

Article

Deep Learning Based on EfficientNet for Multiorgan Segmentation of Thoracic Structures on a 0.35 T MR-Linac Radiation Therapy System

Mohammed Chekroun ¹, Youssef Mourchid ^{1,*} , Igor Bessières ² and Alain Lalande ^{3,4} ¹ CESI LINEACT Laboratory, UR 7527, 21000 Dijon, France; chekrouoss@gmail.com² Centre Georges-François Leclerc, 21000 Dijon, France; ibessieres@cgfl.fr³ IFTIM, ICMUB Laboratory, UMR CNRS 6302, University of Burgundy, 21000 Dijon, France; alain.lalande@u-bourgogne.fr⁴ Medical Imaging Department, University Hospital of Dijon, 21000 Dijon, France

* Correspondence: ymourchid@cesi.fr

Abstract: The advent of the 0.35 T MR-Linac (MRIdian, ViewRay) system in radiation therapy allows precise tumor targeting for moving lesions. However, the lack of an automatic volume segmentation function in the MR-Linac's treatment planning system poses a challenge. In this paper, we propose a deep-learning-based multiorgan segmentation approach for the thoracic region, using EfficientNet as the backbone for the network architecture. The objectives of this approach include accurate segmentation of critical organs, such as the left and right lungs, the heart, the spinal cord, and the esophagus, essential for minimizing radiation toxicity during external radiation therapy. Our proposed approach, when evaluated on an internal dataset comprising 81 patients, demonstrated superior performance compared to other state-of-the-art methods. Specifically, the results for our approach with a 2.5D strategy were as follows: a dice similarity coefficient (DSC) of 0.820 ± 0.041 , an intersection over union (IoU) of 0.725 ± 0.052 , and a 3D Hausdorff distance (HD) of 10.353 ± 4.974 mm. Notably, the 2.5D strategy surpassed the 2D strategy in all three metrics, exhibiting higher DSC and IoU values, as well as lower HD values. This improvement strongly suggests that our proposed approach with the 2.5D strategy may hold promise in achieving more precise and accurate segmentations when compared to the conventional 2D strategy. Our work has practical implications in the improvement of treatment planning precision, aligning with the evolution of medical imaging and innovative strategies for multiorgan segmentation tasks.

Keywords: organs-at-risk; segmentation; MRI; deep learning; radiation therapy; 0.35 T MR-Linac



Citation: Chekroun, M.; Mourchid, Y.; Bessières, I.; Lalande, A. Deep Learning Based on EfficientNet for Multiorgan Segmentation of Thoracic Structures on a 0.35 T MR-Linac Radiation Therapy System.

Algorithms **2023**, *16*, 564. <https://doi.org/10.3390/a16120564>

Academic Editor: Frank Werner

Received: 31 October 2023

Revised: 1 December 2023

Accepted: 9 December 2023

Published: 12 December 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

Since 2017, magnetic-resonance-guided radiotherapy (MRgRT) has been enabled thanks to a dedicated linear accelerator (Linac) with embedded MR imaging, called MR-Linac [1]. This type of RT device offers alternative image quality in comparison to traditional X-ray-based imaging. In this specific workflow, MR images for treatment planning and delivery must be acquired with the same MR-Linac, and then by extension, the MRI for planning cannot be performed on another classical MRI system. In particular, due to the non-radiative nature of MR imaging, continuous imaging can be performed without any limitation during treatment delivery in order to verify the tumor position. This option is particularly relevant for lung tumors that are often subject to breathing movements. MRgRT for lung tumors has been performed since 2019 at the Centre Georges-François Leclerc (CGFL) with the MRIdian (Viewray Inc., Oakwood Village, OH, USA) 0.35 T MR-Linac representing around 15% of the patients treated on this system [2]. The success of this treatment and the decrease in toxicity strongly rely on the exact definition of the different organs at risk (OARs). Indeed, the delineation of the target volume and the

OARs on a reference set of MR images is required for planning and is usually performed manually by radiation oncologists. Nevertheless, manual segmentation is subject to inter- or intra-observer variability, particularly according to the level of expertise. Moreover, it represents a very time-consuming step in the RT workflow, requiring several dozens of minutes depending on the localization.

Several studies have been proposed for the segmentation of various anatomical regions from MR images. Tong et al. [3] introduced a method based on the generative adversarial network (GAN) with a shape constraint (SC-GAN) developed for fully automated head and neck OARs segmentation on a low-field MRI. Their proposed method underwent testing on 25 0.35 T MR images obtained from a dedicated MR-guided radiotherapy system. The OARs comprised the brainstem, the optical chiasm, the larynx, the mandible, and the pharynx. Additionally, they validated their method on other CT scans for different organs. For the same anatomical region (head and neck), Liu et al. [4] proposed a head and neck multi-organ automatic segmentation on CT images. Nineteen OARs were targeted and the MR scans were acquired on a 1.5 T scanner (T1-weighted images). They achieved better results for the segmentation of the mandible, parotid, optic nerve, and chiasm compared to those obtained in the study performed by Tong et al. [3]. Chen et al. [5] developed an automated deep-learning-based technique for the segmentation of abdominal organs based on a two-dimensional U-net and a densely connected network structure. The images were acquired on a 3T system.

As for the thoracic region, numerous approaches have been proposed for the automatic segmentation of thoracic organs on CT images, including atlas-based methods, level-set methods, and morphological methods [6–8]. The majority of existing studies concerning the segmentation of thoracic organs have primarily concentrated on CT scans [9–12]. While there are several publicly available CT datasets that are widely used for multiorgan segmentation in the thoracic region, such as the MSD (Medical Segmentation Decathlon) lung and heart dataset [13] and the SegTHOR (Thoracic Organs at Risk) dataset [14], there is a lack of publicly available MRI datasets that specifically address the segmentation of thoracic organs. This disparity in data availability underscores the need to explore innovative segmentation methods. Traditionally, medical image segmentation has been explored using various “statistical” and “prior-knowledge-based” methods such as thresholding, region growing, and atlas-based segmentation. However, these methods exhibit limitations, such as sensitivity to parameter settings, poor adaptability to diverse data, manual intervention requirements, and difficulty with complex structures. Recently, deep learning (DL) methodologies have emerged as state-of-the-art approaches in medical imaging, offering enhanced performance, flexibility, and automation using learning features directly from the data and capturing contextual information. Notably, the UNet architecture has gained significant recognition for its effectiveness in semantic segmentation tasks [15]. Researchers have extended the original UNet architecture to improve its performance. For instance, the Residual Attention UNet [16] incorporates residual attention modules to enhance the model’s capability to focus on relevant regions while suppressing noise and irrelevant features. TransUNet [17] integrates transformer-based self-attention mechanisms into the UNet architecture, allowing the model to capture long-range dependencies and spatial relationships effectively. One notable encoder architecture is EfficientNet [18], which has demonstrated remarkable performance in various computer vision tasks. EfficientNet is known for its efficient and scalable design, achieving excellent results with fewer parameters [19]. The combination of UNet’s skip connections and EfficientNet’s feature extraction capabilities enables the model to capture both local and global contextual information, leading to more accurate and robust segmentation. In the context of multiorgan segmentation, the 2D multi-class strategies independently segment 2D slices, and have proven effective in various medical imaging applications. Studies of thoracic CT scans consistently favor 2D approaches, as demonstrated by Trullo et al. [20]. Their work highlights 2D’s superiority over 3D, emphasizing its precision in multiorgan segmentation. However, 2D multi-class strategies are susceptible to losing 3D spatial context, thereby influencing

the volumetric interrelationships. They may encounter difficulties when confronted with intricate anatomical structures that extend across multiple slices and may face challenges in the segmentation of irregularly shaped or variable organs. To overcome these limitations, a 2.5D strategy is proposed [21], which combines contextual awareness with 2D efficiency. It integrates neighboring slice data, compensating for 3D context loss in 2D strategies. This improved context enhances precision, especially with complex or irregular organs.

