Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 1, February 2026, pp. 424~437
ISSN: 2302-9285, DOI: 10.11591/eei.v15i1.10839

Deep learning architectures for medical image segmentation: an organized analysis of CNN-based models and uses

Cherifa Abdellaoui[1,3], Samia Belkacem[2], Noureddine Messaoudi[1]
[1]LIST Laboratory, Department of Electrical Systems Engineering, Faculty of Technology, University of Boumerdes, Boumerdès, Algeria
[2]Department of Electrical Systems Engineering, Faculty of Technology, University of Boumerdes, Boumerdès, Algeria
[3]Research Center in Industrial Technologies (CRTI), Signal Processing and Imagery Division, Embedded System Laboratory, Algiers, Algeria

Article Info

Article history:

Received Jun 11, 2025
Revised Oct 12, 2025
Accepted Dec 6, 2025

Keywords:

Convolutional neural network
Deep learning
Medical image
Object detection
Segmentation

ABSTRACT

Numerous techniques, especially those based on deep learning (DL), have been developed and applied to a wide range of tasks, including image recognition, classification, object detection, and image segmentation, as a result of extensive research in the field of image processing. Image processing has become crucial in the medical field, with segmentation emerging as a crucial method for organ identification, disease detection, and abnormality analysis in medical images. Convolutional neural networks (CNNs), one of the many approaches, have recently demonstrated great promise in resolving intricate problems associated with medical image analysis because of their capacity to automatically learn hierarchical features. In this review, we discuss recent developments in deep CNNs for medical image segmentation. The architectures and features of the most popular CNN-based models are examined, along with the different publicly accessible medical imaging datasets that are used in studies and the evaluation metrics that are frequently used to gauge segmentation performance and accuracy, also the advantages and disadvantages of each one. In addition, we look at comparative research and the shortcomings of existing approaches, offering suggestions for future lines of clinical relevance.

Corresponding Author:

Cherifa Abdellaoui
LIST Laboratory, Department of Electrical Systems Engineering, Faculty of Technology
University of Boumerdes
Frantz Fanon City, Avenue de l’indépendance, Boumerdès 35000, Boumerdès, Algeria
Email: c.abdellaoui@univ-boumerdes.dz

1. INTRODUCTION

Data generated by medical imaging studies already spread with an extremely high pace in the last years due to the fast evolution of medical imaging technologies and their high utilization. Commitment to diversity in health care. The breadth and high value of these data make them invaluable resources in healthcare. With continuous attempts to incorporate new theories and technologies to create a general segmentation algorithm applicable to different kinds of images, image segmentation techniques are currently moving toward increased speed and accuracy [1]. Image segmentation is a key methodology in these methods, that aims at partitioning the image into homogeneous and meaningful regions, and can be considered as a first step to enable automatic or semi-automatic delineation or analysis of anatomical structures. This step is crucial as it permits to accurately extract important elements such as organs, skin and tumors, providing the physician with valid and reliable means in both diagnosis, therapeutic follow-up and medical research [2].

In recent years, deep learning (DL) has become a transformative force in medical image segmentation, driving significant improvements in both accuracy and efficiency. Imaging techniques as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), X-ray, ultrasound as well as RGB-based methods such as retinal imaging and microscopy have gained increasing application in clinical practice. These imaging modalities play a pivotal role in the diagnosis of disease, with CT and related modalities being the cornerstone for the clinicians' approach to patients [3].

Especially convolutional neural networks (CNNs) have shown strong capability in feature extraction for the segments both organs and lesions [4] and fully convolutional network (FCN) is an important DL architecture for medical image segmentation which replace the fully connected layers in CNN with convolutional layers and has the merits of flexible input sizes and pixel- wise prediction for such tasks as organ and lesion delineation [5]. U-Net developed an encoder–decoder architecture with skip connections and has proven very effective for biomedical image segmentation, especially with limited-annotated data [6]. In addition to these, Bernal *et al*. [7] have demonstrated the usefulness of region-based convolutional neural network (R-CNN) in medical imaging applications with sequential data, such as temporal variation in videos for surgical workflow analysis or tracking of structures through image slices, and other works have shown further advances of generative adversarial networks (GANs) in segmentation refinement through adversarial learning [8]. Medical image autoencoders trained on normal data with only perceptual loss can detect subtle anomalies by identifying content dissimilarity rather than reconstructing at pixel-level [9]. Hybrid techniques that combine DL and conventional segmentation methods have been recently introduced to tackle challenges such as weak edges, noise, and varying intensity distributions, providing better performance in robustness to imaging modalities [10]. This review summarizes these developments and outlines the future directions in medical image segmentation for medical image analysis as well as clinical reparation.

This paper aims is to present a review of scientific literature about the recent approaches on DL for medical image segmentation to address the limitations and the research challenges on this field. It presents state-of-art techniques and answers what are the key challenges that need to be addressed for DL to be more effective for medical image segmentation. The review is strictly casted on the last guys developments with respect to the good old ones appeared in the previous. ensuring the coverage of how DL put its feet in the field here. The remainder of this review is organized as follows. The general concepts in DL are given in section 2. Note that this section could be skipped by an experienced reader. After this section, the state-of-the-art of medical imaging are surveyed and analyzed regarding advantages and disadvantages in section 3. In section 4, evaluations and comparisons of the works are performed, based on the reported numerical results. We wrap up this work with a discussion that points to potential directions for the field in section 5. Section 6 provides the conclusion of this study.

2. DEEP LEARNING

2.1. Overview of deep neural network for image segmentation

CNN's structure is growing increasingly complex and varied as a result of the DL technology's quick advancement. As a result, it gradually supplants conventional machine learning techniques [11]. The advancement of image recognition tasks, especially in the area of medical image analysis, has been greatly impacted by the development of deep CNNs. Even though the fundamentals of neural networks were established in the 1980s and 1990s, Krizhevsky *et al*. [12] creation of AlexNet marked the beginning of deep CNNs as the most widely used technique in computer vision. This framework achieved ground-breaking results in the ImageNet large-scale visual recognition test, proving that deep networks could be trained with GPUs and large datasets [12]. In domain of processing image DL is a much known technique that has been usedin lot of articles.

A basic task in computer vision is image segmentation, which entails dividing an image into meaningful regions that frequently correspond to distinct objects or structures. The accuracy and efficiency of image segmentation tasks have been greatly improved by deep neural networks (DNNs), especially CNNs.

2.2. Convolutional neural network

CNNs are a subgroup of DNNs however CNNs specialize in processing grid like data, which images are. They are really good at applying multiple layers of learnable filters and operations to input data-sets to capture spatial hierarchies. Due to the rapid progress in DL technology, the structure of CNN is becoming more and more complex and diverse. Consequently, it gradually replaces the traditional machine learning methods [13].

Architecture of a CNN is shown in Figure 1 [13]. It leverages a sequence of convolutional and pooling layers to transform the raw input image and to generate hierarchical feature representations which are then used for classification and other purposes. illustrating convolution pooling and fully connected layers to be created from raw image input to a feature generator output [14]. This architecture is at the root of many of the DL models employed for modality-based imaging analysis.

