International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Sep-24)

http://dx.doi.org/10.12785/ijcds/160183

Overview of Medical Image Segmentation Techniques through Artificial Intelligence and Computer Vision

Hanan SABBAR[1], Hassan SILKAN[1], Khalid ABBAD[2], El Mehdi Bellfkih[3] and Imrane Chems Eddine Idrissi[4]

[1]*LAROSERI, Department of Computer Science, Chouaib Doukkali University, Faculty of Sciences, EL Jadida, Morocco*
[2]*LSIA, Department of Mathematics and Computer Science, Sidi Mohamed Ben Abdellah University, Fes, Morocco*
[3]*LAMS, Department of Mathematics and Computer Science, Hassan II University, Casablanca, Morocco*
[4]*LTIM, Department of Mathematics and Computer Science, Hassan II University, Casablanca, Morocco*

Received 22 Feb. 2024, Revised 21 May 2024, Accepted 22 May 2024, Published 1 Sep. 2024

Abstract: Medical image segmentation is a crucial task in computer vision with significant implications in diagnostics, treatment planning, and medical research. This study comprehensively explores various methodologies employed for image segmentation within the medical field, ranging from traditional techniques such as thresholding, edge detection, region-based methods, and clustering, to advanced artificial intelligence strategies, particularly deep learning. Each method's strengths and limitations are thoroughly examined to provide a clear perspective on their effectiveness. The paper focuses on analyzing different architectures specifically used for medical image segmentation, evaluating their performance meticulously. It aims to delve deeply into the varied segmentation techniques, providing a comparative analysis that highlights their effectiveness across different scenarios. Additionally, the study addresses the latest technological advancements in segmentation, emphasizing breakthroughs that have the potential to transform the accuracy and efficiency of medical image analysis. An exhaustive compilation and detailed critique of results from employing various segmentation strategies are presented, offering insights into the outcomes of diverse approaches. This includes an in-depth discussion of the inherent strengths and weaknesses of the techniques used in medical image segmentation. The research enhances understanding of how these methodologies can be applied effectively within the medical sector, particularly in areas leveraging computer vision. By advancing knowledge in this field, the study paves the way for future research that could further improve the capabilities and applications of image segmentation technology in medicine, potentially leading to better patient outcomes and more efficient medical practices. This research enhances the comprehension of how these methods can be applied within the medical sector, especially in the area of computer vision.

Keywords: Segmentation, Computer vision, Medical image, Machine learning, X-rays, Deep learning.

1. INTRODUCTION

Image segmentation involves dividing an image into separate regions or segments to identify objects or areas of interest [1]. This technique has been crucial for diagnosing diseases for many years. This approach is designed to transform an image into clear and visually interpretable sections, facilitating the recognition process across different medical applications. Today, numerous modalities of medical imaging such as radiography, MRI, computed tomography (CT), ultrasound, and more. The choice of imaging modality depends on factors such as acquisition speed, image resolution, and patient comfort. These technologies are instrumental in detecting and diagnosing conditions at their earliest, ensuring timely and crucial interventions when they are most needed. Once a medical image is acquired, the real challenge begins. A healthcare professional must painstakingly examine each image to detect possible diseases and determine their potential causes. This manual inspection is not only time-consuming often taking hours to days depending on the complexity of the case but also prone to human error, particularly in cases involving subtle anomalies or when under time constraints. Professionals need to assess the size of organs, identify any anomalies, and decide on the necessary treatments. These tasks are performed by identifying regions of interest, a process where segmentation is implicitly crucial but not always clearly defined or standardized. This is where the critical role of medical image segmentation becomes apparent. Segmentation assists in the detection and quantification of abnormalities, aids in the creation of precise surgi-

E-mail address: sabbar.h@ucd.ac.ma, Silkan_h@yahoo.fr , Khalid.abbad@usmba.ac.ma

cal plans, and tracks the advancement of diseases. More importantly, it can significantly lighten the workload of healthcare workers by automatically pinpointing regions of interest within medical imagery. However, medical image segmentation faces numerous inherent challenges such as low contrast, high levels of noise, and the presence of artifacts, which can complicate the extraction of accurate and reliable information. Historically, conventional methods have been utilized to address the challenges associated with medical image analysis, yet they frequently struggle to cope with the intricacy and diversity of such images. Recently, deep learning techniques[2] have proven to be highly effective, significantly enhancing the precision and efficiency of these analyses.

The integration of deep learning into artificial intelligence has substantially improved the functionality of medical image segmentation. These deep learning strategies are adept at autonomously extracting detailed hierarchical features from complex data, delivering superior performance over traditional machine learning and computer vision methods in terms of both accuracy and processing speed. Such progress not only alleviates the burden on medical professionals by automating the identification of critical areas within medical images but also leads to more precise and streamlined diagnoses and treatment protocols. Additionally, deep learning models excel at detecting minute or early-stage pathological changes that manual methods may miss, thereby enhancing treatment outcomes and facilitating prompt medical interventions.

This review thoroughly examines the field of medical image segmentation, detailing the advantages and drawbacks of various segmentation techniques and their application across different imaging modalities. The selection of specific methods or algorithms often depends on the type of imaging modality used and the particular challenges presented by the medical condition being examined. The evolution of segmentation strategies in medical imaging has often been explored in literature reviews [3], [4].

Techniques for segmenting medical images can be divided into two main categories: traditional methods that utilize machine learning, and innovative strategies that employ artificial intelligence. Below is a depiction of the predominant medical image segmentation techniques found in each category, as presented in Figure 1:

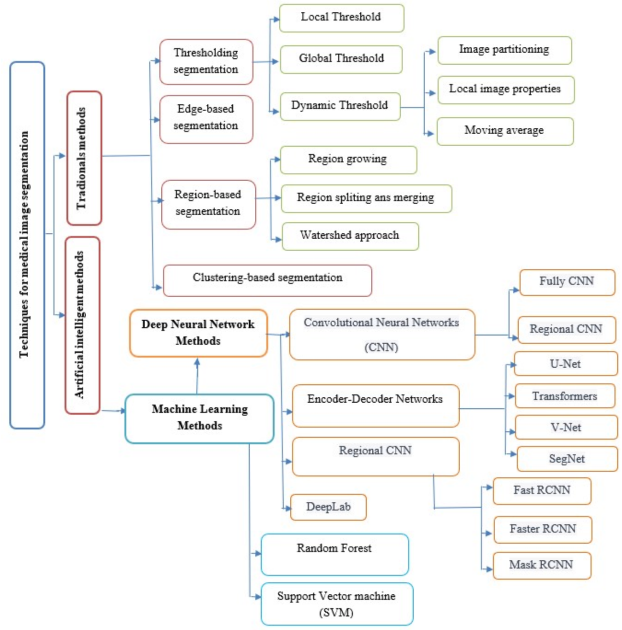

Figure 1. Techniques for Segmenting Medical Images.

