IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 June 2023, accepted 12 July 2023, date of publication 20 July 2023, date of current version 31 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297506

RESEARCH ARTICLE

Interactive Segmentation for COVID-19 Infection Quantification on Longitudinal CT Scans

MICHELLE XIAO-LIN FOO[1], SEONG TAE KIM[2], (Member, IEEE), MAGDALINI PASCHALI[3], LEILI GOLI[4], EGON BURIAN[5], MARCUS MAKOWSKI[5], RICKMER BRAREN[5], NASSIR NAVAB[6], (Fellow, IEEE), AND THOMAS WENDLER[6]

[1]Department of Informatics, Ludwig Maximilian University of Munich, 80539 Munich, Germany
[2]Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea
[3]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada
[5]Institute for Diagnostic and Interventional Neuroradiology, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany
[6]Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Garching, Germany

Corresponding author: Seong Tae Kim (st.kim@khu.ac.kr)

This work was supported in part by the Bavarian Research Foundation (BFS) under Grant AZ-1429-20; in part by the Bavarian State Ministry for Economics, Development and Energy (BayStMW), under Grant DIK 0127/02; in part by the Korean Government (MSIT) through the Institute of Information and Communications Technology Planning and Evaluation (IITP) under Grant 2021-0-02068; and in part by the National Research Foundation of Korea (NRF) under Grant 2021R1G1A1094990.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Technical University of Munich under Approval No. 111/20 S-KH.

ABSTRACT Consistent segmentation of CT scans in COVID-19 patients across multiple time points is important to accurately evaluate disease progression and therapeutic response. In medical domains, previous interactive segmentation studies have been mainly conducted on data from a single time point. However, the valuable segmentation information from previous time points is often underutilized in assisting the segmentation of a patient's follow-up scans. Moreover, fully automatic segmentation techniques frequently produce results that would need further refinement for clinical applicability. In this study, we propose a novel single-network model for interactive segmentation that fully leverages all available past information to refine the segmentation of follow-up scans. In the first segmentation round, our model takes concatenated slices of 3D volumes from two-time points (target and reference), employing the segmentation results from the reference time point as a guide for segmenting the target scan. Subsequent refinement rounds incorporate user feedback in the form of scribbles that rectify the segmentation, in addition to incorporating the previous segmentation results of the target scan. This iterative process ensures the preservation of segmentation information from prior refinement rounds. Experimental results obtained from our in-house multiclass longitudinal COVID-19 dataset demonstrate the effectiveness of the proposed method compared to its static counterpart, thus providing valuable assistance in localizing COVID-19 infections in patients' follow-up scans.

INDEX TERMS Interactive segmentation, longitudinal segmentation, COVID-19 quantification.

I. INTRODUCTION

In December 2019, the first cases of a new coronavirus disease, COVID-19, a severe acute respiratory illness, emerged in Wuhan, China [1]. This highly infectious respiratory virus rapidly spread worldwide and threw the world into a global pandemic. As per the data provided by the COVID-19 monitoring site at Johns Hopkins University,[1] as of March 10th, 2023, the number of confirmed infections has exceeded 676 million, with a staggering 6.88 million fatalities resulting from complications arising from the virus.

During the initial outbreak of COVID-19, there was an urgent need for expeditious annotation of medical scans to gain deeper insights into the disease. Many researchers

The associate editor coordinating the review of this manuscript and approving it for publication was Gyorgy Eigner.

[1]https://coronavirus.jhu.edu/map.html

addressed this demand by employing deep learning-based methodologies [2], [3], [4]. Computed tomography (CT) scans provide crucial diagnostic information in the assessment and treatment of COVID-19 patients [5], [6]. However, medical reports indicate that the imaging characteristics of COVID-19 exhibit heterogeneity among patients, and the evolution of radiological patterns throughout the disease progression is inconsistent [1], [7], [8]. The subtle anatomical boundaries, as well as the variations in imaging devices and protocols, pose a challenge for automatic segmentation techniques.

Human interactions coupled with deep learning models can offer a way to overcome the challenges faced by automatic segmentation models and improve segmentation results as shown by [9], [10], and [11]. However, previous studies on interactive segmentation have primarily focused on utilizing single time point data for segmentation purpose. The potential benefit of incorporating readily available segmentation information from previous time points to facilitate the segmentation of follow-up scans for individual patients remains unexplored.

To address the limitations of static automatic segmentation models, we propose an interactive segmentation method that segments COVID-19 infection on longitudinal CT scans. The proposed method aims to harness longitudinal information and user feedback to improve the quality of segmentation results. The main contributions of this paper can be summarized as follows:

- We propose a new segmentation approach that utilizes information from previous time point, past segmentation refinement rounds, and user feedback to improve the quality of segmentation, which can be used for infection quantification. To the best of our knowledge, this work represents the pioneering effort in developing an interactive segmentation technique specifically tailored for longitudinal CT scans. Our method is also the first interactive segmentation of COVID-19 infection using longitudinal CT scans.
- Our method offers a way to extend existing static models for longitudinal interactive segmentation, requiring minimal additional effort. By incorporating our approach into the existing framework, the capabilities of static models can be enhanced, which effectively adapts them for longitudinal interactive segmentation tasks.
- To demonstrate the enhanced performance of our longitudinal interactive segmentation model in comparison to a static interactive segmentation model, we conducted a comprehensive study on our in-house longitudinal COVID-19 dataset. Experimental results show the effectiveness of the proposed method.

II. RELATED WORKS

A. COVID-19 INFECTION CT SEGMENTATION

After the outbreak of COVID-19, several studies have been conducted to comprehend the disease and provide support to radiologists. Wang et al. [2] introduced a noise-robust training method for the task of learning from noisy labels to segment COVID-19 pneumonia lesions from lung CT scans. Shan et al. [3] proposed VB-Net, a modified 3D convolutional neural network that combines V-Net [12], a fully convolutional neural network for volumetric medical image segmentation with a bottleneck deep residual learning framework for quantitative assessment of COVID-19 infection. Fan et al. [4] presented a semi-supervised learning approach for segmenting diverse radiological patterns, including ground-glass opacity and consolidation, from lung CT scans. However, due to factors such as subtle anatomical boundaries, pleural-based location, and high variations in infection characteristics and imaging methods, the automated identification and quantification of CT image findings related to COVID-19 remain challenging [1], [7].

B. INTERACTIVE SEGMENTATION

In interactive segmentation, user feedback plays a crucial role in enhancing the predictions of machine learning models. This human-in-the-loop method shows potential for improving segmentation results [9]. However, for interactive segmentation to be valuable in a clinical setting, it must exhibit both accuracy and efficiency. Xu et al. [13] introduced a deep interactive object selection method that transforms user-provided clicks into Euclidean distance maps. While this approach lacks the exploitation of image context information, it serves as a foundation for further advancements. Criminisi et al. [14] introduced the geodesic distance transform, which encodes spatial regularization and contrast-sensitivity information, but its sensitivity to noise poses a challenge. Wang et al. [10] presented a deep learning-based interactive segmentation framework that incorporates user-provided bounding boxes and scribbles. By integrating this user feedback, they demonstrated an improvement in segmentation performance. Zhou al. [11] showed that with a small number of user interactions, segmentation accuracy can be substantially improved. Kitrungrotsakul et al. [15] proposed a segmentation refinement module that can be appended to automatic segmentation networks and utilized a skip connection attention module to improve important features for both segmentation and refinement tasks. However, it should be noted that the aforementioned methods are primarily designed for single-time point data and may not address the specific challenges of longitudinal interactive segmentation nor exploit its advantages.

