Intelligent System for Stability Assessment of Chest X-Ray Segmentation Using Generative Adversarial Network Model with Wavelet Transforms

Omar El Mansouri[1], Mohamed Ouriha[2], Wadiai Younes[3], Yousef El Mourabit[4], Youssef El Habouz[5], Boujemaa Nassiri[6]

TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco[1, 2, 4]
Laboratory of Innovative Systems Engineering-National School of Applied Sciences of Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco[3]
IGDR, UMR 6290 CNRS, Rennes University, Rennes, France[5]
InterDisciplinaire Applied Research Laboratory- LIDRA International University of Agadir -Unversiapolis, Agadir, Morocco[6]

***Abstract*—Accurate segmentation of chest X-rays is essential for effective medical image analysis, but challenges arise due to inherent stability issues caused by factors such as poor image quality, anatomical variations, and disease-related abnormalities. While Generative Adversarial Networks (GANs) offer automated segmentation, their stability remains a significant limitation. In this paper, we introduce a novel approach to address segmentation stability by integrating GANs with wavelet transforms. Our proposed model features a two-network architecture (generator and discriminator). The discriminator differentiates between the original mask and the mask generated after the generator is trained to produce a mask from a given image. The model was implemented and evaluated on two X-ray datasets, utilizing both original images and perturbed images, the latter generated by adding noise via the Gaussian noise method. A comparative analysis with traditional GANs reveals that our proposed model, which combines GANs with wavelet transforms, outperforms in terms of stability, accuracy, and efficiency. The results highlight the efficacy of our model in overcoming stability limitations in chest X-ray segmentation, potentially advancing subsequent tasks in medical image analysis. This approach provides a valuable tool for clinicians and researchers in the field of medical image analysis.**

Keywords*—Deep learning; X-rays; segmentation; medical imaging; Generative Adversarial Networks; wavelet transforms

I. Introduction

In recent years, the application of deep learning (DL) has seen significant strides in the field of medical imaging, revolutionizing the automation and enhancement of tasks such as detection, classification, and segmentation of medical images. At the forefront of these advancements are neural networks, sophisticated machine learning algorithms designed with layers of interconnected neurons. This architectural mimicry of the human brain enables neural networks to interpret incoming data, extracting intricate patterns and features with a level of sophistication that traditional algorithms struggle to achieve [1, 2].

Neural networks are trained on annotated medical images to detect anatomical structures and abnormalities [1]. Neural networks can be used to automate disease detection, such as identifying lung nodules on chest X-rays or brain tumors on MRI images. It's an effective tool to improve the quality of medical images, with backpropagation function which removes the noise and enhance image resolution. Neural networks can detect specific anatomical structures in images, for example, to isolate a tumor in an MRI image or to measure the dimensions of an organ in an ultrasound image [1, 2]. Deep learning is widely used in medical image analysis to analyze images from a variety of modalities, such as Computed Tomography (CT), X-ray, Positron Emission Tomography (PET), Ultrasound, Magnetic Resonance Imaging (MRI), Optical Coherence Tomography (OCT) [3, 6, 7]. The X-rays are considered as a type of medical imaging that uses ionizing radiation to produce images of bones and other dense body structures. X-rays are commonly used for diagnosis and treatment planning for conditions such as fractures, joint problems, dental issues and chest [3]. They are fast, widely available, and inexpensive compared to other medical imaging modalities.

In this article, we used in our experiments the chest X-rays (CXRs) [3], a type of medical imaging that employs X-rays to create images of the chest, including the heart, lungs, and blood arteries. They are frequently employed to identify and keep track of illnesses like pneumonia, lung cancer, tuberculosis, heart issues, and fluid retention in the lungs. Compared to other imaging techniques used in medicine, CXRs are quick, accessible, and affordable [4]. The detection of COVID-19 has also been tried in this manner [4]. Chest X-rays (CXRs) are used more frequently for early triage of ARDS patients who are already exhibiting COVID-19 symptoms and acquiring accurate because they are simpler to obtain than computed tomography (CT) [5].

Segmentation of a Chest X-ray is the process of separating the foreground (such as the lung and heart) from the background (such as the chest wall) in a chest X-ray image. This can be done using various techniques such as thresholding, morphological operations, and machine learning algorithms [6]. The purpose of chest X-ray segmentation is to accurately identify and isolate specific structures in the image,

which can be used for diagnosis, treatment planning, and measurement of structures such as lung volume. Segmenting chest X-rays is a challenging task due to the diverse range of anatomy, disorders, and appearances that can be present. However, it is a crucial step for enhancing the accuracy and efficiency of medical image analysis [8]. On the other hand, the stability of images segmentation is very important to improve the data security against attackers on medical images field. Images contain sensitive and confidential patient information, and their unauthorized access or manipulation can have serious consequences [9]. Segmentation of medical images, such as chest X-rays, can help improve their security against attackers by reducing the amount of data that needs to be protected. By isolating relevant structures and removing unnecessary information, the risk of unauthorized access or manipulation of sensitive and confidential patient information can be reduced. To further protect the segmented images, encryption can be used to scramble the data using secure algorithms. This makes it difficult for unauthorized users to access or view the confidential information contained in the images. Access to segmented images can also be controlled through authentication and authorization procedures, such as role-based access control or password protection. This helps ensure that only authorized users have access to the data, further reducing the risk of data breaches.

It's important to note that while these techniques can provide some level of security against attackers, no single method can provide complete security. A combination of these techniques, along with robust security protocols throughout the life cycle of medical X-ray images, is recommended for the best results [10]. Deep learning algorithms have shown promising results in both segmentation tasks and security against potential attackers, particularly for chest X-rays.There are several deep learning algorithms used for medical image segmentation, including Convolutional Neural Networks (CNNs). [10] CNNs are a type of neural network specifically designed for image analysis tasks. They can be trained to perform image segmentation by learning the relationships between image pixels and object boundaries. CNNs have achieved notable success in semantic segmentation in [11], Long et al. Semantic segmentation was suggested in early work on integer convolutional networks (FCNs). Where the authors have tested the label map using deconvolutional layers to acquire classification results for each pixel, they have substituted the conventional fully connected CNN layers with convolutional layers to generate a coarse label map. However, CNNs have some drawbacks, they present good results with clear and well-defined images and poor performance on security applications, typical low-resolution, noisy, or occluded images [11]. The U-Net algorithm [12], a well-known CNN architecture created expressly for medical image segmentation, is the alternative segmentation algorithm. In order to record both high-level and low-level characteristics in the image, it employs an encoder-decoder architecture.