The organization of this paper is outlined as follows: Section 2 presents an examination of the current literature, encompassing earlier related research efforts. Section 3 provides a summary of different conventional frameworks applied in the segmentation of medical images. Section 4 explores the recent frameworks utilizing artificial intelligence for segmenting medical images. Section 5 conducts a comparative analysis between various deep learning models and conventional frameworks. In conclusion, Section 6 wraps up the paper, highlighting potential avenues for future studies and applications within the realm of 'biomedical image segmentation.

2. Review of Literature

Various methodologies for medical image segmentation have been explored, with Lee. 2007 [5] introducing a statistical approach incorporating morphological operations and Gaussian mixture modeling, demonstrating efficacy in CT image segmentation. Similarly, Ashwani et al. [6] developed a technique based on thresholding and morphology for brain MRI segmentation, validated through CT Angiography. This approach achieved performance ratings of 95.4% for brain MRIs and 95.8% for CT-Angiography, assessed by completeness.

In a recent study, Bhosle et al. 2023[7] evaluated binary adaptive and Otsu thresholding techniques for lung segmentation in CT images, identifying adaptive thresholding as the superior method with a 78.69% accuracy rate. Binary inverse thresholding followed closely at 75.59%, while Otsu's method, despite its computational simplicity, only achieved 61.70% accuracy due to its lower efficacy in handling images with diverse pixel intensities. This research provides essential insights for selecting the optimal thresholding technique for image segmentation, balancing accuracy, and the particular demands of varying image types.

Zhou et al. 2018 [8] explored the efficacy of an innovative segmentation method for identifying multiple organs in computed tomography (CT) images, leveraging a Convolutional Neural Network (CNN) architecture. Their evaluation focused on Mean Accuracy and the Jaccard Similarity Index (JSI), revealing that the method achieved a mean JSI of 79% with a 3D deep CNN and 67% using a 2D deep CNN across seventeen organ types. This indicates the technique's versatility and high performance in segmenting a variety of organs. Jia et al. 2017[9], introduced an approach based on Fully Convolutional Networks (FCN) for segmenting histopathology images using deep weak supervision. This method innovatively utilized superpixels rather than standard pixels, effectively enhancing the preservation of natural tissue boundaries. A key outcome of this approach was its superior performance in segmentation accuracy, as evidenced by an F1 score of 83.6%. This score notably exceeded that of other existing algorithms under weak supervision, marking a significant advancement in the field. Fully convolutional networks (FCN) [10], including models like U-Net [11], DeepMedic [12], and holistically nested networks [13], have proven to be effective and accurate in a range of segmentation challenges, covering areas such as cardiac magnetic resonance (MR) [14], brain tumors [15], and abdominal CT scans [16], Inspired by DenseNet architecture [17].

Ummadi (2022) [18] reviewed U-Net and its derivatives (UNet++, R2UNet, Attention UNet, TransUNet), underscoring their pivotal role in facilitating non-invasive diagnoses through high performance across diverse biomedical segmentation tasks. Inspired by the foundational work of GoogleNet [19], [20], Gu et al. [21] developed CE-Net, integrating the inception model into the domain of medical imaging segmentation. This integration augments feature extraction capabilities through the use of atrous convolution, allowing for an expanded capture of spatial details. Additionally, CE-Net utilizes 1×1 convolutions within its feature maps to incorporate the inception design, albeit this intricacy introduces hurdles in terms of model flexibility. Dosovitskiy et al. [22] introduced the Vision Transformer (ViT), marking a breakthrough in medical image analysis by providing an innovative alternative to conventional convolutional neural networks (CNNs). Originating from advancements in natural language processing, ViT has been successfully implemented in the segmentation of medical images, as evidenced by recent implementations such as TransUnet (2021) [23], UTNet (2021) [24], and Swin-unet (2021) [25]. These applications highlight ViT's capability to manage complex interdependencies that exceed CNNs' scope. Combining ViT with the CNN framework is emerging as an effective method for enhancing the precision and efficiency of segmentation in medical imaging. In the realm of medical image segmentation, the latest breakthroughs have been aimed at improving the precision of organ and lesion outlines.

Cui et al. (2023) [26] developed an advanced cardiac segmentation technique, CFUN+, by integrating Faster R-CNN with 3D U-Net, addressing GPU memory constraints for high-resolution 3D data. Their approach, using a new Complete Intersection over Union (CIoU) loss and edge loss, significantly enhances accuracy and speeds up the segmentation process. The method achieved a notable 5.2% improvement in Dice score over the baseline and reduced segmentation time to under six seconds.

Wu et al. (2024) [27] introduce MedSegDiff-V2, a new framework that combines UNet architectures and vision transformers. This approach demonstrates marked superiority over previous methodologies in 20 medical image segmentation tasks.

Chen et al. (2023) [28] developed and introduced TransAttUnet, marking noteworthy progress in medical image segmentation technology. This attention-based network boosts semantic segmentation by merging multi-level attention mechanisms and multi-scale connectivity within the U-Net structure.