C. LONGITUDINAL IMAGE SEGMENTATION

To comprehensively study the progression of a disease, consistent segmentation of affected regions using scans from multiple time points can provide valuable insights. Birenbaum and Greenspan [16] propose the use of multiple longitudinal networks to process longitudinal patches from different views, where the model concatenates the encoder outputs to perform Multiple Sclerosis (MS) lesion segmentation. Their method demonstrates the benefits of incorporating information from multiple time points into the segmentation

process. To address the segmentation challenges posed by minute structural differences in MS lesions across different time points, [17] proposes a longitudinal network with an early fusion of two-time points scans to implicitly encode the structural differences. However, achieving high accuracy remains a challenge for the network due to the subtle variations. Kim et al. [18] present a framework that leverages spatio-temporal cues between longitudinal scans to improve quantitative assessment of the progression of COVID-19 infection in chest CT scans. Despite these advancements, the mentioned implementations do not utilize the available segmentation mask from previous time point scans to segment a patient's follow-up scans.

III. PROPOSED METHOD

A. INTERACTIVE SEGMENTATION NETWORK

The proposed method builds upon the baseline longitudinal network initially introduced by Denner et al. [17] to effectively leverage longitudinal information and user feedback for interactive segmentation. The baseline longitudinal model, also referred to as a 2.5D model, incorporates the global context of the CT volume by combining per-slice predictions (2D) from three anatomical planes (coronal, sagittal, and axial views) to generate segmentations for individual voxels. Each slice is processed using FC-DenseNet56 [19], a fully convolutional dense network for 2D segmentation. Previous studies [20], [21], [22], [23] have demonstrated that 2.5D approaches can achieve state-of-the-art results in various medical image segmentation tasks. This is mainly because fully 3D approaches tend to entail high computational costs and are data greedy, while patch-based 3D approaches sacrifice global structural information within the individual slices. Thus, our baseline model adopts the 2.5D framework, utilizing two-time points that consist of stacked 2D slices from three anatomical views. This approach effectively preserves global information along two axes while incorporating local information from the third axis [18], [21].

Let $\mathbf{V}_{t+1} \in \mathbb{R}^{h\times w\times s}$ and $\mathbf{V}_t \in \mathbb{R}^{h\times w\times s}$ denote the follow-up target CT volume and the reference previous volume, respectively. h and w are the height and width of the input, and s is the number of slices in the volume. $\mathbf{X}_{t+1}$ and $\mathbf{X}_t$ indicate the individual slices of the volumes. In this study, we assume that the segmentation masks for the reference previous slice $\mathbf{S}_t \in \{0,1\}^{C\times h\times w}$ is available. C represents the number of foreground classes, and it is set to 2 in this study since we are targeting the segmentation of two foreground classes (i.e., ground-glass opacity and consolidation). In addition, let $\mathbf{E}_{t+1} \in \{-1,1\}^{C\times h\times w}$ denote the editing masks on the target segmentation during the user feedback. Note that the segmentation masks for the reference previous volume and the editing masks are concatenated to the input.

1) TRAINING

To facilitate the adaptation of the model to various combinations of input information during inference, we employ a training strategy that involves randomly training the model with two different inputs. This approach allows the model to learn and adjust its parameters to handle different input configurations effectively. The process is visually illustrated in Figure 1.

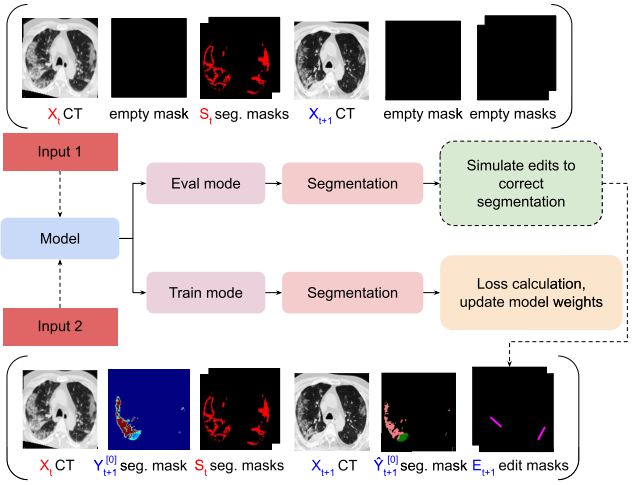


FIGURE 1. Training flow of the proposed interactive longitudinal segmentation model. Training alternates between two training inputs that represent different scenarios: *Input 1* for initial segmentation round. *Input 2* for interactive segmentation rounds. Note that $\mathbf{S}_t$ and $\mathbf{E}_{t+1}$ two channels in our case as $C = 2$. Accordingly if $\mathbf{E}_{t+1}$ is not available as it in the scenario of *Input 1*, $\mathbf{E}_{t+1}$ is a pair of empty masks.

In our framework, *input 1* corresponds to the initial stage of interactive segmentation, where the user provides the data to the model to generate the first segmentation for the target slice whereas *input 2* represents the input data in the subsequent editing rounds. In both cases, the scans from the two-time points are concatenated along the channel dimension so that structural changes that are evident between them are utilized by the model to improve its segmentation performance. Empty masks are used in place of information that is not available in the first segmentation round, such as user feedback and target prediction.

Let $\mathbf{I}^{[T]}_{t+1} \in \{0,1\}^{h\times w\times 8}$ denote the input tensor for the segmentation round T. The input $\mathbf{I}^{[T]}_{t+1}$ consists of the reference previous CT slice $\mathbf{X}_t$, the segmentation masks on the reference previous CT slice $\mathbf{S}_t$, the target CT slice $\mathbf{X}_{t+1}$, the highest class probability per pixel on the target slice from the previous segmentation round $\mathbf{Y}^{[T-1]}_{t+1} \in \{0,1\}^{h\times w}$, the predicted class on the target slice from the previous segmentation round $\hat{\mathbf{Y}}^{[T-1]}_{t+1} \in \{0,1,2\}^{h\times w}$ (0 for background, 1 for ground-glass opacity, and 2 for consolidation), the editing masks $\mathbf{E}_{t+1}$ as shown in Figure 1. For *Input 2*, the empty mask is used for $\mathbf{Y}^{[T-1]}_{t+1}$, $\hat{\mathbf{Y}}^{[T-1]}_{t+1}$, $\mathbf{E}_{t+1}$. The interactive segmentation refinement network (ISR), f_{ISR} outputs the per-slice segmentation of the target image as follows:

$$\mathbf{Y}^{[T]}_{t+1} = f_{ISR}(\mathbf{I}^{[T]}_{t+1}) \in \{0,1\}^{h\times w} \quad (1)$$

To produce the simulated edits for training using *input 2*, the model is first set to evaluation mode, preventing any updates to its weights. An initial segmentation of the target

Algorithm 1 Training of Proposed Interactive Longitudinal Segmentation Model, f_{ISR}

```
for 1 ≤ e ≤ epochs do
    for i = 1 to N do
        Get 1st predictions for input batch i
        Generate random number, Z ∈ [0, 1]
        if Z > 0.5 then
            Generate simulated edits for the predictions
            Append simulated edits & outputs from
          1st prediction round to inputs
            Get 2nd predictions
            Calculate loss using 2nd predictions
        else
            Calculate loss using 1st predictions
        end if
        Backpropagate & update model weights
    end for
end for
```

slice is generated. This is then used to produce the simulated edits. The edit simulation process will be further introduced in Section III-B.