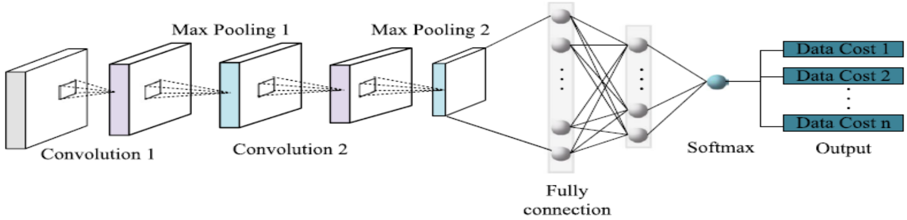

Figure 1. General architecture of a CNN

All of the fundamental units in the CNN architecture play a different role for processing and learning from the image data. These include fully connected layers to make decisions, pooling layers to reduce dimension, and convolutional layers to extract features. Additional components (e.g., batch normalization and dropout) are often appended to increase training stability and model generalization.

2.3. Application of deep learning in medical image segmentation for various human organ

In human body, there are many what-a-ma-call-it s. All of these elements are different by their nature. For example, there is a large area of segmentation required for diagnosis of lung nodule and brain tumor, and the vascular segmentation is required for retinal blood images. For the latter, more accurate segmentation is required. To improve segmentation accuracy, researchers adopt ideas within these messages and develop the segmentation techniques for different organs. Here, we will describe the optimal approach to segment different types of organs. We extracted segmentation methods of brain, eyes, chest, abdomen, and heart, from literature reading [15]. The CNNs structures generally used for image segmentation have different methods as Fully CNN, Regional CNN, U-Net, V-Net, R-CNN, DeepLab, and semantic segmentation [16].

Medical image segmentation, an essential operation in medical imaging for pixel-wise labeling of anatomical structures, is paramount to enable accurate diagnosis, treatment planning, and quantitative analysis. Segmentation determines the class of each pixel, with structural and spatial context, as opposed to image-level classification of whole images. U-Net (Ronneberger *et al.* [6]), a symmetric encoder-decoder network with U-shaped connections has demonstrated excellent performance across various modalities including medical images and it is also designed particularly for biomedical image segmentation with scarce training samples. For the diagnosis and treatment planning of tumor, accurate brain tumor segmentation in MRI is necessary, but it remains challenging due to tumor heterogeneity, indistinct boundary, and intensity variation. Research by Zhao *et al.* [17] proposes a deep leaning model which combines conditional random fields (CRFs) and fully convolutional neural networks (FCNNs) to enhance spatial and appearance consistency of segmentation. After training 2D patches across axes and views using FCNNs, the model employs CRFs in the form of CRF-RNNs for global optimization. Other advancements were achieved by extending U-Net with attention mechanisms, an example is presented in Oktay *et al.* [18] attention U-Net, which helps the model focus on relevant regions in complex anatomical scenarios. Research by Zhang *et al.* [19], mitigates these limitations by proposing TransFuse, a parallel-in-branch architecture that integrates CNNs and transformers. TransFuse takes advantage of a two-branch architecture: one branch of transformer to record long-range dependencies and another branch of shallow CNN to preserve fine-grained spatial details. There is an effective global-local feature synthesis from both branches by the BiFusion module through hierarchical multi-scale feature fusion with bi-channel-wise, spatial-wise, and Hadamard product wise attentions. Evaluated across a number of 2D and 3D medical image analysis tasks prostate, hip, skin lesion, and polyp.

3. SUPERVISED LEARNING ALGORITHMS FOR MEDICAL IMAGE SEGMENTATION

3.1. Fully convolutional networks

Specifically designed for high-density prediction tasks like image segmentation, FCNs are a variant of CNNs. The concept of FCNs was first proposed by Shelhamer *et al.* [20], who concentrated on image segmentation. In contrast to conventional CNNs, which employ fully connected layers for classification, FCNs do away with these layers and only employ convolutional layers, making pixel-level predictions possible. Standard convolutional layers are used to identify the captured image's attributes at the beginning of the process. These layers detect a wide range of patterns, from simple features like contours to more complex shapes in deeper layers. After convolution, pooling layers reduce the feature maps' spatial dimensions, enabling effective image down sampling while condensing the key features. Examined are fully convolutional networks' learning and inference procedures. By limiting the input to just the foreground, just the background, or just the shape, masking experiments are carried out to examine the functions of context and shape. To determine whether learning a background classifier is required for semantic segmentation, a "null" background model is defined. To further optimize whole-image learning, a detailed approximation between batch size and momentum is presented. Lastly, task accuracy bounds for specified output resolutions are measured to show that notable gains are still possible, Figure 2 show the structure of FCN.

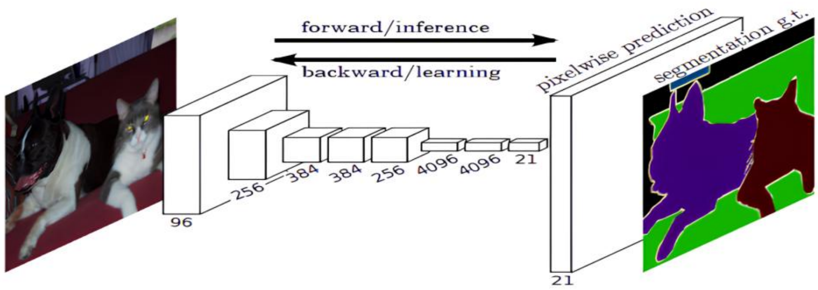

Figure 2. The structure of the FCN

3.2. Regional convolutional neural network

R-CNN is a DL architecture developed based on pairing up CNNs with region proposal algorithms for object detection. It actually requires something in order to work: region proposals – finding candidate regions (bounding boxes) in an image that could potentially contain an object using something like selective search. Feature extraction generating fixed length feature vector passing each region through CNN. Categorizing things: label each region by a set of class-specific SVM classifiers. Location refinement: Separate bounding box regressor to refine the accuracy in localization of the positive anchors [21].

3.2.1. Fast regional convolutional neural network

Researcher subsequently improved the R-CNN model and developed a new model Fast R-CNN to solve some problems that are left by the original R-CNN algorithm Figure 3. This improved method reduces the processing pipeline by feeding region proposals directly into the network. Shared convolutional feature map is generated by passing the entire input image through a CNN. A region of interest (RoI) pooling layer [9] on this feature map is subsequently used for refining the projected region proposals. These representations are then passed over fully connected layers. Finally, a SoftMax layer is employed for predicting bounding-box and classification scores. Fast R-CNN In particular, a key benefit of Fast R-CNN is that it eliminates the need to run the CNN forwards for each of the around 2,000 region proposals. Instead, the feature map is obtained in a single forward pass of the whole image, dramatically improving computational efficiency [22].

