lesion in CBCT Images

Da-El Kim[1*] dkim3@snu.ac.kr
Su Yang[1*] s8431@snu.ac.kr
SeRyong Kang[1] seryongkang@snu.ac.kr
Jin Kim[1] kimjin116@snu.ac.kr
SoYoung Chun[1] soyljh@snu.ac.kr
MinHyuk Choi[1] cmh9101@snu.ac.kr
Won-Jin Yi[1] wjyi@snu.ac.kr
[1] *Seoul National University, Seoul, South Korea.*

Editors: Under Review for MIDL 2022

Abstract

To alleviate the workload of data annotation, we propose a label-free approach (SinusNet) for the segmentation of maxillary sinus lesions in CBCT images via learning the anomaly features from diverse synthetic lesions within the normal maxillary sinus. Our SinusNet achieved average F1 of $80.9 \pm 11.6\%$, precision of $82.7 \pm 9.1\%$, and recall of $80.1 \pm 15.0\%$, respectively, and comparable performance with those of previous supervised approaches.

Keywords: Deep learning, maxillary sinus lesion, label-free segmentation.

1. Introduction

Segmentation and volume measurement of the maxillary sinus (MS) and maxillary sinus lesion (MSL) in cone-beam CT (CBCT) can provide an assessment for the diagnosis of sinusitis and help with the preoperative planning of maxillary implants and bone grafts. However, the manual segmentation of MS and MSL is complex and time-consuming because the MS and MSL are adjacent to other important anatomical structures and are sometimes affected by metal artifacts. Also, the MSL is observed with high variations in texture, size, and position within the MS region. Recently, deep learning-based methods (Jung et al., 2021; Hung et al., 2022) have been proposed for successful MS and MSL segmentation in CBCT images. Despite such success, they all rely on the large-scale well-labeled dataset to model the diversities of the MS and MSL in CBCT images.

In this paper, we propose a label-free approach for multi-label segmentation of MS, maxillary air region (MA), and MSL in CBCT images, requiring no lesion annotation. Our contributions are follows: 1) We propose a label-free approach (SinusNet) for MSL segmentation to alleviate the workload of data annotation in CBCT images. 2) Motivated by the observation that the MSL exists in the MS with various patterns of high-intensity ranges, we synthesize lesion masks considering the shape and texture features of MSL using morphological operations and insert these into the normal MS to generate training pairs. 3) To the best of our knowledge, we are the first to attempt a deep learning-based label-free segmentation of MSL in CBCT images.

