International Journal of Pattern Recognition and Artificial Intelligence
© World Scientific Publishing Company

A 3D GAN and TransUNet based approach for Self-supervised region-aware segmentation of COVID-19 CT scans

Adam M. KHALI , Aymane DHIMEN , Aymane EL FAHSI, EL ANSARI Mostapha , Selma KOUDIA

Ecole Centrale of Casablanca
adam.khali@centrale-casablanca.ma
aymane.dhimen@centrale-casablanca.ma
aymane.elfahsi@centrale-casablanca.ma
mostapha.elansari@centrale-casablanca.ma
selma.koudia@centrale-casablanca.ma

BANOUAR Oumayma

Faculty of Sciences and Technics, Cady Ayyad University
o.banouar@uca.ac.ma

Abstract: Segmenting medical images is paramount in various medical applications, catering to diagnostic purposes, disease identification, and guidance for medical procedures such as surgery. This study introduces groundbreaking contributions in constructing a self-supervised model for COVID-19 image segmentation. Initially, a memory-efficient GAN (Generative Adversarial Network) is employed to systematically enhance feature resolutions. This entails training the model to craft a 3D pseudo-mask by subtracting synthesized healthy images from original COVID-19 CT scans. Subsequently, our approach incorporates a TransUNet model, seamlessly integrating a transformer and a U-Net, for lesion prediction. The GAN significantly contributes by generating an extensive array of synthetic lung images, thereby amplifying the segmentation model's performance. Concurrently, the TransUNet outperforms the U-Net in segmentation tasks owing to its integration of transformer layers, facilitating efficient processing of large images while capturing intricate details. We tested our new approach on the Mosmed dataset, and the dice score, sensitivity and specificity scores were much better compared to using only UNET for the segmentation.

Keywords: Medical Image Segmentation, COVID-19 CT Scans, Generative Adversarial Networks (GAN), TransUNet, UNet, Self-Supervised Learning, 3D Image Processing

Introduction

Medical image segmentation [1] is an indispensable component of clinical analysis, involving the identification and categorization of specific regions within images. However, a significant challenge in this field is the demand for a well-annotated dataset. This issue is particularly pronounced in medical imaging, where the creation of datasets is both resource-intensive and time-consuming due to the sensitive nature of the data and the requirement for highly specialized domain expertise in annotation. To address this limitation, researchers have introduced innovative unsupervised and weakly supervised methods, eliminating the need for manual pixel-level annotation.

Unsupervised methods [2] leverage pretext tasks to generate supervisory signals, enabling the model to learn representations without relying on annotated pixel-level labels. In the context of medical image segmentation, self-supervised techniques harness intrinsic data properties, enhancing model generalization and reducing reliance on extensively labeled datasets. Semi-supervised techniques [3] bridge the gap between fully supervised and unsupervised learning by combining labeled and unlabeled data. In medical image segmentation, utilizing a smaller set of annotated data alongside a larger pool of unlabeled data enhances model performance and scalability, overcoming challenges in obtaining extensive annotations within the medical domain. While offering improved model performance and reduced labeling efforts, this approach introduces challenges such as label uncertainty, label bias, and domain-specific complexities. Weakly-supervised methods [4] alleviate annotation burdens by relying on less detailed or partial annotations, often using image-level annotations or rough labels. This reduces annotation efforts while guiding the model to infer accurate segmentations. Despite the benefits of reduced annotation cost and time and the ability to leverage larger datasets, challenges such as incomplete information, noise, ambiguity, and domain-specific considerations persist.

Transformers [5] , built upon attention mechanisms, have emerged as powerful tools. In various fields, including image segmentation, their capability to capture long-range dependencies and contextual information makes them promising for medical image analysis. This study explores the integration of Transformers within the segmentation framework, leveraging their attention mechanisms to enhance the model's understanding of relationships between different image regions, potentially improving segmentation accuracy and robustness.

Our proposed pipeline adopts a pioneering 3D GAN architecture to generate a pseudo-mask through a self-supervised process [6]. This process involves eliminating infected regions from COVID-19 images and generating synthetic healthy images while preserving the 3D structure of the lung. The 3D pseudo-mask is subsequently created by subtracting synthesized healthy images from the original COVID-19 CT scans. We further refine pseudo-masks using a contrastive-aware learning approach, constructing a region-aware segmentation model that prioritizes the infected area. The final segmentation model demonstrates its efficacy in predicting lesions in COVID-19 CT images without the need for manual annotation at the pixel level.

Our approach undergoes meticulous validation, surpassing existing unsupervised and weakly-supervised segmentation techniques across selected data from the version of the Mosmed dataset we used. Notably, we achieve substantial enhancements in segmentation results for CT images with low infection (50 validation cases with their corresponding ground-truth lesion region masks), augmenting sensitivity mean among validation set up to 98.52% and the specificity score up to 99.99% against GAN+UNET respective 72.89% and 99.63%. The proposed pipeline adeptly overcomes major limitations associated with existing unsupervised segmentation approaches, highlighting its potential for diverse applications in medical image segmentation and heralding new horizons in this field.

1. Related Works

Medical image segmentation stands as a crucial domain within computer vision, captivating the attention of numerous research works.

Several studies have explored self-supervised learning techniques in the context of predicting the COVID-19 status of CT lung images. Chen et al[7] introduced a framework employing contrastive self-supervised learning, incorporating three key components: data augmentation, representation learning, and few-shot classification. Their data augmentation strategy involved cropping two segments from CT lung images. Additionally, representation learning aimed to enhance the similarity score, where cropped images from the same CT lung image yielded higher scores, while those from different CT lung images obtained lower scores.

In recent years, there has been widespread exploration of GAN-based unsupervised generation frameworks in medical image processing. GANs offer data enhancement methods to address the challenge of fragmented datasets. Han et al[8] introduced a three-dimensional multi-condition GAN (MCGAN) capable of generating authentic and diverse nodules, naturally placing them on lung CT images. This approach helps mitigate the impact of variations in lesion locations, sizes, and attenuation. Previous research [9] [10] indicates that unsupervised GANs alleviate the need for labor-intensive data labeling tasks, showcasing significant potential applications. These applications extend to various medical scenarios, including retinal images [11] , skin lesion images , pulmonary nodules , among others. The use of GAN-generated images of pulmonary nodules has demonstrated the potential to enhance radiologists' diagnostic efficiency. Zhao C et al [12] proposed a patched 3D U-NET and context convolutional neural network (CNN) that can automatically segment and classify pulmonary nodules. The application of GAN-enhanced model training further improves the performance of nodule image extraction.