The main objective of this work is to facilitate treatment planning by reducing organ contouring time, taking advantage of the accuracy of DL-based techniques. Our proposed method for segmenting OARs in the thoracic region involves the use of MR-Linac-acquired images. Our approach is based on an internal dataset and is driven by clinical requirements. The primary focus is placed on five critical organs: the heart, the right lung, the left lung, the spinal cord, and the esophagus. Essential preprocessing steps are applied to prepare the dataset for training, employing various techniques. Model selection and the choice of associated loss functions are carefully considered, prioritizing performance evaluation through five specific metrics: the Dice similarity coefficient (DSC) [22], intersection over union (IoU) [23], the Hausdorff distance (HD) [24], results from Bland–Altman plots concerning the volumes of the organs [25], and correlation studies, also concerning the volumes of the organs [26], that assess organ volume relationships. Our proposed method aims to contribute to the field of multiorgan segmentation in the thoracic region using MR images. The key contributions can be summarized as follows:

- A preprocessing approach is presented for MRI datasets, utilizing YOLO for region-of-interest detection, then removal of unwanted regions, followed by adaptive histogram equalization and thresholding, to enhance dataset quality and completeness.
- A novel approach for multiorgan segmentation in the thoracic region of MR images is proposed based on the EfficientNet as an encoder for UNet with a 2.5D strategy in the context of a 0.35 T MR-Linac radiation therapy system.
- The efficiency of the proposed model is demonstrated through extensive experimentation on an internal dataset of 81 patients.

2. Materials and Methods

2.1. Dataset

A total of 81 MRI studies of the thoracic region were collected from patients treated for lung cancer at the CGFL in Dijon, France. The MRIdian 0.35 T MR-Linac (Viewray Inc., Oakwood Village, OH, USA) was used for image acquisition and for the treatments. A balanced steady-state free precession (bSSFP) sequence producing T2/T1-weighted images was used for the MRI acquisitions with dedicated thoracic coils. In order to limit movement artifacts, all acquisitions were performed with the patient lying supine in a reproducible breath-hold position without any contrast media or patient-specific conditioning. This bSSFP sequence had an echo-time of 1.27 ms, a repetition-time of 3.00 ms, and a total acquisition time of 25 s. The full field of view had the following dimensions: $465 \times 540 \times 429 \text{ mm}^3$, with a voxel size of $1.5 \times 1.5 \times 3.0 \text{ mm}^3$ (matrix of $310 \times 360 \times 143$). Subsequent to each MRI exam, the radiation therapy structure set (RT structure) was retrieved. This set encompassed all the segmented targets and the OARs created by experts using the treatment planning software of the MR-Linac machine. Both the MRI exams and the RT structure were stored as DICOM files.

Figure 1 presents a sample of the original DICOM images along with the corresponding structures segmented by medical experts for five organs: the left lung, the right lung, the esophagus, the spinal cord, and the heart. For each patient, there were up to 144 slices, creating 6852 images in our dataset. Generally, having a larger dataset can contribute to better generalization and performance of the deep learning model. In the context of multiorgan segmentation, where nuances and variations can be critical, a dataset of 6852 images is considered relatively large and should provide a solid foundation for training our multiorgan segmentation model, potentially leading to more accurate and reliable results.

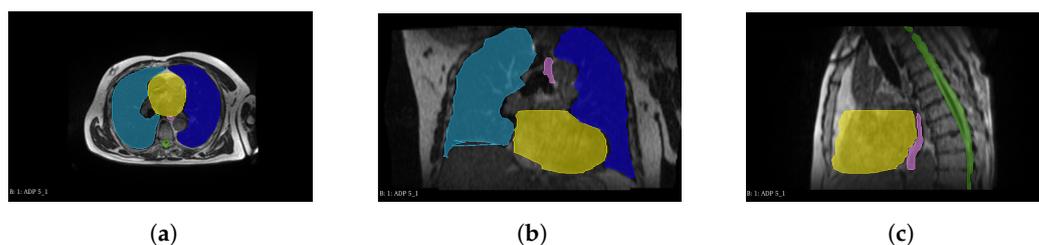


Figure 1. A preview of an MRI study in (a) axial orientation, (b) coronal orientation, and (c) sagittal orientation. The segmented organs include the left and right lung, the heart, the spinal cord, and the esophagus, depicted in light blue, blue, yellow, green, and pink colors, respectively. The delineation of the OARs can be rough as for the heart in this example, and then rectified to train our model.

2.2. Preprocessing

Some problems were related to the image content and quality. In some slices, the presence of noise and the patients' arms in the background was noticeable (Figure 2). Moreover, some slices exhibited low luminosity, resulting in poor visualization of the organs. These issues can hinder the effectiveness of our automatic segmentation model. Low luminosity reduces image contrast, making it challenging for the model to distinguish organ boundaries accurately, while image noise or artifacts can introduce spurious details that lead to false positives or negatives. Additionally, the presence of the patient's arms in the background can interfere with accurate segmentation by potentially including them in the segmented regions.

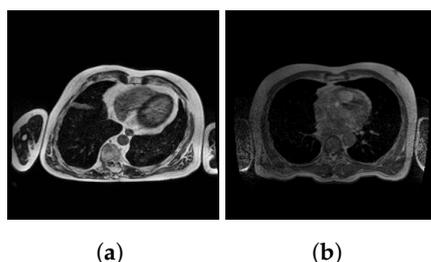


Figure 2. Problematic slices. (a): Slice with the presence of patients' arms. (b): A slice with low luminosity.

To address the content and image quality issues, we propose some preprocessing steps that are summarized in Figure 3.

The initial and crucial step in the preprocessing workflow involves the precise elimination of extraneous anatomical structures beyond the designated region of interest. This strategic elimination not only improves the visual output but also significantly enhances the overall accuracy of the model. YOLO version 5 [27] was employed for this purpose. The selection of YOLO version 5 was driven by its advanced capabilities in accurately detecting bounding boxes and efficiently capturing the regions of interest. In our workflow, it served as a reliable means to eliminate extraneous anatomical structures. For training the network, we used 200 manually annotated images with bounding boxes outlining the regions of interest (ROI). YOLO version 5's streamlined architecture ensures swift processing, facilitating real-time detection and preprocessing of MR images. Moreover, the model benefits from robust transfer learning techniques, being pretrained on extensive natural image datasets, which enhances its adaptability to diverse domains, including medical images. This transfer learning leads to exceptional organ detection performance. Achieving accuracy in delineating organs and removing background and irrelevant structures, like arms, is accomplished through training with image sizes of 640×640 , a batch size of 16, and YOLO version 5 small weights for 20 epochs. This process not only contributes to a cleaner and more visually appealing output but, more importantly, substantially enhances the overall accuracy of the model by ensuring that the segmentation focuses exclusively on the critical

structures relevant to radiation therapy planning. This improvement is fundamental in minimizing the risk of radiation toxicity and optimizing treatment planning precision.