II. Related Work

The section provides an overview of existing research, methodologies, and advancements relevant to this study. By examining prior contributions, this section highlights the current state of knowledge, identifies gaps in the literature, and demonstrates how this work builds upon or diverges from earlier studies. The review also contextualizes this research within the broader academic discourse, ensuring clarity in its contributions and alignment with ongoing developments in the field. In recent years, medical image segmentation has become a critical area of research in medical imaging and computer vision. Numerous studies have proposed innovative methods to improve the accuracy and reliability of segmentation models, especially for challenging tasks involving small, unclear, or complex anatomical structures. The U-Net architecture for medical image segmentation is extended in the work for auteur [12], introducing an attention mechanism to enhance performance on challenging objects like the pancreas. In order to enable the network to recognize key features for the target item, the proposed Attention U-Net employs a gating mechanism to dynamically weight feature mappings in the network's encoding and decoding sections. For a dataset of CT scans used for pancreatic segmentation, the Attention U-Net performed better than the regular U-Net and a number of cutting-edge segmentation techniques. Auteur [12] presents, a deep learning framework for U-Net architecture-based segmentation of chest X-ray images. To more effectively handle the present small and hazy structures, the authors suggested changes to the basic U-Net architecture. For improving the accuracy and realism of the segmentation results, they used adversarial segmentation, it's a type of image segmentation technique that uses adversarial training. Adversarial training is a machine learning technique that trains a model using adversarial examples, which are inputs designed to mislead the model. In the context of image segmentation, adversarial training is used to guide the segmentation model towards producing accurate and visually plausible segmentation masks. The model is trained using a combination of a segmentation loss and an adversarial loss, which measures the realism of the generated segmentation masks. Authors of [13], explore the application of SegAN for medical image segmentation and suggest using adversarial training to increase the segmentation accuracy and realism of the segmentation masks created. Two parts make up SegAN: a generator and a discriminator. The segmentation masks are created by the generator, and the discriminator assesses how realistic they are. In order to increase the segmentation accuracy, the authors also suggest a multi-scale loss function that takes into consideration background data from several scales. In order to train the SegAN, the multi-scale loss function is combined with the adversarial loss. For numerous medical imaging datasets, experimental results demonstrate that SegAN is superior to other cutting-edge approaches. Additionally, the authors demonstrate how SegAN can deal with small and hazy structures. The authors of [14] demonstrate how SegAN can be used for unsupervised adversarial training on medical image segmentation. Using a number of medical imaging datasets, they assess the performance of the unsupervised SegAN and demonstrate that it can produce results that are on par with or even superior to those of the conventional supervised SegAN.

The paper in [15], presents an interesting extension of SegAN by incorporating unsupervised adversarial training, which can help address the limitations of supervised adversarial training, such as the need for large amounts of annotated data. However, more research is needed to fully

understand the benefits and limitations of unsupervised adversarial training for medical image segmentation. While the unsupervised adversarial training approach has advantages, it comes with limitations. In this method, the generator is trained without using ground truth masks, potentially resulting in suboptimal performance. This differs from supervised adversarial training, where ground truth masks are employed. Generative Adversarial Networks (GANs) pose challenges in training and can experience stability issues, such as mode collapse, leading to suboptimal performance. The authors evaluated the performance of unsupervised SegAN on various medical imaging datasets. However, it's important to note that these datasets may not fully represent all medical imaging modalities and applications. This version maintains the key points while presenting them in shorter, more digestible sentences. Authors in study [16], present a new approach to medical image segmentation that leverages the advantages of both multimodal imaging and adversarial learning. The authors propose a multi-modal GAN called SegAN, which is trained using both modalities and adversarial learning. The network is trained to segment the target structures in both modalities while maintaining the multi-modal information in the generated results. However, the limitations of the proposed approach include the need for large amounts of annotated data for training, the stability issues associated with GANs, and the lack of interpretability of the model's predictions. In study [17], to improve the performance of adversarial networks for medical image segmentation tasks, the authors propose to use self-supervised learning, where the model can learn from the input data without requiring manual annotations, The SS-GAN is trained using a combination of adversarial loss and self-supervised loss, which helps the network to generate accurate and plausible segmentation mask. The experimental results demonstrate that the proposed SS-GAN's more effective than other state-of-the-art methods for medical image segmentation. The limitations are not specified in the article [18]. However, some common limitations of self-supervised learning and adversarial networks can be the model's performance might be affected by the presence of outliers or data with different distributions, the model may not perform well on unseen data and generalize poorly to new medical imaging modalities, the stability issues associated with GANs, and the lack of interpretability of the model's predictions.

A new paradigm for medical image segmentation employing an adversarial attention network is proposed by S. Kim et al. in [19] (AAN). The AAN enhances segmentation performance by combining adversarial learning with an attention mechanism. Within the network, both a generator and a discriminator coexist. The generator divides the image into segments, while the discriminator assesses the segmentation output. The generator is given access to the attention mechanism to aid with segmentation accuracy and focus on the target regions. The technique was put to the test on several medical image datasets, and the results indicated enhanced performance when compared to other segmentation techniques already in use. X. Song et al. [20], proposes a multiple adversarial network (MAN) architecture for medical image segmentation. The MAN architecture is made up of a number of adversarial sub-networks, each of which is trained to provide various aspects of the input medical image. These features are then combined to give the segmentation result. The results of the experiments the authors undertake on two medical image segmentation datasets show that the MAN architecture is superior to other cutting-edge techniques. The paper [21], does not provide any constraints. However, some typical drawbacks of self-supervised learning and adversarial networks include the instability of GANs, the unintelligibility of the model's predictions, and the model's performance being impacted by the existence of outliers or data with different distributions. In study [22], authors presents a decision support system based on a GAN model for brain tumor segmentation. The authors aim to improve the accuracy and efficiency of brain tumor segmentation using GANs. The proposed system is tested and evaluated using a dataset of brain MRI scans, and the results demonstrate its effectiveness and superiority compared to traditional methods. Some common limitations of the proposed system include stability issues associated with adding noise to the images. The model may not perform well when incorporating such noise.

