Self-supervised region-aware segmentation of COVID-19 CT images using 3D GAN and contrastive learning

Siyavash Shabani [a],[1], Morteza Homayounfar [a],[1], Varut Vardhanabhuti [b], Mohammad-Ali Nikouei Mahani [c], Mohamad Koohi-Moghadam [a],[*]

[a] Applied Oral Sciences and Community Dental Care, The University of Hong Kong, Hong Kong S.A.R., PR China
[b] Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong S.A.R., PR China
[c] School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

ABSTRACT

Medical image segmentation is a key initial step in several therapeutic applications. While most of the automatic segmentation models are supervised, which require a well-annotated paired dataset, we introduce a novel annotation-free pipeline to perform segmentation of COVID-19 CT images. Our pipeline consists of three main subtasks: automatically generating a 3D pseudo-mask in self-supervised mode using a generative adversarial network (GAN), leveraging the quality of the pseudo-mask, and building a multi-objective segmentation model to predict lesions. Our proposed 3D GAN architecture removes infected regions from COVID-19 images and generates synthesized healthy images while keeping the 3D structure of the lung the same. Then, a 3D pseudo-mask is generated by subtracting the synthesized healthy images from the original COVID-19 CT images. We enhanced pseudo-masks using a contrastive learning approach to build a region-aware segmentation model to focus more on the infected area. The final segmentation model can be used to predict lesions in COVID-19 CT images without any manual annotation at the pixel level. We show that our approach outperforms the existing state-of-the-art unsupervised and weakly-supervised segmentation techniques on three datasets by a reasonable margin. Specifically, our method improves the segmentation results for the CT images with low infection by increasing sensitivity by 20% and the dice score up to 4%. The proposed pipeline overcomes some of the major limitations of existing unsupervised segmentation approaches and opens up a novel horizon for different applications of medical image segmentation.

1. Introduction

Medical image segmentation is the process of dividing an image into several sections to identify prognostic and clinical characteristics of the images. This is a crucial initial step in several therapeutic applications, such as tumor localization and lesion volume measurement. Several effective approaches for medical image segmentation have been reported in the literature, with the most notable advances of fully supervised convolutional neural networks (CNNs) methods. These methods train the model on paired data sets consisting of the original medical images and those annotated by a radiologist, and their results frequently outperform human experts [1–5]. However, one significant limitation of these approaches is the requirement of having a well-annotated dataset. This is not always possible, particularly in the medical image domain, where creating medical image datasets is expensive and time-consuming due to the sensitive nature of the data and the need for highly specialized domain expertise to annotate them. To overcome this limitation, the community has introduced novel unsupervised and weakly-supervised methods that do not require manual annotation at the pixel level.

Here, we aim to propose a dual-model framework for segmenting COVID-19 computed tomography (CT) scan without any annotation at the pixel level. The core idea is to use a 3D GAN to synthesize healthy images from the COVID-19 CT scans. The healthy images generated by GAN would then be subtracted from the original CT images to extract the 3D infected area. This approach allows us to generate pseudo masks without the need for human annotators. We then train a multi-objective segmentation network through pseudo masks and contrastive losses.

Listed below is a summary of the main contributions of our work:
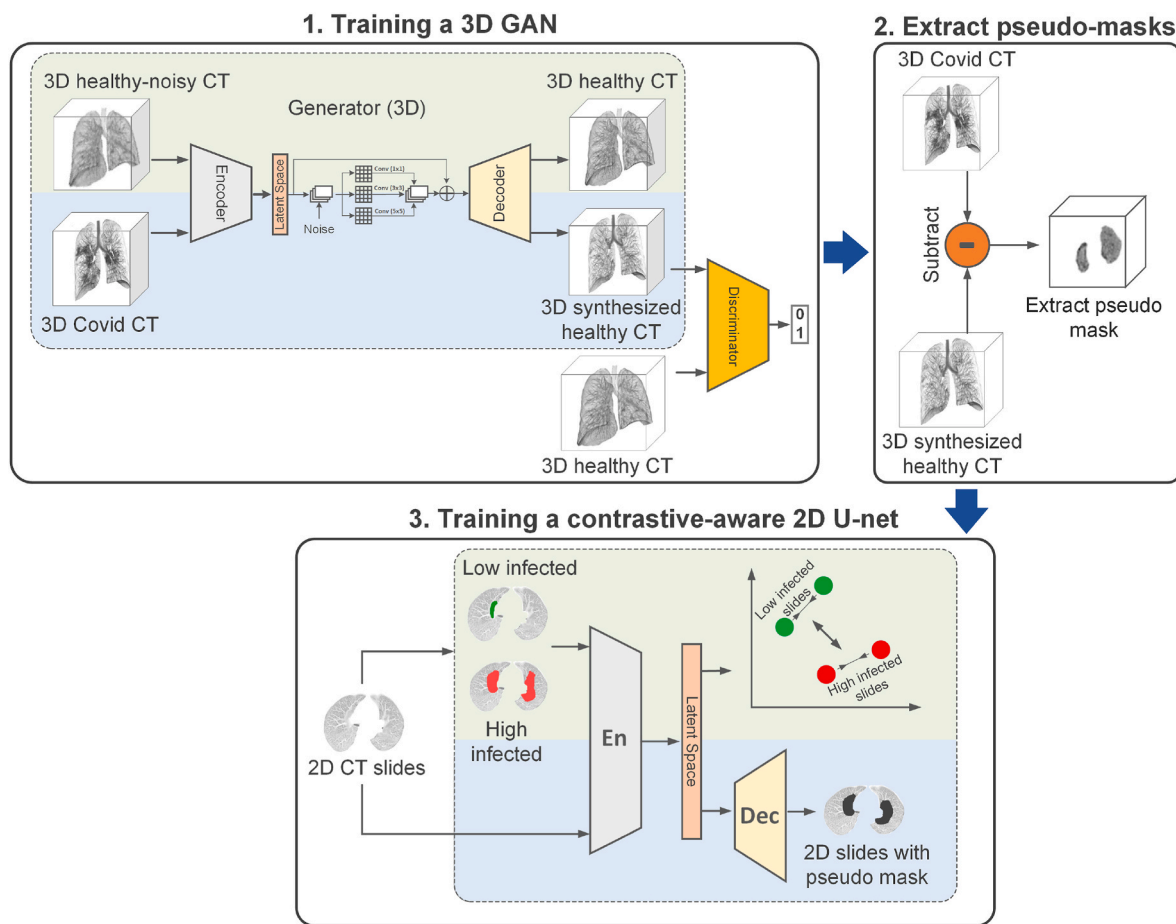