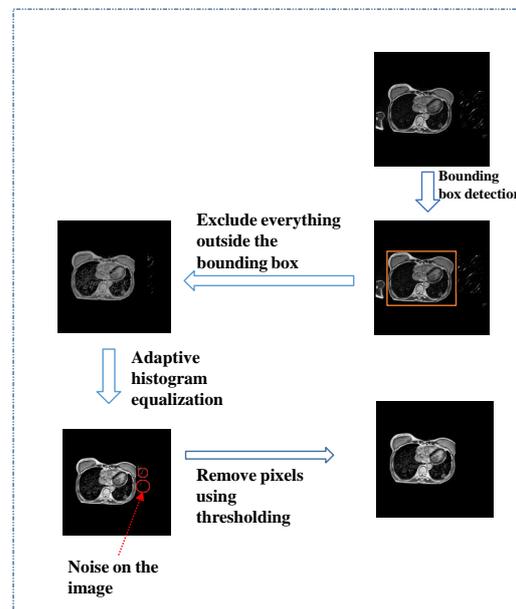


Figure 3. Overview of the preprocessing steps.

To address the issue of low luminosity in the remaining images, adaptive histogram equalization (AHE) was employed [28]. AHE is a technique used to enhance the contrast and visibility of details in images with low luminosity. By applying AHE to the images, we can ensure that subtle structures and features within the images become more pronounced and distinguishable. This, in turn, leads to better image quality, which is crucial for accurate organ segmentation. Improved image quality resulting from AHE can enhance the model's ability to detect and delineate organ boundaries, making the segmentation process more precise and reliable. Figure 4 illustrates the impact of adaptive histogram equalization on image quality.

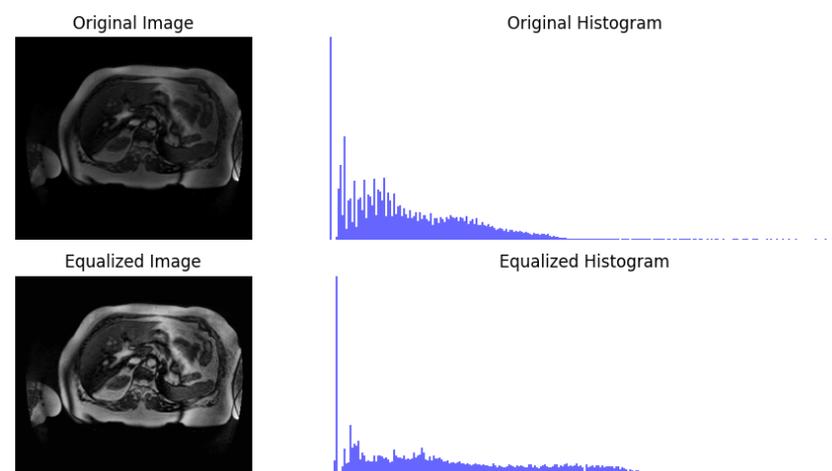


Figure 4. Impact of adaptive histogram equalization.

Some remaining noise was removed through thresholding, which eliminates pixels near zero intensity within the bounding boxes.

2.3. Efficient-UNet Model for Multiorgan Segmentation

The UNet architecture, known for its effectiveness in medical image segmentation tasks, is employed to accurately segment the five OARs. The UNet architecture comprises an encoder and a decoder, where the encoder captures high-level features from the input images and the decoder generates the segmentation masks. In our model, we employed the EfficientNet B4 model as an encoder, which is recognized for its efficiency and strong performance in various computer vision tasks. By using the pretrained weights of EfficientNet B4, the learned representations can be leveraged to enhance the segmentation accuracy. A detailed explanation of the developed architecture will be provided in this section. An overview of the network scheme is presented in Figure 5.

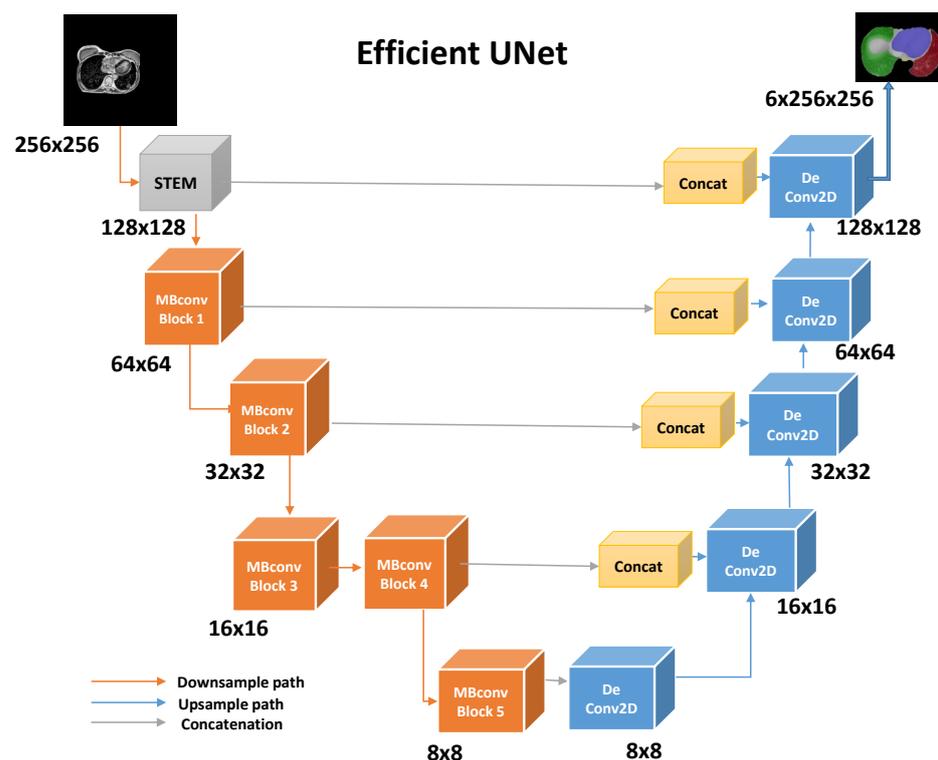


Figure 5. Overview of the used Efficient UNet network scheme. MBConvblocks: Mobile Inverted Residual Bottleneck blocks.

2.3.1. Encoder Blocks

The architecture consists of six blocks, known as MBConvBlocks (Mobile Inverted Residual Bottleneck blocks), each responsible for extracting and encoding features from the input data. These blocks progressively capture more complex and abstract information as we move through the encoder. The choice of different filter sizes in each encoder block allows capture of features at different spatial scales and capture of a range of details. This diversity in filter sizes enables the model to extract relevant information across multiple resolutions. The MBConvBlock can enhance automatic image segmentation for MRI images by offering several advantages. First, the depthwise convolution operation helps capture spatial details within the MRI images while preserving input channel information, making it particularly useful for identifying fine anatomical structures. Second, the pointwise convolution operation, or 1×1 convolution, enables effective channel-wise combination, facilitating dimensionality reduction or expansion. This is vital in MRI image segmentation because it allows the model to focus on relevant features and reduce computational complexity. Finally, the expansion operation increases the model's representational capacity, allowing it to learn more complex features, which can be particularly beneficial for capturing nuanced organ boundaries. Overall, the MBConvBlock's design, which balances

accuracy and computational efficiency, can improve the segmentation process for images by efficiently capturing spatial information, reducing the parameter count, and enabling a more compact yet effective model representation. This makes it well-suited for accurate and efficient automatic organ segmentation in MRI images.

