applied sciences

MDPI

Review

Enhancing 3D Lung Infection Segmentation with 2D U-Shaped Deep Learning Variants

Anindya Apriliyanti Pravitasari *, Mohammad Hamid Asnawi, Farid Azhar Lutfi Nugraha, Gumgum Darmawan and Triyani Hendrawati

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia; mohammad19011@mail.unpad.ac.id (M.H.A.); farid19002@mail.unpad.ac.id (F.A.L.N.); gumgum@unpad.ac.id (G.D.); triyani.hendrawati@unpad.ac.id (T.H.)
* Correspondence: anindya.apriliyanti@unpad.ac.id

Abstract: Accurate lung segmentation plays a vital role in generating 3D projections of lung infections, which contribute to the diagnosis and treatment planning of various lung diseases, including cases like COVID-19. This study capitalizes on the capabilities of deep learning techniques to reconstruct 3D lung projections from CT-scans. In this pursuit, we employ well-established 2D architectural frameworks like UNet, LinkNet, Attention UNet, UNet 3+, and TransUNet. The dataset used comprises 20 3D CT-scans from COVID-19 patients, resulting in over 2900 raw 2D slices. Following preprocessing, the dataset is refined to encompass 2560 2D slices tailored for modeling. Preprocessing procedures involve mask refinement, image resizing, contrast limited adaptive histogram equalization (CLAHE), and image augmentation to enhance the data quality and diversity. Evaluation metrics, including Intersection over Union (IoU) and dice scores, are used to assess the models' performance. Among the models tested, Attention UNet stands out, demonstrating the highest performance. Its key trait of harnessing attention mechanisms enhances its ability to focus on crucial features. This translates to exceptional results, with an IoU score of 85.36% and dice score of 91.49%. These findings provide valuable insights into guiding the selection of an appropriate architecture tailored to specific requirements, considering factors such as segmentation accuracy and computational resources, in the context of 3D lung projection reconstruction.

Keywords: 3D projection; CT-Scan; image segmentation; UNet; LinkNet; Attention UNet; UNet 3+; TransUNet; image segmentation; lung disease; deep learning; convolutional neural network

check for updates

Citation: Pravitasari, A.A.; Asnawi, M.H.; Nugraha, F.A.L.; Darmawan, G.; Hendrawati, T. Enhancing 3D Lung Infection Segmentation with 2D U-Shaped Deep Learning Variants. *Appl. Sci.* **2023**, *13*, 11640. https://doi.org/10.3390/app132111640

Academic Editor: Jan Egger

Received: 20 September 2023
Revised: 20 October 2023
Accepted: 21 October 2023
Published: 24 October 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

In the field of medical imaging, the ability to accurately reconstruct three-dimensional (3D) representations from two-dimensional (2D) images is a critical challenge with significant implications for diagnosis and treatment. Traditional medical imaging techniques, such as computed tomography (CT), generate images that provide valuable insights into anatomical structures and disease manifestations. However, relying solely on 2D images limits our ability to fully comprehend the complex spatial relationships and structural characteristics of the imaged regions. This limitation is particularly evident in the analysis of lung diseases, where the precise localization and characterization of infection areas are crucial for assessing disease severity and guiding appropriate treatment strategies. Consequently, the development of robust and efficient methods for 3D projection reconstruction from 2D images has become an active area of research, with the potential to significantly enhance our understanding of lung pathologies and improve patient care. In this study, we aim to address this challenge by leveraging deep learning techniques and advanced image processing methods to construct 3D lung projections from CT-scans of COVID-19 patients. By doing so, we hope to contribute to the growing body of knowledge in medical imaging and provide healthcare professionals with a powerful tool for precise disease characterization and personalized treatment planning.

Coronavirus Disease 2019 (COVID-19), which is caused by the SARS-CoV-2 virus, leads to a respiratory infection, particularly affecting the lungs [1]. The areas defined as infection in the lungs are generally Ground-Glass Opacities (GGOs) or areas that have increased attenuation on CT-scans of the lungs [2]. The area of lung infection in a COVID-19 patient is an important aspect of examining the patient's condition. By knowing the area of infection, medical personnel can determine the severity of the patient's condition, the condition of the patient's lungs, and most importantly, use it as a factor to determine the progress of the COVID-19 patient's condition. To determine the area of infection in the lungs, one of the biomedical imaging technologies, namely CT-scan, can be used.

The CT-scan is a product of radiographic imaging, where it is produced using the Computed Tomography technique [3]. In the case of COVID-19, a CT-scan is also used as a tool to diagnose whether the patient has COVID-19 or not. According to research by Ai et al. [4] and Simpson et al. [5], the CT-scan has far better accuracy than RT-PCR, which is the global standard technique for diagnosing COVID-19 [6,7]. The diagnosis of COVID-19 using a CT-scan has a high sensitivity of 97%, while RT-PCR only has a sensitivity in the range of 42% to 71% [4,5]. Despite the high accuracy of CT-scans in diagnosing COVID-19, it is not used as a global standard due to its high cost and the need for a radiologist to interpret the radiological features of COVID-19 on CT-scan images. Additionally, CT-scans are typically reserved for hospitalized COVID-19 patients for further examination and treatment purposes, including assessing the progress of the patient's condition. The further examination process using CT-scans is carried out directly by a radiologist to determine the condition of the lungs, the progress of healing, and the volume of the area of infection, as well as the severity of COVID-19.

During the further examination process using CT-scan imaging, radiologists manually search for radiological characteristics, including areas of infection [3]. Analyzing the area of infection or the entire lung requires a high level of focus because CT-scan images are grayscale, and the radiological characteristics, particularly infections, are often indistinct and difficult to detect. To assist medical personnel, particularly radiologists, in analyzing these subtle and unclear images, we utilize machine learning techniques, specifically image segmentation. This approach enables us to identify lung and infection areas on CT-scans, reconstructing 3D projections that provide a more focused view of the regions of interest. This reconstruction allows medical professionals to gain insights into the precise locations and extents of lung structures and infections. By visualizing the 3D projections, radiologists can compare the size and spatial relationships between the lung and infection areas, aiding in more accurate diagnosis and treatment planning. Looking forward, this method also holds the potential for calculating the volume of specific areas, such as the volume of the infection, further enhancing the quantitative analysis capabilities for medical practitioners.

Image segmentation is one of the most critical tasks in medical image analysis, involving the recognition and separation of various components in an image [8]. With the dynamic strides in machine learning, particularly in the revolutionary domain of deep learning, substantial headway has been achieved in this field. Among these advancements, the U-shaped deep learning architectures, originally stemming from UNet [9], have risen to prominence as prevalent frameworks for image segmentation, unequivocally showcasing their prowess across diverse medical imaging scenarios. In our research, we will harness the power of multiple U-shaped deep learning architectures. This inclusive lineup encompasses its precursor, UNet, along with its subsequent developments: LinkNet [10], Attention UNet [11], UNet 3+ [12], and TransUNet [13]. We will utilize these architectures, specifically targeting the identification and separation of the lung and infection areas, using data obtained from COVID-19 lung patients. Our aim is to evaluate and compare the performance and efficacy of these five 2D U-shaped deep learning architectures in generating accurate 3D lung reconstructions within this specific context.

UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet stand as prominent deep learning models extensively employed within the domain of medical image analysis, particularly in the intricate realm of semantic segmentation tasks. These versatile models

have demonstrated remarkable efficacy across a diverse spectrum of medical imaging applications, including image segmentation in CT-scans, MRI scans, and microscopy images [9,11–18]. In this paper, we focus on reconstructing the 3D projection of lung CT-scans using 2D deep learning architectures, considering the advantages they offer. While 3D models are possible for this task, employing 2D architectures provides computational efficiency during training and inference. By leveraging 2D models, we can achieve comparable results to 3D models, albeit with the need for additional post-processing steps to convert the 2D output to a 3D projection, unlike 3D models, which are more straightforward and do not require any post-processing steps [14]. This approach strikes a balance between computational efficiency and accurate visualization, which is crucial for CT-scan analysis. Moreover, using 2D architectures allows us to capture both local and global context while preserving high-resolution features. This methodology enables faster analysis and interpretation of lung CT-scans, providing valuable insights for medical personnel.

In summary, we contribute to the domain of medical imaging by leveraging a comprehensive array of advanced deep learning techniques for the paramount task of 3D projection reconstruction from 2D lung CT-scans. Our approach not only addresses the critical challenge of accurately representing the complex spatial relationships within these scans but also presents a unique amalgamation of cutting-edge methodologies. By seamlessly integrating the power of U-shaped deep learning architectures, encompassing UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet, we transcend the limitations of 2D images and propel the field forward. Furthermore, we bridge the computational efficiency of 2D models with the precision of 3D projection through innovative post-processing strategies, offering a compelling balance between analytical speed and comprehensive visualization. Our work not only enriches the arsenal of medical imaging tools but also facilitates rapid and accurate lung diseases' diagnosis, making significant strides toward personalized treatment and improved patient care.

The primary contributions of the research are as follows:

1. Our study will contribute to the advancement of medical image analysis by exploring the potential of 2D architectures in reconstructing 3D lung projections. By evaluating the performance of UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet in this specific context, we aim to enhance the understanding of how 2D architectures can effectively capture lung structures and infection areas, enabling accurate diagnosis and treatment planning for lung diseases, specifically COVID-19.
2. To contribute to the academic literature, our study carries practical implications for healthcare professionals and researchers working on lung disease diagnosis and treatment. By identifying the most suitable 2D architecture for reconstructing 3D lung projections, we can provide valuable guidance on selecting appropriate techniques and tools for accurate assessment of disease severity and monitoring disease progression. This can lead to the development of more efficient and reliable clinical protocols for lung diseases, particularly COVID-19 cases.
3. In addition, this study provides a comprehensive comparison of UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet for visualizing the 3D lung construction of CT-scans, specifically focusing on the effectiveness of 2D architectures in reconstructing 3D projections. This aspect has not been extensively studied in the literature, and our findings will shed light on the advantages and limitations of using 2D architectures for this task.

The remainder of this paper is structured as follows: Section 2 delves into a comprehensive review of various research papers pertaining to the utilization of UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet in their respective studies. Section 3 provides an explanation of the dataset, data preprocessing, model setting, metrics evaluation, proposed architecture, and postprocessing steps utilized in this study. In Section 4, the results of every architecture are explained and analyzed.

