STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, May 2024, pp 646–659.
Published online in International Academic Press (www.IAPress.org)

GAN based approaches for self-supervised segmentation: A comparative study

Zohair Elmourabit *, Oumayma Banouar

Laboratory of Computer and System Engineering, Faculty of Sciences and Technics, Marrakesh, Morocco

Abstract Image segmentation is a fundamental image processing technique that involves di-viding an image into distinct regions or segments to enable the analysis and extrac-tion of valuable information. It finds applications in various fields including medi-cine, pattern recognition, computer vision, and security surveillance. Different types of image segmentation techniques can be employed based on specific application requirements, including thresholding, region-based segmentation, edge-based segmen-tation, clustering-based segmentation, active contour models (Snakes), graph-based segmentation, watershed segmentation, and deep learning-based segmentation. the latter has become a powerful tool in image segmentation. Deep learning (DL) models can be trained on large datasets of images, allowing them to learn intricate relation-ships between pixels and object classes. This is particularly beneficial for challenging segmentation tasks. However, one significant challenge is the scarcity of labeled training data. To address this issue, self-supervised approaches using generative mod-els like Generative Adversarial Networks (GANs) provide a solution. GANs can gener-ate synthetic training data, which is useful for tasks where acquiring labeled training data is difficult or expensive, such as in medical imaging or remote sensing applica-tions. Additionally, GANs can generate realistic segmentation masks, which are cru-cial for tasks like medical imaging. In this study, we conducted a comparative analy-sis of DL-based segmentation approaches for COVID-19 CT scans. The evaluated approaches were assessed using metrics such as Dice Score, Specificity, and Sensitivi-ty. These metrics provide quantitative measures of the segmentation performance, allowing for an objective evaluation and comparison of the different techniques.

Keywords Image Segmentation, Automatic Segmentation, Deep Learning, GAN, Self-supervised approaches.

DOI: 10.19139/soic-2310-5070-1928

1. Introduction

Medical image segmentation is a crucial step in many therapeutic applications, such as early disease detection, disease progression monitoring, and personalized treatment planning for patients. Segmentation involves dividing an image into different regions or segments [29], which facilitates analysis and the extraction of useful information from the image. However, manual segmentation is a tedious and time-consuming task that requires the expertise of a radiologist or medical imaging specialist. That's why automatic segmentation techniques have been developed to reduce the expert's workload and improve the efficiency of medical image analysis [28]. Automatic segmentation methods can be supervised [23, 30], semi-supervised [24], or unsupervised [24]. Supervised methods require a well-annotated paired dataset to train the segmentation model [23, 28], while unsupervised methods do not require manual annotations. Automatic segmentation techniques have made significant advancements in recent years, thanks to the use of DL networks. Techniques such as magnetic resonance imaging (MRI), computed tomography (CT scanning), and digi-tal mammography allow for non-invasive visualization of an individual's internal anatomy. These technological advancements have greatly improved our understanding of normal and pathological anatomy, promoted medical research and facilitated accurate diagnosis and treatment planning. With

*Correspondence to: Zohair Elmourabit (Email: zohair.elmoura@gmail.com), Oumayma Banouar (Email: o.banouar@uca.ac.ma)

ISSN 2310-5070 (online) ISSN 2311-004X (print)

these imaging modalities, doctors can obtain detailed information about organs, tissues, and anatomical structures, enabling them to detect abnormalities, lo-cate pathologies, quantify tissue volumes, and assess treatment efficacy. However, due to the increasing quantity and size of medical images, it has become necessary to utilize computers to facilitate the processing and analysis of this data. In this work, the objective is to develop an unsupervised DL model for medical image segmentation, specifically focusing on CT images related to COVID-19. The images are medical images obtained from a CT scanner, which uses X-rays to pro-duce detailed cross-sectional images of the body's interior [25]. They are commonly used to diagnose and monitor various diseases, including respiratory diseases such as COVID-19. The use of an unsupervised model in this context aims to reduce the workload of collecting and manually annotating data, as well as accommodate vari-ations and inaccuracies in medical imaging data. The adopted method for this pur-pose is a GAN, which is a DL framework consisting of two neural networks: a genera-tor and a discriminator. GANs have gained significant popularity in various fields. Regarding medical image processing, GANs have the potential to be employed for the creation of data masks, which can be used to augment images and train alternative models for image segmentation. In this work, GANs are employed to synthesize healthy images from infected images. These generated images are then used to pro-duce pseudo masks by subtracting them from the original infected images. Finally, these masks are employed to train a U-Net model with contrastive learning, aiming to enhance the quality of the image segmentation. The primary objective of employing an unsupervised approach in this study is to tackle the issue of insufficient labeled data. Labeling medical images for segmenta-tion tasks can be a resource-intensive and time-consuming process. By using an un-supervised model, the need for manual annotation is eliminated or reduced, which saves significant time and resources. In this research, the model underwent training and evaluation using three openly accessible datasets: coronacases, radiopedia, and mosmed. These datasets contain CT images related to COVID-19 and are commonly used in medical imaging research. The subsequent sections of this paper are organized as follows: Section 2 introduc-es and defines DL techniques for self-supervised image segmentation. Section 3 out-lines the experimental methodology, presents the results obtained, and discusses them. Finally, in Section 4, the paper concludes.

2. Deep learning-based approaches for self-supervised image seg-mentation

2.1. Image segmentation using GAN

GAN is a neural network architecture comprising two primary components: a generator (G) and a discriminator (D). The generator network accepts random noise as input and generates synthetic data samples, such as images, using that noise as a basis. On the other hand, the discriminator network takes both real data samples and the synthetic samples generated by the generator as input and strives to differentiate between them [5]. The training of a GAN involves a competitive process between the two networks G and D. The objective of the generator is to produce synthetic samples that closely resemble real samples, making them difficult to distinguish. On the other hand, the discriminator aims to enhance its ability to discern between real and synthetic samples. These two networks are trained iteratively, with the generator(G) aiming to fool the discriminator(D), and the discriminator aiming to accurately classify the samples. The goal of training a GAN is to achieve a generator that can produce realistic and high-quality samples that are like real data distribution. In simpler terms, D and G engage in a competitive game where they try to outsmart each other. This game involves a minimax strategy and a value function V(G,D)[7]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \tag{1}$$

GANs have been successfully applied in various domains, including image synthe-sis [6], image translation, and image segmentation, among others.