3. Traditional methods

Traditional medical image segmentation methods encompass a variety of classical image processing and machine learning techniques, each with distinct advantages and limitations. These methods often require manual or semi-automatic intervention, relying on predefined rules, handcrafted features, and mathematical algorithms. Key traditional approaches include thresholding [29], which is simple and practical but can struggle with medical images containing diverse regions, leading to noise and over-segmentation issues. Advanced thresholding techniques like the OTSU method [30] aim to refine this process using local statistical information. Edge-based segmentation [31] accurately detects transitions in image properties but is sensitive to noise, whereas region-based techniques like region growing [32]and the watershed approach [33] group pixels based on similarity, offering diverse segmentation methods but potentially lacking in precision. Clustering-based segmentation groups similar pixels based on intensity or feature similarity. Popular algorithms like K-means or ISODATA [34], fuzzy c-means [35], and the expectation-maximization (EM) algorithm [36]vary in their approach to grouping data, with K-means focusing on mean intensities [37] and fuzzy c-means offering soft segmentations [38]. The EM algorithm assumes Gaussian mixture models to estimate mixture components and posterior probabilities. Each method showcases a unique array of advantages, making them suitable for specific image types and segmentation challenges. However, these approaches also come with inherent limitations, particularly when addressing the complexity of medical images that demand highly accurate segmentation. Often, enhancements are required to achieve greater precision and specificity across various medical imaging applications. The comparative tableI below offers an overview of these traditional methods, highlighting their strengths and limitations within the context of medical image segmentation.

TABLE I. Comparative Overview of Traditional Medical Image Segmentation Methods

Techniques	Advantages	Limitations
Thresholding [39]	-Among the simplest and most effective methods.	-Ineffective for images with complex intensity distributions. -Struggles with images that have histograms close to unimodal.
Edge Detection [40]	-Works well for images with clear edges.	-Not applicable to images with many edges. -Inadequate for images where edges are not well-defined.
Region Detection [41]	-Ideal for images with distinct regions.	-Not applicable to images with many edges. -Ineffective for images where region borders are not clear.

4. Intelligence artificial methods

Amid swift progress in artificial intelligence, alongside machine learning and deep learning techniques, the approach to segmentation has undergone a transformative shift. Nevertheless, the advent of sophisticated neural networks, including Convolutional Neural Networks (CNN) and encoder-decoder architectures, has markedly enhanced segmentation efficacy. These advanced deep learning models are adept at extracting intricate features and identifying distinctive patterns within extensive datasets, resulting in segmentations that are both more precise and reliable. In the subsequent sections, an outline of traditional machine learning and contemporary deep learning approaches to medical image segmentation will be provided.

A. Machine learning methods

Machine learning techniques for segmentation are a crucial component of medical image analysis, facilitating the automated extraction and recognition of crucial structures and areas within medical images. The segmentation methods in medical imaging are based on machine learning principles, focusing on Support Vector Machine (SVM) and Random Forest algorithms. SVM, a powerful learning system widely used in pattern recognition, computer vision, and bioinformatics, has demonstrated superior performance compared to traditional classifiers [42]. In medical imaging, SVMs utilize supervised learning to discern complex boundaries between structures, ensuring accurate segmentation of tissues or lesions. Meanwhile, Random Forest, another robust machine learning algorithm for medical imaging, relies on labeled training data, which can be challenging to obtain in medical domains. To address this challenge, semi-supervised learning methods like semi-supervised random forest [43], CoForest [44], and semi-supervised super-pixel method [45] have been introduced, integrating unlabeled data to enhance performance and optimize segmentation accuracy. These techniques represent significant advancements in automating medical image segmentation, enabling precise analysis and diagnosis.

B. Deep Neural Network Methods

Deep learning has achieved remarkable advancements in the field of image segmentation, outperforming traditional approaches. Subsequent parts will provide a detailed examination of diverse deep-learning strategies for segmenting medical images. This includes Convolutional Neural Networks (CNNs) like R-CNN, and encoder-decoder frameworks such as U-Net, V-Net, and SegNet, alongside DeepLab-based segmentation networks and Transformer models.

1) Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have become widely recognized in the domains of computer vision and medical image analysis for their capacity to autonomously identify pertinent features within images, leading to remarkable performance in segmenting anatomical structures and abnormalities in medical images [46]. CNNs ((see Figure 2) consists of three main layers: the convolutional layer, which detects distinct features in images through mathematical operations; the pooling layer, which reduces spatial dimensions without changing depth, reducing computational requirements for subsequent layers; and the fully connected layer, where high-level reasoning and integration of feature responses occur, enabling accurate image analysis. These network architectures have demonstrated remarkable efficacy in medical imaging, transforming the field and substantially enhancing the accuracy of image segmentation. CNNs facilitate meticulous segmentation of anatomical structures across diverse imaging modalities, including MRI, CT, and X-rays [46].

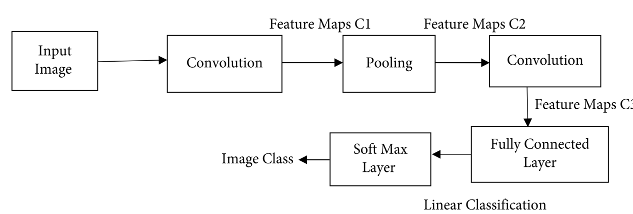

Figure 2. Convolutional neural network architecture.

As CNN models and architectures have continued to advance, medical image segmentation has achieved unprecedented levels of accuracy and efficiency. Notable deep neural network architectures for image segmentation, including U-Net, V-net, and DeepLab (illustrated in Figure 1), have played a pivotal role in this progress. These CNN-based segmentation techniques are in a constant state of evolution, continually enhancing segmentation outcomes and broadening the scope of clinical applications. Additionally, recent developments have introduced techniques like TransUNet, TransFuse, MedT, and TransAttUnet, which combine the power of Transformers and CNNs to further elevate the state of medical image segmentation. These hybrid methodologies have demonstrated the potential to address intricate segmentation challenges within the domain of medical image analysis.

