Computers in Biology and Medicine 198 (2025) 111171

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Advances in medical image segmentation: A comprehensive survey with a focus on lumbar spine applications

Ahmed Kabil [a], Ghada Khoriba [a], Mina Yousef [a], Essam A. Rashed [b,c,*]

[a] Center for Informatics Science, School of Information Technology and Computer Science, Nile University, 12588, Giza, Egypt
[b] Graduate School of Information Science, University of Hyogo, Kobe, 650-0047, Japan
[c] Advanced Medical Engineering Research Institute, University of Hyogo, Himeji, 670-0836, Japan

ARTICLE INFO

Keywords:
Medical image segmentation
Semantic segmentation
Deep learning
Lumbar spine segmentation
Active learning
Transformer networks
Federated learning

ABSTRACT

Medical Image Segmentation (MIS) stands as a cornerstone in medical image analysis, playing a pivotal role in precise diagnostics, treatment planning, and monitoring of various medical conditions. This paper presents a comprehensive and systematic survey of MIS methodologies, bridging the gap between traditional image processing techniques and modern deep learning approaches. The survey encompasses thresholding, edge detection, region-based segmentation, clustering algorithms, and model-based techniques while also delving into state-of-the-art deep learning architectures such as Convolutional Neural Networks (CNNs), Fully Convolutional Networks (FCNs), and the widely adopted U-Net and its variants. Moreover, integrating attention mechanisms, semi-supervised learning, generative adversarial networks (GANs), and Transformer-based models is thoroughly explored.

In addition to covering established methods, this survey highlights emerging trends, including hybrid architectures, cross-modality learning, federated and distributed learning frameworks, and active learning strategies, which aim to address challenges such as limited labeled datasets, computational complexity, and model generalizability across diverse imaging modalities. Furthermore, a specialized case study on lumbar spine segmentation is presented, offering insights into the challenges and advancements in this relatively underexplored anatomical region.

Despite significant progress in the field, critical challenges persist, including dataset bias, domain adaptation, interpretability of deep learning models, and integration into real-world clinical workflows. This survey serves as both a tutorial and a reference guide, particularly for early-career researchers, by providing a holistic understanding of the landscape of MIS and identifying promising directions for future research. Through this work, we aim to contribute to the development of more robust, efficient, and clinically applicable medical image segmentation systems.

1. Introduction

Year after year, the volume of medical data grows exponentially as patients seek clinicians' consulting services for diagnostic and therapeutic services. Various imaging modalities have historically been used for noninvasive medical imaging to provide medical doctors with information on the patient's state; these include computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound (US), optical coherence tomography (OCT), and X-rays [1]. The choice of imaging modality depends on the organ and pathology investigated [2].

In addition to being a laborious and time-consuming process, human interpretation of medical images is vulnerable to subjectivity due to different levels of experience, as well as the possible mental and physical fatigue of the clinical expert [2], which could potentially lead to variations in the interpretation of the same medical image by two different experts. Intra and interexpert variability is further exacerbated by irregularities in the targeted structures, morphological variations, and pathological deformities between patients [3]. This problematic irregularity in human perception led to the development of computer-aided tools that facilitate image analysis, producing precise, fast, repeatable, and objective measurements. Image analysis can be further subcategorized into image enhancement, registration, classification, segmentation, detection, localization [4], and visualization [5].

* Corresponding author at: Graduate School of Information Science, University of Hyogo, Kobe, 650-0047, Japan.
E-mail addresses: a.mohamad2116@nu.edu.eg (A. Kabil), ghadakhoriba@nu.edu.eg (G. Khoriba), myousef@nu.edu.eg (M. Yousef), rashed@gsis.u-hyogo.ac.jp (E.A. Rashed).

https://doi.org/10.1016/j.compbiomed.2025.111171
Received 5 January 2025; Received in revised form 16 July 2025; Accepted 30 September 2025

Medical Image Segmentation (MIS) is crucial for separating anatomical and pathological structures by identifying their boundaries [6]. This is done by dividing the image into homogeneous (or correlated) regions that share similar characteristics. The goal is to separate anatomically distinct organs, tissues, and lesions. This process converts the image into a more meaningful representation that can be easily analyzed [2,7].

Given the wide range of imaging modalities, each presenting unique challenges and requiring specific strategies, medical image segmentation (MIS) must adapt to modality-specific characteristics and underlying physics [8,9]. Computed Tomography (CT) relies on high-contrast density differences, making threshold-based and region-growing methods effective. Magnetic Resonance Imaging (MRI), with its rich soft tissue contrast, often benefits from deep learning models that account for spatial and intensity variations. Positron Emission Tomography (PET), which provides functional imaging, is typically segmented using fusion techniques that integrate anatomical information from CT or MRI. Ultrasound (US) images are affected by speckle noise and variability, necessitating adaptive filtering and machine learning approaches. Optical Coherence Tomography (OCT), frequently used in ophthalmology, requires layer segmentation techniques using graph-based or deep learning methods. X-ray images, often challenged by overlapping structures, benefit from deep convolutional networks and attention-based models. Understanding these modality-specific differences is essential for designing robust segmentation algorithms suited to diverse clinical applications.

Two major types of MIS are commonly used and distinguished: semantic segmentation and instance segmentation [10]. Semantic segmentation involves classifying each pixel in an image into a predefined class. This means all pixels belonging to a particular class (e.g., tumor, organ) are labeled with the same identifier. Instance segmentation classifies each pixel and distinguishes between different instances of the same class. For example, it can differentiate between multiple tumors in a single image. While both semantic and instance segmentation are crucial for medical image analysis, they serve different purposes. Semantic segmentation focuses on classifying each pixel into a class, which helps identify regions of interest. In contrast, instance segmentation goes further by distinguishing between individual instances of the same class, which is essential for detailed analysis and quantification tasks [11]. It can facilitate the detection of microcalcifications in mammograms and tumor volume and automatic counting of blood cells [2]. Furthermore, it can provide the necessary spatial and volumetric information to assist in quantitative analysis and the consequent diagnosis of the patient state [12]. It is particularly essential in Radiotherapy (RT) as it can provide the needed segmentation of CT scans based on which the physician decides the dose to administer based on a computerized radiotherapy planning system (RTPS) [7].

In short, MIS plays a pivotal role in the region of interest (ROI) extraction, lesion quantification, and 3D reconstruction [13]. However, accurate computer-aided segmentation faces multiple challenges [2, 13]:

- Being of an interdisciplinary nature, the field of MIS requires extensive cooperation between clinicians and machine learning scientists.
- Noise and other acquisition-associated artifacts in medical images make it difficult to be processed than natural images. This challenge is exacerbated by discrepancies between different imaging modalities and variations within the same modality (e.g., different X-ray machines producing images with varying contrasts and noise levels).
- Existing medical data are limited due to the difficulty and costliness of acquiring annotated medical images. Traditional MIS Approaches (discussed in Section 3) require substantial data to process images effectively.
- Inherent fallibility of specific imaging modalities. For example, soft tissues and lesions are ambiguous in CT scans, and the anatomical structure of bones is not well delineated in MRI images. This challenges segmenting regions with missing edges and a lack of texture contrast.
- Due to variations in spatial characteristics of images, as well as in the objective behind segmentation and the nature of the original image taken, it is challenging to develop a universally applicable segmentation method that can be executed in clinical trials.

Due to these challenges, recent research has proposed a range of solutions using diverse deep learning neural network architectures. This work aims to present a comprehensive tutorial survey covering various mechanisms and paradigms related to deep learning-based medical image segmentation (MIS), from pre-training strategies to advanced models. While most existing surveys tend to emphasize specific branches of deep learning, this paper adopts a broader perspective. It provides balanced attention to traditional, current, and emerging trends, making it especially useful for junior researchers entering the field. Additionally, we conclude the paper with a case study on lumbar spine segmentation. The main contributions of this paper are as follows:

- Reviews related literature and outlines this survey's unique contributions (Section 2).
- Provides a tutorial on traditional image segmentation methods that remain relevant today, including thresholding, edge detection, region-based segmentation, clustering, model-based segmentation, and graph-based approaches (Section 3).
- Offers a chronological overview of deep learning-based MIS, from conventional convolutional neural networks to fully convolutional networks and U-Nets, which form the foundation of many contemporary MIS techniques (Section 4).
- Examines semi-supervised learning methods in MIS, covering pseudo-labeling, unsupervised regularization, prior knowledge embedding, generative adversarial networks (GANs), and contrastive learning (Section 5).
- Highlights current and emerging trends in MIS, including cascaded networks, attention mechanisms, medical transformers, neural architecture search, cross-modality segmentation, distributed learning, active learning, uncertainty quantification, and lightweight networks (Section 6).
- Presents a case study on vertebral lumbar spine segmentation, including medical context, segmentation techniques, and recent contributions utilizing U-Nets and autoencoders (Section 7).
- Discusses limitations in current research and proposes possible future research directions.

Medical image segmentation is a cornerstone of modern medical imaging, enabling precise diagnostics and effective treatment planning. Despite notable progress, continued research is essential to overcome existing challenges and enhance the efficiency, robustness, and accuracy of segmentation techniques. This survey aims to serve as a foundational tutorial for both traditional and state-of-the-art approaches in the field, encouraging informed contributions and future advancements in medical image segmentation research.

2. Related works

Many survey papers have attempted to delineate various aspects of contemporary MIS research. In [5], the classification was modality-based; the authors distinguished between six imaging modalities (CT, MRI, US, X-ray, OCT, and PET scans) and surveyed the works in each. A survey of recent breakthroughs in the field of MIS was conducted in [14]. This survey focuses on technical challenges and emerging research trends, such as knowledge distillation, contrastive learning, medical Transformers, prior knowledge embedding, cross-modality

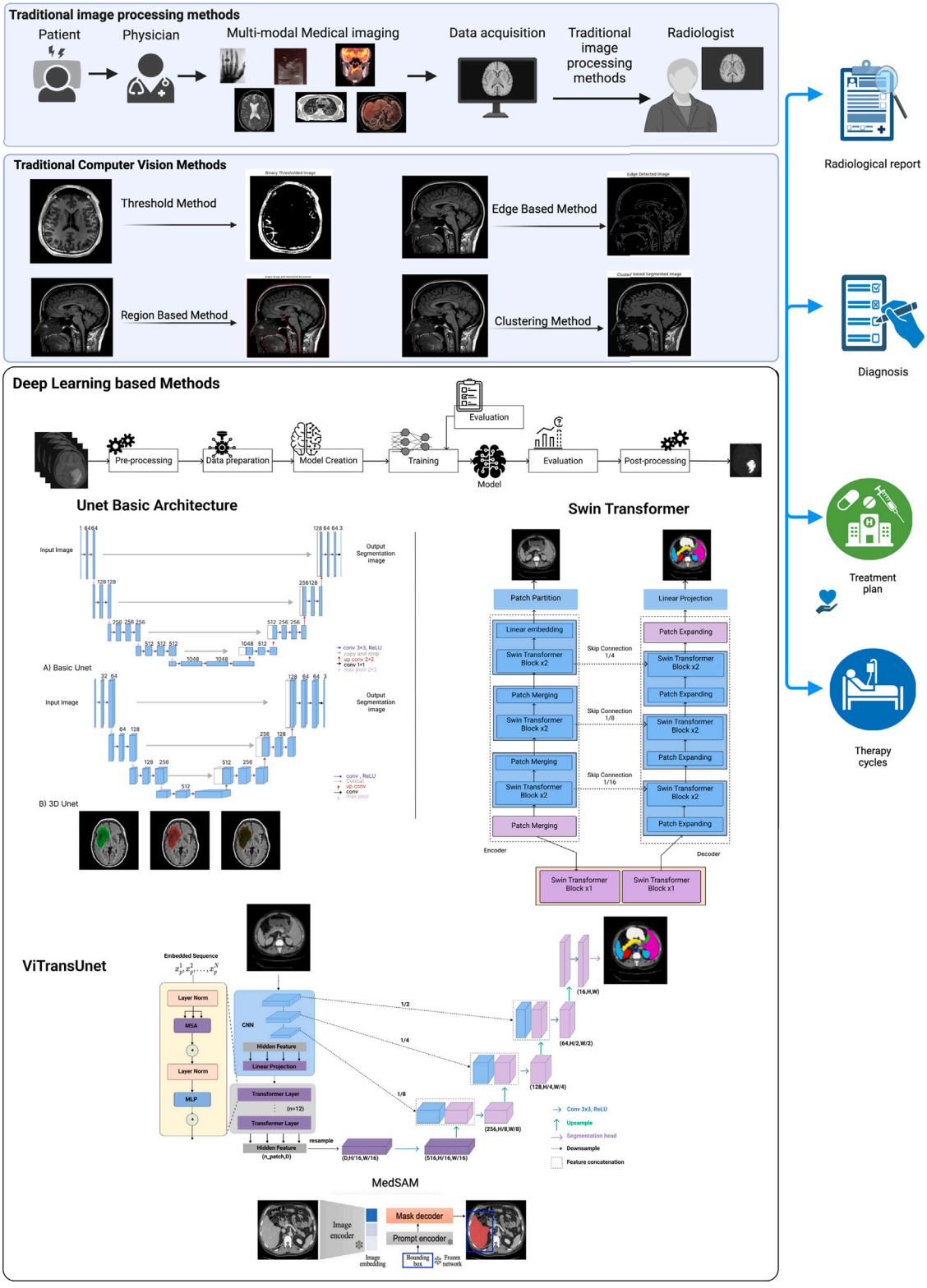

Fig. 1. Medical image segmentation methods summary. Traditional image processing methods are mainly data-driven approaches. Traditional computer vision methods are based on extracting spatial image features. Deep learning methods are commonly based on complex neural networks that can extract image features without prior hand-crafting.