Fig. 1. Schematic of the proposed pipeline. The proposed pipeline consist of three phases: 1. Training a 3D GAN with a multi-objective loss: a standard GAN objective loss to remove infected regions from COVID-19 CT images (blue part), and a reconstruction loss for generation of the healthy images from healthy-noisy image (green part). The latter one regularizes the GAN in order to produce an image with the same lung structure as the input image. 2. Extract pseudo-masks: we extracted the 3D infected area using pixel-to-pixel subtraction of the synthesized healthy image from the original COVID-19 image. 3. A contrastive-aware 2D segmentation model: we trained a multi-objective 2D segmentation model using the extracted pseudo-masks from last previous step to predict infected areas. The 2D segmentation model has been trained jointly with an end-to-end pixel-wise MSE loss (blue part) and a contrastive loss (green part), which was applied only to the encoder part of the network.

- Our pipeline does not require any paired data or manual annotation at the pixel level for model training and validation.
- Our self-supervised GAN is robust enough to synthesize healthy images from COVID-19 CT scans to generate pseudo-mask. The proposed 3D GAN architecture could potentially be used in other medical applications to generate pseudo-paired data.
- We employ a contrastive loss to perform region-aware segmentation on the CT slices. Our results show that using contrastive loss enhances the quality of segmentation, particularly for the slices with low infection level.
- Our technique outperforms the state-of-the-art methods of unsupervised (without any real mask) and weakly supervised (which used mask partially in their training) segmentation approaches for COVID-19 CT segmentation by a reasonable margin.
- Our source code and data are publicly available on Github (htt ps://github.com/ClinicalAI/3DGAN_CT_Segmentation).

2. Background

Medical image segmentation. Computer-aided diagnosis (CAD) refers to automatic approaches that help doctors to interpret medical images. One of the popular approaches to provide an efficient CAD platform is to use deep learning for medical image segmentation. It allows the radiologists and other medical professionals to analyze and evaluate a large amount of data in a short period of time [6]. Image segmentation is used in several areas of medical image processing such as brain-tumor segmentation, liver-tumor segmentation, retina segmentation, and lung segmentation. Xue et al. (19) used U-Net as the generator part of a GAN and proposed a new multi-scale loss function to effectively segment brain tumors. Kim et al. (90) used Cycle-GAN to overcome the mode collapse phenomenon and improved the speed of segmentation for liver cancer images. However, the performance of their model is lower than supervised models. Lahiri et al. (39) segmented retinal images using GAN, where the generator maps the latent noise space to the realistic images, and discriminator was a multiclass classifier for the real versus fake images as well as vessel versus background images. Iqbal et al. (42) synthesized the retinal images and masks using a GAN with a complex loss function in order to produce a more realistic retinal image and improve segmentation performance. Tan et al. (65) implemented an EM distance-based loss function with a GAN to segment the lung from the input CT scans.

Abdel-Basset et al. [7] used a few-shot weakly supervised method to segment infection in COVID-19 CT slices. They employ a dual-path network consisting of two encoder-decoder networks linked by a recombination and recalibration block, as a limitation the performance of their model with ground glass opacity (GGO) is not the same as consolidation infection class and the performance of model is decreased with GGO infection class. Hu et al. [8] proposed a convolutional model that requires less annotated data. Using their multi-scale network, they distinguished COVID-19 patients from healthy individuals by focusing on infected lung regions. Laradji et al. present a pre-trained network that uses the maximum arguments in an uncertainty map, and they use active

learning to add new labels in the subsequent training phase, however their model required a large number of samples for training process [9]. Xu et al. [10] use a segmentor block in a GAN to segment lesions in COVID-19 CT images. Using this model, they generate pseudo-healthy images from COVID-19 CT images. For training the discriminator, the pseudo-healthy data is added to the healthy data, which enhanced the performance of the generator in removing lesion regions. Yang et al. [11] present a weakly-supervised model for federated learning. The model handles unannotated images in different clients for COVID-19 segmentation, however, the results of this model are lower than other similar studies. Zheng et al. [12] applied representation learning and k-mean clustering method on a private dataset to locate lung infected area. Yao et al. [13] develop a model for free-mask infection segmentation in COVID-19 CT-scan images. They created a noise generator that is used to generate the pseudo-infected data by adding the different types of noise to the healthy one. Since they did not use a mask in their model, the model's outputs are discontinuities, necessitating a post-processing step to enhance the final output. Xu et al. [14] develop an unsupervised model based on anomaly detection for segmenting the infected regions. It is worth pointing out that they used only healthy data to train the model. Ding et al. [15] proposed a teacher-student architecture that is trained on both annotated and unannotated images for segmenting infected areas on the Mosmed dataset. They used a noise-aware loss function to improve the performance of segmentation [16]. Liu et al. [17] came up with a two-path framework that includes both a student and a teacher. The teacher model is trained with imperfect masks, and then the student model is trained with perfect masks to perform segmentation precisely.

Contrastive learning for medical image segmentation. The basic idea behind contrastive learning is to use the contrastive loss to contrast the similarities of pairs of samples in the representation space by bringing positive pairs' representations together and pushing negative pairs' representations apart [18]. Several studies have utilized contrastive learning to improve the accuracy of classification medical image [19–22]. Zeng et al. [23] demonstrate how contrastive learning can be used to improve medical image segmentation accuracy when only a subset of the dataset has been annotated. Additionally, the correct selection of positive and negative pairs is essential for contrastive learning performance. Zeng and et al. [24] demonstrate that segmentation performance can be enhanced by utilizing follow-up X-ray images of patient as positive and others as negative pairs. Their strategy of paring data has improved the performance of two widely used contrastive learning frameworks, MoCo [25] and SimCLR [26].

3. Methods

The proposed model consists of three phases:

1-Training a multi-objective 3D GAN to transform the COVID-19 CT images into healthy CT images. In order to train the generator, the 3D GAN simultaneously employs two losses: the first loss for the mapping of healthy-to-healthy images and the second loss for generating semi-healthy images from infected images. The goal of the discriminator is to distinguish the real healthy CT images from the synthesized healthy images.

2 -Synthesizing pseudo masks by subtracting the generated healthy images from the original COVID-19 CT images. These pseudo masks emphasize the infected regions of CT images.