In 2020, Zhanwei Xu et al. proposed GASNet [4], a Semi-automated framework for segmenting COVID-19-related lesions. It is a type of neural network based on GANs. It has been used for unsupervised medical image segmentation. GASNet com-prises 3 networks: a generator (G), a discriminator (D), and a segmenter (S). The S is utilized to generate segmentation masks, while the G replaces the predicted lesion region with a generated region that resembles the non-infected region while preserving the non-infected region's characteristics. The G and S are trained in an adversarial manner with the D, whose role is to differentiate between synthetic volumes and real

healthy volumes. The synthetic volume is formulated by combining the generated volume and the actual diseased volume based on the segmentation mask predicted by the S. GASNet was evaluated on the Mosmed and Radiopedia datasets, achieving Dice scores of 54% and 60%, respectively, for the two datasets.

2.2. Image segmentation using Contrastive learning

Contrastive learning is an approach to self-supervised learning that focuses on ex-tracting valuable representations from unlabeled data. It achieves this by maximizing the similarity between augmented views of the same sample and minimizing the simi-larity between views from different samples. It has demonstrated successful applica-tion in diverse fields such as computer vision and NLP. [9]. In DL, contrastive learn-ing typically involves training a neural network encoder to map input data into a lower-dimensional embedding space. entails generating augmented versions of input samples and considering them as positive pairs, while treating other samples as nega-tive pairs. The objective is to maximize the similarity between positive pairs and min-imize the similarity between negative pairs within the embedding space. as visualized in figure 1. The contrastive loss function is a frequently employed method for training the network. It incentivizes the network to minimize the distance between positive pairs, bringing them closer together, while simultaneously encouraging the network to maximize the distance between negative pairs, pushing them further apart. Various techniques, such as instance discrimination or InfoNCE loss, have been proposed to formulate the contrastive loss and enhance the learning process.

Figure 1. Contrastive learning reduces the distance between an anchor and a positive sample, which share the same identity, while increasing the distance between the anchor and a negative sample from a different identity [8].

In 2021, Yao et al. proposed a label-free method called ”Label-free” [3] for seg-menting COVID-19 lesions in lung CT scans. The objective of this method was to develop an automated and efficient segmentation method that does not require manual annotation of lesion regions. The proposed method utilizes a DL architecture based on convolutional neural networks (CNNs) and a self-attention mechanism to identify and segment COVID-19 lesions in CT scans.

CNNs are highly specialized ANNs widely utilized for tasks revolving around image recognition and processing. These networks draw inspiration from the intricate struc-ture and functionality of the visual cortex [26], a component of the brain responsible for processing visual information. Comprising multiple layers, each layer assumes a specific role within the CNN architecture. Typically, the initial layer in a CNN is a convolution layer responsible for extracting pertinent features from the input image. Following this, a pooling layer is commonly employed to reduce the image dimen-sions while retaining crucial features. This iterative process of convolution and pool-ing persists until the final layer of the CNN delivers the desired prediction [27].

The self-attention mechanism is used to selectively focus on the most relevant fea-tures for the segmentation task, while CNNs are used to extract and process image features. Label-free achieved Dice scores of 69%, 59%, and 61% on the Coro-nacases, Radiopedia, and UESTC datasets, respectively, during evaluation.

2.3. U-net architecture for image segmentation

U-Net is a CNN developed for biomedical image segmentation Its U-shaped architec-ture as shown in the figure 2. It was first proposed in 2015 by Olaf Ronneberger [11]. The architecture comprises three primary components: the encoder, the decoder, and the bridge. The encoder is tasked with extracting image features and reducing its size to minimize the number of network parameters. It uses convolutional layers to pro-gressively extract increasingly

abstract information from the input image. The bridge connects the encoder and the decoder. Convolutional layers are commonly em-ployed in the composition of the bridge, facilitating the integration of information from both the encoder and the decoder [10]. The decoder is responsible for precise localization through transposed convolutions and also for restoring the initial size of the image [12]. It uses transposed convolution operations (also known as deconvolu-tion) to gradually increase the size and restore the original spatial resolution. This produces a detailed segmentation map that corresponds to the input image.

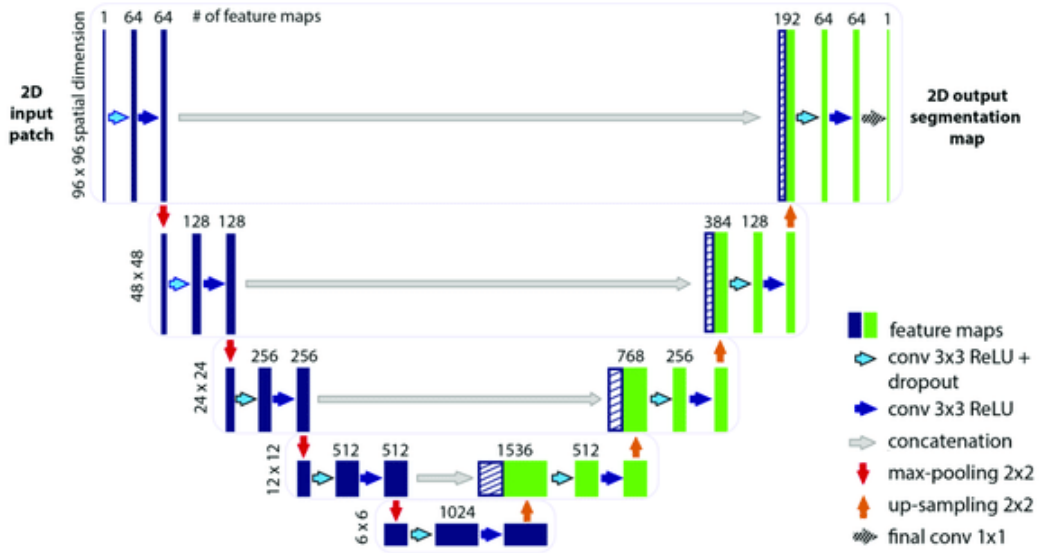

Figure 2. U-Net architecture [13].

In 2020, Wang Guotai et al. introduced nnU-Net [2], a DL-based segmentation technique that automatically adapts to new tasks by configuring itself, including preprocessing, network architecture, training, and post-processing. The architecture of nnU-Net consists of several key modules. Firstly, nnU-Net employs a (CNN)-based encoder to extract informative characteristics from input images. This part of the architecture uses convolutional and pooling layers to capture contextual information at different scales. Next, nnU-Net integrates an attention mechanism that focuses on the most relevant regions of interest for segmentation. This adaptive attention mechanism allows the model to selectively emphasize crucial details. Another notable aspect of nnU-Net is its approach to handling class imbalance. It uses adapted sam-pling strategies to deal with cases where the classes to be segmented are imbalanced in terms of size or number of pixels. Finally, nnU-Net employs a decoder architecture that reconstructs segmentation masks from the features extracted by the encoder. This part of the architecture uses deconvolution and up sampling layers to enlarge the features and obtain accurate segmentation masks. nnU-Net achieved Dice scores of 81% and 76% on the UESTC and Radiopedia datasets, respectively. In the two pre-viously mentioned works, the authors proposed supervised approaches that require manually generated masks and human intervention for performing segmentation. However, this task can be time-consuming and resource intensive as it requires more annotated data. Next, we will present some works that are less supervised and only require manually generated masks.