During the training phase, the segmentation of individual slices is treated as a 2D segmentation problem. During the inference phase with user feedback, the predictions obtained from slices in the three anatomical orientations are combined to generate the final segmentation output for each voxel. The pseudocode outlining the training process of the proposed longitudinal interactive segmentation model is presented in Algorithm 1. Algorithm 1 provides a step-by-step description of the procedures and operations involved in training the model.

B. EDIT SIMULATION DURING TRAINING

During the training process, simulated user edits are automatically generated for regions that are incorrectly segmented, which can include areas that are under-segmented or over-segmented. Incorrectly segmented regions are areas that are under- or over-segmented. The segmentation output from the model is compared with the ground truth to choose the slice region for simulating the user edits. Lines are automatically drawn on the selected regions as simulated feedback. As mentioned before, the edit information is concatenated to the CT scans as additional channels, one for each class, with foreground interaction having a value of 1 and background interaction −1. Because the model input is 2.5D instead of 3D, edits are simulated in the axial, coronal, and sagittal slices as opposed to only the axial slices as in [11]. Zhou et al. [11] simulate edits only for the most extensive 2D incorrectly segmented slice region. However, due to the scattered nature of the COVID-19 infections in our case, the top-5 largest wrongly segmented regions in each slice is used for edit simulation. The total number of generated edits, independent of the different classes are limited to prevent

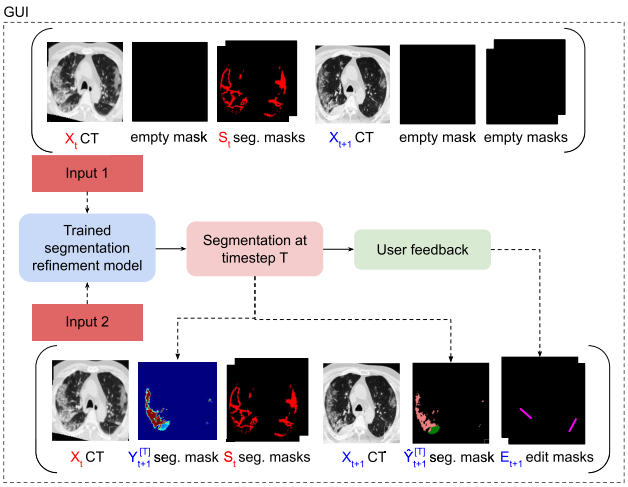


FIGURE 2. Overview of the interactive segmentation flow with the GUI. Input to the segmentation model prior to the first user interaction resembles *Input 1* that is used during training. Accordingly, the input of subsequent editing rounds resembles *Input 2* as in Figure 1 which improves segmentation accuracy by incorporating user feedback via the GUI.

the model from overfitting and also to avoid considerable slow-down in the training process caused by the long edit simulation time needed when large numbers of incorrectly segmented regions are detected.

C. INFERENCE WITH GUI FOR USER FEEDBACK

Figure 2 illustrates the interactive segmentation refinement stage, where an editing graphical user interface (GUI) is employed. The GUI is implemented using Qt Designer[2] as shown in Figure 3. During start-up, the GUI automatically loads the trained segmentation refinement model, which has been trained to improve segmentation results. The user can then load the CT volume that requires segmentation. After the initial segmentation round, the predicted segmentation is overlaid on the CT scans for visual inspection. This allows the user to assess the accuracy and quality of the segmentation. The user can utilize the brush tool provided in the GUI to edit incorrectly segmented areas, making necessary adjustments and refinements. The data can then be processed through the model again to generate an updated segmentation. This iterative process of editing and re-segmenting can be repeated as many times as necessary to achieve the desired segmentation outcome.

During each segmentation refinement round, the user feedback pertaining to incorrectly segmented regions from both the current round and previous rounds is aggregated into an edit mask for each slice. This edit mask captures the cumulative feedback from multiple editing iterations. To prioritize the current round's user feedback, it is multiplied by two before being added to the previous user edits. This amplification ensures that the most recent feedback has a greater impact on the refinement process.

[2]https://doc.qt.io/qt-5/qtdesigner-manual.html

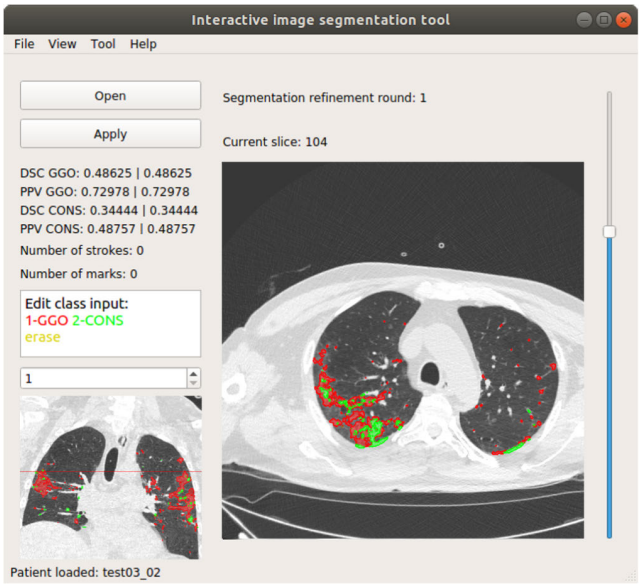


FIGURE 3. Our GUI for COVID-19 lung infection interactive segmentation. The spin box in the sidebar shows the class for the current brush input, here 1, i.e., the brush for ground-glass opacity (GGO).

To preserve the previous edit information and avoid its loss, the values in the edit mask are clipped to the range of $[-1, 1]$. This clipping operation ensures that the edit mask retains both positive and negative feedback signals. Subsequently, the edit mask is concatenated with its corresponding slice image, creating an augmented input that includes both the image data and the edit mask. This augmented input is then fed into the segmentation refinement model, which leverages this combined information to improve the segmentation results during each refinement round.

IV. EXPERIMENTS

A. IMPLEMENTATION

In this study, we utilized a modified version of the baseline longitudinal segmentation model proposed by [17]. The model is implemented as an end-to-end 2.5D segmentation network based on FC-DenseNet56 [19] and implemented using PyTorch 1.4 [24]. For training, we employed the mean squared error (MSE) loss function and utilized the Adam optimizer with AMSGrad [25]. The learning rate used during training was set to 0.0001. The inference time for processing a COVID-19 patient's 2.5D data with a size of $3\times(150 \times 150\times150)$ takes 15 seconds on an NVIDIA GeForce RTX 2080 Ti with 11GB GPU.

B. COVID-19 SEGMENTATION DATASET AND PREPROCESSING

For this study, we utilized an in-house clinical dataset collected from the Radiology Department of the Technical University of Munich during the first wave of the COVID-19 pandemic (March-June 2020). The dataset consists of 30 longitudinal low-dose native CT scans obtained from patients aged between 46 and 82 years old. All patients had

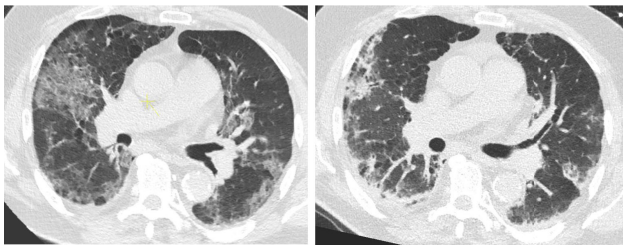


FIGURE 4. Deformable registration example of reference scan to target scan. (a) Reference scan, (b) target scan.