Deep neural networks (DNNs) have shown excellent performance for many automatic image segmentation tasks. The U-Net architecture, introduced by Ronneberger et al [13] , relies on the encoder-decoder structure commonly employed in medical image segmentation for its impressive performance. Through the incorporation of skip connections, it establishes a link between the high-level, low-resolution semantic feature map and the low-level, high-resolution structural feature map of both the encoder and decoder. This integration enhances the spatial resolution of the network output. Oktay et al[14] introduced the attention gate model to the U-Net, resulting in improved sensitivity and prediction accuracy without a notable increase in computational cost. UNet++ [15] building upon the U-NET framework, utilizes a set of nested and dense skip paths to connect the encoder and decoder sub-networks. This modification further diminishes the semantic gap between the encoder and decoder, leading to enhanced performance in liver segmentation tasks. Several research groups reduce manual delineation time, utilizing noisy labels, and implementing semi-supervised learning. VB-Net [16] demonstrates remarkable effectiveness in segmenting COVID-19 infection regions. The mean percentage of infection (POI) estimation error for automatic segmentation versus manual segmentation on the verification set is a mere 0.3 Wang et al. [17] introduced the noise-robust Dice loss and applied it in COPLENet, surpassing other anti-noise training methods in learning COVID-19 pneumonia lesion segmentation with noisy labels. Inf-Net [18] employs a parallel partial decoder to aggregate high-level features and generate a global map to enhance the boundary area. Additionally, it utilizes a semi-supervised segmentation framework, achieving excellent performance in lung infection area segmentation.

In [6] ,authors presented a pipeline that comprises three main subtasks: automatically generating a 3D pseudo-mask in a self-supervised mode using a generative adversarial network (GAN), leveraging the quality of the pseudo-mask, and constructing a multi-objective segmentation model to predict lesions. The proposed 3D GAN architecture removes infected regions from COVID-19 images, generating synthesized healthy images while preserving the 3D structure of the lungs. Subsequently, a 3D pseudo-mask is generated by subtracting the synthesized healthy images from the original COVID-19 CT images. The authors enhanced pseudo-masks using a contrastive learning approach to construct a region-aware segmentation model that focuses more on the infected area. The final segmentation model can predict lesions in COVID-19 CT images without any manual annotation at the pixel level. The authors demonstrate that this approach outperforms existing state-of-the-art unsupervised and weakly-supervised segmentation techniques on three datasets by a significant margin. Specifically, their method improves the segmentation results for CT images with low infection, increasing sensitivity by 20% and the dice score by up to 4%.

Our proposed pipeline showcases also a remarkable effectiveness in segmenting COVID-19 infection regions by conserving the 3D GAN architecture and substituting UNet with TransUNet. The integration of TransUNet into the segmentation phase represents a key enhancement in their approach. By incorporating the TransUNet architecture, we aim to achieve highly accurate segmentation of COVID-19-infected regions. TransUNet's capabilities are particularly highlighted for effectively delineating these regions with exceptional accuracy.

2. Proposed approach for self supervised segmentation of Covid 19 CT scans

Our proposed approach focuses primarily on enhancing the segmentation model developped in [19], leveraging the state-of-the-art TransUNET architecture. While the three-phase pipeline of the original research article provided a valuable starting point, we refined it and augment the segmentation component to achieve more accurate and robust results by:

(1) Using TransUNET for segmentation.
(2) Enhancing 3D GAN training through the integration of noisy healthy images for pseudo-healthy generation (using Perlin Noise).
(3) Optimizing hyperparameters throughout the training phases.

Our proposed approach is as follows:

- **Phase 1: 3D GAN for Synthesizing Pseudo-Masks**
  - ⋆ Train a multi-objective 3D GAN to transform COVID-19 CT images into healthy CT images.
  - ⋆ Employ two losses simultaneously: one for mapping healthy-to-healthy images and another for generating semi-healthy images from infected ones.
  - ⋆ Distinguish real healthy CT images from synthesized healthy images.
- **Phase 2: Synthesizing Pseudo Masks**
  - ⋆ Synthesize pseudo masks by subtracting the generated healthy CT images from the original COVID-19 CT scans.
  - ⋆ Pseudo masks emphasize the infected regions within the CT images.
- **Phase 3: Segmenting the CT scans using TransUNET**
  - ⋆ Adapt the segmentation phase by incorporating the TransUNET architecture for precise COVID-19 CT image segmentation.
  - ⋆ Utilize the generated pseudo masks from the previous step as input for TransUNET.
  - ⋆ Highlight TransUNET's capabilities in effectively segmenting COVID-19-infected regions with exceptional accuracy.

The complete workflow of our pipeline is visually depicted in Figure 1, and a detailed algorithmic explanation can be found in Algorithm 1. These three phases provide a comprehensive foundation for our approach.