[*] Contributed equally

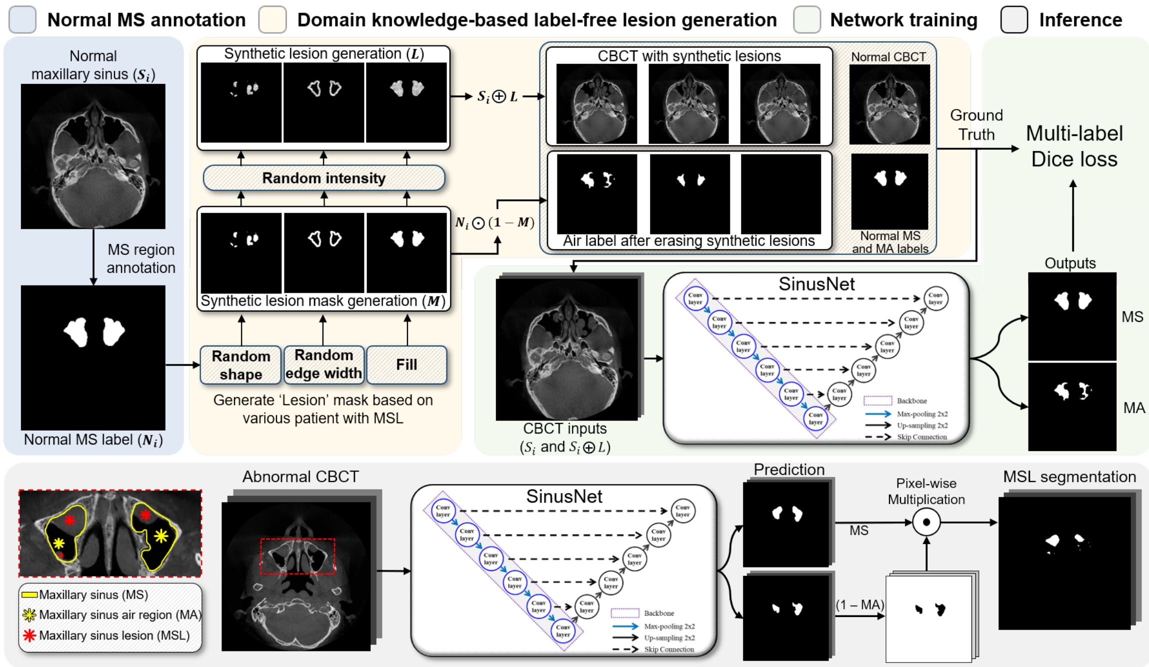

Figure 1: Overview of proposed SinusNet framework. We insert synthetic lesions in the normal MS and train SinusNet to segment the MS and MA. In the inference, the SinusNet first segment the MS and MA, then the segmentation results of MSL are generated by pixel-wise multiplication of the predicted MS and inverse MA.

2. Methods

Data description. We built the CBCT dataset acquired 38 patients, consisting of 18 and 6 normal MS for training and validation sets, respectively, 14 patients with lesions for test sets. The normal ground truth was manually annotated by one radiologist. This manual task took approximately 2 hours per CBCT scan to complete. The 640×640 size of CBCT images was resized to 256×256 and normalized according to min-max normalization.

Synthetic lesion generation. We used three synthetic lesion generation methods based on our domain knowledge to represent various shapes of the MSL (Figure 1). First, we generated a curve line with random width (3-5 pixel diameter) along the inside of the MS boundary, starting from the edge of the MS in most cases. Second, we generated a random number of shapes (ellipses) with various sizes and orientations into the normal MS region. In our domain knowledge, we observed that the MSL is mostly shown as ellipse shapes in the axial plane of CBCT images. For the last method, we filled the whole air region of the MS to represent patients filled with the MSL. After generating different shape types to represent synthetic MSL, we apply pixel noise to closely resemble the real MSL's intensity range (30-80 pixel values). The optimal range of pixel values was determined on an empirical basis. We obtained air labels for our synthetic lesions by simple pixel-wise multiplication of inverse synthetic lesion masks and paired normal air labels. CBCT inputs were generated by pixel-wise addition with synthetic lesions to the MS region in normal CBCT images.

Network architecture. The network architecture was motivated by the U-Net consisting of an encoder and decoder. We used three backbones of ResNet101, DenseNet121, and EfficientNetB4 with initial weights of ImageNet for transfer learning. The *Sigmoid* function was used as the output activation for multi-label segmentation of the MS and MA.

Multi-label segmentation loss. For multi-label segmentation, we used soft Dice loss (SDL) for both the MS and MA segmentation. The SDL is defined as: $SDL(G,P) = 1 - \frac{1}{K}\sum_{j=0}^{K}(2\sum_i^n G_i^j P_i^j)/(\sum_i^n G_i^j + \sum_i^n P_i^j)$, where n is number of pixels, K is the number of classes (background, MA, MS), G is the ground truth, and P is the prediction result.
Training setup. We implemented the deep networks on Tensorflow with NVIDIA GeForce GTX 1080Ti. For training, Adam optimizer was used with an initial learning rate of 10^{-3}. Deep networks were trained for 200 epochs with a batch size of 16. We adopt three data augmentations, including rotation ($\pm 10°$), brightness ($\pm 20\%$), Gaussian noise (0.05).