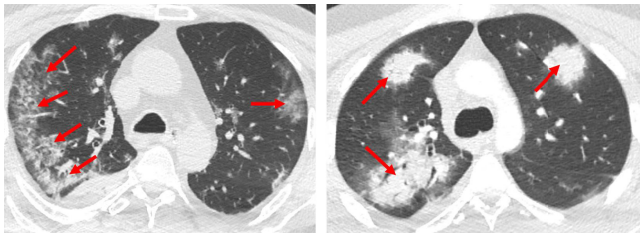
a confirmed positive polymerase chain reaction (PCR) test for COVID-19. The time gap between the follow-up scans and the previous scans is 17 ± 10 days. These scans were performed during the patients' admission and hospitalization, with an average duration of 33 ± 21 days (ranging from 0 to 71 days).

Two different CT imaging devices were used for data acquisition: the IQon Spectral CT and iCT 256, both from Philips Healthcare, located in Best, the Netherlands. The imaging parameters were consistent across both devices, including an X-ray current of 140-210 mA, voltage of 120 kV peak, and slice thickness of 0.9 mm. The scans covered the entire lung region. The collection and usage of this dataset were carried out in compliance with ethical considerations and guidelines. The study received approval from the institutional review board of the Technical University of Munich (ethics approval 111/20 S-KH).

The annotation process for the dataset involved an expert rater, a radiologist with four years of experience. The annotations were performed at the voxel level using ImFusion Labels software, developed by ImFusion GmbH in Munich, Germany.([3]) The dataset includes lung masks, which distinguish lung parenchyma from other tissues, as well as pathology masks for four classes: healthy lung (HL), ground-glass opacity (GGO), consolidation (CONS), and pleural effusion (PLEFF). These masks were generated by the expert rater.

To ensure optimal training and evaluation, preprocessing steps were applied to the raw CT volumes due to variations in intensity range, size, and alignment. The volumes were first cropped to focus on the lung regions based on the manually annotated lung masks. Next, intensity values outside the range of (-1024, 600) were clipped. A min-max normalization technique was then applied to the volumes. To achieve a consistent input size for the segmentation model, the volumes were resized to dimensions of $150 \times 150\times150$ pixels. Additionally, slices that exhibited a voxel-value variation smaller than 0.001% between their minimum and maximum values were considered almost empty slices and subsequently removed from the dataset to ensure data integrity and avoid introducing noise into the training process.

[3]https://www.imfusion.com/



(a) (b)

FIGURE 5. Examples of GGO and CONS on axial slices of lung CT scans from COVID-19 patients. Arrows point to the areas affected by the infection. (a) GGO, (b) CONS.

In line with the approach taken by [18], we utilized the deformable registration algorithm introduced by [26]. This algorithm employs a B-Spline Transform with a sparse set of grid points that overlay the fixed domain of the image. By deforming the image through this transformation, the algorithm achieves registration of the reference scan to the follow-up scan. This process helps to resolve misalignment errors between scans. To ensure accurate registration, we performed the registration process specifically on the lung masks rather than the entire image. This approach minimizes the risk of registration errors arising from pathological changes in the lung parenchyma. An example of the aligned CT scans from different time points is presented in Figure 4, showing the effectiveness of the registration process in aligning the images and compensating for any misalignments that may have occurred during acquisition or due to patient motion.

According to [27], GGO is the most common finding in CT scans of COVID-19 patients, followed by CONS. Figure 5 shows examples of GGO and CONS from our dataset, illustrating the visual characteristics of these findings. For our experiments, we only segment GGO and CONS, due to the low occurrences of PLEFF in the patient cohort of the dataset. After the registration, the average structural similarity index (SSIM) [28] between the scans from different time points was calculated to be 29.71%. This indicates that the CT scans taken at different time points exhibit significant perceptual differences, suggesting variations in disease progression and lung conditions over time. Besides that, the average change in the percentage of GGO and CONS in the patients' lung CTs from different time points is 13.68% and 6.59%, respectively. This indicates the noticeable difference in the disease progression over time in the dataset. Table 1 shows the percentage of GGO and CONS in the lungs of the patients at each timestep.

The training set consists of data from 16 patients, which corresponds to a total of 37 volumes of CT scans, with 12 patients (28 volumes) used for training and the remaining 4 patients (9 volumes) allocated for validation purposes. To assess the performance of our model, we conducted testing on an independent test set comprising data from 14 patients, resulting in a total of 28 volumes of CT scans. This test set was separate from the training and validation sets, ensuring the evaluation of the model's generalization ability and its performance on unseen data.

TABLE 1. Percentage of GGO and CONS in the lungs of patients. *T* refers to the timestep of the scans. *n* is the number of patient volumes at each timestep.

Radiomic	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.
	T-1 (n=30)		T-2 (n=30)		T-3 (n=5)	
GGO	15.23%	14.76%	20.17%	18.07%	15.02%	11.91%
	T-1 (n=30)		T-2 (n=30)		T-3 (n=5)	
CONS	6.52%	7.07%	8.15%	11.51%	11.28%	11.07%

TABLE 2. Input for different models tested in ablation study.

Model name	Model input				
	Target image	Reference image	Ref. manual segmentation	Edit mask	Target previous segmentation
Baseline static network	✓				
Baseline long. network	✓	✓			
Baseline long.+ref_seg	✓	✓	✓		
static_edit	✓			✓	✓
long_edit+ref_seg	✓	✓	✓	✓	
Proposed	✓	✓	✓	✓	✓

C. EXPERIMENTAL SETTINGS

1) EVALUATION METRICS

The segmentation performance of the models is evaluated using the following metrics. Dice Similarity Coefficient (DSC) is a statistical measure of the similarity between two segmentations.

$$DSC = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{2}$$

Positive Predictive Value (PPV) shows the fraction of correctly segmented regions over predicted segmentations.

$$PPV = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

True Positive Rate (TPR) shows the proportion of correct segmentation outputs with respect to the ground truth.

$$TPR = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

Volume Difference (VD) is calculated as the absolute difference in the predicted lesion segmentation volume and ground truth lesion segmentation volume over the ground truth lesion segmentation volume.

$$VD = 100 \times \frac{\left|\text{Lesion\_volume}_{pred} - \text{Lesion\_volume}_{gt}\right|}{\text{Lesion\_volume}_{gt}} \tag{5}$$

D. EXPERIMENTAL RESULTS

1) ABLATION STUDY

In our ablation study, we aimed to investigate the impact of incorporating additional information into the segmentation model's inputs on its performance. For this purpose, we utilized our longitudinal COVID-19 dataset and compared different model configurations. The baseline model referred

TABLE 3. Evaluation results on the test set before user interactions. Values displayed are the mean and standard errors. Bold values represent the best results for each metric. Empty masks are used in place of the edit mask and target image previous segmentation mask for static_edit, long_edit+ref_seg and the proposed model.

Model	Dice (%)		PPV (%)		TPR (%)		VD (%)	
	GGO	CONS	GGO	CONS	GGO	CONS	GGO	CONS
Non interactive methods								
Baseline static network	44.15 ± 3.33	19.75 ± 4.89	62.12 ± 2.89	**42.86** ± 8.53	38.98 ± 5.09	14.96 ± 3.60	49.98 ± 6.07	102.0 ± 28.84
Baseline long. network	45.42 ± 3.03	27.63 ± 5.61	**67.34** ± 3.03	41.35 ± 8.01	37.96 ± 4.46	23.54 ± 4.32	48.56 ± 6.09	93.18 ± 38.04
Baseline long.+ref_seg	**46.86** ± 3.12	28.39 ± 6.39	62.61 ± 3.86	42.12 ± 8.53	**41.63** ± 4.16	24.36 ± 5.61	**43.31** ± 6.53	90.36 ± 39.43
Interactive methods								
static_edit	35.97 ± 4.20	13.99 ± 3.51	66.82 ± 3.01	41.20 ± 8.45	29.03 ± 4.99	9.28 ± 2.27	59.12 ± 7.20	98.11 ± 22.21
long_edit+ref_seg	37.70 ± 3.89	**29.69** ± 5.86	64.66 ± 3.03	37.85 ± 7.60	30.39 ± 4.46	**26.80** ± 5.35	55.46 ± 6.75	64.88 ± 27.62
Proposed	44.42 ± 3.61	23.94 ± 6.12	62.65 ± 3.49	41.89 ± 8.86	40.38 ± 5.42	19.62 ± 5.92	51.84 ± 7.40	**49.71** ± 7.94

TABLE 4. Evaluation results of the interactive methods on the test set with one round user interaction. Values displayed are the mean and standard errors. Bold values represent the best results for each metric.