Through our review of several studies focusing on the segmentation by GAN's algorithms, we discovered that the biggest limitations was the stability constraints related to the GAN's network. In this paper, we proposed a new approach to chest X-ray segmentation based on the GAN with wavelet transforms. The GAN consists of two neural networks: the generator and the discriminator. The generator is responsible for generating a mask for a given original chest X-ray image, while the discriminator distinguishes between the original mask and the generated mask. To evaluate the proposed approach, we implemented their model on two datasets of chest X-ray images. One dataset consisted of original images, while the other dataset consisted of perturbed images with added noise using the Gaussian noise method [38]. The performance of the proposed approach was compared to that of the traditional GAN algorithm. The results showed that the proposed approach based on the GAN algorithm with wavelet transforms outperformed the traditional GAN algorithm in terms of stability, accuracy, efficiency, and it can help to improve the accuracy of subsequent medical image analysis tasks. The proposed approach has the potential to be a valuable tool for clinicians and researchers in the field of medical image analysis. It can help improve the accuracy and efficiency of medical image analysis tasks, which can ultimately lead to better diagnosis and treatment outcomes for patients. The rest of this paper is organized as follows: In Section III, we show methods. The proposed approach of our work in Section IV. The experimental results are discussed in Section V. Finally, a conclusion are presented in Section VI.

III. Methodology

This section, provides an overview of the GAN architecture. Subsequently, we will introduce two models for GAN-based segmentation, with the first model being our initial approach and the second model being a novel technique to overcome stability limitations of GAN.

A. Generative Adversarial Networks (GAN)

Goodfellow et al. [23] were the ones who first proposed the traditional GAN architecture. When creating new synthetic data that is similar to a training set, GANs combine two

separate types of neural networks generator G and a discriminator D. The role of the D is to learn how to distinguish between genuine and fake samples, whereas the G is responsible for generating artificial samples. The min-max game involves two networks competing against each other, where one network aims to maximize the value function 1 while the other seeks to minimize it. In a game-theoretic framework, the two networks are trained concurrently while the generator strives to generate samples that deceive the discriminator, and the discriminator tries to properly identify the generated samples. The discriminator gets better at spotting fraudulent samples, and the generator gets better over time at producing synthetic samples that are close to the training set.

Below, Formula 1 present an illustration of how network G and network D can be understood mathematically with value function V (G, D):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{Z\sim p_Z(Z)}[\log(1-D(G(Z)))] \quad (1)$$

where:

- $\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)]$: log probability of D predicting that real-world data is real.
- $\mathbb{E}_{Z\sim p_Z(Z)}[\log(1-D(G(Z)))]$: log probability of D predicting that G's generated data is not real.
- x: with real images drawn from $p_{data}(x)$ real data images distribution.
- Z: the prior input noise from $p_Z(Z)$.

GANs can be used for a variety of tasks, like image synthesis by use type of GAN called conditional-GAN (cGAN) [24], text to image generation [24], Data augmentation [25], classification, and segmentation [39].

B. GAN for Image Segmentation

The discriminator network is trained to differentiate between real and fake (mask segmented) masks, while the generator network is trained to create realistically segmented images known as mask segmented. The two networks play a game in which the generator tries to produce masks that the discriminator cannot distinguish from real ones, and the discriminator tries to correctly identify the generated masks. Over time, the generator improves, producing more and more realistic masks, while the discriminator becomes better at identifying generated masks. In this way, GANs can be used for image segmentation by training the generator to produce accurate segmentations of input images. The general structure of the GAN for segmentation is shown in Fig. 1.

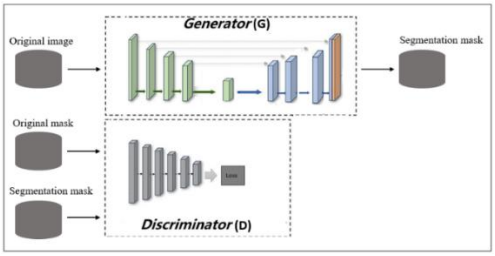

Fig. 1. Architecture GAN.

a) Generator: Our goal is to enhance the preliminary: segmented output of the segmentation component by leveraging the generative component, which is known for its advanced optimization capabilities. To achieve this, we employed the U-Net neural network as the generator, given its exceptional performance in previous works [23]. A U-Net generator for segmentation is a type of deep learning neural network architecture designed for image segmentation tasks. The U-Net architecture is based on a fully convolutional network, with a "U" shaped structure that merges feature from lower-level to higher-level layers in the network [12]. In the U-Net architecture, the contracting path, represented by the left half of the "U" shape called encoder, is used to extract features and reduce the spatial resolution of the input image, and the expanding path, represented by the right half of the "U" shape called decoder, is used to upsample and recover the spatial resolution, while also combining features from the contracting path to make segmentation predictions [12]. The U-Net generator is commonly used in medical imaging to segment organs, tumors, or other structures of interest. However, it can also be applied to other types of image data, such as satellite imagery or aerial photographs, for tasks such as object detection and semantic segmentation. The general structure of the generator in our model is shown in Fig. 2.

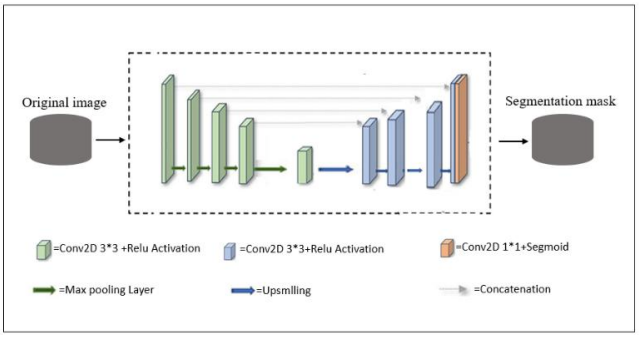

Fig. 2. Structure of the generator.

The network is composed of five layers for the encoder and decoder: first one is the input layer, next three layers of the Relu convolutional batch normalization model are hidden, and the last one is the final output layer. The architecture is comprised of a total of 10 layers, with the encoder and decoder each containing five layers. The encoder consists of five layers, with each layer being composed in the following manner: The first layer contains an input size of $512 * 512$, Conv2D is a 2D convolution layer with a $3 * 3 * 32$ filter size for identifying spatial patterns in an image, such as edges and activation the max pooling layer of size $2 * 2$ is used to minimize representation size and speed up computations in Relu, a piecewise linear function that outputs the input directly if it is positive. With a multiplication filter $*2$, the remaining layers share the same structure. The decoder constitutes the latter half of the architecture, comprising of five layers. Each compound layer is structured in the following manner: Transposition-based deconvolution is an upsampling technique used to increase the size of an image. To achieve this, the transposed convolution is applied, and the resulting image is concatenated with the corresponding image from the contracted path. This

process creates an image of the same size as the input. In the Conv2D class, the padding parameter can have one of two values: "valid" or "same". When set to "valid", convolution can be applied to reduce the spatial dimensions of the input volume, assuming that it is not empty. When set to "same", convolution can be applied to reduce the spatial dimensions, and the input volume is assumed not to be empty. Including prior information in the process increases the accuracy of the output. In our work, we utilized "same" padding, and the final layer consists of a convolutional layer with a filter size of 1 × 1. The output of the sigmoid activation function is specified on real numbers and ranges between 0 and 1. It can be interpreted as a probability value [26].

b) Discriminator: Within our work, the discriminator: is implemented as a deep convolutional neural network that takes in two images one that is original and another that is generated by the generator. The objective of the discriminator is to categorize each image as either "real" or "fake". Fig. 3 provides a visual representation of the overall structure of the discriminator that was employed in our work.

