

Article

MWG-UNet: Hybrid Deep Learning Framework for Lung Fields and Heart Segmentation in Chest X-ray Images

Yu Lyu  and Xiaolin Tian *

School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau 999078, China; 2109853gii30002@student.must.edu.mo

* Correspondence: xltian@must.edu.mo

Abstract: Deep learning technology has achieved breakthrough research results in the fields of medical computer vision and image processing. Generative adversarial networks (GANs) have demonstrated a capacity for image generation and expression ability. This paper proposes a new method called MWG-UNet (multiple tasking Wasserstein generative adversarial network U-shape network) as a lung field and heart segmentation model, which takes advantages of the attention mechanism to enhance the segmentation accuracy of the generator so as to improve the performance. In particular, the Dice similarity, precision, and F1 score of the proposed method outperform other models, reaching 95.28%, 96.41%, and 95.90%, respectively, and the specificity surpasses the sub-optimal models by 0.28%, 0.90%, 0.24%, and 0.90%. However, the value of the IoU is inferior to the optimal model by 0.69%. The results show the proposed method has considerable ability in lung field segmentation. Our multi-organ segmentation results for the heart achieve Dice similarity and IoU values of 71.16% and 74.56%. The segmentation results on lung fields achieve Dice similarity and IoU values of 85.18% and 81.36%.

Keywords: WGAN; organ segmentation; computer auxiliary diagnosis



Citation: Lyu, Y.; Tian, X.

MWG-UNet: Hybrid Deep Learning Framework for Lung Fields and Heart Segmentation in Chest X-ray Images. *Bioengineering* **2023**, *10*, 1091. <https://doi.org/10.3390/bioengineering10091091>

Academic Editors: Alan Wang, Sibusiso Mdletshe, Brady Williamson and Guizhi Xu

Received: 14 August 2023

Revised: 13 September 2023

Accepted: 15 September 2023

Published: 18 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the improvement in medicine, medical image analysis has become an important auxiliary diagnosis and treatment method in medical imaging. The anatomical imaging of organs obtained through medical imaging technologies can objectively reflect tissue structure and pathological change. Medical image segmentation improves efficiency for doctors, from positioning and obtaining evidence to guiding treatment. Deep learning technology has made breakthrough research results in the field of computer vision and image processing. Its powerful feature learning ability has attracted widespread attention. The application of artificial intelligence to medical image processing can not only improve the processing efficiency, but also play an auxiliary role, with doctors subsequently analyzing the condition [1]. As a key branch in medical imaging research, medical image semantic segmentation tasks have always played a pivotal role in clinical diagnosis [2]. The mainstream method studies the segmentation of the lesion area in the frontal image. In this paper, we mainly study the segmentation of multi-position medical images.

Artificial intelligence is increasingly popular with the advancement of massive electronic data and improved technology. In the area of combining artificial intelligence and medical treatment, multiple auxiliary diagnosis systems based on convolutional neural networks have become an inevitable trend in developing new medical models [3], such as the incorporation of positioning, medical image segmentation, and classification. To facilitate accurate and detailed observations of lesions, machine-learning-based automatic recognition and segmentation of medical images can enhance processing efficiency and provide supplementary assistance to doctors to subsequently analyze patients' conditions [4]. Due to the COVID-19 pandemic, there has been a massive research focus on lung field

segmentation with chest X-ray or computed tomography (CT) images for auxiliary diagnosis. The CT images show clearer edges without overlapping organs compared to chest X-ray images. Chest X-ray images are low contrast and have blurred borders for organ segmentation, but chest X-ray images are in common use in hospitals because of their low price and quick image generation. There is a great need to process large amounts of information, which can be achieved through automatic segmentation. The combination of lung field and heart segmentation improves the efficiency, showing the relative positions of the organs and sizes of the lung fields and heart at the same time. The positions of the lung fields and heart also provides information on the potential risk of the heart and lung fields. On the one hand, lung fields and heart segmentation identifies the relative position and sizes of the lung fields and heart to clarify the illness. On the other hand, automatic segmentation improves the speed of doctors' ability to deal with diseases. The automatic segmentation of chest X-ray images has improved the image review quality and the speed of disease diagnosis. Our interest is in helping patients to detect potential risk with chest X-ray images and improving the speed of doctors' diagnoses. For these reasons, this paper focus on lung field and heart segmentation with chest X-ray images.

Medical image segmentation is a critical task in the field of medical imaging. It involves the identification and delineation of specific structures or regions of interest within medical images captured using techniques such as magnetic resonance imaging (MRI), CT, or ultrasound scans. The purpose of segmentation is to extract accurate boundaries or contours of anatomical structures, tumors, lesions, or other areas of clinical significance. Accurate segmentation plays a crucial role in various medical applications, including disease diagnosis, treatment planning, surgical guidance, and monitoring of disease progression. It allows clinicians to analyze and quantify specific regions, measure volumes, track changes over time, and assist in making informed decisions for patient care. However, medical image segmentation is a challenging task due to several factors. Firstly, medical images often exhibit low contrast, noise, and variations in intensity, making it difficult to distinguish between different structures or tissues. Additionally, the shapes and sizes of anatomical structures can vary significantly across patients and even within the same patient, further complicating the segmentation process. To address these challenges, various segmentation techniques have been developed, ranging from traditional methods to more advanced deep learning approaches. Traditional techniques include threshold, region-based methods, active contour models, and graph cuts. These methods rely on image characteristics, statistical information, or prior knowledge to delineate structures.

GANs are capable of generating synthetic data that closely resemble real data, making them useful in various applications such as image synthesis, video generation, and text generation. GANs can learn from unlabeled data, which eliminates the need for manually labeled training examples. This makes GANs flexible and adaptable to a wide range of domains and datasets. In novel data creation, GANs have the potential to generate entirely new and unseen data samples. This can be valuable for creative tasks, generating unique artwork or exploring uncharted areas of data distribution. GANs can be used to augment existing datasets to increase their size and diversity. This helps improve the generalization and performance of machine learning models trained on limited data. GANs can be employed for domain adaptation, where they learn to generate data from a source domain to match the statistics of a target domain. This facilitates transferring knowledge learned from one domain to another.