Model	Dice (%)		PPV (%)		TPR (%)		VD (%)	
	GGO	CONS	GGO	CONS	GGO	CONS	GGO	CONS
Initial segmentation with user edits								
static_edit	40.86 ± 3.16	24.33 ± 3.14	67.80 ± 3.01	54.37 ± 7.45	32.77 ± 4.33	17.16 ± 1.97	54.15 ± 6.31	87.99 ± 20.73
long_edit+ref_seg	44.34 ± 2.34	36.71 ± 5.57	67.03 ± 3.30	44.38 ± 6.84	35.72 ± 3.39	33.91 ± 5.05	48.76 ± 5.41	54.67 ± 20.78
Proposed	49.22 ± 2.58	36.33 ± 5.19	64.34 ± 3.88	50.50 ± 7.10	44.64 ± 4.50	30.91 ± 5.13	46.63 ± 6.54	40.08 ± 6.03
Output segmentation after one round of segmentation refinement by model								
static_edit	53.59 ± 2.33	55.48 ± 4.48	**71.63** ± 4.12	**71.34** ± 4.34	44.62 ± 2.98	47.01 ± 4.68	40.21 ± 3.76	**33.68** ± 5.66
long_edit+ref_seg	54.79 ± 1.95	54.01 ± 5.03	66.70 ± 3.62	51.14 ± 5.16	48.54 ± 2.54	**64.44** ± 5.67	**34.74** ± 4.55	74.43 ± 36.10
Proposed	**59.86** ± 2.57	**58.81** ± 4.36	62.06 ± 4.48	58.18 ± 5.15	**61.86** ± 3.00	62.39 ± 4.06	36.39 ± 11.28	34.11 ± 12.75

to as "Baseline longitudinal network" [17], served as the starting point for our study. This model was designed to handle longitudinal data and produce initial segmentations. We also implemented a static version of the model as a baseline for comparison. To evaluate the effect of incorporating additional information, we extended the baseline longitudinal model with an interactive segmentation approach. This extended model, denoted as "long_edit+ref_seg," included past prediction outputs of the target as supplementary information to guide subsequent segmentations. Additionally, we implemented a static version of the proposed model without reference information, labeled as "static_edit". Table 2 provides an overview of the tested models and their corresponding inputs. The reference manual segmentation refers to the ground truth masks of the reference images, while the edit masks contain user feedback on the target segmentation.

As shown in Table 3, the baseline models performed better than the interactive segmentation models in the first segmentation round where there is no edit. The baseline longitudinal model achieves higher Dice scores compared to the baseline static model. Among the longitudinal baseline models, concatenating reference segmentation to the input CT scans can improve its GGO Dice by 1.44% and CONS Dice by 0.76%. Comparing the interactive segmentation models, we found that the longitudinal interactive segmentation models achieved better initial segmentations compared to the static model. Specifically, the proposed model demonstrated an 8.45% increase in GGO Dice compared to the static model, while the long_edit+ref_seg model exhibited a 15.7% improvement in CONS Dice over the static model.

The evaluation results in Table 3 further revealed that the CONS true positive rate (TPR) for the static models is considerably lower than the longitudinal models, with the noninteractive methods having a difference of 8.58% between the baseline static network and baseline longitudinal network. This suggests that the longitudinal models are better suited for segmenting the more complex CONS. In the case of interactive segmentation models, the static_edit model exhibits a CONS TPR that is 10.34% lower than the longitudinal interactive model with a lower TPR. These results further emphasize the advantage of longitudinal models in improving the segmentation performance, particularly for challenging classes such as CONS.

2) QUANTITATIVE RESULTS

The evaluations are conducted using the GUI to simulate user feedback on the segmentation output. The interactive segmentation model aims to assist and reduce the user's workload during the segmentation process by refining the initial segmentation based on user interactions. Therefore, its desired function is to take in rough user interactions and improve the segmentation output. In order to evaluate the effectiveness of the segmentation refinement model, a comparison is made between the manually corrected initial segmentation and the segmentation output refined by the

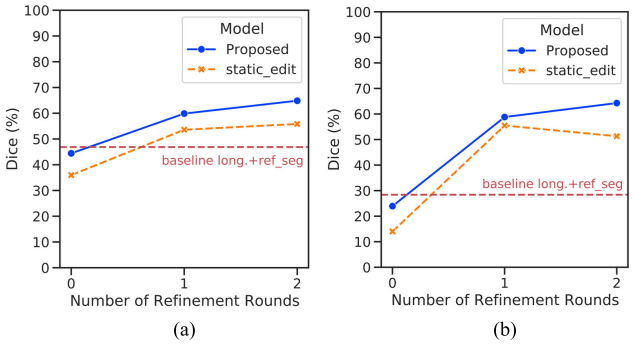


FIGURE 6. Average change in Dice vs Number of Refinement Rounds for 14 test patients. (a) Results on GGO, (b) results on CONS.

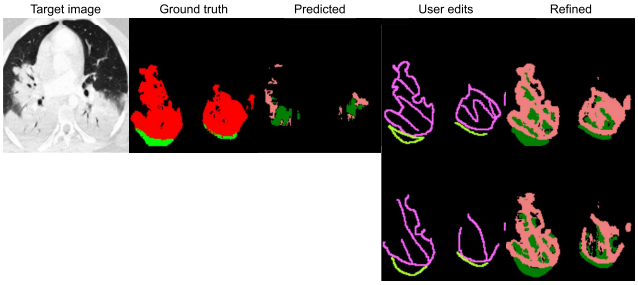


FIGURE 7. Example of how the refined segmentation differs when different types of user edits (top and lower rows) are drawn to refine the segmentation. (*Red:* GGO ground truth, *Green:* CONS ground truth, *Pink:* predicted GGO, *Dark green:* predicted CONS, *Magneta:* foreground edit for GGO, *Neon green:* foreground edit for CONS).

model. The results are shown in Table 4. As shown in the table, the interactive segmentation refinement model effectively utilizes both the previous segmentation results and user feedback to improve the segmentation of the target scans. Among the three models compared, the proposed model achieves the highest Dice scores for GGO and CONS after one round of segmentation refinement. However, when considering the change in Dice between the edited initial segmentation by the user and the refined segmentation by the model, it is found that the static interactive segmentation model shows the most substantial improvement in segmentation output. It demonstrates a 12.73% increase in GGO Dice and a 31.15% increase in CONS Dice, while the proposed model shows a 10.64% increase in GGO Dice and a 22.48% increase in CONS Dice.