analysis, federated learning, and active learning. The authors in [15] primarily focused on different U-net architectures (see Section 4 for further details on the concept); they reviewed variants of U-nets and their applicability to various imaging modalities. The most recent trends and optimizations of image thresholding, a traditional segmentation technique still relevant to contemporary literature, were surveyed in [2]. The work in [12] focuses exclusively on semi-supervised learning methods for MIS and categorizes them into pseudo-labels,

unsupervised regularization, and knowledge priors. They also discussed the limitations and research directions of existing semi-supervised approaches. The state-of-the-art deep learning techniques were reviewed in [4], and special attention was given to their application in radiology. In [7], contemporary optimizations to traditional image segmentation methods were reviewed, including region-based techniques, clustering techniques, edge detection, and model-based techniques; furthermore, the Lattice Boltzmann method was given special attention. The authors in [10] presented a comprehensive survey of recent trends in deep learning, including neural architecture search, graph convolutional networks, multi-modality data fusion, and medical Transformers. In [13], particular attention was given to recent works on clustering algorithms as well optimizations to U-net, and a discussion ensued about extending traditional clustering algorithms using U-net. The work in [16] offers a comprehensive review of DL-based image segmentation, focusing on supervised frameworks like fully convolutional networks and U-nets and unsupervised frameworks like generative adversarial networks. They have also extensively classified the currently available medical datasets for deep learning research.

While most of these works offer extensive and insightful overviews of current works, they often emphasize a few branches of the field over others, missing out on the inter-relatedness of different branches. As such, our survey aims to present the broad spectrum of ongoing research in the MIS field and to delineate the regions in which they overlap and those in which they complement each other (Fig. 1). As such, the primary contribution of this work lies in its relative comprehensiveness, bringing together other previous works and highlighting their interrelatedness, as well as its unique distinction of being tutorial in nature, to aid researchers in the field aiming to produce novel contributions.

3. Traditional medical image segmentation approaches

Before the rise of data-driven segmentation methods, most research focused on mathematical models and low-level image processing techniques [17], including thresholding, edge detection, region-based clustering algorithms, graph theory, and model-based segmentation.

3.1. Thresholding

The thresholding process entails dividing the image into three categories of pixels that are either smaller than, equal to, or greater than a predetermined threshold value [2]. There are two thresholding techniques: global and local (adaptive) thresholding.

Global thresholding views the image as a bimodal histogram, a deep valley between two distinct peaks [2], where one peak represents the target object and the peak represents the background. Object extraction then occurs by comparing pixel values with a threshold [18], yielding a binary image with pixel values taking either 0 (background) or 255 (object). Global thresholding techniques include the Otsu method [19], the Kittler–Illingworth method [20], and entropy-based global thresholding [21]. While computationally simple and fast, global thresholding only works well for images that contain objects with uniform intensity values on a contrasting background, however, it fails if the image is noisy, the contrast is low or the background intensity varies significantly across the image [22].

Local thresholding involves either splitting the image into sub-images and calculating thresholds for each or examining intensity in the neighborhood of each pixel and then determining the threshold based on intensity distribution [2]. It is more computationally expensive than global thresholding but works well in significant background variations or when the object is small.

3.2. Edge detection

Relying on sudden changes in color or intensity, edge detection is one of the most fundamental image segmentation methods. It uses derivatives to determine edge pixels in an image first, then connects them to form boundaries [23]. First derivatives are usually calculated using the Roberts, Prewitt, and Sobel operators, whereas second derivative operators include Laplace and Kirch [13]. The Canny edge detector applies non-maximum suppression followed by hysteresis thresholding, achieving higher accuracy and fewer broken edges [24].

3.3. Region-based segmentation

The region-based segmentation method uses the local spatial information of the image to combine pixel similarity into segmentation results [25]. This is achieved using region growing, region split and merge, and the Watershed approach [7].

In region growing, a seed point is first selected in the image. Then, the region expands by accepting neighboring pixels that share specific criteria with the seed, including pixel intensity, color, or spatial proximity [26].

Region split and merge process entails first dividing the image into large regions, using grid partitioning or quadtree decomposition [7]. Each region is recursively subdivided into smaller regions until each small region has pixels sharing similar qualities based on predefined criteria. Finally, to avoid over-segmentation, the small regions are merged based on similar attributes or neighborhoods. It is computationally expensive for complex images; however, the hierarchical nature of the process aids multi-level analysis of the image.

Particularly useful for segmenting adjacent regions, the watershed approach relies on intensity gradients to delineate boundaries between objects, treating pixel intensities as topographical features [7]. The idea is that low-intensity pixels are viewed as valleys and high-intensity pixels are viewed as peaks [27], water flows from the peaks to the valleys, and where water from two different peaks meets at the same valley, a boundary is created. This method is particularly effective in boundary preservation and works well with complex images without prior information. However, it is computationally expensive and vulnerable to over-segmentation if markers (peaks) are not accurately chosen.

3.4. Clustering

Clustering is the process of dividing pixels into homogeneous regions sharing similar values. It is an unsupervised, efficient, and self-adaptive process [28]. It can be sub-categorized into hard clustering and soft (fuzzy) clustering.

The traditional method of hard clustering is the K-means clustering algorithm. The idea is to choose K centroids and then assign each data point in the image to a specific centroid based on Euclidean distance, which segments the image into K clusters. By taking the average value of the coordinates of all points in a particular cluster, a new cluster centroid can be specified and repeated until no significant change occurs between two successive iterations [13]. Improvements to K-means include crowd-based (FGO) optimization [29], median filtering, Sobel edge detection and morphological operations [30], two-stage fuzzy K-means [31], combining K-means with discrete wavelet transform [32] and Darwinian particle swarm optimization [33].

Contrary to hard clustering, in soft clustering algorithms, data points can belong to multiple clusters simultaneously, offering a more nuanced representation of cluster points. It is more robust to noise than K means and allows for greater flexibility since the degree of overlap (fuzziness) can be controlled. The traditional soft clustering algorithm is the fuzzy C-means algorithm, improvements to which include fuzzy local intensity clustering, fuzzy clustering based on spatial information,

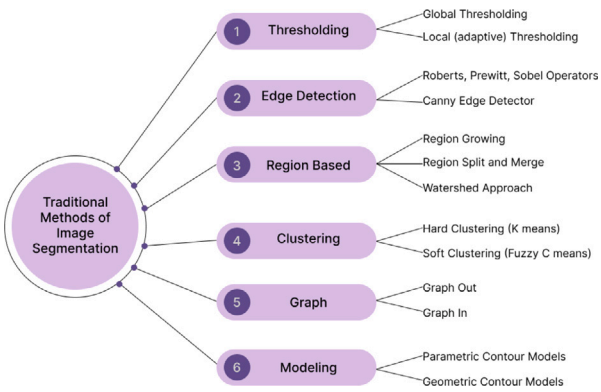

Fig. 2. Traditional methods of image segmentation.

and new fuzzy clustering algorithms, which are reviewed in detail in [13].

A graph-based representation of an image attributes each pixel to a node on the graph, and the connections between adjacent pixels are represented as edges. The weight of the edge is a measure of similarity between neighboring pixels in terms of gray level, color, or texture. The target then becomes to divide the graph into sub-graphs that share the maximum possible feature similarity [13]. Two possible techniques are graph cut [34] and grab cut [35], where the former is a one-time energy minimization algorithm, and the latter is a more interactive iterative process.

In model-based segmentation, continuous curves express the target edges in an image using parametric and geometric contour models [36].

Active Shape Models (ASM) is a statistical shape model constructed from a training set of different positions and orientations of the same object to capture its variations and constraints [37]. Then, an iterative process commences with the intention of finding the best model to fit the statistical shape and the new image.

Active Appearance Models (AAM) use two statistical models: shape and appearance (including texture). This yields a more robust object representation at the expense of greater computational complexity [38].

Rather than using mathematical functions to describe the contours, geometric contours comprise points and vertices, yielding greater accuracy at the expense of reduced flexibility compared to parametric contours [39]. Geometric contours outperform parametric contours in their ability to handle topological changes in curves and their insensitivity to initial positions. Fig. 2 summarizes traditional methods of image segmentation.

3.5. Strengths and limitations

Traditional medical image segmentation approaches provide a foundational framework for extracting meaningful information from medical images, offering both notable strengths and limitations. One of their primary advantages lies in their simplicity and interoperability, supported by well-established mathematical principles. Techniques such as thresholding and edge detection are computationally efficient, easy to implement, and require minimal data, making them suitable for real-time applications and resource-constrained environments. Region-based segmentation methods leverage spatial coherence, enhancing segmentation accuracy when objects exhibit clear intensity homogeneity. Clustering techniques, particularly K-means and fuzzy C-means, offer flexible, self-adaptive solutions that require limited prior knowledge. Despite these advantages, traditional methods face several critical limitations. Global thresholding often performs poorly under non-uniform illumination or low contrast, leading to inadequate object-background separation. Edge detection is highly sensitive to noise and artifacts, frequently producing fragmented or incomplete boundaries that necessitate post-processing. Region-based methods can be computationally intensive and prone to over-segmentation, especially in noisy or highly textured images. Clustering techniques struggle with determining the optimal number of clusters and may be influenced by initialization biases. Moreover, traditional approaches generally lack adaptability to complex, high-dimensional medical images, reducing their effectiveness in addressing anatomical variability, pathological complexity, and multimodal datasets. These limitations have contributed to the growing adoption of machine learning and deep learning-based methods, which are capable of learning high-level representations and adapting to the complexity of medical imaging challenges.

Table 1 summarizes a comparison of traditional medical image processing techniques (such as thresholding and edge detection) with deep learning methods (such as CNNs and U-Nets) in terms of advantages, limitations, and use cases, including metrics like accuracy, speed, and robustness to noise.

4. Deep learning based medical image segmentation

While traditional MIS methods are not contemporarily irrelevant, their frequency in recent literature is diminishing compared to data-driven segmentation techniques based on deep learning. The inherent capabilities of DL allow it to learn about abstract features of data at different levels, enabling the detection of image morphology and texture patterns [5]. Good at observing hidden patterns in images [4], DL has facilitated robust image segmentation across various diseases, anatomies, and imaging modalities [40,42,49] by enabling quantitative analysis and 3D visualization of medical images [5]. Therapeutic planning, follow-up, prognostic, dosimetric, and radionics applications have been concretely affected by the development of DL models [50]. However, while inputs and outputs to a DL network are precise, the behavior of hidden layers is ambiguous, and their functionality cannot be easily replicated or understood, which is why DL has not yet been applied to any large-scale real-world medical trials despite exhibiting tremendous promise [15].

4.1. Convolutional Neural Networks

Deep learning simulates the human brain's learning process using neural networks [16], enabling it to extract features from large-scale data without human supervision. The Convolution Neural Network (CNN) is the classic DL model used in image processing [51]. Many accomplishments have been achieved in image feature extraction, pattern recognition, and classification owing to CNNs [16]. Proposed in 2012, AlexNet [52] was distinguished for its success in image classification. CNN have become a popular and effective tool for this purpose due to their ability to learn and extract features from images. However, limitations of CNNs include their overreliance on geometric priors, rendering it difficult to fully capture the intrinsic relationships between different objects using extracted local features [10]. Graph CNNs were proposed as a powerful and intuitive alternative for non-Euclidean spaces, which enabled the exploitation of intrinsic relationships to extract otherwise invisible connections between objects.