2) U-NET architecture

Ronneberger et al.[47] introduced the U-Net model at the MICCAI conference in 2015(see Figure 3), marking a significant advancement in leveraging deep learning for segmenting medical images. The U-Net model, a tailored Fully Convolutional Network (FCN) for the segmentation of biomedical images, features an encoder, a bottleneck module, and a decoder. Its design has been widely embraced due to its capability to meet the complex requirements of segmenting medical imagery. UNet is frequently used for segmenting various types of medical images, such as MRIs, microscopy images, and CT scans, where detail precision is essential. Figure 3 depicts the U-Net framework. Furthermore, a variety of fundamental U-Net models have been modified for the segmentation of medical images, seeing extensive application. These U-Net modifications and related deep learning frameworks strive to improve segmentation quality by increasing accuracy and computational efficiency, which is facilitated by adjustments in the network architecture and the incorporation of innovative modules. Subsequent iterations of U-Net, such as U-Net++, R2U-Net, Attention U-Net, and Trans U-Net, represent progressive enhancements to the original architecture, tailored to improve the accuracy and operational efficiency in medical image segmentation tasks. U-Net++ introduces nested connections to facilitate a more nuanced semantic interpretation and a smoother gradient propagation. R2U-Net merges residual with recurrent connections, enhancing the model's capability in handling temporal sequence data. Attention U-Net incorporates attention mechanisms to concentrate on particular areas of interest, and Trans U-Net amalgamates transformer network elements, boosting performance in complex segmentation tasks. These U-Net variations have shown remarkable efficacy, even when trained on limited datasets, proving their high precision in biomedical segmentation endeavors [47].

3) SegNet architecture

The SegNet architecture, known as the CNN encoder-decoder, has demonstrated effectiveness in handling medical semantic image segmentation, as outlined in the work of Salem et al [48]. This design features a symmetrical structure, comprising five encoders and five decoders, each equipped with convolution layers, batch normalization, a

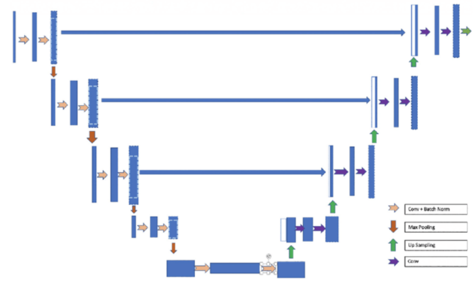

Figure 3. The structure of U-Net [47].

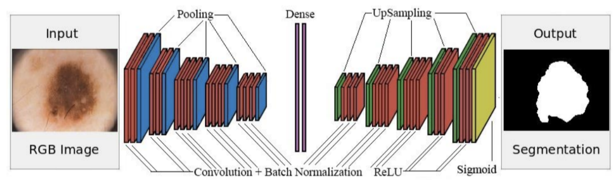

Figure 4. Polyp image segmentation with Segnet.

.

rectified linear unit (ReLU) layer, a max-pooling layer, upsampling, and a SoftMax classifier, as illustrated in the referenced figure 4. SegNet is an advanced medical image segmentation technique based on Convolutional Neural Networks (CNNs). Its fundamental principle revolves around the use of encoder-decoder architecture, where the image is encoded into low-level features and then decoded to produce segmentation. Unlike other architectures, SegNet employs an indexing mechanism during the decoding step. Indices of important pixels obtained during encoding are reused to produce accurate segmentation during decoding. This process enables SegNet to maintain crucial spatial information while reducing the number of parameters, making it efficient for complex medical image segmentation. The SegNet framework, in comparison to other models such as U-Net [11] and FCN [49], stands out due to its efficient memory usage and reduced processing time.

SegNet stands out for its ability to retain important details through the use of max-pooling indices, thereby optimizing segmentation quality without the need for post-processing. This architecture reduces complexity and resource requirements, making the processing of high-resolution images more efficient. By producing smooth images directly, SegNet streamlines the segmentation process, providing a streamlined and precise approach to image segmentation, notably within the medical sector. It is adept at accurately delineating diverse anatomical features and abnormalities, including tumors, blood vessels, and other pathologies across various imaging modalities such as MRI, CT, and Ultrasound.

4) VNet architecture

Milletari et al. [50] introduced the V-Net architecture, an adaptation of the U-Net framework, featuring a 3D deformation structure suitable for images acquired by MRI

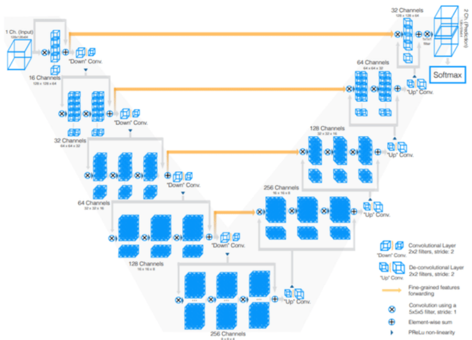

Figure 5. The structure of V-Net.

or CT. The design of this network is depicted in Figure 5. V-Net is also an encoder-decoder architecture but specifically designed for 3D segmentation. Unlike U-Net, V-Net employs residual connections in both the encoder and decoder, allowing it to capture information at different spatial scales. V-Net is primarily used for 3D medical image segmentation, such as images from CT scans or MRIs. Due to its ability to handle 3D volumes, it is particularly suited for tasks where the 3D structure of the objects to be segmented is crucial.

5) R-CNN architecture.

The R-CNN technique [51] represents a groundbreaking application of deep learning in the field of object detection. Initially, this method involved creating a feasible number of potential object regions. For each of these regions, R-CNN utilized DNNs to extract features. Subsequently, enhancements to R-CNN enabled a focus on Regions of Interest (RoIs) within feature maps through RoIPool. This advancement resulted in increased processing speed and improved precision. Several iterations of R-CNN have emerged in the research community, including Fast R-CNN, Faster R-CNN, and Mask R-CNN, each contributing unique enhancements and developments to the domain. These architectures are well-suited for segmentation tasks that require the localization and segmentation of specific regions in images, such as identifying lesions in radiographs or detecting nodules in lung scans. These models are effective at processing large images with localized precision. TableII presented below outlines the progression of R-CNN technology, detailing the innovations and refinements each version offers. This comparison illuminates the distinct benefits and obstacles associated with each, showcasing the continuous advancement in the area of image segmentation.

6) DeepLab architecture.