3-Training a contrastive-aware 2D U-Net [27] segmentation network using the generated pseudo masks from the previous phase.

The procedure of the model is illustrated in Fig. 1 and well explained in Algorithm 1. We will discuss all three training phases in detail in separate subsections.

3.1. 3D GAN

Given a CT scan image of a COVID-19 patient, denoted as $I_C \in$

$R^{W \times H \times L}$, the goal of the 3D GAN model is to synthesize a healthy 3D CT scan image with the same size, denoted as $\widehat{I}_H \in R^{W \times H \times L}$. We proposed an encoder-decoder model with 3D CNN based on DenseUNet [28] as the generator part, $G : (I_C \rightarrow \widehat{I}_H)$, where G is the 3D generator network. The discriminator is simply a dense 3D CNN with four layers. We had to address two main issues to build a robust 3D GAN: First, the model needs to be general enough to remove a wide variety of infections. Second, we observed that the generated healthy samples did not match the original lung structure in some cases. Therefore, we needed to enhance the 3D GAN in such a way that the structure of the synthesized healthy images is analogous to the original ones. To address the former issue, we used a noise adder operator in the latent space of our encoder-decoder model. The vanilla encoder-decoder 3D generator model was modified as follows:

$$\widehat{I}_H = G(I_I) = De\big(En(I_C) + U^N\big), \ En : I_C^{W \times H \times L} \rightarrow L^N, \ De : L^N \rightarrow \widehat{I}_H^{W \times H \times L} \quad (1)$$

where L^N represents the latent space, *De* indicates the decoder, *En* shows the encoder, and U^N is the noise operator. The noise adder operator perturbs the latent space and trains the generator to remove a varied set of infections. It employs different filter sizes to increase the variation in latent space. The input of the noise adder is a uniform noise between $[-0.5, 0.5]$. To address the second issue, we updated the model's weights by incorporating healthy-noisy images into the training phase. The healthy-noisy images were created by adding Perlin noise to the healthy 3D CT scans.

Each training iteration include two weights update steps: the first one with the standard GAN objectives (blue part) and the second to reconstruct healthy images from healthy-noisy images (green part). The proposed multi-objective training approach helps the model to remove the lung infections without affecting the lung's structure. In other words, we modified the standard GAN loss function by adding a new term which leads to regulating the generator to map each voxel/pixel to the same structure of the lung. The total loss function of the proposed 3D GAN is defined as below:

$$Max_D \ V(D, \ G) = E_{I_H \sim p_{healthy}}[logD(I_H)] + E_{I_C \sim p_{covid}}[log(1 - D(G(I_I)))] \\ + E_{I_H^{Pe} \sim p_{noisyHealthy}}\big[G\big(I_H^{Pe}\big) - I_{H2}\big] \quad (2)$$

where I_H indicates the real healthy 3D volume, and I_H^{Pe} is the healthy-noisy image with Perlin noise. By training the 3D GAN with both losses, the generator learns to synthesize a variety of analogous healthy outputs associated with the COVID-19 input images. The first two terms of Eq (2). show the standard GAN loss which its convergence has been shown by Kullback–Leibler divergence [29]. The second component of the loss function optimizes the reconstruction of healthy images from noisy-healthy images. We illustrated the performance of the proposed 3D GAN through the PCA visualization of the synthesized image (see Results).

3.2. Synthesized pseudo mask

The proposed 3D GAN model is capable of synthesizing the healthy CT images from the COVID-19 CT images and maps. The final goal of the segmentation model is to segment the infected areas of a given COVID-19 CT image. We generated the pseudo mask, $\widehat{M}$, for each COVID-19 sample by subtracting the synthesized healthy image from the original CT image:

$$\widehat{M} = I_I - \widehat{I}_H \quad (3)$$

The 3D GAN enables to generate as much pseudo mask samples as needed without requiring any real ground-truth data. Later, we employed the synthesized ground-truth data to train a segmentation model.

3.3. Segmentation network

Although we used a 3D GAN to generate pseudo mask, our final segmentation model is a 2D UNet model. A 2D segmentation model allows us to deal with CT images with different numbers of slices. Using this approach, we can ensure that our segmentation model is general enough to be compatible with a variety of CT image modalities. Furthermore, the 2D model has fewer trainable parameters, allowing for more efficient training.

Given a COVID-19 2D CT scan slice, $I_C^{W \times H}$, our model has to generate a mask of infected areas, $S^{W \times H}$. We used a 2D UNet encoder-decoder model:

$$UNet_{2D} : I_C^{W \times H} \rightarrow S^{W \times H} \tag{4}$$

$$S^{W \times H} = De_{2D}\left(En_{2D}\left(I_C^{W \times H}\right)\right) \tag{5}$$

where the En_{2D} indicates the encoder module, and De_{2D} shows the decoder module. The 2D UNet [30] is trained in a hybrid fashion. In each epoch, first the En_{2D} module is trained using the contrastive learning, then both En_{2D} and De_{2D} are fine tuned end-to-end with the pseudo synthesized masks from the previous phase.

Given a set of 2D CT slices $\{I_{Ci}\}$, we generated a list of positive and negative pseudo labels, $y_i \in \{y_{Cp}, y_{Cn}\}$ where y_{Cp} is the high infected and y_{Cn} shows the low infected CT slices. We determined these high and low sets by sorting slices according to the sum of their pseudo mask pixels counts. By using slices with high and low infection level as positive and negative samples, we tried to focus more on the infection area than other parts in our segmentation model.

The goal of contrastive learning is to train the encoder, $En_{2D} : I_C \rightarrow R^d$ such that it encodes each example into a feature vector of size d, where examples of the same class are close and examples of different classes are far from each other. Given the labeled slices, we incorporated a supervised contrastive loss in our segmentation model [31]. The supervised contrastive loss contrasts the positive samples from the same class to the set of negative samples from the rest of the batch. With the assumption of $z \in R^d$ as the projection of latent space, the contrastive loss function is defined as follows for all indexes of the dataset:

$$L_{cont}(i) = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(z_i \cdot \frac{z_p}{\tau}\right)}{\sum_{l \in C(i)} \exp\left(z_i \cdot \frac{z_l}{\tau}\right)} \tag{6}$$

For anchor i as an index of positive and l for a negative sample where $|P(i)|$ stands for the cardinality of $P(i) \equiv \{p \in C(i) : \tilde{y}_p = \tilde{y}_i\}$ by considering C(i) as all high and low infected COVID-19 indexes except the anchor. We set $\tau = 0.05$ for the constant temperature parameter. We calculated the summation of L_{cont} for all samples to train the contrastive loss. Our ablation study shows that incorporating contrastive loss significantly improves the performance of our segmentation model (see Results).