GAN training can be challenging and unstable at times. It involves a delicate balance between the generator and discriminator networks and finding this equilibrium can be difficult. GANs may suffer from issues like mode collapse, where the generator fails to explore the entire data space, or vanishing gradients. GANs are prone to skipping modes or failing to capture the complete data distribution. This means that certain aspects or modes of the data may not be adequately represented by the generated samples. Assessing the quality of GAN-generated samples objectively is still an open research problem. While subjective evaluation by human observers is often used, developing reliable quantitative

metrics for GAN evaluation remains a challenge. Training GANs can be computationally demanding and time consuming, requiring powerful hardware resources like GPUs (graphics processing units) and extensive training iterations to achieve desirable results. The realistic nature of GAN-generated data raises ethical concerns about potential misuse, such as deep-fake technology or generating deceptive content. Ensuring the responsible use of GANs is an ongoing area of concern and research.

In recent years, deep learning approaches, particularly convolutional neural networks (CNNs), have shown remarkable success in medical image segmentation. CNN-based architectures, such as U-Net, SegNet, and DeepLab, have demonstrated high accuracy and efficiency in segmenting diverse medical structures with minimal manual intervention. These models can learn hierarchical features directly from the images and capture complex patterns for accurate segmentation. Supervised classifier learning is also employed. Refs. [5,6] introduced the concept of generative adversarial networks, with impressive results on image-generation-quality benchmarks. This approach involves the interplay between a generator and a discriminator, enabling comprehensive training of the discriminator to achieve autonomous segmentation [7]. Furthermore, deep learning techniques incorporating multi-modal fusion are extensively employed in medical imaging for accurate medical image segmentation [8]. In the book *Intelligent Data Analysis for Biomedical Application* [9], the authors [10] utilized machine learning to successfully classify myocardial ischemia using delayed contrast enhancement techniques of magnetic resonance imaging. Additionally, addressing the domain shift problem, a new unsupervised domain adaptive framework called the collaborative image and feature adaptive algorithm (SIFA) [11] was proposed and found to be effective. In another study, Ref. [12] directly extracted features from the frequency data of vibration signals and evaluated the performance of feature learning from the original data, and spectrum and time–frequency combined data. This approach successfully applied deep learning in feature extraction for machine-based diagnosis. Furthermore, Ref. [13] introduced a graph convolution method that utilized multi-resolution pools to capture local and contextual features. Their method enabled the learning of spatially related features in irregular domains within a multi-resolution space. A graph-based convolution method employing product neural networks was proposed for position and direction classification, resulting in improved pose parameter estimation and segmentation performance. Another innovative approach, presented by [14], is the multi-receiving domain CNN (MRFNet) method. MRFNet employs an encoder–decoder module (EDM) with sub-net, providing optimal receiving fields for each sub-net and generating context information at the functional map level. MRFNet exhibits exceptional performance across all three medical image datasets.

Edge accuracy is often a key aspect and challenge in segmentation [15]. One proposal suggests utilizing a parameter model with correlation probability density to describe the integration method in the largest posterior form [16]. In 2019, Hiroki Tsud introduced a method that utilizes generative adversarial networks (GANs) with multiple functions for cell image segmentation [17]. This method demonstrates improved segmentation accuracy compared to the traditional pix2pix approach [18]. The field of medical imaging technology generates vast amounts of data. To address this, a heterogeneous framework for multi-core learning based on support vector machines (SVMs) was proposed by [19]. They investigated the flexibility of this method in comparison to using SVMs and other classifiers to process single features which can enhance the learning ability of the Math Kernel Library (MKL) algorithm. Additionally, Nilanjan Dey proposed several medical applications based on meta-heuristics for segmentation [20]. CT images are commonly employed not only for analyzing X-ray films but also for examining the internal structure of the heart, necessitating segmentation of the heart region itself [21,22]. Due to the complexity and significance of large blood vessels attached around the atria and ventricles, a team led by Lohendran Baskaran devised a method for multi-organ segmentation from coronary computed tomography angiography images [23].

In contrast, Olivier Ecabert presented a model that encompasses four heart chambers and interconnected large blood vessels [24]. Avi Ben-cohen introduced a method utilizing a fully convolutional network (FCN) with global context and local plaque level analysis based on super-pixel sparse classification for the detection of liver metastases [25]. Addressing the complementary nature of salient edge and object information, Ref. [26] proposed an edge navigation network that employs a three-step approach to simultaneously incorporate and model these two complementary pieces of information within a single network. The experimental results demonstrate improved performance, particularly in scenarios involving rough object boundaries. Additionally, Ref. [27] proposed a multi-layer densely connected super-resolution 3D network with training guided by generative adversarial networks (GANs). This approach enables fast training and reasoning that outperforms other popular deep learning methods by achieving four times higher image resolution restoration while running six times faster. Furthermore, a novel system has been developed that utilizes CT scans to generate positron emission tomography (PET) virtual images. This system offers the potential for cardiac sarcoidosis evaluation [28]. Ref. [29] proposed multiple tasks deep learning model for detection of peripherally inserted central catheter (PICC) which aids in accurate identification and placement of catheters for medical procedures. Zhongrong Wang proposed pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI [30]. LF-SegNet [31] is a fully convolutional encoder-decoder network designed for lung fields segmentation from chest radiographs images which assists in automated analysis and diagnosis of respiratory conditions for enabling more efficient medical interventions.

While significant progress has been made in organ segmentation using deep learning methods, there are still several gaps and challenges that exist. The availability of large-scale annotated datasets for organ segmentation is limited. Creating accurate and comprehensive annotations requires significant time and expertise, leading to a scarcity of labeled data. This constraint hinders the development and evaluation of robust models. The organs can exhibit significant variations in shape, size, and appearance across individuals and even within the same individual due to factors such as pathology or imaging artifacts. Existing algorithms often struggle to handle this variability, leading to sub-optimal segmentation results. Certain organs may have indistinct boundaries or overlap with neighboring structures, making their precise delineation challenging. Algorithms need to effectively handle these ambiguous cases and accurately differentiate organ boundaries from surrounding tissues. Deep learning methods primarily rely on data-driven learning without explicitly incorporating prior anatomical knowledge. Integrating prior knowledge, such as anatomical atlases or spatial constraints, into the segmentation process could improve accuracy and consistency.

The major contributions of this paper are outlined as follows.