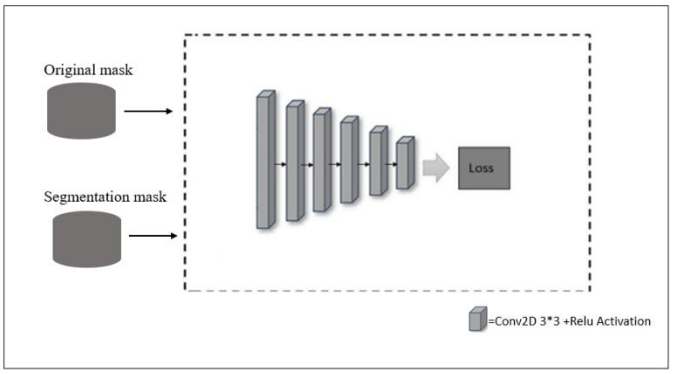

Fig. 3. Structure of the discriminator.

In our work, the discriminator examines two sequences to determine whether they are real or fake by using two images: a original mask and a mask produced by the generator. For the discriminator in our model, we used a total of five layers, starting with the input layer, followed by four Conv2D layers each with a size of 3 × 3. Hidden and dropout layers were incorporated, implementing the Relu activation function and batch normalization. The fifth layer was another Conv2D layer, this time with a size of 3 × 3, utilizing the Relu activation function, and employing max pooling layers. The 2-dimensional arrays resulting from the pooled feature maps are then flattened in the final layer before being passed on as a continuous linear vector to the fully connected layer, which generates the final output [27, 28].

C. Wavelet Transforms

A subfield of mathematics known as the wavelet transform (WT) started to develop gradually in the 1980s [40]. The wavelet transform is another outstanding example of the ideal fusion of pure and applied mathematics, following the fourier transform. It shares the mathematical microscope renown. The wavelet transform has significantly advanced approaches in nonlinear science, engineering technology, signal processing, image processing, and computer applications in recent years. It is one of the most effective and popular time frequency analysis techniques, and it has been utilized extensively in signal and image processing [31, 32]. WT is classified into two types: continuous WT (CWT) and discrete WT (DWT). In general, DWT is more useful than CWT for resolving practical issues. We used the Discrete Wavelet Transform (DWT) to implement our new approach. The DWT can be used to decompose signals into different frequency components and provide a multiresolution representation of the signal [32]. This can be useful in image processing and computer vision applications, where the goal is to extract important features and remove noise from the signals [29, 32]. The data is split into high and low pass bands by the DWT algorithm, with or without information loss. It is built on high and low pass filters that are sub-sampled. Functions over a finite range are defined as DWT. The goal of DWT is to convert data from the time-space domain to the time-frequency domain in order to improve compression efficiency. We have simplified the DWT by describing how it works as follows in Fig. 4:

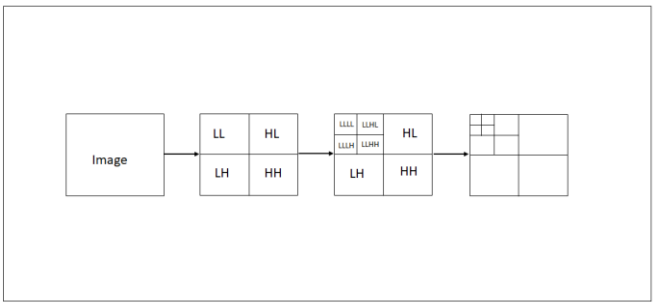

Fig. 4. Block diagram of 3-level wavelet transform.

Low-Low (LL), Low-High (LH), High-Low (HL), and High-High are the names of the four bands that are created (HH). As the LL band still contains information that resembles an image, it is possible to use the same set of wavelet filters that were used on the original image. The image can be divided into sub bands for as many levels as needed (up to the image resolution), however for image compression, only 4 or 5 levels are often used [33]. An image is a two-dimensional (2-D) signal with rows and columns. The wavelet decomposition of an image (rows and columns) can be represented by decomposing one-dimensional signals. The low-frequency components in horizontal and vertical directions (LL), low-frequency components in horizontal and vertical directions (LH), high-frequency components in horizontal and vertical directions (HL), and high-frequency components in horizontal and vertical directions (HH) are obtained following a one-layer wavelet decomposition of an image [33]. The DWT performs two steps of calculation on pairs of data items from the signal after it has been processed, as presented on following Formula 2.

$$\begin{cases} L_i = \frac{1}{\sqrt{2}}(X_{2i} + X_{2i+1}) \\ H_i = \frac{1}{\sqrt{2}}(X_{2i} - X_{2i+1}) \end{cases} \quad (2)$$

An approximation (average) of the two data items is what comes out of the first step, and an approximation (difference) of the two data items is what comes out of the second step.

Where :

- i=0...N/2
- X=the signal data of length N.
- L=the low subband of length N.
- H=the high subband of length N.

The inverse Wavelet transform (IDWT), in contrast, fuses four sub-images to the original image by up-sampling while utilizing the same filters. As shown on Formula 3.

$$\begin{cases} X_{2i} = \frac{1}{\sqrt{2}}(L_i + H_i) \\ X_{2i+1} = \frac{1}{\sqrt{2}}(L_i - H_i) \end{cases} \quad (3)$$

a) Wavelet transforms for generative adversrial network: The GAN-based image segmentation model: that incorporates wavelet transforms is a deep learning architecture that combines two powerful techniques to achieve accurate segmentation results. In this approach, the wavelet coefficients of the input image are used as features to train a GAN model, which generates a segmented image output. The generator network is designed to optimize the segmentation performance by minimizing the difference between the generated output and the ground truth segmentation mask, while the discriminator network objective is to distinguish between the generated segmentation mask and the ground truth. By iteratively training the generator and discriminator, the GAN model can learn to accurately segment complex images [30]. A GAN based on Wavelet coefficients can help improve the stability of image segmentation by using the wavelet coefficients as inputs to the network. The wavelet transform allows the decomposition of an image into different frequency components, which can provide a more robust representation of the image compared to using raw pixel values. By training the GAN on these wavelet coefficients, it can learn to produce stable segmentations even with variations in the input image. This approach can help improve the robustness and generalization ability of the image segmentation model, it's that's what we actually got by implementing our new approach based on a GAN with a wavelet and comparing it to a simple GAN.

