Eff-Unet for Trachea Segmentation on CT Scans

Arthur Guilherme Santos Fernandes(✉), Geraldo Braz Junior, João Otávio Bandeira Diniz, Marcos Melo Ferreira, José Ribamar Durand Rodrigues Junior, Mackele Lourrane Jurema Da Silva, and Lucas Araújo Gonçalves

Applied Computing Group, Federal University of Maranhão, Av. Portugueses 1996, São Luís, MA, Brazil
{arthurgsf,geraldo,joao.bandeira}@nca.ufma.br,
{marcos.ferreira,durand.jose,mackele.ljs,araujo.lucas}@discente.ufma.br

Abstract. Organ at Risk segmentation has an important role in the meticulous planning of radiotherapy for cancer treatment. Its primary objective is to safeguard the surrounding healthy tissues while precisely directing radiation to target cancer cells. Currently, this task needs manual intervention by physicians, a process that can be time-consuming and susceptible to errors. Consequently, the integration of automatic segmentation methods offers the potential to accelerate the delineation of organs during radiotherapy planning. In this study, we applied Eff-Unet, a fully convolutional neural network model, and trained it to perform the semantic segmentation of trachea in computed tomography images. This approach yielded a noteworthy 78.9% dice score, underscoring its capability to enhance the efficiency and precision of organ segmentation during the radiotherapy planning process.

Keywords: Deep Learning · Semantic Segmentation · Radiotherapy · Fully Convolutional Neural Networks · Organs at Risk

1 Introduction

Radiation therapy represents a form of cancer treatment wherein substantial doses of intense radiation are directed towards malignant cells, destroying it or reducing the tumor's size [5]. The initial phase of radiation therapy planning revolves around pinpointing the precise location of the tumor and the adjacent organs. This crucial step is instrumental in mitigating the counter effects of

This work received the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001, National Council for Scientific and Technological Development (CNPq), Maranhão Research Support Foundation (FAPEMA) and Brazilian Company of Hospital Services (Ebserh) (Proc. 409593/2021-4).

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 139–150, 2024.
https://doi.org/10.1007/978-3-031-60665-6_10

ionizing radiation over healthy cells. The healthy organs that encircle the tumor are commonly referred to as "Organs at Risk" (OaR) [7,14]. The task of manually delineating these OaRs from images acquired through Computed Tomography (CT) falls under the purview of a medical practitioner.

CT is a diagnostic imaging technique that helps physicians to assess the condition of soft and hard tissues within the body. This method involves an X-ray machine that circumnavigates the patient while radiation detectors on the opposing end register radiation levels as the X-rays pass through the biological tissue and ultimately convert them into a digitized image [2,11]. The outcome is a three-dimensional image that is further dissected into two-dimensional slices for enhanced visualization [9]. For illustrative purposes, Fig. 1 showcases instances of CT cross-sections.

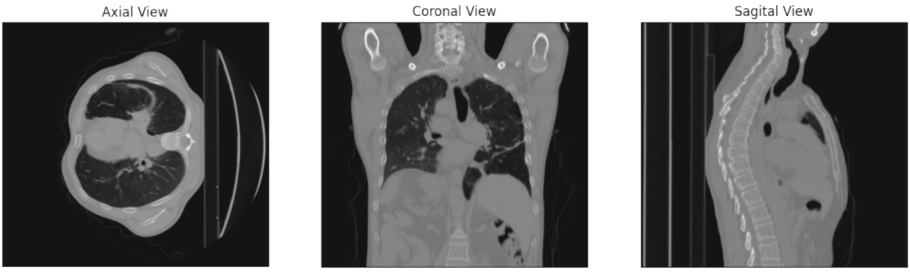

Fig. 1. Examples of 2D images obtained from a CT scan.

Accurate manual segmentation of Organs at Risk (OaR) holds significant importance, as it provides precise information regarding the proximity of healthy tissue to the tumor. This information allows specialists to precisely target the radiation beam exclusively at the cancerous tissue, effectively minimizing unwarranted X-ray exposure and decreasing the likelihood of adverse side effects from radiation therapy, which may encompass issues such as inflammation, fibrosis, ulceration, and in severe cases, organ failure [18]. However, this process relies on the physician's clinical experience and can be time-consuming and susceptible to errors. Small organs, such as the trachea, present unique challenges for segmentation due to their size, textural similarities with surrounding tissues, and limited detail. Consequently, recent research endeavors concentrated on developing automated methods for segmenting organs at risk (OaR).

