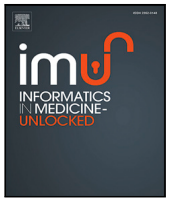
Informatics in Medicine Unlocked 47 (2024) 101504

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



Deep learning for medical image segmentation: State-of-the-art advancements and challenges

Md. Eshmam Rayed [a], S.M. Sajibul Islam [a], Sadia Islam Niha [a], Jamin Rahman Jim [a], Md Mohsin Kabir [b], M.F. Mridha [a,*]

[a] Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh
[b] Department of Computer Science & Engineering, Bangladesh University of Business & Technology, Dhaka 1216, Bangladesh

ARTICLE INFO

Keywords:
Image segmentation
Medical imaging
Image pre-processing
Data augmentation
Deep learning

ABSTRACT

Image segmentation, a crucial process of dividing images into distinct parts or objects, has witnessed remarkable advancements with the emergence of deep learning (DL) techniques. The use of layers in deep neural networks, like object form recognition in higher layers and basic edge identification in lower layers, has markedly improved the quality and accuracy of image segmentation. Consequently, DL using picture segmentation has become commonplace, video analysis, facial recognition, etc. Grasping the applications, algorithms, current performance, and challenges are crucial for advancing DL-based medical image segmentation. However, there is a lack of studies delving into the latest state-of-the-art developments in this field. Therefore, this survey aimed to thoroughly explore the most recent applications of DL-based medical image segmentation, encompassing an in-depth analysis of various commonly used datasets, pre-processing techniques and DL algorithms. This study also investigated the state-of-the-art advancement done in DL-based medical image segmentation by analyzing their results and experimental details. Finally, this study discussed the challenges and future research directions of DL-based medical image segmentation. Overall, this survey provides a comprehensive insight into DL-based medical image segmentation by covering its application domains, model exploration, analysis of state-of-the-art results, challenges, and research directions—a valuable resource for multidisciplinary studies.

1. Introduction

Image segmentation involves dividing an image into regions or segments to identify objects or areas of interest [1]. The objective of image segmentation is to streamline the depiction of an image into coherent and visually meaningful sections. These segments find applications in diverse fields, including object recognition, image manipulation, and medical image analysis. The process initiates with the intake of an original image, followed by a series of pre-processing steps such as resizing, data augmentation, scaling, and cropping. Subsequently, feature extraction and image segmentation techniques are applied. Post-segmentation, further pre-processing steps are employed, culminating in the final stage of result evaluation, which yields the segmented image output. This systematic approach ensures the accurate delineation of relevant portions within the image.

Rapid advances in artificial intelligence, particularly deep learning [2], have significantly improved image segmentation methods. Deep learning-based techniques have shown remarkable results in image segmentation, outperforming traditional machine learning and computer vision approaches in terms of both accuracy and speed. In particular, using DL for medical image segmentation helps doctors accurately determine tumor sizes and quantitatively assess treatment effects. This advancement not only reduces physicians workload but also contributes to more accurate and efficient medical diagnoses and treatment planning. Furthermore, DL algorithms make it easier to detect small or early pathological changes that are often difficult to identify manually, improve patient outcomes and enable timely intervention [3]. Its applications include scene understanding, medical image analysis, and augmented reality. On the other hand, deep learning (DL) is a subset of machine learning, that uses artificial neural networks to learn hierarchical representations of data. DL has proven to be highly effective for image segmentation because it can automatically extract complex hierarchical features from data, enabling it to capture intricate patterns and relationships in images [4]. The connection between image segmentation and DL lies in the fact that DL models can directly learn complex features and patterns from raw pixel values, making them well suited for image segmentation tasks. Specific

* Corresponding author.
E-mail address: firoz.mridha@aiub.edu (M.F. Mridha).

https://doi.org/10.1016/j.imu.2024.101504
Received 10 January 2024; Received in revised form 17 April 2024; Accepted 18 April 2024
Available online 22 April 2024

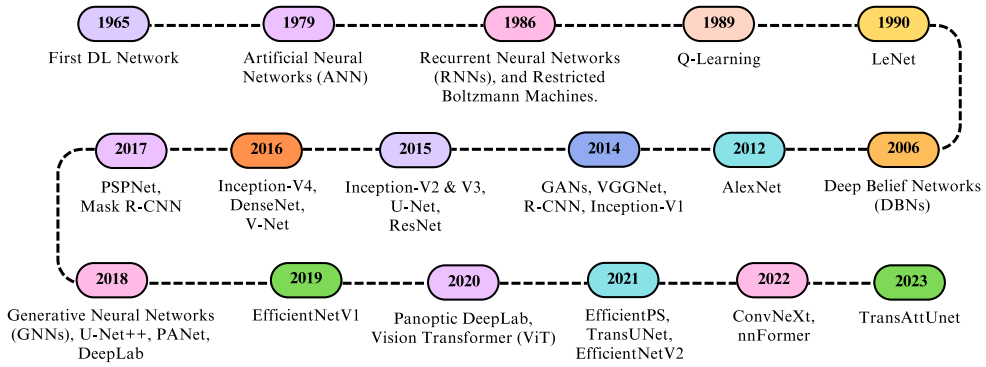


Fig. 1. Timeline of the most popular DL algorithms in medical image segmentation.

examples of DL applications in image segmentation include using DL for scene understanding to identify objects and their relationships in images, which aids in navigation and object detection. In medical imaging, DL models are used in medical imaging to segment tissues and organs, which aids in illness diagnosis and treatment planning.

DL-based image segmentation is becoming increasingly crucial in the medical field because of its ability to accurately and efficiently analyze medical images. It provides an accurate quantitative analysis and offers objective and standardized indicators for clinical trials. In addition, multi-modal image fusion allows the integration of data from different imaging modalities, enabling a comprehensive assessment of a patient's health. This technique has gained prominence in research and medical applications because of its effectiveness in detecting and diagnosing diseases, as well as in guiding image-based interventions. This ability is crucial for segmenting medical images that often exhibit noise and wide variations. The potential for transforming medical image segmentation is considerable. Automating the segmentation process can reduce the workload of radiologists and medical experts, thereby enabling them to concentrate on diagnosis and treatment planning. In addition to DL, it is possible to develop new segmentation algorithms that are more reliable and accurate than traditional methods.

Image segmentation has made a remarkable evolutionary journey from traditional machine learning (ML) methodologies to its current unprecedented state of development within the field of DL. Initially, image segmentation techniques relied mainly on traditional ML algorithms, where hand-crafted features and heuristic rules were used to partition images into different regions [5]. However, the emergence of DL, in particular Convolutional Neural Networks (CNNs), revolutionized image segmentation by enabling automatic feature extraction and learning of hierarchical representations [6]. This paradigm shift has opened a new era of image segmentation, allowing models to capture complex spatial relationships and contextual information, thus yielding more accurate and versatile segmentation results in various applications. Fig. 1 showcases the progression of DL models for medical image segmentation, reflecting continuous advancements in both quality and necessity. These developments have significantly enhanced the capabilities of DL over time. End-to-end DL, 3D capabilities, and transfer learning have led to accurate segmentation, while approaches such as weak supervision, self-supervision, and ensemble modeling have addressed challenges such as limited labeled data and class imbalance. Real-time deployment, continuous learning, and adaptability further enhance clinical utility where models like U-NET [7], U-NET++ [8], FCN (Fully Convolutional Network) [9] has evolved in the domain of medical image segmentation. This continuous progress and development reflect the growing importance and applicability of these techniques for medical applications.

The growing development and investigation of the segmentation of medical images using DL underlines its growing importance. Facilitating the launch of research efforts on DL-based medical image segmentation requires a comprehensive and illuminating study that encompasses the advances and challenges of using DL techniques. This study provides a thorough review of the different DL methods used in medical image segmentation, accompanied by scenarios, applications, pre-processing techniques, datasets, and their challenges with future research directions in different medical imaging fields to overcome these challenges. Although there are some surveys on medical image segmentation, none of them specifically address all the areas that have been mentioned. More specifically, the most widely used pre-processing techniques for image segmentation in the medical field are currently not covered in any review article. Therefore, to better understand the current state of medical image segmentation, we only used research articles published between 2022 and 2023 to conduct this systematic review. Table 1 compares our findings with those of existing surveys. This article aims to support researchers in developing DL-based medical image segmentation, while providing insights into current advancements in the field.

The presented study makes several noteworthy contributions, including:

1. **Comprehensive source of databases:** Our review focuses exclusively on high-caliber scholarly articles sourced from reputable databases such as ScienceDirect, SpringerLink, MDPI, Wiley, IEEE Xplore, and other prominent databases in the field.
2. **Criteria for systematic review:** We carefully selected relevant research articles that were published only in 2022 and 2023, all of which adhered to the strict PRISMA guidelines.
3. **Timeline:** In our review, we considered a timeline primarily consisting of frequently employed DL algorithms for medical imaging up until 2023.
4. **Pre-processing & datasets:** Our study provides a thorough analysis of pre-processing techniques and commonly used datasets in the realm of DL-based medical image segmentation.
5. **Algorithms:** We have included an in-depth examination of popular DL algorithms used in medical segmentation by highlighting each algorithm's unique advantages and limitations.
6. **Advancement analysis:** This review analyzes the most recent developments and contributions made by researchers, showcasing the latest experimental findings that influence the field's landscape.
7. **Challenges and future work:** We identified a wide range of challenges related to DL-based medical image segmentation, and analyzed potential future research directions to address and overcome these challenges.

The remainder of this paper is structured as follows. Section 2 outlines the methodology employed for this systematic review. Section 3 explores the wide range of application domains of medical image segmentation, and Section 4 covers frequently used datasets & pre-processing techniques. Section 5 discusses some of the most commonly used evaluation metrics for assessing the model performance. Section 6

Table 1
Comparative performance and limitations of existing DL-based medical image segmentation survey papers from 2022 & 2023.

Ref.	Application	Datasets	Pre-Processing	Algorithm	Results analysis	Challenges & Future work	Medical image	Contribution
Minaee et al. [1]	✗	✓	✗	✓	✓	✓	✗	Analyzed a thorough review on DL algorithms for image segmentation, along with an analysis of their similarity, strengths, challenges, datasets, result analysis, and future research directions.
Liu et al. [3]	✓	✓	✗	✓	✗	✓	✓	Conducted a novel thematic survey, classifying DL into a multi-level structure. Provided thorough insights for advancing the field by examining datasets, challenges, and outlining future directions.
Aljabri et al. [10]	✓	✓	✗	✓	✓	✓	✓	Offered a thorough analysis of more than 150 applications and DL algorithms in medical image segmentation, highlighting current challenges and outlining potential future research directions.
Qureshi et al. [11]	✓	✓	✗	✗	✓	✓	✓	Presented a comprehensive review of DL approaches for medical image segmentation, discussesd datasets, analyzed results for various diseases, and discusses future directions and challenges.
Yu et al. [12]	✗	✗	✗	✓	✓	✓	✗	Provided a systematic review of image segmentation techniques, covering classic, collaborative, and semantic segmentation, examined key algorithms, compared models, and discussed challenges and future developments in the field.
This Paper	✓	✓	✓	✓	✓	✓	✓	Provides a comprehensive analysis covering all the mentioned areas in DL-based medical image segmentation, focusing on the most recent (2022 & 2023) state-of-the-art research publications.

analyzes popular DL algorithms in medical image segmentation in detail. Section 7 presents advancements in recent state-of-the-art papers, and Section 8 identifies challenges and future research opportunities. Finally, Section 9 concludes the paper.

2. Survey methodology

This study was conducted using the systematic literature review (SLR) proposed by Keele et al. [13,14]. This survey was conducted to explore the application of DL in medical image segmentation. It is specifically confined to high-caliber scholarly articles sourced from esteemed databases, such as ScienceDirect, SpringerLink, ACM Digital Library, and IEEE Xplore. The essential resources for this survey were meticulously gathered in accordance with the guidelines depicted in Fig. 2, adhering to the Preferred Reporting Items for Deep Learning in Medical Image Segmentation (PRISMA) framework. Furthermore, as outlined in the PRISMA diagram, the inclusion and exclusion criteria are comprehensively explained in Table 2. This table methodically outlines the parameters employed to ascertain a paper's eligibility for inclusion or exclusion from this review.

In this research, 524 papers were identified for review. Table 3 lists the various keywords employed for selecting articles from different databases along with the corresponding count of articles chosen for review against each keyword. Fig. 3 shows the number of papers selected from each bibliographic database, with the majority being sourced from Science Direct, followed by Springer Link and IEEE Xplore, indicating a concentration of relevant literature in these databases for the field of study.

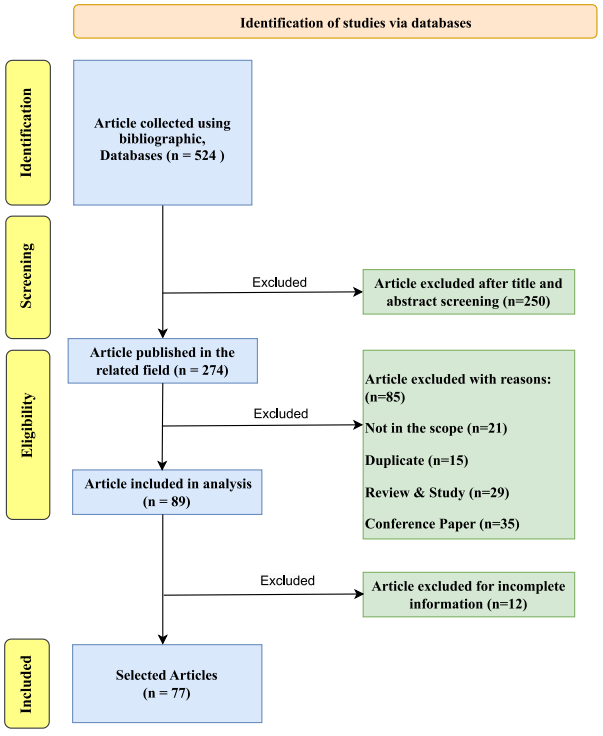


Fig. 2. PRISMA diagram illustrating the process of selecting articles for the applications and state-of-the-art advancements.

3. Applications

Medical image segmentation is a cutting-edge field that fundamentally changes the diagnosis and treatment of healthcare problems. The process essentially entails the accurate identification and removal of

Table 2
The table discusses the including and excluding criteria for selecting articles.

	Inclusion criteria	Exclusion criteria
Types of study	Original and review articles.	Thesis, white papers, communication letters, reports and editorials.
Language	Research articles written in English.	Duplicate and non-English articles.
Publication year	Articles published in 2021–2023 (For applications and results analysis part).	Not related to the theme of the review.
Source	Articles published in academic journals and conferences.	Articles lacking information and review papers.
Intervention	Deep Learning methods.	Traditional and statistical methods.
Region	Not restricted to a particular region.	–
Settings	DL methods in medical image segmentation	Not in DL methods in other image segmentation settings.

Table 3
The table presents a formal overview of the keywords utilized for article selection across various databases and the corresponding paper count for each keyword.

Keywords	Paper count
Image segmentation	250
Deep learning algorithms	85
Medical image survey	20
Health informatics	23
AI-Assisted diagnosis	31
Machine vision in medicine	19
Image-based diagnostics	41
Data-driven healthcare	29
Neural network applications in medicine	26

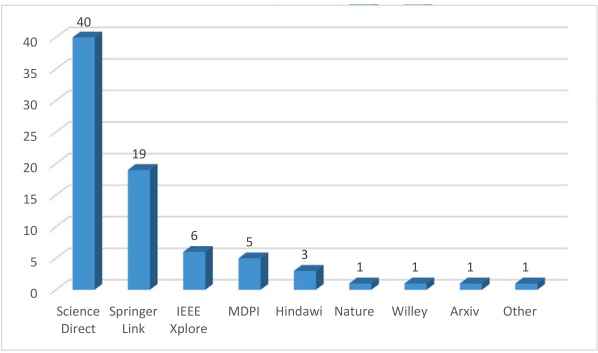


Fig. 3. Number of papers selected from each bibliographic database.

certain objects or regions of interest from medical images, such as X-rays, MRIs, and CT scans. At its foundation, medical image segmentation uses state-of-the-art technologies such as deep learning and complex algorithms such as CNNs, autoencoders, and U-Nets. This technology has become an essential part of many medical applications because it offers new insights never before possible and enables more accurate analyses. One of the primary applications of medical image segmentation lies in disease diagnosis. Medical professionals can learn more about a patient's condition by segmenting organs, tissues, or pathological anomalies in medical images. This ability is essential for accurate anomaly localization, early identification of the disease, and the development of more efficient treatments. For example, in the field of oncology, segmentation plays a crucial role in enabling prompt and accurate tumor identification and measurement, which in turn influences treatment choices and patient outcomes. Similarly, it facilitates the analysis of brain scans in neurology to identify and track neurological disorders.

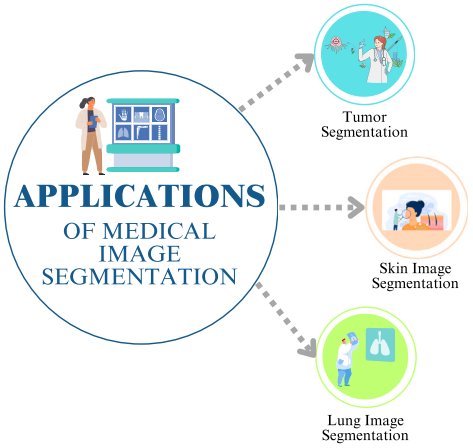


Fig. 4. Applications of medical image segmentation.

Currently, the field faces challenges related to handling complex anatomical variations in models and the requirement for large, diverse, and annotated datasets. This field is expected to make more strides in the future, including the incorporation of 3D imaging methods and the use of segmentation in personalized medicine. Fig. 4 shows some of the core applications of medical image segmentation that we are going to explore. These applications are briefly discussed below.

3.1. Tumor segmentation

DL has significantly influenced tumor segmentation in medical imaging, which is a crucial phase in the identification and management of cancer. In different imaging modalities, such as MRI and CT scans, techniques like CNNs, autoencoders, GANs, and RNNs have been used to differentiate tumor tissues from normal tissues. These methods make it possible to thoroughly analyze the growth patterns, size, and shape of tumors, all of which are crucial for determining the stage of cancer and developing treatment plans. In particular, GANs improve the model robustness by producing artificial medical images for training, which is important in situations where real-world data are limited. With the help of autoencoders for dimensionality reduction, it is vital to manage high-resolution medical images. RNNs have advantages when it comes to sequential data analysis, which is useful for monitoring tumor changes over time. A combination of these deep learning methods in tumor segmentation results in enhanced identification accuracy, thereby accelerating timely diagnosis and customized treatments. This method helps in the research and development of new cancer therapies, in addition to improving patient outcomes.