2. Related Works

Image segmentation has emerged as a significant research domain in recent times, with the application of various deep learning techniques for this task. Among the numerous deep learning techniques for image segmentation, the U-shaped architectures have gained a lot of attention due to their efficacy in performing segmentation on medical images. This section presents works related to our research that utilized the UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet architectures for image segmentation.

The authors in [19] presented four different semantic segmentation networks based on the UNet architecture with different backbones, namely ResNet50-UNet, DenseNet169-UNet, SE-ResNext50-UNet, and EfficientNetB4-UNet, which were employed for the segmentation of chest X-ray images to diagnose Pneumothorax. The evaluation of these models achieved an IoU score of around 78% and a mean dice similarity coefficient (DSC) score of 90%. Sabir et al. [20], modified the UNet architecture, substituting the convolutions layer with residual blocks and developed an architecture called ResUNet. A deep residual UNet was used to segment liver tumor CT-scans and obtained a DSC score of 89.3%, an accuracy of 97%, a precision of 95%, and a specificity 95.7%. Asnawi et al. [14] proposed various different encoder architectures for 3D UNet for segmenting the COVID-19 CT Lung and Infection Dataset. The evaluated architectures include 3D UNet, 3D ResUNet, 3D VGGUNet, and 3D DenseUNet. The 3D UNet architecture demonstrated high accuracy, producing an IoU score of 81.58% and a dice score of 88.61% for lung and infection segmentation.

The LinkNet algorithm was used in [21] to develop a Multi-Task network on LinkNet-based architecture (MTLN) with multi-scale inputs. The architecture achieves a DSC score of 96.84% by performing semantic segmentation and estimating fetal head circumference (HC) in 2D ultrasound images. Akyel et al. [22] presented a new approach to the LinkNet-based algorithm using EfficientNetB7 as the encoder, called LinkNet-B7, for noise removal and segmentation of skin cancer images. The proposed model obtained a dice coefficient of 95.72% and 96.70%, respectively. The work presented in [23] proposes a different approach for the LinkNet-based algorithm, where LinkNet is used with a pre-trained encoder as its backbone, and additional dilated convolution layers are added in the center part of the network.

Attention UNet [11], a model that introduces attention gates, has made a considerable impact and served as an inspiration for numerous researchers in the realm of image segmentation. Initially proposed to address segmentation challenges in medical contexts, Attention UNet has significantly advanced the field by harnessing attention mechanisms. In its seminal work, the application of Attention UNet was demonstrated on abdominal CT-scans, specifically the multi-class abdominal CT-150, as well as the TCIA Pancreas CT-82 dataset [24]. Notably, Attention UNet consistently outperformed UNet, excelling particularly in the demanding task of pancreas segmentation. In the same study, Attention UNet exhibited its prowess against benchmark models on the CT-82 dataset, including hierarchical 3D FCN [25], dense-dilated FCN [26], holistically nested 2D FCN stage-1 and stage-2 [27], 2D FCN [28], 2D FCN + recurrent Networks [28], and single and multi-model 2D FCN [29]. It achieved comparable performance despite using lower resolution and omitting post-processing steps. The influence of this model extends beyond abdominal CT-scans. Nguyen et al. [30] employed Attention UNet with Active Contour Based Hybrid Loss for brain tumor segmentation, where it exhibited superiority over SegNet [31] and UNet, achieving dice and IoU scores of 89% and 81%, respectively. The groundbreaking performance of Attention UNet has catalyzed the development of several derivative models, including Attention-augmented UNet (AA-UNet) [32], Multi-Scale Dilated Attention UNet (MDA-Unet) [33], and much more, as researchers continue to draw inspiration from its proven efficacy.

UNet 3+ [12], a model derived from the UNet++ architecture, stands as another integral U-shaped framework that holds significance in this study and has garnered widespread attention. UNet 3+ is designed to harness the full potential of multi-scale features extracted across various levels of both encoder and decoder pathways. Its debut saw applications

on two distinct datasets: the liver segmentation dataset sourced from the ISBI LiTS 2017 challenge and the spleen dataset. Remarkably, in comparison with five state-of-the-art methods, including PSPNet [34], Deeplab variations [35–37], and Attention UNet [11], UNet 3+ showcased its supremacy on both datasets. Employing a hybrid loss with Conditional Gated Module (CGM), UNet 3+ achieved the highest performance, attaining remarkable evaluation scores of 96.75% and 96.20% for the liver and spleen datasets, respectively. The prowess of UNet 3+ extends beyond its initial applications. Widely embraced, this model has found utility in various studies, not only as a primary method but also as a robust benchmark in various studies.

TransUNet [13], a pivotal advancement within the landscape of image segmentation, marks the fusion of two potent domains, the Transformer architecture and UNet. Notably, TransUNet introduces self-attention mechanisms to meticulously capture the global context from input data, setting the stage for its successful application in medical image segmentation. This remarkable model embarks on the segmentation journey across two diverse medical datasets: the Synapse dataset, encompassing various structures, and the ACDC dataset, focused on cardiac MRI scans. In the scrutiny of the Synapse dataset, TransUNet takes on established contenders including VNet [38], DARR [39], UNet (with ResNet50 as the encoder), Attention UNet (with ResNet50 as the encoder) [40], and various ViT variations. Impressively, TransUNet emerges triumphant, securing the highest dice score of 77.48%. Similarly, in the context of the ACDC dataset, TransUNet goes head-to-head with UNet (with ResNet50 as the encoder), Attention UNet (with ResNet50 as the encoder), and assorted ViT [41] variations, once again clinching the top spot with a remarkable dice score of 89.71%. The profound influence of TransUNet reverberates throughout the medical imaging community, culminating in the emergence of derivative architectures such as Multilevel TransUNet (MTUNet) [42], MT-TransUNet [43], TransUNet++ [44], and a host of other innovative creations, solidifying TransUNet as a cornerstone and blueprint for future developments.

Previous studies have demonstrated the widespread utilization of UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet architectures for achieving high accuracy results in semantic segmentation tasks. This study aims to address the analysis of these five U-shaped architectures, specifically evaluating their performance and effectiveness, while also exploring the advantages and limitations of employing 2D architectures for generating 3D projections.

3. Material and Methods

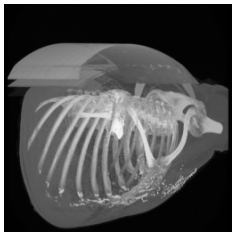
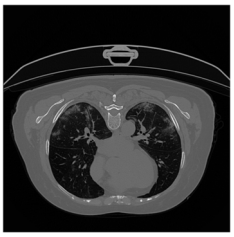
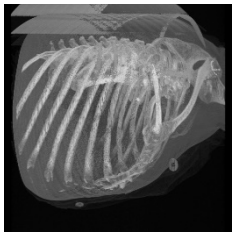
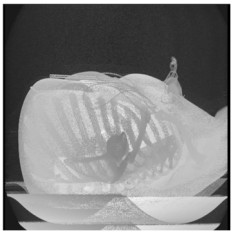
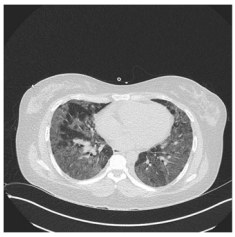
Our suggested strategy and the resources we will use for this study are presented in this section. We begin by outlining the dataset that will be applied to this study. The proposed method or architecture that will be used in this study will then be described, as well as the metrics evaluation that will be used in the evaluation of our models. Finally, the model experimental setting for each architecture will be addressed.

3.1. Dataset

The lung CT-scan dataset from Ma et al. [45] served as the foundation for training and testing the CT-scan segmentation models in this study. This dataset comprises 20 CT-scans acquired from radiopaedia [46] and the Corona Cases Initiative (RAIOSS) [47], featuring COVID-19 patients. Beyond supplying the CT-scan data, the reference [45] also offers three segmentation masks, denoted as the “lung mask”, “infection mask”, and “lung and infection mask”. Within our investigation, we specifically leveraged the “lung and infection mask” to perform segmentation of both the lung area and the infected regions in the CT-scans. Notably, this dataset boasts comprehensive annotations manually crafted by two radiologists and validated by a more experienced radiologist, in accordance with the work by Ma et al. [48]. Furthermore, the utilization of this dataset is particularly advantageous due to its ample size, inclusion of various levels of disease severity, and the

availability of ground truth masks. A comprehensive overview of the CT-scan dataset's attributes is provided in Table 1 [14].

Table 1. Three samples from the used dataset [14].

3D Projection of CT-Scan	CT-Scan Slice	Lung and Infection Mask
		
		
		

Every CT-scan in the [45] dataset has a unique width, height, depth (slice), and infection severity level. The more thorough CT-scan profile for each patient is displayed in Table 2.

3.2. Data Preprocessing

In the data preprocessing stage, several essential steps were employed to ensure the suitability of the data for subsequent analysis. This subsection outlines the preprocessing techniques applied, including mask preprocessing, image resizing, the CLAHE method, and image augmentation. These preprocessing steps are crucial for enhancing the quality of the data, optimizing the performance of the segmentation models, and improving the accuracy of the results.