3.2.2. Faster regional convolutional neural network

One of the most popular DL approaches for extracting image detections is the Faster R-CNN, which includes a region proposal network (RPN) on top of the convolutional features layers to speed up the computational operation. With the generation of region proposals built into the network, the RPN gets rid of the selective search process in previous R-CNN implementations. For Faster R-CNN, each of the input image has to resized to a fixed size before processing. Each image is passed through a convolutional network, inspired by VGG-16, which has 13 layers with convolution and 3 fully connected layers in total. As we know the classification accuracy of VGG-16 perform better than ZF-Net [23]. 3×3 convolutional kernels are used in the architecture of VGG-16. The model generates nine anchors at each location on the convolutional layers with three scales (128^2, 256^2, and 512^2 pixels) and three aspect ratios (1:1, 1:2, and 2:1). These anchors will generate 512×9 pixel feature maps, and therefore the sizes and shapes of the bounding boxes are adjusted with the sizes and shapes of the objects in the image. Then these feature maps go into the fully connected layers and the RoI pooling layer. A class probability is computed by a SoftMax layer and a two class scores are output which are object and background. In addition, the bounding box regressor estimates the coordinates and size of the detected object. This final output, called the proposal region, is then input to the detection module [23].

3.2.3. Mask regional convolutional neural network

Mask R-CNN is a powerful instance segmentation model; it extends the Faster R-CNN by adding a branch for predicting segmentation masks on each RoI, in parallel with the existing branch for classification and bounding box regression. It performs in two main stages: first, the input image is analyzed to detect candidate regions, and in the second, regions are classified and appropriate segmentation masks and bounding boxes are created. For instance-level object detection and segmentation, Mask R-CNN is a model that is able to both detect object classes and segment objects at the pixel level. Its architecture is a Fast R-CNN-based network that incorporates enhanced components such as the ROIAlign algorithm to improve the localization, and FPN to produce multi-scale features. The 6 major components of the Mask R-CNN model consist of the ROIAlign module, the RPN, input processing, the feature extraction backbone, the FPN, and the output heads for mask prediction, classification, and bounding box regression (box, class, and mask) [24]. The Mask R-CNN architecture is illustrated in a schema in Figure 3.

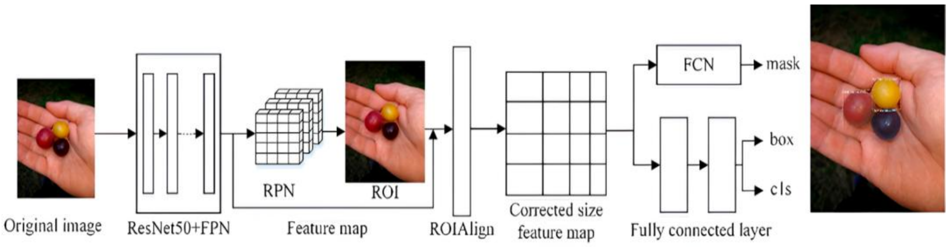

Figure 3. The structure of R-CNN

3.3. DeepLab

Several versions of the DeepLab technique, including DeepLabv1, DeepLabv2, DeepLabv3, and DeepLabv3+, have been proposed. The supplied image passes through a deep CNN layer with one or two atrous convolution layers in DeepLabv1 [25]. In this way, it produces an incomplete feature map which is then upsampled to the size of the original image using a bilinear interpolation technique. Before obtaining a segmented image, the entire CRF is connected using interpolated data.

The DeepLabv2 model uses several attenuated convolutions on the input feature map with different expansion rates. Outputs are merged and objects at different scales are segmented to capture information from a large effective field with fewer parameters and less computational complexity [25].

DeepLabv3 [26] is an extension to DeepLabv2 that has added image-level functionality to the atrous spatial pyramid pooling (ASPP) module. To easily form a network, it uses batch normalization. DeepLabv3's ASPP module is combined with an encoder and decoder structure in the DeepLabv3+ model. For faster computation, the model also uses at reuse and depth-separable convolution. Low-level and high-level functionalities corresponding to structural details and semantic information are combined in the decoder section.

A decoding module and an encoding module are included in DeepLabv3+ [27]. Using an atrous convolutional network and backbone network such as MobileNetv2, PNASNet, ResNet, and Xception, the encoder path extracts the required information from an input image. Using the information from the input image, the decoder path reconstructs the output in the appropriate dimensions.

3.4. U-Net

For per-pixel prediction, U-Net is a popular image segmentation technique. A U-channel and an ignored connection make up U-Net [16]. The structure of the U-channel is similar to that of the SegNet encoder-decoder. Each of the encoder's four submodules contains two convolutional layers. There is a maximum pool for sub-sampling after the sub-modules [7]. The decoder consists of four submodules. Oversampling progressively increases resolution. It then makes predictions for each pixel. This network comprises only convolution and subsampling, rather than a fully connected layer. The result of oversampling is connected to the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder via an ignored network connection [15]. Figure 4 show structure of U-Net [28].