The CNN comprises multiple hidden layers sandwiched between an input layer and an output layer and is responsible for various tasks such as convolution, pooling, and activation [16]. The input layer is connected to the input image, comprising several neurons matching the pixels of the image. The second layer is a convolution layer, which performs feature extraction on the input data to yield a feature map; this is determined by parameter setup in the convolution kernel. The next layer is a pooling layer, which is responsible for filtering and selecting feature maps to simplify the computational complexity of the network. Then, a fully connected layer connects all neurons in the previous layer to yield an output, which is finally sent to the

Table 1
Comparison between traditional and deep learning methods.

Aspect	Traditional methods	Deep learning methods
Advantages	• Simple and easy to implement [6,40] • Low computational cost [40] • Effective for well-defined edges and contrasts [40]	• High accuracy and generalizability [6,40–42] • Robust feature extraction [6,40,41] • Adaptable to various medical imaging tasks [6,41,42]
Limitations	• Low classification accuracy [40] • Poor robustness to noise and variations [6,40] • Limited to simple segmentation tasks [6,40]	• High computational complexity [42] • Requires large annotated datasets [43,44] • Potential overfitting and generalizability issues [40,42]
Use cases	• Basic segmentation tasks [6,40] • Initial preprocessing steps [45]	• Complex segmentation tasks (e.g., tumors, organs) [6,42,46] • Disease detection and classification ([47,48], ?)
Accuracy	Generally lower accuracy [6,40]	Higher accuracy [48]
Speed	Faster for simple tasks [40]	Slower due to high computational demands [42]
Robustness to noise	Poor robustness [6,40]	Better robustness with advanced architectures [40,42,49]

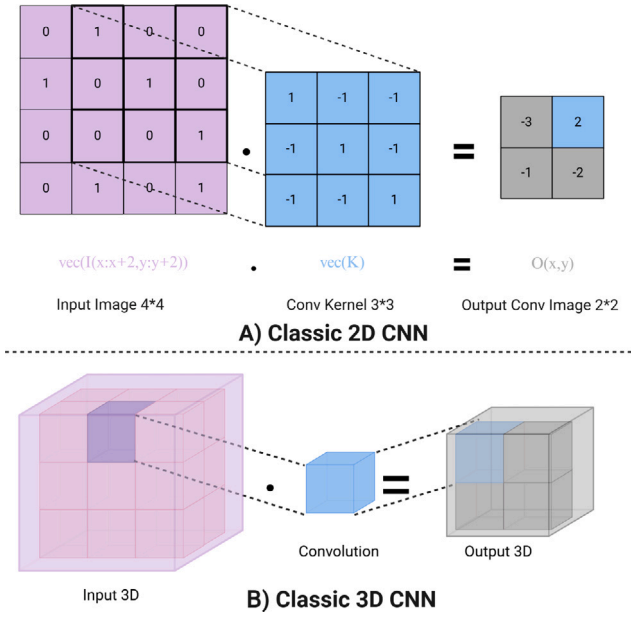

Fig. 3. Classic (a) 2D and (b) 3D CNN and the convolution operation [16].

classifier [16]. Examples of classic 2D CNNs are ResNet [53] and Visual Geometry Group (VGG) [54]. The classic 2D CNN is shown in Fig. 3 (a).

Medical images are essentially three-dimensional, even though clinicians tend to analyze 2D slices of these images. However, to effectively investigate these images, the convolution kernel must be 3D as well [55], which enables the extraction of more powerful volumetric representations and spatial considerations. An example of 3D CNN is shown in Fig. 3(b).

While increasing the number of network layers or the network depth can yield better results, the network is vulnerable to other problems like overfitting and vanishing gradients [16]. To solve this problem, GoogleNet [56] proposed an inception structure that increases the depth and width of the network while maintaining or reducing the number of parameters, which is achieved using multiple convolution kernels of different sizes and adding pooling. ResNet [53] also solved the problems associated with network depth using residual blocks, where each module consists of several consecutive layers and a shortcut that connects the input and output layers of the module before ReLU activation. Finally, squeeze and excitation blocks [57] (Fig. 4) improve the expressive ability of the network.

The encoder–decoder architecture, enhanced with CNNs and Transformers, has proven effective for image segmentation tasks. Innovations in boundary detection, attention mechanisms, and multi-scale feature fusion continue to improve the accuracy and efficiency of these models, making them suitable for various applications, particularly in medical imaging. SegNet [59] is a deep CNN architecture designed for semantic image segmentation. It employs an encoder–decoder structure to perform pixel-wise classification, making it suitable for various applications such as urban scene understanding, medical image processing, and autonomous driving. These include (PSP Net) [60], in which a pyramid pool module and a pyramid scene parsing network were proposed to aggregate context information of different regions into global context information in a pyramid-like scheme. In [61], an instance segmentation framework (Mask R-CNN) is proposed.

4.2. Fully Convolutional Networks (FCN)

The traditional CNN structure comprises convolutional layers followed by fully connected layers, which means that the network's final output is one-dimensional [16]; this makes it suitable for image classification and object detection tasks. However, in image segmentation, we require pixel-wise prediction rather than categorizing the image as a whole [62]. Fully Convolutional Networks (FCNs) improve traditional CNNs by removing the fully connected layers from their architecture and replacing them with convolutional layers instead; this enables the attainment of a 2D feature map of each pixel. Furthermore, FCNs can accept any image size at the input and use deconvolution layers to upsample the last convolution layer's feature map and restore the input image's size. However, upsampling in the traditional FCN yields fuzzy results insensitive to the image's details. Thus, improvements to the conventional FCN were proposed, including DeepLab v1 [63], which is inspired by VGG16 with atrous convolutions and Conditional Random Field (CRF), DeepLab v2 [64], which is inspired by spatial pyramid pooling but introduces atrous spatial pyramid pooling (ASPP) with a parallel convolutional sampling of holes at different sampling rates on a given input, and DeepLab v3 which introduced a cascaded atrous convolution module. Furthermore, a 2.5D approach to FCN was implemented in [65], which implements three FCNs for each 2D profile but works better for larger organs [16]. In [66], it was proposed to apply focal loss on FCN to reduce the number of false positives in medical images due to class imbalance. The FCN schematic is shown in Fig. 5. FCNs for 3D MIS have evolved significantly, with various innovative approaches addressing the challenges of volumetric data, computational complexity, and limited training samples. Models like AdaEn-Net, CMV convs, and UNETR represent the forefront of this field, achieving state-of-the-art results across multiple medical imaging tasks [67–69].

4.3. U-Net

In 2015, inspired by FCNs, Ronneberger et al. [70] designed a U-Net network for MIS. It is composed of a U-shaped channel similar in structure to the SegNet encoder–decoder architecture, where the encoder employs successive pooling layers to reduce spatial dimension, and the decoder progressively recovers object resolution using

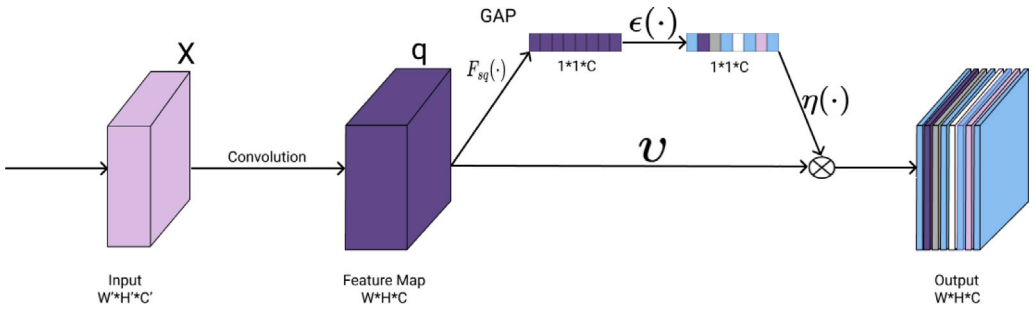

Fig. 4. Flowchart diagram of squeeze and excitation block [58]. GAP stands for global average pooling.

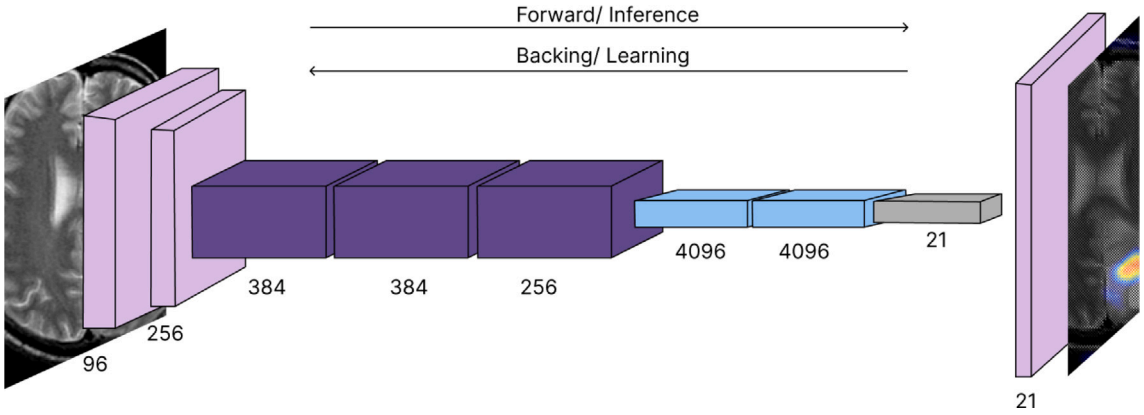

Fig. 5. Structure of FCN used for brain tumor segmentation in MRI [62].

upsampling. The contracting path encoder typically consists of sequential 3 × 3 convolutional layers followed by batch normalization (BN) and rectified linear unit (ReLU) activation. Spatial size is reduced using 2 × 2 max-pooling layers. The decoder is built symmetrically concerning the encoder, the only exception being that max-pooling layers are replaced with transpose convolution, bilinear interpolation, or any other upsampling operation. A final layer with SoftMax activation achieves pixel-wise segmentation at the original resolution. Skip connections improve localization accuracy and convergence speed by concatenating features between contracting and expanding paths [14]. U-Net's suitability for MIS stems from its ability to combine low-level information (for accuracy) with high-level information (to extract complex features), therefore propagating contextual information along the network, which allows it to segment objects in an area using context from a larger overlapping area [15]. The U-Net architecture is shown in Fig. 6(a). U-net received rapacious attention from the MIS community, prompting many improvements and developments that gave credence to the base U-net as their starting point. They are summarized next.

Three-dimensional U-Net and its variants have significantly advanced 3D MIS, offering high accuracy and efficiency in handling complex medical imaging tasks. The 3D U-Net architecture offers significant advantages, making it highly effective for medical imaging tasks. It is specifically designed to handle volumetric data, which is crucial for processing medical imaging modalities like MRI and CT scans that provide 3D images [71,72]. The architecture supports end-to-end training, optimizing the segmentation process using objective functions such as the Dice coefficient to address the imbalance between foreground and background voxels [72,73]. Additionally, 3D U-Net excels in feature extraction of multiple adjacent slices as input to capture more contextual information, which improves segmentation performance [71].

Despite its advantages, 3D U-Net faces challenges, particularly its computational complexity and the large amount of data required. To mitigate these issues, various modifications and optimizations have been introduced. For instance, the Self-Excited Compressed Dilated Convolution (SECDC) module reduces computational load while maintaining high segmentation accuracy by combining normal and dilated convolutions [74]. The Convolutional Block Attention Module (CBAM) enhances feature representation and segmentation accuracy through attention mechanisms [75]. Furthermore, contour loss refines segmentation by incorporating additional distance information to improve output precision [72].

Several variants and modifications of the 3D U-Net have been developed to enhance performance. The Y-Net variant employs dilated convolutions to capture features at multiple scales, improving the segmentation of small anatomical structures [76]. The Half-UNet simplifies the encoder and decoder components, reducing computational requirements while maintaining performance [77]. The Enhanced 3D U-Net also integrates spatial attention mechanisms to improve feature extraction and model robustness [78].

These advancements have enabled 3D U-Net models to achieve remarkable results in various applications. For brain tumor segmentation, enhanced models have demonstrated high Dice coefficients and robust performance [74,75]. In knee MRI segmentation, adjacent slices' modifications have improved accuracy and performance [71]. Similarly, dual 3D U-Net structures have precisely segmented the left atrium, with high sensitivity and specificity [72].

This combination of innovations and applications highlights the versatility and impact of 3D U-Net in advancing MIS. Fig. 6 shows the architecture of U-Net (A), 3D U-Net (B), and V-Net (C).

Attention U-nets [79,80] exploit attention gates to trim features irrelevant to the current task, focusing specifically on essential objects in an image [15]. Each layer in the expansive path has an attention gate through which the corresponding features from the contracting path must pass before concatenating with the upsampled features. Improved segmentation occurs due to localized classification information rather than global information without a significant increase in computational complexity. Fig. 7 shows an additive attention gate.