IV. Proposed Approach

Using DWT and IDWT in our approach, we were able to develop a reliable segmentation model that could handle variations in input images. Our results indicate that our approach enhanced the robustness and generalizability of the original image segmentation model. Unlike the original generator that used convolutions to transform its 3 channel input image, Fig. 2. We used convolutions to convert the original images into 12-channel and 16-channel data within the generator, along with DWT and IDWT. We then employed convolutions iteratively, replacing max pooling and upsampling with DWT and IDWT, respectively. By adopting DWT, we could retrieve multiscale edge characteristics while reducing the data by a fourth and multiplying the number of channels by four. We adjusted the convolution stride to reduce the data size, unlike the original method with max pooling, which discards partial original data and loses minor edge features. We also substituted DWT for the sampling operations max pooling to enhance the segmentation capabilities of our discriminator. Fig. 5 illustrates the general structure of our generator using the wavelet transform.

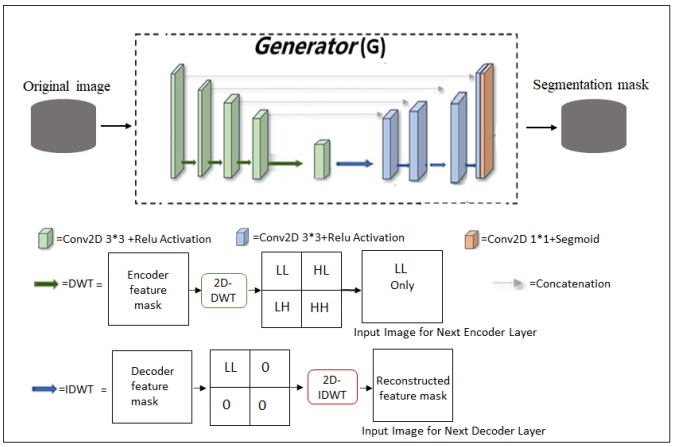

Fig. 5. Generator with wavelet transforms structure.

In our study, we employed two approaches. Firstly, we compared the effectiveness and advanced nature of two models: GAN and GAN with DWT. This evaluation aimed to assess the performance and capabilities of the proposed approach. Secondly, we introduced noise to the testing data and compared the performance of the two models to address stability limitations. This comparison was conducted to analyze the models' ability to handle and mitigate the effects of noise.

A. Evaluation Metrics

In our work, we used three metrics below to assess the effectiveness of our approach.

1) Accuracy [34]: is the ratio of correct predictions to the total number of predictions made by the classifier and can be found using Eq. (4).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

Where:

- TP (True positive): The model predicts an already present object.
- FP (False positive): The model predicts an object that is not actually present.
- FN (False negative): The model is unable to predict an already present object.
- TN (True Negative): The model correctly predicts the negative class.

2) IoU (Intersection over union) [35]: The intersection over union also known as Jaccard is the area of overlap between the predicted mask and the original mask divided by the area of union between the two sets and can be found using Eq. (5).

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} = \frac{Areaof Overlap}{Areaof Union} \quad (5)$$

3) DSC (Dice Score Coefficient) [34]: The Dice Score Coefficient or what is said as the F1-Score is the arithmetic mean of precision and recall and can be used to better assess the segmentation effect and can be found using Eq. (6).

$$\text{DSC} = \frac{2\times\text{P}\times\text{R}}{\text{P+R}} \quad (6)$$

Where:

P is the proportion of samples among the correctly categorized samples that are classed as positive samples.

$$P = \frac{\text{TP}}{TP+FP} \quad (7)$$

R is a measure of how many correctly identified positive samples there are compared to all positive samples.

$$R = \frac{TP}{TP+FN} \quad (8)$$

Simply put, the Dice Coefficient is 2 ∗ the Area of Overlap between the predicted mask and the original mask divided by the total number of pixels in both images.

$$\text{DSC} = \frac{2\times\text{AreaofOverlap}}{\text{Totalnumberofpixels}} \quad (9)$$

B. Datasets

For our experiments, we used two datasets for training and testing:

- The first is the Shenzhen Hospital (SH) dataset [36],

which consists of 662 images, 336 of which are from abnormal individuals displaying various tuberculosis signs, and 326 from healthy individuals. Hospitals in Shenzhen, Guangdong province, China, collected the JPEG formatted X-ray images used in this data set. These segmentation masks for the Shenzhen Hospital X-ray Set were manually created by in structors and students from the National Technological University of Ukraine's Computer Engineering Department of the Faculty of Informatics and Computer Engineering.

- The second, COVID-19 Chest X-ray images and

Lung masks Database [37]. It contains chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images in cooperation with medical professionals, created by a group of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh, as well as collaborators from Pakistan and Malaysia. The collection includes 2905 chest X-ray (CXR) images with the segmentation mask, 219 of which are from abnormal patients showing COVID-19 positive, 1341 normal, and 1345 images with viral pneumonia.

We scale the images in the training and test sets to 512*512 for training and testing. We use 20% for testing, 10% for validation, and 70% for training. A train-test split was used to divide the data as shown in Table I.

TABLE I. DATASETS

Datasets	Total images	Format	Original Dimension	Training	Testing	Validation
SH	662	JPEG	(512,512,3)	70%	20%	10%
COVID-19	2905	PNG	(512,512,3)			

C. Experimental Setup and Training Parameters

After data enhancement, we trained two GAN models (Traditional GAN and GAN with DWT) for 200 epochs. Experiments were conducted on a server equipped with a GPU P100. The two models we provided for our dataset were trained on 200 epochs with a batch size of 3, and one model's training might take up to 6 hours. The implementation of our work is based on the Keras and Tensorflow libraries. We used the Adam algorithm to optimise the network with a learning rate of 0.0003 and a decay rate of 0.5.