3.2. Lung image segmentation

Lung image segmentation, particularly in the context of diseases, such as lung cancer, pulmonary fibrosis, and emphysema, has witnessed significant advancements in DL. In addition to CNNs, attention-based models and U-Nets have proven useful for analyzing lung CT scans. Because U-Nets can precisely localize images and capture contextual information, they are well known for their efficiency in medical image segmentation. This is especially helpful for locating lung nodules, describing their dimensions, and determining the severity of the lung conditions. Neural networks with integrated attention mechanisms improve the capacity of the model to concentrate on relevant features in lung images, such as abnormal tissue patterns or nodules. These developments have made it possible to segment the lungs more precisely and

thoroughly, which is essential for the identification and management of a variety of pulmonary disorders. In addition to streamlining patient care procedures, the increased accuracy of lung image segmentation enables quicker and more efficient treatment of lung-related medical conditions. Recent developments have resulted in the incorporation of advanced DL models, leading to significant improvements in diagnostic procedures and treatment planning.

3.3. Skin image segmentation

In dermatology, DL has improved skin image segmentation, which is a key element in the diagnosis and treatment of skin diseases. The accurate identification of different skin structures and lesions has advanced significantly with the use of CNNs, transfer learning, and deep residual networks (ResNets). By leveraging pretrained models on large datasets, accurate segmentation can be achieved with relatively limited dermatological data. ResNets, known for their deep structures, are effective for learning complex patterns in skin images, which is crucial for identifying and differentiating between benign and malignant lesions. This precision in segmentation assists dermatologists in monitoring disease progression and in planning effective treatments, including surgical interventions. The progress made in deep learning for skin image segmentation has lead to better health outcomes for patients with skin conditions by increasing the efficiency of dermatologists and improving diagnostic accuracy. Recent developments in this area, such as the use of advanced DL models, have led to significant improvements in the early identification and treatment of various skin diseases.

Following a detailed exploration of tumor, lung, and skin image segmentation, Section 7 delves into the recent advancements in these specific applications, offering an in-depth perspective on the latest innovations in this field.

4. Datasets and pre-processing

This section is divided into two parts. The first concerns about datasets and explores well-known datasets used in the domain of medical image segmentation. The second subsection focuses on pre-processing techniques in the context of image segmentation.

4.1. Datasets

In the field of image segmentation, datasets play a crucial role as a standard reference against which novel segmentation methods can be compared. These datasets serve as the benchmarks for assessing the effectiveness of innovative approaches in the field. Some of the widely used datasets for DL-based image segmentation are listed in Table 4.

4.2. Pre-processing techniques

Pre-processing is essential to make raw data for meaningful analysis. This includes a set of methods for cleaning, modifying, and improving the quality of the dataset. This section explores critical pre-processing techniques such as scaling, noise removal, grayscale conversion, data augmentation, and image resizing. These methods are essential for ensuring accurate and reliable results across different areas of data science. Table 5 lists the various pre-processing methods used for image segmentation.

5. Evaluation metrics

When evaluating the effectiveness of a deep learning model, it is critical to comprehend and apply evaluation measures. This is an essential phase in evaluating deep learning algorithms and is required for the execution of any research. These metrics act as quantitative measurements and provide insightful information regarding the effectiveness and caliber of a model. Numerous evaluation metrics exist in the field of DL, each of which has a specific function. The importance of these metrics cannot be overstated; they are essential to ensure that the model is running correctly and efficiently. The effective use of these metrics allows us to review the performance of the model and identify areas in need of development to ensure that the model fulfills efficiency and quality requirements. Accuracy, precision, recall, and F1-score are some of the popular measures used to evaluate a system's performance (see Table 6).

The letters TP, TN, FP, FN, and FN represent, the numbers of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) samples.

In Mean IoU: N represents the number of classes, S1 is the set of pixels classified as class i in the model's output, Gi is the set of pixels actually belonging to class i in the ground truth, Intersection(Si, Gi) is the number of pixels correctly classified as class i, and Union(Si, Gi) is the total number of pixels that are in either Si, Gi, or both.

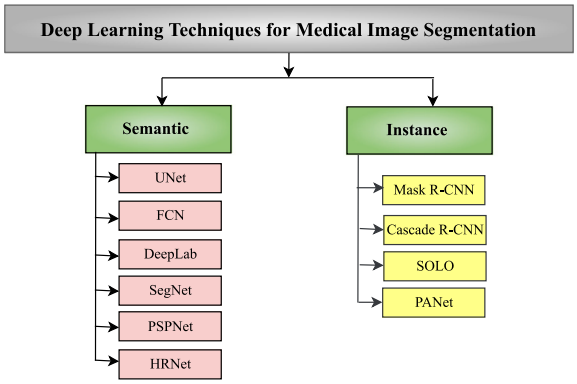


Fig. 5. The taxonomy of the different DL techniques used for medical image segmentation.

6. Algorithms for Medical Image Segmentation

Within the domain of Medical Image Segmentation, various algorithms play an important role in the extraction of essential information from complex images. Convolutional Neural Networks (CNNs), including notable examples such as U-Net, have been used to capture complex structures. Sophisticated techniques, such as DeepLab and PSPNet, go beyond traditional segmentation and provide methods for capturing fine details, incorporating the global context, and semantic understanding. Semantic segmentation and instance segmentation offer sophisticated methods for segmentation, unique instance delineation, and comprehensive image comprehension, in addition to traditional segmentation. These algorithms effectively address the unique challenges of medical image data, thereby enabling accurate and thorough segmentation for diagnostic and research applications. The DL algorithms used in medical image segmentation are briefly categorized in Fig. 5.

6.1. Semantic segmentation

Semantic segmentation involves classifying and labeling individual pixels in an image based on their semantic meaning. The goal was to categorize each pixel in the image and divide it into meaningful segments. This process is essential for tasks such as medical imaging, where pinpointing specific structures is crucial for diagnosis, and autonomous driving, which demands precise object and road delineation. The capabilities of semantic segmentation have advanced significantly owing to widely used techniques, such as U-Net, FCN, PSPNet, and DeepLab.

Table 4
Overview of popular medical image segmentation datasets: A comprehensive compilation of diverse imaging data across various medical areas. The table illustrates the type of data set, its body area, and its features, including the quantity and dimensions of images.

Dataset	Type	Domain	Details	Used by
BUSI	Ultrasound images	Tumor	Grayscale pictures make up the images. This data collection contains 780 photos of 600 women. The photos are 500 × 500 pixels on average and are in the PNG format. Benign, Malignant, and Normal are the three different types of classes. Dataset evaluation was conducted using five-fold cross-validation to assess the test performance of various methods. Additionally, the size of the input images was standardized at 256 × 256 pixels for all techniques.	[15]
STS	PET/CT	Tumor	Soft Tissue Sarcoma dataset includes 51 confirmed cases, with FDG-PET scans performed using a PET/CT scanner. Tumor regions were marked by a radiation oncologist on T2FS scans and transferred to FDG-PET images using MIM software. For patients with edema, non-visible edema data was also used. In preprocessing, PET and CT data were resized to 128 × 128 pixels, adjusted for Standard Uptake Values and Hounsfield Units, and normalized. To overcome data limitations and avoid overfitting, augmentation techniques like flipping and rotation were applied to PET/CT images.	[16]
BOT (Gastric Cancer)	WSI	Tumor	This dataset comprises of 700 digital pathology images focusing on gastric cancer, including 560 cancer and 140 non-cancer samples. A selection of 500 images (comprising 400 cancer and 100 non-cancer samples) forms the training set, accounting for approximately 71% of the total dataset. The remaining 29%, which includes 160 cancer and 40 non-cancer samples, is equally divided for testing and validation purposes. Each image in the dataset is of the size 2048 × 2048 pixels. These images are provided in the TIFF format, complete with pixel-level annotations.	[17]
LIDC-IDRI	CT	Lung	The dataset was divided into two groups. The first group consisted of 452 subjects, making up 66.57% of the total, and was used for training and validation. The second group included the remaining 227 subjects, accounting for 33.43%, and was specifically set aside for testing purposes.	[18]
Decathlon (MSD)	3D CT	Lung	Includes 96 segmented 3D CT scan sets. This dataset is split into two segments: a training subset with 64 3D volumes and a testing subset containing 32 3D volumes. The ratio of the data split between the training and testing subsets in the Decathlon lung dataset is 2:1.	[19]
Montgomery County (MC)	X-rays	Lung	The dataset comprises 138 chest X-ray images, categorized into two groups: 80 images representing healthy individuals and 58 depicting cases of tuberculosis. Upon request, these images can be provided in the Digital Imaging and Communications in Medicine (DICOM) format. Each X-ray image possesses a resolution of 4020 × 4892 pixels.	[20–22]
JSRT	X-rays	Lung	The Japanese Society of Radiological Technology (JSRT); dataset encompasses 247 chest X-ray images, each with a resolution of 2048 × 2048 pixels. Among these, 93 images are classified as normal, while the remaining 154 exhibit abnormalities indicative of tuberculosis (TB). These images are maintained in PNG format, with each image having a 2048 × 2048 pixel resolution and represented in 12-bit grayscale.	[20–22]
CAMUS	Echocardio-graphic images	Lung	A dataset made up of two- and four-chamber images in two dimensions that were taken from the apical sequences of 500 patients. Each sample was individually adjusted by subtracting the average value of the image and then dividing it by the standard deviation. No post-processing was employed in this study, which differs from other methodologies where such processes are applied.	[23]
ImageCLEF 2019	3D CT	Lung	The dataset consists of 335 CT images, with 218 for training and 117 for testing, each sized 512 × 512 pixels. The training images underwent random transformations (3D shifting, rotation, cropping, shearing, and scaling) with a 50% probability, also binarizing the non-lung areas. This augmentation process across the 218 training images resulted in a total of 1308 images for training.	[24]
LUNA16	CT	Lung	The dataset is systematically partitioned into two splits: one for training and another for testing. The training part is subjected to a process of data augmentation, which involves the modification of images to enhance the diversity of the dataset. This augmentation is achieved by flipping the images along an axis and rotating them.	[25]
ISIC challenge	Dermoscopic	Skin	The ISIC dataset is an extensive open-source archive of annotated skin images, designed to aid in skin cancer research and AI development. It emphasizes high standards in digital imaging and engages the computer vision community through workshops and challenges.	[26–33]

6.1.1. Convolutional Neural Network (CNN)

CNN [38] is a widely used model that combines image processing and DL. It has achieved numerous achievements in the area of image processing and analysis as a highly representative neural network within the domain of DL technology. CNNs are among the most popular and successful architectures in the DL field, particularly for computer vision problems. In 1959, Hubel et al. [39] discovered that animal visual cortex cells are responsible for sensing light in the receptive fields. In 1980, Fukushima unveiled the neocognitron as a result of this discovery [40], which is often considered the predecessor of CNN. LeCun et al. [41] built a CNN architecture for document recognition, whereas Waibel et al. [42] used backpropagation training to recognize phonemes and introduced CNNs with shared weights across the temporal receptive fields.

As illustrated in Fig. 6, a standard CNN comprises an input layer, output layer, and series of functional layers that transform an input

Table 5
The table discusses the commonly used pre-processing methods in image segmentation.

Pre-processing method	Description
Scaling	DL models achieve optimal performance when the original data is transformed to a specific range, typically between 0 and 1, as these models work most efficiently with small input values. There are two widely used techniques for scaling data: normalization and standardization. Normalization moves data from the original range to a new scale ranging from 0 to 1. Normalization transforms data from its initial range to a fresh scale ranging from 0 to 1. It achieves this by utilizing the minimum and maximum values min(x) and max(x) found within the input vector x. Consequently, the transformed values, denoted as x′, are confined within the range of 0 to 1. This process can be mathematically represented in (1) as: $x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ (1)
Data normalization	Normalization is a commonly employed method for preparing data for DL and involves three key steps: local normalization, global contrast normalization (GCN), and histogram equalization. In a GCN, every pixel value in an image is adjusted by subtracting the mean value and then dividing it by a known error. The formula for GCN is applied to each pixel (m, n, l) in (2). $Z'_{(m,n,l)} = s \frac{Z_{(m,n,l)} - \bar{Z}}{\max\left\{\varepsilon, \lambda + \frac{1}{3rc}\sum_{m=1}^{r}\sum_{n=1}^{c}\sum_{l=1}^{3}\left(Z_{m,n,l} - \bar{Z}\right)^2\right\}}$ (2)
Noise elimination	Image data frequently include undesirable noise, disrupting the segmentation process and causing inaccuracies in the results. To address this problem, noise elimination techniques are employed. Popular methods involve the use of filters, such as median filters, Gaussian filters, or morphological operations. These filters function by smoothing the image and diminishing the noise while retaining the vital features. An elementary noise elimination technique such as using a median filter, can be expressed in (3) as follows: $I_{\text{filtered}}(x,y) = \sum_{i=-k}^{k}\sum_{j=-k}^{k} I(x+i, y+j) \times \text{Kernel}(i,j)$ (3)
Grayscale conversion	Grayscale conversion is a standard pre-processing technique in image segmentation, where colored images are converted to grayscale versions. This simplification streamlines the calculation process and guarantees consistent pixel intensity, thereby improving the efficiency of segmentation algorithms. There are different methods for grayscale conversion; however, the luminosity method is a popular approach. In this equation, "Gray" signifies the intensity of the grayscale, while "Red," "Green," and "Blue" denote the respective color channels. The grayscale conversion is generated using (4). $\text{Gray} = 0.299 \times \text{Red} + 0.587 \times \text{Green} + 0.114 \times \text{Blue}$ (4)
Image resizing	It standardizes image dimensions, enhances computational efficiency, and ensures consistency.
Data augmentation	It is an essential pre-processing method that, by performing different transformations, increases the quantity and variety of training data, strengthening the machine learning models' resilience and generalizability.
Clipping	Its primary purpose is to restrict extreme values, thus guaranteeing that the data stays within a predefined range. This ensures that subsequent analysis is more robust and yields more meaningful results.
Cropping	Cropping involve isolating a specific part of an image region of interest (ROI). This method increases the efficiency of neural networks in image segmentation by removing irrelevant background details. This enables these networks to identify and outline objects within the selected area more accurately.
Data re-sampling	It alters the distribution of a data set by either oversampling or under-sampling to address the class imbalance, enhance model performance, and adjust data granularity.
Image enhancement	Image enhancement is a crucial pre-processing method that enhances the visual quality and information content of digital images, enabling better analysis and interpretation in various applications, including computer vision and image processing.
Background removal	Its primary purpose is to isolate specific subjects or objects by removing unnecessary elements, thereby improving visual emphasis and supporting a wide range of subsequent uses.
Linear warping	Linear warping stand as a robust pre-processing technique within image and signal processing, facilitating data transformation through applying a linear function to improve or rectify particular characteristics.

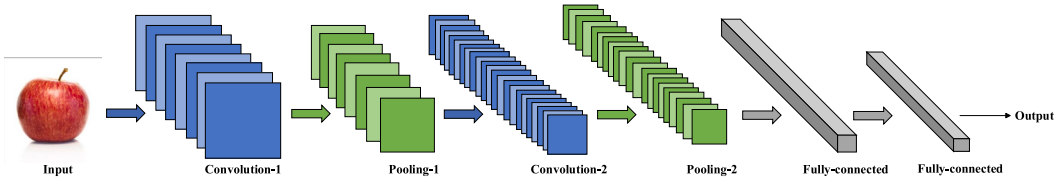


Fig. 6. Architecture of convolutional neural network (CNN) [41].

into a specific form, such as vectors. These functional layers typically include convolutional, pooling, and fully connected layers.

Kayalibay et al. [43] presented a method that uses convolutional neural networks (CNN) with three-dimensional filters for the segmentation of MRI images of hands and brains. It describes two adaptations to a current CNN framework to overcome problems such as the limited availability of labeled data, significant class imbalance, and significant memory requirements of three-dimensional images. In addition, Pereira et al. [44] utilized Convolutional Neural Networks to Segment brain tumors in MRI scans. Several methods have been developed since 2006 to solve the difficulties in training deep CNNs [45,46]. Krizhevsky et al. [47] developed a traditional CNN architecture and demonstrated significant improvements over earlier approaches to image classification. LeNet-5 and AlexNet share similar overall architectures; however, AlexNet has a deeper structure. Since AlexNet [47] was successful, various initiatives have been suggested to enhance its functionality. The following are among the most renowned CNN architectures: VGGNet [48], GoogLeNet [49], ResNet [50], MobileNet [51], DenseNet [52].

6.1.2. *Fully Convolutional Network (FCN)*

The FCN is a widely used neural network model for semantic segmentation of computer vision. Semantic segmentation involves labeling individual pixels in an image to create discrete regions corresponding to different classes of objects. FCNs have significantly contributed to the development of semantic segmentation and related fields in computer vision [53]. The FCN architecture has proven to be very useful in medical image segmentation, enabling precise segmentation of complex structures and significant regions in complex medical images. FCNs

Table 6
Comprehensive overview of key evaluation metrics in image segmentations.