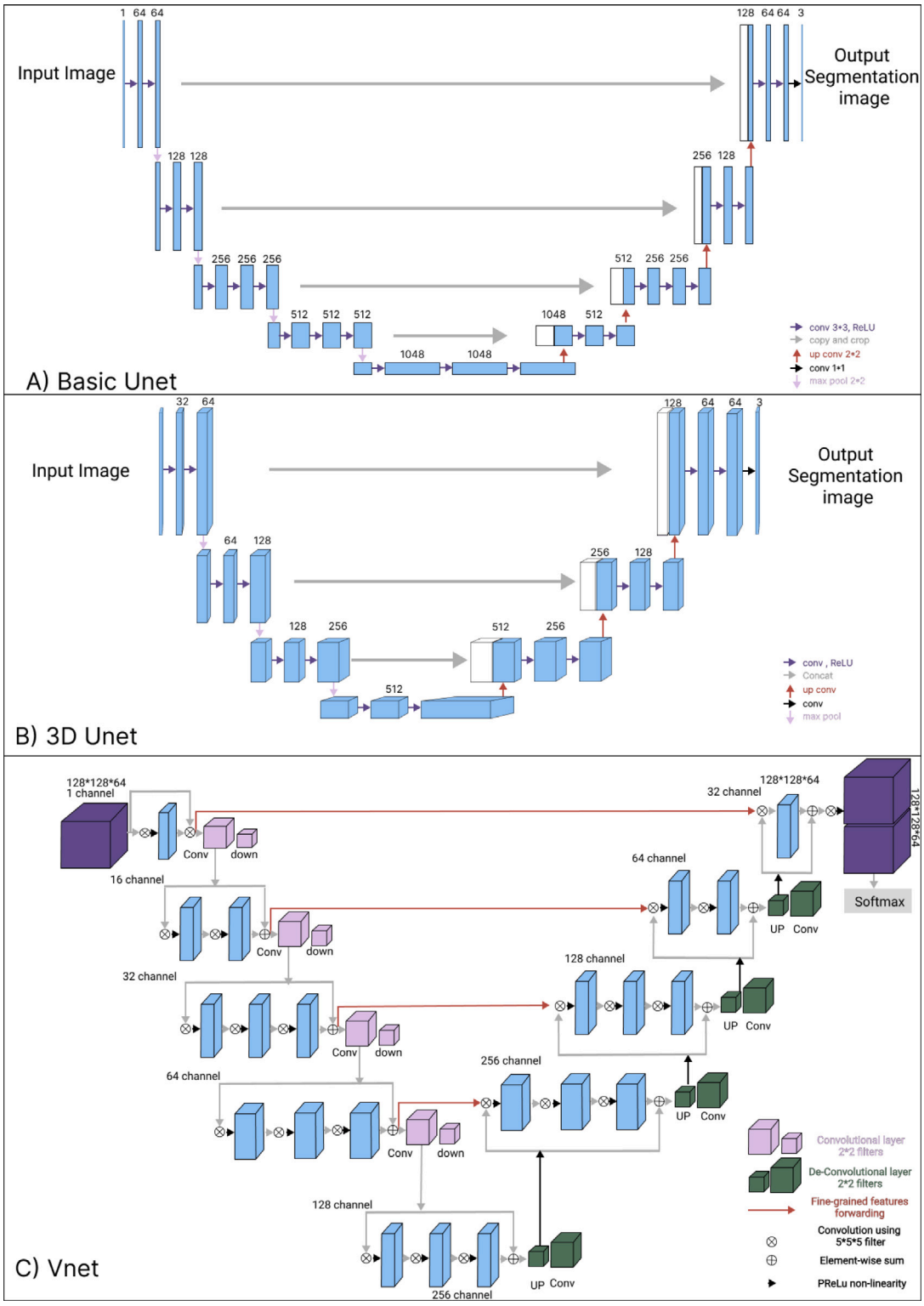

Fig. 6. (A) Basic U-net architecture. Blue boxes represent the feature map, and gray boxes represent the cropped feature maps at each layer [15]. (B) 3D Unet architecture. (C) V-net architecture.

It is often the case that segmentation requires looking at images with considerable variations in shapes and sizes; this is achievable using inception networks. Inception networks can effectively analyze images with different salient regions because they use multiple-sized filters on the same layer in the network [15]. GoogleNet [56] proposed the original inception network. This was followed by multiple improvements that achieved equivalent performance at substantial computational cost reduction, simply by replacing 5×5 convolutions with two successive 3×3 convolutions. Fig. 8 presents two configurations of the inception module.

The ResNet inspires residual U-nets [53] architecture, whose primary motivation was to overcome difficulties associated with training deep networks. The basic idea is that increasing the depth of the network yields faster convergence but at the expense of performance

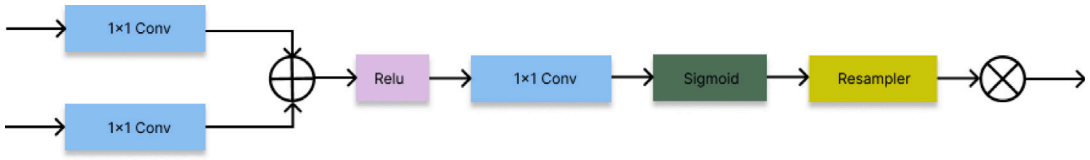

Fig. 7. Attention block in Unet [79].

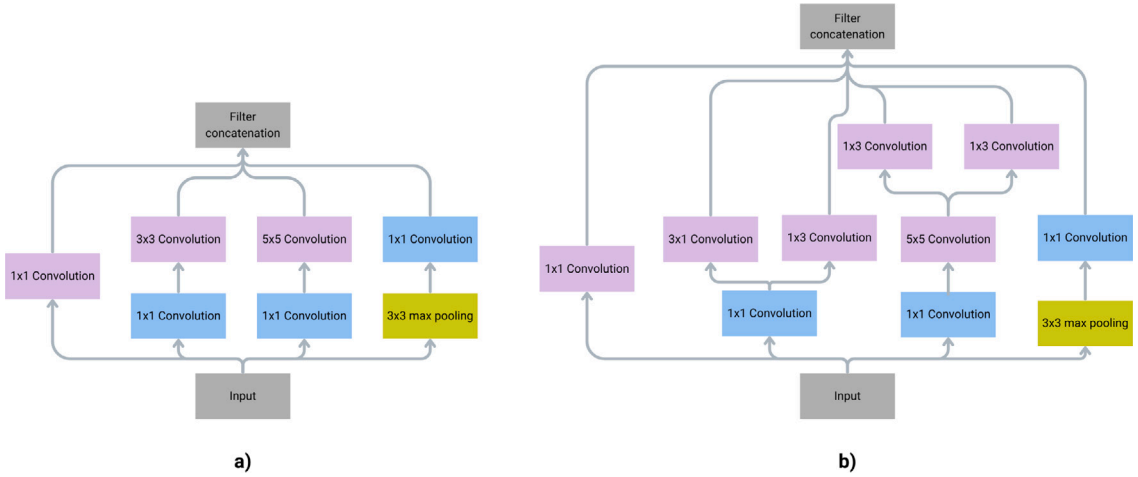

Fig. 8. The original inception module used in GoogleNet (b) Improved inception block with factorized filters, it yields an equivalent effect to (a) but at less computational power [15].

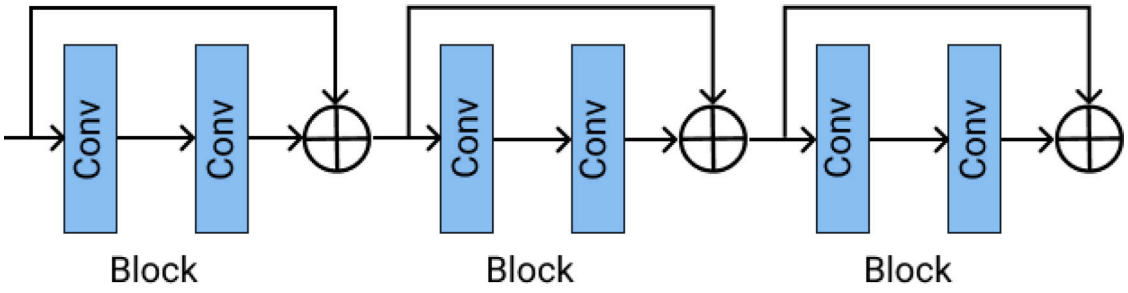

Fig. 9. Successive ResNet blocks with skip connections.

degradation due to the loss of feature identities caused by diminishing gradients [15]. ResNet tackles this problem by utilizing skip connections, which add the feature map of one layer to another layer deeper in the network, hence preserving the feature map. Fig. 9 displays the basic residual block architecture.

Initially designed to analyze sequential data, recurrent neural networks (RNNs) employ feedback loops (or recurrent connections) so that a node's output is affected by the previous output [15]. The recurrent U-net employs recurrent convolutional neural networks (RCNNs) [81], which allows units to use context from adjacent units to update their feature maps. The architecture is shown in Fig. 10.

DenseNet [82] is an extension of ResNet to better resolve the problem of vanishing gradients. This is achieved by complementing every layer in a block with feature maps from all preceding layers and combining feature maps into tensors using channel-wise concatenations rather than element-wise additions, as in ResNet. This significantly promotes gradient propagation, and each layer can have fewer channels as information is better preserved across layers [15]. This is shown in Fig. 11.

Inspired by DenseNet, Unet^{++} (Fig. 12) employs a dense network of skip connections between the contracting and expanding paths, aiding in the propagation of more semantic information between the two paths. It is beyond the scope of this paper to detail the vast

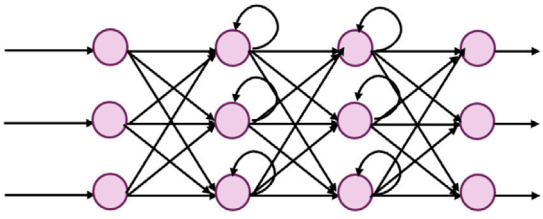

Fig. 10. RNN architecture.

number of different U-net architectures proposed in the literature, and we restricted ourselves to delineating the fundamental building blocks upon which other U-net architectures are based. To name a few, ensemble U-nets (or cascaded U-nets) are architectures combining two or more U-nets, where the first U-net performs a high-level segmentation and then each successive U-net performs segmentation on smaller objects [15]. Parallel U-nets were proposed in [83] and their results were aggregated for improved accuracy, a 2.5D U-net employs 3 U-net networks running in parallel on three different 2D projections of a 3D image yielding a 3D segmentation map at a smaller computational cost than a 3D U-net. 3D attention context Unet was

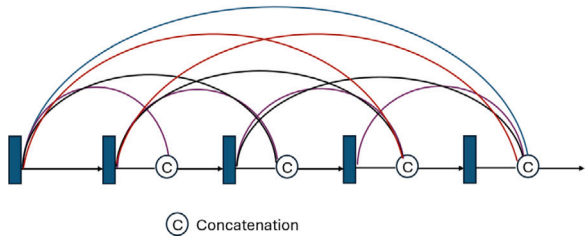

Fig. 11. DenseNet architecture.

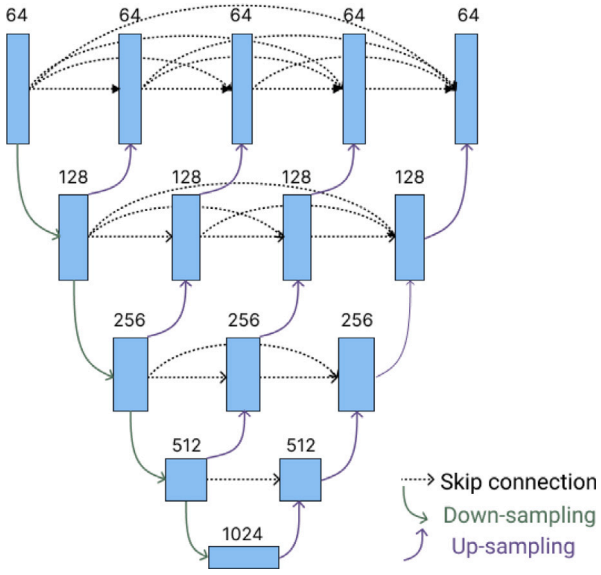

Fig. 12. Unet^{++} architecture [15]. An improved version of basic Unet (Fig. 6 A) with additional skip connections that enable better capture fine-grained features and reduce the semantic gap between encoder and decoder feature maps.

proposed in [84] with spatial attention blocks to aid in 3D context guidance [13]. Multi-scale nested U-net (MSN U-net) [85] is conceptually similar to Dense U-net in that the semantic gap is alleviated; however, they achieve this using a multi-scale context fusion block that combines the top and bottom layers [13]. RU-Net [86] combines the base U-net architecture with a multi-level boundary detection network using an image segmentation algorithm that employs a multi-layer boundary perception self-attention mechanism. Finally, no new Unet (nnU-net) [87] is a standardization attempt designed to tackle data set diversity by condensing and automating key decisions for designing a successful pipeline for any given dataset. nnUnet surpasses most existing approaches and achieves state-of-the-art performance while automatically configuring the strategies of pre-processing, training, inference, and post-processing to any arbitrary dataset [12]. nnU-Net automatically configures itself for any new task, including preprocessing, network architecture, training, and post-processing, without requiring manual intervention [87,88]. One limitation of nnU-Net is the lack of uncertainty measures, which can be problematic in heterogeneous datasets. Recent advancements have introduced methods to estimate uncertainty without altering the original architecture, enhancing segmentation accuracy and quality control [89].