In 2020, Guotai Wang et al. [1] proposed a novel noise-robust framework for ad-dressing noisy labels in the segmentation task. They introduced a noise-aware Dice loss, which extends the conventional Dice loss for segmentation, and the Mean Abso-lute Error (MAE) loss to enhance noise robustness. Additionally, they developed COPLE-Net, a new network specifically designed for segmenting COVID-19 pneu-monia lesions, aiming to handle lesions with diverse scales and appearances. The noise-aware Dice loss and COPLE-Net were integrated into an adaptive self-ensembling framework for training. The noise-robust Dice loss is a modified loss func-tion that is more resilient to noise in the training data compared to the standard Dice loss commonly used in image

segmentation tasks (proposed by Milletari et al. [27]). As part of the framework, a teacher model was created by applying an Exponential Moving Average (EMA) to the student model, which was dynamically updated by excluding the contribution of the student when the student exhibited high training loss. The student model also adapted its learning from the teacher only when the teacher outperformed the student. COPLE-Net was evaluated on two datasets, UESTC and Radiopedia, resulting in different Dice scores: 83% for UESTC and 59% for Radiopedia. This indicates that COPLE-Net's performance may be influenced by dataset variations and may not consistently provide stable results across different data sources.

A study of works related to the segmentation of medical images, particularly CT images, conducted between 2020 and 2022, shows that supervised models have yielded good results in terms of Dice scores, slightly outperforming unsupervised models, and have demonstrated stability across different datasets, with the exception of COPLE-Net, which showed different results between Radiopedia and UESTC, indicating some instability.

2.4. GAN and U-Net combination for image segmentation

In medical imaging, the process of accurately outlining or segmenting specific struc-tures or regions of interest within medical images is crucial for various diagnostic and treatment purposes. However, obtaining annotated or labeled data for training ma-chine learning algorithms in medical image segmentation can be particularly chal-lenging and limited in availability. The segmentation is a complex task that requires high-quality data and professional expertise for annotation. Moreover, obtaining annotated data is often challenging due to the need to comply with patient confiden-tiality protocols and health data protection regulations [14]. This limits the amount of data available for training, which can result in overfitting or low performance when applied to new data All these factors highlight the importance of exploring machine learning methods that can learn from unlabeled or sparsely labeled data. The goal of our approach is to build a pipeline based on unsupervised learning. Our pipeline involves a dual-model framework for segmenting CT scans without any pixel-level annotations. The main idea is to use a 3D generative adversarial network.

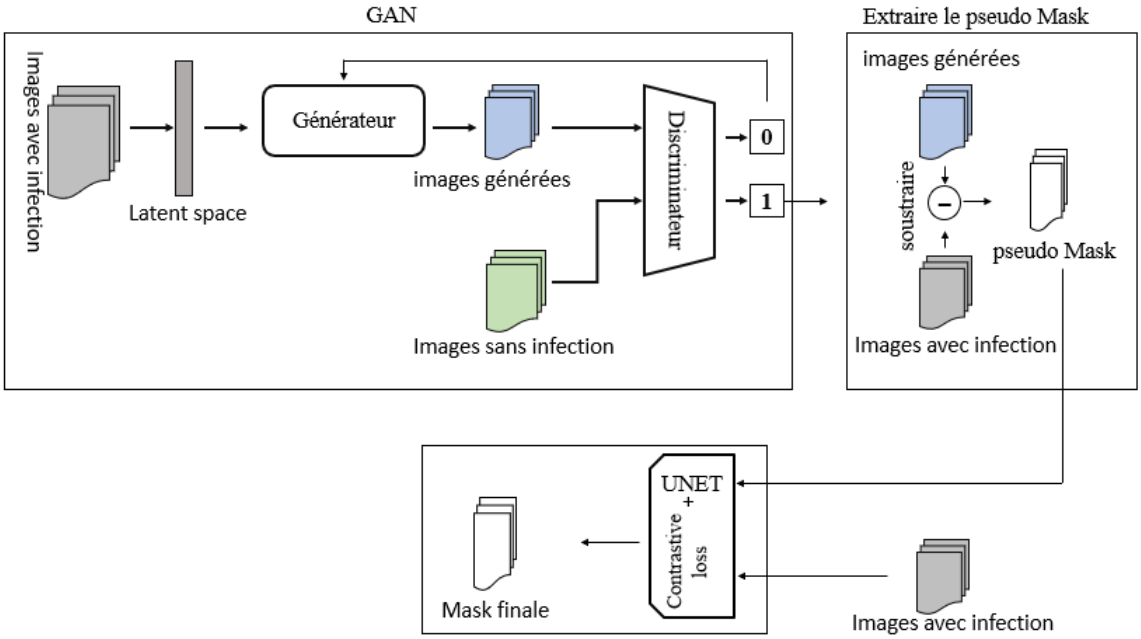

Figure 3. Our pipeline architecture comprises three primary components: a 3D GAN model that generates synthetic infected images, the extraction of a pseudo mask without pixel-level annotations, and the training of a 2D UNet model.

Our approach involves utilizing a Generative Adversarial Network (GAN) to gener-ate healthy images based on COVID-19 CT scans. These generated healthy images are subtracted from the original CT scans to extract the

infected area in three dimen-sions (3D). This process enables us to obtain pseudo-masks without relying on human annotators. Subsequently, we train a U-Net segmentation network using these pseu-do-masks and incorporate contrastive loss. Figure 3 illustrates the schematic represen-tation of our proposed approach : Our approach consists of three main parts: Part 1: During the training process, we utilize a 3D GAN with a multi-objective loss function. This loss function consists of a standard GAN objective, which aims to remove infected regions from CT COVID-19 images, and a reconstruction loss that aids in generating healthy images from noisy versions of the original healthy images (i.e., CT COVID-19 images with infected areas removed). The reconstruction loss serves as a regularization term, ensuring that the GAN generates images with the same underlying lung structure as the input images. The cost function for our 3D GAN is:

$$\mathrm{Max}_D\, V(D,G) = E_{I_H \sim p_{\text{healthy}}}\left[\log D\left(I_H\right)\right] + E_{I_C \sim p_{\text{covid}}}\left[\log\left(1 - D\left(G\left(I_C\right)\right)\right)\right] \tag{2}$$

Part 2: Pseudo-mask extraction, To extract the infected area in three dimensions (3D), we employed a pixel-wise subtraction technique. Specifically, we subtracted the synthesized healthy image, generated by the GAN, from the original COVID-19 image.
This subtraction process allowed us to obtain a pseudo-mask representing the infect-ed regions within the image.

$$M_{pseudo} = I_C - I_{HS} \tag{3}$$

With: M_{psuedo} extracted pseudo mask, I_c infected image and I_{HS} generated healthy image.