Metric	Details	Equation	Used In
Accuracy	Accuracy is the proportion of pixels in the picture that were properly classified to all of the image's pixels.	$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$	[15,17–22,34–36]
F1-score	The Dice Coefficient (F1-score), which is a balanced indicator of the model's performance, is the harmonic mean of accuracy and recall.	$F1 - score = 2\frac{Precision*Recall}{Precision+Recall}$	[15,17–22,24,34,35,37]
Hausdorff distance (HD)	It is a metric for assessing the accuracy of segmentation in image processing. It measures the edge morphology of an object by comparing the distances between the perimeters of the actual object (A1) and the segmented object (A2). HD determines the maximum distance from a point on one object to the closest point on the other object. A higher HD value signifies a larger discrepancy between the actual and segmented object boundaries, indicating lower segmentation precision. This metric is particularly useful for identifying errors concentrated along the edges of objects. HD is computed as follows:	$HD(A_1, A_2) = \max(h(A_1, A_2), h(A_2, A_1))$	[15,18]
Jaccard Index (IoU)	One of the most frequently used metrics in semantic segmentation is intersection over union (IoU), also known as the Jaccard Index. It may be calculated by dividing the area of intersection between the anticipated segmentation map and the ground truth by the area of union between the two, where A and B stand for the ground truth and the predicted segmentation maps, respectively. It has a range of 0 to 1.	$IoU = J(A, B) = \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$	[15–17,20,22,35–37]
Mean IoU	Another often used measurement is Mean IoU, which is the average IoU across all classes. It is frequently employed in reporting on the effectiveness of contemporary segmentation algorithms.	Mean IoU $= \frac{1}{N}\sum_{i=1}^{N}\frac{\text{Intersection}(S_i,G_i)}{\text{Union}(S_i,G_i)}$	[34]
Precision	Precision is defined as the ratio of the total number of positive samples predicted by the model to the genuine positive samples.	$Precision = \frac{TP}{TP+FP}$	[15,17,34,37]
Recall	Recall is defined as the proportion of real positive samples to all positive samples in the data set. Recall is also called sensitivity.	$Recall = \frac{TP}{TP+FN}$	[15–20,24,25,34,36,37]
Specificity	Specificity is the ability of the segmentation process to correctly identify negative cases, that is, to accurately mark areas of an image that do not belong to the object of interest. Specificity is a measure of how well the segmentation algorithm avoids false positives.	$specificity = \frac{TN}{TN+FP}$	[15,17–20,24,34,36]

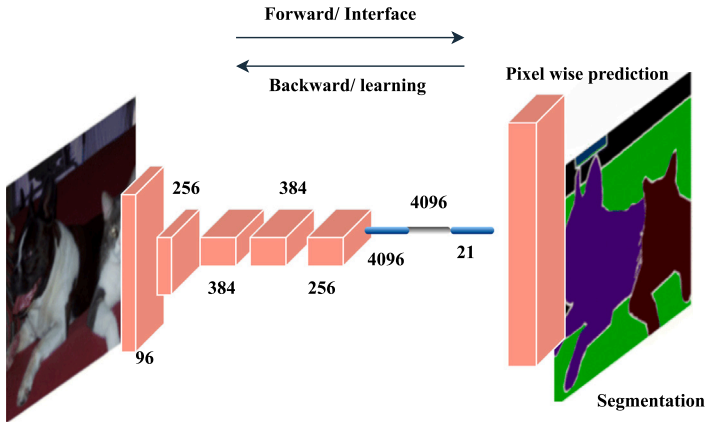


Fig. 7. For per-pixel tasks such as semantic segmentation, fully convolutional networks can effectively learn to generate dense predictions [9].

have various applications in this context, including organ segmentation, tumor detection and segmentation, vascular segmentation, cell and nucleus segmentation, lesion detection, and anatomical landmark localization. FCNs have risen to the top in tasks that involve segmenting natural images and representing features for a comprehensive classification. Fig. 7 shows that this comprehensive classification approach can also be used to predict the probability of individual pixels being associated with specific anatomical structures in the images.

Sun et al. [54] on their paper presented a method for segmenting multimodal brain tumor images using a multi-pathway architecture approach based on 3D Fully Convolutional Networks (FCNs)

In his research on tumor identification, Kumar et al. [55] did not use traditional 2D detection methods, such as dice cut. Instead, they used 3D segmentation for detection, which increases the accuracy of the process. To achieve this, they used a Fully Convolutional Network (FCN), which replaces the CONV layers, leading to a significant reduction in the number of parameters. By implementing skip connections, a technique in which feature maps from the model's final layers are gathered and fused with feature maps from preceding layers and, blends semantic understanding (extracted from deeper, broader layers) with visual attributes (derived from shallower, more detailed layers), as illustrated in Fig. 8. This fusion enables the generation of precise and intricate segmentation. The model's efficacy was assessed across various datasets including NYUDv2, PASCAL VOC, and SIFT Flow, and it demonstrated superior segmentation performance compared to other approaches [9].

Deep networks can trained be end-to-end on variable-sized images for semantic segmentation. For example, the introduction of

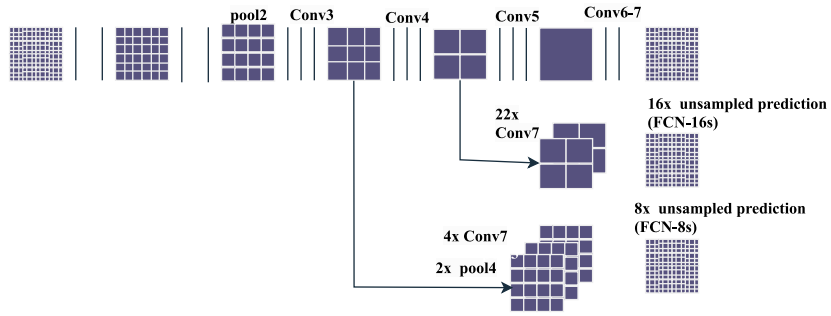


Fig. 8. Skip connections bring together both fine-grained, low-level information and coarse, high-level information [9].

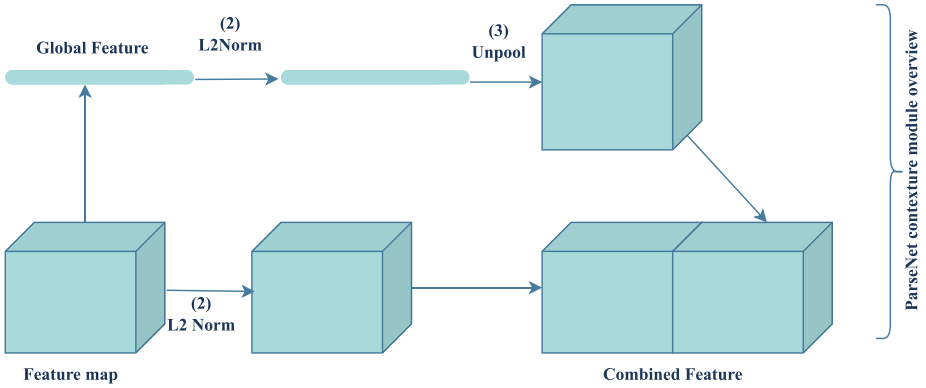


Fig. 9. ParseNet incorporates additional global context to produce smoother segmentation [1].

ParseNet [56] addressed a limitation in FCNs, which is a disregard for global context information. To overcome this, ParseNet integrates the global context into FCNs by enhancing the features at each location with the average feature from a layer. This process involves aggregating the feature map off a layer across the entire image, resulting in a context vector. This vector was then normalized and employed to generate new feature maps of the same dimensions as the original ones. These feature maps were subsequently interconnected. In summary, Fig. 9 demonstrates that ParseNet can be regarded as an FCN wherein the module described replaces the convolutional layers. FCNs have been applied in diverse segmentation tasks, such as delineating brain tumors [57], achieving instance-aware semantic segmentation [58], segmenting skin lesions [59], and delineating irises [60].

6.1.3. U-Net

Using FCN as a foundation, Ronneberger et al. [7] created the U-Net network in 2015, which is a specialized tool that excels in handling images in the medical industry or any field where detailed image analysis is important. U-Net was developed to break down an image into smaller components, analyze those components, and reassemble the image. This is particularly helpful for finding and highlighting specific objects in medical images, such as organs or tumors. U-Net is a useful tool for researchers and anyone working with images that require careful analysis because of its exceptional capacity to simultaneously see the big picture and concentrate on small details.

The U-Net network has two components: U-channel and skip connection. The encoder–decoder architecture of SegNet [61] bears resemblance the structure of the U-Net model. Each of the four submodules of the encoder has two convolutional layers. Max pooling was employed for downsampling after each submodule in the process. The decoder consists of four submodules, the resolution is progressively increased through incremental upsampling. Finally, predictions were generated for each pixel. The network architecture is dillustrated in Fig. 10. The input was 572 × 572, and the output was 388 × 388. The figure shows that this network only uses convolution and downsampling, and does not have a fully connected layer. The network also employs a skip connection to link the upsampling outcome to the output of the encoder submodule with the same resolution as the input of the decoder's subsequent submodule [3]. Each blue box is associated with a multi channel feature map, and the top of the box displays the channel count. The x-y size is provided in the lower-left corner of the box. The copied feature maps are represented by white boxes. The various operations are indicated by the arrows. Ahmad et al. [62] developed and evaluated the MH UNet, and a sophisticated multi-scale hierarchical architecture tailored for medical image segmentation was conducted, showing its performance on the BraTS 2018, 2019, and 2020 Magnetic Resonance Imaging (MRI) validation datasets. On the other hand, Deng et al. [63] use the ELU-Net which is a highly efficient and lightweight version of U-Net, enhanced with deep skip connections that encompass both same- and large-scale links from the encoder for comprehensive feature extraction. In addition, the use of various loss functions, specifically targeting whole tumor (WT), tumor core (TC), and enhanced tumor (ET) segmentation in brain tumor learning, introduces a novel loss function named DFK.

6.1.4. U-Net++

Another powerful version of the U-Net architecture, which draws inspiration from DenseNet [52], is U-Net++. In 2018, Zhou et al. [64] developed a nested U-Net architecture (U-Net++). U-Net++ was designed to handle images more intelligently, particularly when there are numerous details to record. This is achieved by adding more layers that cooperate to fully understand the images. It is comparable to having a group of specialists who concentrate on a different picture aspect to gain a thorough understanding.

Zhou et al. [65] applied UNet++ across six different medical image segmentation datasets, including different imaging techniques, such as CT, MRI, and EM. This shows that UNet++ reliably outperforms the baseline models in semantic segmentation tasks across multiple datasets and backbone architectures. U-Net++ connects the encoder

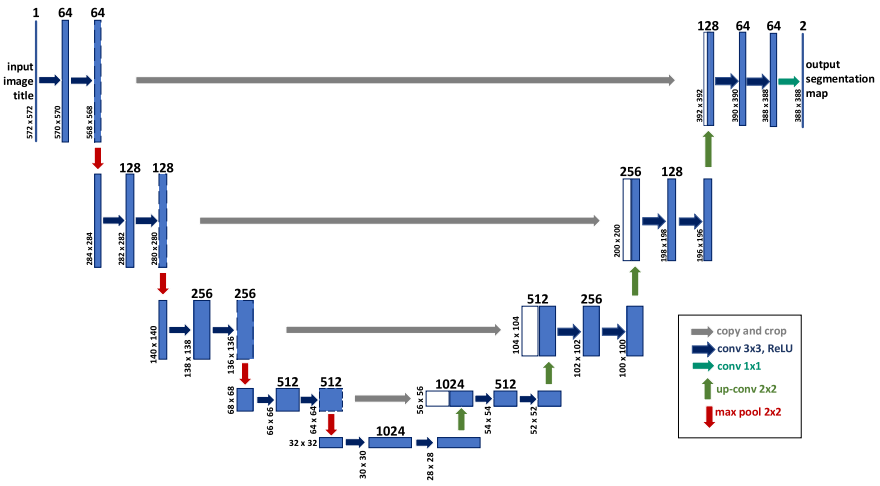


Fig. 10. The U-Net architecture [7]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and decoder components using a sequence of nested dense convolutional blocks. Before fusion, the primary objective of U-Net++ was to bridge the semantic gap between the feature maps of the encoder and decoder. As shown in Fig. 11, U-Net++ incorporates several skip connection nodes between corresponding layers. Each skip connection unit receives an upsampled feature map from its immediate lower unit, in addition to all feature maps from the previous units at the same level. Consequently, each level corresponded to a dense block. This arrangement minimizes the loss of semantic information between two paths [64]. Zhou et al. [64] formulated the operation of the skip connection unit, where x represents the feature map and i and j represent the indices along the contracting path and across the skip connections, respectively. This is defined in Eq. (5) as follows:

$$x^{i,j} = \begin{cases} \mathscr{H}\left(x^{i-1,j}\right), & j = 0 \\ \mathscr{H}\left(\left[x^{i,k}\right]_{k=0}^{j-1}, \mathscr{U}\left(x^{i+1,j-1}\right)\right), & j > 0 \end{cases} \tag{5}$$

Here, $H(\cdot)$ denotes the convolution and activation operation, $U(\cdot)$ signifies the upsampling operation, and [] represents the concatenation. While traversing the contracting path, the count of intermediary skip connection units decreases linearly and depends on the number of layers [64].

6.1.5. DeepLab

Dilatation convolution, also known as "rous" convolution, adds a new parameter to the convolutional layer, called the dilatation rate. As illustrated in Fig. 12, this concept involves applying dilation convolution to the signal *x(i)*, which is defined in Eq. (6) as follows:

$$y_i = \sum_{k=1}^{K} x[i + rk]w[k] \tag{6}$$

In this context, r represents the dilation rate, which establishes a gap between the core weights w. For example, a 3 × 3 core with a dilation rate of 2 will have a receptive field of the same size as a 5 × 5 core while using only nine parameters. This effectively expands the receptive field without incurring additional computational cost. Extended convolutions have gained prominence in real-time segmentation, and several recent publications have highlighted their use. Notable examples include the DeepLab series [66], strategies involving context aggregation at multiple scales [67], techniques such as dense upsampling convolution and hybrid dilated convolution (DUC-HDC) [68], densely connected Atrous Spatial Pyramid Pooling (DenseASPP) [69], and efficient neural networks (ENet) [70].

DeepLabv1 [71] and DeepLabv2 [66] are among the most widely recognized methods for image segmentation. The latter introduces

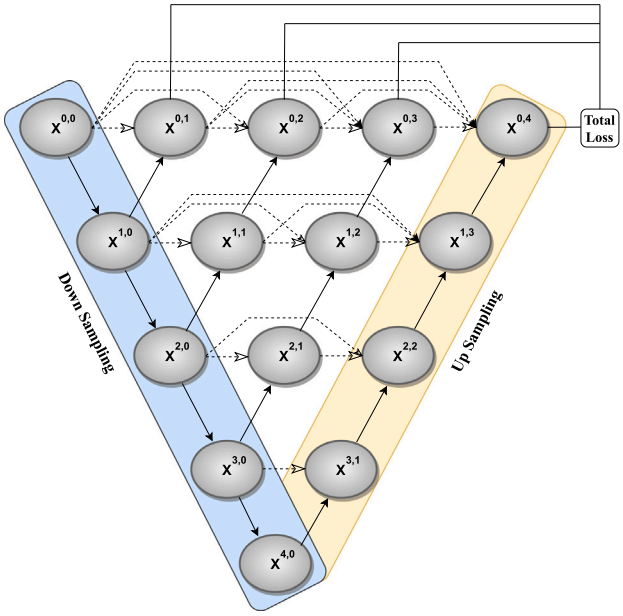


Fig. 11. The U-Net++ architecture [65].

three key characteristics. Initially, it utilized dilated convolution to counteract the reduction in resolution within the network caused by max-pooling and striding. Second, it employs Atrous Spatial Pyramid Pooling (ASPP), a technique that examines incoming convolutional feature layers using filters with diverse sampling rates. This strategy facilitates the comprehensive detection of objects and image context across multiple scales, thereby improving the segmentation of objects with robustness across various scales. Third, the approach improves object boundary localization by synergizing techniques from probabilistic graphical models and deep convolutional neural networks (CNNs). Fig. 13 visually represents the DeepLab model, which bears similarities to the reference but distinguishes itself through the use of ASPP and dilated convolution.

Subsequently, Chen et al. [72] introduced DeepLabv3, which merges cascaded and parallel modules of extended convolutions. The ASPP is structured using parallel convolution modules. Improvements include the incorporation of a 1 × 1 convolution and batch normalization into the ASPP structure. The results from these modules were concatenated and subsequently processed through another 1 × 1 convolution,

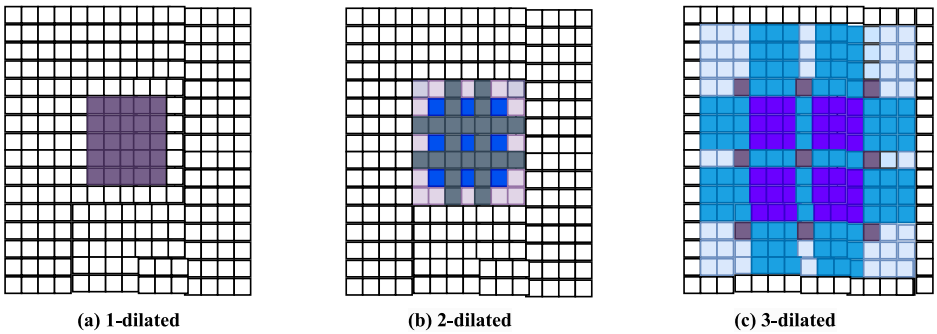


Fig. 12. Dilated convolution 3 × 3 kernel at different dilation rates smoother segmentation [1].

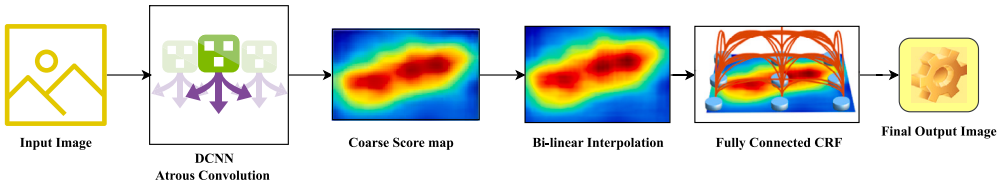


Fig. 13. DeepLab model architecture [66].

resulting in the final output with logits assigned to each pixel. Subsequently, in 2018, Chen et al. [73] unveiled Deeplabv3+, which utilizes an encoder–decoder architecture and introduces atrous separable convolution. This convolution technique comprises two steps: a depth convolution that operates spatially on each input channel, followed by a point convolution (1 × 1 convolution) with the output of the depth convolution. The DeepLabv3 framework served as the encoder. The most notable model features a modified Xception backbone with an increased number of layers, replacing max pooling and batch normalization with extended-depth separable convolution.

Wang et al. [74] utilized the DeepLab v3+ model for pathological slices of gastric cancer and, reported a sensitivity of 91.45%, specificity of 92.31%, and accuracy of 95.76%. The Dice coefficient of the model is 91.66%. Sun et al. [75] employed the Deeplabv3+ model specifically for the segmentation of liver tumors and the associated prediction of subsequent risks.