Given the significance of Organ-at-Risk (OaR) segmentation in cancer treatment and the papers available on automated medical image processing, this research tackles the challenge of automating the segmentation of the trachea from CT exams by applying image preprocessing techniques alongside a convolutional neural network architecture known as Eff-Unet [4]. The primary contributions of this are: **(a)** A methodology for automated trachea segmentation utilizing Eff-Unet, **(b)** evaluating the performance of the Eff-Unet [4] neural network for the trachea segmentation task, **(c)** minimizing neural network resource

usage by applying the eff-Unet [4] architecture. The EfficientNet [17] is used as a feature extractor backbone, leading to a leaner model suitable for embedded or mobile systems.

2 Related Work

Automated segmentation facilitates the data processing time and guides clinicians by providing task-specific visualizations and measurements [3]. Semantic segmentation of Organs at Risk serves as an initial phase in radiotherapy planning, aiding the identification of task-relevant specific regions. Specifically about trachea region segmentation, certain studies have devised techniques for trachea segmentation by employing encoder/decoder fully convolutional networks or by integrating transformers within these network architectures.

In response to recent advancements in transformer architecture applied to image segmentation, [21] demonstrated that these techniques paired with U-Net [15] achieved relative success segmenting various Organs at Risk (OaR), including the trachea.

Other approaches harness the 3D characteristics of tomography to extract contextual information from CT scans, as seen in [19] and [20].

Alternately, some methods combined U-Net [15] with other convolutional network architectures for OaR segmentation. For instance, [8] employs a contextual pyramid fusion module that extracts features from various scales and merges them, while [22] utilizes Generative Adversarial Networks in conjunction with U-Net [15] to perform semantic segmentation of various organs in CT scans.

Efforts have been made to curate publicly available datasets for OaR segmentation. In the case of [13], the authors prepared a CT scan collection containing 60 volumes from different patients that were diagnosed with lung cancer. With this data, they trained a simplified U-NET [15] model. Out of these 60, 40 CT scans have been made publicly accessible, forming the SEGTHOR dataset.

Within the existing literature, many methodologies have sprung from the U-NET [15] architecture, frequently without placing a central emphasis on network efficiency concerning the number of trainable parameters. In pursuit of a leaner model, this study assesses the adoption of the Eff-Unet [4] fully convolutional network design. It incorporates the robust yet compact EfficientNet [17] as an encoder within the U-Net [15] encoder/decoder framework, aiming to achieve semantic segmentation of tracheal regions in computed tomography images with lower resource consumption, enabling the model to be embedded into larger medical systems.

3 Materials and Methods

This section introduces the adopted methodology, outlining its stages. Figure 2 provides an overview of the overall workflow in this study, commencing with image acquisition, progressing through preprocessing, Eff-Unet [4] model training, and concluding with the generation of predicted segmentation masks.

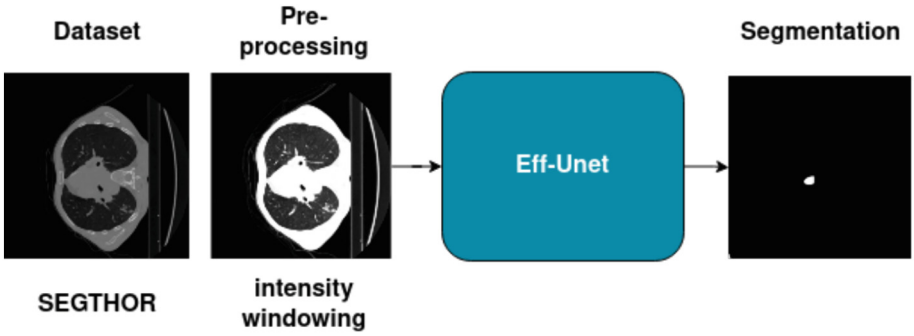

Fig. 2. Methodology Overview

3.1 Dataset

The images were obtained from the SEGTHOR [13] dataset, which contains 60 CT volumes obtained from lung cancer patients. From those 60 patients, only 40 were made public by the dataset owners. This yields 7420 2D cross-section images, where 1987 images depict the trachea, while the remaining 5433 do not, resulting in an imbalanced dataset. Slices without trachea do not participate during the training phase. The training, test and validation were separated patient-wise as follows: 30% (12) of the 40 patients were used for testing, while 62,5% (25) were separated for training, and the remaining 7,5% (3) patients were used for validation (Fig. 3).