2.3.2. Decoder

The decoder section of the Efficient-UNet architecture comprises a series of decoding blocks. Each decoding block incorporates convolutional layers, residual blocks, and activation functions, such as ReLU. These blocks play a crucial role in recovering spatial information and refining the segmentation masks generated by the encoder. To obtain the final segmentation maps, a segmentation head that consists of a 1×1 convolution layer is used. This configuration allows the network to generate probability distributions for each pixel across the five organ classes.

2.4. Training Strategies

To segment the five organs within thoracic structures, a 2.5D strategy was implemented, combining the strengths of 2D and 3D methods. This strategy processes consecutive 2D slices, capturing context efficiently. It improves organ segmentation accuracy by considering spatial relationships and contextual cues from neighboring slices. Notably, it overcomes 3D method limitations, like high computational costs and memory requirements. Instead of processing the entire 3D volume, the focus is placed on a limited number of adjacent slices. The strategy introduces a novel and refined approach in the context of radiation therapy planning using the MR-Linac system for multiorgan image segmentation. Our proposed 2.5D methodology involves feeding sets of consecutive slices into the EfficientNet encoder. This design choice allows the model to harness the advantages of both 2D and limited 3D contextual information. The novelty lies in the careful integration of spatial context from neighboring slices, leveraging the benefits of consecutive slices to enhance the model's understanding of three-dimensional structures. The use of EfficientNet as the backbone further contributes to the proposed model's efficiency and parameter optimization. This approach effectively addresses the limitations of 2D deep learning approaches, particularly in capturing volumetric information and spatial context. By overcoming these inadequacies, the proposed 2.5D approach aims to provide a more comprehensive understanding of anatomical structures, crucial for precise segmentation in the MR-Linac system. The incorporation of contextual information from adjacent slices bridges the gap between 2D and fully 3D approaches, offering an effective solution for tumor targeting and minimizing radiation toxicity during external radiation therapy in complex medical imaging scenarios. Through extensive experimentation, we determined that using precisely five adjacent slices strikes an optimal balance between precision and computational efficiency (Figure 6). Each set of five consecutive slices is treated as a single multi-slice input during training and inference. The model processes these slices collectively, leveraging the information from neighboring slices. Specifically, during training, our proposed Efficient-UNet is trained on batches containing these 2.5D input sets. The model learns to capture relevant features and contextual information across consecutive slices. During inference, when a single slice needs to be segmented, the model takes five adjacent slices as input and produces a segmentation map for the central slice. This strategy allows the model to incorporate spatial context from neighboring slices, which is particularly beneficial for tasks where volumetric information is crucial.

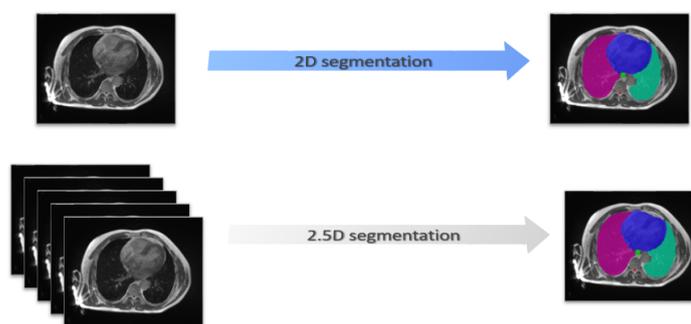


Figure 6. Training strategies. Difference between 2D and 2.5D strategies.

2.5. Evaluation Metrics

In the context of multiorgan segmentation, five commonly used metrics were employed to assess the segmentation performance. The chosen metrics included DSC, IoU, HD, and two metrics comparing the volumes obtained with the ground truth and with the automatic segmentation: the mean of the differences using a Bland–Altman study and the correlation coefficient.

DSC is a measure of the similarity between two sets. IoU is another measure of overlap between automatically segmented regions and reference regions. HD is a metric that quantifies the maximum distance between points in one set to the nearest point in the other set. It provides a measure of the dissimilarity between two sets of points.

The Bland–Altman method used for the evaluation of the organ volume calculation is a statistical technique used to assess the agreement between two sets of measurements obtained with the predicted and ground truth segmentations. The coefficient of correlation is a metric that quantifies the correlation between two measures—in our case, the volumes of the segmented regions and the reference regions. It provides a measure of how well the volumes are preserved in the segmentation process.

Higher DSC and IoU values indicate better results, while lower HD values also suggest better performance. The Bland–Altman method and the correlation coefficient provide insights into the agreement between the volumes of the segmented regions and the reference regions in multiorgan segmentation. All these metrics are computed in 3D, where the ground truths and masks are combined for each patient to create a 3D image. This approach provides a genuine assessment of the error between the masks and the ground truths. The metrics presented in Section 3 are all calculated in this manner to facilitate a comprehensive evaluation of the entire organ segmentation process.

2.6. Implementation

After the completion of the dataset preprocessing, the dataset is divided into three sets: a training set consisting of 6221 images (from 74 patients), a validation set with 160 images (from 2 patients), and a testing set with 464 images (from 5 patients). The selection of five patients for the test set was made through a rigorous randomization process. The limited number of patients in the validation and in the test set was due to the limited dataset in terms of patients and we decided to reinforce the training set. The subsequent step involved the selection of the model and the training strategy (2D vs. 2.5D). After a series of experiments, the chosen loss function was the focal loss. While other loss functions have proven effective for multiclass segmentation, the focal loss has gained attention for its ability to handle class imbalance issues. The focal loss assigns higher weights to misclassified samples, thereby emphasizing the training on challenging examples. This property makes it particularly useful in scenarios where certain classes are underrepresented. The focal loss can be defined as follows:

$$\text{Focal Loss} = - \sum_i^N (1 - p_i)^\gamma g_i \log(p_i)$$

where p_i represents the predicted mask of class i , g_i represents the ground truth mask of class i , and γ is a tunable focusing parameter that modulates the weight assigned to each sample.

The training process was performed on a T4 GPU, featuring 40 streaming multiprocessors (SMs) with a 6 MB L2 cache shared by all SMs, and 16 GB high-bandwidth memory (GDDR6) connected to the processor, enabling accelerated computations and reducing training times. The model parameters, including the batch size (16), the loss function (focal loss), the number of epochs (40), and the learning rate (0.0002), were carefully chosen based on their performance during a series of experiments. The Adam optimization algorithm, as proposed by Kingma and Ba [29], dynamically adjusts the learning rate for each model parameter based on gradients, combining the advantages of AdaGrad and RMSProp. The chosen learning rate for the Adam algorithm was set at 0.0002. Additionally, the training process incorporates a cosine annealing learning rate scheduler, which gradually adjusts the learning rate over epochs. The synergy between Adam and the cosine annealing scheduler contributes to effective convergence and exploration of the model's parameter space, ensuring optimal performance in multiorgan segmentation tasks.

To improve efficiency and to reduce memory usage, mixed precision training was employed, performing certain operations with lower precision while maintaining higher precision for others. The cosine annealing LR scheduler adjusts the learning rate gradually, aiding the model in exploring different loss landscapes. The exponential moving average (EMA) was used to stabilize and improve the model's performance by smoothing out parameter fluctuations. Moreover, by comparing the 2D and 2.5D strategies in our model, the impact of contextual information from neighboring slices could be evaluated.

3. Experiments and Results

3.1. Experiment 1: Model Selection

To evaluate the effectiveness of the proposed model, a quantitative assessment was conducted by comparing it with several existing state-of-the-art approaches, like UNet and the Residual Attention UNet, using the DSC, IoU, and HD mean values for the five organs on the test set.