V. Results and Discussion

A. Training Results

We used the two architectures on both the SH dataset and the COVID-19 dataset, and we compared them to evaluate the effectiveness and advanced nature of the proposed approach. The following Fig. 6 and Fig. 7 present the results of the training and validate the Dice score coefficient of our models.

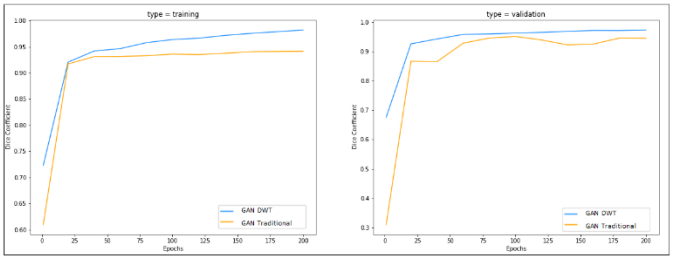

Fig. 6. Training results in the SH dataset.

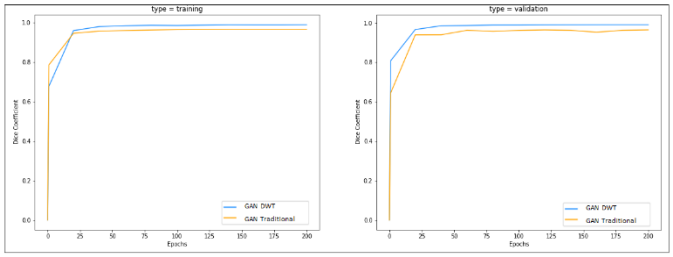

Fig. 7. Training results in the COVID-19 dataset.

It can be seen from the above results. It shows that our algorithm GAN with DWT has better robustness and stability and has given the best performance for the training in two datasets. So, it can be said through our comparison between the two training models in these datasets, a GAN based on Wavelet coefficients can help improve the performance for the training of image segmentation.

B. Stability Results

To see if GAN with DWT prevents the stability limit of a GAN Traditional. We have added multiple ε levels of noise density to the original images in the testing phase. The Table II flow shows the results of testing the dice score coefficient of our models in two datasets at every ε level of noise. To add ε noise into the testing data, we utilized the Gaussian noise method. This approach entailed introducing random values sampled from a Gaussian distribution and adding them to the original data. By employing the Gaussian noise method, we were able to simulate realworld noise sources and introduce random variations characterized by a normal distribution. To apply Gaussian noise to an image, we use the following equation:

$$\mathbf{noisy_{image}} = \mathbf{original_{image}} + \boldsymbol{\varepsilon} \times \mathbf{N(0, sigma^2)} \quad (10)$$

- $\text{noisy}_{\text{image}}$:represents the resulting image after adding Gaussian noise.
- $\text{original}_{\text{image}}$: is the original input image.
- ε : is a scaling factor that controls the intensity of the noise.
- $N(0, \text{sigma}^2)$: represents a random variable drawn from a Gaussian distribution with.

TABLE II. STABILITY ON TESTING PHASE

	SH Dataset		COVID-19 dataset	
ε	GAN DWT	GAN Traditional	GAN DWT	GAN Traditional
0	0.9685	0.9384	0.9906	0.9744
0.10	0.9685	0.0071	0.9904	0.0012
0.20	0.9635	0.0068	0.9904	0.0012
0.30	0.9636	0.0067	0.9900	0.0011

We used the two architectures on both the SH dataset and the COVID-19 dataset, and we compared them to evaluate the effectiveness and advanced nature of the proposed approach. The following Fig. 8 present the results of Stability on testing phase our models.

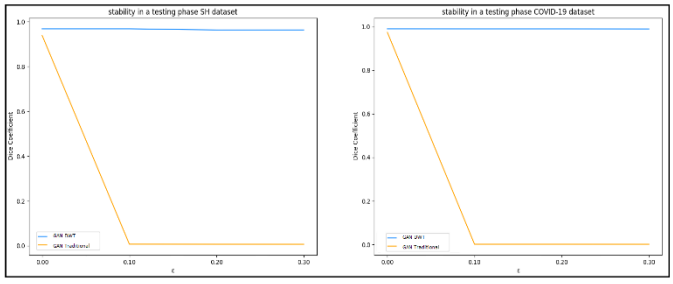

Fig. 8. Stability on testing phase.

According to the results, when adding ε=0.10, 0.20, and 0.30 levels of noise density to the original images in the testing phase. The performance in the dice score coefficient of algorithm GAN with DWT remained stable compared to GAN Traditional. So, it can be said through our comparison between the two models in these datasets, a GAN based on Wavelet coefficients can help to prevent stability limitation in image segmentation of the GAN Traditional. The prediction segmentation results of the two models for datasets SH and COVID-19 for ε levels of noise density added are shown in the Fig. 9 to 16 are as follows:

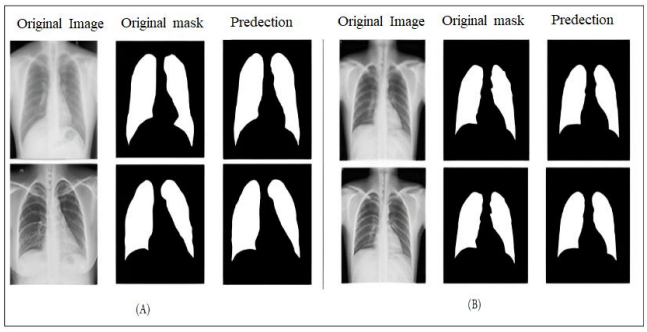

Fig. 9. Results of SH dataset with 0 noise added: (A) GAN with DWT; (B) GAN traditional.

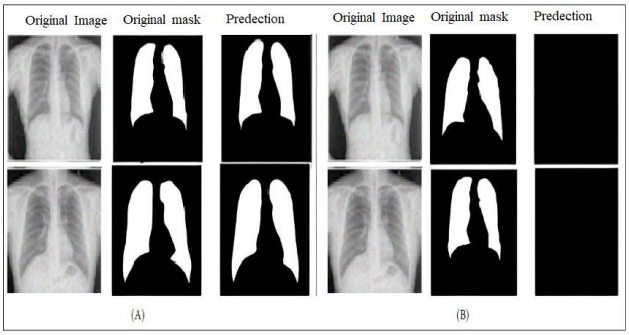

Fig. 10. Results of SH dataset with 0.10 noise added: (A) GAN with DWT; (B) GAN traditional.