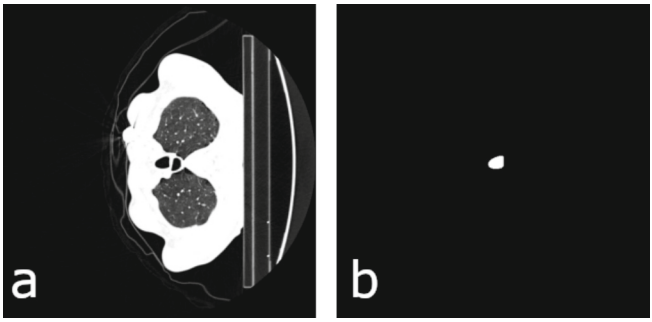

Fig. 3. a) Example of a slice that contains trachea; b) Physician-made delineation of the trachea.

Pre-processing: We implemented an intensity windowing technique to enhance contrast, restricting pixel values within a specific range and eliminating irrelevant information. Following this step, we conducted image normalization and resizing.

Intensity Windowing: In radiology, intensity windowing is a valuable tool to enhance CT imaging. It involves clipping the pixel values outside a given interval, say $[\alpha, \beta]$, to the closest boundary value while letting pixels within the interval remain unaffected. These intervals are known as "windows", and each one enhances different tissues. Figure 4 depicts the application of various windows to the same image.

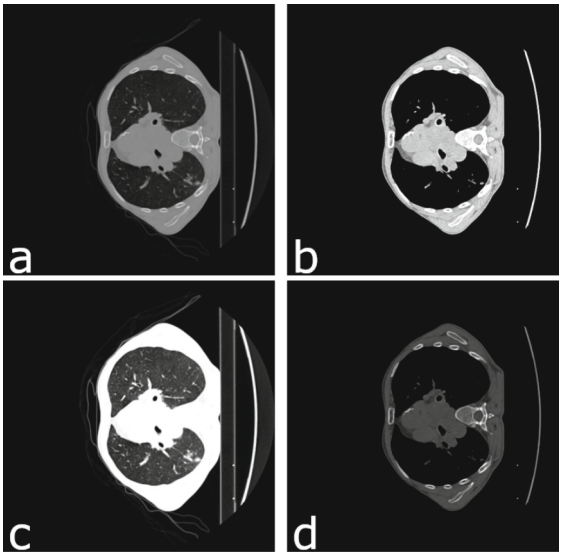

Fig. 4. Example of various windows clipping the same image intensities.

According to [6], an intensity interval of $[-1000, 60]$ is commonly used to visualize the trachea, as it effectively raises contrast between muscle and air. These boundaries are appropriate for human visualization but might not be optimal for convolutional network training, so a hyper-parameter optimization of the interval $[\alpha, \beta]$ is applied to find the best window for trachea segmentation.

We performed the hyper-parameter optimization using Bayesian optimization [1]. The optimization was executed for 100 trials, estimating the α and β boundaries for the intensity windowing preprocessing step. The best parameters were found in trial 48, obtaining **78.35%** dice score for the test with $\alpha = -890$ and $\beta = 66$. The learning rate for the ADAM was also optimized. The best parameter found was 0.0008. Figure 5 demonstrates the dice score behavior along the 100 trials.

Normalization: Following the application of intensity windowing, a subsequent step involves normalization. It entails subtracting the image mean from pixel values and dividing by the image's standard deviation. This process aims to create more uniform data, facilitating the model's learning process.