Several U-Net variations have been proposed to address challenges in MIS, such as low contrast, class imbalance, and variability in anatomical structures. As discussed above, a widely used variants are Attention U-Net, ResUNet, and nnU-Net. These models demonstrate different features. Attention U-Net is known to enhances segmentation near complex boundaries and small structures which is useful in organ delineation and it requires slightly larger memory size in comparison with vanilla U-Net. ResUNet is known to improves gradient flow and paralytically performs well in noisy datasets and for multimodality applications. However, it requires additional memory resources to accommodate with residual blocks. nnU-Net is considered a leading architecture that reported state-of-the-art segmentation accuracy performance in several applications. Nevertheless, auto-configuration leads to complex pipelines that requires relatively high memory and computational powers.

Among their many superior attributes, U-nets accept images of arbitrary size, solve problems related to shadow and overlap [5], and create highly detailed segmentation maps using samples with minimal annotation [15]. This is achieved through the random elastic deformation of data [70] as well as the use of context-based learning. Another advantage is its incredible modularity and mutability, facilitating an avalanche of new and improved versions that improve its structure without changing its essential foundational structure. However, compressing the network model without reducing stability remains an open problem [13], and they are still dependent on high-quality labeled datasets. The semi-supervised learning approach discussed in the following section is an alternative philosophy to reduce this dependence. Table 2 compares the MIS key architectures.

4.4. Challenges

Deep learning has significantly advanced the field of MIS, offering automated and accurate delineation of anatomical and pathological structures. However, despite its success, the application of deep learning methods in MIS presents several notable challenges that limit their broader clinical adoption. One of the primary limitations is the dependency on large, high-quality annotated datasets. Medical image annotation is both time-consuming and resource-intensive, often requiring expert radiologists or clinicians. While some open source online resources are now available (REF), it still limited in terms of number of images and variety of clinical applications. The complexity increases when dealing with rare diseases or subtle pathologies, where inter-observer variability and limited sample availability exacerbate the challenge. Moreover, data privacy regulations and institutional barriers further restrict data sharing, making it difficult to compile diverse, representative training sets.

Another key issue is domain shift. Deep learning models trained on data from a specific institution or scanner often struggle to generalize to new environments. Differences in imaging protocols, scanner vendors, patient populations, and acquisition settings can lead to significant drops in performance when the model is applied outside of its original context. Without robust domain adaptation techniques, such models may fail in real-world clinical scenarios where consistency and reliability are critical.

Interpretability is also a pressing concern. Deep learning models, especially those based on convolutional neural networks and Transformers, are often viewed as black boxes. In clinical applications, understanding why a model made a certain segmentation decision is essential for trust and accountability. Although recent efforts in explainable AI provide some insights through attention maps or activation visualizations, these methods are still limited in terms of clinical transparency and interpretability.

Furthermore, many segmentation tasks suffer from class imbalance. Structures of interest, such as tumors or lesions, often occupy only a small portion of the image. This imbalance can bias models toward the background or majority classes, resulting in missed detections or reduced sensitivity for clinically important findings. Coupled with the high computational demands of training and deploying deep models, these issues can be prohibitive, especially in resource-limited clinical settings.

Finally, integrating deep learning models into clinical workflows involves overcoming regulatory, infrastructural, and evaluation challenges. Obtaining regulatory approval (e.g., FDA or CE certification),

Table 2
Comparison of key network architectures.

Architecture	Key features	Advantages	Limitations
U-Net	Encoder–decoder with skip connections	High accuracy, widely used, effective for 2D and 3D images	May struggle with complex tasks without modifications
U-Net++	Nested, dense skip pathways	Reduces semantic gap, improves performance	Increased complexity and computational cost
Attention U-Net	Attention mechanisms	Focuses on relevant regions, better for complex tasks	May require more computational resources
3D U-Net	3D convolutional layers	Suitable for volumetric data	Higher computational and memory requirements
ViT	Transformer-based	Powerful performance, captures global context	Limited generalizability, requires adaptation for different tasks
Hybrid Models	Combines CNNs and Transformers	Balances local and global features, efficient	Complexity in design and implementation
Active Contour	Deformable models	Maintains smooth boundaries, good for complex structures	May require manual tuning and initialization
SAM	Zero-shot learning, generalizes across modalities	Versatile, adaptable	Initial performance may be lower without fine-tuning

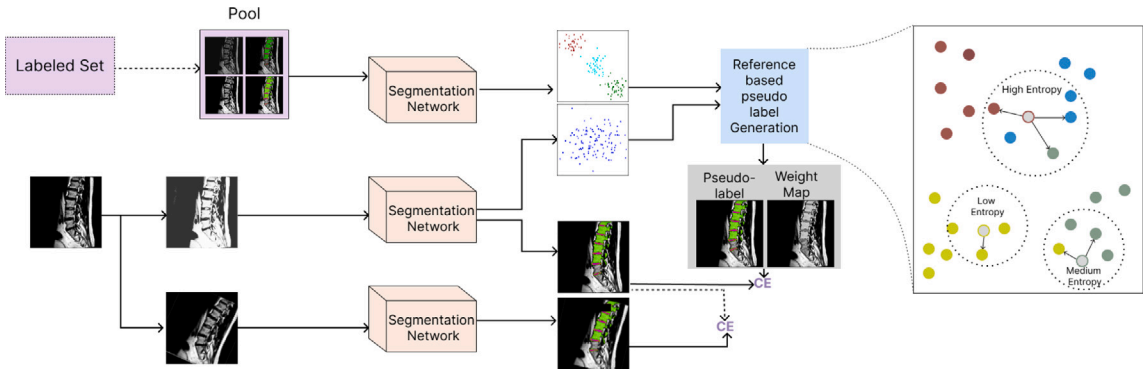

Fig. 13. Pseudo-label generation model [90].

ensuring data privacy compliance (e.g., HIPAA, GDPR), and aligning model outputs with clinically meaningful metrics are non-trivial tasks. Traditional performance metrics such as the Dice coefficient or Intersection over Union (IoU) may not fully capture the clinical relevance of segmentation results, underscoring the need for more standardized and clinically validated evaluation protocols.

5. Semi-supervised learning in medical image segmentation

Despite the considerable success of DL techniques in general, and U-nets in particular, in realizing accurate image segmentation, the impracticality of obtaining large-scale carefully labeled datasets remains problematic [12]. To ease the burden of manual labeling, considerable efforts have been dedicated to annotation-efficient DL techniques, including data augmentation, conditional generative adversarial networks, contrastive learning, and others, all succinctly described as semi-supervised learning techniques. In semi-supervised settings, we aim at building a model of comparable performance to fully supervised models but with a relatively limited amount of labeled data and a vast amount of unlabeled data. This section covers the various techniques and strategies for achieving semi-supervised medical image segmentation (SSMIS).

5.1. Pseudo labels

Pseudo-labeling is a popular semi-supervised learning strategy in which pseudo annotations are generated for unlabeled data and iteratively used to improve the segmentation model. Initially, a model is trained on a small amount of labeled data and then applied to the unlabeled set to produce pseudo-segmentation masks. These pseudo-labeled samples are then combined with the original labeled data to retrain and refine the model [12]. This process is repeated over multiple iterations to progressively enhance model performance. A key challenge in pseudo-labeling is handling noisy predictions, as inaccurate pseudo-labels can hinder training convergence [91]. Various methods have been proposed to mitigate this, differing in strategies for label quality assurance and model initialization.

Online pseudo-labeling involves using high-confidence predictions during training as temporary labels. For example, Jiao et al. [12] applied this approach to the Synapse multi-organ segmentation dataset, demonstrating that using a confidence threshold improved label precision. While annotation efficiency was significantly improved—reducing the need for manual labeling by over 60%, the quality of pseudo-labels was still contingent on threshold calibration and model stability.

Offline pseudo-labeling can be achieved through label propagation, with methods such as:

(a) Prototype learning [92]: This approach computes distances between feature vectors of unlabeled images and class prototypes. Using morphological post-processing, high-quality pseudo labels are generated. Han et al. [92] used the Left Atrium (LA) dataset and reported that pseudo-labeling helped reduce expert annotation time by nearly 50% while maintaining strong segmentation accuracy.

(b) Nearest neighbor matching [93]: Here, pseudo labels are assigned based on embedding similarity to neighboring labeled instances. This method was validated on the NIH pancreas CT dataset, showing that even with only 20% labeled data, performance approached that of fully supervised methods.

Fig. 13 illustrates a typical pseudo-label generation framework. While pseudo-labeling methods offer substantial annotation savings,

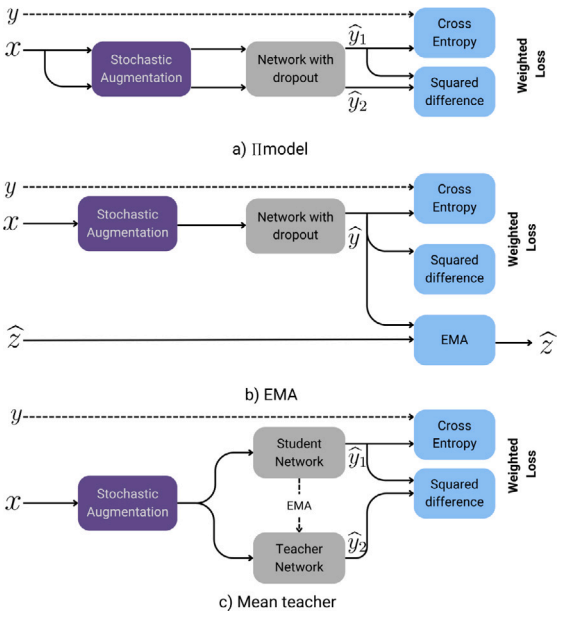

Fig. 14. Architectures of consistency learning. (a) *Π* model, (b) EMA, (c) Mean teacher [12].

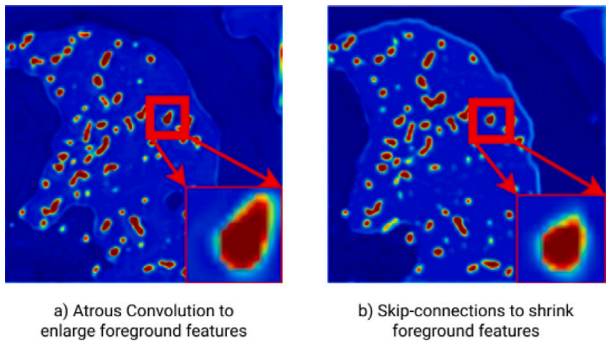

Fig. 15. Perturbations used in consistency learning. (a) Atrous convolutions to enlarge foreground features [96], (b) Skip connections to shrink foreground features [53].

their effectiveness heavily depends on model initialization, data distribution, and robustness to noisy predictions.

5.2. Unsupervised regularization

This approach attempts to incorporate unlabeled data into the training process with an unsupervised loss function [12]. Different choices of the unsupervised loss function and regularization term lead to various models.

Consistency learning enforces an invariance of predictions under different perturbations and pushes the decision boundary to low-density regions [12], based on the assumption that perturbations would not change the model's output. The consistency between two objects is determined using similarity measures like Kullback–Leibler divergence, mean square error, or Jensen–Shannon divergence [12]. Consistency Learning suffers from sensitivity to noise, dependence on the choice of parameters, and dependence on the choice of perturbations. There are different architectures to achieve consistent learning:

(a) *Π* model [94]: two random augmentations of a sample are created for both labeled and unlabeled data. The model expects consistency in the output of the same unlabeled sample propagating twice under different perturbations.
(b) EMA [95]: Exponential moving average predictions are used as consistency targets for unlabeled data. However, maintaining EMA during the training is a heavy burden.
(c) Mean Teacher Architecture: Alleviates the burden of maintaining EMA by utilizing teacher and student networks where the consistency of predictions from perturbed inputs between teacher and student is enforced. Fig. 14 shows the different architectures, and Fig. 15 shows different perturbations; for further detail on different perturbation types, refer to [12].