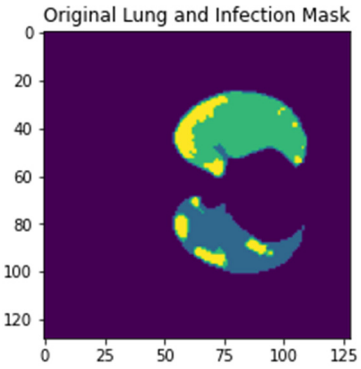
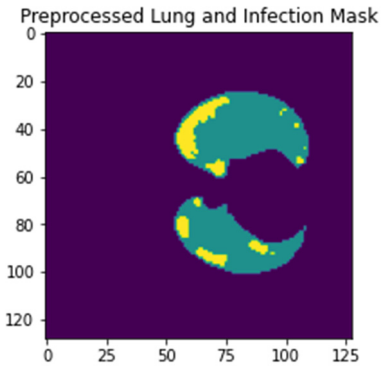
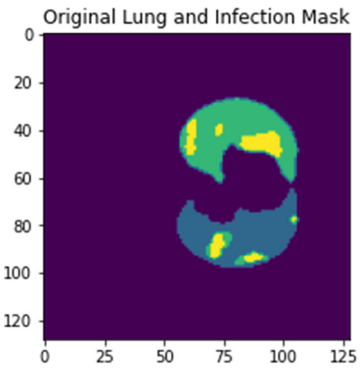
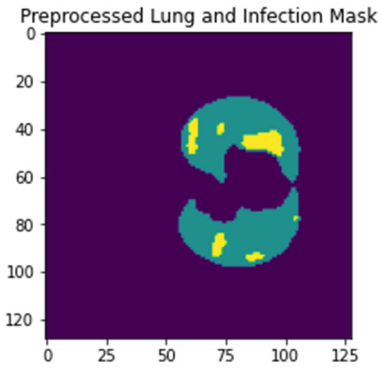
3.2.1. Mask Preprocessing

Mask preprocessing plays a crucial role in consolidating the left and right lung labels into a single mask for enhanced segmentation accuracy. The merging process involves equating the pixel values of the left lung mask and right lung mask, resulting in a unified lung mask that includes both lung regions. By incorporating this preprocessing step, the resulting lung and infection masks exhibit three distinct pixel values: 0 represents the background, 1 represents the lung, and 2 represents the infection. This merging process aims to assist the model in avoiding errors in distinguishing between the left and right lungs. The identification of the left lung and the right lung can be seen by projecting the prediction results of the model in 3D. The following in Table 3 is an example of a lung and infection mask before and after mask preprocessing.

Table 2. Patient ID, source, and size information for each patient's CT-scan.

Patient ID	Source	Size (w × h × d)
radiopaedia_4_85506_1	Radiopaedia	630 × 630 × 39
radiopaedia_7_85703_0	Radiopaedia	630 × 630 × 45
radiopaedia_10_85902_1	Radiopaedia	630 × 630 × 39
radiopaedia_10_85902_3	Radiopaedia	630 × 630 × 418
radiopaedia_14_85914_0	Radiopaedia	630 × 401 × 110
radiopaedia_27_86410_0	Radiopaedia	630 × 630 × 66
radiopaedia_29_86490_1	Radiopaedia	630 × 630 × 42
radiopaedia_29_86491_1	Radiopaedia	630 × 630 × 42
radiopaedia_36_86526_0	Radiopaedia	630 × 630 × 45
radiopaedia_40_86625_0	Radiopaedia	630 × 630 × 93
coronacases_001	RAIOSS	512 × 512 × 301
coronacases_002	RAIOSS	512 × 512 × 200
coronacases_003	RAIOSS	512 × 512 × 200
coronacases_004	RAIOSS	512 × 512 × 270
coronacases_005	RAIOSS	512 × 512 × 290
coronacases_006	RAIOSS	512 × 512 × 213
coronacases_007	RAIOSS	512 × 512 × 249
coronacases_008	RAIOSS	512 × 512 × 301
coronacases_009	RAIOSS	512 × 512 × 256
coronacases_010	RAIOSS	512 × 512 × 301

Table 3. Comparison table of lung and infection masks before and after mask preprocessing.

Before Mask Preprocessing	After Mask Preprocessing
	
	

3.2.2. Image Resizing and Pixel Scaling

To ensure uniformity in the CT-scan size and facilitate effective model training, a two-step process involving image resizing and pixel scaling is implemented. Initially, all data samples are resized to a consistent size of 128 × 128 pixels, ensuring standardization

across the dataset. Following the resizing step, the pixel values of the CT-scans undergo scaling using the MinMax scaler. This transformation maps the original pixel values to a normalized range of 0 to 1. By applying this scaling process, the data values are homogenized, promoting consistency and facilitating a more efficient model training process. The MinMax scaler method can be represented by Equation (1), as shown below:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x_{scaled} represents the scaled pixel value, x denotes the original pixel value, x_{min} represents the minimum pixel value in the CT-scan, and x_{max} denotes the maximum pixel value in the CT-scan.

3.2.3. Contrast Limited Adaptive Histogram Equalization (CLAHE)

To address contrast-related issues, such as noise and grayscale inhomogeneity commonly observed in CT-scans, the Contrast Limited Adaptive Histogram Equalization (CLAHE) [49] method is implemented following the data-scaling step. CLAHE is specifically utilized to alleviate these problems and enhance the visualization of CT-scans. By applying the CLAHE method, image noise is effectively reduced, while simultaneously generating pixels with improved grayscale contrast [50]. Consequently, CT-scans become more visually discernible, aiding in the accurate analysis and interpretation of the images. A comparative illustration of CT-scan samples before and after the application of the CLAHE method is presented in Table 4 [14], highlighting the significant enhancement achieved through this preprocessing technique.

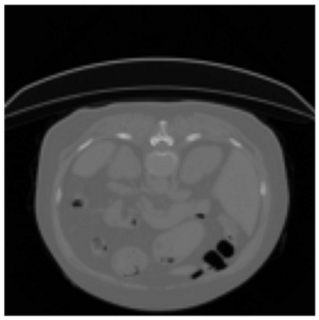
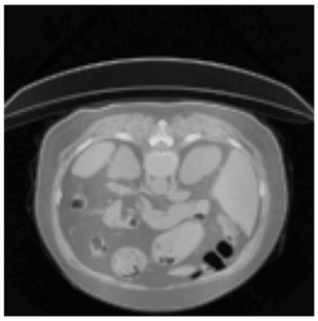
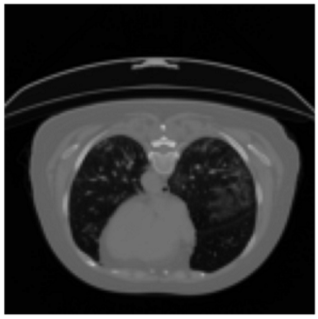
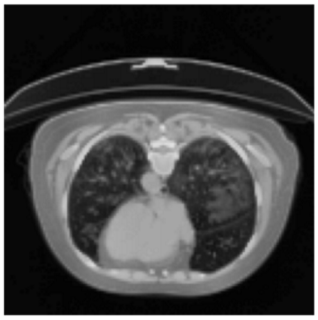
3.2.4. Image Augmentation

In the data preprocessing step, the final step involves applying image augmentation techniques to the dataset. Image augmentation is a powerful approach that enhances the diversity and robustness of the training data, thereby improving the model's ability to generalize to unseen examples. For this purpose, the Tensorflow's ImageDataGenerator from the Keras API is employed. During the image segmentation step, various augmentation parameters are set to introduce controlled transformations to the images. These transformations include rotation, shifting, shearing, zooming, horizontal flipping, and vertical flipping. Specifically, the rotation range parameter is set to 90 degrees, the width and height shift ranges are both set to 0.3, the shear range is set to 0.5, the zoom range is set to 0.3, and the fill mode is set to 'reflect'. These values are carefully chosen to introduce realistic variations in the data, mimicking potential variations in the real-world scenarios. By augmenting the data with such transformations, the model becomes more resilient to variations in the input images, improving its ability to accurately segment the lung and infection areas.

3.3. Segmentation with 2D Architectures

In this subsection, we will discuss five deep learning architectures for image segmentation: UNet, LinkNet, Attention UNet, UNET 3+, and TransUNet. These architectures all share a common foundation, characterized by a U-shaped architecture. This lineage of U-shaped architectures, originally inspired by UNet, has been meticulously adapted to address specific nuances within the field of medical imaging. In this paper, we will apply these U-shaped architecture variants for lung infection segmentation. Departing from the foundational UNet architecture, each of the developed algorithms incorporates distinctive elements. In the subsequent sections, we will explore the key features of each architecture.

Table 4. CT-scan comparison before and after CLAHE application [14].

Before CLAHE Applied	After CLAHE Applied
	
	
	

3.3.1. UNet

UNet [9] is a widely recognized and extensively used deep learning architecture based on Convolutional Neural Networks (CNNs) for medical image segmentation. Specifically, in the context of lung segmentation, the UNet architecture, which combines CNN encoder and decoder techniques, has been employed to achieve rapid and precise segmentation of lung structures [14,51,52].

The key strength of the UNet architecture lies in its ability to extract features at multiple levels of resolution and reconstruct the output segmentation map with the required dimensions, thanks to its symmetrical encoder–decoder structure. The architecture of UNet comprises two main routes: the contraction path (also known as the encoder or analytic path) and the expansion path (also known as the decoder or synthesis path) [53].

The contraction path resembles a standard convolutional network and extracts classification information from the input image. On the other hand, the expansion path consists of up-convolutions and concatenations with features from the contraction path. This enables the propagation of features from the early layers, preserving fine input details, to the deeper layers that aggregate high-level information. This is particularly important as deep layers tend to lose small image details due to the presence of intermediate pooling layers [54,55]. The skip connections between the encoder and decoder routes in the UNet architecture facilitate the integration of low-level and high-level features, promoting the accurate delin-

eation of lung structures [9]. Please refer to Figure 1, which is adapted from [9], for a visual representation of the UNet architecture.

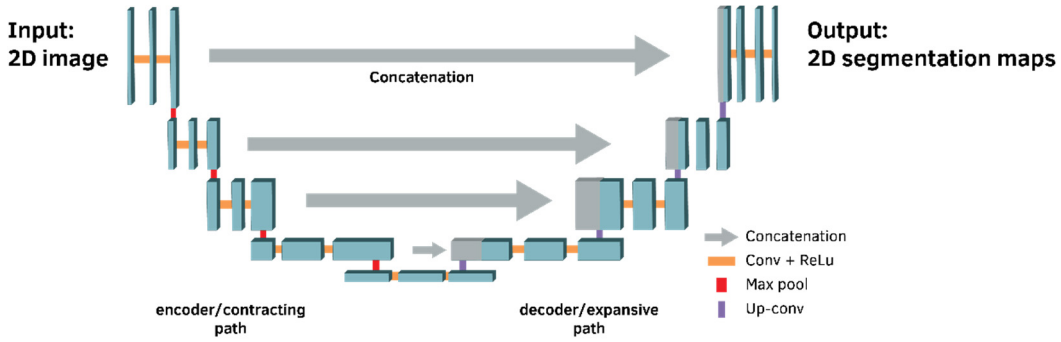


Figure 1. UNet architecture.