Algorithm 1: Procedure of our pipeline

Inputs:

- I_C = COVID-19 3D volume

- I_H = Health 3D volume

- I_p = High infected COVID-19 2D slices as positive images

- I_n = low infected COVID-19 2D slices as negative images

Abbreviations:

- G: 3D Generator

- D: 3D Discriminator

- U_{ed} : 2D UNet (U_e refers to encoder part of UNet only)

- CL: contrastive loss

for $Epochs_1$ **do:**

- freeze G, unfreeze D

- calculate generated healthy volume ($\hat{I}_H$) using $G(I_c)$

- update D using I_H and $\hat{I}_H$

$$\nabla_{\theta_D} E\left(\left[\log D(I_H) + \log\left(1 - D(\hat{I}_H)\right)\right]\right)$$

- freeze D, unfreeze G;

- update G using I_C

$$\nabla_{\theta_G} E\left(\log\left(1 - D(\hat{I}_H)\right)\right)$$

- add Perlin noise to healthy volumes I_H^{Pe}

- update G using MSE loss of pairs of I_H^{Pe} and I_h as - input and output respectively

$$\nabla_{\theta_G} E([\|G(I_H^{Pe}) - I_H\|_2])$$

for $Epochs_2$ **do:**

- freeze G and unfreeze U_{ed}

- calculate healthy volume (I_H) using G COVID inputs (I_C)

- extract the infections using subtraction $I_{inf} = I_C - I_H$

- update U_{ed} using MSE loss of pairs of I_c and I_{inf} as input and output respectively

$$\nabla_{\theta_{U_{ed}}} E\left(\left[\|B(I_C) - I_{inf}\|_2\right]\right)$$

- update U_e using the contrastive loss of pairs of positive and negative pairs

$$\nabla_{\theta_{U_e}} E([CL(I_p, I_n)])$$

Table 1
Data description.

Datasets	Task	Cases numbers	COVID/ Healthy	Average rate of infections per case (%)
MosmedData [32]	Train	254	Healthy	–
MosmedData [32]	Train	806	COVID	–
CoronaCases [28,33]	Validation	10	COVID	7.92
MosmedData [32]	Test	50	COVID	1.47
UESTC [35,36]	Test	120	COVID	5.67
Radiopeadia [31]	Test	9	COVID	10.33

Also, our 2D UNet model is trained with the simple pixel-wise Mean Squared Error (MSE) loss:

$$L_{MSE}(I_{Ci}, \widehat{S}_i) = \frac{1}{W \times H} \sum_{v=0}^{H} \sum_{t=0}^{W} [I_{Ci}(v, t) - \widehat{S}_i(v, t)]^2 \qquad (7)$$

where $(I_{Ci}, \widehat{S}_i)$ is the ith example with 2D input slice indicated by I_{Ci} and the corresponding output heatmap $\widehat{S}_i$. This loss has been applied in each epoch to train both encoder and decoder of the final 2D model. Our synthesized pseudo masks contain continuous values of the infection probability, rather than categorical labels of the infected areas. Therefore, we trained the 2D UNet based on the continuous values to benefit from the full range of information from pseudo masks rather than the simplified categorical labels. Using our validation dataset, we calibrated a simple threshold to generate the final segmentation result from the infection probability heatmap. The final output mask S_i, is the thresholded output of $\widehat{S}_i$. In each iteration, the encoder part of the model is trained with L_{cont}, then the whole model is trained end-to-end with L_{MSE}. The training process of the model is summarized in Algorithm1.

4. Experiments

We describe the implementation details of our model in this section and provide a detailed analysis using ablation studies. Then, we compare our model to state-of-the-art weakly and unsupervised models for COVID-19 CT segmentation. We obtain experimental results using Tensorflow with an NVIDIA A100 GPU.

4.1. Implementation details

Database and evaluation. To train, validate, and test our pipeline, we use four different public datasets, namely MosmedData [32], CoronaCases [33,34], UESTC [35,36], and Radiopeadia [37]. MosmedData has both healthy and COVID-19 3D CT images. In total, we used 254 healthy and 806 COVID-19 3D CT images to train the GAN from this dataset. We also used CoronaCases dataset to validate our model, which includes 10 annotated 3D CT images. Finally, we tested our model on 179 annotated 3D CT images from MosmedData, UESTC, and Radiopeadia datasets (Table 1). The average infection rate for each dataset is shown in Table 1. Radiopeadia is the most infected dataset, while MosmedData has the least amount of infection.

To evaluate the performance of our model, we use three metrics including, dice score (DSC) [38], specificity, and sensitivity, which are commonly used to evaluate image segmentation methods.

Preprocessing of CT images. In the first step, all the pixels spacing of all CT images is mapped to a constant value (1 mm for each direction). The Hounsfield Unit (HU) of each pixel is limited to the range (−1000,400), then normalized to the range (0,1) for all pixels. Using an open-source pre-trained UNet model [39], we extracted the lung area for each CT image. The image is then cropped to place the lung in the center of the image. Finally, we resize all the 3D volumes to a size (128, 128, 32) that is the desired shape for input into our model.

GAN network. The architecture of our 3D GAN consists of two networks, including a generator and a discriminator. We used a modified version of the UNet++ model [40] to implement our generator model. For the discriminator, we use a simple model consisting of five convolution layers and four dense layers. In addition, for a more robust result, we include a noise adder in the model (see Methods). The noise adder block consists of three different convolutional layers with one, three, and five filter sizes. We train our GAN for 10,000 iterations with a batch size of 8 and Adam optimizer with a learning rate of 5e-5. We removed the cases where the number of slices was lower than 32. Our 3D GAN model is trained using 226 healthy cases and 664 COVID-19 cases.