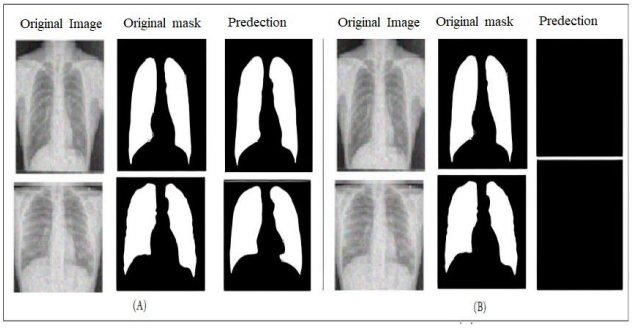

Fig. 11. Results of SH dataset of 0.20 noise added : (A) GAN with DWT; (B) GAN traditional.

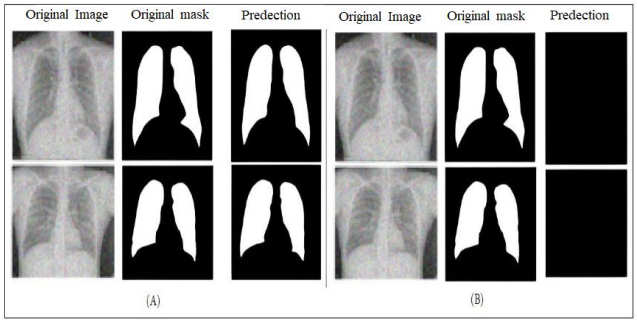

Fig. 12. Results of SH dataset of 0.30 noise added: (A) GAN with DWT; (B) GAN traditional.

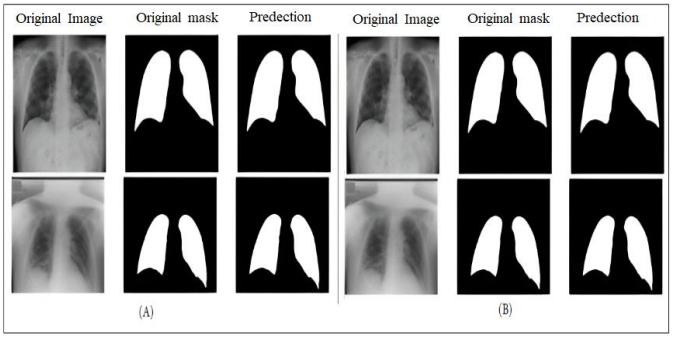

Fig. 13. Results of COVID-19 dataset with 0 noise added: (A) GAN with DWT; (B) GAN traditional.

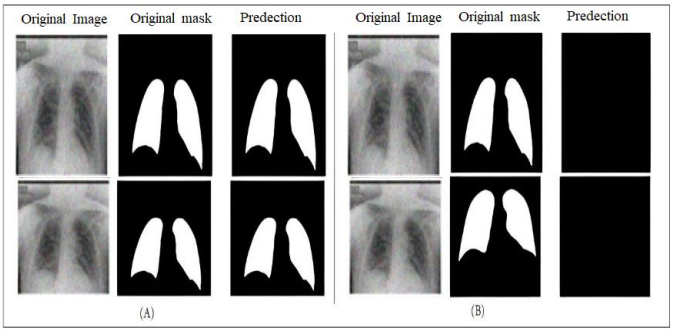

Fig. 14. Results of COVID-19 dataset of 0.10 noise added. (A) GAN with DWT; (B) GAN traditional.

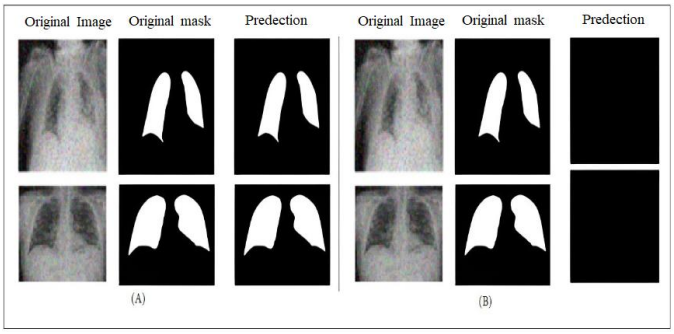

Fig. 15. Results of COVID-19 dataset with 0.20 noise added: (A) GAN with DWT; (B) GAN traditional.

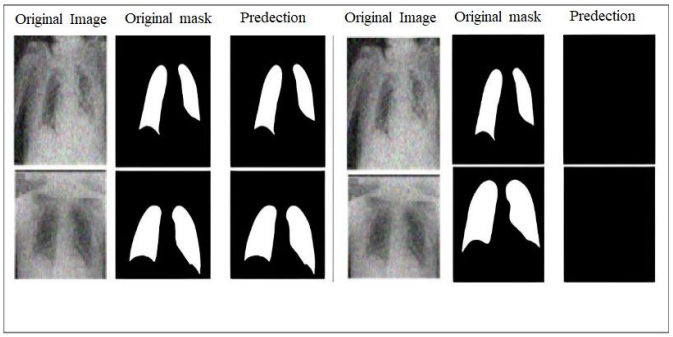

Fig. 16. Results of COVID-19 dataset with 0.30 noise added: (A) GAN with DWT; (B) GAN traditional.

VI. Conclusion

This paper present an innovative segmentation approach that leverages the power of GAN combined with wavelet transforms. The objective was to tackle stability issues commonly encountered in the image segmentation process when using GAN Traditional. By incorporating wavelet transforms into the approach, we aimed to enhance the robustness and reliability of the segmentation results. The wavelet transforms played a crucial role in effectively capturing multi-scale features and spatial details within the images, enabling more accurate segmentation outcomes. Additionally, we recognized that image segmentation with traditional GAN can introduce instability, resulting in artifacts and inaccurate boundaries. To overcome this challenge, our proposed approach offered an alternative approach to pooling that mitigated stability issues, thereby improving the overall quality and consistency of the segmentation results. Our approach results demonstrate the effectiveness of our approach to achieve high segmentation performance compared to existing methods. The integration of GANs, wavelet transforms, and the alternative pooling technique showcased the potential for significant advancements in image segmentation accuracy and stability. The proposed approach not only provided a robust solution to stability challenges but also opened up new avenues for exploring the synergistic benefits of combining GANs, wavelet transforms, and alternative pooling strategies in the image segmentation field.

References

[1] Suganyadevi, S., Seethalakshmi, V. Balasamy, K., 2022. A review on deep learning in medical image analy-sis. Int. J Multimed. Info. Retr. 11, 19-38.