6.1.6. V-Net

Milletari et al. [76] introduced a 3D deformation structure known as V-Net based on the U-Net network architecture. The V-Net structure is shown in Fig. 14. The Dice coefficient loss function was used instead of the conventional cross-entropy loss function. This network employs a 3D convolution kernel for image convolution and reduces the channel dimension using a 1 × 1 × 1 × 1 convolution kernel. The left-hand side of the network is a progressively compressed path consisting of several stages, each containing one of the three convolutional layers. To guarantee the efficient learning of parameter functions in each stage, the inputs and outputs from each stage are combined to facilitate the learning of residual functions. The size of the convolution kernel utilized in each step of the convolution operations was 5 × 5 × 5 × 5. These convolution operations are utilized to extract data features while reducing the data resolution at the end of each step through appropriate step sizes. On the right-hand side of the network, there is a progressively decompressed path. This path extracts features and extends the spatial support of lower-resolution feature maps to collect and integrate necessary information to generate a dual-channel volume segmentation output. The final output size of the network is identical to its initial input size.

Türk et al. [77] leveraged the advanced capabilities of existing V-Net models, enhanced the hybrid V-Net model, and demonstrated impressive average Dice coefficients, achieving 97.7% for kidney segmentation and 86.5% for tumor segmentation. This indicates its potential as a reliable technique for segmenting soft-tissue organs.

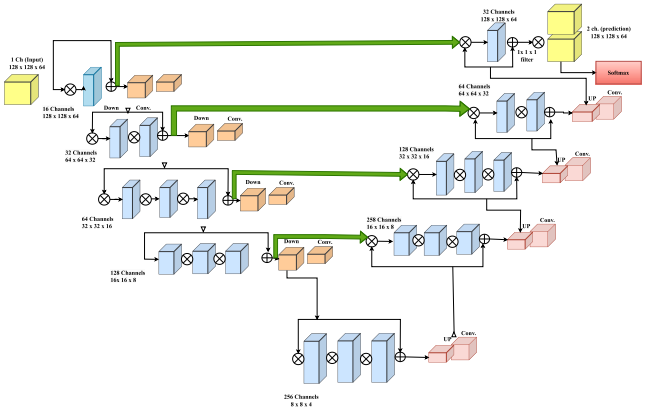


Fig. 14. Structure of V-Net [76].

6.1.7. Pyramid Scene Parsing Network (PSPNet)

PSPNet is a semantic segmentation model that uses a pyramid parsing module to combine contexts based on various areas to leverage global context information. The pixel-level classification network based on the FCN serves as the foundation for the pyramid scene parsing network [78]. Using multi-scale pooling layers, the feature maps from a ResNet-101 network are transformed into activations with various resolutions, which are then upsampled and combined with the initial characteristic map to accomplish segmentation. Auxiliary classifiers have been used to improve the learning process in deep networks, such as ResNet. The various pooling mechanism types are concentrated on various regions of the activation map. The spatial pooling pyramid is built by examining different zones of the activation map using pooling kernels of various sizes, such as 1 × 1, 2 × 2, 3 × 3, and 6 × 6. An illustration of the PSPNet is shown in Fig. 15.

Ye et al. [79] used this PSP-Net deep learning network to employ prostate tumor-based dataset sample identification, achieving segmentation accuracy near the Dice similarity coefficient and Hausdorff distance, surpassing conventional methods with a 91.3% Dice index, and faster processing. The predicted tumor markers closely matched those identified manually by the clinicians. Yan et al. [80] applied this model to prostate MRI, and PSP-Net achieved a leading segmentation accuracy of 98.65%.

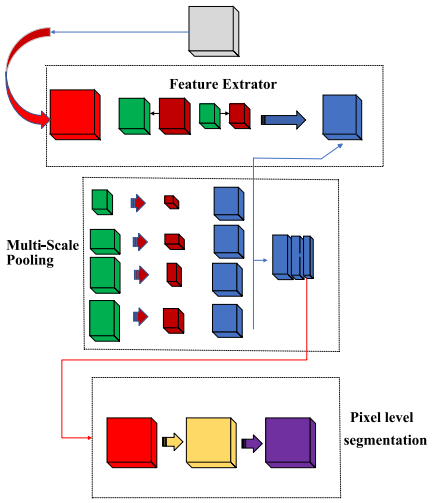


Fig. 15. An illustration of the PSPNet.

6.2. Instance segmentation

Instance segmentation is a computer vision task that involves identifying and precisely outlining individual objects at the pixel level within an image. As opposed to object detection, instance segmentation provides each object with a thorough segmentation mask, allowing for a more in-depth comprehension of the scene. Accurate object localization and segmentation are crucial for applications in robotics, autonomous vehicles, and medical image analysis. The capabilities of instance segmentation have significantly improved owing to frameworks such as Mask R-CNN, PANet.

6.2.1. Mask R-CNN

Region-based Convolutional Neural Network (R-CNN) [81] can identify objects in an image; however, it is unable to identify which pixel in the image belongs to an object. Hence, He et al. [82] proposed a mask R-CNN for instance segmentation in computer vision applications. It is an enhancement over the original Faster R-CNN [83] architecture, adding the capacity to predict segmentation masks at the pixel level.

The mask R-CNN comprises two heads: a bounding box head for object detection and a mask head for pixel-level segmentation. The backbone CNN was used for feature extraction, and the Region Proposal Network (RPN) was used to generate region proposals. For accurate feature extraction from the proposed regions, the ROI Align module was presented, and the mask head generated segmentation masks for every object. This architecture makes Mask R-CNN a flexible solution for tasks requiring fine-grained object understanding in computer vision. It can simultaneously identify and classify objects and provide detailed pixel-wise segmentation masks. Fig. 16 shows the architecture of mask R-CNN.

Shu et al. [84] used an improved Mask R-CNN to enhance multi-organ segmentation to support esophageal radiation therapy. Jeong et al. [85] suggested an advanced 3D Mask R-CNN approach for automatic segmentation of brain tumors in Dynamic Susceptibility Contrast Enhanced (DSCE) MRI perfusion images.

6.2.2. Path Aggregation Network (PANet)

The PANet was proposed by Liu et al. [86] in 2018. As an extension of the Mask R-CNN architecture, PANet was introduced to address problems with feature aggregation for improved instance segmentation. It focuses on capturing high-level contextual information as well as fine-grained details through top-down and bottom-up feature aggregation paths.

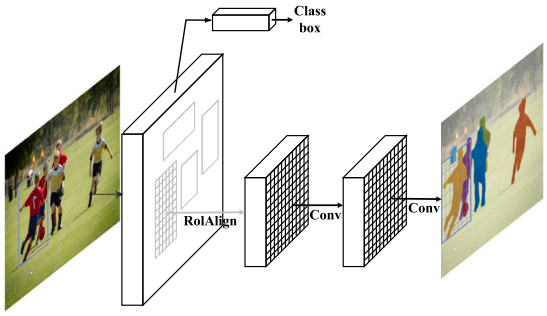


Fig. 16. The mask R-CNN framework [82].

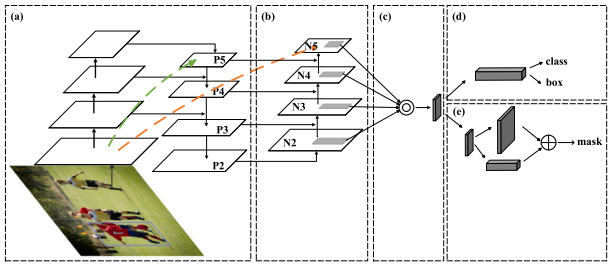


Fig. 17. The framework of PANet [86].

Fig. 17 shows an overview of the PANet architecture. Beginning with (a), the FPN backbone, which is recognized for its Feature Pyramid Network, serves as the foundation for feature extraction at various scales. Proceeding to (b), a pathway dedicated to aggregating fine-grained details from lower-level convolutional layers was introduced by Bottom-Up Path Augmentation. (c) A dynamic pooling operation is suggested using adaptive feature pooling, which is essential for maintaining spatial information during feature pooling. (d) The Box Branch is involved in object detection; it manages tasks such as class labels and bounding box prediction. Finally, (e) connected fusion is the process of combining features via fully connected layers, guaranteeing a thorough fusion of data from various channels. When combined, these elements highlight the novel methodology of PANet by highlighting the importance of local and global feature aggregation to improve instance segmentation performance.

Zhao et al. [87] applied PANet, adopting a coarse-to-fine approach, for multi-lesion segmentation in medical imagery. This framework has been explored in two key applications: 2D segmentation of various lung infections in CT scans and 3D segmentation of multiple lesions in brain MRIs.

Here, Table 7 discusses the advantages and limitations of the discussed algorithms.

7. Recent advancements

Here, we provide a brief overview of recent developments and notable achievements in the field of medical image segmentation. We demonstrate the present landscape that characterizes the cutting edge of image segmentation systems in the medical industry by shedding light on cutting-edge methodologies, technological advancements, and research findings. This study offers insightful information on the most recent advancements, innovative approaches, and enduring difficulties that researchers must deal with. Their joint efforts were intended to increase the usefulness of medical image segmentation techniques by improving their accuracy.

Moreover, recent advancements in medical image segmentation have introduced innovative models, each contributing uniquely to the evolution of the field. The relentless pace of innovation underscores

Table 7
The table presents the advantages and limitations of the algorithms used in medical image segmentation.

Algorithm	Advantages	Limitations
CNN	CNNs excel at learning features through weight sharing, enabling efficient hierarchical representations. CNNs are perfect for tasks like image recognition and classification because of their ability to recognize complex patterns.	CNNs, are prone to overfitting, particularly on limited data, due to numerous parameters and hierarchical features, increasing the risk of fitting noise. As a result, the model's capacity to generalize to new data is reduced.
FCN	FCNs preserve spatial information via upsampling, like transposed convolutions, a key advantage in semantic segmentation for preserving fine details and ensuring accurate pixel-wise classification.	FCNs often require large amounts of memory for both training and inference, which presents problems in situations where resources are limited. This restricts their applicability on memory-constrained devices.
U-Net	U-Net's unique U-shaped architecture captures fine details in images through skip connections, maintaining high-resolution data in both encoding and decoding. It excels at tasks like segmenting biomedical images.	U-Net frequently encounters situations involving imbalanced class distributions. The segmentation performance may be affected as a result of its inability to reliably distinguish under-represented classes.
U-Net++	U-Net++ improve information flow and feature learning by implementing nested skip pathways, leading to increased performance. This architecture improves segmentation accuracy, especially when it comes to more accurately representing complex structures	U-Net++ uses nested skip pathways to increase computational complexity, improving performance but using more resources. This could make it less useful in environments with limited processing power.
DeepLab	DeepLab extracts dense features, which are essential for tasks like semantic segmentation, using dilated convolutions. This improves the ability of the model to identify complex patterns by enabling finer input analysis.	DepLab experiences computational expenses, especially with high-resolution inputs. Dilated convolutions increase the computational load, demanding significant processing and memory resources.
V-Net	V-Net is designed for 3D medical image segmentation and it efficiently captures volumetric data with skip connections and 3D convolutions. It excels in tasks like organ segmentation, highlighting its importance in medical imaging.	When processing large 3D volumes, V-Net encounters computational limitations. The use of 3D convolutions and skip connections increases the computational load, posing challenges in handling extensive datasets.
PSPNet	PSPNet capture multi-scale contextual information using pyramid pooling modules, enabling simultaneous analysis at various scales. This improves the model's comprehension of the global context in scene parsing tasks.	Pyramid pooling modules increase the computational cost of PSPNet. The model's enhanced multi-scale information capture demands additional computations, increasing resource requirements in both training and inference.
Mask R-CNN	Mask R-CNN excels in providing precise pixel-level masks for individual objects, crucial for tasks requiring accurate instance-level segmentation.	The mask generation stage of Mask R-CNN can be computationally demanding, which presents difficulties for real-time deployment and on devices with limited resources.
PANet	Through the use of a path aggregation module, PANet improves information flow between network layers, enhancing contextual awareness and improving object detection performance.	The additional modules in PANet make it more complex and require more computing power, which may limit its applicability when computational power is limited.

a transformative period in medical informatics, marking significant progress toward precision medicine, in which diagnostics and treatments are precisely tailored for individual patients. With the increase in computational capabilities and refinement of algorithmic sophistication, the vision of fully automated, highly accurate medical imaging analysis approaches reality. This advancement is pivotal for enhancing disease identification, refining treatment planning, and effectively monitoring therapeutic outcomes, which are fundamental aspects of patient-centric healthcare. The integration of these state-of-the-art technologies into clinical workflows promises overhaul healthcare delivery, leading to improved patient outcomes, streamlined operations for medical professionals, and the emergence of groundbreaking diagnostic techniques. These concerted efforts are intended to increase the usefulness of medical image segmentation techniques by improving their accuracy, ensuring that each step forward in the technology translates into a leap forward in patient care and clinical outcomes.

The continuous evolution of these medical image segmentation technologies heralds a new era in healthcare analytics, in which big data and advanced algorithms converge to provide deeper insights into patient health. The integration of predictive analytics with segmentation technologies can significantly improve predictive evaluations, enabling earlier interventions and personalized treatment pathways. Advances in federated learning also promise to harness the power of decentralized data while preserving patient privacy, thereby offering a scalable solution for global collaborative research. Furthermore, the robustness of AI systems against adversarial attacks and their ability to adapt to new emerging medical conditions will become paramount, necessitating ongoing research and development. These future-oriented efforts will not only enhance the clinical efficacy of medical image segmentation but will also pave the way for broader adoption in routine clinical practice, ultimately contributing to a more resilient and responsive healthcare system. These innovative models include Multi-resolution Attention and Multi-scale Convolution Network (MAMC-Net), Automatic weighted Dilated convolutional network (AD-Net), Residual Attention based Uncertainty-Guided Mean Teacher (RA-UGMT), Joint Gate Attention Residual U-Net (JGate-AttResUNet), Context-coordination multi-atlas Boundary-Aware UNet-(C^2BA-UNet), Edge U-Net, Deep Residual U-Net (dResU-Net), Attention mechanism based Cystoscopic images Segmentation (ACS), Hierarchical analysis and Weight Adjustment Segmentation Network 3 (HWA-SegNet3), ChimeraNet, Efficient Group Enhanced UNet (EGE-UNet), Multi-class Dilated D-net (MD^2N) framework, Psoriasis Lesion Segmentation NetV2 (PsLSNetV2), Mayfly Optimizer Fuzzy U-Net (MFO-Fuzzy U-net), Residual Bi-directional Convolutional U-Net (ResBCU-Net), Modified ChimeraNet, Backchannel Filling Convolutional Neural Network (BCF-CNN) + Level Sets Segmentation (LS), Fully Convolution Encoder Decoder Network (FCEDN), Localization and Enhancement (LE) + U-Net, Lung Dense Attention Network (LDANet), Modified U-Net, Optimized Net, U-Net with EfficientNet-b4 Encoder, Glonal + Regional U-Nets, Debnath Bhattacharyya-Network (DB-NET), hyper-dense VGG16 U-Net, Dual encoder–decoder CNN architecture (DED-CNN). Table 8 provides a detailed analysis of the outcomes achieved by some state-of-the-art medical image segmentation applications.

7.1. Performance analysis

In this section, we assess the performance of the recent advancements listed in Table 8, focusing solely on their performance metrics when applied to the datasets on which they were originally tested. Researchers have employed a variety of datasets in the field of tumor

Table 8
Comparative performance and limitations of existing DL-based medical image segmentation survey papers from 2022 & 2023.

Reference	Application	Dataset	Pre-processing method	Model	Results	Device used	Limitations
Luo et al. [16]	Tumor segmentation	STS & Own Dataset (Lymphoma)	Resizing, Normalization, Data augmentation	C^2BA-UNet	Dice: STS: 78.1%, Own Dataset: 83.4% IoU: STS: 64.9%, Own Dataset: 72.2%	NVIDIA RTX 3070 GPU	Limited capacity to learn from 3D imaging data, affecting comprehensive tumor segmentation.
Zeng et al. [17]	Tumor segmentation	CAMELYON 2017 & BOT Challenge	Overlapping, Resizing, Data augmentation	MAMC-Net	Accuracy: CAMELYON 2017: From 91.7% to 94.7% BOT Challenge: From 82.6% to 92.9% Dice: CAMELYON 2017: From 88.3% to 92.9% BOT Challenge: From 80.8% to 89%	NVIDIA GTX 2080 Titan GPU	Due to variations in color, illumination, staining quality, and image acquisition techniques in whole-slide images from various medical institutions, model generalization across a range of medical imaging conditions needs to be improved. Future enhancements to address these variations are necessary.
Tyagi et al. [88]	Tumor segmentation	NSCLC-Radiomics & Own Dataset	Resizing, Clipping, Data augmentation	Based on CNN and ViT	Dice: NSCLC-Radiomics: 74.68% Own Dataset: 68.47%	16 GB NVIDIA P100 GPU	Dice loss, used for class-imbalanced data in the context of binary cross-entropy, focal loss, or their combinations, exhibits the limitation of an unstable gradient. This may impact the model's performance and dependability, especially in scenarios where class imbalance is significant.
Dar et al. [89]	Tumor segmentation	BUSI & Own Dataset	Resizing, Normalization, Data augmentation	EfficientU-Net	Dice: BUSI: 90.45% ± 0.7% Own Dataset: 92.27% ± 1.1% IoU: BUSI: 82.58% ± 1.2% Own Dataset: 85.67% ± 2.0%	T4 and P100 GPUs	While the proposed model shows promising results in the specific context of breast tumor segmentation in ultrasound images, more research and validation may be necessary to assess its efficiency and adaptability in a wider range of medical imaging scenarios, including those involving various cancer types or imaging modalities.
Peng et al. [90]	Tumor segmentation	BraTS19, BraTS20 & FeTS21	Randomization, Normalization, Data augmentation	AD-Net	Dice: BraTS19: ET: 76%, WT: 90%, TC: 91% FeTS21: ET: 80%, WT: 91%, TC: 85%	NVIDIA RTX 2080 Ti	Limited loss optimization in comparative analysis due to architectural limitations of reproduced models, leading to potential errors from reliance on cross entropy loss.
Farooq et al. [91]	Tumor segmentation	BUSI & UDIAT	Handling Speckle Noise, Improving Visibility	RA-UGMT	Dice: BUSI: 80.82%, UDIAT: 83.56% Precision: BUSI: 85.26%, UDIAT: 91.49%	NVIDIA RTX 3090 GPU	A flaw in the basic network architecture negatively impacts segmentation performance, particularly in complex breast tumor cases in 2D ultrasound images.
Ruba et al. [92]	Tumor segmentation	BraTS 2015 & BraTS 2019	Resizing, Normalization	JGate-AttResUNet	Accuracy: BraTS 2015: 94.99%, BraTS 2019: 93.69% Dice: BraTS 2015: 89.60%, BraTS 2019: 91.31%	N/A	Limited generalizability due to focus on BRATS 2015 and 2019 datasets. Further testing on diverse datasets needed for robustness and suitability confirmation in various clinical contexts.
Allah et al. [93]	Tumor segmentation	Brain Tumor dataset by [94]	Denoising, Resizing, Data augmentation	Edge U-Net	Dice: Meningioma: 88.8%, Glioma: 91.76%, Pituitary: 87.28% Accuracy: Meningioma: 99.51%, Glioma: 99.23%, Pituitary: 99.35%	NVIDIA Tesla P100 -PCIE GPU	Intensity variation in brain tumors due to different imaging modalities and MRI noise, causing imprecise textures and blurred boundaries.