Unsupervised regularization with a co-training framework assumes two or more different views of each datum and that each view has sufficient information to generate independent predictions [97]. First, it learns a separate segmentation model for each view on labeled data; then, it gradually adds model predictions on unlabeled data to the training set to continue training. One view is assumed redundant to other views; the models are encouraged to have consistent predictions on all views [12]. It diverges from self-training models in that pseudo labels are added from one view to the training set and then used to supervise other views. It differs from consistency training in that all models in co-training undergo gradient-descent-based updates. This approach is limited by the assumption that each view is sufficient and independent enough to generate its predictions, the risk of overfitting if the two models are too similar, and sensitivity to noisy pseudo labels. Fig. 16 shows an example of co-training architecture.

5.3. Prior knowledge embedding

Different types of information can be included as prior knowledge of DL frameworks, including shape constraints, topology specifications, edge polarity, or inter-region adjacency rules [14]. In [79,99], the authors used autoencoders (AE) to demonstrate the learning of anatomical shape variations from medical images. Autoencoders follow an encoder–decoder architecture, where the encoder maps the input to a low-dimensional feature space (smaller than the input dimension), and the decoder reconstructs the image from the feature space. A typical autoencoder architecture is shown in Fig. 17. Priors are not restricted to shape only, in fact, texture, topology and size might be more meaningful priors to incorporate into a DL network for increased robustness [100]. Another example of prior knowledge embedding was featured in [101], where the authors designed anatomically constrained neural networks (ACNN) for MIS tasks. Limitations of knowledge priors include overfitting (if prior knowledge is too specific to the training data) and non-differentiable models (if knowledge priors are too complex, like region connectivity, convexity, and symmetry).

5.4. Generative Adversarial Networks (GAN)

Goodfellow et al. [102] introduced Generative Adversarial Networks (GANs), which consist of two neural networks, the generator and the discriminator, engaged in a competitive game. The generator creates synthetic data from random noise, while the discriminator attempts to distinguish real from fake data [16]. This adversarial process leads to progressively more realistic outputs from the generator (see Fig. 18).

The application of GANs to semantic segmentation was pioneered by Luc et al. [103], who utilized a CNN-based generator to synthesize segmentation masks that closely resemble the ground truth. While initially developed using the Cityscapes dataset, this approach showed that even with a limited number of labeled examples, the model could learn realistic structure, thus improving annotation efficiency by reducing dependence on pixel-wise annotations.

Segmentation Adversarial Network (SegAN) [104] incorporated a U-Net generator to handle medical image segmentation, where adversarial learning was used to refine outputs by minimizing the L1 loss. Evaluated on the ISBI 2012 EM segmentation dataset, SegAN

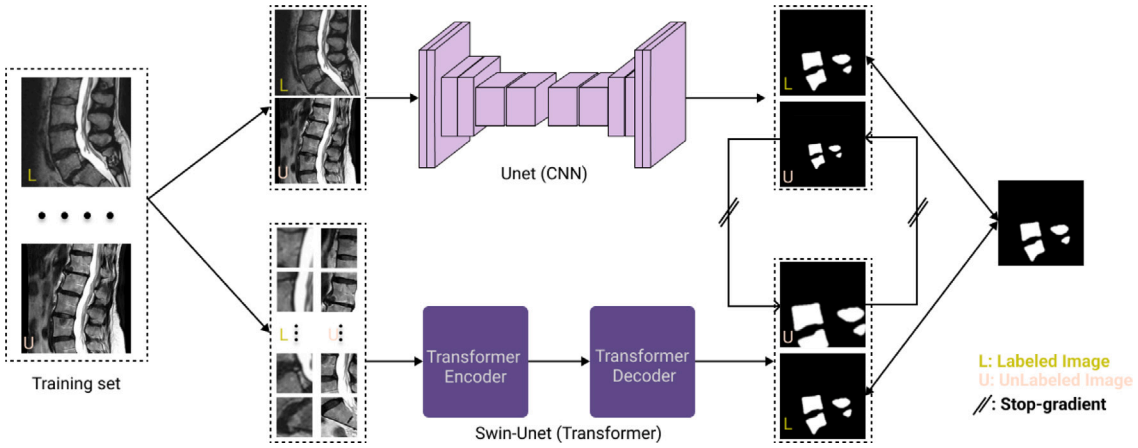

Fig. 16. A co-training framework [98].

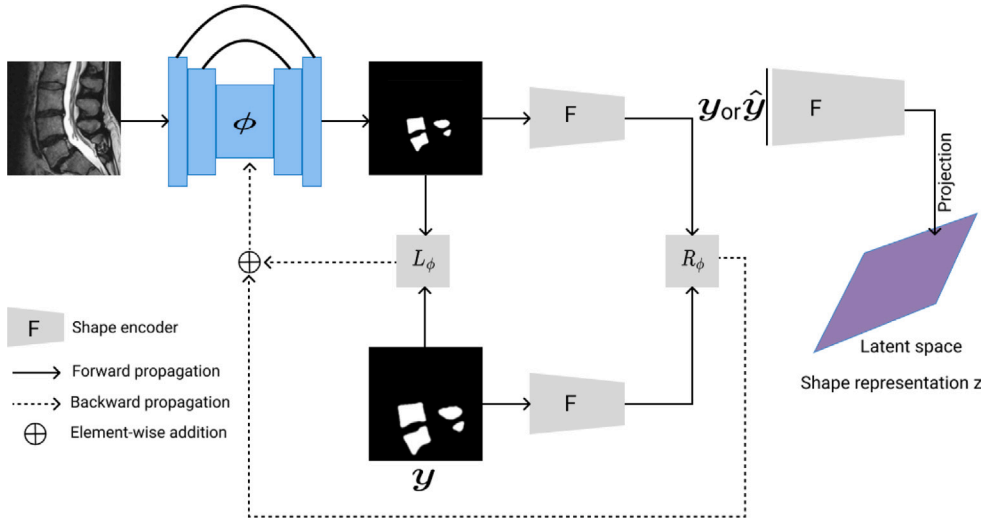

Fig. 17. Autoencoder architecture [14].

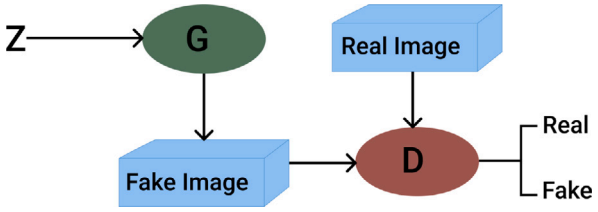

Fig. 18. Generative Adversarial Network basic architecture.

achieved comparable performance to fully supervised methods using only 50% of the labeled data. Structure Correcting Adversarial Network (SCAN) [105] extended this framework by using fully convolutional networks (FCNs) for both generator and discriminator. The discriminator imposed anatomical plausibility by enforcing structural consistency based on human physiology. Applied to the NIH Pancreas CT dataset, SCAN demonstrated superior boundary accuracy while reducing manual annotation effort by approximately 40%. Projective Adversarial Network (PAN) [106] introduced 3D context through 2D projections, avoiding the computational burden of 3D segmentation. The dual-discriminator design addressed global and local inconsistencies. PAN was tested on the PROMISE12 prostate MRI dataset, achieving competitive results while using only partial annotations during training.

Beyond segmentation, GANs have been employed for privacy-preserving synthetic data generation. In AsynDGAN [107], a distributed framework was introduced where a central generator collaborated with multiple site-specific discriminators. Applied to the BraTS brain tumor dataset, AsynDGAN enabled multi-institutional model training without sharing patient data, effectively eliminating the need for direct annotation at centralized locations.

Conditional GANs (cGANs) [108] offer another direction, where the generation process is guided by a condition (e.g., a class label or segmentation map). Singh et al. [109] used cGANs for breast tumor segmentation with limited labeled samples, reporting strong segmentation performance while relying on only 20% of the annotated dataset. Similarly, Wang et al. [10] showed how cGAN-based augmentation could reduce annotation needs by generating diverse and anatomically consistent samples for rare pathologies. A more complex setup was presented in [110], where a cascaded architecture of a GAN and cGAN generated label maps and corresponding synthetic images. The model was validated on synthetic datasets and later applied to real anatomical structures, showing potential for use in annotation bootstrapping workflows. A general illustration of cGAN is shown in Fig. 19.

Although many advanced GAN architectures exist, detailing all is beyond the scope of this review. We direct the reader to [111–113] for additional promising implementations in 3D medical image segmentation (MIS). In general, GAN-based frameworks significantly improve annotation efficiency, particularly in settings with limited ground truth availability or strict data privacy requirements.

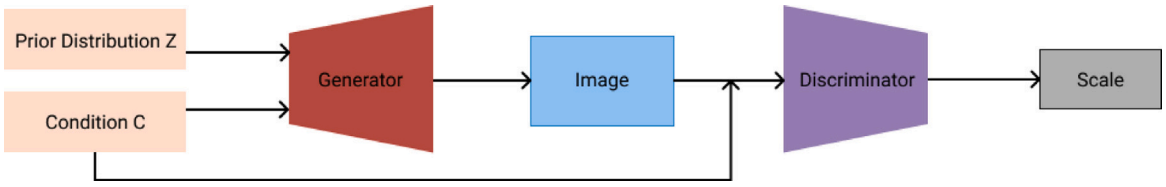

Fig. 19. A general sketch of cGAN architecture [114].

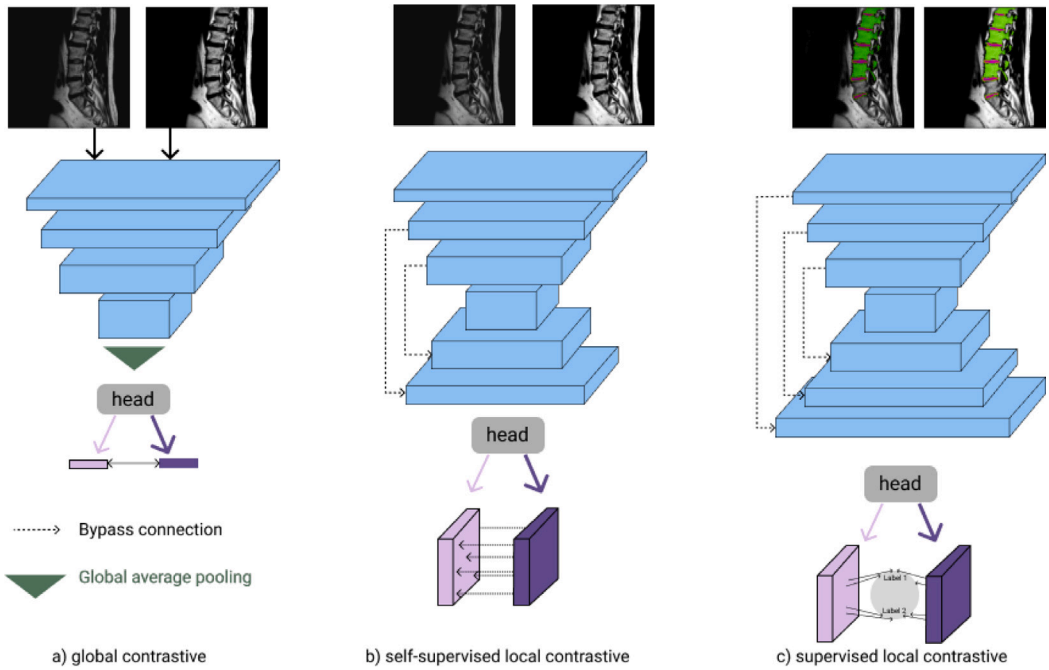

Fig. 20. Different examples of constructive learning architectures [117].

5.5. Contrastive Learning

Contrastive Learning (CL) is a technique used to improve the quality of visual representations by contrasting semantically similar and dissimilar pairs of samples. This method has shown significant promise in medical image segmentation tasks, including vertebra segmentation, by enhancing the ability to distinguish between anatomical structures without relying heavily on labeled data. ARCO [115] is a semi-supervised contrastive learning framework that uses variance-reduction techniques to improve pixel/voxel-level segmentation tasks with limited labels. Another example is the OBCL [116] Own-background Contrastive Learning framework that effectively incorporates background pixels into CL to handle imbalanced data distributions.

In addition, local contrastive learning can help students learn distinctive representations of local regions instead of relying on a global representation. Suitable for image segmentation. Fig. 20 shows contrastive learning architectures.