The DeepLab framework represents the forefront of semantic image segmentation, particularly prevalent in medical image analysis. This model leverages a pre-trained CNN, such as ResNet-101 or VGG-16, incorporating atrous convolution to efficiently extract image features [55]. Atrous convolutions offer key advantages, enabling precise feature resolution control and transforming image classifiers into dense feature extractors without extra parameters. Additionally, DeepLab utilizes conditional random fields (CRF) for detailed segmentation output. Despite these strengths, DeepLab requires meticulous hyperparameter tuning for optimal performance in medical image segmentation tasks. DeepLab is particularly useful in imaging scenarios where a clear distinction of contours is critical, such as in endoscopic or dermatological images. It adapts well to the challenges posed by variations in texture and contrast, which are often encountered in these modalities. Various versions of DeepLab have been introduced in research, such as DeepLabv1 [56], DeepLabv2, DeepLabv3[57], and DeepLabv3+ [58]. Each version brings improvements and advancements in the field. The comparative table III below illustrates the evolution of the DeepLab architecture, highlighting the innovations and optimizations introduced by each version. It sheds light on the specific advantages and challenges, reflecting the ongoing progress in image segmentation.

7) Transformers

Transformer-based techniques have revolutionized medical image segmentation, delivering groundbreaking results through innovative neural network architectures, Yuan et al. [76] designed a complementary network combining Convolutional Neural Networks (CNNs) with Transformers to improve the accuracy of medical image segmentation. This method harnesses the strengths of CNNs for extracting local features and Transformers for capturing global context, leading to more accurate and efficient segmentation. The origin of Transformer-based segmentation techniques is traced to the work of Vaswani et al., who introduced the Transformer with attention mechanisms, achieving remarkable outcomes in various language processing tasks [77]. In the medical segmentation context, Chen et al. demonstrated successful segmentation of medical images by integrating Transformers with U-Net, significantly enhancing both localization and contextual understanding [78]. Notable developments in this field include Zhang et al.'s TransFuse, a framework that merges Transformers and CNNs to achieve top-tier results across various medical image segmentation contexts [79]. The integration of Gated Axial-Attention into MedT by Valanarasu et al. surpassed prior methods in medical image segmentation, setting new benchmarks [80]

To augment semantic segmentation, Chen et al. developed TransAttUnet, employing guided attention to notably advance medical image segmentation efforts [56]. Lin and collaborators presented DS-TransUNet, which integrates the Swin Transformer with U-Net, signifying a notable advancement in the domain[81]. The realm of medical imaging has witnessed a paradigm shift with these advanced neural network architectures coming into play. Among these innovations, ViTransUNet emerges as a pioneering innovation that combines the strengths of Vision Transformers (ViT) with the traditional U-Net structure, improving the accuracy and efficiency of image segmentation tasks essential for a wide range of medical purposes. The architecture

TABLE II. Comparative Overview of R-CNN Architectures: Fast R-CNN, Faster R-CNN, and Mask R-CNN.

Criteria	Fast R-CNN	Faster R-CNN	Mask R-CNN
Architecture	Utilizes the Region Proposal Network (RPN) and a CNN for feature extraction.[52]	Utilizes RPN and a CNN, but with optimizations in region proposal method.[53]	Builds on Faster R-CNN by adding a segmentation branch with RoI-Align for precise pixel-level instance segmentation.[54]
Applications	Employed across a range of medical computer vision tasks, such as identifying tumors and segmenting organs within radiographic images.	Suited for cases where speed and precision are critical, such as computer-assisted surgery and real-time anomaly detection.	Instance segmentation and object detection for complex image analyses.
Speed	Relatively slow but offers strong performance in accurate segmentation of medical objects but The computation time is significantly increased.	Improved for increased speed compared to Fast R-CNN, suitable for real-time medical applications. The computation time is reduced.	Faster than R-CNN; additional computation for mask segmentation. Training time is significantly extended.
Accuracy	Provides high accuracy in detecting and segmenting complex medical objects.	Provides high accuracy in detecting and segmenting complex medical objects.	High accuracy for detection and instance segmentation, superior to previous models.

TABLE III. Comparative Table of DeepLab Versions.

‘Version	Description	Advantages	Limitations
DeepLabv1	Uses atrous convolution to extract features from an image and applies a Conditional Random Field (CRF) to refine object contours.[59]	The use of atrous convolution is effective at capturing contexts at various scales, while Conditional Random Fields (CRF) enhance the accuracy of object contours.	Use of CRF increases computational complexity, making the algorithm slower.
DeepLabv2	Introduces Atrous Spatial Pyramid Pooling (ASPP) which applies atrous convolutions at different sampling rates and fuses them.[60]	ASPP enhances the segmentation of objects across different scales, proving robust for objects of varying sizes.	Challenges in capturing precise fine object contours.
DeepLabv3	Utilizes atrous separable convolution to better capture object boundaries.[61]	Atrous separable convolution enables precise capture of object contours, leading to improved segmentation accuracy.	Despite improvements, challenges remain in refining object contours.
DeepLabv3+	Extends DeepLabv3 by adding a decoder module to refine segmentation results along object boundaries.[62]	achieves refined delineation of object boundaries and enhances overall segmentation precision through its advanced atrous separable convolution technique.	Model complexity requires significant GPU memory for training on high-resolution images and batch sizes.

of ViTransUNet is illustrated in Figure 6.

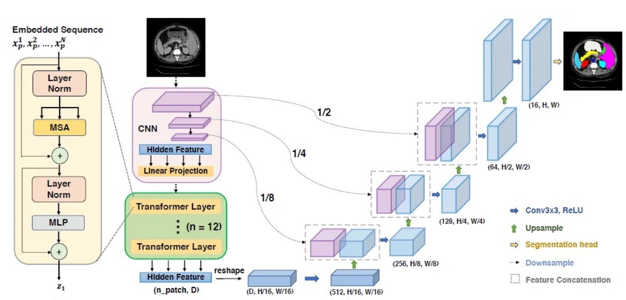

Figure 6. ViTransUNet architecture.

TABLE IV. Evaluation of Segmentation Techniques Against Cutting-Edge Benchmarks on the JSRT and Montgomery (MC) Datasets.