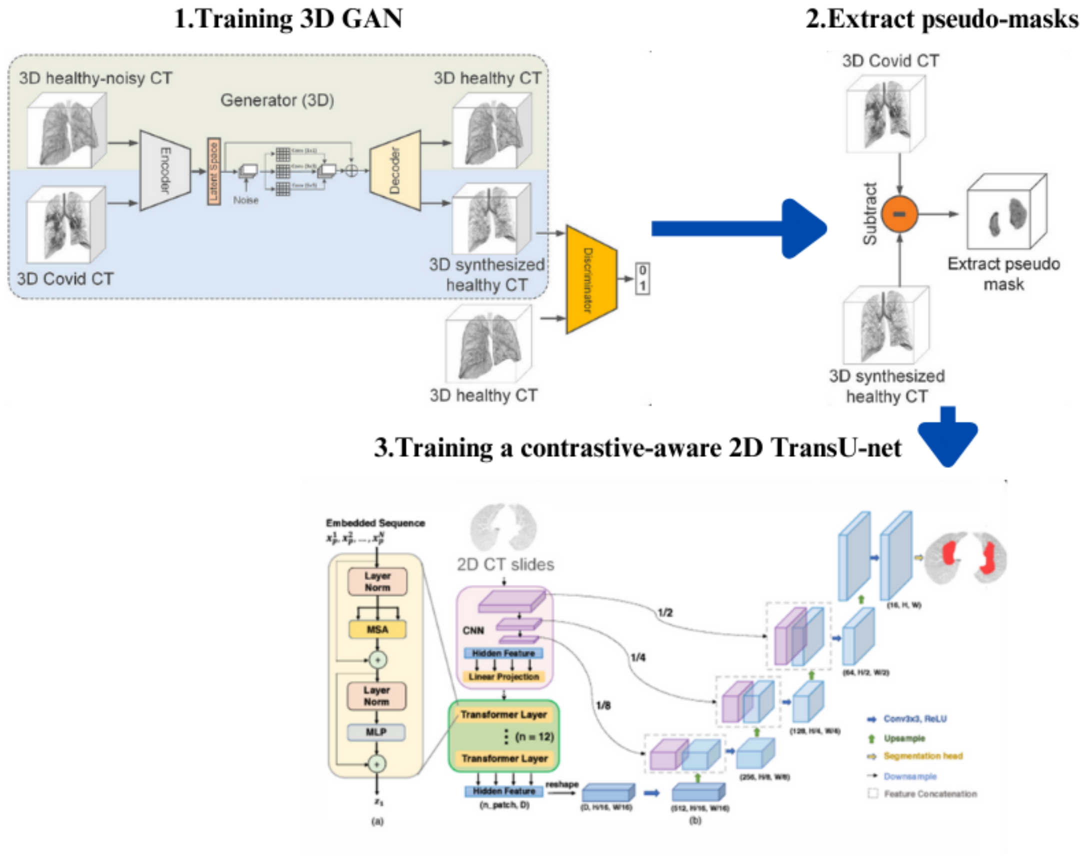


Fig. 1. Our Pipeline: GAN + contrastive-aware TransUNET Model

3. Methods

3.1. *Training the 3D GAN Model*

In our study, we used a 3D Generative Adversarial Network (GAN) to transform CT scan images of lungs affected by COVID-19 into their healthy counterparts, following the same procedure as in [19]. The objective is to generate a healthy 3D CT scan image, denoted as $\hat{I}_H \in \mathbb{R}^{W \times H \times L}$, from a given COVID-19 patient's CT scan, represented as $I_C \in \mathbb{R}^{W \times H \times L}$.

The encoder-decoder framework is using a 3D CNN based on DenseUNet as the generator part, $G : (I_C \rightarrow \hat{I}_H)$. The discriminator is designed as a dense 3D CNN with four layers. To build an effective 3D GAN, two main challenges were adressed: ensuring the model's generality for various infections and maintaining the original lung structure in the generated images.

Algorithm 1 Enhanced TransUNET Segmentation Pipeline Incorporating Transformers

```
Inputs: $I_c$: COVID-19 3D volume, $I_h$: Health 3D volume
Input: $I_p$: High infected COVID-19 2D slices, $I_n$: Low infected COVID-19 2D slices
Initialize: $G$: 3D Generator, $D$: 3D Discriminator
Initialize: $U_{ed}$: TransUNET (UNet with Transformer modules), $U_e$ refering to the encoder-part only
Initialize: $SEL$: Sensitivity Enhanced Loss, $ECL$: Enhanced Contrastive Loss
for each epoch in first training phase (3D GAN G-D Model training) do
    Freeze $G$, Unfreeze $D$
    Calculate generated healthy volumes $\hat{I}_H$ using $G(I_c)$
    Update $D$ using $I_H$ and $\hat{I}_H$, by calculating adversarial loss for $D$ as
        $\nabla_{\theta_D}\mathbb{E}[(\log D(I_h) + \log(1 - D(\hat{I}_h)))]$
    Freeze $D$, Unfreeze $G$
    Update G using $I_c$, by calculating adversarial loss for $G$ as
        $\nabla_{\theta_G}(\log(1 - D(\hat{I}_h)))$
    Add Perlin Noise to the healthy CT volumes to obtain $I_H^{Pe}$
    Update $G$ using MSE loss of pairs of $\hat{I}_h^{Pe}$ and $I_h$ as input and output respectively
        $\nabla_{\theta_G}\mathbb{E}[(\|G(\hat{I}_h^{Pe}) - I_h\|_2)]$
end for
for each epoch in second training phase(2D Segmentation Model training) do
    Perform forward pass of the batch through TransUNet
    Calculate the contrastive loss
    Calculate the mean squared error (MSE) loss
    Calculate the sensitivity-enhanced loss
    Combine losses and perform backpropagation
    Update model weights with the optimizer
    Periodically plot loss graphs and sample segmentations
    Save model checkpoints
end for
Output: Segmented infection volumes
Output: Updated $G$ and $U_{ed}$ parameters
```

To adress the first challenge, the model's architecture was modified to include a noise adder (Perlin Noise) operator in the encoder-decoder's latent space, which is represented as follows:

$$\hat{I}_H = G(I_C) = De\left(En(I_C) + U_N\right), \tag{1}$$

where $En : I_C \in \mathbb{R}^{W\times H\times L} \to L_N$ symbolizes the encoder, $De : L_N \to \hat{I}_H \in \mathbb{R}^{W\times H\times L}$ is the decoder, and U_N is the noise operator. The noise adder perturbs the latent space to train the generator to remove a wide variety of infections.

To address the second challenge, we updated the model's weights by incorporating healthy-noisy images into the training phase. These images were created by adding Perlin noise to healthy 3D CT scans. The training process includes two weight update steps: one with the standard GAN objectives and the second for reconstructing healthy images from healthy-noisy images. This approach ensures that the lung structure remains intact in the synthesized images.

The loss function for the proposed 3D GAN is defined as:

$$\begin{aligned} \max_D V(D,G) = & \mathbb{E}_{I_H \sim p_{\text{healthy}}}[\log D(I_H)] \\ & + \mathbb{E}_{I_C \sim p_{\text{covid}}}[\log(1 - D(G(I_C)))] \\ & + \mathbb{E}_{I_H^{Pe} \sim p_{\text{noisy\_Healthy}}}[G(I_H^{Pe}) - I_{H_2}], \end{aligned} \tag{2}$$

where I_H denotes the real healthy 3D volume, and I_H^{Pe} is the healthy-noisy image with Perlin noise. By training with both losses, the 3D GAN learns to generate a variety of healthy outputs corresponding to the COVID-19 input images.