U-Net is a convolutional neural network architecture widely used for image segmentation tasks. Introduced by Ronneberger et al. in 2015 [15], U-Net features a symmetric encoder–decoder structure. The encoder captures the context and extracts the features, while the decoder reconstructs the segmentation map. Skip connections between corresponding layers in the encoder and decoder facilitate the integration of high-resolution information during segmentation, making it effective for tasks like medical image segmentation.

The ResAtt U-Net, or Residual Attention U-Net [16], is an extension of the U-Net architecture that incorporates residual blocks and attention mechanisms. The residual blocks help to mitigate the vanishing gradient problem, aiding in the training of deeper networks. Attention mechanisms enhance the model's focus on relevant regions during feature extraction, contributing to improved performance in capturing the intricate details. The combination of residual connections and attention mechanisms aims to boost the network's capability in handling complex and nuanced segmentation tasks.

The number of parameters, the training time, and the inference time are provided for the three models. Table 1 showcases a comparison between the three models. We observed that the Residual Attention UNet and the original UNet models showed poor performance, leading to lower DSC and IoU metrics. Despite having lower HD values, the Residual Attention UNet yielded modest results visually. In contrast, the Efficient-UNet exhibited good performance due to its efficiency, shortened training and inference durations, and a reduced number of parameters. The Efficient-UNet emerged as a superior

choice, integrating the advantages of a pretrained encoder with efficient training and inference processes.

Table 1. Comparison of UNet, Residual Attention UNet, and Efficient-UNet on the test set based on DSC, IoU, HD, number of parameters, training time (in hours), and inference time (in seconds). The best results are highlighted in bold. All metrics are computed in 3D.

Model	DSC	IoU	HD (mm)	# Parameters	Training Time	Inference Time
Efficient-UNet	0.804 ± 0.058	0.711 ± 0.062	25.663 ± 18.724	20,225,550	4 h	3.71 s
UNet	0.761 ± 0.078	0.657 ± 0.086	35.915 ± 13.632	30,106,806	3.9 h	7.07 s
ResAtt UNet	0.677 ± 0.105	0.561 ± 0.115	21.536 ± 8.803	34,877,746	13.9 h	5.37 s

Figure 7 displays the results of the different tested models on the test images. Efficient-UNet exhibits superior segmentation performance compared to the other models. In the first row, Efficient UNet excels in segmenting the heart, the spinal cord, and the esophagus when compared to UNet and Residual Attention UNet. Moving to the second row, we focus on the segmentation of the left lung for another exam. Here, UNet misclassifies a portion of the left lung as the right lung, while both Efficient UNet and Residual Attention UNet demonstrate accurate segmentation. However, Residual Attention UNet falls short in completing the segmentation of the right lung. In the last row, representing the segmentation on a distal slide from another exam, Efficient UNet tends to over-segment the left lung. In contrast, UNet incorrectly segments a part above the left lung as the right lung. Residual Attention UNet fails to segment any organ.

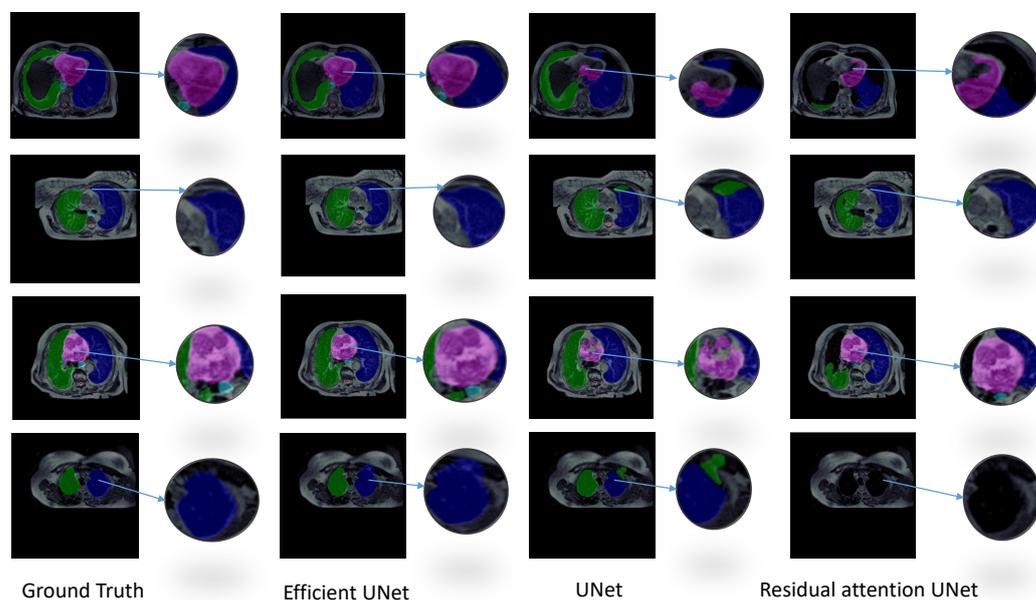


Figure 7. Visual comparison of the results from Efficient-UNet, Residual Attention UNet, and UNet models on the test dataset alongside the corresponding ground truth using the 2D strategy. The heart is in pink, the spinal cord is in red, the left lung is in blue, the right lung is in green, and the esophagus is in light blue.

3.2. Experiment 2: Training Strategy Selection

In Section 3.1, a high HD value was observed for Efficient-UNet when using 2D images. This issue can be mitigated by implementing a 2.5D strategy. The high HD values in 2D may be attributed to the consideration of individual slices, particularly affecting certain organs, like the left lung, due to their sensitivity to noise. However, 2.5D strategies, which

incorporate contextual information from neighboring slices, can help alleviate the impact of noise and enhance segmentation performance, ultimately resulting in improved HD values.

Table 2 summarizes the metrics employed to evaluate both strategies for all organs within the test dataset, while Figure 8 presents the visual results of the 2D Efficient-UNet and the 2.5D Efficient-UNet across several patients in the same test dataset. A notable observation is that the 2.5D strategy consistently exhibited superior performance across the majority of metrics in comparison with the 2D strategies, with a notable advantage in terms of the average Hausdorff distance (HD), signifying enhanced boundary accuracy. Furthermore, the segmentation results visually reinforce this pattern, underscoring the effectiveness of the 2.5D approaches.

Table 2. Comparison of 2D and 2.5D strategies utilizing the Efficient-UNet model. All metrics are calculated in a 3D context. The comparison of automatically and manually obtained volumes (computed in pixels) was performed using a Bland–Altman plot (fourth column), as well as the correlation coefficient (fifth column) for organ volumes. The best results are highlighted in bold.

Strategy	DSC	IoU	HD (mm)	Volume (Bland–Altman)	Volume (Correlation)
2.5D	0.820 ± 0.041	0.725 ± 0.052	10.353 ± 4.974	153.943 ± 50.149	0.734 ± 0.090
2D	0.804 ± 0.058	0.711 ± 0.062	25.663 ± 18.724	151.915 ± 55.747	0.726 ± 0.073

Analyzing organ-specific performance as presented in Table 3, it is evident that the 2.5D strategy consistently yielded lower HD values for the left lung, the right lung, and the spinal cord in contrast to the 2D strategy. This suggests that the 2.5D strategy exhibits effective boundary capture with reduced deviations from the ground truth. While the 2D strategy achieved higher IoU and DSC scores for the right lung and the heart compared to the 2.5D strategy, it still resulted in potential boundary inaccuracies. In terms of anatomical metrics, specifically the volume measurements, the 2.5D strategy demonstrated an overall higher correlation when compared to the 2D strategy, albeit with a minimal difference. However, when considering the Bland–Altman analysis, the 2D strategy appeared to be more favorable. Nevertheless, upon examining the organ-specific values, it is evident that the volume metrics closely align across the majority of organs, exhibiting minimal variation. The sole exception is observed in the case of the spinal cord, where all the metrics suggest that the 2.5D strategy outperformed the 2D strategy. Overall, there is some indication that the 2.5D strategy may exhibit improved organ segmentation performance, particularly in boundary delineation, when compared to the 2D strategy, though we should acknowledge the possibility of certain organ-specific variations in the performance metrics.