Methodology	Dataset	Dice	Accuracy	Recall	Precision
U-Net[11]	JSRT	96.17	98.21	94.94	97.50
FCNN [63]	JSRT& MC	95.1	97.7	95.1	98.0
Encoder-Decoder Structure [64]	JSRT	96.0	-	95.1	-
Improved Segnet [65]	JSRT	-	98.7	-	-
Edge Detection & Morphology [66]	JSRT	-	82.9	-	-
Thresholding [6]	JSRT	-	89.63	88.75	78.76
Fuzzy C-Means (FCM) [6]	JSRT	-	93.34	85.14	92.02
ResUNet [67]	JSRT	97.12	98.64	96.61	97.70
Attention U-Net [68]	JSRT	97.59	98.81	98.82	96.41
UNet++ [69]	JSRT	97.84	98.93	99.28	96.47
ResUNet++ [67]	JSRT	97.92	98.68	98.48	98.48
Swin-Unet [70]	JSRT	97.67	98.71	95.42	98.36
Improved U-Net[71]	JSRT	97.7	98.9	-	-
Improved U-Net[71]	MC	97.9	98.5	-	-
DED-CNN[72]	JSRT	97.60	-	-	-
TransUNet [73]	JSRT& MC	-	98.36	-	-
UCTransNet [74]	JSRT	98.32	99.37	-	-
TransAttUnet [75]	JSRT	**98.88**	**98.41**	**98.88**	**99.04**

Incorporating transformers into segmentation frameworks like TransUnet and Swin-Unet has led to notable improvements in segmentation accuracy, especially in demanding tasks such as accurately delineating organs and lesions. This progress is not merely a technological leap; it represents a significant stride towards achieving more precise and minimally invasive diagnostics in healthcare.

5. Comparative Study

In this study, we evaluate the effectiveness of various segmentation techniques for lung field segmentation from chest X-rays, utilizing the JSRT (Japanese Society of Radiological Technology Database) and Montgomery (MC) datasets [82]. We analyze several advanced deep-learning models alongside traditional methods to assess their performance in terms of Dice Coefficient, Accuracy, Recall, and Precision.

The results, detailed in Table IV and visually summarized in Figure 7, highlight the exceptional performance of the TransAttUnet model, which achieves the highest Dice score of 98.88% on the JSRT dataset. This score represents a significant improvement over other advanced models such as U-Net, which recorded a Dice score of 96.17%. This comparison underlines a 2.71% increase in Dice performance, illustrating the benefits of integrating attention mechanisms and multi-scale skip connections. These features enable TransAttUnet to excel by enhancing detail recognition and segmentation accuracy, crucial for medical diagnostics.

We also compare TransAttUnet against other high-performing models listed in the table, including UNet++, which achieved a Dice score of 97.84%, ResUNet++, which scored 97.92%, and UCTransNet, which scored 98.32%.

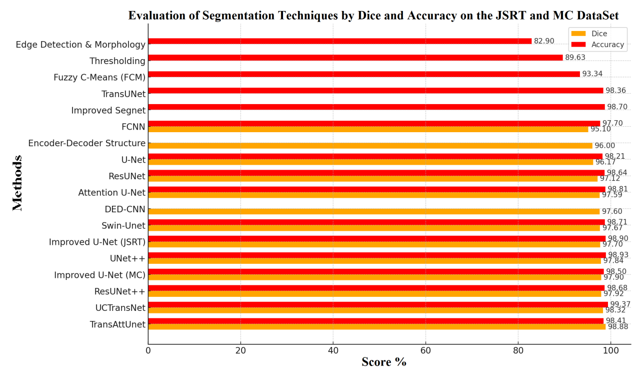

Figure 7. Performance of Segmentation Techniques by Coefficients Dice and Accuracy on the JSRT and Montgomery Datasets

While these models also demonstrate high efficacy, TransAttUnet's use of transformer-based architecture provides superior recognition of complex patterns in chest X-rays, as evidenced by its superior metrics.

Furthermore, the comparative analysis includes traditional methods such as Thresholding and Edge Detection. These methods show considerably lower performance, with Edge Detection scoring an Accuracy of only 82.9% on the JSRT dataset, significantly below the benchmarks set by deep learning models. This stark contrast emphasizes the evolution of medical image segmentation techniques from traditional approaches to more sophisticated deep-learning methodologies.

The superiority of deep learning over traditional techniques is attributed to its flexibility and ability to adapt to the specifics of medical images. Conventional methods, limited by unchangeable parameters, struggle to handle the complex variability of medical data. In contrast, deep learning adjusts its models for precise segmentation, efficiently leveraging the diversity of features and anomalies present.

This juxtaposition not only validates the advancements brought about by deep learning in the analysis of medical images but also emphasizes the pivotal role of attention mechanisms in enhancing model sensitivity to relevant features for segmentation. The comparison reveals that while traditional techniques and early neural network models provided a foundational approach for segmentation, the integration of attention mechanisms and advanced neural architectures such as TransAttUnet offers a significant enhancement in segmentation precision. This becomes particularly clear in complex endeavors such as segmenting lung fields from chest X-rays, where accurately outlining the lung edges is vital for correct diagnosis and treatment formulation.

In conclusion, the deep learning models, particularly TransAttUnet, outperform traditional segmentation methods by a substantial margin. This not only demonstrates the advancements in artificial intelligence applications in medical imaging but also reinforces the need for continued research and development in this field to leverage the full potential of deep learning technologies for medical diagnostics.

6. Conclusions and Future Work

In this paper, we have conducted a comprehensive evaluation of traditional and AI-based image segmentation methods, delineating their respective benefits and limitations within the realm of medical imaging. Traditional techniques are valued for their simplicity and low data requirements. However, they often fall short in complex imaging scenarios characterized by significant noise and variations in intensity. On the other hand, AI-based methods, especially those utilizing Convolutional Neural Networks (CNNs) and attention mechanisms, consistently achieve higher accuracy. These sophisticated approaches leverage large datasets to adeptly identify pertinent features and effectively manage the inherent variability in medical images. The decision between traditional and AI-based methods will hinge on the specific requirements of the segmentation task, the resources available, and the desired level of performance.