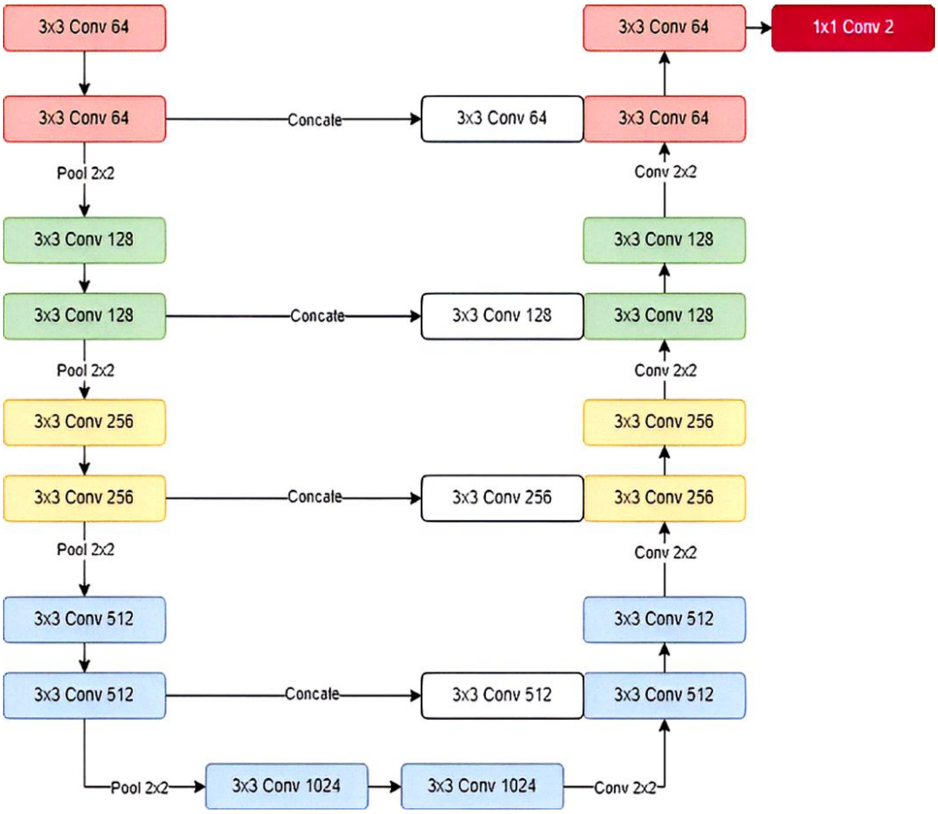

Figure 4. The structure of U-Net

3.5. V-Net

A DL network known as V-Net was specially designed for 3D medical imaging data, such as CT and MRI images. V-Net also cautiously considers the issue of complex semantic oriented boundary information, so it also makes some architecture changes, and the V-Net model still retains the overall framework of the U-Net model. It employs strides for down sampling of convolutional layers in place of pooling layers. We make such design choice so as to reduce the dimension of volumetric features and alleviate the bottleneck problem in the training process. Due to these structural differences, the V-Net was particularly effective in dealing with the challenges in volumetric medical image segmentation [29]. The V-Net is a widely used 3D CNN in biomedical imaging analysis. The high resolution feature representations produced by its encoder offers the precise localization of target structures. To refine the final segmentation result, in this work, the model establishes skip connections between these hi-resolution features and decoder outputs. V-Net can effectively overcome the well-known issue of class imbalance, i.e. a huge difference between the pixel numbers of the foreground and background, in medical imaging [30], as shown in Figure 5.

3.6. Other U-Net structures

This has resulted in various extensions to the U-Net architecture over the years to enhance segmentation accuracy and automatically learning rich features across different medical imaging tasks. To reduce the semantic gap between encoder and decoder feature map. To enhance the performance for images with complex or noisy backgrounds, another model attention U-Net [18] employs attention gate to guide the network to concentrate on important part of the input. In order to bridge the semantic gap between encoder and decoder feature maps and enable more precise medical image segmentation, UNet++ [31] introduces nested and dense skip pathways. Additionally, it integrates deep supervision, which allows for optional model pruning to enable both high-accuracy and high-speed inference modes. Residual U-Net [32] (incorporating residual connections within its convolutional (conv) blocks) allows for ease of augmenting network depth in order to facilitate a smooth learning of gradients. An effective CT and MRI volume segmentation is made available by 3D U-Net [33], the 2D U-Net is extended to three dimensions for volumetric data. The motivation behind these modifications is to enhance robustness, localization accuracy and feature representation across medical image segmentation tasks. Additionally, the U-Net architecture is enhanced with dense connectivity as that in the DenseNet architecture in the dense U-Net [34]. This configuration is designed to promote the reuse of features; while also mitigating the vanishing gradient problem by having each block inner-layers take input from all the previous layers. Therefore, the medical image segmentation network is more robust and parameter-saving. With respect to the recurrent residual U-Net (R2U-Net) [35], which uses recurrent convolutional layers and residual learning in U-Net. By enabling network to preserve spatial dependencies over multiple layers, this design enhances the representation of contextual features, and enhances segmentation accuracy in complex medical images.

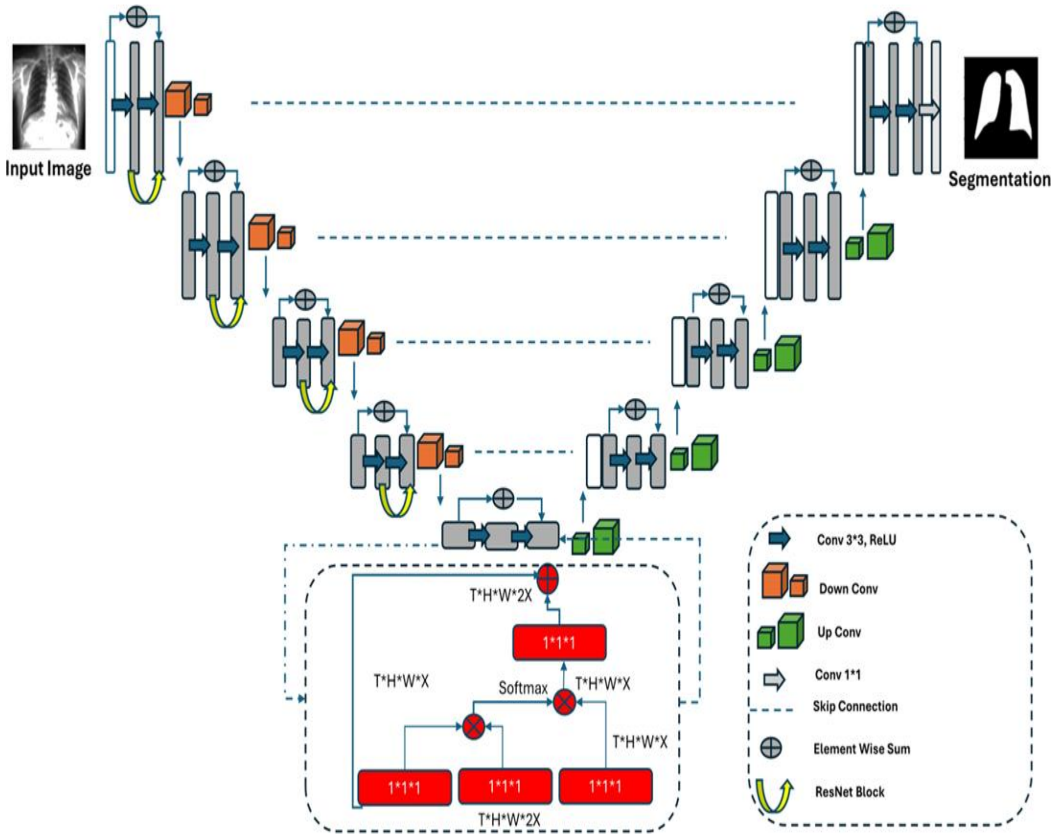

Figure 5. The structure of V-Net

3.7. Generative adversarial networks

GANs is composed of two competing networks: a generator and a discriminator tasked to generate realistic data. The generator generates the samples that look like the real ones, and the discriminator differentiates the real from the fake, and as such, GANs can learn complex data distributions. This adversarial training has been shown to be effective in medical imaging, plagued by data scarcity and low quality. GAN models can be utilized to synthesize images for complementing scarce data, restore high-resolution details, improve segmentation performance, and identify anomalies in representing normal data distributions in the diagnosis and treatment assistance [36]. GANs are a popular family of DL models which comprise of a generator and a discriminator network competing in an adversarial game to generate highly realistic synthetic data. In medical imaging, GANs are of particular interest to solve data scarcity, class-imbalance, and privacy issue problems, where high-quality synthetic data are generated for data augmentation, segmentation, reconstruction, and domain translation [37].

4. MEDICAL IMAGE SEGMENTATION BASED ON DEEP LEARNING

Under a given input image of DL, includes a lot of application areas of image, objects, and so on, including object detection, image segmentation, image recognition, classification and the like. Medical image segmentation is also a widely studied topic, given that it plays a critical role in computerized diagnosis and treatment planning. Several imaging modalities have been utilized for the studies, such as CT, ultrasound (US), X-ray and MRI, and each one of them is suitable for a specific clinical context. Depending on the application, various organs and anatomical structures such as brain, lungs, polyp, heart, and tumors are each aimed to be segmented. Each study commonly relies on a segmentation method to solve the problem. The proposed methods are also evaluated with Clinical data or public datasets and quantitative measurements such as Hausdorff distance (HD), precision and recall, intersection over union (IoU), and dice similarity coefficient are used to verify the effectiveness and accuracy of the proposed algorithms. The challenging and variable nature of medical image segmentation can be assessed through studies that demonstrate the complexity and adaptability required for successful solutions.