In our implementation, the UNet model comprises a total of 8.5 million parameters. It consists of a total of 63 layers, including layers with parameters and layers without parameters. Specifically, there are 35 layers with parameters only. For a more detailed scheme of UNet, please refer to [9].

3.3.2. LinkNet

LinkNet [10] is an efficient neural network architecture designed for semantic segmentation tasks. It incorporates skip connections, residual blocks, and an encoder–decoder structure. Similar to UNet, LinkNet adopts a U-shaped architecture, but with some key differences. In LinkNet, each level of the encoder and decoder is replaced with a residual module (res-block) instead of a conventional convolution structure. Additionally, the fusion of deep and shallow features in LinkNet follows an "adding" approach rather than "stacking", as used in UNet [23].

One of the key features of LinkNet is its rich shortcut connections, which facilitate the transmission of shallow information to deeper layers in the network. By employing residual skip connections, LinkNet effectively combines features from the encoder and decoder using element-wise summation. This approach avoids increasing the number of input channels and network parameters, unlike concatenation used in UNet skip connections.

The design principles of LinkNet focus on enhancing network efficiency while maintaining competitive segmentation performance. By leveraging residual blocks and efficient skip connections, LinkNet achieves effective feature fusion and preserves spatial information during the segmentation process. Figure 2, adapted from [10], shows the general architecture of LinkNet.

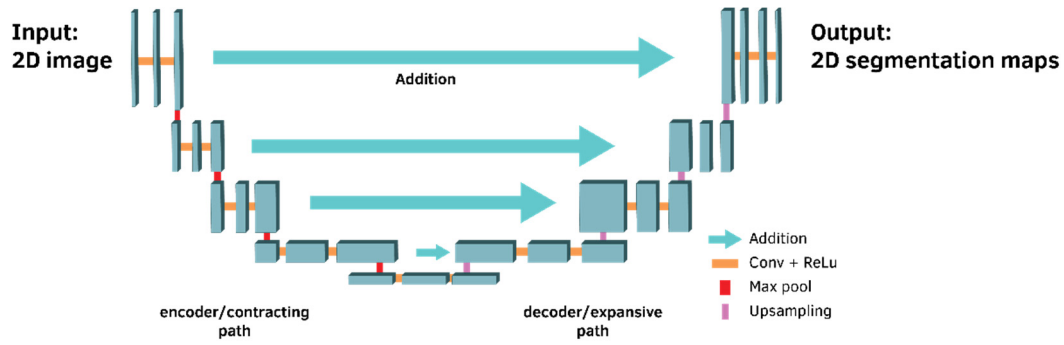


Figure 2. LinkNet architecture.

In our implementation of the LinkNet architecture, the model comprises a total of 11.5 million parameters. There are 120 layers in all, including both parameterized and

parameterless layers. There are precisely 71 layers with parameters. For a more thorough explanation of LinkNet's architecture, see [10].

3.3.3. Attention UNet

Attention UNet [11] introduces an innovative approach to improve the effectiveness of the UNet architecture. With a core objective of enhancing the model's ability to capture relevant information, Attention UNet aims to refine the segmentation process by leveraging an attention gate mechanism. This mechanism facilitates the network's capacity to focus on the most pertinent features within the input or output for a given segmentation task. Particularly, Attention UNet employs this attention mechanism between the skip connections, an essential component of the UNet architecture.

The attention mechanism in Attention UNet holds a fundamental role in reshaping how the model processes and utilizes information. This mechanism operates by filtering the features extracted from the encoder before these features are concatenated with those from the decoder. As the decoder provides the contextual information, the attention mechanism dynamically assigns varying weights to different regions of the feature map. This weight assignment is contextually informed by the decoder, enabling the network to accentuate significant features while attenuating noise or irrelevant information.

A prominent contribution of Attention UNet lies in its capacity to eliminate the need for an external organ localization module. Unlike many segmentation frameworks that often incorporate such modules to simplify tasks, Attention UNet achieves this without additional supervision or computational complexity. By harnessing attention gates, Attention UNet autonomously learns to localize and segment the target structures. Illustrated in Figure 3, adapted from [11], is the overall structure of Attention UNet.

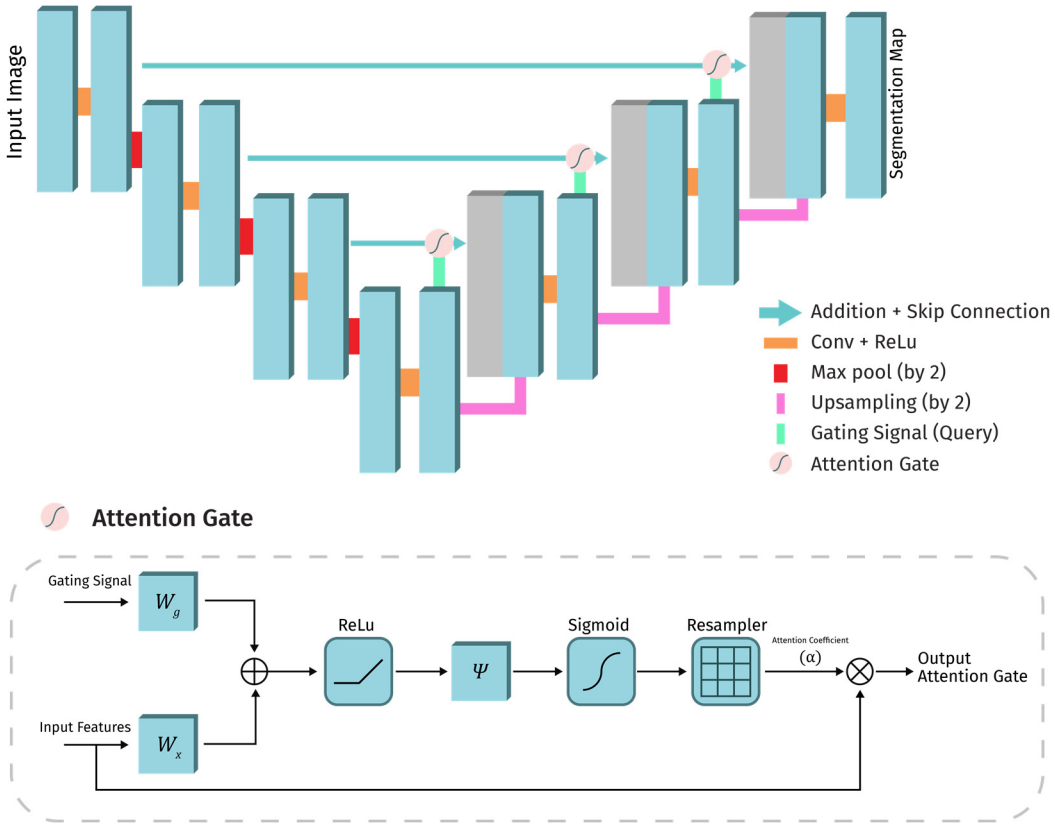


Figure 3. Attention UNet architecture [11].

In our implementation of Attention UNet, the model consists of a total of 75 layers, with 38 of these layers containing parameters. The model has a total of 7.9 million parameters. For a more detailed architectural overview of Attention UNet, please refer to [11].

3.3.4. UNet 3+

UNet 3+ [12] introduces a refined approach that augments the capabilities of both the original UNet and its extension, UNet++. The motivation behind the development of UNet 3+ emerges from the inherent drawbacks observed in the earlier UNet++ architecture. Particularly, UNet 3+ endeavors to rectify the issue of insufficient information exploration across full scales. This limitation in UNet++ prevents the model from fully harnessing the rich details and semantics available at various scales, a challenge that UNet 3+ sets out to overcome.

UNet 3+'s pioneering feature is the incorporation of full-scale skip connections. These skip connections serve as a key differentiator from conventional and nested skip connections utilized in prior models. The full-scale skip connections in UNet 3+ amalgamate low-level details with high-level semantics taken from feature maps at different scales, in contrast to earlier methods that merely combined features from different levels. This unique fusion empowers the model to capture intricate fine-grained details alongside broader coarse-grained semantics, culminating in segmentation outcomes that are more precise and comprehensive.

Complementing the innovation of full-scale skip connections, UNet 3+ leverages a full-scale deep supervision strategy. This strategy underpins a hybrid loss function designed to capture information at pixel, patch, and map-levels. The fusion of this loss function with the full-scale skip connections results in the model's capacity to learn hierarchical representations from the aggregated feature maps across the full scale. This synergy heightens the discernment of region of interest boundaries and positions, contributing to the model's enhanced segmentation accuracy. Figure 4, which has been adapted from [12], shows UNet 3+'s general architecture.

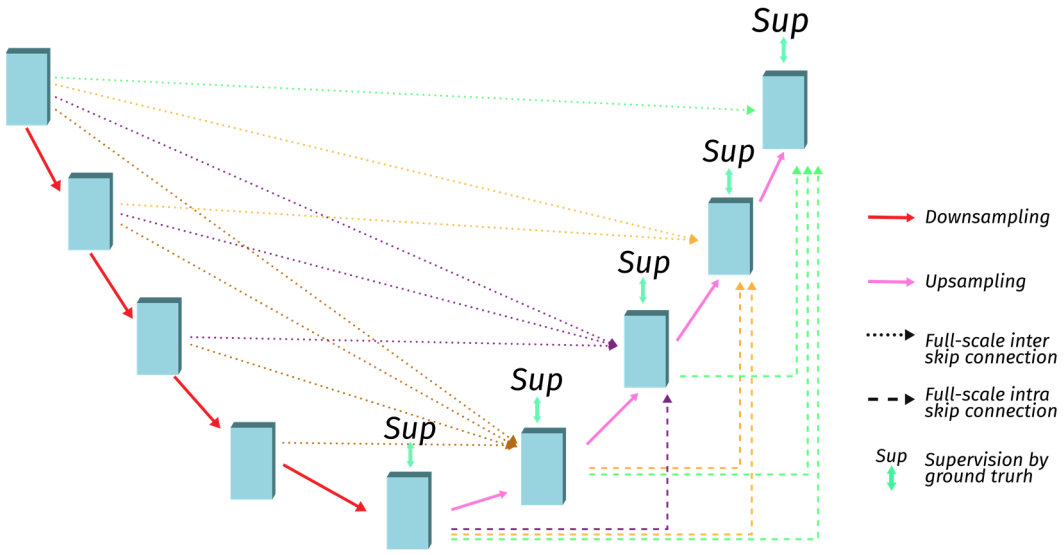


Figure 4. UNet 3+ architecture [12].