[2] Zhou, S. Kevin, Hayit Greenspan, and Dinggang Shen, eds, 2017: Deep learning for medical image analysis. Academic Press.

[3] Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.J., Kim, N., 2019. Deep Learning in Medical Imaging. Neurospine. 16(4), 657.

[4] Hendee, W. R., Chien, S., Maynard, C. D., Dean, D. J., 2002. The National Institute of Biomedical Imaging and Bioengineering: history, status, and potential impact. Annals of biomedical engineering, 30, 2-10.

[5] Bassi, P. R., Attux, R. 2021. A deep convolutional neural network for COVID-19 detection using chest X-rays. Research on Biomedical Engineering, 1-10.

[6] Bhattacharya, S., Maddikunta, P. K. R., Pham, Q. V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., Piran, M. J. 2021. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 65, 102589.

[7] Pirimoglu, B., Sade, R., Ogul, H., Kantarci, M., Eren, S., Levent, A. 2016. How can new imaging modalities help in the practice of radiology?. The Eurasian journal of medicine, 48(3), 213.

[8] Smith, J., Johnson, A., Brown, C., 2022. Segmentation Techniques for Chest X-ray Images. Medical Imaging Journal, 15(3), 123-137.

[9] Jacobi, A., Chung, M., Bernheim, A., Eber, C. 2020. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. Clinical imaging, 64, 35-42.

[10] Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., Ng, M. Y. 2020. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology, 296(2), E72-E78.

[11] Magdy, M., Hosny, K.M., Ghali, N.I. et al. 2022. Security of medical images for telemedicine: a systematic review. Multimed Tools Appl 81, 25101–25145.

[12] Muhammad, K., Hussain, T., Tanveer, M., Sannino, G., de Albuquerque, V. H. C. 2019. Cost-effective video summarization using deep CNN with hierarchical weighted fusion for IoT surveillance networks. IEEE Internet of Things Journal, 7(5), 4455-4463.

[13] Long, J., Shelhamer, E., Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 3431-3440.

[14] Ronneberger, O., Fischer, P., Brox, T. 2015. 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, Springer, pp. 234-241.

[15] Yinglei Liang, Yudong Liu, Junyu Dong, and Wei Cheng (2019). Chest X-Ray Segmentation with a U-Net-Based Deep Learning Framework, de Jinfeng Yang.

[16] Xue, Y., Xu, T., Zhang, H., Long, L. R., Huang, X. 2018. Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics, 16, 383-392.

[17] Chen, T., Zhai, X., Ritter, M., Lucic, M., Houlsby, N. 2019. Self-supervised gans via auxiliary rotation loss. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12154-12163).

[18] Nahar, N., Soomro, S., Monrat, A. A. 2021. A GAN based Framework for Multi-Modal Medical Image Segmentation.

[19] Long Chen, Jie Chen, Wei Li, Xiaolei Huang, and Pheng-Ann Heng (2019). ”Adversarial Learning for Multi-modal Medical Image Segmentation with SegAN”.

[20] Zeng, Z., Xulei, Y., Qiyun, Y., Meng, Y., Le, Z. 2019. Sese-net: Self-supervised deep learning for segmentation. Pattern Recognition Letters, 128, 23-29.

[21] Zhao, J., Hou, X., Pan, M., Zhang, H. 2022. Attention-based generative adversarial network in medical imaging: A narrative review. Computers in Biology and Medicine, 105948.

[22] El Mansouri, O., El Mourabit, Y., El Habouz, Y., Boujemaa, N., Ouriha, M. 2022. Intelligent System Based on GAN Model for Decision Support in Brain Tumor Segmentation. In: Fakir, M., Baslam, M., El Ayachi, R. (eds) Business Intelligence. CBI 2022. Lecture Notes in Business Information Processing, vol 449. Springer, Cham.

[23] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. 2020. Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[24] Mirza, M., Osindero, S. 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[25] Sawant, Ronit and Shaikh, Asadullah and Sabat, Sunil and Bhole, Varsha, Text to Image Generation using GAN (July 8, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems.

[26] Sandfort, V., Yan, K., Pickhardt, P.J. et al. 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 9, 16884 .

[27] S. A. Israel et al., Generative Adversarial Networks for Classification, 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2017, pp. 1-4.

[28] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, 2012. Geoffrey: ImageNet Classification with Deep Convolu-tional Neural Networks. In: Neural Information Processing Systems.

[29] Peng SL, Li BB, Hu XY. Wavelet and filter bank design: theory and application. Beijing: Tsinghua univer-sity press, 2017, pp. 95–220.

[30] Li, Y., Zhang, Y., Zhang, H., Yang, Y. 2020. GAN-Based Image Segmentation Model with Wavelet Transforms.

[31] Ran QW. Wavelet transform and fractional Fourier transform theory and applications. Harbin: Harbin Institute of Technology Press, 2001, pp. 1-3.

[32] Luo Y, Li JP, Cheng LZ, et al. Wavelet packet transform and the design cost function. Math Theory Appl 2011; 31(3): 65-70.

[33] Smith, J., Johnson, A., Brown, C. (2022). Introduction to Wavelet Transforms and Image Compression. Digital Signal Processing Journal, 15(3), 123-137.

[34] Mittal, A.; Hooda, R.; Sofat, S. Lung field segmentation in chest radiographs: A historical review, current status, and expectations from deep learning. IET Image Process 2017, 11, 937–952.

[35] Bosdelekidis, V.; Ioakeimidis, N.S. Lung Field Segmentation in Chest X-rays: A Deformation-Tolerant Procedure Based on the Approximation of Rib Cage Seed Points. Appl. Sci. 2020, 10, 6264.

[36] Jaeger, S.; Candemir, S.; Antani, S.; Wang, Y.X.J.; Lu, P.X.; Thoma, G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg. 2014, 4, 475.

[37] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, M. T. Islam, 2020, Can AI Help in Screening Viral and COVID-19 Pneumonia, IEEE Access, vol. 8, pp. 132665-132676.

[38] F. Luisier, T. Blu and M. Unser,March 2011. Image Denoising in Mixed Poisson–Gaussian Noise.

[39] XUN, Siyi, LI, Dengwang, ZHU, Hui, et al. 2022, Generative adversarial networks in medical image segmentation: A review. Computers in biology and medicine, vol. 140, p. 105063.

[40] YIN, Ming, LIU, Wei, SHUI, Jun, et al.2012, Quaternion wavelet analysis and application in image denoising. Mathematical Problems in Engineering, 2012, vol.