The proposed model and static_edit model are further evaluated by conducting an additional round of segmentation refinement. Figure 6 shows the changes in DICE score for 14 patients after each round of segmentation refinement, incorporating user edits. The baseline model, baseline long.+ref_seg, is included for comparison. Experimental results show that the proposed model achieves a significant improvement in Dice scores after just two rounds of segmentation refinement with user feedback. The average increase in Dice scores for GGO and CONS after two rounds of refinement is 20.44% and 40.33%, respectively. There is an average increase of 4.99% for GGO and 5.46% for CONS between the first and second rounds of segmentation refinement. In comparison, the static_edit model shows lower Dice scores for GGO and CONS after two rounds of segmentation refinement. Specifically, the GGO and CONS Dice scores are 9.03% and 12.98% lower, respectively, compared to the proposed model. These results demonstrate the effectiveness of the proposed model in improving segmentation accuracy through iterative refinement based on user edits. The model's ability to adapt to user feedback and incorporate it into the segmentation process leads to significant improvements in the segmentation performance. From the plots in Figure 6, it is shown that the proposed model outperforms the static interactive segmentation model in terms of segmentation refinement performance. The static_edit model fails to correctly refine the segmentation at the second refinement step, as observed by the decrease in CONS segmentation Dice. Upon closer examination of the segmentation output, it is observed that the static_edit model is more prone to misclassifying regions, leading to a decrease in Dice. On the other hand, the proposed model benefits from the inclusion of reference scan information, resulting in improved segmentation.

The number of rough user edit strokes required to modify the initial segmentation varies from 1 to 10 per edited slice for the proposed model, depending on the severity of infection in the test patients. In most cases, the increase in Dice is larger for CONS after refinement. Qualitative results of the initial segmentation reveal that CONS is frequently under-segmented due to its high similarity to the background class. The presence of similar Hounsfield unit histograms in blood vessels and airway walls makes it challenging for the model to accurately segment the CONS region without user guidance.

3) QUALITATIVE RESULTS

In this section, we present qualitative results obtained from the proposed model. Figure 7 shows the segmentation refinement output for various types of user edits. In the lower segmentation output, it is observed that simply drawing the outer outline of the infected regions is sufficient to separate them from the background. The model is able to further segment the unedited areas, although it may make some incorrect classifications. As shown in Figure 8, this incorrect segmentation can be corrected in subsequent refinement rounds.

The feature similarity between GGO and CONS in CT scans can indeed pose a challenge for accurate segmentation. One of the key distinguishing factors between GGO and CONS is their location within the lung CT. CONS is typically found at the bottom of the axial lung CT, but in severe cases, it can extend to higher regions, where its borders may overlap with GGO regions. This overlapping region can make it difficult for the model to accurately differentiate between the two classes. As shown in the second row of Figure 8, the CONS region sometimes extends higher up in

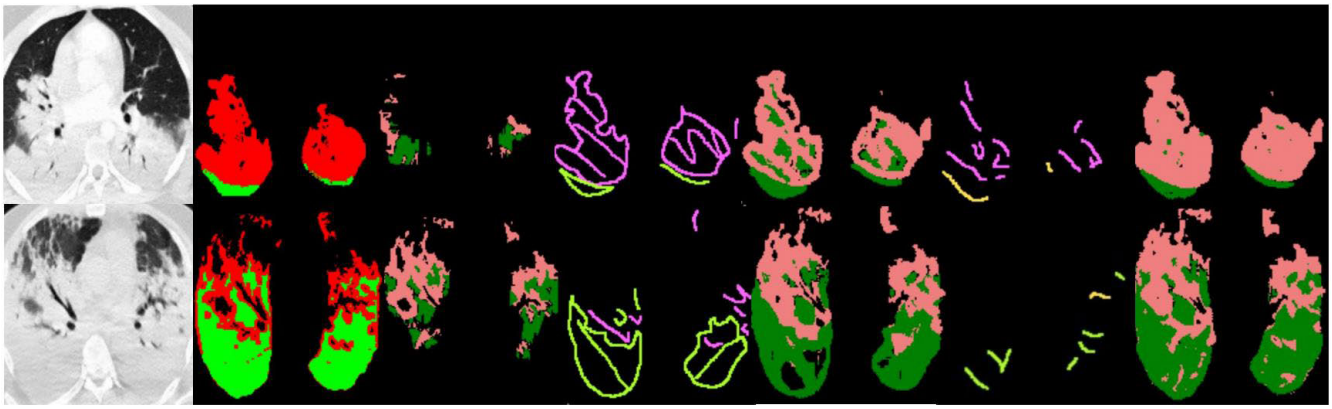


FIGURE 8. Examples of two rounds of segmentation refinement, resulting in increasingly better results. (a) Target images, (b) ground truth masks, (c) predicted masks, (d) user edits in the first round, (e) refined results from first round interaction, (f) user edits in the second round, (g) refined results from first round interaction.

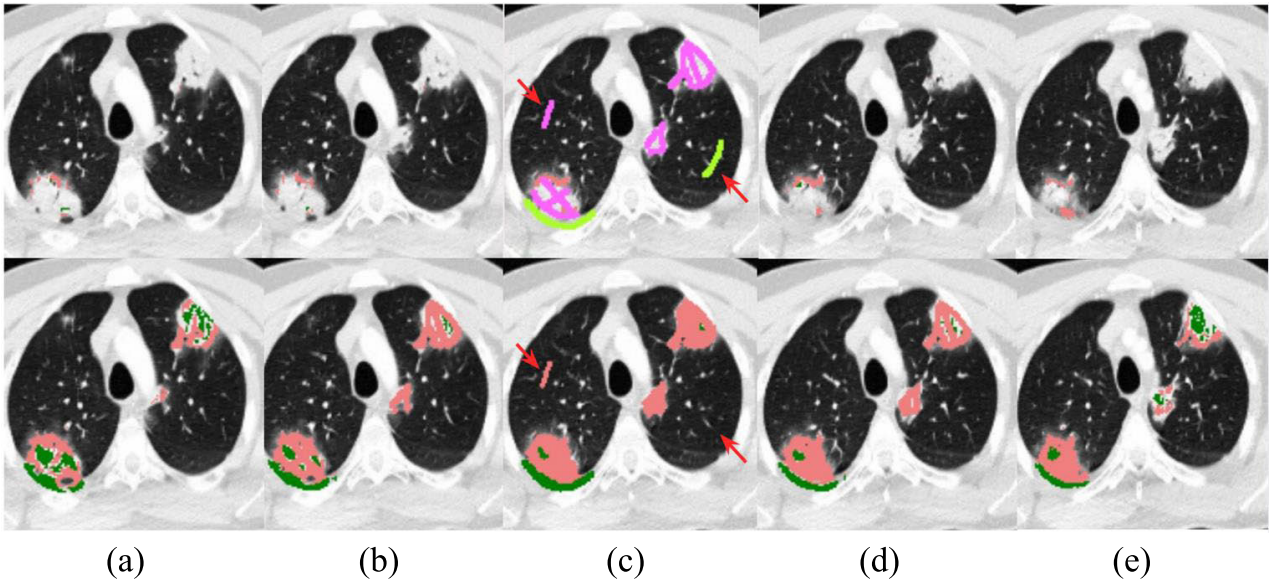


FIGURE 9. Example of how edits on one slice are automatically propagated to other slices. (a) Image before the target image (slice index i-2), (b) image before the target image (slice index i-1), (c) target image (slice index i), (d) image after the target image (slice index i+1), (e) image after the target image (slice index i+2). The first row represents the images before refinement (initial prediction) and the second row represents the images after the refinement. After the initial prediction, edits are drawn on one slice. Red arrows point to false edit locations. Interestingly, the model learns that false edits should not be propagated further than one slice by the model.

the lung and its borders coincide with GGO regions. In such cases, the model may initially misclassify these regions, but with the iterative refinement process, it can learn from user edits and correct the segmentation to accurately differentiate between GGO and CONS. This highlights the importance of user guidance and refinement iterations in cases where there is feature similarity or ambiguity between different classes. By incorporating user feedback and iterative refinement, the proposed model can adapt and improve its segmentation output, leading to more accurate and precise results, even in challenging scenarios like the overlapping GGO and CONS regions.