In our implementation of the UNet 3+ architecture, the model has a total of 26 million parameters and comprises a total of 114 layers, including layers with parameters and layers without parameters. Among these layers, 61 are equipped with parameters. For a more detailed scheme of UNet 3+, please refer to [12].

3.3.5. TransUNet

TransUNet [13] emerges as a transformative model within the domain of image segmentation, combining the strengths of global self-attention from Transformers and the local spatial awareness of Convolutional Neural Networks. Its primary aim is to tackle the intricate challenges presented by images, characterized by complex structures and significant variations.

The overarching goal of TransUNet is to leverage the inherent capabilities of both Transformers and CNNs to achieve precise segmentation results on medical images. With medical images often exhibiting intricate anatomical details and varying shapes, TransUNet seeks to exploit the global self-attention mechanisms of Transformers for understanding broader contextual relationships, while simultaneously harnessing the local spatial information extraction prowess of CNNs. TransUNet employs a hybrid CNN-Transformer encoder, effectively extracting high-level semantic features from image patches. Complementing this, TransUNet features a cascaded upsampler (CUP) decoder, responsible for upscaling the features and amalgamating them with high-resolution CNN features through the integration of skip-connections.

TransUNet also introduces an innovative enhancement through the inclusion of additive Transformers within the skip-connections. This strategic integration further augments the model's capability to capture salient features and intricate structures, ultimately enhancing the quality of the segmentation results. Figure 5, adapted from [13], illustrates the general structure of TransUNet.

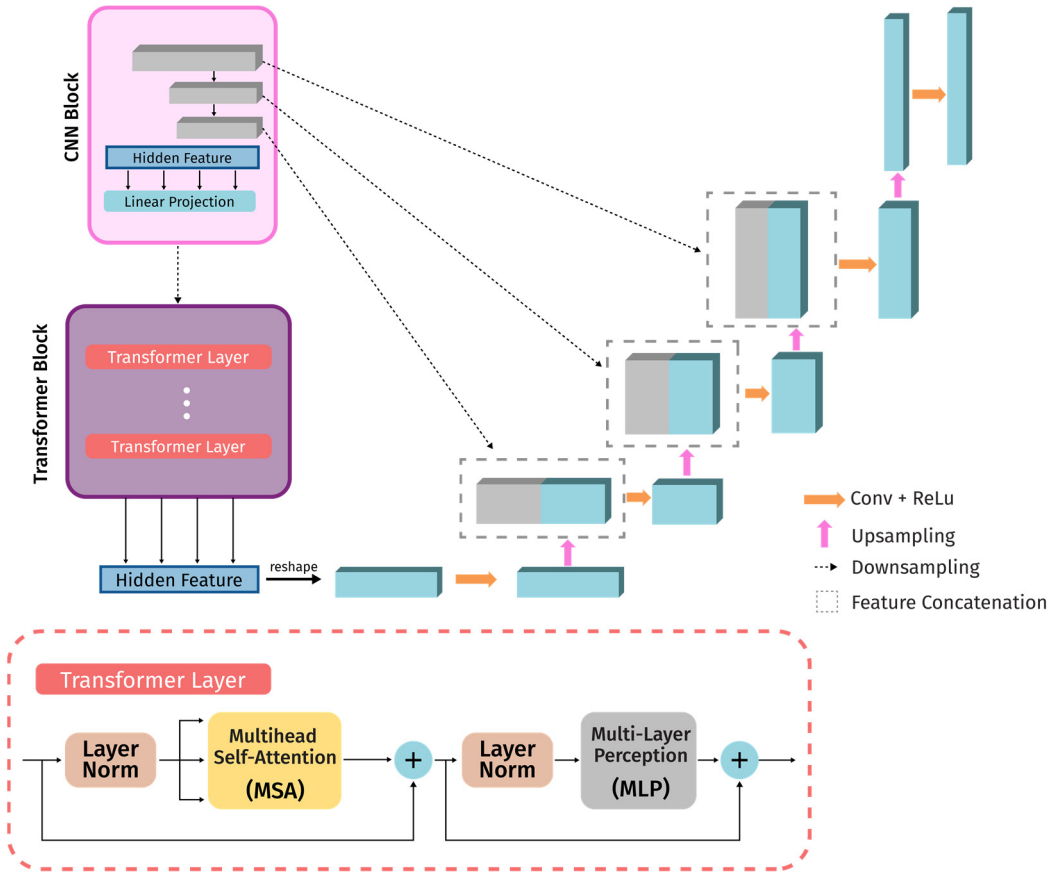


Figure 5. TransUNet architecture [13].

In our implementation of the TransUNet architecture, the model is constructed with a substantial total of 406.5 million parameters and a total of 176 layers, with 98 of these layers containing parameters. For a more detailed scheme of TransUNet, please refer to [13].

3.4. Metrics Evaluation

In this study, the performance of the five architectures is assessed using two types of metrics: similarity metrics and classification metrics. The similarity metrics used are the Intersection over Union (IoU) score and the dice score. These metrics are widely used in image segmentation studies and are recommended for evaluating segmentation models [9,56]. The IoU score and dice score measure the similarity between the predicted segmentation and the ground truth by comparing the overlapping area and the overall area of the segmentations.

The IoU score, also known as the Jaccard Index, is calculated as the intersection of the predicted and ground truth segmentations divided by the union of the two. It is represented by the formula:

$$\text{IoU} = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} \tag{2}$$

The dice score is calculated as twice the intersection of the predicted and ground truth segmentations divided by the sum of the sizes of the predicted and ground truth segmentations. It is represented by the formula:

$$\text{Dsc} = \frac{2|A_1 \cap A_2|}{|A_1| + |A_2|} \tag{3}$$

In addition to similarity metrics, this study also uses three classification metrics: accuracy, F1 score, and specificity. These classification metrics are used in this segmentation case because image segmentation is sometimes called, or also known as, pixel-level classification tasks. Although these metrics may not capture the full complexity of segmentation evaluation, they provide valuable insights into the classification performance.

The accuracy measures the overall correctness of the segmentation by calculating the ratio of correctly classified pixels to the total number of pixels. It is represented by the formula:

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

The F1 score is the harmonic mean of precision and recall, where precision represents the proportion of true positive predictions among all positive predictions, and recall represents the proportion of true positive predictions among all actual positive samples. It is represented by the formula:

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

The sensitivity, also known as recall or the true positive rate (TPR), assesses the model's capability to correctly identify all positive instances within the dataset. In the context of image segmentation, sensitivity is crucial for recognizing and not missing relevant regions. It quantifies the proportion of true positive predictions, which are the instances correctly identified as positive, among all actual positive samples. Sensitivity is represented by the formula:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

Lastly, the specificity measures the model's ability to correctly identify the true negative cases. It is calculated as the ratio of true negative predictions to the sum of true negative and false positive predictions. It is represented by the formula:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

From the five equations above, it is necessary to know that A_1 represents the ground truth area and A_2 represents the area of the model's prediction results. TP, TN, FP, and FN represent the number of classification results for each pixel.

3.5. Experimental Setting

In the pursuit of our objective to model 3D lung infection using 2D variants, it is crucial to ensure that all models are evaluated under the same settings to maintain fairness in the comparative analysis. For this purpose, the models were trained using a standardized configuration. They were compiled with the ADAM optimizer [57] using a learning rate of 1×10^{-4}. To assess the maximum capability of each architecture, a high number of epochs (5000) is set for both models, along with an early stopping mechanism. The early stopping operator monitors the validation loss with a patience of 250 epochs.

In this study, a multi-loss strategy inspired by [58] is employed. This approach utilizes multiple loss functions to enhance the model's generalization capability by incorporating diverse theoretical motivations. The total loss used in this study consists of the focal loss and dice loss. The dice loss is commonly used in image segmentation tasks as it quantifies the overlap between the predicted segmentation mask and the ground truth mask. By penalizing the discrepancies between the predicted and ground truth masks, the dice loss encourages the model to generate accurate segmentations. In addition, the focal loss is applied to address the challenge of class imbalance, which is often encountered in segmentation tasks. It assigns higher weights to misclassified pixels, prioritizing the challenging examples and effectively mitigating the class imbalance issue. By combining the dice loss and focal loss, we can leverage the advantages of both. The dice loss helps to improve overall segmentation accuracy, while the focal loss addresses class imbalance and enhances the model's handling of challenging examples. Notably, we discovered through multiple experiments that the segmentation results achieved with the combination of dice loss and focal loss were better than those obtained with either one alone.

Notably, all five models were executed without the utilization of any predefined backbones or pre-trained weights during training. This deliberate choice underscores the core objective of evaluating the pure learning capability of each model architecture, devoid of any external prior knowledge embedded within pre-trained weights. By initiating the models from scratch and allowing them to learn from the data without any pre-established biases, we aim to gain a comprehensive understanding of each architecture's innate ability to adapt to the specific task of image segmentation.

Hold-out validation is utilized for the training and testing of models. The data are split, with 75% of the data allocated for model training and 25% for model testing. This study's models are all run on Google Colab Pro, which makes use of the GPU as a hardware accelerator and high RAM usage to maximize runtime efficiency. The modeling schemes used in this study are shown in Figure 6.

3.6. Post-Processing: 3D Projection

To generate a 3D projection from the 2D output, a post-processing step is required. This step involves aggregating the CT image slices and transforming them into a coherent 3D representation. In this process, the Mayavi [59] library in Python serves as a valuable tool for facilitating the conversion of the 2D model output into a 3D representation.

The process begins by obtaining segmentation masks for each individual slice of the CT-scan. These masks provide information about the different regions within the lung. Next, the segmented slices are merged to create a unified representation. This merging process involves swapping the axes of the segmented data, ensuring that the resulting 3D projection can be visualized as a cohesive volume.