(*continued on next page*)

Table 8 (*continued*).

Raza et al. [95]	Tumor segmentation	BraTS 2020	Resizing, Standardization, Normalization	dResU-Net	Dice: TC: 83.57%, WT: 86.60%, ET: 80.04%	Tesla T4 GPU	Initial overfitting with residual blocks at each encoder stage; balanced by using them only in the first four levels.
Zhang et al. [96]	Tumor segmentation	Own Dataset	Denoising, Partitioning, Resizing, Data augmentation	ACS	Dice: 82.7% MIoU: 69%	NVIDIA RTX 3090 GPU	Training on a particular dataset may limit generalizability and high computational requirements may restrict clinical use.
Mohakud et al. [26]	Skin image segmentation	ISIC 2016 & ISIC 2017	Resizing	FCEDN	Dice: ISIC2016: 98.483% ISIC2017: 87.234% Jaccard: ISIC2016: 96.413% ISIC2017: 86.85%	Tesla P100 GPU	Effectiveness may vary among datasets, especially with varying image quality, highlighting the need for more testing and validation.
Badshah et al. [27]	Skin image segmentation	ISIC 2018	Resizing	ResBCU-Net	Jaccard: 94% Specificity: 98%	N/A	Insufficient dataset size for effective training; early overfitting mitigated by dropout layers and Gaussian noise, yet a challenge for training and generalizability.
Han et al. [28]	Skin image segmentation	ISIC Challenge	Normalization, Resizing, Data augmentation	HWA-SegNet3	Accuracy: 96.33% Dice: 91.38%	NVIDIA RTX 3090 GPU	Focus on gray-scale features limits generalizability and applicability in diverse dermatological conditions.
Ruan et al. [29]	Skin image segmentation	ISIC 2017 & ISIC 2018	Resizing, Normalization, Data augmentation	EGE-UNet	Dice: ISIC2017: 88.77% ± 0.06% ISIC2018: 89.46% ± 0.07% MIoU: ISIC2017: 79.81% ± 0.10% ISIC2018: 80.94% ± 0.11%	NVIDIA RTX A6000 GPU	Model specific to skin lesion segmentation, limiting application to other tasks. Future work aims to broaden its scope.
Kumar et al. [30]	Skin image segmentation	ISIC Challenge	Denoising, Image Enhancement, Data augmentation	MD^2N	Accuracy: 97.462% Precision: 93.627% Recall: 99.721%	Intel Core i3-6100CPU	Performance variation on different datasets; necessitates further testing and validation for robustness and applicability.
Bindhu et al. [31]	Skin image segmentation	ISIC 2016 & ISIC 2017	Denoising	MFO-Fuzzy U-net	Accuracy: 99.13% Specificity: 94.65%	N/A	Limited generalizability, optimized for ISIC 2016 and 2017; may not adapt well to varying lesion types and conditions.
Lama et al. [32]	Skin image segmentation	ISIC 2017	Resizing, Normalization, Data augmentation	Modified ChimeraNet	Accuracy: 94.8% Jaccard: 80.7% Dice: 88%	32 GB Nvidia V100 GPU	Shortcomings include single dermatologist evaluation and lack of new ground truths for noisy data segments, affecting noise level evaluation.
Huang et al. [33]	Skin image segmentation	ISIC 2016 & ISIC 2017	N/A	BCF-CNN + LS	Dice: ISIC2016: 91.2% ISIC2017: 85.5% Jaccard: ISIC2016: 85.1% ISIC2017: 77% Accuracy: ISIC2016: 95.3% ISIC2017: 93.0%	NVIDIA RTX 2080 Ti	Difficulty in segmenting areas with dark color and similar texture, and lesions with blurry borders.
Raj et al. [97]	Skin image segmentation	Own Dataset	Resizing, Normalization	PsLSNetV2	Dice: 97.43% Jaccard: 95.05%	NVIDIA Quadro P4000 GPU	Dataset exclusions and variability due to uncontrolled environments; focuses on Indian patients with plaque psoriasis.
Rahman et al. [20]	Lung image segmentation	MC, JSRT & UTMB	Resizing, Normalization	Modified U-Net	Accuracy: 96.88% Dice: 95.32% IoU: 93.15%	NVIDIA GTX 1080 Ti GPU	Requires many images, addressed by patching; time-consuming post-processing; more suited for bacterial infections, sensitive to small datasets.

(*continued on next page*)

segmentation, including NSCLC-Radiomics, BUSI, CAMELYON 2017, BOT Challenge, BraTS19, BraTS20, FeTS21, UDIAT, BraTS 2015, STS, and private or own datasets. Notably, BraTS and BUSI datasets have been commonly used in multiple studies.

Focusing on the BraTS20 dataset, Peng et al. [90] introduced AD-Net, a model that leveraged an automatic weighted dilated convolutional network. This model efficiently extracts multimodal features that are crucial for accurate brain tumor segmentation. The methodology includes dual-scale convolutional feature maps for channel feature separation and deep supervision training techniques. For preprocessing, AD-Net employs a randomization strategy, including 3D random clipping, rotation, intensity enhancement, mirror inversion, and normalization. The results were impressive, with Dice scores of 90%, 80%, and 76% for the whole tumor (WT), tumor core (TC), and

Table 8 (*continued*).

Arvind et al. [21]	Lung image segmentation	MC, JSRT & Shenzhen Chest X-rays	Resizing, Cropping, Data augmentation	Optimized U-Net	Accuracy: 93.87% Dice: 91.87%	Nvidia GTX 1050 Mobile	Limited number of images may not capture full range of lung conditions, affecting model's generalizability.
Liu et al. [22]	Lung image segmentation	MC & JSRT	Resizing, Normalization, Horizontal Flip	U-Net with EfcientNet-b4 Encoder	Accuracy: MC: 98.9% JSRT: 98.5% Dice: MC: 97.7% JSRT: 97.9% Jaccard: MC: 95.5% JSRT: 95.8%	NVIDIA RTX 3060 GPU	Varying accuracy depending on disease type in CXR images, with challenges in segmenting certain conditions.
Alebiosu et al. [24]	Lung image segmentation	ImageCLEF 2019 TB tasks	Resizing, Data augmentation	DAvoU-Net	Dice: 98.34% Sensitivity: 97.82% Specificity: 99.45%	NVIDIA Tesla K40c GPU	Converting 2D weights to 3D may not capture spatial intricacies of 3D images, impacting accuracy.
Bhat-tacharyya et al. [25]	Lung image segmentation	LUNA16	Labeling, Data augmentation	DB-NET	Dice: 88.89% ± 11.71% Sensitivity: 90.24% ± 13.15%	N/A	BiFPN with ReLU increases computational demands and overhead, impacting efficiency in resource-limited settings.
Ilhan et al. [98]	Lung image segmentation	Own dataset	Masking, Resizing, Localization, Enhancement	LE method + U-Net	Accuracy: 97.75% Sensitivity: 87.46% Precision: 82.06% Dice: 85% Jaccard: 74%	NVIDIA RTX 2080 GPU	Limitations include training on segmented images only and a limited number of images, affecting segmentation performance.
Chen et al. [99]	Lung image segmentation	LIDC-IDRI & CO-Seg	Resizing, Labeling, Data augmentation	LDANet	Dice: LIDC-IDRI: 98.43% CO-Seg: 98.31% Accuracy: LIDC-IDRI: 99.52% CO-Seg: 99.52%	NVIDIA Quadro RTX 5000 GPU	2D input limits spatial information capture, affecting RSA mechanism; long training time is a concern.
Park et al. [100]	Lung image segmentation	Own Dataset	Cropping	Global + Regional U-Nets	Mean Dice: 78% Median Dice: 90%	N/A	2D slice output may lead to discontinuity in binary segments in coronal and sagittal slices.
Alshmrani et al. [101]	Lung image segmentation	STS	Resizing, Normalization, Data augmentation	Hyper-dense VGG16 U-Net	Dice: 73.01% Accuracy: 98.10% IoU: 55.66%	NVIDIA Tesla P100-PCIE GPU	Untested generalizability across different medical imaging modalities and conditions.
Ullah et al. [102]	Lung image segmentation	JSRT, MCCXR & SCXR	Normalization, Resizing	DED-CNN	Dice: JSRT: 97.60% MCCXR: 97.67% SCXR: 95.68% IoU: JSRT: 95.50% MCCXR: 95.48% SCXR: 91.84%	NVIDIA 1080Ti GPU	Model only accomplishes segmentation tasks and is computationally expensive due to dual encoder and decoder architecture.

enhancing tumor(ET), respectively, showing good performance in multimodal learning.

In another study, Raza et al. [95] proposed dResU-Net, a hybrid model that combines a deep residual network with the U-Net architecture. It was designed to overcome the vanishing gradient problem and retain low-level features by using residual blocks and adaptive skip connections. This approach enhances both low-level and high-level feature representations, which are crucial for segmenting brain tumors. Preprocessing was minimal, as the dataset had already undergone standard preprocessing phases. dResU-Net achieved Dice scores of 83.57%, 86.60%, and 80.04% for the tumor core (TC), whole tumor (WT), and enhancing tumor(ET), respectively, on the BraTS 2020 dataset, demonstrating its effectiveness in handling multimodal MRI data with considerable complexity.

Ruba et al. [92] showed a significant leap in brain tumor segmentation using JGate-AttResUNet. This model integrates a novel joint gate-attention mechanism within the residual U-Net architecture. The attention gates in the model are key to enhancing the tumor localization and segmentation accuracy. Intensity normalization during preprocessing ensured the consistency of the input data. JGate-AttResUNet outshined the mean Dice values of 0.896 and 0.913 for the BRATS 2015 and 2019 datasets, respectively. These results were attributed to its advanced attention mechanism, which is effective in capturing intricate tumor boundaries and improving segmentation accuracy across diverse datasets.

The JGate-AttResUNet model demonstrated superior performance primarily because of its innovative use of joint gate-attention modules. These modules provided enhanced feature extraction and precise tumor localization, which are crucial for the complex anatomy of the brain. Additionally, the attention mechanism in the model refined the focus on relevant regions, substantially reducing false positives and improving the overall accuracy. The preprocessing technique of intensity normalization further contributes to the robustness of the model, ensuring standardized input data quality across different datasets. Consequently, the ability of JGate-AttResUNet achieve higher Dice scores, especially in complex scenarios presented in the BRATS 2015 and 2019 datasets, underscores its superiority in brain tumor segmentation.

In the context of the BUSI dataset, Farooq et al. [91] developed Residual-Attention U-Net, a cutting-edge model for segmenting breast masses in ultrasound images. This model integrates residual learning with attention mechanisms to optimize the deep network processing and focus during learning. Utilizing a semi-supervised approach, it leverages both labeled and unlabeled data to address the scarcity of labeled medical imaging data. Preprocessing included adaptive histogram equalization and speckle noise reduction. On the BUSI dataset, the Residual-Attention U-Net achieved a Dice score of 80.82%, attributing its success to its innovative architecture that efficiently processes

high and low-level features and its effective use of unlabeled data for improved segmentation accuracy.

On the other hand, Dar et al. [89] introduced EfficientU-Net, a model designed for breast tumor detection in ultrasound images. This model innovates the traditional U-Net architecture by integrating efficient convolutional networks and, enhancing feature extraction and learning efficiency. It incorporates multiscale feature fusion to combine features from various levels of the network, providing a comprehensive feature representation. During preprocessing, EfficientU-Net applies contrast enhancement, noise reduction, and image normalization. The model's performance on the BUSI dataset was remarkable, achieving a Dice score of 90.45%, notably outperforming Farooq et al. [91]'s model. EfficientU-Net demonstrated impressive results with a high Dice score, demonstrating its capability to accurately segment breast tumors in diverse ultrasound images.

EfficientU-Net and Residual-Attention U-Net are notable for their advancements in breast tumor segmentation in ultrasound images. Although both models achieve high levels of results, their core strengths lie in different areas. EfficientU-Net excels owing to its efficient convolutional networks and multi-scale feature fusion, making it highly effective in feature extraction and learning. In contrast, Residual-attention U-Net stands out for its combination of residual learning and attention mechanisms, along with its semi-supervised learning approach. This allows it to effectively utilize a wider range of data, including unlabeled samples, thereby enhancing its learning capability.

Researchers have explored a variety of datasets in the area of skin image segmentation, with a notable emphasis on the ISIC Challenge datasets (spanning different years, such as ISIC 2016, ISIC 2017, and ISIC 2018), as well as other resources such HAM10000 and private or own custom datasets. In particular, the ISIC datasets have been a popular choice for multiple studies because of their extensive collection of skin lesion images.

Han et al. [28] developed HWA-SegNet3 to tackle the inherent challenges of skin lesion image segmentation by, incorporating a discrete Fourier transform (DFT) for data enrichment and a hierarchical dilated analysis module. This combination is designed to understand the complex semantic features in multi-channel data and to fine-tune the pre-prediction results through a weight adjustment structure. On the ISIC Challenge dataset, HWA-SegNet3 demonstrated a notable improvement in segmentation, achieving an accuracy of 96.33%, surpassing the U-Net model. On the other hand, Kumar et al. [30] proposed the Multi-class Dilated D-Net (MD^2N) framework, a deep learning model for the segmentation of skin cancer. It employs a two-stage process involving an adaptive anisotropic diffusion filter for pre-processing and a multi-class dilated D-Net for segmentation tasks. This model was specifically tailored to minimize feature information loss while accurately recognizing small skin lesion patches. The MD^2N model showed remarkable performance metrics, achieving an accuracy of 97.462%. This level of performance outperformed other contemporary models, including DenseNet and EfficientNetB4.

The superior performance of the MD^2N model can be ascribed to its innovative approach for handling the challenges of skin cancer segmentation. The adaptive anisotropic diffusion filter significantly enhances the quality of the input images, reduces noise, and improves classification performance. The multi-class Dilated D-Net framework, meticulously developed with various layers and filter sizes, increased efficacy and performance, particularly in dealing with the complex variability of skin lesion images. The model's focus on reducing feature information loss and accurately identifying small lesion patches addressed critical challenges in skin cancer segmentation, leading to outstanding performance.

Focusing on the ISIC 2016 and 2017 datasets, Mohakud et al. [26] proposed Fully Convolution Encoder Decoder Network (FCEDN). The proposed model integrates a novel Exponential Neighborhood Gray Wolf Optimization (EN-GWO) algorithm for optimizing network hyper-parameters. This approach balances exploration and exploitation strategies to enhance segmentation performance. The FCEDN only employs resizing to preprocessing the dataset. The FCEDN achieved a remarkable Dice score of 98.483% on the ISIC 2016 dataset and a Dice score of 87.234% on the ISIC 2017 dataset. Huang et al. [33] introduced the BCF-CNN + LS model. This method combines CNNs with level sets by introducing a backchannel filling (BCF) method. It leverages a CNN to estimate the target position and integrates it with level set segmentation for evolution and accuracy enhancement. The CNN-Level Set combined approach achieved a Dice score of 91.2% on the ISIC 2016 dataset and a Dice score of 85.5% on the ISIC 207 dataset. On the other hand, Bindhu et al. [31] presented a novel deep learning model called MFO-Fuzzy U-Net for the segmentation process to extract the affected area from skin cancer images. This model combines the Fuzzy U-Net architecture with the May Fly Optimizer (MFO) to improve segmentation accuracy. Fuzzy U-Net is designed to handle segmentation by processing pre-processed images, which are optimized using the May Fly Optimizer to enhance the accuracy range of the model. The author only used the denoising technique for preprocessing and achieved an accuracy of 99.13%. It outperformed traditional networks such as YOLO net, SegNet, Mask RCNN, U-net, LinkNet-B7, and FCNs in terms of accuracy and other performance metrics.

The FCEDN model performed best on the ISIC 2016 dataset. Its success is attributed to the EN-GWO algorithm, which effectively optimized the network hyper-parameters, resulting in higher segmentation accuracy, particularly in terms of the Jaccard and Dice coefficients. On the ISIC 2017 dataset, the MFO-Fuzzy U-net model exhibited superior performance. The use of bilateral filtering for noise reduction and optimization by the May Fly Optimizer contributed significantly to its high performance.

In the field of lung image segmentation, researchers have utilized various datasets, including LIDC-IDRI, MC, JSRT, and STS, along with their own. Notably, MC and JSRT datasets have been commonly used in multiple studies.

Rahman et al. [20] proposed a novel framework for segmenting the lung regions in chest X-ray images. This methodology employs a "divide and conquer" strategy, involving two distinct deep learning models: a CNN and a modified U-Net architecture. The CNN classified small image patches as lung or non-lung areas, whereas the adapted U-Net model segmented these patches. Both models work in tandem with their pre-segmented outputs combined using a binary disjunction operation. This ensemble approach uses preprocessing techniques, such as resizing and normalization, coupled with post-processing techniques, such as erosion and dilation, to enhance the accuracy and robustness of lung region segmentation, particularly in images with complex pulmonary conditions. The model performed remarkably well on the MC and JSRT datasets, achieving a Dice score of 95.32%. In another study on the same datasets, Arvind et al. [21] developed a modified and lightweight version of the U-Net model for lung segmentation of chest X-ray images. The key features of this improved model include the incorporation of dropout layers in both convolutional and deconvolutional layers to prevent overfitting and the application of batch normalization in each convolution layer to handle real-world image variations. In addition, the model reduces the number of filters in comparison to the standard U-Net, which minimizes its complexity while maintaining high segmentation performance. This design makes the model both efficient in terms of computational resource usage and effective in clinical applications, particularly where computational resources are limited. For preprocessing, the optimized U-Net employed resizing, cropping, and data augmentation, and achieved a Dice score of 91.87%. Finally, Liu et al. [22] proposed an enhanced version of the U-Net model, integrating a pre-trained EfficientNet-b4 as the encoder and employing residual blocks and the LeakyReLU activation function in the decoder. This improved U-Net model was designed for the automatic segmentation of lung regions in chest X-ray (CXR) images. The integration of EfficientNet-b4 enables efficient extraction of lung field features, and the residual blocks and LeakyReLU help to avoid gradient instability caused by the multiplication effect in gradient

backpropagation. The preprocessing techniques employed in this study were resizing, normalization, and data augmentation. The proposed model achieved the highest Dice score of 97.7% on the MC dataset and a Dice score of 97.9% on the JSRT dataset among the studies we covered.