- A designed U-Net with an SE block called AR-UNet, which takes advantages of the attention mechanism to enhance the segmentation accuracy of the generator so as to improve the performance.
- By applying the AR-UNet as the generator of the MWG-UNet structure, the proposed model can both promote accurate lung field segmentation and enhance the stability in model training.
- The proposed MWG-UNet is comprehensively evaluated on the JRST and Shenzhen Hospital datasets and achieves the optimal performance for most evaluation metrics except IoU.

The remainder of this paper is organized as follows. Section 2 introduces the proposed method in detail. Section 3 presents the experimental results. Section 4 presents a discussion of our method and others. In Section 5, conclusions are drawn.

2. Materials and Methods

The methodology includes three parts: the flaws of GANs, Wasserstein GAN, and the improvement of WGAN. The flaws of GANs introduces the shortcomings of GANs

from the perspective of mathematical principles. Wasserstein GAN introduces the working principles of WGAN and the improvement compared with GAN. The improvement of WGAN introduces the improvements that have been made based on WGAN.

2.1. GANs

Generative adversarial networks (GANs) are generative models and their training is in the form of a confrontational game. Equation (1) is the objective loss function of the confrontation.

$$V(D, G) = \mathbb{E}_{x \sim P_r} [\log D(x)] + \mathbb{E}_{z \sim P_g} [\log(1 - D(G(z)))] \quad (1)$$

where P_r is the true sample distribution, and P_g is the sample distribution of the generator.

$$D(x) = \frac{P_r(x)}{P_r(x) + P_g(x)} \quad (2)$$

Equation (2) represents the simplified optimal discriminator. It is derived by setting the derivative of Equation (1) to 0 and expresses the loss function that captures the relative proportion of the true distribution and the probability of generating the distribution.

One challenge with GAN training is the need to avoid over-training the discriminator. If the discriminator becomes too proficient, it hampers the reduction in the generator's loss function during the experimental phase. Another issue encountered in GANs is the limited diversity of the generated samples. In scenarios where the true sample distribution and the generator's sample distribution exist as low-dimensional manifolds within a high-dimensional space, the probability of their overlap being negligible approaches 1. Consequently, regardless of how dissimilar they are, the JS divergence remains constant, resulting in the generator's gradient approaching (approximately) 0 and effectively disappearing.

In summary, GANs encounter challenges related to the discriminator's proficiency and the lack of diversity in generated outputs. Mitigating these challenges is crucial to achieving better performance and diversity in GAN-based architecture. The Wasserstein distance is defined as follows:

$$KL(P_g \| P_r) = 2JS(P_r \| P_g) \quad (3)$$

There are two significant issues with this equivalent minimization objective. Firstly, it aims to minimize the KL divergence between the true distribution and the generator's distribution, while simultaneously maximizing the JS divergence between them. This approach is flawed, since the Kullback–Leibler (KL) divergence is not a symmetrical measure, leading to intuitive absurdity and numerical gradient instability.

Furthermore, the generator tends to prioritize generating repetitive and safe samples over diverse samples, exacerbating the challenges within GAN frameworks. These fundamental problems can be attributed to the unreasonable nature of the equivalent optimization distance metric and the generator's loss function, which result in unstable gradients and imbalanced penalties between diversity and accuracy.

To address the latter concern, a proposed solution involves introducing noise to both the generated and real samples. Intuitively, this noise “diffuses” the original low-dimensional manifolds into the entire high-dimensional space, compelling them to have a noticeable overlap. Once an overlap exists, enabling the presence of a gradient, the JS divergence can effectively operate. Notably, as the two distributions become closer, the diffusion of overlap reduces the JS divergence. However, despite these improvements the quest for a quantitative indicator to measure training progress remains unresolved.

2.2. Wasserstein GAN

The Wasserstein distance, introduced in WGAN, addresses the problem of gradient disappearance in theory. Unlike KL divergence and JS divergence, which exhibit abrupt

changes and can only be the largest or smallest values, the Wasserstein distance maintains smoothness. When researchers aim to optimize parameters using the gradient descent method, KL and JS divergences fail to provide gradients altogether, while the Wasserstein distance remains capable of providing a gradient.

Similarly, in a high-dimensional space, if two distributions lack overlap or have a negligible overlap, KL and JS divergences cannot accurately represent the distance or offer gradients. In contrast, the Wasserstein distance can provide meaningful gradients in such scenarios. The key advantage of the Wasserstein distance over KL divergence and JS divergence is its ability to reflect the distance between two distributions even when there is no overlap.

By utilizing the Wasserstein distance instead of JS divergence, stable training and progress indicators can be simultaneously achieved. The Wasserstein distance is defined as follows:

$$W(P_r, P_g) = \inf_{r \in \Pi(P_r, P_g)} E_{(x,y) \sim r} [\|x - y\|] \quad (4)$$

Formula (4) represents the marginal distribution of each distribution for the true sample distribution and the sample distribution of the generator. For each possible joint distribution γ , the expected value of the distance of the sample under the joint distribution γ can be calculated. The Wasserstein distance cannot be drawn directly and the loss function becomes the following formula:

$$W(P_r, P_g) = \frac{1}{K} \sup_{\|f\|_L} \quad (5)$$

When the Lipschitz continuity condition limits the maximum local variation in a continuous function, a parameter can be constructed. When the last layer of the discriminator network is not a non-linear activation layer, the loss function will approximate the distance between the true distribution and the generated distribution. Due to the excellent nature of the Wasserstein distance, there will be no problem with the disappearance of the generator gradient. The loss functions of the generator and discriminator are as follows:

$$\min L_D(P_{r+\epsilon}, P_{g+\epsilon}) = -\mathbb{E}_{x \sim P_{r+\epsilon}} [\log D^x(x)] - \mathbb{E}_{x \sim P_{g+\epsilon}} [\log(1 - D^x(x))] \quad (6)$$

According to Equation (6), the smaller the value the better the training. After the improvement of WGAN, there are several improvements compared to the original GAN algorithm:

- The sigmoid function is not applicable in the discriminator. GAN's discriminator performs a two-classification task of true and false. The discriminator in WGAN approximates the Wasserstein distance. So there is no need for the sigmoid function.
- The loss of the generator and discriminator does not take the log function.
- It limits all parameters of the neural network to no more than a certain range.
- It completely solves the problem of GAN training instability; there is no longer a need to carefully balance the training level of the generator and the discriminator.
- During training there is a value like cross-entropy to indicate the progress of the training.