The utilization of the Mayavi library streamlines the transition of the 2D model output to a 3D representation. This library offers robust functionality for transforming and displaying the segmented data in a three-dimensional space. By employing Mayavi, the proposed approach enhances diagnostic capabilities, providing clinicians with a comprehensive 3D visualization of lung structures and infection areas.

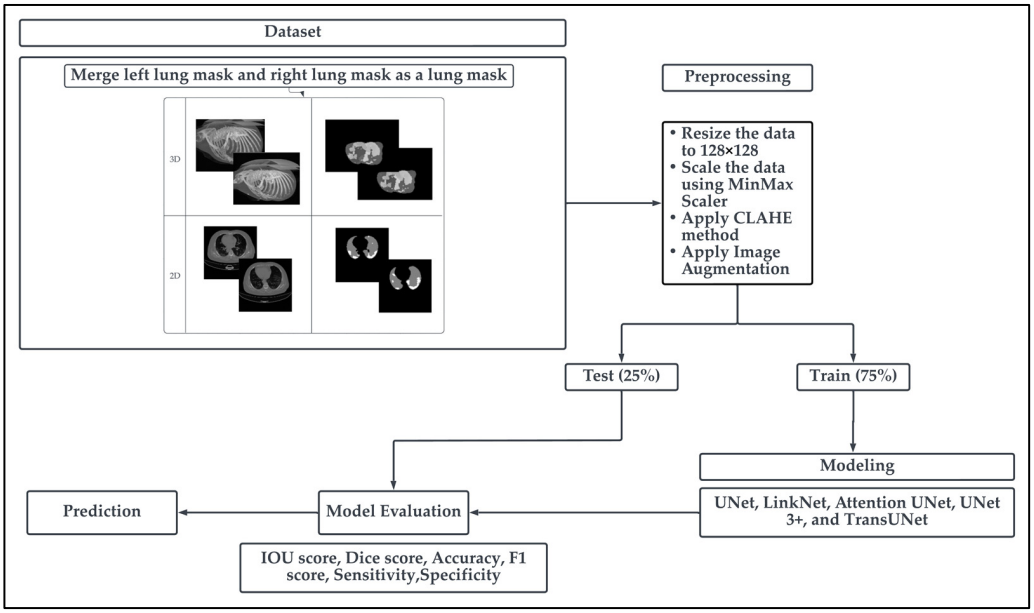


Figure 6. The modeling schemes.

4. Results

In this section, we showcase the results of employing advanced deep learning techniques, specifically UNet, Linknet, Attention UNet, UNet 3+, and TransUNet architectures, for 3D projection reconstruction in 2D lung infection diagnosis. Our focus lies on visualizing the intricate lung structures affected by COVID-19 through segmentation experiments conducted on CT-scans. To assess the effectiveness of these architectures, we utilize various evaluation metrics including specificity, sensitivity F1 score, accuracy dice score, and IoU score. Additionally, we analyze the training process by comparing the loss graphs of UNet, Linknet, Attention UNet, UNet 3+, and TransUNet to gain further insights into their respective convergence behaviors.

4.1. Training Process

The training process for lung and infection segmentation on CT-scans was conducted using UNet, Linknet, Attention UNet, UNet 3+, and TransUNet architectures. The hyperparameters and dataset were standardized to ensure a fair and appropriate comparison. Figure 7 depicts the loss and the IoU score graphs of UNet, Linknet, Attention UNet, UNet 3+, and TransUNet during the training process.

From Figure 7, it is evident that all five algorithms exhibit the expected decreasing pattern. Notably, both UNet and Attention UNet stand out for their smoother convergence trajectories, distinct from the other algorithms that exhibit sporadic anomalies with intermittent high loss values across various epochs. Taking a closer look at the convergence behaviors, LinkNet, while exhibiting more abrupt fluctuations with occasional spikes in loss values, demonstrates a relatively rapid pace in reaching the learning limit. Specifically, LinkNet achieves learning convergence in just 542 epochs, making it the quickest to approach the learning limit within the experiment. On the other hand, among the other models, UNet 3+ requires 665 epochs, standard UNet requires 948 epochs, Attention UNet needs 953 epochs, and TransUNet exhibits the longest convergence trajectory, taking 1789 epochs to reach the learning limit. Similarly, from the IoU score graph in Figure 7, all five algorithms exhibit the expected increasing pattern, with UNet and Attention UNet standing out for their smoother convergence trajectories.

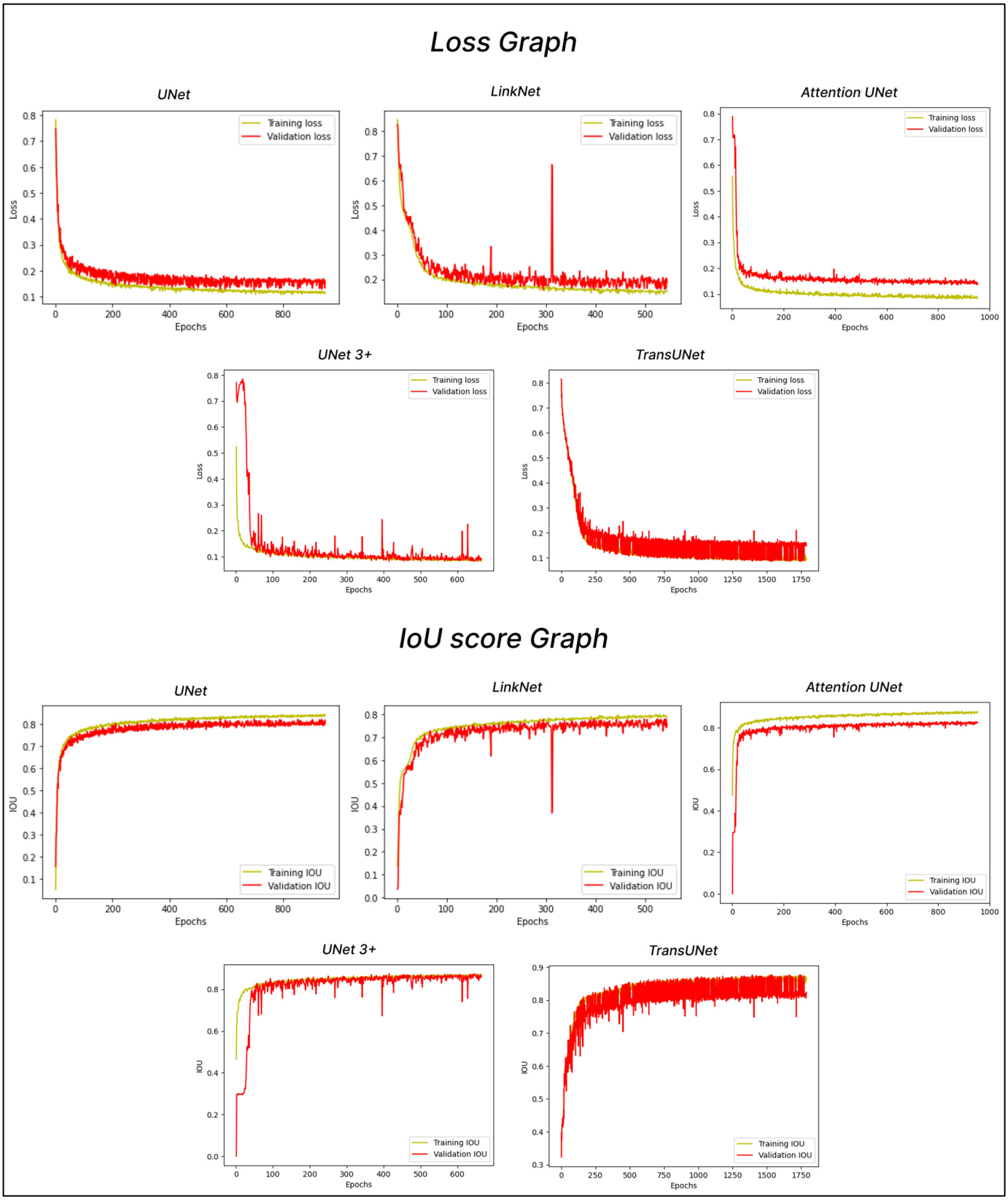


Figure 7. Comparison of the loss and the IoU score graphs of the UNet, Linknet, Attention UNet, UNet 3+, and TransUNet algorithms during the training process.

4.2. Metrics Evaluation and Prediction

Table 5 provides a comprehensive overview of the metrics evaluation outcomes for the COVID-19 3D lung construction segmentation task using UNet, LinkNet, Attention UNet, UNet 3+, and TransUNet architectures. Focusing on the architecture with the most prominent performance, Attention UNet emerges as the standout performer in this analysis. Among the evaluated metrics, Attention UNet attains superior results, underscoring its effectiveness in segmenting the 3D lung construction affected by COVID-19. Specifically, Attention UNet achieves an IoU score of 85.36%, a dice score of 91.49%, an accuracy rate of 98.63%, an F1 score of 98.64%, a sensitivity of 98.63%, and an impressive specificity of 99.32%.

Table 5. Metrics evaluation of UNet, Linknet, Attention UNet, UNet 3+, and TransUNet.

Metrics	UNet	LinkNet	Attention UNet	Unet 3+	TransUNet
IoU score	84.82%	79.31%	**85.36%**	84.77%	83.52%
Dice score	91.07%	86.71%	**91.49%**	91.02%	90.27%
Accuracy	98.58%	98.22%	98.63%	98.63%	**98.96%**
F1 score	98.58%	98.23%	**98.64%**	98.64%	98.33%
Sensitivity	98.58%	98.21%	98.63%	98.63%	**99.81%**
Specificity	99.29%	99.12%	**99.32%**	99.32%	98.41%
Max Epoch [1]	698	**294**	703	415	1539

[1] The Max Epoch is determined by reducing the total training epochs using the early stopping patience value. The bolded numbers in the table represent the best values in comparison to other architectures.

In Table 6, a comparison of the ground truth and prediction results of each model is presented in 2D, along with the 3D projection of each model. The visual analysis of the predictions provides additional insights into the performance of each model. The visualizations demonstrate that Attention UNet captures the intricate details of the lung construction more accurately, resulting in more precise segmentation of the infected areas compared to other models.