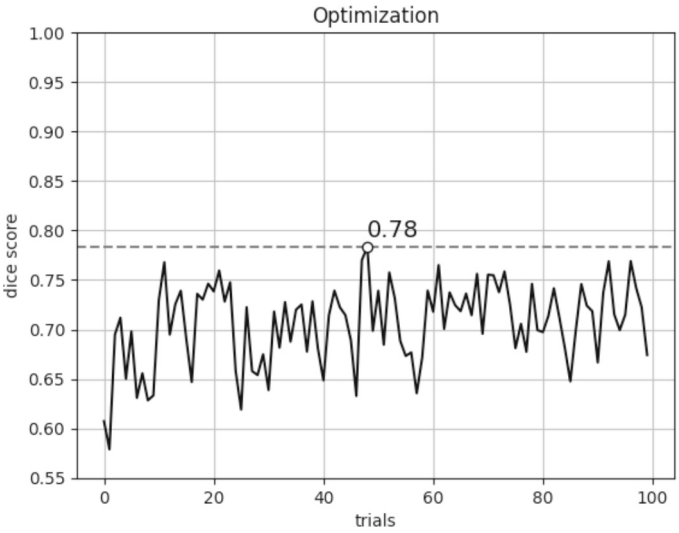

Fig. 5. The dice score curve along the optimization trials.

Resizing: After normalization, the images underwent resizing, transitioning from 512×512 dimensions to 256×256.

3.2 Eff-Unet

The Eff-Unet [4] architecture was applied to perform tracheal region segmentation. It is composed of an EfficientNet [17] feature extraction backbone acting as the encoder in the U-Net [15] framework, merging feature maps from different layers, and restoring them to the input size via transposed convolution, the result is a segmentation mask. Figure 6 depicts the model architectural overview.

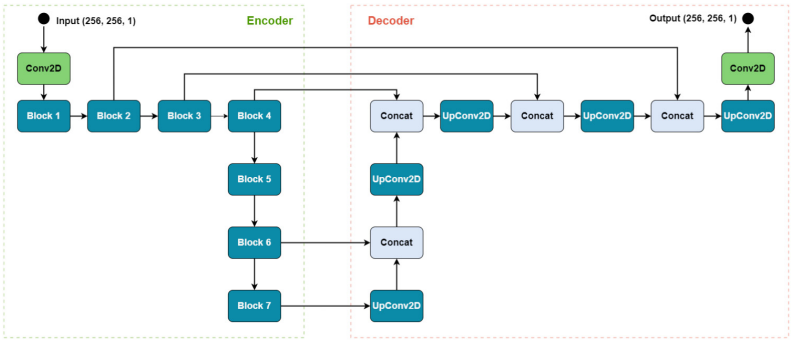

Fig. 6. Architectural overview of Eff-Unet.

EfficientNet: Engineered with efficiency resource usage as a main consideration, this network family comprises variants derived from a base model obtained via Neural Architecture Search (NAS) [23]. The foundational network is denoted as "b0", while additional models span from "b1" to "b7". These models vary in terms of input resolution, channel count, and layer configurations within each stage. Consequently, the numerical designation in the network's name corresponds to its size, with larger numbers indicating more extensive models in terms of parameters. In this work, the **EfficientNetB0** was chosen as the feature extractor for the U-Net. A concise overview of the core structure of the "b0" model is provided in Table 1.

Table 1. EfficientNet B0 architecture [17].

Stage	Operator	Resolution	#Channels	#Layers
1	Conv 3×3	224×224	32	1
2	MBConv1 3×3	112×112	16	1
3	MBConv6 3×3	112×112	24	2
4	MBConv6 5×5	56×56	40	2
5	MBConv6 3×3	28×28	80	3
6	MBConv6 5×5	14×14	112	3
7	MBConv6 5×5	14×14	192	4
8	MBConv6 3×3	7×7	320	1
9	Conv 1×1 & Pooling & FC	7×7	1280	1

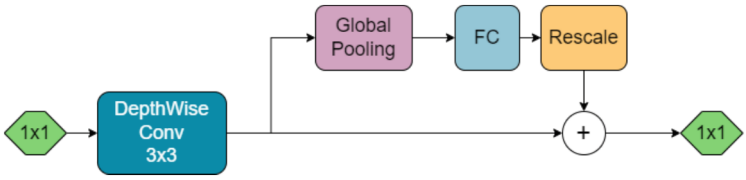

Fig. 7. Schematic diagram of MBConv.