4.1. State of the art

Research by Tseng *et al.* [38] an semi-supervised CNN bridging dense block with U-Net architecture for cardiac MRI segmentation named DNetUnet were introduced and it was trained using ACDC2017 dataset. By introducing GAN architecture on unlabeled data samples, focal and boundary loss on imbalanced classes and knowledge distillation for mobile deployment, the model surpassed U-Net and ResUNet in hearts' structures segmentation task, and thus providing trustworthy automation for diagnosis of diseases such as myocardial infarction and cardiomyopathy. Cao *et al.* [39] proposed Swin-Unet, the first U-shaped architecture for medical image segmentation consisting only of transformer blocks, benefiting from hierarchical Swin Transformer blocks, shifted-window self-attention, and convolution-free up-sampling. When applied to Synapse CT and ACDC MRI data, it performed advancing of CNN and hybrid models, which demonstrates the success of pure transformer that at the cost of increased computation would hold as considerable potential for medical imaging project of the future. Yue *et al.* [40] developed an automatic DL pipeline for bronchiectasis identification on LDCT, in a workflow that consisted of image enhancement (ACER), automatic lung-lobe segmentation (RDU-Net), and lesion classification (HDC Mask R-CNN). Show that a system that was taught using 1992 annotated images was competitive in terms of accuracy, could produce output in approximately 1 second and showed promise as a fast and reliable tool to help radiologists detect bronchiectasis. Azad *et al.* [41] developed TransDeepLab, a transformer-modified architecture (based on DeepLabv3+) TransDeepLab autoregressively replaces CNN layers with Swin-Transformer blocks, and incorporates a Swin SPP module and cross-contextual attention; delivering better multi-scale fusion and boundary segmentation. It also obtained state-of-the-art performance on Synapse CT and skin lesion datasets, demonstrating that pure transformer structures can achieve quite useful representations of 3D volumes despite having to rely on pre-training with ImageNet and taking 2D as input. Sun *et al.* [42] introduced DA-TransUNet, which is a Transformer U-Net variant integrated with PAM and CAM attention modules in order to capture the spatial (PAM) and channel (CAM) characteristics and alleviate the gaps between the encoder and decoder and the redundancy of the filters. The network was tested on Synapse, CVC-ClinicDB, and ISIC2018 datasets where it achieved the state-of-the-art performance on organ, polyp, and skin lesion segmentation, providing accurate boundary localization with efficiently learned parameters but with additional architectural complexity. Bukhori *et al.* [28] has idea to predeveloped a paper regarding the application pose of breast cancer image segmentation based on CNN model to the U-Net architecture. In this paper, we focus on the easier task of diagnosing breast cancer in the mammogram as it is a challenging one because of its low contrast and high noise, and use ultrasound USG images to ensure a better accuracy. The proposed U-Net model consists of the encoder, bottleneck, and decoder components and is evaluated for various scenarios involving varying training to testing data ratio. It is noted that the best Mean IU attains the conclusion that the U-Net is effective for breast cancer segmentation. Innani *et al.* [43] presented EGAN and MGAN, both GAN-based unsupervised methods for skin lesion segmentation on ISIC 2018. EGAN adopts a squeeze-and-excitation encoder-decoder structure that combines a patchGAN discriminator with morphology-based loss for accurate segmentation, while MGAN integrates ASPP with MobileNetV2 for efficient deployment. Both EGAN and MGAN outperform CNN and transformer baselines, with EGAN mainly better in accuracy and MGAN often better in efficiency, although they face challenges in computational cost and cross-dataset generalization. Wan *et al.* [44] also use a glioma-special MRI segmentation architecture based on the improved DeepLabv3+ and RegNet, together with an attribute encoder for multimodal data fusion. Furthermore, achieving state-of-the-art results on the LGG dataset (94.36 Dice and 91.83 IoU) by utilizing cross-entropy, Dice, and outlier-region loss, we demonstrated the effectiveness of multimodal fusion and the use of sophisticated loss functions, leaving some room for improvement in multi-class and 3D segmentation. Qin *et al.* [45] proposed DB-SAM – a dual-branch variant of the Segment Anything Model for universal medical image segmentation. DB-SAM gains advantage over both SAM and MedSAM on as many as 30 segmentation tasks (87.05% Dice and 85.31% NSD) by integrating one ViT body for high-order features and a convolutional counterpart for domain-specific details with bilateral cross-attention for feature fusion. It is capable to efficiently process small or complex structures and multimodal data, however it remains problematic for 3D volume reconstruction and increased computational expenses. Iqbal *et al.* [46], which proposed TBConvL-Net, a hybrid CNN–ViT–BConvLSTM network for medical image segmentation. It employs CNNs for local detail extraction, transformers for global context, BConvLSTM for temporal feature learning, Swin blocks in skip connections and composite loss for boundary accuracy. It surpassed state-of-the-art methods in Dice, Jaccard and sensitivity on ten different datasets, while being lightweight and fast (19.1 ms inference), showing strong generalization and robustness. Ruan *et al.* [47] proposed VM-UNet, a segmentation model based on the Mamba state space model (SSM) framework with Visual State Space (VSS) blocks for effective long-range dependency modeling. By using a content specific asymmetric encoder decoder and efficient skip connections rather than convolution or self-attention, it outperforms state-of-the-art methods on ISIC17, ISIC18, and Synapase datasets and is much more parameter/ computation efficient. Although it is not good at capturing fine boundaries and large inputs, it serves as the first pure SSM-based baseline for medical image segmentation.

Table 1 compares 11 representative methods and highlights their salient features, advantages, and limitations to facilitate the selection of a suitable method by the researchers and the clinicians. We have selected a subset of these strategies for further study in this work based on the summarized knowledge for each method.