2.3. The Overall Framework of MWG-UNet

The basic framework of our proposed method, called MWG-UNet, is shown in Figure 1. The architecture of MWG-UNet contains a discriminator for distinguishing real and fake data and a generator for making the distribution of generated samples close to the real data. In the generator of MWG-UNet, we use an improved U-Net called ARU-Net, with a squeeze and excitation (SE) block. The generator is responsible for creating synthetic data samples and attempting to generate realistic outputs that can deceive the discriminator. The discriminator is tasked with distinguishing between the real and fake data samples produced by the generator. The results of the generator and the real data are input into the discriminator to distinguish the real data from the fake data to finally realize accurate

segmentation. The parameter from the discriminator updates the generator to improve the performance of the generator.

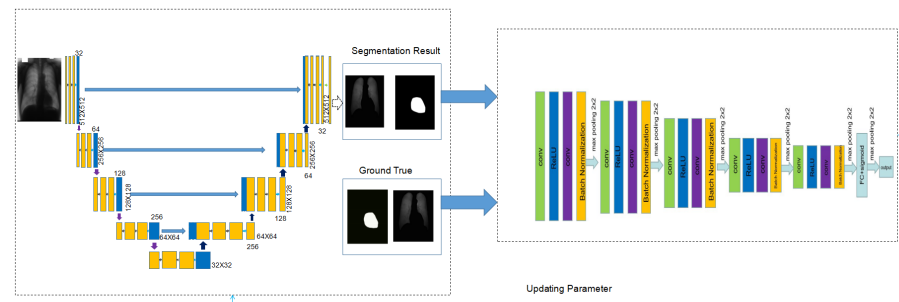


Figure 1. Structure of MWG-UNet.

Figure 2 shows the details of the designed AR-UNet as the generator of MWG-UNet. AR-UNet is U-Net transformed with an SE block and residual connection. The structure of AR-UNet comprises an encoder and decoder. The encoder captures the high-level features of an input image and reduces its spatial dimensions, while the decoder recovers the spatial information and generates a segmentation mask by upsampling the encoded features to match the original input resolution. For every step of downsampling, we take 3×3 convolution layers to extract the features and the SE block for boosting the discriminated power of each channel, improving both the accuracy and efficiency in various computer vision tasks. After that, we use a 2×2 max pooling operation to extract high-dimensional features. On the contrary, for upsampling we use a 2×2 upsampling operation to enlarge the feature image. Then, we use the same method as every step of downsampling before max pooling for image processing. There are four steps for the upsampling operation and downsampling operation. In order to map the feature, we use a 1×1 convolution layer at the last layers. The skip connection enables the direct flow of low-level spatial information from the encoder to the decoder for facilitating precise localization and fine-grained details in the output and helps mitigate the problem of information loss during max pooling, improving the overall segmentation accuracy of U-Net. For the sake of improving the efficacy in medical image segmentation to obtain spatial features, we take non-linear addition at the final step with a 1×1 convolution layer for the output.

The discriminator of MWG-UNet is shown in Figure 3. In every step of the discriminator, after every two convolutions, the downsampling operator is implemented. After four steps of the max pooling operation, the final output is given by the fully connected layer. Through a series of convolutional or fully connected layers, the discriminator learns to discern subtle patterns and features. The discriminator guides the generator's learning process by providing feedback on how well it can deceive the discriminator. In this study, a simple discriminator is used for accurate segmentation and saving unnecessary computation.

Unstable training is a common issue with GANs. Although WGAN has made significant strides in achieving stability, there are instances where it generates poor samples or struggles to converge. The introduction of Wasserstein GAN shifts the measurement of probability distributions in GANs from f-divergence to Wasserstein distance, resulting in improved stability during training and generally higher-quality generated outputs. However, WGAN relies on a weight clipping strategy to enforce the Lipschitz constraint on the critic, which can lead to undesired behavior during the training process. To address this limitation, this paper proposes a different truncation pruning strategy known as gradient penalty. This strategy penalizes the gradient's norm with respect to the critic's input. By incorporating gradient penalty, the training of WGAN becomes more stable, and the quality of the the generated images improves.

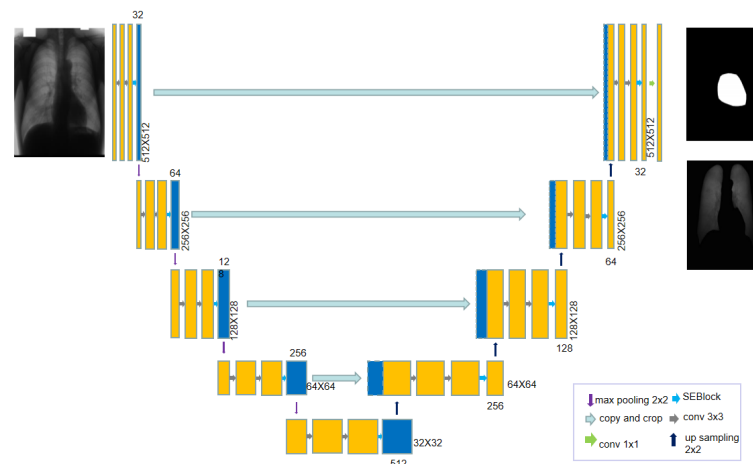


Figure 2. The architecture of AR-UNet.

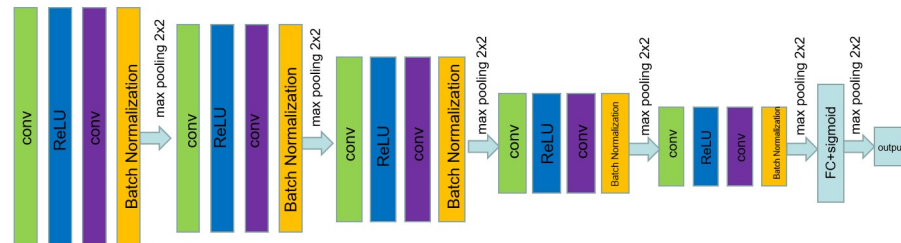


Figure 3. The discriminator of MWG-UNet.