Table 3. Comparison of 2D and 2.5D strategies for organ segmentation using the Efficient-UNet. All metrics are computed in 3D. The comparison of the volumes (computed in pixels) obtained automatically and manually was performed by applying the Bland–Altman plot (fourth column), as well as the correlation coefficient (fifth column) for the organ volumes. The best results are highlighted in bold.

Organ	Strategy	Metrics				
		IoU	DSC	HD (mm)	Volume	
					Bland–Altman	Correlation
Left Lung	2D	0.895 ± 0.033	0.944 ± 0.018	14.492 ± 13.280	225.900 ± 101.430	0.978 ± 0.009
	2.5D	0.895 ± 0.031	0.944 ± 0.017	9.443 ± 1.953	216.580 ± 78.003	0.976 ± 0.008
Right Lung	2D	0.912 ± 0.011	0.953 ± 0.006	70.157 ± 43.920	209.380 ± 47.580	0.994 ± 0.001
	2.5D	0.904 ± 0.017	0.949 ± 0.009	9.739 ± 2.853	236.69 ± 69.311	0.992 ± 0.004

Table 3. Cont.

Organ	Strategy	Metrics				
		IoU	DSC	HD (mm)	Volume	
					Bland-Altman	Correlation
Heart	2D	0.860 ± 0.018	0.924 ± 0.010	19.403 ± 19.860	235.939 ± 108.410	0.983 ± 0.009
	2.5D	0.856 ± 0.037	0.922 ± 0.021	11.270 ± 8.454	237.638 ± 84.649	0.975 ± 0.018
Esophagus	2D	0.356 ± 0.116	0.513 ± 0.134	10.772 ± 5.280	74.510 ± 16.860	0.454 ± 0.205
	2.5D	0.384 ± 0.078	0.551 ± 0.080	11.710 ± 5.689	66.017 ± 13.650	0.422 ± 0.254
Spinal cord	2D	0.534 ± 0.133	0.685 ± 0.121	13.490 ± 11.250	13.835 ± 4.430	0.223 ± 0.138
	2.5D	0.585 ± 0.096	0.733 ± 0.077	9.600 ± 5.920	12.770 ± 5.130	0.304 ± 0.164

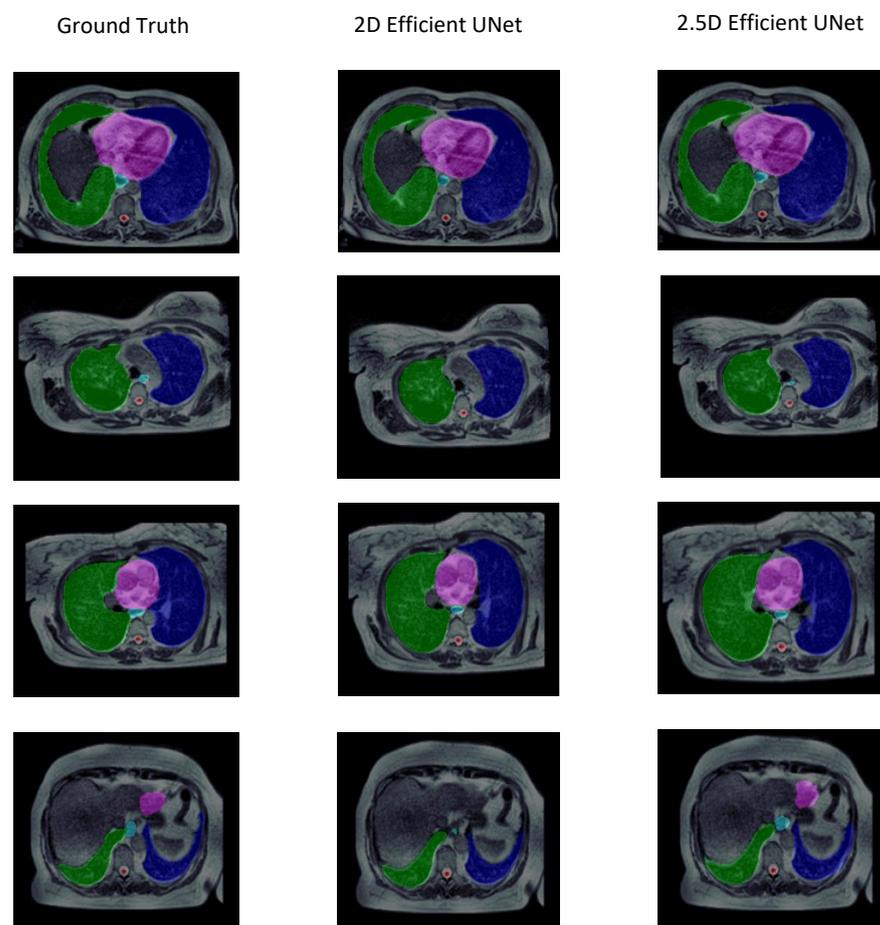


Figure 8. Comparison of training strategies using the Efficient UNet architecture for different patients in the test set. The first column displays the ground truth image, the second column shows the 2D strategy, and the third column presents the 2.5D strategy.

4. Discussion

In our study on multiorgan segmentation using MR-Linac-acquired images in the thoracic region, we focused on critical organs, like the heart, the right lung, the left lung, the spinal cord, and the esophagus, driven by clinical requirements. Our dataset, consisting of 81 MRI studies and radiation therapy structure sets, formed a solid foundation for training our Efficient-UNet model. We eliminated extraneous anatomical structures, applied image enhancement techniques, and utilized an EfficientNet B4 encoder for our automatic segmentation task. The complexity in the anatomy of the organs reflects different

performances of our method. Firstly very good segmentation accuracy is observed for certain organs, especially for the lung and the heart, as demonstrated by the evaluation metrics. On the other hand, more questionable results are obtained for the spinal cord and the esophagus. A first reason could relate to the size of the organs. The IoU and the DSC are sensitive to the number of pixels and a small error in small anatomic structure segmentation can have a big impact on these metrics. In particular, the spinal cord is defined with a few pixels in each image plane. Concerning the esophagus, the manual segmentation of this organ could be difficult due to its complexity and may generate high inter-observer variation. Imprecise manual segmentation could result in compromised training of our network. Future work will seek to include a consensus of experts for the segmentation of this organ in order to create a robust ground truth.

Our 2.5D training strategy, combining 2D and 3D methods, improved the boundary segmentation accuracy and the overall segmentation performance. Overall, the Efficient-UNet model and the 2.5D strategy appear to be the better option (Figure 9), even if segmentation with a 2D strategy sometimes provides better metrics results (such as the DSC for the lung). However, in cases where there is a discrepancy between the DSC and the HD, the better results for the DSC with a 2D strategy is compensated by an important decrease in the HD with a 2.5 strategy (as for the segmentation of the right lung). Sometimes a difference between the results obtained with the 2D and 2.5D strategies is not clearly evident. However, the benefit of the 2.5D approach may be more apparent in challenging cases, such as those with complex anatomical variations or pathologies. Concerning the calculation of the volume of each organ, broadly speaking, the results are similar between the two strategies. We note, in particular, good agreement for the spinal cord that counterbalances the low DSC and which demonstrates the impact of organ size in the calculation of the DSC.