Table 1. Performance comparison of supervised medical image segmentation methods

Method	Core features	Advantage	Limitations
DNetUnet	Dense block feature extractor to preserve relevant features Semi-supervised learning using labelled and unlabeled data Focal loss and boundary loss	Superior segmentation accuracy, effective handling details, reduced training	High computational requirements for training, limited performance improvement
Swin-Unet	Pure transformer architecture Shifted window mechanism Patch expanding layer Hierarchical design	Global context modeling, high accuracy, edge preservation	Computational cost, pre-training dependency, and 3D data handling
Mask R-CNN	ACER image enhancement RDU-net segmentation HDC Mask R-CNN	High accuracy, fast processing, lobe-wise scoring	Limited dataset, hardware constraints, and tongue lobe not segmented separately.
TransDeepLab	Shifted window mechanism SSPP module	Multi-scale fusion and computational efficiency Generalizability: effective across modalities	Pre-training dependency 2D focus: not evaluated on 3D High-resolution demands
DA-TRANSUNET	Dual attention (DA-Blocks) Transformer integration Optimized skip connections	Superior accuracy Multi-scale feature fusion Generalizability: effective across diverse Robust skip connections	Computational cost Pre-training dependency
U-Net	Data augmentation Hyperparameter testing Mean IoU metric	High accuracy, Automated segmentation Robust to noise Flexible training	Dependence on data quality Moderate performance in some cases Limited generalization Computational cost
GAN	Unsupervised learning Boundary-aware segmentation Multi-scale feature extraction Computational efficiency	High accuracy Data efficiency Real-time deployment Robustness	Computational cost Generalizability Dependence on ISIC data
DeepLabv3+	Multimodal fusion Advanced loss functions ASPP module	Higher accuracy Noise reduction Multimodal integration Efficiency: uses 2D CNNs (faster than 3D CNNs)	Small dataset Transformer underperformance Noise size dependency No 3D segmentation
DB-SAM	Dual-branch encoder (ViT+CNN) Channel attention blocks in ViT to enhance medical-specific features Efficient adaptation	Superior accuracy Handles complex structures Multimodal compatibility	2D processing for 3D data Dependence on bounding box prompts
TBConvL-Net	Depth-wise separable convolutions Dense connections BConvLSTM in skip connections Swin Transformer blocks Composite loss function Transfer learning	High segmentation accuracy Efficient computation Robust feature learning Strong generalization	Transformer component still computationally heavier than pure CNNs Requires moderate hardware Preprocessing needed No 3D segmentation
VM-UNet	Pure SSM-based design VSS SS2D module Asymmetric encoder-decoder Simple skip connections	High segmentation accuracy Linear computational complexity Parameter-efficient Strong generalization Robust in complex scenarios	Sensitivity to fine details Performance degradation at higher resolutions Generalization limitations of SSMs No attention or multi-scale fusion

Table 2 provides a comparison of the 11 selected methods, with emphasis on four key aspects: reference source, imaging modalities utilized, anatomical regions segmented, and evaluation metrics considered. Through these facilitating factors like Dice coefficients, accuracy, and sensitivity, these schemes offer a complete evaluation of the quality of segmentation in various medical imaging domains. Collecting this information, Table 2 enables a better comparison between the effectiveness and appropriateness of these methodologies for certain medical imaging applications.

Table 2. Performances evaluation of a list of papers

Ref	Modalities	Segmented area	Evaluations
Tseng *et al.* 2020 [38]	Cardiac MRI images	Segmenting left ventricle, right ventricle, and myocardium	Achieves dice coefficients of 0.942 (LV), 0.908 (RV), and 0.894 (MYO) Outperforms existing methods like U-Net, ResUNet, and DenseUNet
Cao *et al.* 2021 [39]	Multi-organ CT/Cardiac MRI	Aorta, gallbladder, kidneys, liver, pancreas, spleen, stomach/Left/right ventricles (LV/RV), myocardium (MYO)	Synapse dataset: 79.13% dice similarity coefficient (DSC), 21.55 HD ACDC dataset: 90.00% DSC (best in LV/MYO segmentation)
Yue *et al.* 2022 [40]	CT	Lung lobes	Classification accuracy: 91.4% IoU: 88.8% Sensitivity: 88.6% Specificity: 85.4% Segmentation IoU (RDU-Net): 98.3%
Azad *et al.* 2022 [41]	CT, MRI (cardiac), dermoscopy, (skin lesions)	Synapse dataset: 8 abdominal organs (aorta, liver, and pancreas) ACDC: cardiac structures (left/right ventricles, myocardium) Skin lesions: ISIC 2017/2018, PH² datasets	Synapse: 80.16% DSC, 21.25 HD Skin lesions: 92.39% Model efficiency: 21.14M SSPP levels: 2-level optimal for DSC; 3-level improves HD Cross-attention: +1% DSC over basic fusion.
Sun *et al.* 2024 [42]	CT (Synapse: abdominal organs) MRI (cardiac structures). X-ray (tuberculosis detection) Endoscopy/dermoscopy (polyps, skin lesions)	Organs: liver, pancreas, kidneys, Lesions: polyps (CVC-ClinicDB), skin cancer (ISIC2018) Pathologies: tuberculosis (Chest X-ray)	DSC: primary overlap accuracy (79.80% on synapse) HD: Boundary precision (23.48 mm on synapse)
Bukhori *et al.* 2023 [28]	Ultrasound (USG) images	Breast tumor	Primary metric: mean IoU
Innani *et al.* 2023 [43]	Dermoscopic images	Skin lesions	Dice coefficient (90.1% for EGAN), Jaccard index (83.6%), Accuracy (94.5%) Sensitivity (93.6%) and specificity (95.5%) for clinical relevance Speed MGAN achieves 13 FPS
Wan *et al.* 2023 [44]	MRI	Brain tumor	DSC: 94.36, IoU: 91.83
Qin *et al.* 2024 [45]	3D: MRI 2D: X-ray, ultrasound, endoscopy, retinal imaging, pathology	Brain: ventricles, tumors, cerebellum Abdomen: liver, pancreas, gallbladder, tumors Thorax: lungs, pleural effusion, heart ventricles, other: prostate, head-neck tumors, retinal vessels, colon glands	DSC: 87.05% (3D), 82.00% (2D), (NSD): 85.31% (3D), 91.81% (2D)
Iqbal *et al.* 2025 [46]	Optical imaging Ultrasound Whole slide imaging (WSI) X-ray – chest radiographs (MC) Fundus photography Microscopy MRI	Skin lesions Thyroid nodules Breast tumors Cell nuclei Optic discs Fluorescent cells Chest and brain abnormalities	Jaccard index (IoU) Dice coefficient Accuracy Sensitivity (recall) Specificity
Ruan *et al.* 2024 [47]	Dermoscopy images CT Images	Skin lesions Aorta, gallbladder, left/right kidneys, liver, pancreas, spleen, stomach	Mean intersection over union (mIoU) DSC Accuracy Sensitivity (recall) Specificity HD95 (95th percentile HD)