The fundamental building blocks at the core of EfficientNet are the MBConv blocks, initially introduced in MobileNetV2 [16]. These blocks represent an inverted residual bottleneck architecture characterized by an expansion of the channel count in intermediate layers, followed by a subsequent reduction to their original size. Figure 7 provides a visual representation of the MBConv block. The numeric value following the name, such as in "MBConv1" and "MBConv6", indicates the number of filters in the initial 1×1 convolutional layer on the left. For instance, if the input channels number 4, then the MBConv1 block will have 4 filters, while the MBConv6 block will have 24 (4 * 6).

3.3 Results

A series of experiments were conducted to assess the viability of our proposed method. The training ran on an Intel Core™ i5-10400F CPU@2.90 GHz computer, complemented by an Nvidia® GPU RTX-3060 with 12GB of memory. The model is implemented in Python, with libraries such as TensorFlow, Keras, and Segmentation Models [10].

The patient-wise division for training, testing, and validation occurs as follows: Out of the 40 patients in the dataset, 30% (12) were designated for the testing set, while 62.5% (25) stands for the training set. The remaining 7.5% (3) of patients were separated for validation. It's crucial to underscore that our results derive from a dataset that incorporates 30% less data than previous studies on the same dataset. This discrepancy originated because 20 CT scans were hidden in a challenge platform and were used only to perform model evaluation, but the challenge ended, and the 20 volumes are still not public.

The train lasted for 100 epochs and employed the ADAM [12] (Adaptive Moment Estimation) to update weights. The chosen loss function was Dice Loss. The best-performing model reached 78.35% test dice score. This result was found with EfficientNetB0, the most compact model of the EfficientNet family, which has only *10 million* trainable parameters, as opposed to the literature's best performer [19], which has *120 million*. Table 2 summarizes the dice score for the main methods of the literature for trachea segmentation in comparison to this paper's method.

Table 2. Contrast with Relevant Studies.

Paper	Method	Dataset	Dice Score (%)	#Params
[19]	MultiRes 3D U-Net	SEGTHOR	92.17	120M
[20]	U-Net 2.5D	SEGTHOR	92.56	23M
[8]	U-Net/SPP Module	SEGTHOR	89.00	26M
[22]	U-Net/ GAN	SEGTHOR	88.30	∼
Current Work	Eff-Unet	SEGTHOR (-30%)	78.35	10M

3.4 Discussion

To visually evaluate what the achieved 78.35% dice score represents for a segmentation model, some case studies are illustrated below. The network demonstrates robust performance in accurately segmenting the central regions of the trachea, as exemplified in Fig. 8. In this example, the Eff-Unet [4] model correctly identified the trachea's shape and precise location. The prediction is indicated in the left image, in gray contour, while the ground truth is on the right image, marked in orange.

Some other regions of the organ poses a more intricate challenge, primarily due to the organ's shape variations depending on its depth within the body. In certain depths, the trachea may bifurcate into two separate tubes. Figure 9

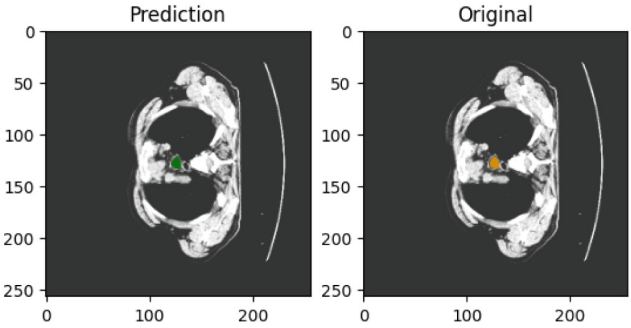

Fig. 8. Case Study: Forecasting the Central Segment of the Trachea. The leftmost image is overlapped with model's prediction, tinted in green. Right image represents the ground truth, provided by a physician, and is marked with orange (Color figure online)