Part 3: 2D U-Net segmentation model. To train our U-Net segmentation model, we utilized the pseudo-masks obtained from the previous step as training data. These pseudo-masks were used as ground truth to predict the infected areas in the CT scans. During training, we employed a combination of Mean Squared Error (MSE) loss and contrast loss. The MSE loss was applied to both the encoder and the decoder parts of the network, while the contrast loss was specifically applied to the encoder. This joint training approach ensured effective learning and accurate segmentation results. Our U-Net architecture is specifically designed to take a 2D slice of a CT image that is infected with COVID-19 as input, noted as $I_C^{w\times h}$ where $w \times h$ is the dimension of the slice. The 2D U-Net model is as follows:

$$\begin{aligned} &UNet_{2D} : I_c^{w\times h} \rightarrow M_s^{w\times h} \\ &M_s^{w\times h} = De_{2D}\left(En_{2D}\left(I_c^{w\times h}\right)\right) \end{aligned} \tag{4}$$

With: De_{2D} , En_{2D} being the decoder and encoder models of the U-Net, respectively, $I_C^{w\times h}$ slice 2D and $M_S^{w\times h}$ the segmented mask. In the training of the encoder, we utilized contrastive learning to transform each image into a d-dimensional feature vector. To apply the contrastive loss as the cost function of the encoder, we generated a list of pseudo-positive and pseudo-negative labels, $y_i \in \{y_p, y_n\}$, from a set of 2D CT slices. Here, y_p represents heavily infected CT slices, and y_n represents lightly infected CT slices. These sets of heavily and lightly infected slices are determined by sorting the slices based on the sum of their pseudo-mask pixels. The encoder model is as follows:

$$\begin{aligned} &\mathrm{En}_{2D} : I_C \rightarrow \mathbb{R}^d \\ &L_{\text{contrast}}(i) = \frac{-1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp\left(z_i \cdot \frac{z_p}{\tau}\right)}{\sum_{l\in\mathcal{C}(i)} \exp\left(z_i \cdot \frac{z_l}{\tau}\right)} \end{aligned} \tag{5}$$

The contrastive loss function [15], as denoted by L_{contrast} , operates in a d-dimensional latent space, where $z \in R^d$ represents the projection of the latent space. T is a scalar temperature used in contrastive learning. This temperature is employed to adjust the ”softness” of the similarity distribution between image pairs. Higher temperature values result in a softer distribution, meaning that image pairs can have closer similarities to be considered positive [15]. The cardinality of positive samples for anchor i is denoted as $|P(i)|$, and $C(i)$ represents the set of both samples. In each iteration, we train the encoder part of the model using the contrastive loss. Subsequently, we train the entire U-Net model (encoder-decoder) using an end-to-end approach, employing the pixel-wise Least Mean Squared

Error ($LMSE$) loss. The $LMSE$ loss is calculated as follows:

$$L_{MSE}\left(I_{Ci}, M_{Si}\right) = \frac{1}{w \times h} \sum_{v=0}^{h} \sum_{t=0}^{w} \left[I_{Ci}(v,t) - M_{Si}(v,t)\right]^2 \tag{6}$$

In conclusion, we have presented a comprehensive pipeline for our proposed ap-proach, which avoids the need for labeled data. Our model has been specifically im-plemented for segmenting 3D lung images infected with COVID-19, but there is po-tential for its application in other medical scenarios, such as segmenting brain tumor images derived from MRI scans.

3. Result and discussion

3.1. Dataset

In our research, we made use of three different datasets: Mosmed dataset, coro-nacases, and Radiopaedia. The Mosmed dataset [16] consists of anonymized CT scans of human lungs, which include scans showing COVID-19-related observations as well as scans without such observations. Some studies within this dataset have been annotated with binary pixel masks to identify specific areas of interest, such as ground-glass opacities and consolidations. These CT scans were collected from medi-cal hospitals in Moscow, Russia, between March 1, 2020, and April 25, 2020. The dataset contains a collection of 1110 studies, and each study represents a distinct patient. Radiopaedia [17] is another dataset we utilized, which contains 10 CT stud-ies obtained from the website radiopaedia.org. Radiopaedia.org serves as an online educational platform that offers a wide range of information on radiology and medi-cal imaging. It provides a library of radiological images, articles, videos, and clinical cases, aiming to enhance the knowledge of healthcare professionals and medical students in the field of radiology. The Coronacases dataset [18] comprises 10 studies that were collected from the official website coronacases.org. Similar to the previous datasets, this dataset also includes CT scans relevant to COVID-19.
To ensure proper evaluation and training of our model, we divided the data ob-tained from these datasets into two sets: training data and test data, table 1. These datasets provided valuable resources for our research, enabling us to analyze and develop effective models for COVID-19 detection and diagnosis:

Table 1. Description of training and test data.

Training / Test	Dataset	Number	Case
Training	Mosmed	850	Infected
	Mosmed	250	healthy
Test	Mosmed	10	Infected
	Radiopedia	10	Infected
	Coronacases	10	infected

3.2. Data preprocessing

Data preprocessing is a vital and indispensable step in the data analysis process. It involves a set of operations aimed at preparing raw data for analysis or further use, ensuring data consistency, high quality, and formatting suitable for analysis. The data has been preprocessed following the following steps:
- Data Reading: The data was retrieved from NIfTI format files. Each file contains a CT study (niftii).
- Pixel Spacing Adjustment: To ensure uniform resolution, the pixel spacing of all CT images was standardized to a constant value of 1 mm in each direction.
- Hounsfield Unit (HU) Normalization: The Hounsfield Unit (HU) is a measurement unit used in computed tomography (CT scan) to measure the density of body tissues [19]. HU values are determined by measuring X-ray absorption in body tissues, which is influenced by tissue density and composition. In the preprocessing of CT

images, the Hounsfield Unit (HU) values of each pixel were constrained to the range (-1000, 400), which is the typical range for human tissues. Subsequently, the HU values were normalized within the range (0, 1) for all pixels.. This step is important to enable easi-er comparison of CT images as it reduces the impact of HU value variation across different samples and facilitates subsequent image analysis. Figure 4 shows four image slices after the 3 preprocessing steps:

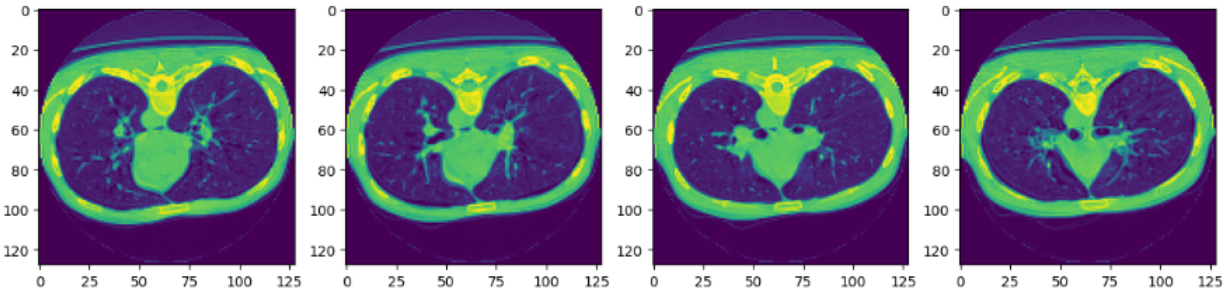

Figure 4. Example of CT image after pixel spacing adjustment and HU normalization.

- Image Cropping: The process includes repositioning the image to center the lungs, which emphasizes the region of interest and minimizes unnecessary data. This center-ing operation allows for focused analysis on the lungs and reduces data complexity, thereby enhancing the performance of analysis models.
- Resizing: To ensure consistent analysis and reduce data complexity, the 3D images are resized to a specific shape of (128, 128, 32). This resizing step ensures that all images have the same size, facilitating uniform processing and analysis..as shown in the figure 5.

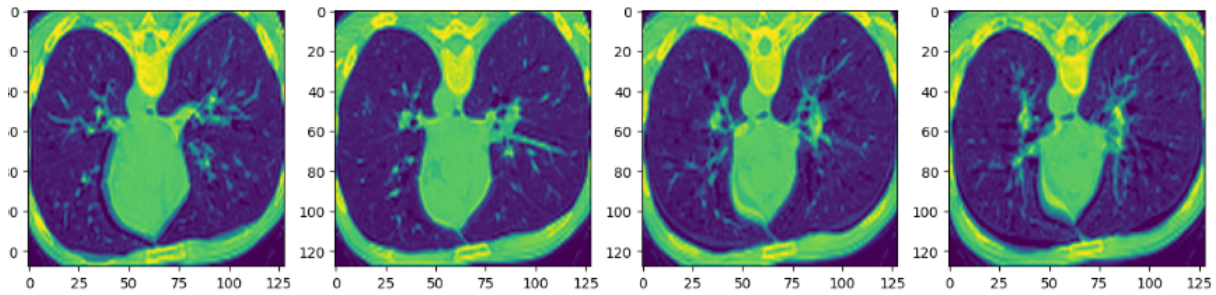

Figure 5. An example of a CT image after complete preprocessing.

3.3. Hyper-parameters

The choice of hyperparameters is an important aspect of neural network design. Hyperparameters are parameters that are not directly learned from the data, but are chosen before training the model and have a significant impact on the model's performance. We have chosen the following hyperparameters for the 3D GAN:

Table 2. The hyperparameters of the 3D GAN model.

Hyper-parameter	Optimizer	learning rate	loss function	batch size	epoch
Value	Adam	0.00005	mse	5	10000

To achieve end-to-end training, we simultaneously trained the encoder and decoder of the U-Net architecture using the hyperparameters listed in Table 3. Subsequently, we incorporated the contrastive loss to further train the encoder, as depicted in Table 4.

Table 3. The hyperparameters of the 2D U-Net model.

Hyper-parameter	Optimizer	learning rate	decay step	decay rate	loss function	batch size	epoch
Valeur	Adam	0.001	1000	0.9	mse	32	1000

Table 4. The hyperparameters of the Encoder model in U-Net..

Hyper-parameter	Optimizer	learning rate	loss function	batch size	epoch
Valeur	Adam	0.0001	Contrastive	32	1000

3.4. Evaluation and Discussion

The evaluation of our segmentation pipeline involves measuring the accuracy of the image segmentation produced by the models (GAN and GAN + U-Net). The eval-uation is conducted qualitatively and quantitatively. Qualitative evaluation is often used to assess segmentation models that produce visual outputs, such as segmented images. In this phase, we will examine the outputs generated by the models. Quanti-tative evaluation involves measuring the model's performance using quantitative metrics. We have used the Dice score [21, 22], sensitivity [22], and specificity [22]. In the qualitative evaluation phase, we will assess our proposed pipeline for tomograph-ic image segmentation. Firstly, we will begin by evaluating the 3D GAN. In this step, we present six examples of image slices from the three datasets (one slice for each dataset), as shown in the following figures:

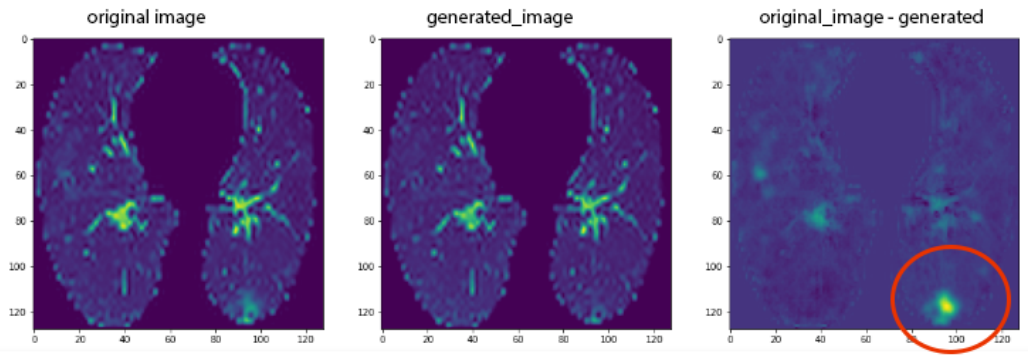

Figure 6. Mosmed: Example of a CT image segmented with GAN.