We trained 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 51 healthy cases and 100 COVID-19 cases.

Algorithm 2 Adam Optimizer

```
Input: Learning rate $\alpha$, $\beta_1$, $\beta_2$, $\epsilon$, Initial parameters $\theta_0$
Initialize $m_0 \leftarrow 0$, $v_0 \leftarrow 0$, $t \leftarrow 0$
while stopping criterion not met do
    $t \leftarrow t + 1$
    Compute gradient $g_t$ using the current minibatch
    $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$    ▷ First moment estimate
    $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$    ▷ Second moment estimate
    $\hat{m}_t \leftarrow \frac{m_t}{1-\beta_1^t}$    ▷ Bias-corrected first moment estimate
    $\hat{v}_t \leftarrow \frac{v_t}{1-\beta_2^t}$    ▷ Bias-corrected second moment estimate
    $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon}$    ▷ Parameter update
end while
```

3.2. *Masks Extraction*

In this section, we describe the process of masks extraction for the segmentation phase. Masks extraction is a crucial step in the development of our approach for COVID-19 CT image segmentation.

The primary objective of the segmentation model is to precisely delineate the regions of infection in a given COVID-19 CT image. To achieve this, we employ the pseudo mask, denoted as $\hat{M}$, which is computed as follows:

$$\hat{M} = I_C - \hat{I}_H \tag{3}$$

Where: - I_C represents the COVID-19 CT image, which is an entry of the trained 3D GAN model generator. - $\hat{I}_H$ represents the output of the 3D GAN, consisting of synthesized pseudo-healthy images.

The subtraction operation results in $\hat{M}$, the pseudo mask, which highlights the areas of infection in the original COVID-19 CT image. This pseudo mask serves as a crucial tool in training our segmentation model later on.

One of the key advantages of our proposed 3D GAN model is its capability to generate synthetic pseudo masks, which are essential for training the segmentation model. Unlike traditional methods that require real ground-truth data for segmentation, our approach eliminates the need for such data. Instead, we leverage the synthesized ground-truth data, including pseudo masks, to train a highly effective segmentation model.

3.3. *Training the 2D TransUNET-based Segmentation Model*

Our TransUNET implementation combines a modified U-Net with a Transformer model. The U-Net provides local context through its encoder-decoder structure, while the Transformer captures long-range dependencies, essential for comprehensive CT scan analysis. Our implementation is an adaption of [20] and [21] to our specific case. The proposed workflow is the following:

3.3.1. *Input Preprocessing and Patch Embedding*

In preparing for our specific use case, we adapted the preprocessing of CT images to meet the requirements of the Transformer model. Each CT image was segmented into fixed-sized patches and subsequently linearly embedded. This preparation was critical to align the image data with the fixed-size input format required by the Transformer layers.

3.3.2. *Encoder-Decoder TransUNET Architecture*

Training Strategy for the TransUNET Model

The training process of our TransUNET model is a complex and carefully orchestrated procedure, designed to maximize the model's performance in segmenting COVID-19 infections in lung CT scans.

- **Model Configuration:** We configured the TransUNET with specific parameters, including image dimensions, channel input, and the structure of the Transformer and U-Net components. This setup was optimized for handling single-channel lung CT images.
  - **Encoder:** The encoder performs down-sampling operations to deepen the feature maps while reducing their spatial dimensions. This process is crucial for extracting complex features from the CT images, which are vital for a detailed analysis of the pulmonary structures.
  - **Decoder:** The decoder inversely mirrors the encoder's operations. By progressively up-sampling the feature maps and fusing them with the corresponding encoder outputs, the decoder retains essential spatial information necessary for precise segmentation.
- **Training Procedure:** We utilized a multi-phase training approach:
  - Initially, the model was trained on contrastive tasks to effectively distinguish between different infection levels.
  - The model then underwent end-to-end training, incorporating all loss functions to refine its segmentation capabilities.

  This approach ensured comprehensive learning, addressing both the global context and the nuanced details of lung CT scans.
- **Training Loop:** The training loop involved processing batches of data, calculating each of the loss components, and updating the model's parameters. We meticulously tracked the performance over epochs and adjusted the learning rate to optimize the training process.
- **Model Evaluation and Saving:** The model's performance was evaluated at regular intervals, allowing us to monitor its progress and make necessary adjustments. The best-performing model configurations were saved for future use and analysis.
- **Visualization and Monitoring:** Throughout the training process, we employed visualization techniques to monitor the loss trends and inspect the segmentation results on sample images. This allowed us to gain insights into the model's learning dynamics and its ability to segment lung CT scans effectively.

This training strategy, combining advanced loss functions and a meticulously designed training loop, was pivotal in developing a TransUNET model that is highly effective in segmenting lung CT scans for COVID-19 diagnosis.

Our advanced segmentation model employs two novel loss functions, each meticulously crafted to enhance the model's performance in segmenting COVID-19 infections in lung CT scans. Here, we present the mathematical formulations of these loss functions.

Enhanced Contrastive Loss The Enhanced Contrastive Loss function is designed to distinguish effectively between different levels of infection. It is mathematically formulated as follows:

$$L_{\text{contrastive}} = -\log \frac{\sum_{i=1}^{N} \exp\left(\frac{\mathbf{f}_i \cdot \mathbf{f}_p}{\tau}\right)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \exp\left(\frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\tau}\right)} \tag{4}$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ are the flattened and normalized feature vectors of the i^{th} and j^{th} samples, respectively, $\mathbf{f}_p$ represents the feature vector of a positive sample, τ is the temperature parameter, and N is the number of samples.