4) ROBUSTNESS TESTING

A robustness test is carried out to assess the ability of the 2.5D model, which utilizes stacked slices from three anatomical views as input, to propagate edits made on one slice to other slices. In Figure 9, it can be observed that the automatic propagation of edits from one slice to adjacent slices is limited but still present. The model is able to correctly segment the regions that were edited in the initial slice, but as the slices move further away from the edited slice, there is a gradual increase in misclassifications and under-segmentation of certain regions. However, it is noteworthy that the model does not propagate false edits

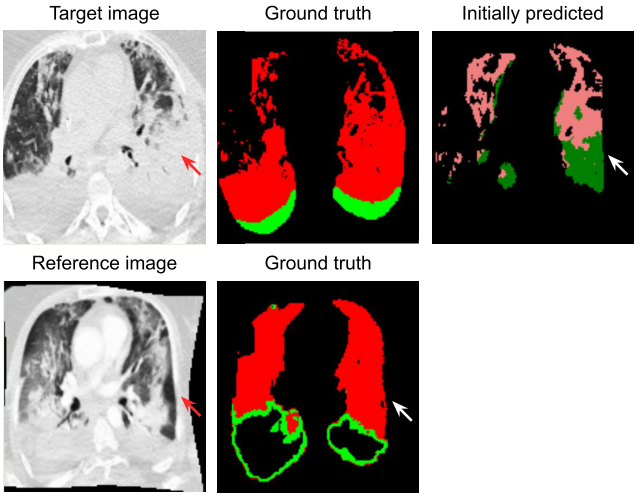


FIGURE 10. Example of how the target's initial prediction is influenced by the reference image in a case where there is a misalignment between the reference and target image. The red arrows point to regions where the CT scans differ the most. The white arrows point to the interested regions.

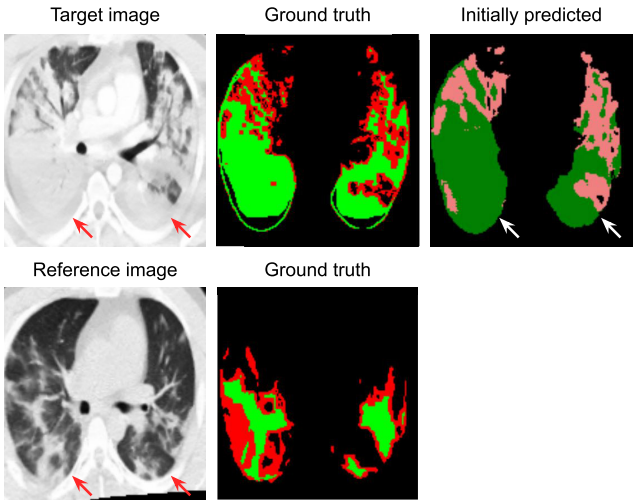


FIGURE 11. Example of a case where the reference image and target image are correctly aligned. The model segmented the difficult areas better compared to Figure 10. The red arrows point to regions where the CT scans differ the most. The white arrows point to the interested regions.

further than one slice away. This indicates that the model has learned to distinguish between correct edits and erroneous edits based on the features present in the CT scans. It demonstrates the model's ability to detect and limit the propagation of incorrect edits, preventing them from influencing the segmentation results of neighboring slices.

5) INFLUENCE OF REFERENCE DATA ON TARGET SEGMENTATION

In some cases, where the alignment of the patient's reference scan with the follow-up scan is not accurate, the initial segmentation of the follow-up scan for challenging regions can be negatively affected by the reference CT. This issue is illustrated in Figure 10, where abnormalities are observed in the initial predicted segmentation of the target image. The infected lower part of the target CT, which is harder to distinguish from the background class, is segmented based on the reference image. As observed in the initial predicted segmentation of the target, the left outline closely resembles that of the reference image. On the other hand, when the reference image is correctly aligned with the target image, as in Figure 11, the reference image can serve as a guide for the model to segment regions that are challenging for the model, such as areas that resemble the background class. In this example, the model produces a more accurate initial segmentation of the target scan.

E. LIMITATIONS AND FUTURE WORK

There are several limitations that could be addressed in future work. Firstly, the availability of a larger longitudinal COVID-19 dataset would be beneficial to further evaluate and generalize the proposed method. With a larger dataset, the model's performance and generalizability can be better assessed. Secondly, label noise is a potential issue in the dataset, particularly in severe cases where even expert radiologists may struggle to accurately separate CONS from the background class and PLEFF. Addressing the presence of label noise and exploring strategies to mitigate its impact could lead to improved segmentation results.

In future research, the proposed method can be extended to other longitudinal medical image datasets and applied in different clinical contexts. Investigating the effectiveness of the method in improving segmentation results in various medical imaging scenarios would provide valuable insights and potential applications. The issue of wrongly classified segmentations in multiclass segmentation can be further explored by utilizing an ensemble of binary interactive segmentation models for each foreground class. This approach could potentially improve the accuracy and reliability of segmentations, especially for challenging classes.

Additionally, testing the proposed method with other state-of-the-art model architectures would be worthwhile. Comparing the performance of the proposed method with different model architectures could provide insights into its effectiveness and applicability across different models. The study also highlights a limitation of the 2.5D model in propagating edits from one slice to other slices that are further away. Future work could explore the implementation of a 3D interactive segmentation method, which may overcome this limitation at the cost of requiring larger training datasets and increased computational resources.

V. CONCLUSION

In this paper, we proposed an interactive segmentation method using a 2.5D longitudinal network. The method leveraged the information from previous time points, including the reference segmentation mask, segmentation outputs from past rounds, and user interactions, to enhance the segmentation results. The proposed method was evaluated on a longitudinal COVID-19 dataset, and the experimental

results demonstrate a significant improvement in the Dice scores for both classes after one round of interactive segmentation refinement. The study highlights the advantages of the proposed longitudinal interactive segmentation refinement model compared to a static version of the interactive model. The longitudinal model achieves superior segmentation performance, indicating the effectiveness of utilizing longitudinal information for segmentation refinement. The findings of this study suggest that existing segmentation models can be easily adapted and trained end-to-end for interactive segmentation refinement using longitudinal data. By incorporating past information and user interactions, the proposed method improves the segmentation results, making it a valuable tool for interactive medical image segmentation tasks.

REFERENCES

[1] S. Zhou, Y. Wang, T. Zhu, and L. Xia, "CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China," *Amer. J. Roentgenology*, vol. 214, no. 6, pp. 1287–1294, Jun. 2020.

[2] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2653–2663, Aug. 2020.

[3] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction," *Med. Phys.*, vol. 48, no. 4, pp. 1633–1645, Mar. 2021.

[4] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-Net: Automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[5] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020.

[6] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management," *Amer. J. Roentgenology*, vol. 214, no. 6, pp. 1280–1286, Jun. 2020.

[7] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, and C. Zheng, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study," *Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, Apr. 2020.

[8] J. Mendel, J. Lee, and D. Rosman, "Current concepts imaging in COVID-19 and the challenges for low and middle income countries," *J. Global Radiol.*, vol. 6, no. 1, pp. 1–10, Jun. 2020.