Segmentation network. Our segmentation model is a 2D UNet model [30], which consists of an encoder and a decoder model. We used MSE loss to train the encoder and decoder end-to-end. Also, the encoder was trained with an additional contrastive loss. To calculate MSE loss, we used the Adam optimizer with an initial learning rate of 0.001, a decay step of 1000, a decay rate of 0.9, and a batch size of 32 for 1000 epochs. We used the Adam optimizer with a learning rate of 0.0001 and a batch size of 32 for 1000 epochs to train the encoder model with contrastive loss.
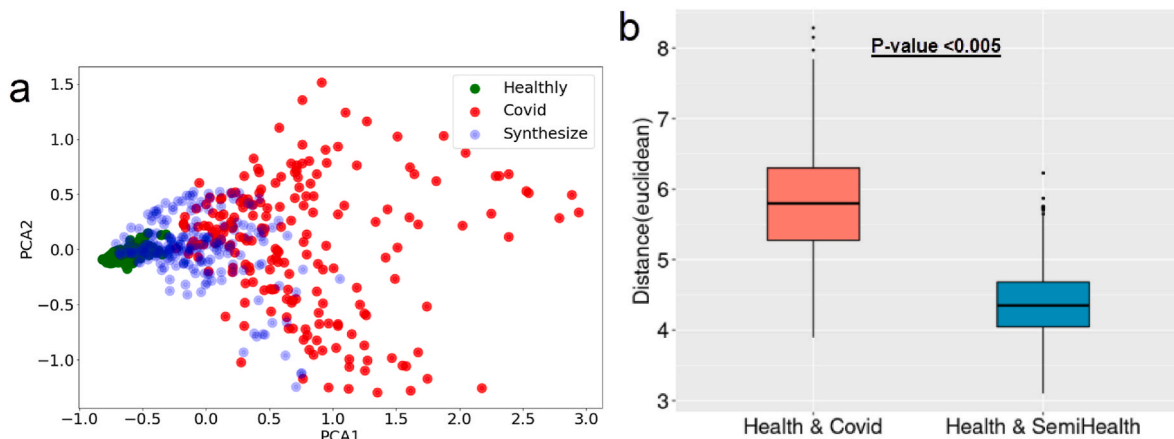

Fig. 2. Evaluate the performance of GAN model. (a) PCA to visualize 2D plot for COVID-19, heathy and synthesize images. The red dots show the original COVID-19 cases and the green ones are the original healthy ones. The blue ones are the healthy synthesized ones. (b) The mean and standard deviation of the distance between COVID-19, healthy and synthesized healthy images.

Table 2
The comparison of our model with similar studies.

		Score	UESTC	Radiopeadia	Mosmed
Weakly and self -supervised model	Our model	DSC	63.69 ± 17.5	61.71 ± 16.8	58.43 ± 19.0
		SPC	97.56 ± 2.7	97.64 ± 2.0	97.74 ± 2.4
		SEN	75.04 ± 15.8	67.21 ± 18.5	76.82 ± 16.7
	GAS net [10]	DSC	–	60.20 ± 23.3	54.20 ± 22.4
		SPC	–	99.30 ± 0.5	99.60 ± 0.2
		SEN	–	66.80 ± 28.9	55.60 ± 28.3
	Label-free [13]	DSC	61.40 ± 19.4	59.30 ± 16.9	–
		SPC	–	–	–
		SEN	77.60 ± 19.6	65.60 ± 18.7	–
Supervised model	COPLE-Net	DSC	83.90 ± 9.47	59.30 ± 17.7	–
		SPC	84.80 ± 10.5	57.90 ± 16.0	–
		SEN	84.60 ± 12.0	63.30 ± 23.0	–
	nnU-Net	DSC	81.60 ± 9.43	76.70 ± 5.81	–
		SPC	83.00 ± 11.7	77.10 ± 14.0	–
		SEN	81.70 ± 11.6	80.50 ± 13.1	–

4.2. Evaluation of GAN performance

To evaluate the performance of the GAN model, we used a 2D projection of the healthy (I_H), COVID (I_C) and synthesized healthy ($\hat{I}_H$) CT images. A Principle Component Analysis (PCA) dimensionality reduction method has been applied to these CT images. The plot shows that the class clusters formed by the synthesized healthy images overlap with the original healthy images while they are separated from the COVID-19 cases (Fig. 2a).

We also used a bootstrapping approach to calculate the Euclidean distance between the synthesized healthy, original healthy, and COVID-19 CT images. Our results show that the synthesized healthy images are closer to the healthy images in all three databases (Fig. 2b). These results show that our GAN successfully transformed the COVID-19 images to healthy data.

4.3. Quantitative comparison with state-of-the-art methods

In Table 2, we compare the performance of our approach with state-of-the-art methods like GASNet [10] and Label-free [13]. The results
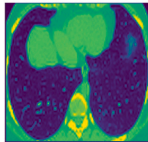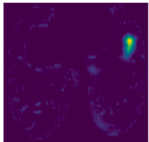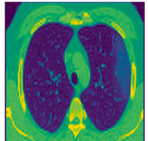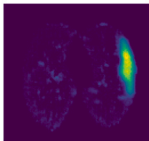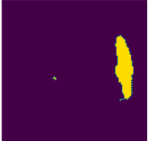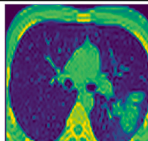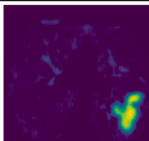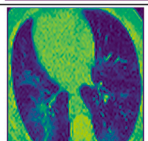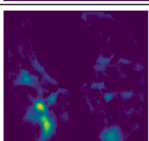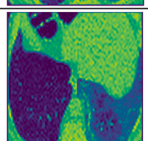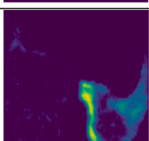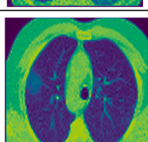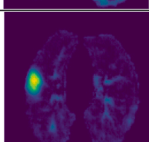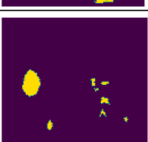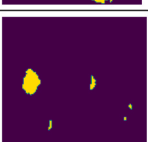
CT images	Output of model	Predicted mask	Real Mask	Dice score (%)
				87.67
				90.98
				90.44
				73.00
				82.44
				83.12