When determining the optimal value for τ, stochastic gradient descent can be employed:

Initialization:

$$\tau_0 \leftarrow \text{initial guess for } \tau$$

$$t \leftarrow 0$$

Optimization Loop:

While a (approximate) minimum is not reached: Generate a random permutation $\sigma = (\sigma_1, \ldots, \sigma_n)^T$ of $\{1, \ldots, n\}$

$$\text{for } b = 1, \ldots, B :$$

$$\tau_{t+1} \leftarrow \tau_t - \frac{\alpha}{n|\mathcal{I}_b|} \sum_{i \in \mathcal{I}_b} \nabla L(\tau_t; x_{\sigma_i})$$

$$t \leftarrow t + 1$$

Output:

$$\tau_{\text{final}}$$

Sensitivity Enhanced Loss The Sensitivity Enhanced Loss function aims to reduce the number of false negatives, crucial in medical diagnostics. This function is defined as:

$$L_{\text{sensitivity}} = \frac{1}{N} \sum_{i=1}^{N} \left[BCE(\mathbf{p}_i, \mathbf{t}_i) \times (1 + \beta \times FN(\mathbf{p}_i, \mathbf{t}_i)) \right] \tag{5}$$

where BCE represents the binary cross-entropy loss, $\mathbf{p}_i$ is the predicted probability for the i^{th} sample, $\mathbf{t}_i$ is the true label, FN denotes the false negative identification function, and β is the weight amplifying the influence of false negatives. N is the number of samples.

These mathematical formulations underpin the loss functions in our model, contributing significantly to its ability to accurately segment and identify COVID-19 related abnormalities in lung CT images.

4. Results and Discussion

4.1. *Dataset Description and Preprocessing*

Our research employed the Mosmed Dataset, which we downloaded on [22] is a comprehensive collection of computed tomography (CT) scans specifically assembled for the study of COVID-19. The dataset is organized into several categories, with 'CT-0' representing CT scans of normal lungs without any signs of infection. For a more nuanced analysis, we combined data from two categories, 'CT-1' and 'CT-2', as they both contain CT scans of lungs with varying levels of COVID-19 infection. This combination was essential to capture a broader spectrum of the disease's manifestations, ranging from mild to severe cases. The scans in 'CT-1' depict a moderate level of lung infection, while 'CT-2' encompasses scans with more severe infection symptoms. Each scan in these categories is paired with corresponding ground truth masks found in a 'masks' folder.

This extensive and diverse dataset, particularly with its 50 meticulously labeled ground truth data points, allowed us to enrich our preprocessing workflow. By leveraging the varied degrees of infection represented in the combined 'CT-1' and 'CT-2' data, we were able to enhance the robustness and accuracy of our COVID-19 lesions segmentation model, ensuring it is well-trained to recognize and differentiate between multiple infection intensities.

We utilized Python on Google Colab Pro for the complete implementation of the pipeline, including its training, testing, and validation phases.

4.1.1. *Preprocessing Mosmed CT scans and Hyperparameters Setting*

In our research, the preprocessing of CT scan images was a multi-faceted and critical phase. We initiated this by reading the scans in NIfTI format, commonly used in medical imaging, and extracted their spatial resolution. To ensure clarity and consistency, we normalized the intensity values of the scans within a specific Hounsfield

Unit window, clipping values outside this range. This normalization accentuated pertinent tissue densities.

Resizing followed, where we adjusted each scan to a uniform size, maintaining the aspect ratio to prevent image distortion. We also performed a resampling of CT pixels, standardizing the pixel spacing across all scans. This step was crucial to ensure each voxel represented consistent physical dimensions, a key factor for accurate analysis.

The next step involved extracting a set number of slices from each scan, standardizing the depth and focusing on the most informative sections. We applied specific techniques to both the scans and their corresponding masks. For the scans, we utilized a combination of advanced image processing techniques that are typically employed in medical image analysis. This included normalization to standardize the intensity levels across different scans, which is essential for consistent model training and analysis. We also applied filtering methods to enhance the clarity of the scans, reducing noise and improving the visibility of critical features. Additionally, techniques like contrast adjustment were used to highlight subtle differences in tissue densities, which is particularly important for identifying and segmenting areas of interest in lung CT scans. These preprocessing steps ensured that our model received the highest quality data, which is crucial for accurate segmentation and analysis.

Lastly, we implemented a lung mask extraction process, applying custom scripts to isolate and highlight lung regions in each scan. This comprehensive preprocessing approach was instrumental in achieving a dataset optimized for detailed and accurate lesion segmentation, setting a solid foundation for our subsequent analysis.

Parameter Values

The parameter τ is very important for measuring the loss function. To find the best value, we used the stochastic gradient descent method, and we found that τ is equal to 0.06.
We also utilized the following parameters:

- **initial_learning_rate**: The starting value for the learning rate, denoted as α_0, it is set to 0.00005.
- **decay_steps**: Specifies how often the learning rate is adjusted. After every 1000 training steps, the learning rate is updated.
- **decay_rate**: Defines the rate at which the learning rate is reduced, denoted as β. After each set of 1000 steps, the learning rate is multiplied by 0.9 to decrease its value.

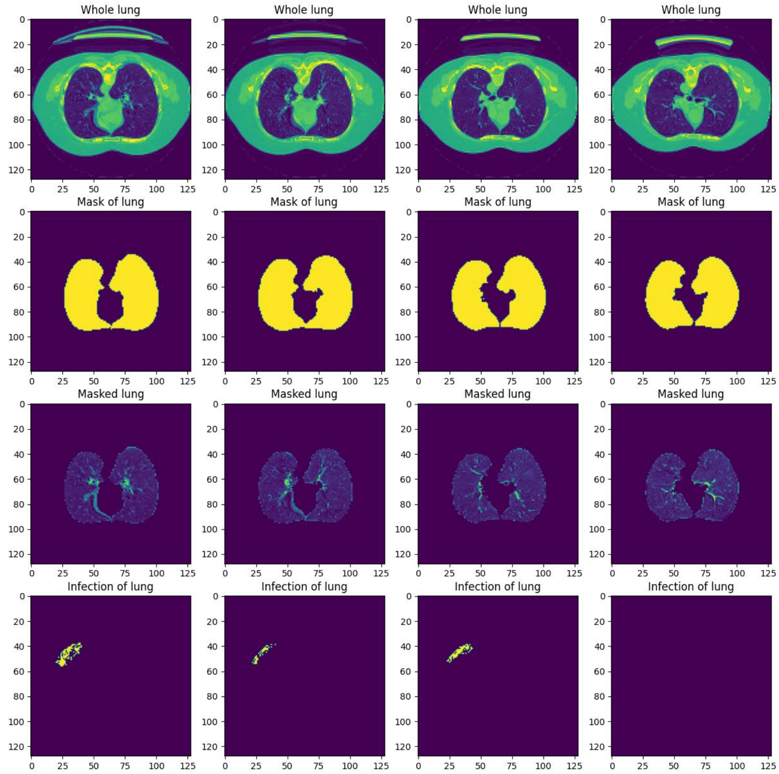


Fig. 2. Cropped Lunges by Extracted Masks - Mosmed Data Preprocessing Phase

4.1.2. *Preprocessing Outputs*

After we had finished preprocessing, we ended up with a set of CT images that were ready for further study. We made sure these images were clear and consistent by adjusting their size and focus **(A 50 Covid-positive CT-scans and 51 Covid-negative CT-scans)**.