[9] H. Ramadan, C. Lachqar, and H. Tairi, "A survey of recent interactive image segmentation methods," *Comput. Vis. Media*, vol. 6, no. 4, pp. 355–384, Aug. 2020.

[10] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.

[11] B. Zhou, L. Chen, and Z. Wang, "Interactive deep editing framework for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent*, Oct. 2019, pp. 329–337.

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016, *arXiv:1606.04797*.

[13] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep interactive object selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 373–381.

[14] A. Criminisi, T. Sharp, and A. Blake, "GeoS: Geodesic image segmentation," in *Computer Vision—ECCV* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2008, pp. 99–112.

[15] T. Kitrungrotsakul, Q. Chen, H. Wu, Y. Iwamoto, H. Hu, W. Zhu, C. Chen, F. Xu, Y. Zhou, L. Lin, R. Tong, J. Li, and Y.-W. Chen, "Attention-RefNet: Interactive attention refinement network for infected area segmentation of COVID-19," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2363–2373, Jul. 2021.

[16] A. Birenbaum and H. Greenspan, "Multi-view longitudinal CNN for multiple sclerosis lesion segmentation," *Eng. Appl. Artif. Intell.*, vol. 65, pp. 111–118, Oct. 2017.

[17] S. Denner, A. Khakzar, M. Sajid, M. Saleh, Z. Spiclin, S. Tae Kim, and N. Navab, "Spatio-temporal learning from longitudinal data for multiple sclerosis lesion segmentation," 2020, *arXiv:2004.03675*.

[18] S. T. Kim, L. Goli, M. Paschali, A. Khakzar, M. Keicher, T. Czempiel, E. Burian, R. Braren, N. Navab, and T. Wendler, "Longitudinal quantitative assessment of COVID-19 infection progression from chest CTs," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent*, 2021, pp. 273–282.

[19] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1175–1183.

[20] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy," *NeuroImage*, vol. 186, pp. 713–727, Feb. 2019.

[21] H. Zhang, A. M. Valcarcel, R. Bakshi, R. Chu, F. Bagnato, R. T. Shinohara, K. Hett, and I. Oguz, "Multiple sclerosis lesion segmentation with Tiramisu and 2.5D stacked slices," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Shenzhen, China: Springer, Oct. 2019, pp. 338–346.

[22] S. Aslani, M. Dayan, L. Storelli, M. Filippi, V. Murino, M. A. Rocca, and D. Sona, "Multi-branch convolutional neural network for multiple sclerosis lesion segmentation," *NeuroImage*, vol. 196, pp. 1–15, Aug. 2019.

[23] R. Alkadi, A. El-Baz, F. Taher, and N. Werghi, "A 2.5D deep learning-based approach for prostate cancer detection on T2-weighted magnetic resonance imaging," in *Proc. Eur. Conf. Comput. Vis.* Munich, Germany: Springer, Jan. 2019, pp. 734–739.

[24] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035.

[25] S. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in *Proc. Int. Conf. Learn. Represent.*, Feb. 2018, pp. 1–21.

[26] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of SimpleITK," *Front. Neuroinform.*, vol. 7, p. 45, Dec. 2013.

[27] M. M. Hefeda, "CT chest findings in patients infected with COVID-19: Review of literature," *Egyptian J. Radiol. Nucl. Med.*, vol. 51, no. 1, p. 239, Nov. 2020.

[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

MICHELLE XIAO-LIN FOO received the B.S. degree from Reutlingen University, in 2018, and the M.S. degree from the Technical University of Munich, Germany, in 2021. She is currently pursuing the Ph.D. degree in artificial intelligence and human–computer interaction with the Ludwig Maximilian University of Munich, Germany. Her current research interests include artificial intelligence, human–computer-interaction, computer vision, and robotics.

SEONG TAE KIM (Member, IEEE) received the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2014 and 2019, respectively. In 2015, he was a Visiting Researcher with the University of Toronto, Canada. From 2019 to 2021, he was a Senior Research Scientist with the Chair for Computer Aided Medical Procedures, Technical University of Munich, Germany. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Kyung Hee University, South Korea. His current research interests include explainable deep learning and data-efficient deep learning.

MAGDALINI PASCHALI received the M.S. and Ph.D. degrees from the Technical University of Munich, Germany, in 2017 and 2021, respectively. She is currently a Postdoctoral Scholar with the Computational Neuroimage Science Laboratory, Stanford University, USA. Her current research interests include machine learning for diagnosis and treatment of neuropsychiatric disorders.

LEILI GOLI received the B.S. degree from the Sharif University of Technology, Iran, in 2021. She is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Toronto, Canada. Her current research interests include 3D computer vision, robotics, and deep learning.

EGON BURIAN is currently a Researcher with the Department of Diagnostic and Interventional Neuroradiology, University Hospital Rechts der Isar, Technical University of Munich, Munich, Germany.

MARCUS MAKOWSKI received the degree from LMU Munich, in 2007, and the Ph.D. degree from the School of Medicine, King's College London, London, in 2010. In 2016, he received an appointment to a W2 Heisenberg Professorship (Charité). He is currently a Professor and the Director of the Department of Diagnostic and Interventional Radiology, Klinikum Rechts der Isar, Technical University of Munich, Germany. His current research interests include quantitative imaging and the application of AI methods in radiology. In 2013, he was awarded the Wilhelm Conrad Röntgen Prize, for the Habilitation degree.

RICKMER BRAREN is currently an Assistant Professor and the Vice-Director of the Department of Diagnostic and Interventional Radiology, Klinikum Rechts der Isar, Technical University of Munich, Germany. His current research interest includes the translational research of machine learning to clinical practice.

NASSIR NAVAB (Fellow, IEEE) received the Ph.D. degree from INRIA, University of Paris XI, France. Before joining Siemens Corporate Research (SCR), in 1994, he was a Postdoctoral Fellow with the MIT Media Laboratory, for two years. He is currently a Full Professor and the Director of the Laboratory for Computer-Aided Medical Procedures, Johns Hopkins University, and the Technical University of Munich. He has also secondary faculty appointments at both affiliated medical schools. He is the author of 100 peer-reviewed scientific articles, with more than 58,000 citations and an H-index of 105 as of December 2022. He is the author of more than 30 awarded articles, including 11 at MICCAI, five at IPCAI, and three at IEEE ISMAR. He is the inventor of 50 granted U.S. patents and more than 50 international ones. His current research interests include medical augmented reality, computer-aided surgery, medical robotics, and machine learning. In 2012, he was elected as a fellow of the MICCAI Society. He has acted as a member of the Board of Directors of the MICCAI Society, from 2007 to 2012 and from 2014 to 2017. He serves on the Steering Committee for the IEEE Symposium on Mixed and Augmented Reality (ISMAR) and Information Processing in Computer-Assisted Interventions (IPCAI). At SCR, he was a Distinguished Member and received the Siemens Inventor of the Year Award, in 2001. He received the SMIT Society Technology Award, in 2010, for the Introduction of Camera Augmented Mobile C-Arm and Freehand SPECT Technologies, and the Ten Years Lasting Impact Award of IEEE ISMAR, in 2015.

THOMAS WENDLER received the Engineering degree in electrical engineering from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2004, and the M.Sc. degree in biomedical engineering and the Ph.D. degree in computer science from the Technical University of Munich, Munich, Germany, in 2007 and 2010, respectively. He is currently the Vice-Director of the Chair for Computer-Aided Medical Procedures and Augmented Reality (CAMP), Technical University of Munich (TUM), and the CEO of ScintHealth GmbH, Munich. His current research interests include the translation of novel computer-aided intervention tools and data-driven medical image analysis methods for clinical decision support into clinical applications.