In the realm of medical image segmentation, U-Net and transformed U-Net are commonly used methods. Our approach aims to enhance the performance and robustness by combining the concepts of U-Net and adversarial networks. The paper introduces the addition of Gaussian noise to the generated images and utilizes batch normalization in the discriminator to achieve higher-quality generated results with improved stability. We use the structure of Wasserstein GAN for the adversarial network. In the generator, we use the UNet structure for feature extraction. UNet, as the generator, has an encoding path for context and extracting features and a decoding path for upsampling to recover the spatial information lost during the encoding phase. The combination of WGAN and UNet improves the performance on lung segmentation with regards to the Jaccard similarity and Dice similarity metrics. Details on the value of the combination are shown in Section 4.

This paper proposes a new architecture called the multi-tasking Wasserstein generative adversarial network U-shape network (MWG-UNet). Multi-tasking refers to the ability of a system or an individual to perform multiple tasks simultaneously or in rapid succession. The goal of multi-task learning is to improve the generalization and performance on each individual task by jointly optimizing the model's parameters across multiple tasks. In the context of computer systems and artificial intelligence, multi-tasking refers to the capability of a program or a model to handle and execute multiple tasks concurrently. For chest X-ray medical images, multi-tasking improves the efficiency of diagnosis with the relative positions and the shapes of the heart and lung fields. The combination of WGAN and UNet improves the performance of lung segmentation with regards to Jaccard similarity and Dice similarity.

3. Results

The Results and Discussion address two areas: image pre-processing and segmentation results. They introduce the results of using different generative adversarial networks.

The training of the methods is conducted on a Linux system version 18.04, GPU environment, cuda10.2, cudnn7.6.5, and Python environment 3.8.

3.1. Data Pre-Processing

The original image for transformation and organs segmentation in Figure 4. The Japanese Society of Radiological Technology (JSRT) [32] is a public dataset of chest X-ray images with accurate annotation for lung fields and heart masks. JSRT encompasses 154 nodule and 93 non-nodule 12-bit gray-scale images with high resolution 2048×2048 . As we know, machine learning needs massive amounts of data to optimize the model and avoid overfitting. The medical images used in the experiment are provided by Shenzhen No.3 Hospital in Shenzhen [33]: 340 normal X-ray images and 275 abnormal X-ray images. The total number of original images is 862. We use geometric transformations and rotation for data augmentation to increase the data to avoid gradient explosion. In this paper, the adaptive histogram equalization (CLAHE) operation is used for image enhancement. Data augmentation is a common method used in image processing for medical image segmentation and classification because public datasets with accurate annotation are commonly small, which can cause overfitting. Common data augmentation techniques can be divided into basic image manipulation and deep learning approaches. The heart part and the lung part of all pictures are marked by the doctor. All algorithms use 70% as the training sample and 30% as the test sample and all images are converted to 512×512 pixels PNG format.

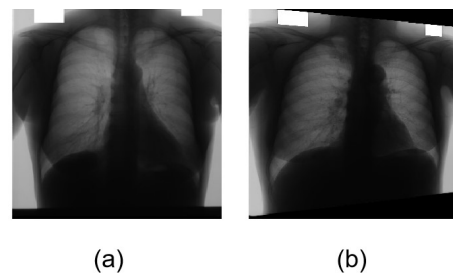


Figure 4. Chest X-ray images. (a) The original chest X-ray medical image. (b) The figure is transformed with rotation.

3.2. Evaluation Metrics

As a binary question for organ segmentation, five measures are displayed, including Dice similarity, IoU, recall, precision, and F1 score. Equations (7)–(11) show their mathematical formulae. These metrics are based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) classification results. In image segmentation, Dice similarity, also known as the Dice coefficient, has commonly been used as a metric to quantify the similarity or overlap between two sets or binary masks. The IoU, also called the Jaccard similarity, reflects the similarity of the segmentation with deep learning and the ground truth, which is the most intuitive indicator to compare the performance with other methods. The Dice similarity, also known as the Dice coefficient, is a statistical measure used to quantify the similarity or overlap between two sets or binary masks. It calculates the ratio of twice the intersection of the sets to the sum of their sizes, providing a value between 0 and 1, where 1 indicates perfect overlap and 0 indicates no overlap. It is commonly used in image segmentation tasks to evaluate the accuracy of predicted segmentation masks by comparing them with ground truth masks. Higher Dice similarity scores indicate better agreement between the predicted and ground truth masks. Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances correctly identified by the model. Precision represents the proportion of predicted positive instances that are actually true positives. The F1 score is the harmonic

mean of recall and precision, providing a balanced measure that considers both metrics. The equations of these five metrics are presented below.

$$\begin{aligned} Dice &= 2 * TP / (FP + FN + 2 * TP) \\ &= 2 * (true * pred_score) / [true * true + pred_score * pred_score] \end{aligned} \quad (7)$$

$$IoU = (TP) / (TP + FP + FN) \quad (8)$$

$$Recall = TP / (TP + FN) \quad (9)$$

$$Precision = TP / (TP + FP) \quad (10)$$

$$F1score = 2 * TP / [(TP + FN) + (TP + FP)] = 2 * (ture * pre) / (true + pred) \quad (11)$$

The subdivision indices of the results generated after the image segmentation are shown in Table 1. This table compares the results of five metrics obtained by different networks of improvised lightweight deep CNNs [34], UNet++ and HardNet [35], UNet and EfficientNet [36], AlexNet and ResNet [37], GAN, and WGAN and MWG-UNet. The results of other models are directly cited from the corresponding literature. The values without citation are trained by our machine. According to the results, it is found that the proposed MWG-UNet achieves the best results in terms of Dice, precision, and F1 score compared with other algorithms. In particular, the Dice similarity, precision, and F1 score of the proposed method outperform other models, reaching 95.28%, 96.41%, and 95.90%, respectively, and the specificity surpasses the sub-optimal models by 0.28%, 0.90%, 0.24%, and 0.90%. However, the value of IoU is inferior to the optimal model by 0.69%. In the image segmentation, the value of Dice demonstrates the average performance and the value of IoU tends to measure the worst performance. Our results are not as good as the optimal model for some difficult cases, which indicates that the model's generalization ability needs to be improved.