Cropped Lung Images: An important part of this process was cropping the images to focus just on the lungs. This helps us see any changes in the lungs more clearly, which is especially useful for spotting signs of COVID-19.

Saving Processed Data: We sorted the images into two groups: one for CT scans showing COVID-19 and another for normal, healthy scans.

We then changed these images and masks into a format that's easy to use for computer-based analysis (numpy arrays). This step makes sure that our data is well-organized and easy to access later.

Data Storage: We saved these numpy arrays separately. This includes arrays for healthy lung images, images showing infection, lung masks, and the original CT images. Keeping these files separate and well-organized is important for our work, especially when we start training our models and later testing the TransUNET-based segmentation model on our testing set.

4.2. *3D GAN Model Training Results*

Figure 3 shows the result of the 3D GAN synthesizing a pseudo-healthy image, to later substract it from the covid-infected image in order to obtain the pseudo-mask highlighting regions of infection for the training of the 2D segmentation model.

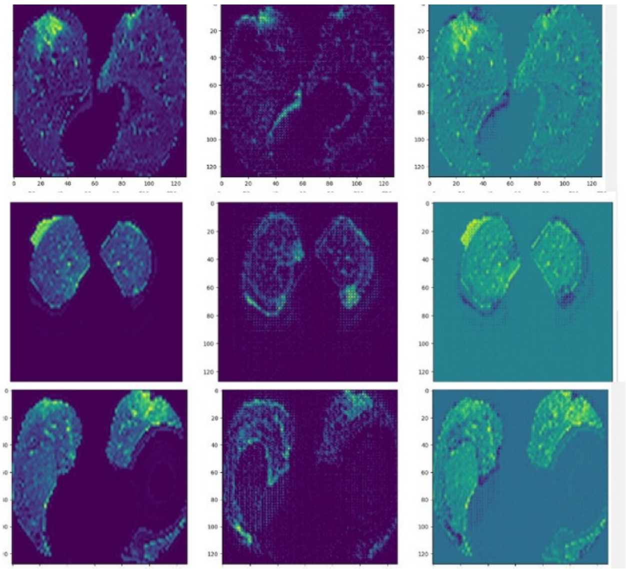
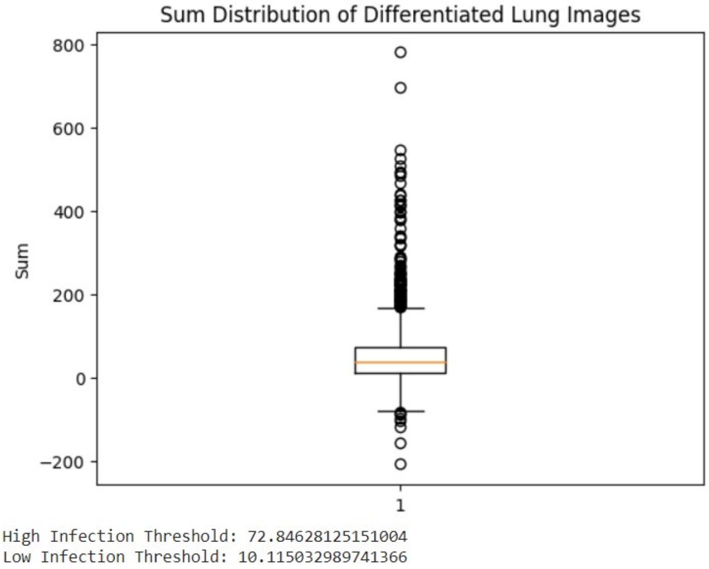


Fig. 3. Example of synthesizing semi-healthy medical images: Original COVID-19 lung CT scan (Left) - Synthesized COVID-19 lung CT scan (Center) - Highlighting the difference to identify infected regions (Right).

4.3. *Preprocessing and Infection Threshold Determination in Lung CT Images*

In our study, we employed a novel approach where we synthesized pseudo-masks by subtracting original COVID scans from the output generated by our 3D GAN model. These pseudo-masks, referred to as 'differentiated images', highlight the regions of lung infection and are critical for quantifying the severity of infections. To assess this severity, we calculated the sum of pixel intensities for each differentiated image. These sums serve as an indicator of lung involvement, with higher values pointing to more severe infections.



```
High Infection Threshold: 72.84628125151004
Low Infection Threshold: 10.115032989741366
```

Fig. 4. Box plot showing the distribution of pixel intensity sums in differentiated lung CT images. The plot delineates median, lower, and upper quartiles, aiding in the establishment of infection severity thresholds.

A box plot was then utilized as can be seen in figure 4 to visualize the distribution of these sums, assisting in identifying the spread and central tendencies of the data. Based on the box plot analysis, we established thresholds for categorizing infection levels. The 'high infection threshold' was set above the upper quartile, indicative of extensive infection, while the 'low infection threshold' was placed between the lower quartile and the median, representative of milder cases. This methodology enabled us to effectively segment the images into distinct categories based on the severity of infection, providing a data-driven foundation for our subsequent analytical endeavors.

4.4. *Segmentation of Lung CT Images into Infection Categories*

We categorize lung CT images into high and low infection groups based on the established thresholds. This involves initializing datasets for each category, classifying images using loop iteration, and tracking the distribution of infection severity. This segmentation process informs our model's training strategy and is pivotal in achieving nuanced segmentation.

4.5. *Architectural Adaptation of Vision Transformer for Lung CT Images*

The original RGB-focused Vision Transformer (ViT) was adapted to process single-channel grayscale CT images. We altered the input layer to handle one-channel input and adjusted patch embedding to accommodate the $128 \times 128 \times 1$ image size. The Transformer Encoder was also adapted to process the embedded patches effectively. These modifications allow the ViT to extract spatial and intensity features relevant to lung CT scans, and integrate smoothly with the U-Net architecture, enhancing the model's segmentation capabilities.