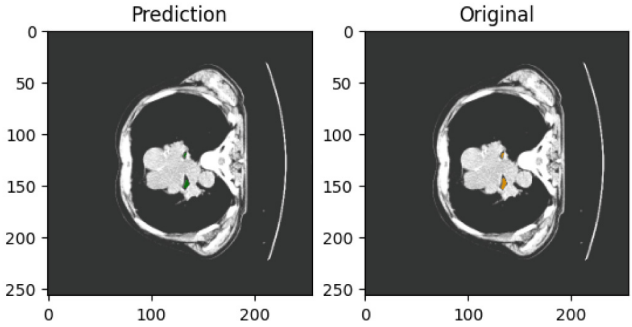

Fig. 9. Case Study: 2D slice featuring trachea division into two tubes. The leftmost image displays the model's prediction, highlighted in green and overlapped with the original. On the right, the ground truth, validated by a physician, is presented in orange. (Color figure online)

contains a 2D slice where the trachea splits into two. Despite the adversity and varying shape, it is clear that the model found no issue labeling the tracheal region pixels in the image.

For peripheral tracheal slices (those near the organ's ends) or images lacking tracheal presence, the model encounters difficulties in providing accurate outputs. This occurs due to specific image regions displaying textural features and shapes closely resembling the trachea. Consequently, the network generates false positives, as illustrated in Fig. 10. It's worth noting that while this CT slice does not contain the trachea, another body structure shares a similar shape and texture, leading the model to misclassify this region as the trachea. This false positive area is highlighted in green.

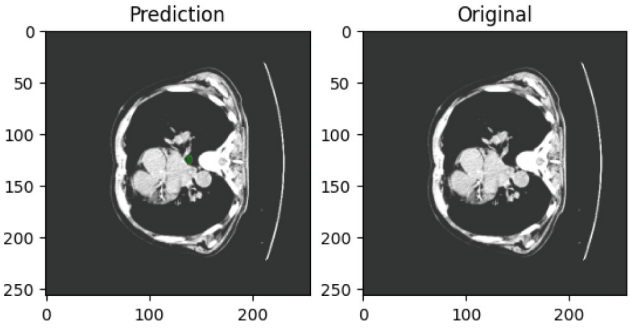

Fig. 10. Case Study: The primary limitations of the method are evident in slices devoid of trachea presence. In the leftmost image, the model's prediction is highlighted in green and overlapped. On the right, the ground truth, supplied by a physician, corresponds to an image without trachea presence, hence no markings are present. (Color figure online)

Precisely delineating slices of this nature poses a central limitation for the model. Nevertheless, this hurdle can be mitigated by incorporating post-processing techniques, specifically selecting solely the largest 3D object. This approach considers the model's predictions across the entire CT scan, ultimately enhancing segmentation accuracy.

While it may not claim the top spot as the highest dice score in the literature, this method consistently delivers commendable results that are in close to those achieved by other studies. Notably, our model achieves this performance with substantially fewer parameters (approximately 12 times fewer than the model presented in [19]). This reduced parameter count signifies a compact fully convolutional neural network, that operates with significantly reduced computational resource demands compared to its counterparts.

This efficiency is achieved by leveraging the EfficientNet [17] architecture, which is renowned for its compact and scalable design philosophy. It has been purposefully engineered to address resource constraints while maintaining robust performance, making it an ideal choice for applications where computational efficiency is a priority.

4 Conclusion

In this study, we introduced an automated approach for trachea segmentation in CT images, employing the Eff-Unet [4] neural network architecture. The obtained results exhibit relative competitiveness within the field. The utilization of Eff-Unet architecture led to a **compact model** for trachea segmentation that can be embedded into broader healthcare systems, being a valuable tool in the realm of medical image analysis, particularly in Organ at Risk segmentation and Radiotherapy planning.

There is ample room for future enhancements in this methodology, with the potential to elevate its effectiveness further. Some promising avenues for improvement are: **(a):** adding neighboring slices information to the network input, enabling a richer understanding of the 3D context and potentially improving segmentation performance, **(b):** implementing post-processing methods to refine the model's output, such as false positive reduction and segmentation refinement.