However, single-organ segmentation demonstrates less information for diagnosis. This paper also expands on a new area for lung field and heart segmentation to improve the efficiency of diagnosis. The relative positions of the heart and lung fields can demonstrate some heart or lung disease. The segmentation of heart and lung fields intuitively displays the sizes and positions, allowing direct diagnosis of related diseases. Our lung field and heart segmentation results achieve Dice similarity and IoU values of 71.16% and 74.56%. The segmentation result of lung fields achieve Dice similarity and IoU values of 85.18% and 81.36%. Figure 5 shows the original medical images and a comparison of the ground truth and transformed images for the heart and lung fields. Column (a) shows the original chest X-ray images. Column (b) demonstrates the ground truth for the chest X-ray images with the segmentation of the heart and lung fields. A comparison with column (c) and column (d) demonstrates an increase in smoothness and clarity. However, the accuracy for each organ in the lung field and heart segmentation has decreased much more than in single-organ segmentation. Figure 5 shows examples of partial segmentation results. Figure 5a shows different non-processed and pre-processed medical images for organ segmentation for the heart and lungs. Figure 5b shows the ground truth of heart and lung segmentation with multiple colors for different parts of organs. The blue part is the left lung. The green part is the right lung of the lung field. The red part shows the heart segmentation. Figure 5c shows the segmentation with WGAN. Figure 5c is the segmentation result using our method, MWG-UNet. Comparing columns (c) and (d), the result of the segmentation in column (d) is slightly better than column (c).

Table 1. Segmentation results of different methods for lung fields.

Method	Dice	IoU	Recall	Precision	F1 Score
Improved lightweight deep CNN	90.64%	86.53%	/	/	/
UNet++ and HardNet	95%	93%	96.5%	94%	95%
UNet and EfficientNet	95%	90%	/	94%	/
AlexNet and ResNet	93.56%	88.07%	/	/	/
GAN	89.16%	86.75%	90.9%	80.69%	85.49%
WGAN	92.23%	89.96%	91.95%	88.93%	90.41%
MWG-UNet	95.28%	92.31%	97.40%	94.24%	95.79 %

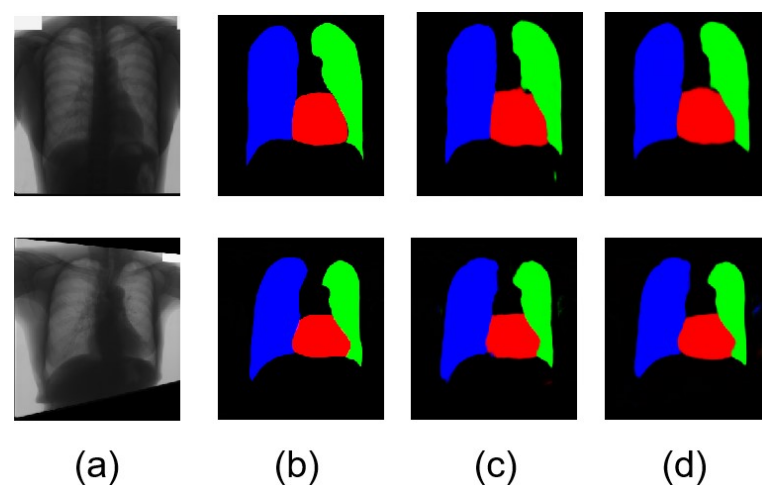


Figure 5. (a) The chest X-ray images and segmentation results. (b) Ground truth for chest X-ray images. (c) Segmentation result using WGAN. (d) Segmentation result using our improved method, MWG-UNet.

4. Discussion

In the era of big data, artificial intelligence is touching all fields, including medical image processing. Deep learning methods for automatic organ segmentation improve the efficiency of diagnosis. However, organ segmentation presents challenges due to irregular shapes, occlusions, image artifacts, and the limited availability of annotated data. Developing robust and efficient algorithms capable of handling these challenges is crucial for enhancing clinical decision making and improving patient care outcomes. Several deep learning methods have been developed for organ segmentation in medical imaging. U-Net is a popular architecture known for its encoder–decoder structure and skip connections, enabling precise localization. Mask R-CNN combines object detection with instance segmentation, providing accurate organ delineation. FCN utilizes fully convolutional layers to generate pixel-wise predictions. V-Net extends U-Net with 3D convolutions, suitable for volumetric data segmentation. DenseNet employs dense connectivity patterns to enhance feature reuse and gradient flow. Attention-based models like DeepLab utilize spatial attention mechanisms to focus on relevant regions. These methods leverage the power of deep neural networks in capturing intricate patterns and have shown promising results in organ segmentation tasks.

In this paper, we proposed a new method, called MWG-UNet, for medical image segmentation. The proposed method uses the designed AR-UNet as the generator and several convolutional and fully connected layers as the discriminator. AR-UNet combines the U-Net architecture for semantic segmentation with an SE block that captures channel-

wise dependencies. The SE block re-calibrates feature maps, enhancing the discriminative power. This integration improves both localization accuracy and feature representation in U-Net, making it more effective for organ segmentation tasks. MWG-UNet with U-Net combines the WGAN framework for stable training of GANs with the AR-UNet architecture for semantic segmentation. This integration enables the generation of realistic and high-quality segmentation masks by leveraging the benefits of both an improved loss function and the ability to capture spatial information and preserve fine details.

The training data used were from JSRT and Shenzhen Hospital, with 862 images. The medical images from Shenzhen No.3 Hospital in Shenzhen provide 340 normal X-ray images and 275 abnormal X-ray images. JSRT encompasses 154 nodule and 93 non-nodule 12-bit gray-scale images with high resolution 2048×2048 . To handle the limited dataset, we use geometric transformations and rotation as data augmentation to increase the data to avoid gradient explosion. The evaluation metrics are Dice similarity, IoU, recall, precision, and F1 score. As mentioned above, our results for the Dice similarity, precision, and F1 score of the proposed method outperform other models, reaching 95.28%, 96.41%, and 95.90% and the specificity surpasses the sub-optimal models by 0.28%, 0.90%, 0.24%, and 0.90%. The result of the IoU is slightly lower than the optimal method. The IoU demonstrates a worse performance in training. In further research, improving the ability of generalization is an important step to upgrade our algorithm. Our lung field and heart segmentation results have Dice similarity and IoU values of 71.16% and 74.56% for the heart segmentation. The segmentation results of the lung fields have Dice similarity and IoU values of 85.18% and 81.36%.