Among the three authors, Liu et al. [22] achieved the best performance. Their enhanced U-Net model, integrating a pre-trained EfficientNet-b4 as the encoder and employing residual blocks and LeakyReLU in the decoder, achieved the highest Dice scores for both datasets. This superior performance is attributed to the advanced architecture of the model, which improves feature extraction and provides greater stability in gradient propagation. Despite the computational efficiency of Arvind et al.'s [21] model and the robust segmentation capability of Rahman et al.'s [20] ensemble approach, Liu et al.'s [22] model stands out for its accuracy, making it the most effective among the three in terms of segmentation accuracy on the datasets.

8. Challenges and future research directions

Addressing the challenges and clearing the path for further research in the field are crucial for its advancement. This section discusses the challenges encountered by researchers and practitioners when, exploring the complex terrain of unresolved issues and obstacles. From data-related challenges to technological limitations, we navigate the current impediments that demand innovative solutions. With an eye toward the future, we show the way forward for future lines of inquiry by identifying unknown areas and new trends.

8.1. Challenges in Medical Image Segmentation

The term "Challenges in Medical Image Segmentation" in medical image segmentation refers to a variety of problems that make it more difficult to extract accurate and clinically meaningful information from medical imaging data. This task is complicated by the complexities of human anatomy, the wide range of disease presentations, and subtle distinctions between normal and abnormal tissues. Technically speaking, these difficulties are made even more difficult owing to the constraints of imaging modalities, where noise, resolution, and contrast can all have a substantial effect on the quality of the images that are captured. Deep learning presents a promising field for automated segmentation; however, there are a number of data-related challenges to overcome, including handling non-uniform data distributions, requiring large annotated datasets, and protecting personal data. A key area of research is the creation of computationally efficient, clinically-interpretable models that generalize well across a wide range of populations and imaging conditions. Another crucial issue that needs to be resolved is the incorporation of these sophisticated computational models into the clinical workflow to enhance rather than interfere with medical practice. These difficulties demonstrate the necessity of interdisciplinary cooperation to progress the field of medical image segmentation. Fig. 18 visualizes some of the major challenges in medical image segmentation and their diverse concerns.

8.1.1. Data diversity and quality

The range and integrity of the data used to train and assess the models are referred to as the data diversity and quality. To ensure that the data accurately represent the variability found in the real world, this includes the diversity and representativeness of patient demographics and disease conditions in the dataset. One of the primary challenges is the relatively small size and lack of diversity in available datasets. Medical datasets often suffer from limited patient numbers, and even fewer instances of rare conditions. Zhang et al. [96] highlighted a critical limitation in the field of bladder tumor segmentation using deep learning methods, noting the challenge posed by limited dataset size and diversity. They emphasized that the scarcity of cystoscopic images restricts the capacity of deep learning model to learn varied tumor shapes and patterns, thereby underlining the necessity for more expansive and diverse cystoscopic datasets.

Diseases can manifest in various manners in different patients. Factors such as age, genetics, lifestyle, and disease progression stage contribute to this variability. Luo et al. [16] acknowledged the challenge posed by variability in disease characteristics, particularly in bladder tumors. They noted distinct morphological differences in tumors across various grades and stages, which manifest in the appearance, color, and size of lesions in cystoscopic images. This variability underscores the complexity of accurately diagnosing and treating bladder cancer, as it requires models capable of interpreting a wide range of tumor presentations. Rahman et al. [20] addressed the significant challenge posed by the variability in disease characteristics, especially in pulmonary diseases. They highlight the difficulties encountered in lung segmentation due to opacities or consolidation caused by pulmonary diseases, fluid, or bacterial infections. This unpredictability makes it difficult for Computer-Aided Diagnosis (CAD) systems to be successful because it causes poor contrast between the damaged lung area and the surrounding region, particularly in patients with pulmonary illness. The challenge presented by the variability in disease characteristics is emphasized by Ilhan et al. [98]. They noted that CT images of COVID-19 patients display diverse characteristics, that are critical for diagnosing the disease and assessing its severity. However, this variability, particularly in irregular shapes and low-contrast occurrences of infected regions, presents significant challenges for automatic lesion segmentation.

Medical images often vary in intensity, and are susceptible to noise. These variations can be due to differences in imaging equipment, settings, or even patient positioning. Such inconsistencies can hinder the ability of a model to learn relevant features for accurate segmentation. Allah et al. [93] identified significant challenges related to intensity variations and noise in imaging, particularly in MRI systems. They noted that variations in brain tumor tissue intensity, influenced by the imaging protocol, imaging modality, and inherent random noise in MRI systems, lead to unclear tissue textures and blurry tumor boundaries.

Labeling medical images requires expert knowledge, which makes it a time-consuming and expensive process. This scarcity of labeled data limits the amount of high-quality training material available for developing deep learning models. Farooq et al. [91] identified the scarcity of publicly available labeled data as a significant obstacle in the development of precise and robust deep-learning models.

Whole-slide imaging presents its own set of challenges, primarily because of the large size and high resolution of the images. Analyzing such large datasets requires substantial computational resources and efficient algorithms to process and extract meaningful information. Zeng et al. [17] acknowledged the complexity inherent in whole-slide image analysis. They highlight the challenges associated with manually annotating tumor areas in WSIs, noting the labor-intensive and time-consuming nature of this process.

In many cases, the available sample size is small, and the features expressed in the images may not be sufficient for deep learning models to effectively learn. When a model overfits, it performs well on training data but poorly on fresh, untested data. This deficiency can lead to overfitting. Critical problems were recognized by Han et al. [28], including a limited sample size and inadequate feature expression. They pointed out that limitations in the collection process, such as lighting and image capture equipment, lead to non-uniform scales and less pronounced features in the images. This, in turn, hampers the ability of the model to effectively extract features, and the relatively small amount of sample information exacerbates the issue, highlighting the need for advanced strategies in data augmentation and feature extraction. This particular challenge in medical image segmentation highlights the need for larger and, more diverse datasets and improved image quality to enhance the effectiveness and accuracy of deep learning models, ensuring that they are robust and applicable across various medical scenarios.

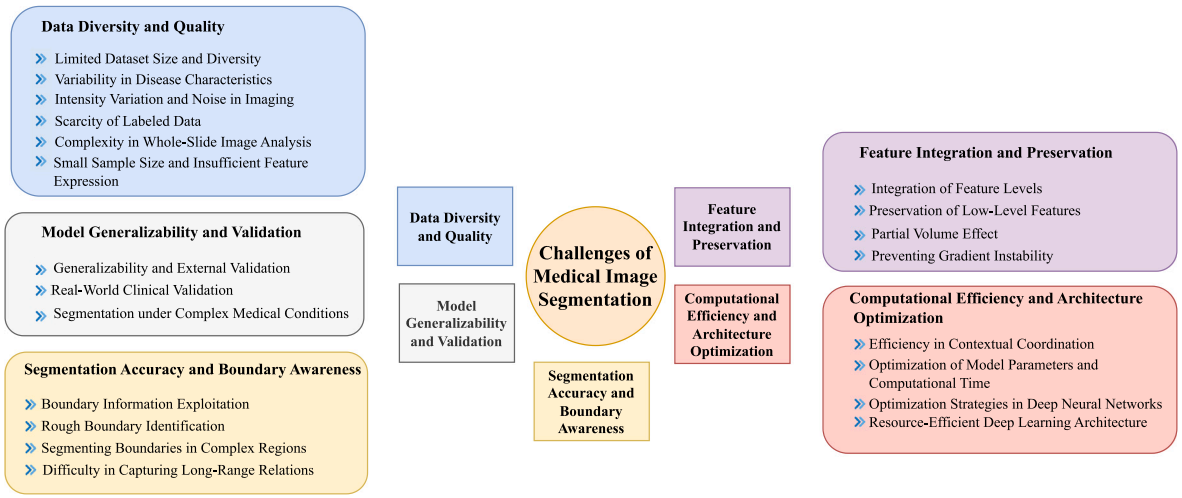


Fig. 18. Challenges of DL in medical image segmentation.

8.1.2. Model generalizability and validation

Model generalizability and validation in medical image segmentation focus on ensuring that deep learning models are not only accurate in controlled or experimental settings, but are also effective and reliable in diverse, real-world clinical environments. It emphasizes the need for models to generalize across different datasets and medical conditions, and highlights the importance of rigorous external and clinical validation. Generalizability is key to ensuring that the models are robust and not overly fitted to the specific characteristics of the training data. Zhang et al. [96] addressed the crucial challenges of generalizability and external validation. They acknowledge that while their proposed model demonstrated excellent performance in internal validation on their collected dataset, there is a potential for bias when applying the model to cystoscopic images from other datasets, owing to differences in equipment.

Real world clinical validation focuses on the necessity of validating deep learning models in actual clinical environments. This goes beyond theoretical or simulation-based validation, emphasizing the need for models to prove their utility and accuracy in real-life medical settings. This validation is crucial to ensure that these models can aid in diagnosis and treatment planning, aligning with clinical workflows and patient-specific factors. Raza et al. [95] briefly touched on the applicability of their model in real-world clinical settings, noting its efficient online test time of only 8 s per subject.

Segmentation under complex medical conditions ensures accurate segmentation in scenarios in which patients have complex medical conditions. These conditions can introduce variability in medical images, such as overlapping symptoms or atypical disease presentations, making segmentation more challenging. Liu et al. [22] addressed the significant challenges in segmentation under complex medical conditions. They reported that while segmentation scores were higher in healthy or mildly affected conditions, the accuracy significantly decreased in cases where the lung field was obscured, either by medical equipment or severe deformation due to serious diseases. This finding highlights the complexities and limitations faced in accurately segmenting medical images in varied and challenging clinical scenarios.

8.1.3. Segmentation accuracy and boundary awareness

Segmentation accuracy and boundary awareness in medical image segmentation focus on the critical aspects of achieving high precision in identifying and delineating boundaries within medical images. The significance of effectively using boundary information in segmentation tasks is covered in boundary information exploitation. Precise segmentation depends on accurately identifying and utilizing boundary information, which guarantees that the model can discriminate between various anatomical structures and diseases. Luo et al. [16] addressed the challenge of boundary information exploitation by introducing C2BA-UNet and its multi-atlas boundary-aware (MABA) module. Using model-driven approaches, this module effectively learns and optimizes uncertain regions between tumors and tissues, thereby significantly enhancing tumor boundary segmentation to closely mimic real tumor boundaries.

The problem of initially recognizing boundaries that are not precisely defined is addressed using rough boundary identification. This is particularly difficult in medical images where there is subtle tissue contrast or when there are overlapping structures. Allah et al. [93] addressed the critical challenge of rough boundary identification in brain tumors. They highlighted how various factors, including intensity variation and noise in MRI systems, contribute to the complexity of accurately identifying and segmenting the rough boundaries of brain tumors, a problem that is central to enhancing the precision and effectiveness of brain tumor segmentation models.

Segmenting boundaries in complex regions delves into the difficulties of segmenting boundaries within anatomically complex regions. These regions often contain intricate structures or may exhibit alterations due to the disease, both of which significantly increase the challenge of accurate segmentation. Luo et al. [16] discussed the challenge of segmenting boundaries in complex regions, particularly in tumor segmentation. They pointed out the lack of focus on tumor boundary information in existing models and highlighted innovative approaches such as radiomics pipelines and edge enhancement supervision that have been developed to enhance the accuracy and certainty of segmenting boundaries between tumors and normal tissues.

The challenge of accurately segmenting boundaries and comprehending the spatial context of medical images is emphasized by the difficulty of capturing long-range relationships. This is particularly challenging with larger images or when important details are dispersed throughout the image. Tyagi et al. [88] addressed the significant challenge of capturing long-range relations in medical imaging. They highlighted the limitations of traditional CNNs in this regard.

8.1.4. Computational efficiency and architecture optimization

Computational efficiency and architecture optimization in the context of medical image segmentation with deep learning focuses on enhancing the performance of models while optimizing their computational demands. This section addresses the intricate balance between maintaining high accuracy in segmentation tasks and ensuring that models are computationally efficient and adaptable to various hardware constraints. Efficiency in contextual coordination focuses on optimizing the interaction between different components of a deep

learning model to improve its overall efficiency. This involves ensuring that various layers and processes within the model work in harmony, thereby enabling faster processing times while maintaining accuracy. This coordination is crucial in medical imaging, where models often need to efficiently analyze complex data. Luo et al. [16] discuss the challenges of efficiency in contextual coordination. They tackled this by developing at context coordination module (CCM). This module fuses multi-scale context information with axial attention to enhance the focus on tumor regions and streamline feature map analysis, demonstrating a significant advancement in efficiently managing contextual information in complex segmentation tasks.

The optimization of model parameters and computational time delves into strategies for refining the model parameters to strike a balance between computational load and segmentation accuracy. This optimization is vital for developing models that can run effectively on the available hardware, especially in time-sensitive clinical environments. Dar et al. [89] specifically addressed the optimization of model parameters and computational time. By incorporating a modified EfficientNet and an atrous convolution block into the U-Net model, they successfully reduced the number of training parameters and, enhance computational efficiency while maintaining the accuracy in tumor boundary localization.

The focus of deep neural network optimization strategies is the development and application of techniques to improve the operational efficiency of deep neural networks. This includes techniques such as pruning, quantization, and knowledge distillation, which can help reduce the size and complexity of the model without compromising its performance. Peng et al. [90] emphasized the significance of optimization strategies for deep neural networks. They demonstrated the effectiveness of deep supervision technology in enhancing training efficiency and the crucial role of pre-processing and architectural choices in achieving high segmentation performance with minimal parameters, thereby underscoring the importance of optimization in both the training process and network design.

A resource-efficient deep learning architecture addresses the need for creating deep learning models that are not only powerful but also resource-efficient. This involves designing architectures that require less computational resources, making them suitable for use in settings with limited hardware capabilities. Bhattacharyya et al. [25] addressed the crucial need for resource-efficient deep learning architectures in medical imaging. To handle complex nodular characteristics while improving segmentation effectiveness through innovative architectural features they proposed a DB-NET model.

8.1.5. Feature integration and preservation

In the field of medical image segmentation, feature integration and preservation deals with the difficulties and strategies of efficiently merging and preserving different tiers of features in deep learning models. It discusses the importance of integrating diverse feature levels to capture both detailed and comprehensive views of medical images, preserving essential low-level details for accuracy, addressing the complexities introduced by the partial volume effect, and ensuring stability in the learning process of the model. The integration of feature levels delves into how different levels of features, from basic edges and textures to more complex patterns, are integrated within deep learning models. This integration is crucial for effectively understanding and segmenting the medical images. The challenge lies in combining these varied feature levels in a manner that allows the model to leverage both local and global information, leading to more accurate and nuanced segmentation. Raza et al. [95] effectively addressed the challenge of integrating feature levels in deep learning architectures. They utilized the U-Net model's ability to combine local and global feature extraction methods, thereby enabling the classifier to simultaneously process low-level and high-level features for more accurate and comprehensive segmentation results. In another study, Alebiosu et al. [24] used a multi-scale block (MRB) to overcome the challenge of integrating feature levels. Together with the residual network, this MRB improves the feature size and contextual information, which helps improve segmentation accuracy by decreasing false positives and improving feature maps.

The preservation of low-level features focuses on the importance of retaining basic yet crucial details in the images, such as edges and textures. Often, as a deep learning model becomes more complex, there is the risk of losing these fine details. Ensuring that these low-level features are preserved is essential for maintaining the accuracy of segmentation, especially in medical imaging, where such details can be indicative of important clinical information. The problem of maintaining low-level features in deep learning architectures was skillfully handled by Raza et al. [95]. Compare to conventional 3D U-Net models, they greatly improve segmentation results by preserving both low level and high-level features simultaneously through the use of adaptive skip connections in the encoder's residual convolutional blocks.

The partial volume effect is a phenomenon commonly encountered in medical imaging, in which a single volume pixel may contain a mixture of tissues or substances, leading to ambiguous intensity values. This effect poses a significant challenge in accurately segmenting medical images, because it can obscure the true boundaries between different structures or pathologies. Allah et al. [93] addressed the complex challenge posed by the partial volume effect in MRI systems. This effect, characterized by the coexistence of multiple tissue types within a single pixel, leads to blurred tumor boundaries and vague tissue textures, thereby complicating the segmentation process and underscoring the need for sophisticated segmentation frameworks that can accurately define tumor boundaries in such challenging scenarios.

Preventing gradient instability addresses a technical challenge in training deep learning models. During the learning process, gradients (which guide the updating of the model parameters) can become unstable, leading to issues in the convergence of the model. Ensuring the stability of these gradients is crucial for achieving effective learning and accurate segmentation. Liu et al. [22] addressed the challenge of preventing gradient instability in deep learning architectures. They incorporated residual blocks to ensure stable gradient propagation across layers, thus solving the vanishing gradient problem and enhancing the network's training effectiveness, particularly in deeper configurations.

8.2. Future research directions in medical image segmentation

The field of diagnostic medicine is about to change owing to the potential of future research directions in deep learning for medical image segmentation. Key areas of focus include developing algorithms to handle vast variations in medical data, enhancing model generalizability across different patient populations, and integrating multimodal imaging information. As the demand for personalized medicine grows, deep learning models must evolve to offer precise, patient-specific insights. Moreover, addressing data privacy, improving the efficiency of algorithms to allow real-time segmentation, and pursuing the incorporation of AI into clinical workflows to augment rather than disrupt them, are all pivotal research directions. By addressing these challenges, the next wave of deep-learning innovations will likely produce more accurate, efficient, and trustworthy tools for medical image analysis, paving the way for groundbreaking advancements in patient care and treatment outcomes. Fig. 19 visualizes some of the key future research opportunities in medical image segmentation.