5. RESULTS AND DISCUSSION

In this review, we demonstrated an in-depth analysis of the performance and developmental journey of DL models applied to supervised segmentation tasks in medical imaging. Even though most of the previous works heavily investigated the efficiency of CNN-based architectures, including but not limited to U-Net, for achievements in specific fields, relatively few studies were dedicated to these networks' clear disadvantages in terms of long-range modeling, low robustness over multiple domains, and real-life applicability. Therefore, one can outline three methodological waves: updated CNN-based architectures, transformer-based models, and developing generalists. Our examination of 11 seminal works demonstrates a direct pattern. The first group, CNN-based models with densely connected/atrous convolutions and specific losses, demonstrate a leading Dice score on small and specific tasks, specifically, 90.18% on the skin lesions dataset with 93.5% accuracy, proving limited effectiveness in complicated scenarios. By contrast, the transformer-based models, Swin-Inst, included DA-TransUNet, confirmed an undoubted advantage in multi-organ tasks and the ACDC cardiac dataset via 90.62% DSC with one head and 79.13% DSC on Synapse. The third wave of models, including state-space architectures like VM-UNet and prompted networks, gave a boost to the real-world deployment and explanation capacity, while losing a battle of generality. Our results indicate that the high level of segmentation quality can be achieved without DL if it is assigned a certain task or issue for local smoothing. For example, boundary-aware loss, included in many articles, increases the segmentation of tiny and irregular bodies, a significant issue of early CNN-based models. The success of the transformer-based models from Swin, Naive, and DA could be easily explained by global context understanding in terms of shifted windows and dual mechanisms, emphasizing the current-voltage oncell masks, which is the same as witnessed by Cao *et al*. [39] and Azad *et al*. [41]. who forgot to mention the massive uptrend of HD. Nevertheless, various paradigms GANs and SSMs demonstrate that the model benefits immensely from the unsupervised part or simplification without a serious cost to accuracy on less labeled tasks. While this study could analyze a broad number of different methods and fields, its result is derived from quick mentioned works, which do not provide a completely integrated analysis of proposed in the compare performance benchmarks. Thus, the need for additional research continues even with the reported limitations: transformer high computational cost, reshuffling of the structures in SSMs, and increased importance of pre-learning, particularly in challenging clinical settings and lower-quality image modalities.

Our findings indicate that the field is shifting towards an equilibrium among accuracy, efficiency, and generalization instead of optimizing a single metric. Forthcoming research could focus on multiple directions: i) cross-modality and multi-institutional benchmarking to inhibit networks from overfitting small and homogeneous cohort of data as well as ensure clinical relevance and portability across populations and equipment; ii) intelligible models that supply a clinical reasoning tool simultaneously with the mask, to give clinicians some degree of trust; and iii) realistic ways to seamlessly insert the tool in the pipeline of diagnosis and prognosis, not as an isolated block but as a part of a device with the specific role of helping the doctor.

Recent studies suggest that the trend of medical image segmentation is moving from tailored CNNs towards context-aware transformers and the direction of easy-to-use models. Our results provide clear evidence that this evolution is driven by a fundamental shift in design priority - the tradeoff between accuracy and computational demand/annotation cost, as opposed to a single architectural innovation. Model selection now becomes largely application-specific: CNNs in resource-constrained environments, transformers in context-driven tasks, and new models such as SSMs for scalable, efficient use.

6. CONCLUSION

In this article, we have presented the long way of DL for medical image segmentation from CNN-embedded architectures, through state-space models and other light-weight methodologies. Our results from the reviewed studies give rise to 3 major implications for developers and clinical practitioners. Firstly, the choice of the model should be application-oriented. CNN-based structures still have a marginal edge in low-data and strict-computational restricted regimes. In contrast, transformer and hybrid models are more appropriate for complex, multi-organ tasks where global contextual information is important. Second, the interest in only the highest accuracy on a benchmark is giving way to the demand for robustness, reproducibility, and validation on a variety of datasets and institutions. Third, computation and annotation efficiency are moving from an engineering convenience to a clinical necessary, for real-time or bedside integration, such as facilitated by light-weighted GANs and state-space models. For the algorithm development community, these results highlight the importance of developing relevant and deployable algorithms. Much future work will need to prioritize large scale cross-modality benchmarking, interpretable models that explain their reasoning, and development of segmentation tools directly in clinical diagnostic pathways. What this means to the clinician, we argue, is that while frontier models may trump in terms of numbered precision, their practicable strength, interpretability and capacity to survive the sheer diversity of real-world clinical imaging are the real benchmark. In this context, semi-supervised CNNs, promptable SAM-like models, and efficient SSMs can be seen as complementary directions to transfer research developments into useful instruments to be actually deployed in the radiology rooms. Finally, the field is going to iterate towards a more subtle trade-off between precision, brevity and interpretability. Striking this balance will require an unprecedented degree of collaboration among computer scientists, clinicians, and regulatory agencies to ensure that these advanced segmentation methods achieve not only state-of-the-art performance but also meaningful, quantifiable improvements in patient outcomes and clinical workflows.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Cherifa Abdellaoui	✓	✓	✓	✓	✓		✓	✓	✓	✓				
Samia Belkacem	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Noureddine Messaoudi	✓	✓		✓		✓	✓			✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019, doi: 10.1016/j.neucom.2019.02.003.

[2] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods," *Tomography*, vol. 11, no. 5, pp. 1–45, 2025, doi: 10.3390/tomography11050052.

[3] D. Shen, G. Wu, and H. Il Suk, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[4] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[5] S. Y. Huang, W. L. Hsu, R. J. Hsu, and D. W. Liu, "Fully Convolutional Network for the Semantic Segmentation of Medical Images: A Survey," *Diagnostics*, vol. 12, no. 11, pp. 1–21, Nov. 2022, doi: 10.3390/diagnostics12112765.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation (1)," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham: Springer international publishing, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[7] J. Bernal *et al.*, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review (1)," *Artificial Intelligence in Medicine*, vol. 95, pp. 64–81, Apr. 2019, doi: 10.1016/j.artmed.2018.08.008.

[8] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, p. 101552, Dec. 2019, doi: 10.1016/j.media.2019.101552.

[9] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, and D. V. Dylov, "Anomaly Detection in Medical Imaging with Deep Perceptual Autoencoders," *IEEE Access*, vol. 9, pp. 118571–118583, 2021, doi: 10.1109/ACCESS.2021.3107163.

[10] S. Graham *et al.*, "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, Dec. 2019, doi: 10.1016/j.media.2019.101563.