Based on the above discussion, it is shown that the proposed MWG-UNet is a lung field and heart segmentation model which takes advantage of the attention mechanism to enhance the segmentation accuracy of the generator so as to improve the performance. Although the proposed method presents sub-optimal results on the IoU value, there is still space for further improvement. Firstly, our novel method uses WGAN and AR-UNet for lung segmentation in chest X-ray images and outperforms other methods. Secondly, we proposed an improved U-Net with an SE block which is responsible for creating synthetic data samples and attempting to generate realistic outputs. Lastly, we use multiple task learning for multiple-organ segmentation to improve the efficiency of diagnosis with the relative positions and sizes of the organs.

5. Conclusions

Medical image segmentation is a critical task in the field of medical imaging. In this paper, we focus on overcoming the difficulties of traditional generative adversarial networks, such as a vanishing gradient and the diversity of generators. Wasserstein GAN was introduced to address the instability of GAN training and the diversity of generated samples. We conduct an in-depth mathematical analysis to highlight the disparities between the mathematical foundation and the practical training of Wasserstein GAN. The article proposes a method, MWG-UNet, for lung field segmentation with chest X-ray images. This method was applied to a public dataset of lung and heart segmented X-ray images. The results for Dice similarity, precision, and F1 score of the proposed method outperform other models, reaching 95.28%, 96.41%, and 95.90%, and the specificity surpasses the sub-optimal models by 0.28%, 0.90%, 0.24%, and 0.90%. The result of the IoU is slightly lower than the optimal method. Meanwhile, the paper explores the frontiers for lung field and heart segmentation with X-ray images. Although the accuracy is lower than for single-organ segmentation, the relative positions of the heart and lung fields provide various information for disease diagnosis. The lung and heart segmentation results achieve Dice similarity and IoU values of 71.16% and 74.56% for the heart segmentation. The segmentation result of the lung fields achieve Dice similarity and IoU values of 85.18% and 81.36%. In future work, we will focus on increasing the accuracy of the lung field and heart segmentation results and on speeding up the training and testing.

Author Contributions: Conceptualization, Y.L. and X.T.; methodology, Y.L.; software, Y.L.; validation, Y.L. and X.T.; formal analysis, Y.L.; investigation, Y.L.; resources, Y.L.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L.; visualization, Y.L.; supervision, X.T.; project administration, X.T.; funding acquisition, X.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Macau Science and Technology Development Fund, grant number 0009/2018/A. The APC was funded by 0009/2018/A.

Institutional Review Board Statement: This retrospective study was approved by the ethics committee of the lead hospital (JCHO Tokyo Shinjuku Medical Center: IRB No. R3-14). This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and the current scientific guidelines.

Informed Consent Statement: This was a retrospective observational study dealing with anonymized processed information. Consent was obtained on an opt-out basis following Japanese ethical regulations, the Japanese Personal Information Protection Law, and the instructions of each ethics committee. In the opt-out method, information regarding the purpose and methodology of the study was notified or disclosed, and opportunities for refusal were guaranteed to the extent possible.

Data Availability Statement: The source code and learned models are available from the authors upon reasonable request.

Acknowledgments: We would like to express our deepest gratitude to all individuals and organizations who have contributed to the successful completion of this research paper. We are grateful to the Science and Technology Development Fund for their valuable guidance and support throughout the project. Our sincere thanks go to the participants and volunteers whose cooperation made this study possible. We are also thankful to our colleagues and peers for their insightful feedback and discussions. Finally, we acknowledge the funding provided by the study on the segmentation method and key technology of Heart X-ray image and CT image of No.0009/2018/A that supported this research endeavor. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GAN	Generative adversarial network
MWG-UNet	Multiple tasking Wasserstein generative adversarial network U-shape network
Multi-organ	Multiple organ
MRI	Magnetic resonance imaging
CT	Computed tomography
PICC	peripherally inserted central catheter
SIFA	Synergistic image and feature alignment
GPUs	Graphics processing units
CNN	Convolutional neural network
MRFNet	Multi-receiving domain convolutional neural network
EDM	Encoder–decoder module
SVM	Support vector machine
MKL	Mathematics kernel library
FCN	Fully convolutional network
PET	Positron emission tomography
SE block	Squeeze and excitation block
KL	Kullback–Leibler
TP	True positive
TN	True negative
FP	False positive
FN	False negative