8.2.1. Explainable Artificial Intelligence (XAI)

XAI is a set of processes and methods that help make the actions of AI systems transparent and understandable to humans. It focuses on shedding light on the decision-making mechanisms of AI, particularly in complex systems, such as deep learning models. In the realm of deep learning for medical image segmentation, a critical frontier for

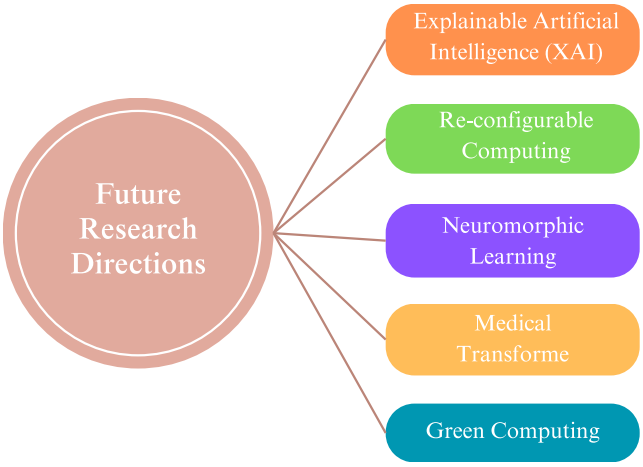


Fig. 19. Future research directions of DL in medical image segmentation.

future research is the XAI. As these advanced models penetrate healthcare and provide medical professionals with decision-making tools, the imperative for transparency becomes crucial. XAI aims to provide a clear understanding of how and why specific segmentation outcomes are achieved by simplifying the inner workings of deep neural networks. This transparency is not merely a technical necessity but also a fundamental component for ethical applications, fostering trust among medical professionals and patients. By integrating XAI principles, future research aims to refine AI systems that are not only precise and efficient but also provide intelligible insights into their processes, thus ensuring that their decisions are accountable and aligned with clinical rationale. Advancements in XAI are expected to bridge the gap between AI's capabilities and human interpretability, ensuring that the segmentation of medical images is both scientifically robust and reliable. XAI is illustrated in Fig. 20 and is widely recognized as essential for the practical implementation of AI models [103].

8.2.2. Re-configurable computing

Incorporating advanced digital image processing algorithms and handling massive amounts of data from sources, such as medical instruments has become crucial [104]. Fig. 21 illustrates that in medical image segmentation, where extensive data-sets are standard, reconfigurable computing platforms such as Field-Programmable Gate Arrays(FPGAs) are essential. These platforms can be tailored to efficiently perform specific tasks within models, thereby significantly accelerating the segmentation process. FPGA enable the quick development of complex algorithms, lower time-to market costs, and streamline the verification and debugging procedures for medical image processing algorithms. Algorithms for medical image segmentation vary greatly depending on imaging modalities, such as CT, MRI, X-ray, and the specific tissue or anomaly being examined. Re-configurable computing enables customization of the hardware to match different segmentation algorithm requirements and ensure optimal performance [105]. Reconfigurable computing is thus indispensable in medical image segmentation with DL, offering efficient, real-time, energy-saving, customizable, adaptable, and parallel processing solutions essential for accurate and timely medical image analysis and diagnosis.

8.2.3. Neuromorphic learning

Neuromorphic learning, which involves neural models in both hardware and software, has the potential to revolutionize medicine [107]. Fig. 22 illustrates that these systems exhibit brain-like behavior, characterized by low power consumption, minimal latency, compact size, and high computing capacity. Inspired by the human brain, they mimic its structure and learning mechanisms, allowing for more efficient processing of complicated medical images and capturing of subtle details that are challenging for traditional algorithms. In addition, they process data in parallel, which is similar to the human brain. However, various challenges, such as the complexity of the models, limited understanding, and lack of standardization, hinder their use in medical image segmentation. To overcome these obstacles, collaboration among neuroscience, computer science, and medical imaging experts is essential. Continuous research is required to enhance the efficiency, interpretation, and scalability of neuromorphic models for various medical image segmentation tasks.

8.2.4. Medical transformer

In recent years, U-shaped structures and skip connections in deep neural networks have gained traction for medical imaging tasks. Despite CNN's excellent performance of CNNs, their ability to capture global and long-range semantic information remains limited. Transformer-based architectures, utilizing self-awareness modules, have become popular, particularly in natural language processing. Dosovitskiy et al. [109] introduced at Vision Transformer (ViT), as shown in Fig. 23, for image classification, and researchers applied transformers to medical image segmentation. Although CNNs excel at extracting basic features, transformers efficiently handle higher-level visual semantic relationships [110]. In ViT, an image from the training data is divided into smaller, consistent pieces known as patches as part of the procedure. Linear embedding was used to transform the visual data for each patch into a structured numerical representation. Positional embeddings were also included to help the model understand the relative placement of these patches and preserve the location data. However, replacing all convolution operators with transformers in computer vision tasks poses challenges [111].

8.2.5. Green computing

Green computing involves adopting practices and techniques that use computing assets in an environmentally friendly manner, while maintaining optimal computing performance. This represents a balanced and sustainable approach to creating a healthier, greener, and safer environment without sacrificing the needs of present and future generations in terms of technology [112]. In the context of medical image segmentation, various critical tasks, such as data pre-processing, normalization, and resizing, require significant computing resources. Green computing manages this resource-intensive process. By advocating energy-efficient algorithms, optimizing hardware consumption, and promoting the use of renewable energy sources, green computing not only improves the environmental sustainability of these operations but also significantly alleviates the computing resource challenges that arise during complex phases such as data pre-processing in medical image segmentation. The benefits of green computing in this context include increasing equipment power density, managing increasing cooling demands, controlling escalating energy costs, overcoming energy supply and access limitations, and improving low server utilization rates [112]. Fig. 24 illustrates advocacy for green computing, emphasizing the utilization of recycled materials and energy-efficient hardware and software solutions. Given the complicated calculations involved in medical image segmentation, a high computing power is essential. Using energy-efficient processors and optimizing algorithms can decrease energy consumption and thus reduce the carbon footprint. In addition, green computing encourages the implementation of virtualization technologies, allowing multiple virtual machines to operate on a single physical server. This consolidation of workloads requires fewer physical servers, leading to lower energy consumption and a reduced physical footprint in data centers.

9. Conclusion

DL-based medical image segmentation utilizes DL algorithms to accurately delineate and identify structures or regions within medical

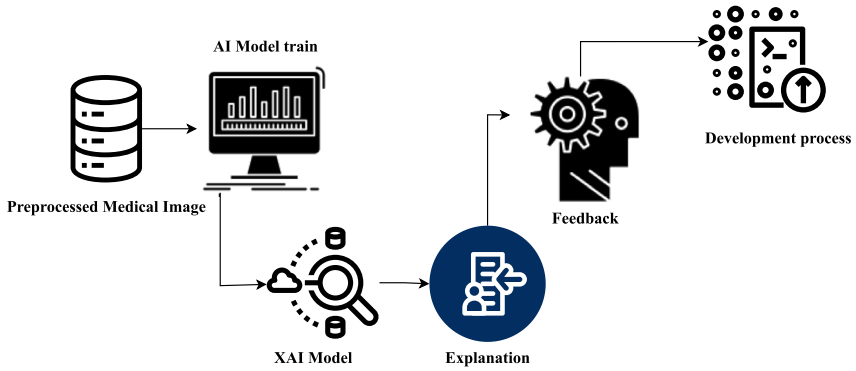


Fig. 20. XAI working process.

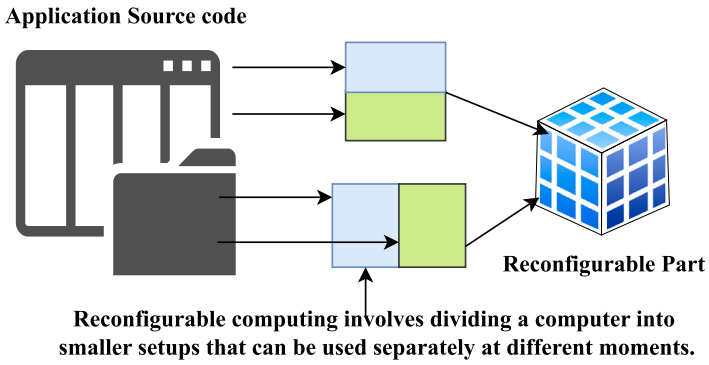


Fig. 21. Re-configurable computing [106].

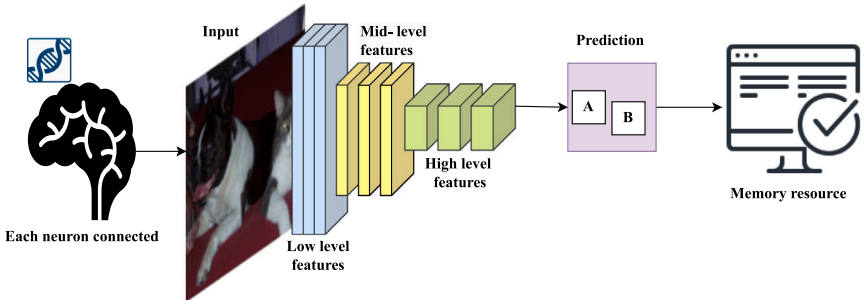


Fig. 22. Neuromorphic learning [108].

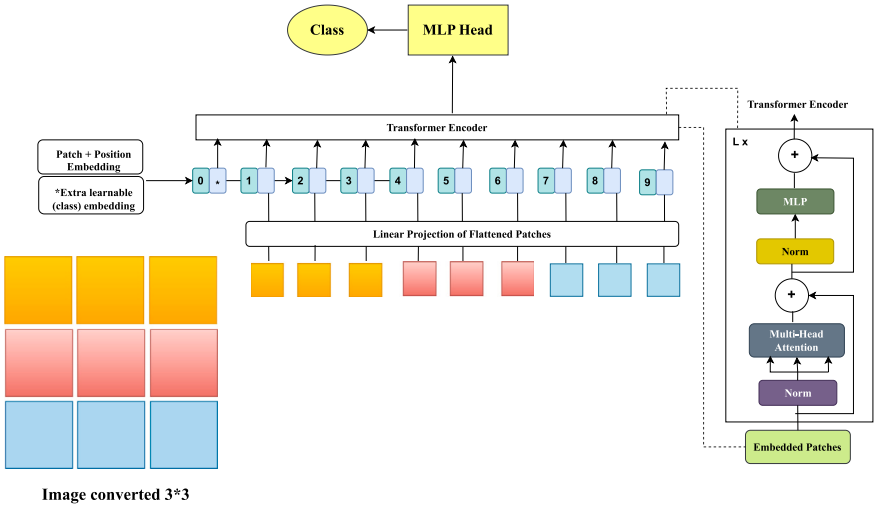


Fig. 23. Vision transformer (ViT).

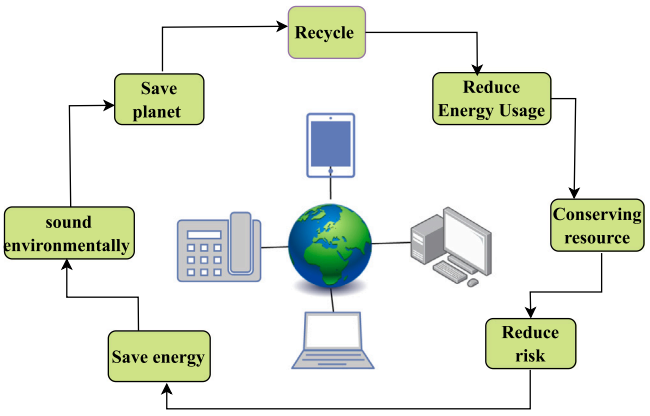


Fig. 24. An eco-friendly green computing plan.

images, aiding in diagnosis and analysis. It significantly enhances the precision of medical diagnosis, treatment planning, and disease monitoring by providing detailed and automated anatomical or pathological region identification in medical images. However, recent investigations in this expansive research domain are lacking. Hence, this study endeavored to conduct a systematic review that offer comprehensive insights into DL-based medical image segmentation and its recent progress. It delved into commonly employed pre-processing techniques and experimental datasets, followed by an examination of prevalent DL algorithms, highlighting their strengths and limitations. Furthermore, it provides an extensive exploration and results analysis of recent state-of-the-art research papers. This paper also addressed challenges in DL-based medical image segmentation and outlines future research directions to overcome these obstacles. This comprehensive review not only serves as a valuable interdisciplinary resource, but also offers insights essential for advancing the DL-based medical image segmentation field, benefiting investors and researchers by fostering innovation and informed decision-making.

CRediT authorship contribution statement

Md. Eshmam Rayed: Writing – original draft. **S.M. Sajibul Islam:** Formal analysis, Conceptualization. **Sadia Islam Niha:** Methodology, Investigation. **Jamin Rahman Jim:** Visualization, Validation. **Md Mohsin Kabir:** Validation, Methodology. **M.F. Mridha:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declared that they have no competing interests in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgment

The authors would like to thank the Advanced Machine Intelligence Research Lab - AMIR Lab for Supervision and Resources.

References

[1] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. IEEE Trans Pattern Anal Mach Intell 2022;44(7):3523–42. http://dx.doi.org/10.1109/TPAMI.2021.3059968.
[2] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
[3] Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. Sustainability 2021;13(3):1224. http://dx.doi.org/10.3390/su13031224.
[4] Ghosh S, Das N, Das I, Maulik U. Understanding deep learning techniques for image segmentation. ACM Comput Surv (CSUR) 2019;52(4). http://dx.doi.org/10.1145/3329784.
[5] Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging 2019;32:582–96. http://dx.doi.org/10.1007/s10278-019-00227-x.
[6] Sultana F, Sufian A, Dutta P. Evolution of image segmentation using deep convolutional neural network: A survey. Knowl-Based Syst 2020;201–202:106062. http://dx.doi.org/10.1016/j.knosys.2020.106062.
[7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–mICCAI 2015: 18th international conference, munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer; 2015, p. 234–41. http://dx.doi.org/10.1007/978-3-319-24574-4_28.
[8] Wu S, Wang Z, Liu C, Zhu C, Wu S, Xiao K. Automatical segmentation of pelvic organs after hysterectomy by using dilated convolution U-Net++. In: 2019 IEEE 19th international conference on software quality, reliability and security companion. QRS-C, 2019, p. 362–7. http://dx.doi.org/10.1109/QRS-C.2019.00074.
[9] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 3431–40.
[10] Aljabri M, AlGhamdi M. A review on the use of deep learning for medical images segmentation. Neurocomputing 2022;506:311–35. http://dx.doi.org/10.1016/j.neucom.2022.07.070.
[11] Qureshi I, Yan J, Abbas Q, Shaheed K, Riaz AB, Wahid A, Khan MWJ, Szczuko P. Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends. Inf Fusion 2023;90:316–52. http://dx.doi.org/10.1016/j.inffus.2022.09.031.
[12] Yu Y, Wang C, Fu Q, Kou R, Huang F, Yang B, Yang T, Gao M. Techniques and challenges of image segmentation: A review. Electronics 2023;12(5):1199. http://dx.doi.org/10.3390/electronics12051199.
[13] Keele S, et al. Guidelines for performing systematic literature reviews in software engineering. 2007.
[14] Kitchenham B. Procedures for performing systematic reviews. vol. 33, Keele, UK, Keele University; 2004, p. 1–26, (2004).
[15] Shareef B, Vakanski A, Freer PE, Xian M. ESTAN: Enhanced small tumor-aware network for breast ultrasound image segmentation. Healthcare 2022;10(11):2262. http://dx.doi.org/10.3390/healthcare10112262.
[16] Luo S, Jiang H, Wang M. C2BA-UNet: A context-coordination multi-atlas boundary-aware UNet-like method for PET/CT images based tumor segmentation. Comput Med Imaging Graph 2023;103:102159. http://dx.doi.org/10.1016/j.compmedimag.2022.102159.
[17] Zeng L, Tang H, Wang W, Xie M, Ai Z, Chen L, Wu Y. MAMC-Net: an effective deep learning framework for whole-slide image tumor segmentation. Multimedia Tools Appl 2023;82:39349–69. http://dx.doi.org/10.1007/s11042-023-15065-x.
[18] Youssef BE, Alksas A, Shalaby A, Mahmoud AH, Van Bogaert E, Alghamdi NS, Neubacher A, Contractor S, Ghazal M, Elmaghraby AS, El-Baz A. Integrated deep learning and stochastic models for accurate segmentation of lung nodules from computed tomography images: A novel framework. IEEE Access 2023;11:99807–21. http://dx.doi.org/10.1109/ACCESS.2023.3313174.
[19] Said Y, Alsheikhy AA, Shawly T, Lahza H. Medical images segmentation for lung cancer diagnosis based on deep learning architectures. Diagnostics 2023;13(3):546. http://dx.doi.org/10.3390/diagnostics13030546.
[20] Rahman MF, Zhuang Y, Tseng T-LB, Pokojovy M, McCaffrey P, Walser E, Moen S, Vo A. Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach. J Vis Commun Image Represent 2022;85:103521. http://dx.doi.org/10.1016/j.jvcir.2022.103521.
[21] Arvind S, Tembhurne JV, Diwan T, Sahare P. Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays. Res Eng 2023;17:100929. http://dx.doi.org/10.1016/j.rineng.2023.100929.
[22] Liu W, Luo J, Yang Y, Wang W, Deng J, Yu L. Automatic lung segmentation in chest X-ray images using improved U-Net. Sci Rep 2022;12(1):8649. http://dx.doi.org/10.1038/s41598-022-12743-y.
[23] Messaoudi H, Belaid A, Ben Salem D, Conze P-H. Cross-dimensional transfer learning in medical image segmentation with deep learning. Med Image Anal 2023;88:102868. http://dx.doi.org/10.1016/j.media.2023.102868.
[24] Alebiosu DO, Dharmaratne A, Lim CH. Improving tuberculosis severity assessment in computed tomography images using novel DAvoU-Net segmentation and deep learning framework. Expert Syst Appl 2023;213:119287. http://dx.doi.org/10.1016/j.eswa.2022.119287.
[25] Bhattacharyya D, Thirupathi Rao N, Joshua ESN, Hu Y-C. A bi-directional deep learning architecture for lung nodule semantic segmentation. Vis Comput 2023;39(11):5245–61. http://dx.doi.org/10.1007/s00371-022-02657-1.
[26] Mohakud R, Dash R. Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN. J King Saud Univ - Comput Inf Sci 2022;34(10, Part B):9889–904. http://dx.doi.org/10.1016/j.jksuci.2021.12.018.
[27] Badshah N, Ahmad A. ResBCU-Net: Deep learning approach for segmentation of skin images. Biomed Signal Process Control 2022;71:103137. http://dx.doi.org/10.1016/j.bspc.2021.103137.