Fig. 3. Segmentation results of unseen slices. The first column is the original CT images and the second column is the output of our model. The predicted mask is shown in the third column after applying thresholding. The ground truth mask of each slice is presented in the last column.
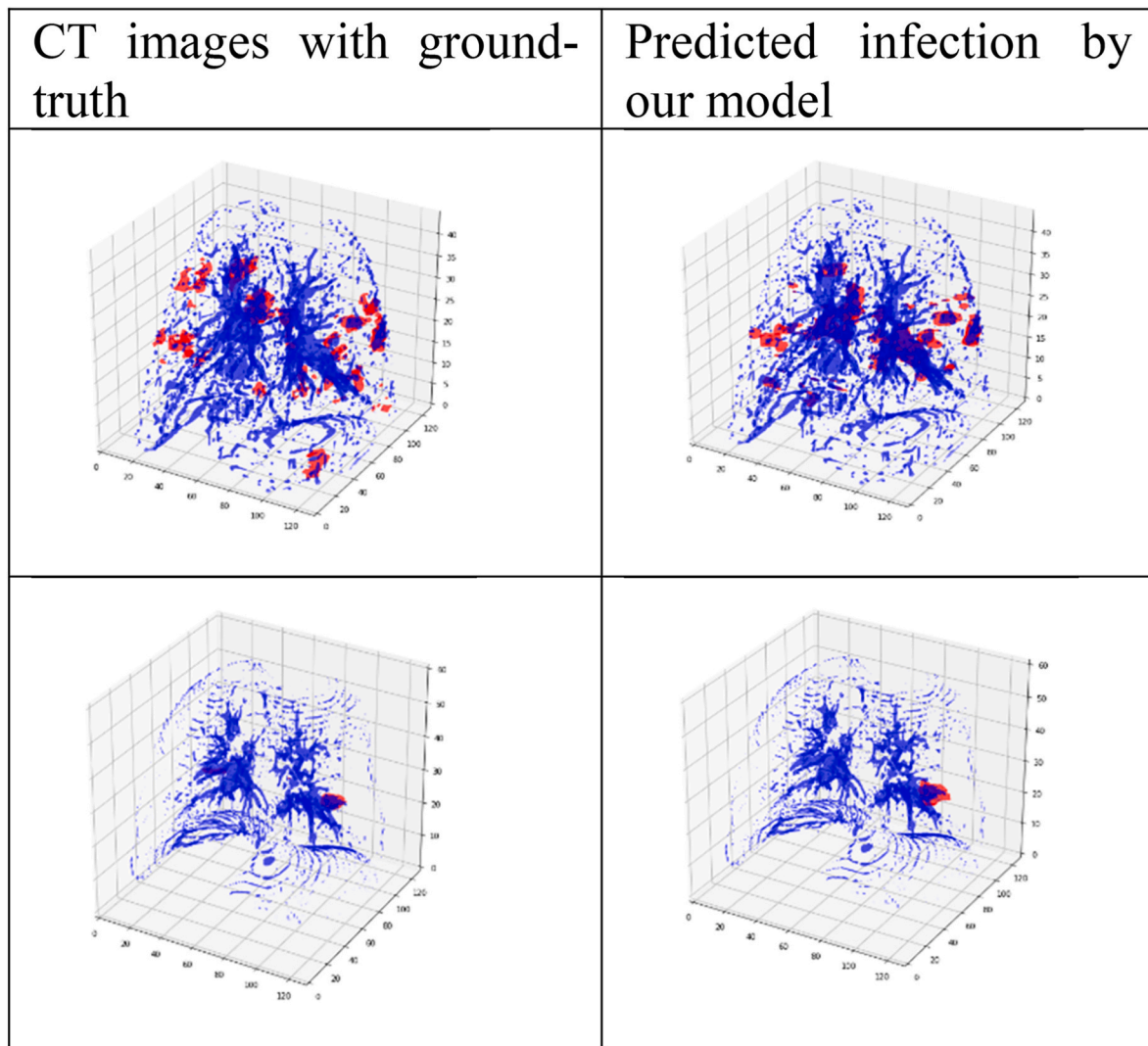
Fig. 4. 3D visualization of infected lung. The blue color is the lung, and the red color shows the infected areas. The predicted infection is compared with the ground truth.

Table 3
Ablation study on the quality of segmentation in different phase of our pipeline.

	UESTC			Radiopeadia			Mosmed		
	DSC (%)	SPC (%)	SEN (%)	DSC (%)	SPC (%)	SEN (%)	DSC (%)	SPC (%)	SEN (%)
GAN	34.9 ± 18.6	96.3 ± 2.6	42.2 ± 20.7	39.3 ± 19.3	96.6 ± 2.5	41.2 ± 22.6	41.2 ± 14.71	97.4 ± 1.3	66.3 ± 17.7
GAN + UNet	60.0 ± 16.3	**97.8 ± 2.2**	66.9 ± 15.0	**62.0.±15.9**	**97.8 ± 2.0**	64.9 ± 16.4	55.7 ± 15.5	**99.0 ± 0.7**	62.9 ± 14.4
GAN + UNet + Contrastive losses	**63.7 ± 17.5**	97.6 ± 2.7	**75.0 ± 15.8**	61.8 ± 16.8	97.6 ± 2.0	**67.2 ± 18.5**	**58.4 ± 19.0**	97.7 ± 2.4	**76.8 ± 16.7**

show that our model outperforms these models by a significant margin, without using any manual annotation at the pixel level. Our model achieves dice scores of 61.71%, 68.43%, and 63.69% for the Radiopeadia, Mosmed, and UESTC, respectively. Our result is better than GasNet and Label-free, which use weakly-supervised and manually generated masks to perform segmentation. Our method significantly improves in the sensitivity of segmentation for the Mosmed dataset, which is highly enriched with low infected CT images [32]. In this dataset, we have an increased sensitivity of 20% and a dice score of 4%. This demonstrates that incorporating of GAN and contrastive learning could be a promising method to perform annotation-free segmentation on COVID-19 infection, particularly in the CT images with low infection.

4.4. Qualitative evaluation

The results of segmentation for seven slices of the Radiopeadia and Mosmed databases are shown in Fig. 3. The results show that our model can almost accurately predict the infected area for various types of infections. We also generated the 3D images of the predicted infection area in DICOM CT files by concatenation of the slices (Fig. 4).