Future research should therefore concentrate on enhancing the practicality and effectiveness of AI-driven medical image segmentation. Priority should be given to advancing data preprocessing techniques, which are crucial for minimizing issues related to noisy data and thus enhancing the quality and precision of segmentation outputs. Additionally, there is an urgent need to develop hybrid models that integrate diverse AI strategies. These models would combine various CNN architectures with machine learning algorithms or incorporate newer AI techniques, utilizing the collective strengths of these systems to foster more accurate and robust segmentation capabilities.

In conclusion, AI-based techniques have demonstrated significant potential in enhancing medical image segmentation, yet the field is still evolving. Ongoing research and development are crucial to surmount existing obstacles and to fully exploit the capabilities of these sophisticated technologies for medical diagnostics and treatment planning. By addressing these challenges, the future of medical imaging is poised to become more precise, efficient, and accessible, ultimately leading to improved patient outcomes and the enhancement of global healthcare services.

References

[1] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2021.3059968

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[3] N. M. Zaitoun and M. J. Aqel, "Survey on image segmentation techniques," *Procedia Computer Science*, vol. 65, pp. 797–806, 2015.

[4] S. Kumar, L. A. Fred, S. Muthukumar, A. H. Kumar, and S. P. Varghese, "A voyage on medical image segmentation algorithms," *Biomedical Research*, pp. 1–12, 2017, special Issue.

[5] S. Seo, Y. Jeong, and J. S. Kenney, "A modified cmos frequency doubler considering delay time matching condition," in *2007 International Symposium on Information Technology Convergence (ISITC 2007)*. Jeonju, Korea: IEEE, 2007, pp. 392–395.

[6] A. K. Yadav, R. Roy, Rajkumar, Vaishali, and D. Somwanshi, "Thresholding and morphological based segmentation techniques for medical images," in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 2016, pp. 1–5.

[7] S. Bhosle, D. Kanade, S. S. Bhosale, V. R. Bho-sale, R. S. Bhople, and S. B. Bhegade, "Comparative analysis between different lung segmentation techniques," in *ITM Web of Conferences*, A. Sumathi, N. Yuvaraj, and N. Ghazali, Eds., vol. 56, 2023, p. 04003.

[8] X. Zhou *et al.*, "Performance evaluation of 2d and 3d deep learning approaches for automatic segmentation of multiple organs on ct images," in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575, 2018, p. 83.

[9] Z. Jia, X. Huang, E. Chang, and Y. Xu, "Constrained deep weak supervision for histopathology image segmentation," *IEEE Trans Med Imag*, vol. 36, no. 11, pp. 2376–2388, 2017.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. Springer, 2015, pp. 234–241.

[12] K. Kamnitsas, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[13] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[14] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A. Lee, N. Aung, E. Lukaschuk, M. Sanghvi *et al.*, "Human-level cmr image analysis with deep fully convolutional networks," *arXiv preprint arXiv:1710.09289*, 2017.

[15] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, Cham, 2018, pp. 450–462.

[16] H. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, K. Misawa, and K. Mori, "Hierarchical 3d fully convolutional networks for multi-organ segmentation," *arXiv preprint arXiv:1704.06382*, 2017.

[17] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P. A. Heng, "H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from ct volumes," *arXiv preprint arXiv:1709.07330*, 2017.

[18] V. Ummadi, "U-net and its variants for medical image segmentation: A short review," https://doi.org/10.48550/ARXIV.2204.08470, 2022.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2015, pp. 1–9.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2016, pp. 2818–2826.

[21] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner *et al.*, "An image is worth 16x16 words: transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations*, 2019.

[23] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:210204306*, 2021.

[24] Y. Gao, M. Zhou, and D. Metaxas, "Utnet: A hybrid transformer architecture for medical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021, pp. 61–71.

[25] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:210505537*, 2021.

[26] H. Cui *et al.*, "An improved combination of faster r-cnn and u-net network for accurate multi-modality whole heart segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3408–3419, 2023.

[27] J. Wu, W. Ji, H. Fu, M. Xu, Y. Jin, and Y. Xu, "Medsegdiff-v2: Diffusion-based medical image segmentation with transformer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6. AAAI, 2024, pp. 6030–6038.

[28] B. Chen, Y. Liu, Z. Zhang, G. Lu, and A. W. K. Kong, "Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–14, 2023.

[29] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[30] X. Yang, G. Xu, and T. Zhou, "An effective approach for ct lung segmentation using region growing," in *Journal of Physics: Conference Series*, vol. 2082, 2021, p. 012001.

[31] N. Otsu, "Discriminant and least square threshold selection," in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, 1978, pp. 592–596.

[32] J. Wu, P. Skip, D. Michael, and V. Marked, "Texture feature-based automated seeded region growing in abdominal mri segmentation," in *Proc IEEE International Conference on Biomedical Engineering and Informatics*, vol. 2. Sanya, China: IEEE, 2008, pp. 263–267.

[33] L. Belaid and M. Walid, "Image segmentation: a watershed transformation algorithm," *Image Analysis Stereology*, vol. 28, pp. 93–102, 2009.

[34] G. Coleman and H. Andrews, "Image segmentation by clustering," in *Proc. IEEE*, vol. 5, 1979, pp. 773–85.

[35] J. Bezdek, L. Hall, and L. Clarke, "Review of mr image segmentation techniques using pattern recognition," *Med. Phys.*, vol. 20, pp. 1033–48, 1993.

[36] T. Lei and W. Sewchand, "Statistical approach to x-ray ct imaging and its applications in image analysis. ii. a new stochastic model-based image segmentation technique for x-ray ct image," *IEEE Trans. Med. Imaging*, vol. 11, no. 1, pp. 62–69, 1992.

[37] A. Jain and R. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[38] L. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–53, 1965.

[39] P. Sivakumar and D. S. Meenakshi, "A review on image segmentation techniques," 2016.

[40] R. Kirti and A. Bhatnagar, "Image segmentation using canny edge detection technique," *International Journal of Techno-Management Research*, vol. 04, no. 04, pp. 8–14, 2017.