Ultimately, these improvements can potentially improve the quality of life for individuals dealing with cancer through the optimization of treatment planning processes.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
2. Amaro, E.J., Yamashita, H.: Aspectos básicos de tomografia computadorizada e ressonância magnética. Braz. J. Psychiatry **23** (2001)
3. Azad, R., et al.: Medical image segmentation review: the success of U-net (2022)
4. Baheti, B., Innani, S., Gajre, S., Talbar, S.: Eff-UNet: a novel architecture for semantic segmentation in unstructured environment. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1473–1481 (2020). https://doi.org/10.1109/CVPRW50498.2020.00187
5. Baskar, R., Lee, K.A., Yeo, R., Yeoh, K.W.: Cancer and radiation therapy: current advances and future directions. Int. J. Med. Sci. **9**(3), 193 (2012)
6. Consídera, D.P., et al.: A tomografia computadorizada de alta resolução na avaliação da toxicidade pulmonar por amiodarona. Radiol. Bras. **39**, 113–118 (2006)
7. Diniz, J., Ferreira, J., Silva, G., Quintanilha, D., Silva, A., Paiva, A.: Segmentação de coração em tomografias computadorizadas utilizando atlas probabilístico e redes neurais convolucionais. In: Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde, pp. 83–94. SBC, Porto Alegre (2021). https://doi.org/10.5753/sbcas.2021.16055, https://sol.sbc.org.br/index.php/sbcas/article/view/16055
8. Feng, S., et al.: CPFNet: context pyramid fusion network for medical image segmentation. IEEE Trans. Med. Imaging **39**(10), 3008–3018 (2020). https://doi.org/10.1109/TMI.2020.2983721
9. Gupta, T., Narayan, C.A.: Image-guided radiation therapy: physician's perspectives. J. Med. Phys./Assoc. Med. Phys. India **37**(4), 174 (2012)
10. Iakubovskii, P.: Segmentation models (2019)
11. Kalender, W.A.: X-ray computed tomography. Phys. Med. Biol. **51**(13), R29 (2006)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2017)
13. Lambert, Z., Petitjean, C., Dubray, B., Ruan, S.: SegTHOR: segmentation of thoracic organs at risk in CT images (2019). https://doi.org/10.48550/ARXIV.1912.05950, https://arxiv.org/abs/1912.05950
14. Noël, G., Antoni, D., Barillot, I., Chauvet, B.: Délinéation des organes à risque et contraintes dosimétriques. Cancer/Radiothérapie **20**, S36–S60 (2016). https://doi.org/10.1016/j.canrad.2016.07.032, https://www.sciencedirect.com/science/article/pii/S1278321816301676. Recorad: Recommandations pour la pratique de la radiothérapie externe et de la curiethérapie

15. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetv2: inverted residuals and linear bottlenecks (2019)
17. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (2019). https://proceedings.mlr.press/v97/tan19a.html
18. Tekatli, H., et al.: Normal tissue complication probability modeling of pulmonary toxicity after stereotactic and hypofractionated radiation therapy for central lung tumors. Int. J. Radiat. Oncol.* Biol.* Phys. **100**(3), 738–747 (2018)
19. Wang, Q., et al.: 3D enhanced multi-scale network for thoracic organs segmentation. SegTHOR@ ISBI **3**(1), 1–5 (2019)
20. Wang, S., et al.: Conquering data variations in resolution: a slice-aware multi-branch decoder network. IEEE Trans. Med. Imaging **39**(12), 4174–4185 (2020). https://doi.org/10.1109/TMI.2020.3014433
21. Yan, X., Tang, H., Sun, S., Ma, H., Kong, D., Xie, X.: After-UNet: axial fusion transformer UNet for medical image segmentation. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 3270–3280. IEEE Computer Society, Los Alamitos (2022). https://doi.org/10.1109/WACV51458.2022.00333, https://doi.ieeecomputersociety.org/10.1109/WACV51458.2022.00333
22. Zhao, W., Chen, H., Lu, Y.: W-net: a network structure for automatic segmentation of organs at risk in thorax computed tomography. In: Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing, IMIP 2020, pp. 66-69. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3399637.3399642
23. Zoph, B., Le, Q.: Neural architecture search with reinforcement learning. In: International Conference on Learning Representations (2017). https://openreview.net/forum?id=r1Ue8Hcxg