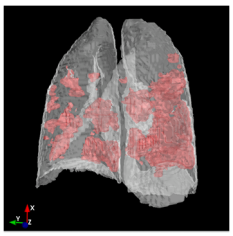
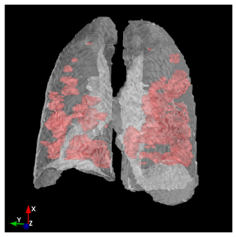
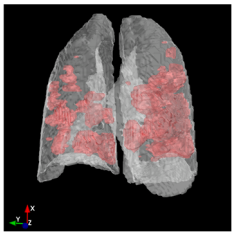
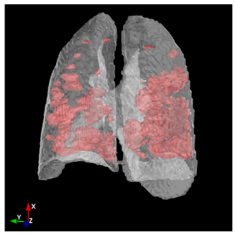
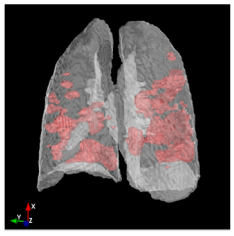
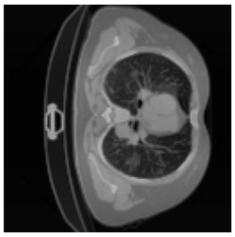
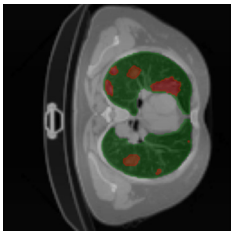
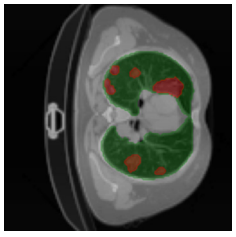
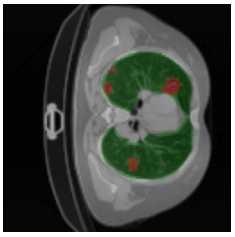
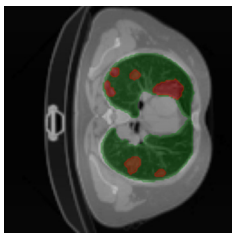
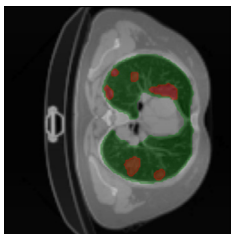
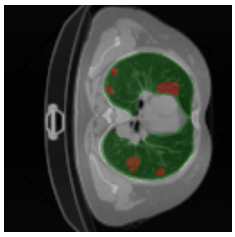
4.3. Result Discussion

Using the diverse set of 2D segmentation architectures investigated in this study, we have achieved significant advancements in the reconstruction of 3D lung and infection projections from CT-scans. Each architecture demonstrates unique characteristics that contribute to their respective performances, enabling a comprehensive evaluation of their applicability in medical imaging tasks.

Attention UNet stands out due to its sophisticated attention mechanisms that allow it to focus on relevant regions and features within the lung structures. This characteristic enhances its ability to capture intricate details specific to COVID-19-affected lungs. The attention mechanisms contribute to Attention UNet's impressive IoU score of 85.36% and dice score of 91.49%, showcasing its potential to provide accurate and finely detailed segmentations. The results suggest that its trait of dynamically assigning different weights to regions, based on their contextual importance, is instrumental in achieving precise and meaningful segmentations in complex medical images.

UNet, characterized by its fundamental U-shaped architecture, adeptly strikes a balance between contextual information and local intricacies. This trait empowers it to attain an impressive IoU score of 84.82% and dice score of 91.07%, underscoring its proficiency in capturing both macroscopic patterns and intricate structural nuances. The consistent convergence observed in UNet could potentially be attributed to its inherent simplicity and directness in combining encoder and decoder features. This straightforward approach aids in achieving a holistic contextual understanding while simultaneously preserving intricate spatial details, ultimately contributing to its commendable performance.

Table 6. Comparison of ground truth and model prediction results using UNet, LinkNet, Attention UNet, UNet 3+, and TransUNet.

Comparison of Ground Truth and Model Predictions in 3D Projections and 2D Slice				
3D Projection	 Original CT-scan	 Ground truth	 UNet	 LinkNet
		 Attention UNet	 UNet 3+	 TransUNet
2D Slice	 Original CT-scan	 Ground Truth	 UNet	 LinkNet
		 Attention UNet	 UNet 3+	 TransUNet

UNet 3+ demonstrates an IoU score of 84.77% and a dice score of 91.02%, showcasing its ability to effectively capture multi-scale features and generate hierarchical representations. This suggests that the architecture's unique traits play a significant role in its performance. The incorporation of full-scale skip connections and deep supervision within UNet 3+ enables it to seamlessly aggregate high-level semantics and intricate details. This approach empowers the model to maintain a harmonious balance between local and global information, ultimately enhancing its segmentation capabilities.

TransUNet employs a distinctive approach by combining the capabilities of Transformers and UNet to effectively integrate global context and local spatial details. Despite requiring more training epochs compared to some other architectures, TransUNet achieves notable segmentation results. With an IoU score of 83.52% and a dice score of 90.27%, TransUNet's unique trait of harnessing Transformers for capturing long-range dependencies, in conjunction with UNet's spatial understanding, comes into play. This synergy

allows TransUNet to effectively capture complex lung structures while considering broader contextual information.

LinkNet, despite slightly lower metrics, showcases the trait of computational efficiency. Its architecture efficiently combines low-level and high-level features, producing an IoU score of 79.31% and dice score of 86.71%. This efficiency is advantageous in resource-constrained settings. The trade-off between performance and resource consumption is essential, as the choice of LinkNet can be particularly valuable when computational resources and time constraints are considerations.

In summary, our experimental results underscore the potential of different architectures for segmenting COVID-19-affected lung constructions. Each architecture's unique traits play a significant role in determining its performance. Attention mechanisms, hierarchical feature fusion, global–local feature integration, and computational efficiency are among the characteristics that contribute to the observed results. The choice of architecture should consider these traits, and further exploration of deep learning architectures and optimization techniques will continue to drive advancements in COVID-19 lung construction visualization and medical image segmentation.

5. Conclusions

In conclusion, this study delved into the potential of 2D architectures—specifically UNet, LinkNet, Attention UNet, UNet 3+, and TransUNet—for reconstructing 3D lung projections in the context of diagnosing lung infections, particularly in COVID-19 patients. The evaluation of these diverse architectures provided invaluable insights into their respective performances and effectiveness in accurately capturing lung structures and infection areas. This endeavor contributes to the advancement of medical image analysis, offering essential guidance to healthcare professionals and researchers in choosing appropriate techniques for assessing disease severity and monitoring its progression. Moreover, the practical implications extend to the development of more efficient clinical protocols for lung diseases.

Our experimentation revealed that each architecture possesses distinct traits that impact its performance. UNet demonstrated remarkable results, achieving a dice score of 91.07% and an IoU score of 84.82%. Its foundational U-shaped architecture, which strikes a balance between contextual understanding and local details, proved to be effective in accurately segmenting lung structures. LinkNet, on the other hand, exhibited slightly lower performance, with a dice score of 86.71% and an IoU score of 79.31%. While its computational efficiency is notable. Attention UNet showcased its capability to focus on salient features by leveraging attention mechanisms, resulting in the highest performance with an IoU score of 85.36% and a dice score of 91.49%. UNet 3+ exploited full-scale skip connections and hierarchical feature fusion, yielding an IoU score of 84.77% and a dice score of 91.02%. TransUNet harnessed Transformers to capture global context, achieving an IoU score of 83.52% and a dice score of 90.27%. Although TransUNet required more epochs to converge due to its trait of integrating long-range dependencies, it still exhibited competitive performance.

Considering the results and the scope of our study, there are potential future developments and research directions that can be explored. Investigating the generalizability and robustness of these architectures by testing them on larger and more diverse datasets would be valuable. Assessing their performance on CT-scans from different populations and varying disease stages can help to identify any limitations or potential biases in the models' segmentation capabilities.

In summary, this study underscores the significance of these 2D architectures in visualizing and understanding lung constructions using CT-scans. These findings provide valuable insights for the development of advanced segmentation models and pave the way for further advancements in accurately visualizing and understanding the lung construction using CT-scans.

Author Contributions: A.A.P., M.H.A., F.A.L.N., G.D. and T.H. comprehended and devised the research plan; A.A.P. and M.H.A. conducted data analysis and composed the initial draft. The draft was then reviewed and revised by all authors, who subsequently granted their approval for the final version of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank Higher Education Research Grant No. 094/E5/PG.02.00.PT/2022 for supporting this research, as well as the Universitas Padjadjaran and the Directorate for Research and Community Service (DRPM) Ministry of Research, Technology, and Higher Education Indonesia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in [Zenodo] at [https://doi.org/10.5281/zenodo.3757476] (accessed on 18 November 2022).

Acknowledgments: The authors acknowledge their gratitude to Universitas Padjadjaran's Research Center for Artificial Intelligence and Big Data (AIDA) for its invaluable assistance and resources.