[11] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in*

Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.
[13] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, p. 99, Mar. 2024, doi: 10.1007/s10462-024-10721-6.
[14] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Information*, vol. 16, no. 3, p. 195, Mar. 2025, doi: 10.3390/info16030195.
[15] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods (2)," *Sustainability*, vol. 13, no. 3, pp. 1–29, Jan. 2021, doi: 10.3390/su13031224.
[16] C. Abdellaoui, S. Belkacem, N. Messaoudi, and R. Drai, "Medical Image Segmentation Based U-Net Architecture Implemented on Raspberry Pi4," in *2024 3rd International Conference on Advanced Electrical Engineering (ICAEE)*, Sidi-Bel-Abbes, Algeria: IEEE, Nov. 2024, pp. 1–5, doi: 10.1109/ICAEE61760.2024.10783388.
[17] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Medical Image Analysis*, vol. 43, pp. 98–111, Jan. 2018, doi: 10.1016/j.media.2017.10.002.
[18] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint*, 2018, doi: 10.48550/arXiv.1804.03999.
[19] Y. Zhang, H. Liu, and Q. Hu, "TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer International Publishing, 2021, pp. 14–24, doi: 10.1007/978-3-030-87193-2_2.
[20] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.
[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
[22] H. N. Huang *et al.*, "Image segmentation using transfer learning and Fast R-CNN for diabetic foot wound treatments," *Frontiers in Public Health*, vol. 10, pp. 1–15, Sep. 2022, doi: 10.3389/fpubh.2022.969846.
[23] S. Wahjuni, Wulandari, and H. Nurarifah, "Faster RCNN based leaf segmentation using stereo images," *Journal of Agriculture and Food Research*, vol. 11, pp. 1–7, Mar. 2023, doi: 10.1016/j.jafr.2023.100514.
[24] S. Wang, G. Sun, B. Zheng, and Y. Du, "A crop image segmentation and extraction algorithm based on mask RCNN," *Entropy*, vol. 23, no. 9, pp. 1–16, Sep. 2021, doi: 10.3390/e23091160.
[25] 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.
[26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv preprint*, 2017, doi: 10.48550/arXiv.1706.05587.
[27] H. Harkat, J. M. P. Nascimento, and A. Bernardino, "Fire segmentation using a DeepLabv3+ architecture," in *Image and Signal Processing for Remote Sensing XXVI*, Eds., SPIE, Sep. 2020, p. 20, doi: 10.1117/12.2573902.
[28] S. Bukhori, M. A. Bariiqy, W. E. Y. R, and J. A. Putra, "Segmentation of Breast Cancer using Convolutional Neural Network and U-Net Architecture," *Journal of Artificial Intelligence and Data Mining (JAIDM)*, vol. 11, no. 3, pp. 477–485, 2023, doi: 10.22044/jadm.2023.12676.2419.
[29] F. Turk and M. Kılıçaslan, "Lung image segmentation with improved U-Net, V-Net and Seg-Net techniques," *PeerJ Computer Science*, vol. 11, p. e2700, Feb. 2025, doi: 10.7717/PEERJ-CS.2700.
[30] Y. Zeng, P.-H. Tsui, W. Wu, Z. Zhou, and S. Wu, "Fetal Ultrasound Image Segmentation for Automatic Head Circumference Biometry Using Deeply Supervised Attention-Gated V-Net," *Journal of Digital Imaging*, vol. 34, no. 1, pp. 134–148, Feb. 2021, doi: 10.1007/s10278-020-00410-5.
[31] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.
[32] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, May 2018, doi: 10.1109/LGRS.2018.2802944.
[33] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, pp. 424–432, doi: 10.1007/978-3-319-46723-8_49.
[34] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017, pp. 11–19.
[35] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," *arXiv preprint*, 2018, doi: 10.48550/arXiv.1802.06955.
[36] S. Kazeminia *et al.*, "GANs for medical image analysis," *Artificial Intelligence in Medicine*, vol. 109, p. 101938, Sep. 2020, doi: 10.1016/j.artmed.2020.101938.
[37] D. N. Sindhura, R. M. Pai, S. N. Bhat, and M. M. M. Pai, "A review of deep learning and Generative Adversarial Networks applications in medical image analysis," *Multimedia Systems*, vol. 30, no. 3, p. 161, Jun. 2024, doi: 10.1007/s00530-024-01349-1.
[38] K. K. Tseng, R. Zhang, C. M. Chen, and M. M. Hassan, "DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service," *Journal of Supercomputing*, vol. 77, no. 4, pp. 3594–3615, Apr. 2021, doi: 10.1007/s11227-020-03407-7.
[39] H. Cao *et al.*, "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation," in *European Conference on Computer Vision*, Cham: Springer Nature Switzerland, 2023, pp. 205–218, doi: 10.1007/978-3-031-25066-8_9.
[40] N. Yue, J. Zhang, J. Zhao, Q. Zhang, X. Lin, and J. Yang, "Detection and Classification of Bronchiectasis Based on Improved Mask-RCNN," *Bioengineering*, vol. 9, no. 8, p. 359, Aug. 2022, doi: 10.3390/bioengineering9080359.
[41] R. Azad *et al.*, "TransDeepLab: Convolution-Free Transformer-Based DeepLab v3+ for Medical Image Segmentation," in *International Workshop on PRedictive Intelligence In MEdicine*, Cham: Springer Nature Switzerland, 2022, pp. 91–102, doi: 10.1007/978-3-031-16919-9_9.
[42] G. Sun *et al.*, "DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1398237, May 2024, doi: 10.3389/fbioe.2024.1398237.
[43] S. Innani *et al.*, "Generative adversarial networks based skin lesion segmentation," *Scientific Reports*, vol. 13, no. 1, p. 13467, Aug. 2023, doi: 10.1038/s41598-023-39648-8.

[44] B. Wan, B. Hu, M. Zhao, K. Li, and X. Ye, "Deep learning-based magnetic resonance image segmentation technique for application to glioma," *Frontiers in Medicine*, vol. 10, p. 1172767, Nov. 2023, doi: 10.3389/fmed.2023.1172767.
[45] C. Qin, J. Cao, H. Fu, F. S. Khan, and R. M. Anwer, "DB-SAM: Delving into High Quality Universal Medical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer Nature Switzerland, 2024, pp. 498–508, doi: 10.1007/978-3-031-72390-2_47.
[46] S. Iqbal, T. M. Khan, S. S. Naqvi, A. Naveed, and E. Meijering, "TBConvL-Net: A hybrid deep learning architecture for robust medical image segmentation," *Pattern Recognition*, vol. 158, p. 111028, Feb. 2025, doi: 10.1016/j.patcog.2024.111028.
[47] J. Ruan, J. Li, and S. Xiang, "VM-UNet: Vision Mamba UNet for Medical Image Segmentation," *ACM Transactions on Multimedia Computing, Communications, and Applications*, Sep. 2024, doi: 10.1145/3767748.

BIOGRAPHIES OF AUTHORS

Cherifa Abdellaoui received the Electronic Engineering degree from University of Saad Dahleb Blida, Algeria in 2011, and Master degree from University of M'hamed BOUGARA Boumerdes, Algeria in 2020. Working as Research Support engineer in Research Center of Industrial Technologies (CRTI) Algeria from 2017, she is a Ph.D. student in M'hamed Bougara University in Boumerdes. She can be contacted at email: c.abdellaoui@univ-boumerdes.dz.

Samia Belkacem is a Professor of Electronic Engineering at M'hamed Bougara University in Boumerdes, Algeria. She received her Master/Ing and Ph.D. degrees in Electronic from Batna University in 2005 and 2015, respectively. She was promoted to full Professor in 2025. Where she supervised research activities in the field of biomedical signals/images and data modeling/cryptography using artificial intelligence and she taught signal and image processing in LMD system as well as Ph.D. student. She has conducted many research activities, reflected in the publication of 14 articles in valuable, peer-reviewed journals. In addition, she has participated in 25 international conference papers, and 9 national conferences. She can be contacted at email: s.belkacem@univ-boumerdes.dz.

Noureddine Messaoudi graduated in electronics engineering from the University of Sétif 1, Algeria, 2003. He received his Magister degree and Ph.D. degree in Electronics from the University of Sétif 1, Algeria, in 2006 and 2017, respectively. Actually, he is a full professor at the University of Boumerdes where he teaches courses in electronics, biomedical engineering, and telecommunications. His research activity covers surface EMG signal modelling and processing and medical diagnosis by artificial intelligence. He can be contacted at email: n.messaoudi@univ-boumerdes.dz.