[28] Han Q, Wang H, Hou M, Weng T, Pei Y, Li Z, Chen G, Tian Y, Qiu Z. HWA-SegNet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment. Comput Biol Med 2023;152:106343. http://dx.doi.org/10.1016/j.compbiomed.2022.106343.

[29] Ruan J, Xie M, Gao J, Liu T, Fu Y. EGE-UNet: An efficient group enhanced UNet for skin lesion segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2023. vol. 14223, Cham: Springer Nature Switzerland; 2023, p. 481–90. http://dx.doi.org/10.1007/978-3-031-43901-8_46.

[30] Kumar MD, Sivanarayana G, Indira D, Raj MP. Skin cancer segmentation with the aid of multi-class dilated D-net (MD2N) framework. Multimedia Tools Appl 2023;82:35995–6018. http://dx.doi.org/10.1007/s11042-023-14605-9.

[31] Bindhu A, Thanammal K. Segmentation of skin cancer using Fuzzy U-network via deep learning. Meas: Sensors 2023;26:100677. http://dx.doi.org/10.1016/j.measen.2023.100677.

[32] Lama N, Hagerty J, Nambisan A, Stanley RJ, Van Stoecker W. Skin lesion segmentation in dermoscopic images with noisy data. J Digit Imaging 2023;36:1712–22. http://dx.doi.org/10.1007/s10278-023-00819-8.

[33] Huang L, Zhao Y-G, Yang T-J. Skin lesion image segmentation by using backchannel filling CNN and level sets. Biomed Signal Process Control 2024;87:105417. http://dx.doi.org/10.1016/j.bspc.2023.105417.

[34] Mostafa AM, Zakariah M, Aldakheel EA. Brain tumor segmentation using deep learning on MRI images. 2023;13(9):1562. http://dx.doi.org/10.3390/diagnostics13091562.

[35] Çetiner H, Metlek S. DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation. J King Saud Univ - Comput Inf Sci 2023;35(8):101663. http://dx.doi.org/10.1016/j.jksuci.2023.101663.

[36] Xu M, Huang K, Qi X. A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification. IEEE Access 2023;11:5377–92. http://dx.doi.org/10.1109/ACCESS.2023.3236693.

[37] Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, Szeskin A, Jacobs C, Mamani GEH, Chartrand G, Lohöfer F, Holch JW, Sommer W, Hofmann F, Hostettler A, Lev-Cohain N, Drozdzal M, Amitai MM, Vivanti R, Sosna J, Ezhov I, Sekuboyina A, Navarro F, Kofler F, Paetzold JC, Shit S, Hu X, Lipková J, Rempfler M, Piraud M, Kirschke J, Wiestler B, Zhang Z, Hülsemeyer C, Beetz M, Ettlinger F, Antonelli M, Bae W, Bellver M, Bi L, Chen H, Chlebus G, Dam EB, Dou Q, Fu C-W, Georgescu B, i Nieto XG, Gruen F, Han X, Heng P-A, Hesser J, Moltz JH, Igel C, Isensee F, Jäger P, Jia F, Kaluva KC, Khened M, Kim I, Kim J-H, Kim S, Kohl S, Konopczynski T, Kori A, Krishnamurthi G, Li F, Li H, Li J, Li X, Lowengrub J, Ma J, Maier-Hein K, Maninis K-K, Meine H, Merhof D, Pai A, Perslev M, Petersen J, Pont-Tuset J, Qi J, Qi X, Rippel O, Roth K, Sarasua I, Schenk A, Shen Z, Torres J, Wachinger C, Wang C, Weninger L, Wu J, Xu D, Yang X, Yu SC-H, Yuan Y, Yue M, Zhang L, Cardoso J, Bakas S, Braren R, Heinemann V, Pal C, Tang A, Kadoury S, Soler L, van Ginneken B, Greenspan H, Joskowicz L, Menze B. The liver tumor segmentation benchmark (LiTS). Med Image Anal 2023;84:102680. http://dx.doi.org/10.1016/j.media.2022.102680.

[38] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T. Recent advances in convolutional neural networks. Pattern Recognit 2018;77:354–77. http://dx.doi.org/10.1016/j.patcog.2017.10.013.

[39] Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. J Physiol 1968;195(1):215–43. http://dx.doi.org/10.1113/jphysiol.1968.sp008455.

[40] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36(4):193–202. http://dx.doi.org/10.1007/BF00344251.

[41] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324. http://dx.doi.org/10.1109/5.726791.

[42] Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ. Phoneme recognition using time-delay neural networks. In: Backpropagation. Psychology Press; 2013, p. 35–61.

[43] Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. 2017, http://dx.doi.org/10.48550/arXiv.1701.03056, arXiv preprint arXiv:1701.03056.

[44] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 2016;35(5):1240–51. http://dx.doi.org/10.1109/TMI.2016.2538465.

[45] Niu X-X, Suen CY. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recognit 2012;45(4):1318–25. http://dx.doi.org/10.1016/j.patcog.2011.09.021.

[46] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. Imagenet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–52. http://dx.doi.org/10.1007/s11263-015-0816-y.

[47] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90. http://dx.doi.org/10.1145/3065386.

[48] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, http://dx.doi.org/10.48550/arXiv.1409.1556, arXiv preprint arXiv:1409.1556.

[49] Szegedy C, Liu W, Jia Y, Sermanet P, ReNG1ed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 1–9.

[50] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.

[51] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017, http://dx.doi.org/10.48550/arXiv.1704.04861, arXiv preprint arXiv:1704.04861.

[52] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 4700–8.

[53] Zhou X, Takayama R, Wang S, Hara T, Fujita H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys 2017;44(10):5221–33. http://dx.doi.org/10.1002/mp.12480.

[54] Sun J, Peng Y, Guo Y, Li D. Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. Neurocomputing 2021;423:34–45. http://dx.doi.org/10.1016/j.neucom.2020.10.031.

[55] Kumar S, Negi A, Singh J, Verma H. A deep learning for brain tumor MRI images semantic segmentation using FCN. In: 2018 4th international conference on computing communication and automation. ICCCA, 2018, p. 1–4. http://dx.doi.org/10.1109/CCAA.2018.8777675.

[56] Liu W, Rabinovich A, Berg AC. Parsenet: Looking wider to see better. 2015, http://dx.doi.org/10.48550/arXiv.1506.04579, arXiv preprint arXiv:1506.04579.

[57] Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. vol. 10670, Cham: Springer International Publishing; 2018, p. 178–90. http://dx.doi.org/10.1007/978-3-319-75238-9_16.

[58] Li Y, Qi H, Dai J, Ji X, Wei Y. Fully convolutional instance-aware semantic segmentation. In: 2017 IEEE conference on computer vision and pattern recognition. CVPR, 2017, p. 4438–46. http://dx.doi.org/10.1109/CVPR.2017.472.

[59] Yuan Y, Chao M, Lo Y-C. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE Trans Med Imaging 2017;36(9):1876–86. http://dx.doi.org/10.1109/TMI.2017.2695227.

[60] Liu N, Li H, Zhang M, Liu J, Sun Z, Tan T. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In: 2016 international conference on biometrics. ICB, 2016, p. 1–8. http://dx.doi.org/10.1109/ICB.2016.7550055.

[61] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39(12):2481–95. http://dx.doi.org/10.1109/TPAMI.2016.2644615.

[62] Ahmad P, Jin H, Alroobaea R, Qamar S, Zheng R, Alnajjar F, Aboudi F. MH UNet: A multi-scale hierarchical based architecture for medical image segmentation. IEEE Access 2021;9:148384–408. http://dx.doi.org/10.1109/ACCESS.2021.3122543.

[63] Deng Y, Hou Y, Yan J, Zeng D. ELU-Net: An efficient and lightweight U-Net for medical image segmentation. IEEE Access 2022;10:35932–41. http://dx.doi.org/10.1109/ACCESS.2022.3163711.

[64] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A nested U-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. vol. 11045, Cham: Springer International Publishing; 2018, p. 3–11. http://dx.doi.org/10.1007/978-3-030-00889-5_1.

[65] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 2020;39(6):1856–67. http://dx.doi.org/10.1109/TMI.2019.2959609.

[66] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40(4):834–48. http://dx.doi.org/10.1109/TPAMI.2017.2699184.

[67] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 2015, http://dx.doi.org/10.48550/arXiv.1511.07122, arXiv preprint arXiv:1511.07122.

[68] Lou A, Loew M. CFPNET: Channel-wise feature pyramid for real-time semantic segmentation. In: 2021 IEEE international conference on image processing. ICIP, 2021, p. 1894–8. http://dx.doi.org/10.1109/ICIP42928.2021.9506485.

[69] Yang M, Yu K, Zhang C, Li Z, Yang K. DenseASPP for semantic segmentation in street scenes. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. 2018, p. 3684–92. http://dx.doi.org/10.1109/CVPR.2018.00388.

[70] Paszke A, Chaurasia A, Kim S, Culurciello E. Enet: A deep neural network architecture for real-time semantic segmentation. 2016, http://dx.doi.org/10.48550/arXiv.1606.02147, arXiv preprint arXiv:1606.02147.

[71] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 2017;40(4):834–48. http://dx.doi.org/10.1109/TPAMI.2017.2699184.

[72] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. 2017, http://dx.doi.org/10.48550/arXiv.1706.05587, arXiv preprint arXiv:1706.05587.

[73] Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Computer vision – ECCV 2018. vol. 11211, Cham: Springer International Publishing; 2018, p. 833–51. http://dx.doi.org/10.1007/978-3-030-01234-2_49.

[74] Wang J, Liu X. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. Comput Methods Programs Biomed 2021;207:106210. http://dx.doi.org/10.1016/j.cmpb.2021.106210.

[75] Sun Y, Shi C. Liver tumor segmentation and subsequent risk prediction based on Deeplabv3+. IOP Conf Ser: Mater Sci Eng 2019;612(2):022051. http://dx.doi.org/10.1088/1757-899X/612/2/022051.

[76] Milletari F, Navab N, Ahmadi S-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision. 3DV, 2016, p. 565–71. http://dx.doi.org/10.1109/3DV.2016.79.

[77] Türk F, Lüy M, Barışçı N. Kidney and renal tumor segmentation using a hybrid V-Net-based model. Mathematics 2020;8(10):1772. http://dx.doi.org/10.3390/math8101772.

[78] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: 2017 IEEE conference on computer vision and pattern recognition. CVPR, 2017, p. 6230–9. http://dx.doi.org/10.1109/CVPR.2017.660.

[79] Ye L-Y, Miao X-Y, Cai W-S, Xu W-J. Medical image diagnosis of prostate tumor based on PSP-Net+VGG16 deep learning network. Comput Methods Programs Biomed 2022;221:106770. http://dx.doi.org/10.1016/j.cmpb.2022.106770.

[80] Yan L, Liu D, Xiang Q, Luo Y, Wang T, Wu D, Chen H, Zhang Y, Li Q. PSP net-based automatic segmentation network model for prostate magnetic resonance imaging. Comput Methods Programs Biomed 2021;207:106211. http://dx.doi.org/10.1016/j.cmpb.2021.106211.

[81] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE conference on computer vision and pattern recognition. 2014, p. 580–7. http://dx.doi.org/10.1109/CVPR.2014.81.

[82] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. IEEE Trans Pattern Anal Mach Intell 2020;42(2):386–97. http://dx.doi.org/10.1109/TPAMI.2018.2844175.

[83] Girshick R. Fast R-CNN. In: 2015 IEEE international conference on computer vision. ICCV, 2015, p. 1440–8. http://dx.doi.org/10.1109/ICCV.2015.169.

[84] Shu J-H, Nian F-D, Yu M-H, Li X. An improved mask R-CNN model for multiorgan segmentation. Math Probl Eng 2020;2020:1–11. http://dx.doi.org/10.1155/2020/8351725.

[85] Jeong J, Lei Y, Kahn S, Liu T, Curran WJ, Shu H-K, Mao H, Yang X. Brain tumor segmentation using 3D Mask R-CNN for dynamic susceptibility contrast enhanced perfusion imaging. Phys Med Biol 2020;65(18):185009. http://dx.doi.org/10.1088/1361-6560/aba6d4.

[86] Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. 2018, p. 8759–68. http://dx.doi.org/10.1109/CVPR.2018.00913.

[87] Zhao X, Zhang P, Song F, Ma C, Fan G, Sun Y, Feng Y, Zhang G. Prior attention network for multi-lesion segmentation in medical images. IEEE Trans Med Imaging 2022;41(12):3812–23. http://dx.doi.org/10.1109/TMI.2022.3197180.

[88] Tyagi S, Kushnure DT, Talbar SN. An amalgamation of vision transformer with convolutional neural network for automatic lung tumor segmentation. Comput Med Imaging Graph 2023;108:102258. http://dx.doi.org/10.1016/j.compmedimag.2023.102258.

[89] Dar MF, Ganivada A. Efficientu-net: a novel deep learning method for breast tumor segmentation and classification in ultrasound images. Neural Process Lett 2023;55:10439–62. http://dx.doi.org/10.1007/s11063-023-11333-x.

[90] Peng Y, Sun J. The multimodal MRI brain tumor segmentation based on AD-Net. Biomed Signal Process Control 2023;80:104336. http://dx.doi.org/10.1016/j.bspc.2022.104336.

[91] Farooq MU, Ullah Z, Gwak J. Residual attention based uncertainty-guided mean teacher model for semi-supervised breast masses segmentation in 2D ultrasonography. Comput Med Imaging Graph 2023;104:102173. http://dx.doi.org/10.1016/j.compmedimag.2022.102173.

[92] Ruba T, Tamilselvi R, Parisa Beham M. Brain tumor segmentation using JGate-AttResUNet – A novel deep learning approach. Biomed Signal Process Control 2023;84:104926. http://dx.doi.org/10.1016/j.bspc.2023.104926.

[93] M. Gab Allah A, M. Sarhan A, M. Elshennawy N. Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information. Expert Syst Appl 2023;213:118833. http://dx.doi.org/10.1016/j.eswa.2022.118833.

[94] Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS One 2015;10(10):e0140381. http://dx.doi.org/10.1371/journal.pone.0140381.

[95] Raza R, Ijaz Bajwa U, Mehmood Y, Waqas Anwar M, Hassan Jamal M. dResU-Net: 3D deep residual U-net based brain tumor segmentation from multimodal MRI. Biomed Signal Process Control 2023;79:103861. http://dx.doi.org/10.1016/j.bspc.2022.103861.

[96] Zhang Q, Liang Y, Zhang Y, Tao Z, Li R, Bi H. A comparative study of attention mechanism based deep learning methods for bladder tumor segmentation. Int J Med Inform 2023;171:104984. http://dx.doi.org/10.1016/j.ijmedinf.2023.104984.

[97] Raj R, Londhe ND, Sonawane R. PsLSNetV2: End to end deep learning system for measurement of area score of psoriasis regions in color images. Biomed Signal Process Control 2023;79:104138. http://dx.doi.org/10.1016/j.bspc.2022.104138.

[98] Ilhan A, Alpan K, Sekeroglu B, Abiyev R. COVID-19 Lung CT image segmentation using localization and enhancement methods with U-Net. Procedia Comput Sci 2023;218:1660–7. http://dx.doi.org/10.1016/j.procs.2023.01.144, International Conference on Machine Learning and Data Engineering.

[99] Chen Y, Feng L, Zheng C, Zhou T, Liu L, Liu P, Chen Y. LDANet: Automatic lung parenchyma segmentation from CT images. Comput Biol Med 2023;155:106659. http://dx.doi.org/10.1016/j.compbiomed.2023.106659.

[100] Park J, Kang SK, Hwang D, Choi H, Ha S, Seo JM, Eo JS, Lee JS. Automatic lung cancer segmentation in [18f] FDG pet/CT using a two-stage deep learning approach. Nucl Med Mol Imaging 2023;57(2):86–93. http://dx.doi.org/10.1007/s13139-022-00745-7.

[101] Alshmrani GM, Ni Q, Jiang R, Muhammed N. Hyper_Dense_Lung_Seg: Multimodal-fusion-based modified U-net for lung tumour segmentation using multimodality of CT-PET scans. Diagnostics 2023;13(22):3481. http://dx.doi.org/10.3390/diagnostics13223481.

[102] Ullah I, Ali F, Shah B, El-Sappagh S, Abuhmed T, Park SH. A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest X-ray images. Sci Rep 2023;13(1):791. http://dx.doi.org/10.1038/s41598-023-27815-w.

[103] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115. http://dx.doi.org/10.1016/j.inffus.2019.12.012.

[104] Chiuchisan I. Implementation of medical image processing algorithm on reconfigurable hardware. In: 2013 e-health and bioengineering conference. EHB, 2013, p. 1–4. http://dx.doi.org/10.1109/EHB.2013.6707298.

[105] Chiuchişan I, Cerlincă M. Implementation of real-time system for medical image processing using verilog hardware description language. In: Proceedings of the 9th international conference on cellular and molecular biology, biophysics and bioengineering. BIO'13, vol. 5125, 2013, p. 66–9.

[106] Compton K, Hauck S. Reconfigurable computing: a survey of systems and software. ACM Comput Surv (csuR) 2002;34(2):171–210. http://dx.doi.org/10.1145/508352.508353.

[107] Aboumerhi K, Güemes A, Liu H, Tenore F, Etienne-Cummings R. Neuromorphic applications in medicine. J Neural Eng 2023;20(4):041004. http://dx.doi.org/10.1088/1741-2552/aceca3.

[108] Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. Nature 2019;575(7784):607–17. http://dx.doi.org/10.1038/s41586-019-1677-2.

[109] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020, http://dx.doi.org/10.48550/arXiv.2010.11929, arXiv preprint arXiv:2010.11929.

[110] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation. 2021, http://dx.doi.org/10.48550/arXiv.2102.04306, arXiv preprint arXiv:2102.04306.

[111] Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: A survey. IET Image Process 2022;16(5):1243–67. http://dx.doi.org/10.1049/ipr2.12419.

[112] Saha B. Green computing. Int J Comput Trends Technol (IJCTT) 2014;14(2):46–50. http://dx.doi.org/10.14445/22312803/IJCTT-V14P112.