References

1. Ibrahim, M.R.; Haworth, J.; Cheng, T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* **2020**, *96*, 102481. [[CrossRef](#)]
2. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. *Radiographics* **2017**, *37*, 505–515. [[CrossRef](#)]
3. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Neural Information Processing Systems, Harrahs and Harveys, Lake Tahoe, CA, USA, 3–8 December 2012.
4. Huang, Y.; Li, X.; Yan, C.; Liu, L.; Dai, H. MIRD-Net for Medical Image Segmentation. In *Advances in Knowledge Discovery and Data Mining*; Lauw, H., Wong, R.W., Ntoulas, A., Lim, E.P., Ng, S.K., Pan, S., Eds.; Springer: Cham, Switzerland, 2020; pp. 207–219.
5. Kazemini, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for medical image analysis. *Artif. Intell. Med.* **2020**, *109*, 101938. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, Z.; Yang, L.; Zheng, Y. Multimodal medical volumes translation and segmentation with generative adversarial network. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, K., Rueckert, D., Fichtinger, G., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 183–204.
7. He, X.; Yang, S.; Li, G.; Li, H.; Chang, H.; Yu, Y. Non-local context encoder: Robust biomedical image segmentation against adversarial attacks. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 8417–8424.
8. Merjulah R.; Chandra, J. Classification of Myocardial Ischemia in Delayed Contrast Enhancement Using Machine Learning. In *Intelligent Data Analysis for Biomedical Applications*; Hemanth D.J., Gupta D., Balas V.E., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 209–235.
9. Zhou, T.; Ruan, S.; Canu, S.A. review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3–4*, 100004. [[CrossRef](#)]
10. Giger, M. L. Machine learning in medical imaging. *J. Am. Coll. Radiol.* **2018**, *15*, 512–520. [[CrossRef](#)] [[PubMed](#)]
11. Xia, Y.; Yang, D.; Yu, Z.; Liu, F.; Cai, J.; Yu, L.; Zhu, Z.; Xu, D.; Yuille, A.; Roth, H. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Med. Image Anal.* **2020**, *65*, 101766. [[CrossRef](#)] [[PubMed](#)]
12. Jing, L.; Zhao, M.; Li, P.; Xu, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **2017**, *111*, 1–10. [[CrossRef](#)]
13. Edwards, M.; Xie, X.; Palmer, R.I.; Tam, G.K.; Alcock, R.; Roobottom, C. Graph convolutional neural network for multi-scale feature learning. *Comput. Vis. Image Underst.* **2020**, *194*, 102881. [[CrossRef](#)]
14. Liu, L.; Wu, F.X.; Wang, Y.P.; Wang, J. Multi-Receptive-Field CNN for Semantic Segmentation of Medical Images. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3215–3225. [[CrossRef](#)] [[PubMed](#)]
15. Tuan, T.M.; Ngan, T.T.; Toan, D.N.; Giap, C.N.; Son, L.H. Semisupervised fuzzy clustering methods for X-ray image segmentation. In *Handbook of Data Science Approaches for Biomedical Engineering*; Balas, V.E., Solanki, V.K., Kumar, R., Khari, M., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 251–289.
16. Metaxas, D.N.; Yan, Z. Deformable models, sparsity and learning-based segmentation for cardiac MRI based analytics. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, K., Rueckert, D., Fichtinger, G., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 273–292.
17. Tsuda, H.; Hotta, K. Cell image segmentation by integrating pix2pixs for each class. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019.
18. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
19. Wani, N.; Raza, K. Multiple Kernel-Learning Approach for Medical Image Analysis. In *Soft Computing Based Medical Image Analysis*; Dey, N., Ashour, A.S., Shi, F., Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 31–47.
20. Dey, N.; Ashour, A.S. Computing in Medical Image Analysis. In *Soft Computing Based Medical Image Analysis*; Dey, N., Ashour, A.S., Shi, F., Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 3–11.
21. Larrey-Ruiz, J.; Morales-Sánchez, J.; Bastida-Jumilla, M.C.; Menchón-Lara, R.M.; Verdú-Monedero, R.; Sancho-Gómez, J.L. Automatic image-based segmentation of the heart from CT scans. *J. Image Video Proc.* **2014**, *2014*, 1–13 [[CrossRef](#)]
22. Ecabert, O.; Peters, J.; Schramm, H.; Lorenz, C.; von Berg, J.; Walker, M.J.; Vembar, M.; Olszewski, M.E.; Subramanyan, K.; Lavi, G.; et al. Automatic model-based segmentation of the heart in CT images. *IEEE Trans. Med. Imaging* **2008**, *27*, 1189–1201. [[CrossRef](#)] [[PubMed](#)]
23. Baskaran, L.; Al'Aref, S.J.; Maliakal, G.; Lee B.C.; Xu Z.; Choi J.W.; Lee S.E.; Sung J.M.; Lin F.Y.; Dunham S.; et al. Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. *PLoS ONE* **2020**, *15*, e0232573. [[CrossRef](#)] [[PubMed](#)]
24. Ecabert, O.; Peters, J.; Walker, M.J.; Ivanc, T.; Lorenz, C.; von Berg, J.; Lessick, J.; Vembar, M.; Weese, J. Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. *Med. Image Anal.* **2011**, *15*, 863–876. [[CrossRef](#)] [[PubMed](#)]
25. Ben-Cohen, A.; Greenspan, H. Liver lesion detection in CT using deep learning techniques. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, K., Rueckert, D., Fichtinger, G., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 65–90.

26. Zhao, J.; Liu, J.; Fan, D.; Cao, Y.; Yang, J.; Cheng, M. EGNNet: Edge Guidance Network for Salient Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8778–8787.
27. Chen, Y.; Shi, F.; Christodoulou, A.G.; Xie, Y.; Zhou, Z.; Li, D. Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*; Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer: Cham, Switzerland, 2018; pp. 91–99.
28. Bois, J.P.; Muser, D.; Chareonthaitawee, P. PET/CT Evaluation of Cardiac Sarcoidosis. *PET Clin.* **2019**, *14*, 223–232. [[CrossRef](#)] [[PubMed](#)]
29. Yu, D.; Zhang, K.; Huang, L.; Zhao, B.; Zhang, X.; Guo, X.; Li, M.; Gu, Z.; Fu, G.; Hu, M.; et al. Detection of peripherally inserted central catheter (PICC) in chest X-ray images: A multi-task deep learning model. *Comput. Methods Programs Biomed.* **2020**, *197*, 105674. [[CrossRef](#)] [[PubMed](#)]
30. Wang, Z.; Xie, L.; Qi, J. Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI. *Magn. Reson Imaging* **2020**, *66*, 131–140. [[CrossRef](#)] [[PubMed](#)]
31. Mittal, A.; Hooda, R.; Sofat, S. LF-SegNet: A fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs. *Wirel. Pers. Commun.* **2018**, *101*, 511–529. [[CrossRef](#)]
32. Digital Image Database. Available online: <http://db.jsrt.or.jp/eng.php> (accessed on 12 January 2020).
33. Tuberculosis Chest X-ray Image Data Sets. Available online: <https://lhncbc.nlm.nih.gov/publication/pub9931> (accessed on 12 January 2020).
34. Arvind, S.; Tembhurne, J.V.; Diwan, T.; Sahare, P. Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays. *Results Eng.* **2023**, *17*, 100929. [[CrossRef](#)]
35. Kim, Y.; Kim, K.; Wu, D.; Ren, H.; Tak, W.Y.; Park, S.Y.; Lee, Y.R.; Kang, M.K.; Park, J.G.; Kim, B.S.; et al. Deep Learning-Based Four-Region Lung Segmentation in Chest Radiography for COVID-19 Diagnosis. *Comput. Diagn.* **2022**, *12*, 101. [[CrossRef](#)] [[PubMed](#)]
36. Karthik, R.; Srinivasan, M.; Chandhru, K. Chapter 10—A deep ensemble network for lung segmentation with stochastic weighted averaging, In *Diagnostic Biomedical Signal and Image Processing Applications with Deep Learning Methods*; Kemal, P., Saban, Ö., Eds.; Academic Press: Cambridge, MA, USA, 2023; pp. 197–214.
37. Souza, J.C.; Diniz, J.O.B.; Ferreira, J.L.; da Silva, G.L.F.; Silva, A.C.; de Paiva, A.C. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Comput. Methods Programs Biomed.* **2019**, *177*, 285–296. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.