5.6. Validation and evaluation of SSMIS

Semi-supervised learning for medical image segmentation is typically validated using standard protocols similar to fully supervised approaches, with careful comparisons to fully-supervised baselines. Researchers often partition the limited labeled data into training and validation sets (or use cross-validation) to tune models, and evaluate final performance on a held-out test set or challenge benchmark. A common strategy is to vary the fraction of labeled data and compare semi-supervised models against fully supervised models trained on the same subset of labels. This reveals how close the semi-supervised method can approach the upper bound of using all labels. Semi-supervised models are expected to match or exceed such metrics relative to low-data baselines. Benchmark datasets are frequently used for evaluation enabling direct comparison to fully supervised competitors. Authors often report that semi-supervised models narrow the gap to fully-supervised performance as unlabeled data are leveraged, sometimes even outperforming a fully-supervised model trained on the same small labeled set [118]. Ultimately, rigorous validation on independent data (e.g. from external institutions or challenge leaderboards) is crucial to demonstrate that semi-supervised segmentation methods maintain accuracy and generalize in clinical scenarios.

While existing SSMIS techniques have achieved comparable results with fully supervised frameworks in specific contexts, they still suffer from certain limitations, such as misaligned distributions and class imbalance, uniform weight for unlabeled data, and integration with annotation-efficient approaches. In SSMIS, the size of unlabeled data far exceeds that of labeled data, and they may follow different statistical distributions, which could diminish performance [12]. Furthermore, the trained model can exhibit bias toward majority classes and, in some cases, completely ignore minority classes [119].

The traditional approach to SSMIS involves supervised loss for labeled data and unsupervised loss/constraints for unlabeled data. In most cases, there is a single weight to balance between supervised and unsupervised loss, which overlooks that not all unlabeled data is equally appropriate for the learning procedure [12]. This could be overcome by assigning individual weights to each unlabeled example to encourage the model to exploit more helpful information from unlabeled data [120].

Since acquiring fully annotated medical images is a complex process, some contributions suggest integrating SSMIS with other

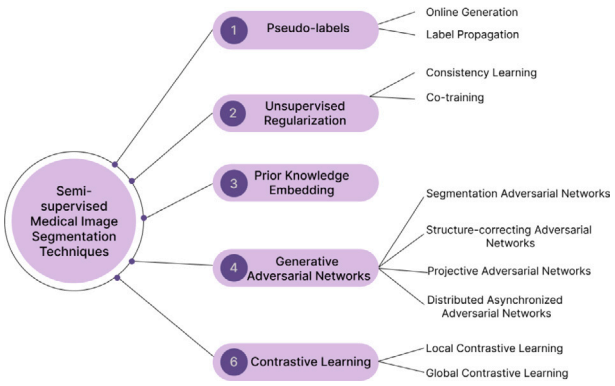

Fig. 21. Summary of Semi-Supervised Medical Image Segmentation Techniques.

annotation-efficient approaches that utilize partially labeled datasets [121], leverage box-level or pixel-level annotations [122] or exploit noisy labeled data [123]. Furthermore, it is suggested that SSMIS be integrated with few-shot segmentation to improve the generalization ability to segment unseen images better. The recent introduction of SAM (segment anything model) [124], which serves as a pseudo-label generator for image segmentation, promises future development in the field of SSMIS. Fig. 21 summarizes different SSMIS techniques

6. Recent and emerging trends in medical image segmentation

This section outlines recent innovations and future research directions in MIS.

6.1. Cascaded networks

Increasing the network depth to utilize larger receptive fields is unsuitable for memory and computational reasons, particularly in the context of medical images' volumetric (3D) nature, and with the added downside of high-resolution detail loss as network depth increases. To overcome this problem, the idea is to utilize cascaded networks in a pyramid-like structure that performs high-resolution segmentation while also considering contextual information from lower resolutions [14]. The weights of the lower resolution model are used as initializers of the higher resolution model through transfer learning (to be detailed subsequently in this section). There are three types of cascaded networks: Coarse-fine segmentation, Detection segmentation.

In coarse-fine segmentation, two 2D convolution networks are cascaded together; the first performs coarse (low-resolution) segmentation, and the second uses the output of the first to achieve fine segmentation [10]. The first network model possibly RCNN [125] or You-Only-Look-Once (YOLO) [126] is used for target location identification. Another network is used for further detailed segmentation based on the results of the first network. Since 2D networks cannot learn temporal data in the third dimension and 3D networks are too expensive in computation and memory, pseudo-3D segmentation methods have been proposed that involve the cascading of multiple 2D networks together for a more efficient version of 3D networks [10].

6.2. Attention mechanisms

Similar to the human visual system's tendency to focus on a tiny portion of highly relevant perceptible information, with disregard for other perceivable stimuli deemed irrelevant, DL frameworks use attention mechanisms to selectively focus on the more essential aspects of an image [14]. The model adaptively weighs its obtained features to focus specifically on those needed for the analytical task, suppressing feature responses in irrelevant backgrounds. The attention problem can be formulated using query (the target image), key (plausible target features), and value (the best matching regions).

The channel attention framework is formulated as each channel represents a feature map that typically denotes distinct objects. Consequently, channel attention strives to calibrate the weight of each channel, selecting the entities that deserve more attention. The squeeze and excitation block [127] is an excellent example of Channel Attention, where global spatial information is captured through a squeeze operation (like global average pooling), and an alignment function (excitation module) captures channel relationships. It outputs an attention vector using fully connected and non-linear layers followed by a sigmoid function. Finally, each channel of the input feature map is multiplied by the corresponding element in the attention vector for contextualization. Similar to Channel Attention, Spatial Attention attempts to adaptively calibrate the weight of each part of the image, choosing where to focus its attention using an adaptive area selection procedure [14]. Mutual attention between different sequences of MRI leads to improvement of low-grade brain tumors [128]. The branch Attention framework separates the attention problem into multiple sub-modules (branches), where each branch focuses on a particular aspect and exchanges significant information with other branches. Spatial Attention ignores inter-channel information variation, and Channel Attention pools global information directly; therefore, mixed attention networks have been designed to combine both advantages in [10]. In [129], a non-local U-net was implemented with a self-attention mechanism and a global aggregation block that extracts full image information during up-sampling and down-sampling. The non-local block is shown in Fig. 22.

6.3. Transformer-based methods

Transformers are attention-based architectures that were originally developed for natural language processing. Over time, they have been adapted for use in image processing tasks as well. A major milestone in this transition was the introduction of Vision Transformers by Dosovitskiy et al. [130]. In this work, the authors explored replacing convolutional neural networks with purely Transformer-based, convolution-free models for image segmentation. Unlike CNNs, Vision Transformers (ViT) offer a complete view of a single layer and facilitate parallel processing. The transformer-based encoder is a sequence of alternating layers of multi-head self-attention (MSA) and multilayer perceptron blocks (MLP). A layer normalization is applied before each block, and a residual connection is applied after [14]. In MIS, Transformer-based segmentation models still adopt a U-net-shaped architecture, as shown in Fig. 23.

A pure Transformer-based encoder enable the global context modeling capabilities of Transformers to effectively capture relationships between spatially distant voxels in medical images. When combined with convolutional upsampling layers and multi-level feature aggregation, this architecture significantly improves segmentation performance [131,132]. Various implementations have emerged to optimize these models for dense prediction tasks, such as segmentation. For instance, hierarchical Vision Transformers (ViTs) are designed to extract features at multiple resolutions without being limited to fixed subregions of an image [133,134]. One notable example is the 3D U-shaped model, which utilizes nested hierarchical Transformers and incorporates global self-attention within non-overlapping blocks [135]. This design streamlines the model while still effectively capturing both local and global contextual information, making it a powerful tool for medical image analysis.

This framework represents Transformers' global context modeling capabilities but with a CNN inductive bias, in that CNNs stack convolution blocks that capture multi-scale context feature maps. In contrast, Transformers capture long-term dependencies among features that can be lost in purely Transformer-based models. Finally, the Transformer

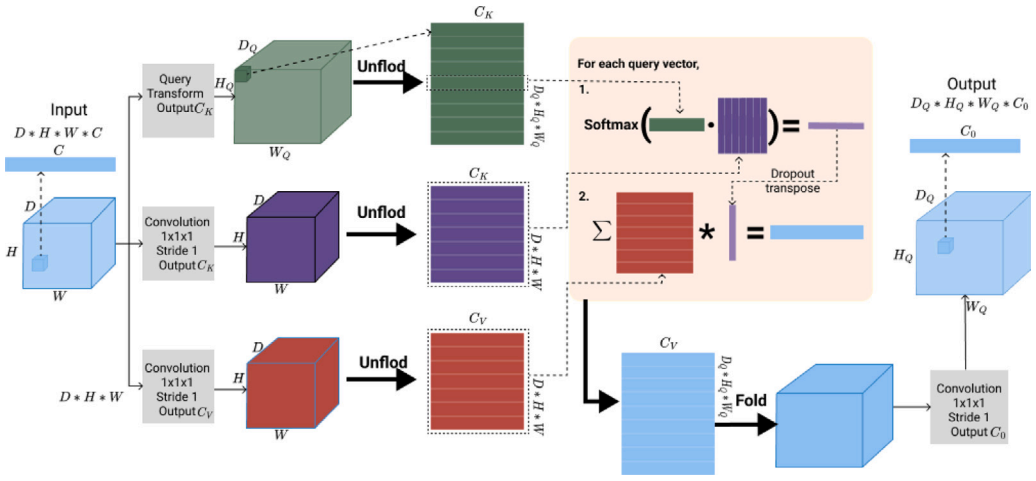

Fig. 22. Global Aggregation block in non-local U-net [129].

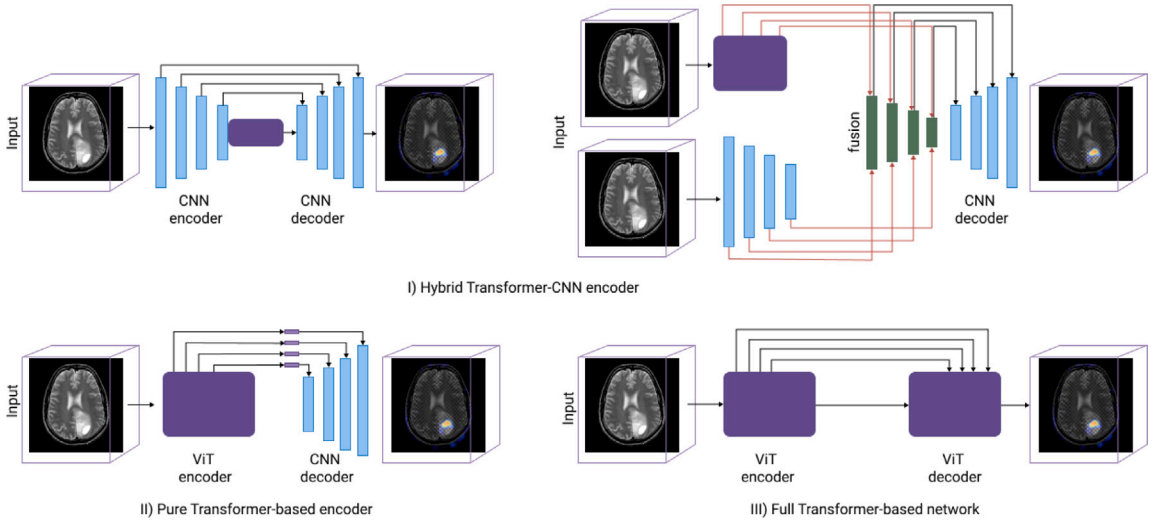

Fig. 23. Transformer architectures. (I) Hybrid Transformer-CNN encoder, (II) Pure Transformer-based encoder, and (III) Full transformer-based network [14].

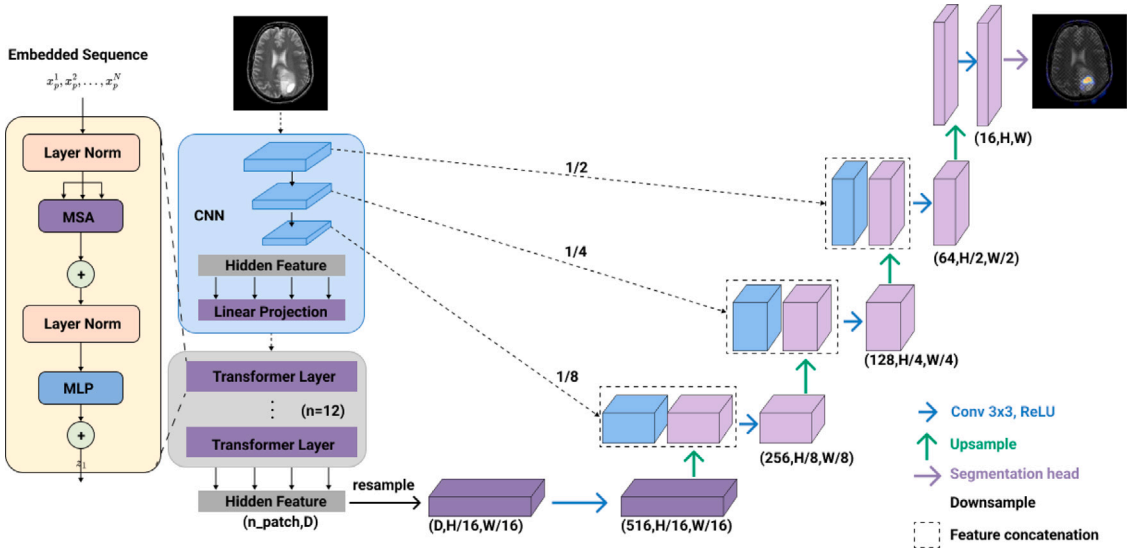

Fig. 24. TransUnet architecture [136].

output is up-sampled into a 4D feature map in a CNN-based decoder to recover the full segmentation mask [14]. This approach is illustrated in Fig. 24.