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

References

1. Dhar, D.; Mohanty, A. Gut Microbiota and COVID-19-Possible Link and Implications. *Virus Res.* **2020**, *285*, 198018. [CrossRef] [PubMed]
2. Oda, M.; Hayashi, Y.; Otake, Y.; Hashimoto, M.; Akashi, T.; Mori, K. Lung Infection and Normal Region Segmentation from CT Volumes of COVID-19 Cases. In Proceedings of the Medical Imaging 2021: Computer-Aided Diagnosis, Online, 15–19 February 2021; p. 103.
3. Shan, F.; Gao, Y.; Wang, J.; Shi, W.; Shi, N.; Han, M.; Xue, Z.; Shen, D.; Shi, Y. Abnormal Lung Quantification in Chest CT Images of COVID-19 Patients with Deep Learning and Its Application to Severity Prediction. *Med. Phys.* **2021**, *48*, 1633–1645. [CrossRef]
4. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* **2020**, *296*, E32–E40. [CrossRef] [PubMed]
5. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P.; et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA—Secondary Publication. *J. Thorac. Imaging* **2020**, *35*, 219–227. [CrossRef]
6. Yin, S.; Deng, H.; Xu, Z.; Zhu, Q.; Cheng, J. SD-UNet: A Novel Segmentation Framework for CT Images of Lung Infections. *Electronics* **2022**, *11*, 130. [CrossRef]
7. Gouda, W.; Almurafeh, M.; Humayun, M.; Jhanjhi, N.Z. Detection of COVID-19 Based on Chest X-rays Using Deep Learning. *Healthcare* **2022**, *10*, 343. [CrossRef] [PubMed]
8. Malmberg, F.; Lindblad, J.; Sladoje, N.; Nyström, I. A Graph-Based Framework for Sub-Pixel Image Segmentation. *Theor. Comput. Sci.* **2011**, *412*, 1338–1349. [CrossRef]
9. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. ISBN 978-3-319-24573-7.
10. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
11. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.0399.
12. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A Full-Scale Connected Unet for Medical Image Segmentation. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
13. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
14. Asnawi, M.H.; Pravitasari, A.A.; Darmawan, G.; Hendrawati, T.; Yulita, I.N.; Suprijadi, J.; Nugraha, F.A.L. Lung and Infection CT-Scan-Based Segmentation with 3D UNet Architecture and Its Modification. *Healthcare* **2023**, *11*, 213. [CrossRef]
15. Efremova, D.B.; Konovalov, D.A.; Siriapisith, T.; Kusakunniran, W.; Haddawy, P. Automatic Segmentation of Kidney and Liver Tumors in CT Images 2019. *arXiv* **2019**, arXiv:1908.01279.
16. Rahman, H.; Bukht, T.F.N.; Imran, A.; Tariq, J.; Tu, S.; Alzahrani, A. A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. *Bioengineering* **2022**, *9*, 368. [CrossRef] [PubMed]

17. Natarajan, V.A.; Sunil Kumar, M.; Patan, R.; Kallam, S.; Noor Mohamed, M.Y. Segmentation of Nuclei in Histopathology Images Using Fully Convolutional Deep Neural Architecture. In Proceedings of the 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, 9 September 2020; pp. 1–7.
18. Islam, M.; Vibashan, V.S.; Jose, V.J.M.; Wijethilake, N.; Utkarsh, U.; Ren, H. Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, 17 October 2019; Revised Selected Papers, Part I 5. Springer: Berlin, Germany, 2020; pp. 262–272.
19. Abedalla, A.; Abdullah, M.; Al-Ayyoub, M.; Benkhelifa, E. Chest X-Ray Pneumothorax Segmentation Using U-Net with EfficientNet and ResNet Architectures. *PeerJ Comput. Sci.* **2021**, *7*, e607. [CrossRef] [PubMed]
20. Sabir, M.W.; Khan, Z.; Saad, N.M.; Khan, D.M.; Al-Khasawneh, M.A.; Perveen, K.; Qayyum, A.; Azhar Ali, S.S. Segmentation of Liver Tumor in CT Scan Using ResU-Net. *Appl. Sci.* **2022**, *12*, 8650. [CrossRef]
21. Sobhaninia, Z.; Rezaei, S.; Noroozi, A.; Ahmadi, M.; Zarrabi, H.; Karimi, N.; Emami, A.; Samavi, S. Brain Tumor Segmentation Using Deep Learning by Type Specific Sorting of Images 2018. *arXiv* **2018**, arXiv:1809.07786.
22. Akyel, C.; Arıcı, N. LinkNet-B7: Noise Removal and Lesion Segmentation in Images of Skin Cancer. *Mathematics* **2022**, *10*, 736. [CrossRef]
23. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–1924.
24. Roth, H.; Farag, A.; Turkbey, E.B.; Lu, L.; Liu, J.; Summers, R.M. Data From Pancreas-CT 2016. *Cancer Imaging Arch.* **2016**. [CrossRef]
25. Roth, H.R.; Oda, H.; Hayashi, Y.; Oda, M.; Shimizu, N.; Fujiwara, M.; Misawa, K.; Mori, K. Hierarchical 3D Fully Convolutional Networks for Multi-Organ Segmentation. *arXiv* **2017**, arXiv:1704.06382.
26. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.R.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Towards Image-Guided Pancreas and Biliary Endoscopy: Automatic Multi-Organ Segmentation on Abdominal CT with Dense Dilated Networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention−MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; Proceedings, Part I 20. Springer: Berlin, Germany, 2017; pp. 728–736.
27. Roth, H.R.; Lu, L.; Lay, N.; Harrison, A.P.; Farag, A.; Sohn, A.; Summers, R.M. Spatial Aggregation of Holistically-Nested Convolutional Neural Networks for Automated Pancreas Localization and Segmentation. *Med. Image Anal.* **2018**, *45*, 94–107. [CrossRef]
28. Cai, J.; Lu, L.; Xie, Y.; Xing, F.; Yang, L. Improving Deep Pancreas Segmentation in CT and MRI Images via Recurrent Neural Contextual Learning and Direct Loss Function. *arXiv* **2017**, arXiv:1707.04912.
29. Zhou, Y.; Xie, L.; Shen, W.; Wang, Y.; Fishman, E.K.; Yuille, A.L. A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scans. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 10–14 September 2017; Springer: Berlin, Germany, 2017; pp. 693–701.
30. Nguyen, D.-T.; Tran, T.-T.; Pham, V.-T. Attention U-Net with Active Contour Based Hybrid Loss for Brain Tumor Segmentation. In *Soft Computing: Biomedical and Related Applications*; Phuong, N.H., Kreinovich, V., Eds.; Studies in Computational Intelligence; Springer International Publishing: Cham, Switzerland, 2021; Volume 981, pp. 35–45. ISBN 978-3-030-76619-1.
31. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
32. Rajamani, K.T.; Rani, P.; Siebert, H.; ElagiriRamalingam, R.; Heinrich, M.P. Attention-Augmented U-Net (AA-U-Net) for Semantic Segmentation. *Signal Image Video Process.* **2023**, *17*, 981–989. [CrossRef] [PubMed]
33. Amer, A.; Lambrou, T.; Ye, X. MDA-Unet: A Multi-Scale Dilated Attention U-Net for Medical Image Segmentation. *Appl. Sci.* **2022**, *12*, 3676. [CrossRef]
34. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
35. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
36. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
37. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
38. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.

39. Fu, S.; Lu, Y.; Wang, Y.; Zhou, Y.; Shen, W.; Fishman, E.; Yuille, A. Domain Adaptive Relational Reasoning for 3d Multi-Organ Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, 4–8 October 2020; Proceedings, Part I 23. Springer: Berlin, Germany, 2020; pp. 656–666.
40. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef]
41. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
42. Wu, T.; Li, B.; Luo, Y.; Wang, Y.; Xiao, C.; Liu, T.; Yang, J.; An, W.; Guo, Y. MTU-Net: Multilevel TransUNet for Space-Based Infrared Tiny Ship Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1. [CrossRef]
43. Chen, J.; Chen, J.; Zhou, Z.; Li, B.; Yuille, A.; Lu, Y. MT-TransUNet: Mediating Multi-Task Tokens in Transformers for Skin Lesion Segmentation and Classification. *arXiv* **2021**, arXiv:2112.01767.
44. Jamali, A.; Roy, S.K.; Li, J.; Ghamisi, P. TransU-Net++: Rethinking Attention Gated TransU-Net for Deforestation Mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103332. [CrossRef]
45. Jun, M.; Cheng, G.; Yixin, W.; Xingle, A.; Jiantao, G.; Ziqi, Y.; Minqing, Z.; Xin, L.; Xueyuan, D.; Shucheng, C.; et al. COVID-19 CT Lung and Infection Segmentation Dataset. *OpenAIRE*. 2020. Available online: https://zenodo.org/records/3757476 (accessed on 18 November 2022).
46. Radiopaedia Pty Ltd. ACN 133 562 722. Available online: https://radiopaedia.org/ (accessed on 23 July 2022).
47. RAIOSS Coronacases. Available online: https://coronacases.org/ (accessed on 23 July 2022).
48. Ma, J.; Wang, Y.; An, X.; Ge, C.; Yu, Z.; Chen, J.; Zhu, Q.; Dong, G.; He, J.; He, Z.; et al. Towards Data-Efficient Learning: A Benchmark for COVID-19 CT Lung and Infection Segmentation. *Med. Phys.* **2021**, *48*, 1197–1210. [CrossRef] [PubMed]
49. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems*; Elsevier: Amsterdam, The Netherlands, 1994; pp. 474–485. ISBN 978-0-12-336156-1.
50. Lin, Z.; Yingjie, Z.; Bochao, D.; Bo, C.; Yangfan, L. Welding Defect Detection Based on Local Image Enhancement. *IET Image Process.* **2019**, *13*, 2647–2658. [CrossRef]
51. Pravitasari, A.A.; Iriawan, N.; Almuhayar, M.; Azmi, T.; Irhamah, I.; Fithriasari, K.; Purnami, S.W.; Ferriastuti, W. UNet-VGG16 with Transfer Learning for MRI-Based Brain Tumor Segmentation. *TELKOMNIKA* **2020**, *18*, 1310. [CrossRef]
52. Mahmoudi, R.; Benameur, N.; Mabrouk, R.; Mohammed, M.A.; Garcia-Zapirain, B.; Bedoui, M.H. A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging. *Appl. Sci.* **2022**, *12*, 4825. [CrossRef]
53. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
54. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
55. Kholiavchenko, M.; Sirazitdinov, I.; Kubrak, K.; Badrutdinova, R.; Kuleev, R.; Yuan, Y.; Vrtovec, T.; Ibragimov, B. Contour-Aware Multi-Label Chest X-Ray Organ Segmentation. *Int. J. CARS* **2020**, *15*, 425–436. [CrossRef]
56. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [CrossRef]
57. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization 2017. *arXiv* **2017**, arXiv:1412.6980.
58. Xu, C.; Lu, C.; Liang, X.; Gao, J.; Zheng, W.; Wang, T.; Yan, S. Multi-Loss Regularized Deep Neural Network. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 2273–2283. [CrossRef]
59. Ramachandran, P.; Varoquaux, G. Mayavi: 3D Visualization of Scientific Data. *Comput. Sci. Eng.* **2011**, *13*, 40–51. [CrossRef]

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.