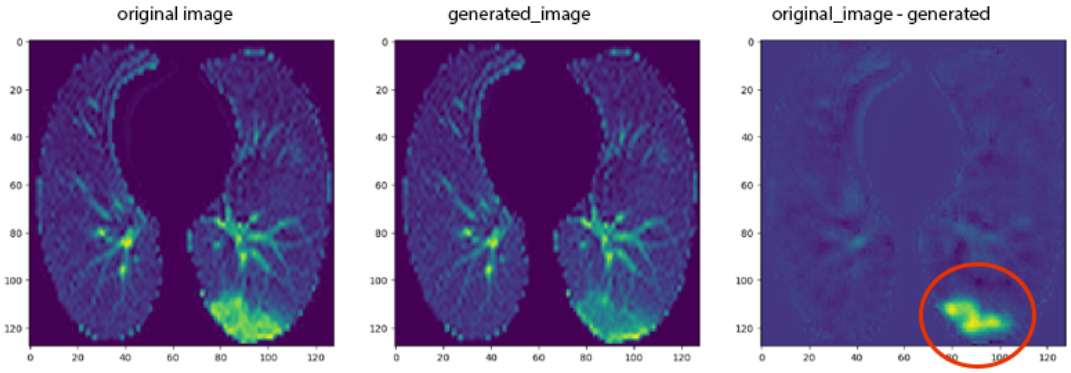

Figure 7. Coronacases: Example of a CT image segmented with GAN.

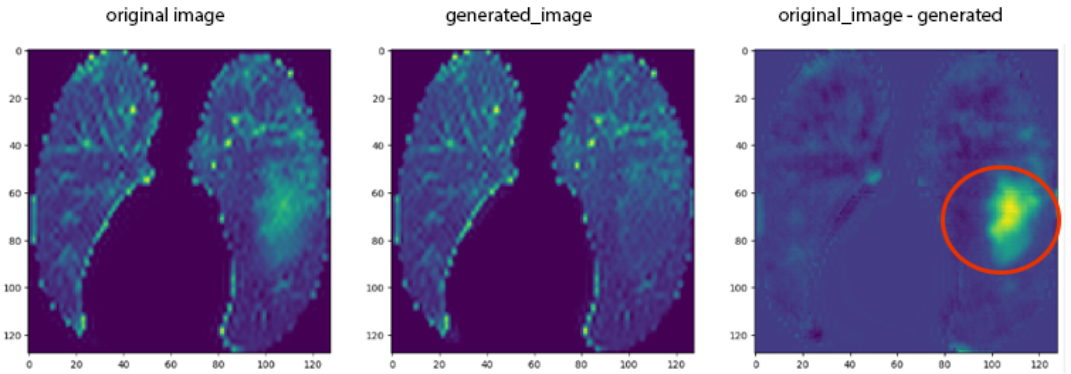

Figure 8. Radiopedia: Example of a CT image segmented with GAN.

The masks are obtained through subtraction of the healthy image generated from the infected tomographic image. Our 3D GAN successfully generates masks containing the infected areas, which are depicted by red ellipses. The following three images depict the qualitative results on the three datasets (one slice for each dataset) from the second part of our pipeline. In the subsequent stage of our pipeline, a 2D U-Net was trained using slices of pseudo-masks generated by the 3D GAN. This approach enhanced CT images segmentation. For each example in the following figures, we have the infected slice on the left and the segmentation result on the right:

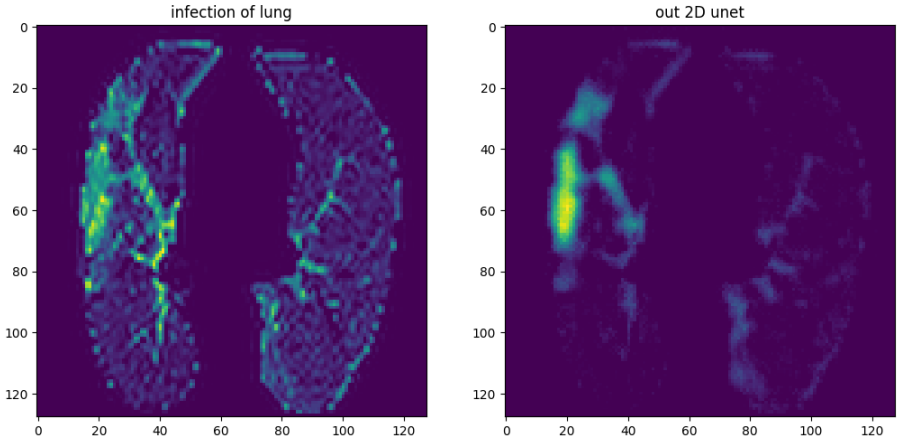

Figure 9. Mosmed: Example of a CT image segmented with GAN + U-Net.

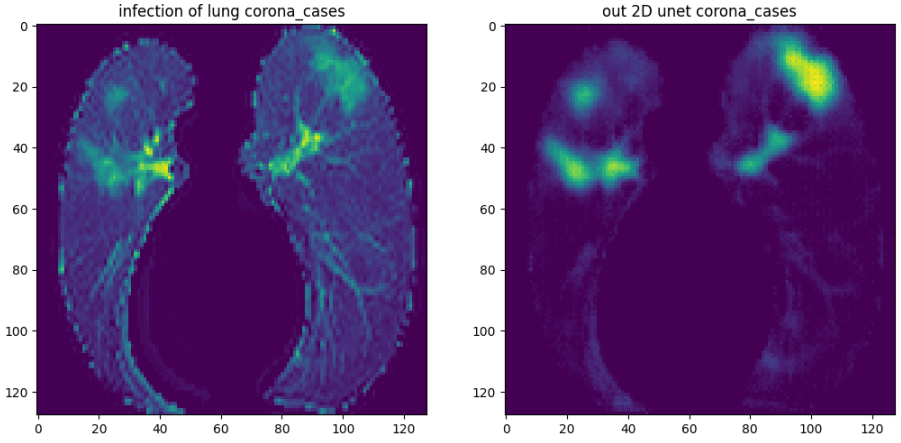

Figure 10. Coronacases: Example of a CT image segmented with GAN + U-Net.

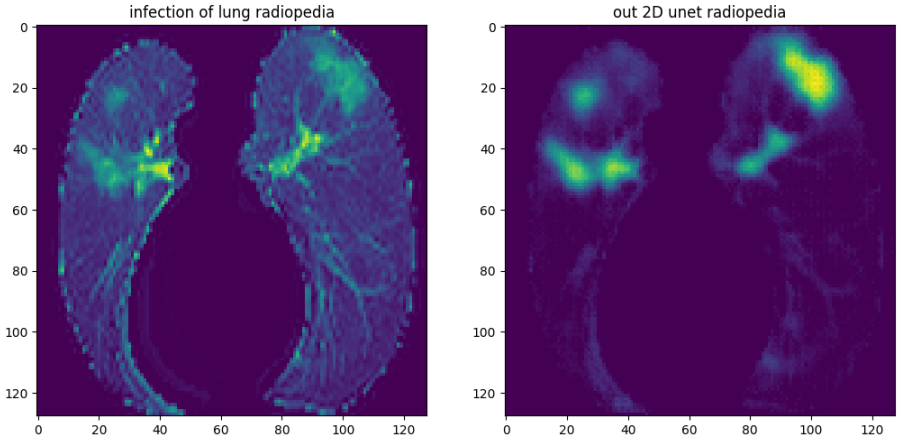

Figure 11. Radiopedia: Example of a CT image segmented with GAN + U-Net.