In the literature, we did not identify any articles addressing the multiorgan segmentation of OARs in the thoracic region, particularly when focusing on the automatic segmentation of the left lung, right lung, heart, spinal cord, and esophagus using MR images or MR images acquired with an MR Linac. However, several studies have been conducted using CT scans for thoracic organs, utilizing either private or publicly available datasets. Zhang et al. [30] introduced a deep learning model utilizing DenseNet and FCN on a private CT dataset. In comparison to their approach, our heart segmentation achieved a higher performance (DSC: 0.924 vs. 0.86). While the results for the lungs are approximately similar, Zhang et al. achieved superior outcomes for the spinal cord and esophagus. The differences in results for the latter two organs can be attributed to the type of modality. Indeed, the esophagus is more distinctly visible in CT scans compared to MR images. Gali et al. [31] proposed a segmentation procedure for thoracic organs on CT images using Dilated UNet. The authors utilized a dilation kernel to achieve the desired resolution in feature maps at any layer, applying it to the network through an integrated model during training. In comparison to their approach, our segmentation of the heart (DSC: 0.924 vs. 0.8597) and the esophagus (DSC: 0.551 vs. 0.4694) exhibited better performance. Larsson et al. [32] proposed an abdominal organ segmentation approach using a regional convolutional neural network. This approach uses a localization algorithm for finding the region of interest. As a second step, a convolutional neural network is applied performing voxelwise classification. These authors also investigated esophagus segmentation, and there was very similar segmentation performance (DSC: 0.585 vs. 0.588). Cao et al. [33] introduced a deep learning network named the cascaded SE-ResUnet, designed for the automatic segmentation of thoracic organs. Their approach involves a two-step process: initially a coarse segmentation network is used to identify the regions of interest, followed by a fine segmentation network to refine the results organ-by-organ. The final segmentation results are obtained through ensembling different configured models. In comparison to our approach, these authors demonstrated superior performance across all organs by employing a complex network and a weighted loss function tailored for 3D images. The differences can also be explained by the use of another image acquisition technique. There are several limitations and potential avenues for improvement that may be identified for our approach.

Notably, we could enhance the complexity of our network by implementing a coarse-to-fine architecture, combining multiple models. Additionally, the current study's patient sample size was relatively small, and increasing it could provide more robust insights. To enhance the evaluation step, cross-validation techniques could be employed to provide a more realistic assessment of our model's performance. Exploring alternative methodologies, such as a 3D strategy, could also be beneficial. Moreover, expanding the scope to include additional thoracic organs could further enhance the comprehensiveness of our segmentation approach. Furthermore, exploring higher field strengths in the MRI could increase the signal-to-noise ratio (SNR). This enhancement is anticipated to refine the segmentation results by amplifying the clarity of the anatomical structures and reducing the background noise. Investigating the impact of increased SNR on the segmentation accuracy could provide valuable insights for optimizing the MRI imaging quality in future research endeavors. Finally, the Efficient UNet model successfully predicted the segmentation of five organs for a single patient in less than 4 s. This stands in stark contrast to the manual approach, which may require more than 30 min (potentially up to 50 min for a patient with 144 slices). These time-saving effects are particularly pertinent in the context of online adaptive radiotherapy, a critical process, especially for lung tumor cases on MR-Linacs. The rapid and efficient nature of the model's predictions could significantly enhance the speed and effectiveness of this treatment modality, addressing a crucial time-related aspect, as highlighted in a study by Gungor et al. [34]. The integration of DL-based automatic segmentation in online adaptive radiotherapy has the potential to produce notable improvements in the overall treatment process.

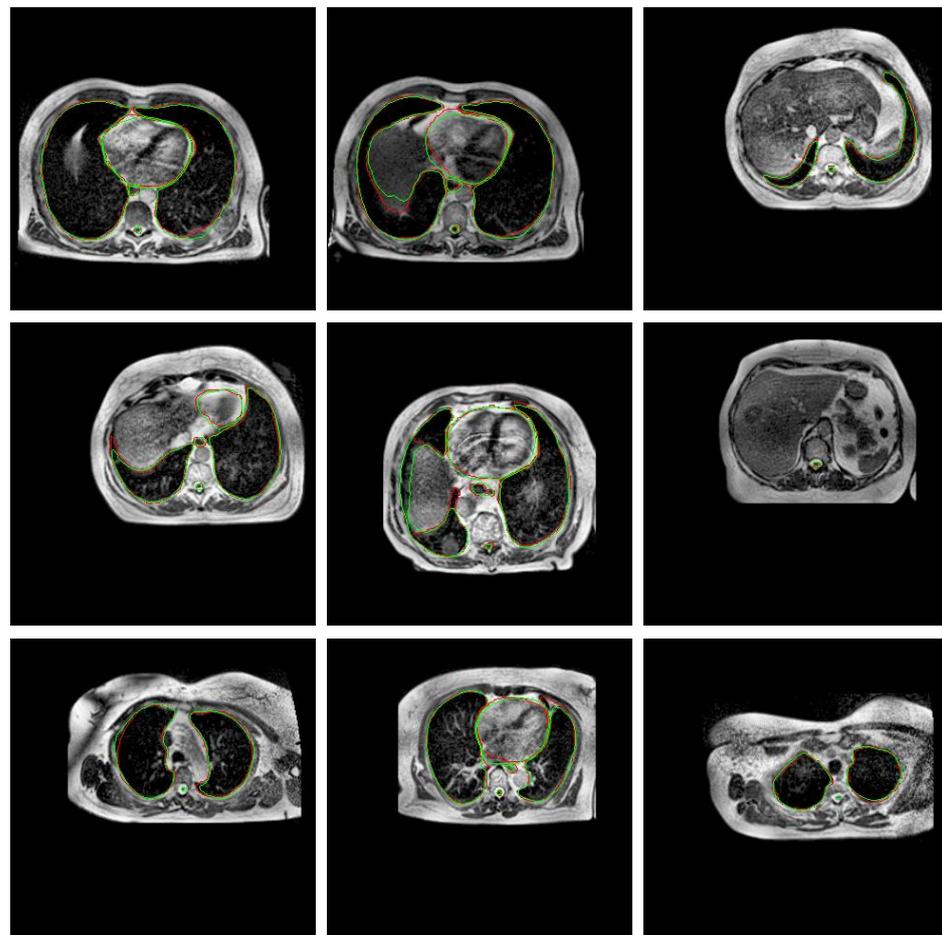


Figure 9. Examples of segmentation using a 2.5D Efficient-UNet model. The ground truth is in red color and the predicted results are in green color.

5. Conclusions

This study introduces an approach for thoracic multiorgan segmentation on a 0.35 T MR-Linac radiation therapy system, utilizing the Efficient-UNet network, known for its balance between computational efficiency and accuracy. By incorporating YOLO and AHE preprocessing techniques, we significantly enhance the image quality and focus only on the thoracic area. Moreover, the integration of a 2.5D strategy in our approach leads to robust performance for OAR segmentation. Our work contributes to the field of thoracic multiorgan segmentation, focusing on MR images from MR-Linac. These findings have practical implications for improving treatment planning precision and patient outcomes, aligning with the evolution of medical imaging and innovative strategies for organ segmentation tasks.