4.5. Ablation studies

Effectiveness of contrastive learning. We conducted an ablation study on our model to determine whether the contrastive loss has a positive impact on the final model results. We compare raw GAN results to segmentation model with and without contrastive loss. The results
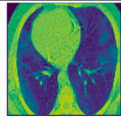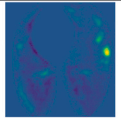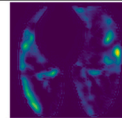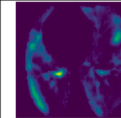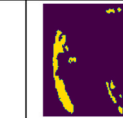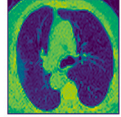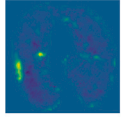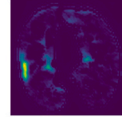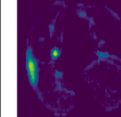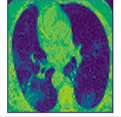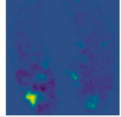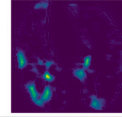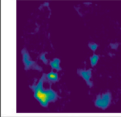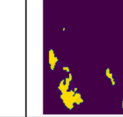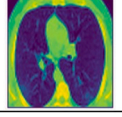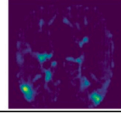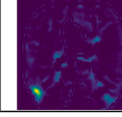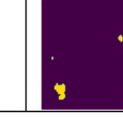
CT images	GAN	GAN+ UNet	GAN+UNet + Contrastive loss	Ground truth	GAN	GAN+ UNet	GAN+UNet+ Contrastive loss
					43.12	69.41	**76.12**
					52.51	55.7	**57.96**
					40.91	56.50	**60.59**
					40.38	48.11	**55.72**

Fig. 5. Segmentation dice score results of our pipeline in different stage. The first column shows the original CT images. The second to fourth columns show the results of the GAN, (GAN + UNet), and (GAN + UNet + Contrastive loss), respectively. The ground-truth mask for each slice is shown in the fifth column. The last three columns show the dice score results for each step output.

Table 4
Effectiveness of the different data pairs in contrastive loss.

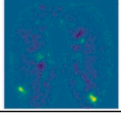|	UESTC			Radiopeadia			Mosmed		
	DSC (%)	SPC (%)	SEN (%)	DSC (%)	SPC (%)	SEN (%)	DSC (%)	SPC (%)	SEN (%)
Low infection vs high infection	63.69 ± 17.5	97.56 ± 2.7	75.04 ± 15.8	61.71 ± 16.84	97.64 ± 2.03	67.21 ± 18.49	58.43 ± 19.0	97.74 ± 2.4	76.82 ± 16.7
Healthy vs High infection	58.84 ± 16.9	98.21 ± 2.1	61.82 ± 18.9	61.50 ± 16.3	98.11 ± 2.06	61.21 ± 17.48	52.47 ± 20.01	98.92 ± 1.2	53.92 ± 22.4
Healthy vs low infection	58.57 ± 17.9	96.45 ± 3.5	74.98 ± 15.3	61.65 ± 18.9	96.89 ± 2.5	72.88 ± 18.4	52.60 ± 19.0	97.35 ± 2.6	72.50 ± 18.6

show that using UNet after GAN helps to improve the GAN output. Table 3 shows that using the UNet model improves dice scores compared to only the GAN model. Also, incorporating contrastive loss into our UNet helped to improve the segmentation accuracy for the Mosmed and UESTC datasets, although no improvement was obtained for the Radiopeadia dataset. In all three datasets, we found that contrastive loss improves our model's sensitivity significantly (Table 3). In Fig. 5, we show the effect of the contrastive loss in our segmentation model for some selected CT slices.

Effectiveness of selecting the positive and negative pairs on contrastive learning. As described in Section 3.2, in each iteration, the encoder part of the segmentation model is trained with L_{cont}, and then the entire model is trained end-to-end with L_{MSE}. To train the L_{cont}, we must use specific negative and positive pairs, which have a direct impact on contrastive learning performance. Here, we investigate different types of data pairing to train our contrastive loss. We examined three different pairs of data, namely, low-infection versus high-infection, healthy versus high-infection slices, and healthy versus low-infection. The results of each data pair are shown in Table 4. We found paired data as low-infection versus high-infection CT slices has the best segmentation performance.

Tuning of the threshold. We applied a fixed simple threshold to the

Table 5
Dice scores of four different values of threshold.

Threshold	0.1	0.15	0.2	0.25	0.3
Dice Score	46.91	59.83	**64.79**	64.56	62.28

segmentation network output to obtain the final mask. We used our validation dataset to tune the threshold value. Our result shows that T = 0.2 is the optimized value to achieve the best segmentation dice score in the validation dataset (Table 5).

5. Limitation

Our pipeline needs healthy data to train the GAN to generate a pseudo mask. This can be a potential limitation as the healthy data is not readily available for all medical applications. Healthy medical images are usually ignored in medical image processing projects as they do not have disease-related features. As a result, healthy data is typically excluded from clinical databases. However, our finding shows that healthy medical images can be potentially used in a generative based pipeline to train an annotation-free model for medical image segmentation. This may draw researchers and clinicians' attention to the importance of healthy data in future medical image processing projects.

Moreover, our pipeline is only compatible with medical images in the 3D DICOM format. It is one of the limitations, as the majority of common medical image databases only contain 2D images. With the advancement of medical equipment and the availability of cloud-based storage, we anticipate that more 3D DICOM-based databases will be released for different medical applications.

6. Conclusion

We propose a pipeline to perform self-supervised segmentation of the COVID-19 lesion. Our method outperformed previously published

weakly supervised and unsupervised models that made use of partial annotation or manual masks. Compared with two similar studies, we obtained a better result on the test dataset. Our improvement was more significant on the CT slices with low infection areas.

Although our GAN is 3D, we ultimately trained a 2D model for final CT slice segmentation. The presented 2D approach hast a number of advantageous: firstly, unlike 3D models, which can only be used with a CT scan of a fixed size, we do not need to change the size of the CT images by adjusting the number of slices to a fixed number in our 2D model. Secondly, since the processing time of a 2D model is less than that of a 3D model, our 2D approach allows clinicians to explore the CT slices quickly. Thirdly, the final model is capable of being retrained for other 2D datasets, such as X-ray images. This contributes to the generalisability of our model, which can be applied to any COVID-19 CT slice regardless of its 3D size. However, the performance of our model is lower than the supervised approaches, but we believe that by having enough training samples, the performance of our pipeline can be competitive with supervised models as well.

The output of our pipeline can be used to perform lesion labeling that can be refined later by radiologists. Moreover, our pipeline can be poetically used for medical image anomaly detection and lesion localization in different medical applications. Using this method, we will save the radiologist's time and energy, which will be especially useful during a pandemic. The application of proposed pipeline is not limited to COVID-19 CT segmentation, but can be extended to other medical image processing applications, such as MRI brain images and liver CT image segmentation.

Data availability

The code and database are publicly available in https://github.com/ClinicalAI/3DGAN_CT_Segmentation.