We will quantitatively evaluate the performance of our pipeline on three different datasets: Mosmed, Radiopedia, and Coronacases, using three performance measures: Dice score, sensitivity, and specificity. The Dice score

measures the extent of agreement or similarity between the segmented lesion and the true labels, indicat-ing how well the segmentation aligns with the true lesion boundaries. Sensitivity and specificity, on the other hand, are commonly used metrics in binary classification tasks, such as lesion detection. Sensitivity, or the true positive rate, measures the pro-portion of correctly detected lesion pixels out of all the actual lesion pixels in the ground truth. It evaluates the algorithm's ability to accurately identify lesion areas. Specificity, on the other hand, quantifies the true negative rate by measuring the proportion of correctly identified non-lesion pixels out of all the actual non-lesion pixels in the ground truth. It reflects the algorithm's capability to accurately identify areas that do not belong to the lesion. Sensitivity focuses on accurately detecting lesion pixels, while specificity emphasizes the correct identification of non-lesion pixels. Both metrics are crucial for evaluating the performance of lesion detection algorithms [28]. The quantitative results obtained are presented in Table 5.

Table 5. The results of GAN + U-Net on the 3 datasets.

Metric	Mosmed	Coronacases	Radiopedia
Dice score	62.04 ± 14.94	62.71 ± 16.08	60.23 ± 15.96
Sensitivity	69.90 ± 10.72	63.07 ± 15.12	73.45 ± 17.46
Specificity	97.62 ± 1.93	97.59 ± 2.33	97.53 ± 2.26

Through the results cited below, we observed that the use of U-Net with contrastive learning had a significant influence on the obtained results. Our approach yielded the following Dice score values: 62.04, 62.71, and 60.23 on the Mosmed, Coronacases, and Radiopedia datasets, respectively, which are significant results. For sensitivity, the model yielded the following results: 69.9, 63.07, and 73.45 on the same datasets. The following values, 97.62, 97.59, and 97.53, represent the specificity values on the Mosmed, Coronacases, and Radiopedia datasets, respectively. Our approach has demonstrated promising performances on different datasets, with a certain level of stability in the results, as shown in figure 12.

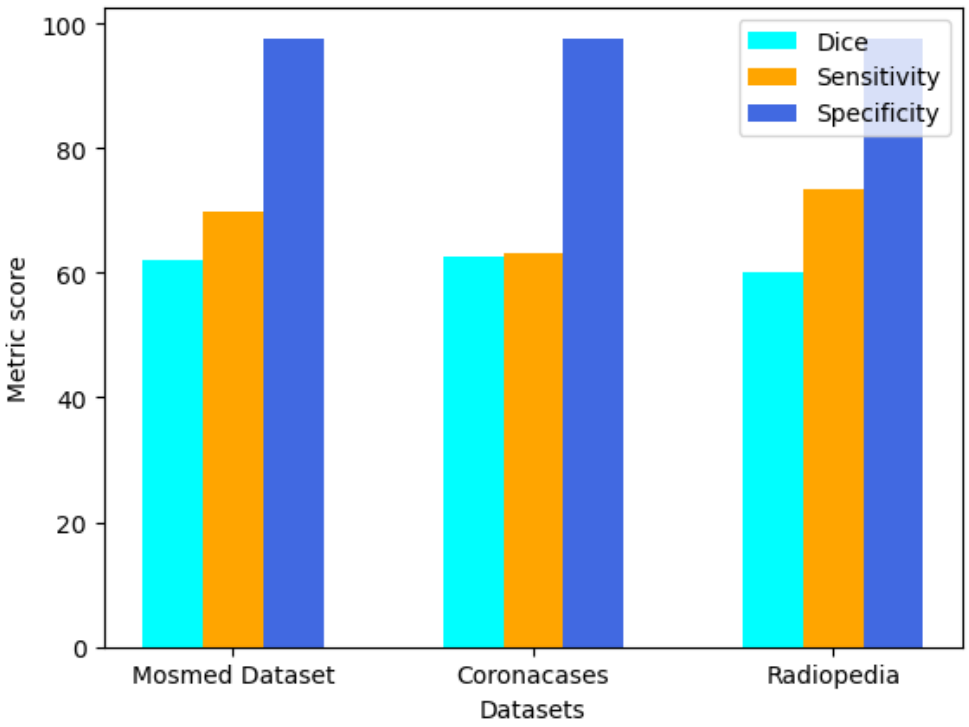

Figure 12. Visulation of the results obtained with GAN + U-Net.

This approach can be reliably and consistently used for medical image segmentation without the need for labeled data, the results demonstrate that our approach outper-forms existing weakly-supervised segmentation techniques GASNet [4] and Label-free [3] from the state-of-the-art on three datasets with a reasonable margin as shows in the table 6.

Table 6. The results of the comparison with weakly-supervised models.

Modèle	Métrique	Mosmed	Coronacases	Radiopedia
Notre Appoche	Dice	**62.04 ± 14.94**	62.71 ± 16.08	**60.23 ± 15.96**
	Sensibilité	**69.90 ± 10.72**	63.07 ± 15.12	**73.45 ± 17.46**
	Spécificité	97.62 ± 1.93	**97.59 ± 2.33**	97.53 ± 2.26
GASNet[4]	Dice	54.20 ± 22.4	-	60.20 ± 23.3
	Sensibilité	55.60 ± 28.3	-	66.80 ± 28.9
	Spécificité	**99.60 ± 0.2**	-	**99.30 ± 0.5**
Label-free [3]	Dice	-	**69.8 ± 15.2**	59.3 ± 16.9
	Sensibilité	-	**66.2 ± 22.2**	65.6 ± 18.7
	Spécificité	-	82.1 ± 8.92	58.3 ± 18.0

4. Conclusion

In our study, we proposed a novel method for CT COVID-19 image segmentation without the need for manual annotation. Our approach consisted of a pipeline with three main sub-tasks: generating a 3D pseudo-mask using a self-supervised GAN, refining the pseudo-masks using a U-Net trained with contrastive learning, and evaluating the performance on three datasets. Our results demonstrated that our method outperformed existing weakly-supervised segmentation techniques by a significant margin, highlighting its effectiveness for annotation-free medical image segmenta-tion. However, we encountered challenges due to the limited availability of healthy data for training the GAN, which could be a potential limitation in some medical applications. To address this, future work aims to improve the reliability and stability of our method. This includes exploring advanced models like TransGAN to enhance the quality of generated images and stabilize GAN training. Additionally, we envision deploying the trained model in a web or desktop application, allowing physicians to easily incorporate this annotation-free CT COVID-19 image segmentation method into their daily practice. This could enhance the speed and efficiency of diagnosis and patient management for COVID-19 cases. Moreover, we are interested in explor-ing the application of this image segmentation method in other medical domains to assist in the detection and treatment of various diseases.