Author Contributions: Conceptualization, Y.M., I.B. and A.L.; methodology, M.C., Y.M., I.B. and A.L.; software, M.C. and Y.M.; data curation, I.B.; writing—original draft preparation, M.C.; writing—review and editing, Y.M., I.B. and A.L.; supervision, Y.M., I.B. and A.L.; project administration, Y.M., I.B. and A.L.; funding acquisition, A.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research presented herein was financially supported by Dijon Metropole through contract DEVECO DM2023-029-20230301.

Data Availability Statement: Ethical reasons prohibit the public availability of the dataset. However, it can be confidentially communicated to the reviewers and the Journal's Editor if necessary.

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

References

1. Winkel, D.; Bol, G.H.; Kroon, P.S.; van Asselen, B.; Hackett, S.S.; Werensteijn-Honingh, A.M.; Intven, M.P.; Eppinga, W.S.; Tijssen, R.H.; Kerkmeijer, L.G.; et al. Adaptive radiotherapy: The Elekta Unity MR-linac concept. *Clin. Transl. Radiat. Oncol.* **2019**, *18*, 54–59. [[CrossRef](#)]
2. Klüter, S. Technical design and concept of a 0.35 T MR-Linac. *Clin. Transl. Radiat. Oncol.* **2019**, *18*, 98–101. [[CrossRef](#)]
3. Tong, N.; Cao, M.; Sheng, K. Shape constrained fully convolutional DenseNet with adversarial training for multi-organ segmentation on head and neck low field MR images. *Int. J. Radiat. Oncol. Biol. Phys.* **2019**, *105*, S93. [[CrossRef](#)]
4. Liu, Y.; Lei, Y.; Fu, Y.; Wang, T.; Zhou, J.; Jiang, X.; McDonald, M.; Beitler, J.J.; Curran, W.J.; Liu, T.; et al. Head and neck multi-organ auto-segmentation on CT images aided by synthetic MRI. *Med. Phys.* **2020**, *47*, 4294–4302. [[CrossRef](#)] [[PubMed](#)]
5. Chen, Y.; Ruan, D.; Xiao, J.; Wang, L.; Sun, B.; Saouaf, R.; Yang, W.; Li, D.; Fan, Z. Fully automated multiorgan segmentation in abdominal magnetic resonance imaging with deep neural networks. *Med. Phys.* **2020**, *47*, 4971–4982. [[CrossRef](#)] [[PubMed](#)]
6. Sharp, G.; Fritscher, K.D.; Pekar, V.; Peroni, M.; Shusharina, N.; Veeraraghavan, H.; Yang, J. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med. Phys.* **2014**, *41*, 050902. [[CrossRef](#)]
7. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)] [[PubMed](#)]
8. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)]
9. Trullo, R.; Petitjean, C.; Ruan, S.; Dubray, B.; Nie, D.; Shen, D. Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields. In Proceedings of the 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 1003–1006.
10. Im, J.H.; Lee, I.J.; Choi, Y.; Sung, J.; Ha, J.S.; Lee, H. Impact of Denoising on Deep-Learning-Based Automatic Segmentation Framework for Breast Cancer Radiotherapy Planning. *Cancers* **2022**, *14*, 3581. [[CrossRef](#)]
11. Khalil, M.I.; Tehsin, S.; Humayun, M.; Jhanjhi, N.; AlZain, M.A. Multi-Scale Network for Thoracic Organs Segmentation. *Comput. Mater. Contin.* **2022**, *70*, 3251–3265.
12. Mahmood, H.; Islam, S.M.S.; Hill, J.; Tay, G. Rapid segmentation of thoracic organs using u-net architecture. In Proceedings of the 2021 Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 29 November–1 December 2021; pp. 1–6.
13. Simpson, A.L.; Antonelli, M.; Bakas, S.; Bilello, M.; Farahani, K.; Van Ginneken, B.; Kopp-Schneider, A.; Landman, B.A.; Litjens, G.; Menze, B.; et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv* **2019**, arXiv:1902.09063.

14. Lambert, Z.; Petitjean, C.; Dubray, B.; Kuan, S. Segthor: Segmentation of thoracic organs at risk in ct images. In Proceedings of the 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France, 9–12 November 2020; pp. 1–6.
15. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
16. Ni, Z.L.; Bian, G.B.; Zhou, X.H.; Hou, Z.G.; Xie, X.L.; Wang, C.; Zhou, Y.J.; Li, R.Q.; Li, Z. Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments. In Proceedings of the International Conference on Neural Information Processing, Sydney, NSW, Australia, 12–15 December 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 139–149.
17. 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.
18. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
19. Baheti, B.; Innani, S.; Gajre, S.; Talbar, S. Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 358–359.
20. Trullo, R.; Petitjean, C.; Nie, D.; Shen, D.; Ruan, S. Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Proceedings of the Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, 14 September 2017*; Proceedings 3; Springer: Berlin/Heidelberg, Germany, 2017; pp. 21–29.
21. Han, L.; Chen, Y.; Li, J.; Zhong, B.; Lei, Y.; Sun, M. Liver segmentation with 2.5 D perpendicular UNets. *Comput. Electr. Eng.* **2021**, *91*, 107118. [[CrossRef](#)]
22. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
23. Costa, L.d.F. Further generalizations of the Jaccard index. *arXiv* **2021**, arXiv:2110.09619.
24. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 850–863. [[CrossRef](#)]
25. Altman, D.G.; Bland, J.M. Measurement in medicine: The analysis of method comparison studies. *J. R. Stat. Soc. Ser. D Stat.* **1983**, *32*, 307–317. [[CrossRef](#)]
26. Mazonakis, M.; Damilakis, J.; Varveris, H.; Prassopoulos, P.; Gourtsoyannis, N. Image segmentation in treatment planning for prostate cancer using the region growing technique. *Br. J. Radiol.* **2001**, *74*, 243–249. [[CrossRef](#)]
27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
28. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [[CrossRef](#)]
29. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
30. Zhang, F.; Wang, Q.; Yang, A.; Lu, N.; Jiang, H.; Chen, D.; Yu, Y.; Wang, Y. Geometric and dosimetric evaluation of the automatic delineation of organs at risk (OARs) in non-small-cell lung cancer radiotherapy based on a modified DenseNet deep learning network. *Front. Oncol.* **2022**, *12*, 861857. [[CrossRef](#)] [[PubMed](#)]
31. Gali, M.S.K.; Garg, N.; Vasamsetti, S. Dilated U-Net based Segmentation of Organs at Risk in Thoracic CT Images. In Proceedings of the SegTHOR@ ISBI, Venice, Italia, 8–11 April 2019.
32. Larsson, M.; Zhang, Y.; Kahl, F. Robust abdominal organ segmentation using regional convolutional neural networks. *Appl. Soft Comput.* **2018**, *70*, 465–471. [[CrossRef](#)]
33. Cao, Z.; Yu, B.; Lei, B.; Ying, H.; Zhang, X.; Chen, D.Z.; Wu, J. Cascaded SE-ResUnet for segmentation of thoracic organs at risk. *Neurocomputing* **2021**, *453*, 357–368. [[CrossRef](#)]
34. Güngör, G.; Serbez, İ.; Temur, B.; Gür, G.; Kayalılar, N.; Mustafayev, T.Z.; Korkmaz, L.; Aydın, G.; Yapıcı, B.; Atalar, B.; et al. Time analysis of online adaptive magnetic resonance-guided radiation therapy workflow according to anatomical sites. *Pract. Radiat. Oncol.* **2021**, *11*, e11–e21. [[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.