[41] S. Abdulateef and M. Salman, "A comprehensive review of image segmentation techniques," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 17, no. 2, p. 166–175, Dec. 2021.

[42] J. Zhou, K. L. Chan, P. Xu, and V. F. H. Chong, "Nasopharyngeal carcinoma lesion segmentation from mr images by support vector machine," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, Arlington, VA, USA, 2006, pp. 1364–1367.

[43] C. Leistner, A. Saffari, J. Santner, and H. Bischof, "Semi-supervised

random forests," in *2009 IEEE 12th International Conference on Computer Vision*, Kyoto, 2009, pp. 506–513.

[44] M. Li and Z. H. Zhou, "Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples," *IEEE Trans. Syst. Man Cybern.*, vol. 37, pp. 1088–1098, 2007.

[45] L. Gu, Y. Zheng, R. Bise, I. Sato, N. Imanishi, and S. Aiso, "Semi-supervised learning for biomedical image segmentation via forest oriented super pixels(voxels)," in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds. Quebec City, QC: Springer International Publishing, 2017, pp. 702–710.

[46] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, 2016, pp. 565–571.

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., Munich, Germany, October 2015, pp. 234–241.

[48] M.-M. Salem, A. Atef, A. Salah, and M. Shams, "Recent survey on medical image segmentation," *Comput Vis Concepts Methodol Tools Appl*, vol. 196953, pp. 129–169, 2018.

[49] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, Dec 2015, pp. 1520–1528.

[50] F. Milletari, N. Navab, and S.-A. Ahmadi, "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: IEEE, 2016, pp. 565–571.

[51] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014.

[52] R. B. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083

[53] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[54] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 2980–2988.

[55] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[56] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018, pp. 801–818.

[57] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv: Computer Vision and Pattern Recognition*, vol. abs/1706.05587, 2017. [Online]. Available: https://arxiv.org/abs/1706.05587

[58] H. Harkat, J. Nascimento, and A. Bernardino, "Fire segmentation using a deeplabv3+ architecture," in *Image and Signal Processing for Remote Sensing XXVI*, vol. 11533, 2020.

[59] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[60] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[61] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[62] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[63] R. Rashid, M. U. Akram, and T. Hassan, "Fully convolutional neural network for lungs segmentation from chest x-rays," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, A. Campilho, F. Karray, and B. Ter Haar Romeny, Eds., vol. 10882. Springer International Publishing, 2018, pp. 71–80.

[64] L. Saidy and C. C. Lee, "Chest x-ray image segmentation using encoder-decoder convolutional network," in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2018.

[65] A. Mittal, R. Hooda, and S. J. W. P. C. Sofat, "Lf-segnet: A fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs," *Wireless Personal Communications*, vol. 101, pp. 511–529, 2018.

[66] M. N. Saad, Z. Muda, N. S. Ashaari, and H. A. Hamid, "Image segmentation for lung region in chest x-ray images using edge detection and morphology," in *2014 IEEE International Conference on Control System, Computing and Engineering*, Penang, Malaysia, 2014.

[67] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2019, pp. 225–2255.

[68] O. Oktay *et al.*, "Attention unet: Learning where to look for the pancreas," *Medical Image Analysis*, vol. 53, p. 2, 2019.

[69] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[70] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537*, 2021.

[71] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu, "Automatic lung segmentation in chest x-ray images using improved u-net," *Scientific Reports*, vol. 12, no. 1, p. 8649, 2022.

[72] I. Ullah, F. Ali, B. Shah, S. El-Sappagh, T. Abuhmed, and S.-H. Park, "A deep learning-based dual encoder–decoder framework for anatomical structure segmentation in chest x-ray images," *Scientific Reports*, vol. 13, no. 1, p. 791, 2023.

[73] R. Ghali and M. A. Akhloufi, "Vision transformers for lung segmentation on cxr images," *SN Computer Science*, vol. 4, no. 4, p. 414, May 2023.

[74] Y. Li, T. Yao, Y. Pan, and T. Mei, "Contextual transformer networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, p. 3552, 2022.

[75] B. Chen, Y. Liu, Z. Zhang, G. Lu, and A. Kong, "Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation," *arXiv*, vol. arXiv:2107.05274, 2021.

[76] F. Yuan, Z. Zhang, and Z. Fang, "An effective cnn and transformer complementary network for medical image segmentation," *Pattern Recognition*, vol. 136, p. 109228, 2023.

[77] A. Vaswani *et al.*, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2023.

[78] J. Chen *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[79] Y. Zhang *et al.*, "Transfuse: Fusing transformers and cnns for medical image segmentation," *arXiv preprint arXiv:2102.08005*, 2021.

[80] J. M. J. Valanarasu *et al.*, "Medical transformer: Gated axial-attention for medical image segmentation," *arXiv preprint arXiv:2102.10662*, 2021.

[81] A. Lin *et al.*, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," *arXiv preprint arXiv:2106.06716*, 2021.

[82] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.

Hanan SABBAR is currently a student researcher in medical imaging at the Faculty of Science, Chouaib Doukkali University in El Jadida, Morocco. Their research interests include the application of artificial intelligence in the medical field.

Hassan SILKAN is currently a research professor at the LAROSERIE laboratory, Chouaib Doukkali University in El Jadida, Morocco. His research focuses on artificial intelligence, advanced signal processing, and image analysis. He has also applied artificial intelligence techniques to detect, segment, and classify medical images.

Khalid ABBAD is currently a research professor the Intelligent Systems & Applications Laboratory Science Department at Sidi Mohamed Ben Abdellah University, Fèz, Morocco. His research focuses on artificial, advanced signal processing, and image analysis. He has also applied machine learning and other artificial intelligence techniques to analyze medical images.

El Mehdi Bellfkih has recently completed his PhD at Hassan II University, specializing in coding theory and artificial intelligence. His research focuses on developing innovative solutions and methodologies in these fields. He can be reached at elmehdi.bellfkih@gmail.com

Imrane Chems eddine Idrissi is currently a PhD holder and a researcher in machine learning and error-correcting codes at Hassan II University. His research involves applying machine learning techniques.