3. Experiments and Discussion

Impact of Label-free approach. To measure the performance of deep networks, we used the F1-score (F1), precision (PR), and recall (RC). In our experiments, ResNet101 achieved the highest performance for the MSL segmentation (Figure 2). As shown in Table 1, experimental results show that the segmentation performance is further improved by training with various morphological and texture features of synthetic lesions generated by our label-free methods. Our SinusNet required about 0.02 seconds per image for inference.
Comparison with supervised approaches. For MSL segmentation in CBCT images, our label-free approach shows comparable performance with those of reported supervised approaches (Jung et al., 2021; Hung et al., 2022) as shown in Figure 2a. Furthermore, we not only use a smaller CBCT dataset than theirs but also do not need any true lesion annotations for training. Our SinusNet may provide a feasible way to be relieved from the labor-intensive manual annotation and allow an automated lesion segmentation in real-time.

Table 1: Ablation study for synthetic lesion generation methods in ResNet101 backbone.

Synthetic method			MS			MA			MSL		
Edge	Shape	Fill	F1	PR	RC	F1	PR	RC	F1	PR	RC
✓	-	-	97.0±1.8	96.7±1.8	97.4±2.4	79.3±18.0	97.5±1.4	70.3±23.8	41.4±18.0	86.1±6.3	29.6±17.0
-	✓	-	97.1±1.4	96.4±2.1	97.9±1.3	87.3±10.1	97.8±1.2	80.2±15.6	61.5±13.5	89.4±8.2	48.4±15.9
-	-	✓	96.4±2.9	97.2±1.6	95.7±4.6	88.6±8.2	97.8±1.3	81.9±12.9	60.2±21.1	86.9±13.8	48.5±21.1
✓	✓	✓	96.6±1.7	95.8±2.4	97.4±2.1	95.0±2.2	94.7±3.5	95.4±2.9	80.9±11.6	82.7±9.1	80.1±15.0

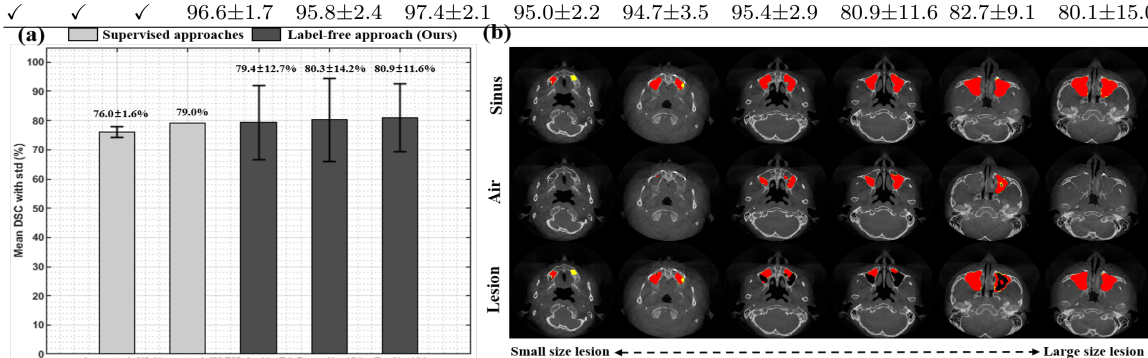

Figure 2: (a) F1 comparison with previous supervised approaches. (b) Comparison of ground truth (yellow) and predicted results (red) from our SinusNet.

Acknowledgments

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health Welfare, the Ministry of Food and Drug Safety) (1711137883, KMDF_PR_20200901_0011, 1711138289, KMDF_PR_20200901_0147).

References

Kuo Feng Hung et al. Automatic detection and segmentation of morphological changes of the maxillary sinus mucosa on cone-beam computed tomography images using a three-dimensional convolutional neural network. *Clinical Oral Investigations*, pages 1–12, 2022.

Seok-Ki Jung et al. Deep active learning for automatic segmentation of maxillary sinus lesions using a convolutional neural network. *Diagnostics*, 11(4):688, 2021.