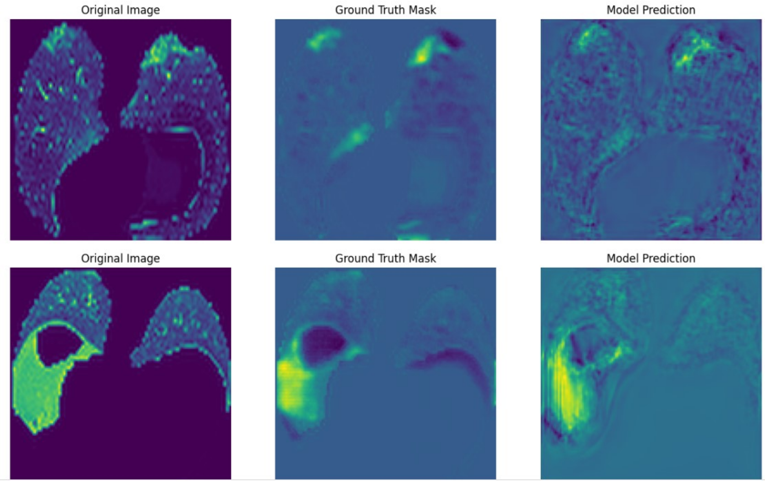


Fig. 5. TransUNET Segmentation Encoder-Decoder results

4.6. *Results on Validation Set Consisting of 50 covid CT scans*

It is important to note that in the Mosmed Dataset we accessed, there were only 50 COVID-19 CT scans with ground-truth masking of the infection regions. Consequently, the preprocessing step resulted in a validation set comprising these 50 COVID-19 CT scans.

Table 1. Comparison of models performances

Metric	**GAN + Perlin Noise + TransUNET**		**GAN + Perlin Noise + UNET**	
	Mean (%)	Std (%)	Mean (%)	Std (%)
Dice Score	50.09	45.44	**72.89**	46.69
Sensitivity	**98.52**	11.29	72.89	46.69
Specificity	**99.99**	1.11e-14	99.63	6.05

In this study, we compared the performance of two different image segmentation models, TransUNET and UNET. The results showed that TransUNET achieved an average Dice score of 47.09%, a lower segmentation accuracy index compared to UNET, but with a standard deviation of 45.44%, indicating less variability and thus being interesting. However, TransUNET was more effective in terms of sensitivity, with an impressive average of 98.52% (in addition to its significantly lower variability compared to UNET), and near-perfect specificity of 100%, suggesting an exceptional ability to accurately identify both positive and negative cases. On the other hand, UNET achieved a higher average Dice score of 72.89%, indicating generally better segmentation accuracy. Nonetheless, its sensitivity and specificity were lower than those of TransUNET, with a sensitivity of 72.89% and a specificity of 99.63%. While these scores are still high, they are noticeably less impressive than those of TransUNET, especially in terms of sensitivity. In conclusion, although UNET demonstrated better overall segmentation accuracy, TransUNET outperformed UNET in terms of sensitivity and specificity. These results suggest that TransUNET could be more reliable for critical applications where accurate detection of every positive case is crucial, such as in the medical diagnosis of conditions like COVID-19. However, if the primary goal is overall segmentation accuracy without specific consideration for sensitivity or specificity, then UNET might be preferable.

5. Conclusion

This research presents an approach to medical image segmentation with a focus on COVID-19 CT image analysis. The integration of a memory-efficient GAN with a TransUNet model forms the cornerstone of this study, addressing the challenge of dataset annotation in medical imaging.

Our method utilizes a 3D GAN to create pseudo-masks by subtracting synthesized healthy images from COVID-19 CT scans. This technique contributes to the segmentation model's performance while preserving the structural details of lung images, which is essential for accurate medical analysis.

The TransUNet, combining transformer layers with the U-Net architecture, has shown promising results in lesion prediction and detailed image processing. The study reveals the potential of TransUNet to perform better in certain aspects, such as sensitivity and specificity, compared to traditional U-Net models.

The validation of our approach across MOSMED dataset indicates an improvement over some existing unsupervised and weakly-supervised segmentation techniques. Particularly noteworthy are the segmentation results in low infection cases, demonstrating the model's capability to handle diverse and complex medical images.

Importantly, this research achieves a higher level of segmentation accuracy without the need for manual pixel-level annotation. This aspect of the study suggests a possible direction for future research in medical image analysis, aiming to lessen the reliance on extensive annotated datasets.

In summary, this study contributes an approach that enhances the efficiency and accuracy of medical image segmentation. The findings have potential applications in improving diagnostic processes, especially in situations like the ongoing COVID-19 pandemic. It is hoped that the insights gained from this research will support further exploration in the field of medical imaging, aiding in the development of more resource-efficient and accurate diagnostic tools.

6. Authors Contributions

Dr. Oumaima BANOUAR, in her role as the team's tutor, proved to be an invaluable asset, playing a pivotal role in steering and bolstering the team's endeavors. She adeptly furnished the team with essential resources and steered them toward prospective approaches to explore. Her supervision extended to the execution of codes, where she meticulously oversaw the intricacies, ensuring a seamless workflow. Additionally, she orchestrated regular meetings, fostering a collaborative environment for the team to delve into advanced problems and collectively brainstorm potential solutions. Through her astute guidance and meticulous oversight, she was instrumental in ensuring the team's consistent progress and the timely achievement of its objectives. Her expertise and leadership were paramount in navigating challenges.

Aymane DHIMEN played a crucial role in coordinating team efforts, particularly addressing challenges in medical image data and managing data generation costs. He was responsible for overseeing the implementation and experimentation processes. His proactive approach fostered a cohesive project outcome, effectively merging theoretical concepts with practical application.

Selma KOUDIA carried out an essential literature review and comparative analysis, highlighting trends and challenges in medical image segmentation and generative models. Her collaboration with the team was pivotal in integrating these findings into the project, ensuring its distinctiveness in the field. Selma meticulously defined the project's pipeline and ensured a well-structured conclusion in the research paper, encapsulating the overall effort and achievements of the implementation and testing phases.