Author contributions

For the work described herein, S.S., M.H., and M.K. conceived the idea. S.S and M.H. implemented the codes. S.S., M.H. and V.V. prepare and validate the dataset. M.K., M.N. performed result validation. S.S., M. H., M.K. and M.N. wrote the paper. V.V. commented on and edited the manuscript. M.K. provided overall project leadership.

Declaration of competing interest

The Authors declare no Competing Financial or Non-Financial Interests.

References

[1] T. Zhou, S. Canu, S. Ruan, Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism, Int. J. Imag. Syst. Technol. 31 (1) (2021) 16–27.

[2] D.-P. Fan, et al., Inf-net: automatic covid-19 lung infection segmentation from ct images, IEEE Trans. Med. Imag. 39 (8) (2020) 2626–2637.

[3] T. Javaheri, et al., CovidCTNet: an open-source deep learning approach to diagnose covid-19 using small cohort of CT images, NPJ digital medicine 4 (1) (2021) 1–10.

[4] J. Liu, et al., COVID-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework, Med. Image Anal. 74 (2021), 102205.

[5] C.-F. Li, et al., MultiR-Net: a novel joint learning network for COVID-19 segmentation and classification, Comput. Biol. Med. (2022), 105340.

[6] M. Firmino, G. Angelo, H. Morais, M.R. Dantas, R. Valentim, Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy, Biomed. Eng. Online 15 (1) (2016) 1–17.

[7] M. Abdel-Basset, V. Chang, H. Hawash, R.K. Chakrabortty, M. Ryan, FSS-2019-nCov: a deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection, Knowl. Base Syst. 212 (2021), 106647.

[8] S. Hu, et al., Weakly supervised deep learning for covid-19 infection detection and classification from ct images, IEEE Access 8 (2020) 118869–118883.

[9] I. Laradji, et al., A Weakly Supervised Region-Based Active Learning Method for Covid-19 Segmentation in Ct Images, 2020 *arXiv preprint arXiv:2007.07012*.

[10] Z. Xu, et al., GASNet: Weakly-Supervised Framework for COVID-19 Lesion Segmentation, 2020 *arXiv preprint arXiv:2010.09456*.

[11] D. Yang, et al., Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan, Med. Image Anal. 70 (2021), 101992.

[12] T. Zheng, et al., Unsupervised segmentation of COVID-19 infected lung clinical CT volumes using image inpainting and representation learning, in: Medical Imaging 2021: Image Processing 11596, International Society for Optics and Photonics, 2021, 115963F.

[13] Q. Yao, L. Xiao, P. Liu, S.K. Zhou, Label-free segmentation of covid-19 lesions in lung ct, IEEE Trans. Med. Imag. 40 (10) (2021) 2808–2819.

[14] R. Xu, et al., Unsupervised detection of pulmonary opacities for computer-aided diagnosis of COVID-19 on CT images, in: 2020 25th International Conference on Pattern Recognition (ICPR, IEEE, 2021, pp. 9007–9014.

[15] W. Ding, M. Abdel-Basset, H. Hawash, RCTE: a reliable and consistent temporal-ensembling framework for semi-supervised segmentation of COVID-19 lesions, Inf. Sci. 578 (2021) 559–573.

[16] Z. Zhang, M. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, Adv. Neural Inf. Process. Syst. 31 (2018).

[17] X. Liu, et al., Weakly supervised segmentation of covid19 infection with scribble annotation on ct images, Pattern Recogn. 122 (2022), 108341.

[18] M. Kang, J. Park, Contragan: Contrastive Learning for Conditional Image Generation, 2020 *arXiv preprint arXiv:2006.12681*.

[19] Y. Zhang, H. Jiang, Y. Miura, C.D. Manning, C.P. Langlotz, Contrastive Learning of Medical Visual Representations from Paired Images and Text, 2020 *arXiv preprint arXiv:2010.00747*.

[20] Y.N.T. Vu, R. Wang, N. Balachandar, C. Liu, A.Y. Ng, P. Rajpurkar, MedAug: Contrastive Learning Leveraging Patient Metadata Improves Representations for Chest X-Ray Interpretation, 2021 *arXiv preprint arXiv:2102.10663*.

[21] B. Li, Y. Li, K.W. Eliceiri, Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14318–14328.

[22] S. Azizi, et al., Big Self-Supervised Models Advance Medical Image Classification, 2021 *arXiv preprint arXiv:2101.05224*.

[23] D. Zeng, et al., Positional Contrastive Learning for VolumetricMedical Image Segmentation, 2021 *arXiv preprint arXiv:2106.09157*.

[24] D. Zeng, J.N. Kheir, P. Zeng, Y. Shi, Contrastive Learning with Temporal Correlated Medical Images: A Case Study Using Lung Segmentation in Chest X-Rays, 2021 *arXiv preprint arXiv:2109.03233*.

[25] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9729–9738.

[26] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.

[27] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[28] Y. Cao, S. Liu, Y. Peng, J. Li, DenseUNet: densely connected UNet for electron microscopy image segmentation, IET Image Process. 14 (12) (2020) 2682–2689.

[29] I. Goodfellow, et al., Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[30] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-directional ConvLSTM U-Net with densley connected convolutions, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, 0-0.

[31] P. Khosla, et al., Supervised Contrastive Learning, 2020 *arXiv preprint arXiv: 2004.11362*.

[32] S.P. Morozov, et al., Mosmeddata: data set of 1110 chest ct scans performed during the covid-19 epidemic, Digital Diagnostics 1 (1) (2020) 49–59.

[33] Jun Ma, et al., Towards Efficient Covid-19 Ct Annotation: A Benchmark for Lung and Infection Segmentation, 2020.

[34] Omir Antunes Paiva, Coronacases database (2020) [online] Available: https://coronacases.org/, 2022. (Accessed 2 February 2022).

[35] G. Wang, et al., A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images, IEEE Trans. Med. Imag. 39 (8) (2020) 2653–2663.

[36] UESTC-COVID-19 dataset [online], https://faculty.uestc.edu.cn/HiLab/en/article/379152/list/index.htm.

[37] Frank Gaillard, Radiopeadia database (2022) [online] Available: https://radiopaedia.org/. (Accessed 2 February 2022).

[38] R.R. Shamir, Y. Duchin, J. Kim, G. Sapiro, N. Harel, Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations, 2019 *arXiv preprint arXiv: 1906.11031*.

[39] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, G. Langs, Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem, European Radiology Experimental 4 (1) (2020) 1–13.

[40] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: a nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 3–11.