REFERENCES

1. ISENSEE, Fabian, PETERSEN, Jens, KLEIN, Andre, et al, *A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Le-sions From CT Images*, In : IEEE Transactions on Medical Imaging 39.8, p. 2653-2663. doi : 10.1109/TMI.2020.3000314.
2. Wang , Guotai et al. (2020a), *nnu-net: Self-adapting frame-work for u-net-based medical image segmentation.*, arXiv preprint arXiv:1809.10486, 2018
3. Yao , Qingsong et al. (2021), *Label-free segmentation of COVID-19 lesions in lung CT.*, In : IEEE transactions on medical imaging 40.10, p. 2808-2819.
4. Xu , Zhanwei et al. (2020), *Gasnet : Weakly-supervised framework for covid-19 lesion segmentation.*, In : arXiv preprint arXiv :2010.09456.
5. Creswell , Antonia et al. (2018), *Generative adversarial networks : An overview.*, In : IEEE signal processing magazine 35.1, p. 53-65.
6. Yang , Huan et al. (2021), *Synthesizing multi-contrast MR images via novel 3D conditional Variational auto-encoding GAN.*, In : Mobile Networks and Applications 26, p. 415-424.
7. Goodfellow, Ian et al. (2020), *Generative adversarial networks*, In : Communications of the ACM 63.11, p. 139-144.

8. Schroff , Florian, Dmitry Kalenichenko et James Philbin (2015), *Facenet : A unified em-bedding for face recognition and clustering*, In : Proceedings of the IEEE conference on computer vision and pattern recognition, p. 815-823.
9. Chuang , Ching-Yao et al, *Debiased contrastive learning*, In : Advances in neural information processing systems 33, p. 8765-8775.
10. Yin , Xiao-Xia et al. (2022), *U-Net-Based medical image segmentation*, In : Journal of Healthcare Engineering 2022.
11. Ronneberger, O., Fischer, P., Brox, T. (2015), *U-net: Convolutional networks for bio-medical image segmentation*, In Medical Image Computing and Computer-Assisted Interven-tion–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.
12. Alom , Md Zahangir et al. (2018), *Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation*, In : arXiv preprint arXiv : 1802.06955.
13. Livne , Michelle et al. (fév. 2019), *U-Net Deep Learning Framework for High Perfor-mance Vessel Segmentation in Patients With Cerebrovascular Disease*, In : Frontiers in Neuroscience 13. doi : 10.3389/fnins.2019.00097.
14. Elnakib, A., Gimel'farb, G., Suri, J. S., El-Baz, A. (2011), *Medical image segmentation: a brief survey*, Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume II, 1-39.
15. Chen , Ting et al. (2020), *Big self-supervised models are strong semi-supervised learners*, In : Advances in neural information processing systems 33, p. 22243-22255.
16. mosmed.ai (2020), *https://www.kaggle.com/datasets/mathurinache/ mosmeddata-chest-ct-scans-with-covid19.*, consulted 06-05-2023
17. radiopedia.org (2020), *https://radiopaedia.org/.*, consulted 06-05-2023
18. Omir Antunes Paiva (2020) Coronacases Dataset, *https : // coronacases . org/.*, consulted 23-05-2023
19. ABADI, Ehsan, SANDERS, Jeremiah, et SAMEI, Ehsan, *Patient-specific quantification of image quality: an automated technique for measuring the distribution of organ Hounsfield units in clinical chest CT images.*, Medical physics, 2017, vol. 44, no 9, p. 4736-4746.
20. Stazio, Alice Victores, Juan Estevez, David Balaguer, Carlos. (2019), *A Study on Machine Vision Techniques for the Inspection of Health Personnels' Protective Suits for the Treatment of Patients in Extreme Isolation.*, Electronics. 8. 743. 10.3390/electronics8070743
21. Wang, Z., Zou, Y., Liu, P. X. (2021), *Hybrid dilation and attention residual U-Net for medical image segmentation.*, Computers in biology and medicine, 134, 104449.
22. Taha, A. A., Hanbury, A. (2015), *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.*, BMC medical imaging, 15(1), 1-28.
23. Belgiu, M., Drăguţ, L. (2014), *Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery.*, ISPRS Journal of Photogrammetry and Remote Sensing, 96, 67-75.
24. Baby, D., Devaraj, S. J., Mathew, S., Anishin Raj, M. M., Karthikeyan, B. (2020), *A per-formance comparison of supervised and unsupervised image segmentation methods.*, SN Computer Science, 1, 1-6.
25. Nikbakhsh, N., Baleghi, Y., Agahi, H. (2021), *A novel approach for unsupervised image segmentation fusion of plant leaves based on G-mutual information.*, Machine Vision and Applications, 32, 1-12.
26. F. Milletari, N. Navab, and S.-A. Ahmadi, *V-Net: Fully convolutional neural networks for volumetric medical image segmentation.*, in IC3DV, 2016, pp. 565–571.
27. Tropea, M., Fedele, G. (2019, October), *Classifiers comparison for convolutional neural networks (CNNs) in image classification.*, In 2019 IEEE/ACM 23rd International Symposi-um on Distributed Simulation and Real Time Applications (DS-RT) (pp. 1-4). IEEE.
28. Filali, Y., Abdelouahed, S., Aarab, A. (2019), *An Improved Segmentation Approach for Skin Lesion Classification.*, Statistics, Optimization & Information Computing, 7(2), 456-467. https://doi.org/10.19139/soic.v7i2.533.
29. Moussaoui, H., Nabil El Akkad, & Mohamed Benslimane. (2023), *A Hybrid Skin Lesions Segmentation Approach Based on Image Processing Methods.*, Statistics, Optimization & In-formation Computing, 11(1), 95-105. https://doi.org/10.19139/soic-2310-5070-1549.27.
30. Bhatt, M. S., Patalia, T. (2019), *Computer Vision Based Content-Based Image Classifica-tion System.*, Statistics, Optimization & Information Computing, 7(4), 840-853. https://doi.org/10.19139/soic-2310-5070-669.