Mostapha EL ANSARI focused on exploring GAN and TransUNET models, proposing the innovative use of TransUNET over UNET. This suggestion was aimed at improving data processing efficiency, offering a potential solution to the scarcity of medical image data and enhancing model performance. His choice of TransUNET brought a fresh perspective that could tackle data availability challenges and improve medical image processing efficiency.

Aymane EL Fahsi's contributions were significant in developing and evaluating the GAN and TransUnet pipeline. He leaded the implementation, adaptation, and testing of the models. His role in the self-supervised, region-aware segmentation method and in drafting the project report was also paramount. Moreover, he also contributed to the planning and writing of the research paper.

Mohammed Adam Khali undertook a mathematical analysis of GAN algorithms and the TransUnet model, supporting the development of our self-supervised region-aware segmentation method. He also contributed to enhancing the U-net Network with attention for segmentation and played a vital role in the manuscript's writing process. Furthermore, his work on optimizing the hyperparameters and parameters of various scripts and algorithms, in collaboration with Aymane EL Fahsi, was essential to the project's success.

7. Declaration of Competing Interests

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

References

1. Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Droit Merhof. Medical image segmentation review: The success of u-net. *arXiv preprint arXiv:2211.14830*, 2022. URL *https://arxiv.org/pdf/2211.14830.pdf*
.
2. *Takayasu Moriya, Holger R. Roth, Shota Nakamura, Hirohisa Oda, Kai Nagara, Masahiro Oda, and Kensaku Mori. Unsupervised segmentation of 3d medical images based on clustering and deep representation learning. 2022. URL https://arxiv.org/pdf/1804.03830.pdf*
.
3. *Tao Song Guotai Wang Shaoting Zhang Xiangde Luo1, Minhao Hu. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. 2021. URL https://arxiv.org/pdf/2112.04894.pdf*
.
4. *Hao Du, Qihua Dong, Yan Xu, and Jing Liao. Weakly-supervised 3d medical image segmentation using geometric prior and contrastive similarity. 2023. URL https://arxiv.org/pdf/2302.02125v1.pdf*
.
5. *Kelei He, Chen Gan, Zhuoyuan Li, Islem Rekik, Zihao Yin, Wen Ji, Yang Gao, Qian Wang, Junfeng Zhang, and Dinggang Shen. Transformers in medical image analysis: A review.* arXiv preprint arXiv:2202.12165v3*, 2022. URL https://arxiv.org/pdf/2202.12165v3.pdf*
.
6. *Varut Vardhanabhuti Mohammad Ali Nikouei Mahani Mohamad Koohi-Moghadam Siyavash Shabani, Morteza Homayounfar. Self-supervised region-aware segmentation of covid-19 ct images using 3d gan and contrastive learning. 2020. URL https://doi.org/10.1016/j.compbiomed.2022.106033*
.

7. *Xiaocong Chen, Lina Yao, Tao Zhou, Jinming Dong, and Yu Zhang. Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images.* Pattern recognition*, 113:107826, 2021.*
8. *Changhee Han, Yoshiro Kitamura, Akira Kudo, Akimichi Ichinose, Leonardo Rundo, Yujiro Furukawa, Kazuki Umemoto, Yuanzhong Li, and Hideki Nakayama. Synthesizing diverse lung nodules wherever massively: 3d multi-conditional gan-based ct image augmentation for object detection. In* 2019 International Conference on 3D Vision (3DV)*, pages 729–737. IEEE, 2019.*
9. *Ping Chai, Lei Hou, Guomin Zhang, Quddus Tushar, and Yang Zou. Generative adversarial networks in construction applications.* Automation in Construction*, 159:105265, 2024.*
10. *Haseeb Nazki, Sook Yoon, Alvaro Fuentes, and Dong Sun Park. Unsupervised image translation using adversarial networks for improved plant disease recognition.* Computers and Electronics in Agriculture*, 168:105117, 2020.*
11. *Kyeong-Beom Park, Sung Ho Choi, and Jae Yeol Lee. M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks.* IEEE Access*, 8:146308–146322, 2020.*
12. *Chen Zhao, Jungang Han, Yang Jia, and Fan Gou. Lung nodule detection via 3d u-net and contextual convolutional neural network. In* 2018 International conference on networking and network applications (NaNA)*, pages 356–361. IEEE, 2018.*
13. *Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In* Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.*
14. *Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas.* arXiv preprint arXiv:1804.03999*, 2018.*
15. *Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming 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: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.*
16. *Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaofei Han, Zhong Xue, Dinggang Shen, and Yuxin Shi. Lung infection quantification of covid-19 in ct images with deep learning.* arXiv preprint arXiv:2003.04655*, 2020.*
17. *Guotai Wang, Xinglong Liu, Chaoping Li, Zhiyong Xu, Jiugen Ruan, Haifeng Zhu, Tao Meng, Kang Li, Ning Huang, and Shaoting Zhang. A noise-robust*

framework for automatic segmentation of covid-19 pneumonia lesions from ct images. IEEE Transactions on Medical Imaging*, 39(8):2653–2663, 2020.*

18. *Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images.* IEEE transactions on medical imaging*, 39(8):2626–2637, 2020.*
19. *Varut Vardhanabhuti Mohammad-Ali Nikouei Mahani Mohamad Koohi-Moghadam Siyavash Shabani, Morteza Homayounfar. Self-supervised region-aware segmentation of covid-19 ct images using 3d gan and contrastive learning.* Computers in Biology and Medicine*, 149(7):106033, 2022. URL https://doi.org/10.1016/j.compbiomed.2022.106033 .*
20. *Varut Vardhanabhuti Mohammad-Ali Nikouei Mahani Siyavash Shabani, Morteza Homayounfar and Mohamad Koohi-Moghadam. Self-supervised region-aware segmentation of covid-19 ct images using 3d gan and contrastive learning.* Computers in Biology and Medicine*, 149, 2021.*
21. *mkara44. transunet implementation for pytorch by mkara44 on github. URL https://github.com/mkara44/transunet$_p$ytorchhttps : //github.com/mkara44/transunet_pytorch .*
22. *Mosmed AI. Mosmeddata-ct-covid19-type vii-v 2. URL https://mosmed.ai/en/datasets/covid191110/https://mosmed.ai/en/datasets/covid191110/ .*