Transformer-based architectures have shown promise in medical image segmentation, but scaling them to large 3D images and clinical datasets poses technical challenges. Memory consumption is a primary concern: the self-attention mechanism has quadratic complexity in the number of input tokens (patches). Medical images (especially 3D volumes) produce far more patches than typical 2D natural images, making a vanilla ViT infeasible without downsampling. Early attempts to apply ViT directly in segmentation resulted in poor accuracy unless heavy downsampling was used. For example, the TransUNet framework addresses this by using a CNN encoder to first reduce spatial resolution, then feeding tokens to a ViT [136]. Similarly, Swin-UNet adopts windowed attention (local self-attention) to curtail memory usage, computing self-attention within small patches instead of globally [137]. Such hierarchical Transformers with local attention greatly reduce memory and FLOPs while approaching the global context modeling of full attention. Even with these optimizations, training speed and model complexity remain concerns. Transformer models often have tens of millions of parameters and require more computation than CNNs for the same input size. In one comparison, pure Transformer encoders (ViT) had significantly higher FLOPs than U-Net or hybrid models. Frameworks like UNETR [138], which uses a ViT as the sole encoder, can exceed 100 million parameters, necessitating multiple high-end GPUs or large memory for training.

The full Transformers network is where Transformers are stacked end-to-end. Neural Architecture Search (NAS) [139] strives to automate the iterative process of network design, traditionally executed manually by researchers [14]. It is a field that overlaps with hyperparameter optimization [140] and meta-learning [141]. The application of NAS to the design and optimization of Transformer-based architectures for medical image segmentation. Given the increasing complexity and scale of Transformer models (e.g., ViT, TransUNet, UNETR), manually designing optimal architectures becomes both time-consuming and prone to suboptimal performance or overfitting. NAS offers a systematic and automated way to explore and optimize design spaces. There are three NAS domains in contemporary research: Search space, search strategy, and performance estimation. This comprises the collection of existing networks to be searched. A global search space represents the search for an entire network structure, whereas a cell-based search space pursues only a few small structures stacked together to form more extensive networks [10]. This framework attempts to execute the fastest possible search within the search space. Possible search strategies include reinforcement-based learning, evolutionary algorithms, and gradients [10].

These are fixed quantifications of performance and rarely vary among different NAS algorithms. The nnUnet [87] is an example of a NAS-based product. Authors argue that excessive manual adjustment to network structure may lead to overfitting and, therefore, offer a new framework that adapts itself to any new dataset. It focuses on preprocessing (resampling and normalization), training (loss, optimization settings, data augmentation), inference (test-time-augmentations and model integration), and post-processing (enhanced single-pass domain) [10].

6.4. Cross-modality segmentation and fusion techniques

Cross-modal information is a potentially valuable but largely unused feature in MIS. However, exploring complementary and redundant information across various imaging modalities can improve segmentation performance. Data could be paired (coming from the same patient) or unpaired, which determines the type of fusion strategy to be adopted (early, mid, or late fusion). This can be observed in Fig. 25.

The most straightforward strategy is the early fusion strategy, where the modalities are integrated at the input level; this has the advantage of simplicity, which allows for more complex segmentation strategies like GAN-based approaches. However, early fusion strategies cannot understand non-linear relationships between low-level features from different modalities, especially when they have significantly different statistical properties. In paired data and mid-fusion strategy, multi-modal data is separately processed in different paths, then mapped into a common latent space via a fusion operation, and finally input into a decoder; the target is to emphasize the most significant features across modalities [14].

(a) Single layer fusion, where each modality has its own encoder, without inter-encoder communication, and a shared decoder. Encoded data is fused by concatenation, addition, or convolution. This technique encourages the network to learn the most correlated features across modalities and the most useful spatial information. Still, the single level of abstraction prevents it from learning within and between modalities.
(b) Multi-layer fusion extends the idea of residual learning by utilizing skip connections that bypass spatial features between modalities. Thus, lower and higher-level features are fused at different abstraction levels, increasing the network's ability to capture complex cues across modalities. Fusing multi-modal contextual information at multi-layer stages is further facilitated by employing attention mechanisms that bridge early feature extraction and late decision-making.

When working with paired data in medical image segmentation, late fusion is a common approach. In this strategy, individual segmentation branches, each processing a different data modality, are integrated during the decoding stage. This involves mapping all computed feature maps into a single, unified feature space using operations like concatenation, averaging, or weighted voting, followed by subsequent convolutional layers to refine the combined information [142]. Both mid- and late-fusion strategies often lead to improved performance because each modality is fed into a separate network, enabling the model to learn complex and complementary feature information unique to that data type. This advantage, however, comes with a trade-off: using multiple networks significantly increases the memory and computational power required [14]. A significant challenge in medical imaging is the difficulty and high cost of collecting large datasets of paired images. This has led to a growing focus on utilizing unpaired datasets. When working with such data, domain adaptation becomes crucial. This specialized area within transfer learning addresses situations where the categories or labels are consistent across different data domains, but the characteristics of the domains themselves vary. The primary goal is to find a suitable transformation that allows a segmentation model, initially trained on a source domain with labeled data, to perform effectively on a target domain, which might contain unlabeled or differently acquired data. Domain adaptation is categorized by the availability of labeled data in the target domain: it can be supervised (both labeled source and target data are available), semi-supervised (labeled source data is available, with partial labeled target data), or unsupervised (only labeled source data is available, with no labeled target data). Cycle Generative Adversarial Networks (CycleGANs) are frequently employed in unsupervised domain adaptation [10]. As illustrated in Fig. 26, a CycleGAN typically uses two generators: one translates an image from a source domain to a target domain, and the other translates the output back, ensuring that the learned transformations preserve the essential content of the images while adapting their style or characteristics to the target domain.

The inter-modality variance problem has been addressed; we will also discuss the multi-domain segmentation as an intra-modality variance problem. For example, two different CT scanners can vary intensity distributions in their output scans. This is particularly relevant if the training data set prepares the model for generalized evaluations

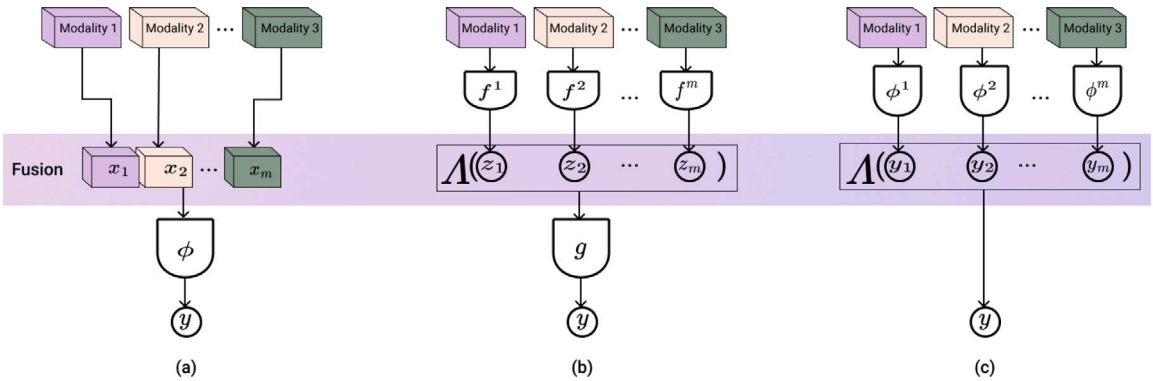

Fig. 25. Cross-modal framework for paired data. (a) Early fusion, (b) mid-fusion, (c) late fusion.

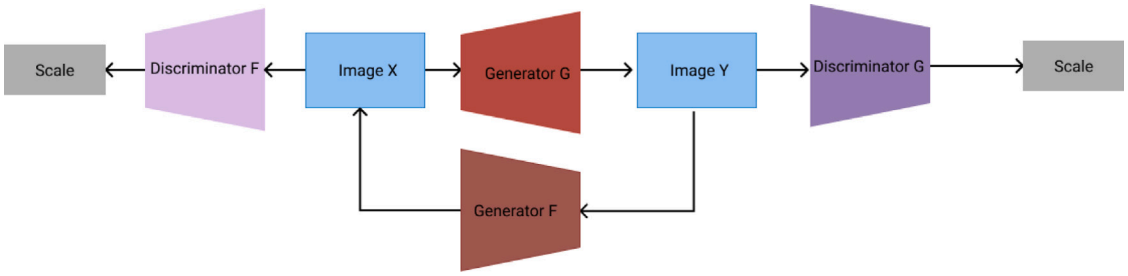

Fig. 26. The Cycle GAN architecture [143].

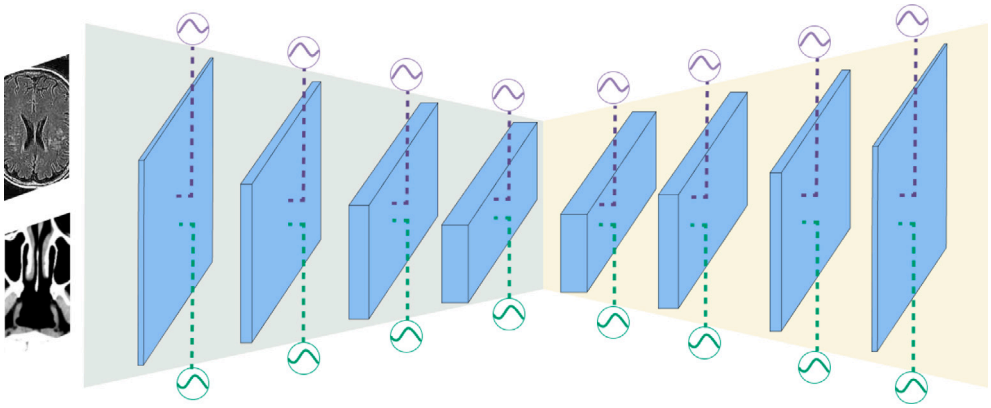

Fig. 27. Shared convolutional kernels and domain specific feature normalization [146].

on different data sets (different instances of the same imaging modality). The underlying assumption is that extraction of robust domain-invariant feature representations is possible if the redundancy between multiple-intensity domains is adequately exploited, enabling the model to perform better than a domain-specific model. Differences in imaging systems, reconstruction settings, and acquisition protocols make multi-domain segmentation essential in real-life medical applications [14]. Attempts to achieve multi-domain segmentation include using adversarial networks to learn domain-invariant features [144] and using transfer learning [145] with a single encoder–decoder segmentation network with shared convolutional kernels but domain-specific feature normalization (Fig. 27).

6.5. Distributed Learning frameworks

Distributed Learning refers to techniques and paradigms that enable deep learning models to be trained across multiple computing nodes or data sources, rather than on a single, centralized dataset. These approaches are critical for unlocking the full potential of deep learning in clinical settings, allowing for robust model development even when data cannot be centrally pooled. Distributed learning in medical image analysis offers several key methodologies, including federated learning for privacy-preserving collaborative model training, transfer learning and domain adaptation for using existing knowledge across different datasets, and knowledge distillation and multi-task learning for efficient model deployment and comprehensive learning from diverse tasks.

Transfer learning is beneficial in a limited datasets context; this strategy entails pre-training the model using existing large-scale datasets before applying it to the problem at hand [4]. Pretrained models include LeNet-5 [147], AlexNet [52], Visual Geometry Group (VGG Net) [54], GoogleNet [56], ResNet [53]. These models are then fine-tuned to MIS by freezing the convolutional base and training some layers only [4]. However, domain adaptation is still required when moving from natural to medical images; furthermore, pretrained models often rely on 2D datasets, and their applicability to 3D data is limited [10].

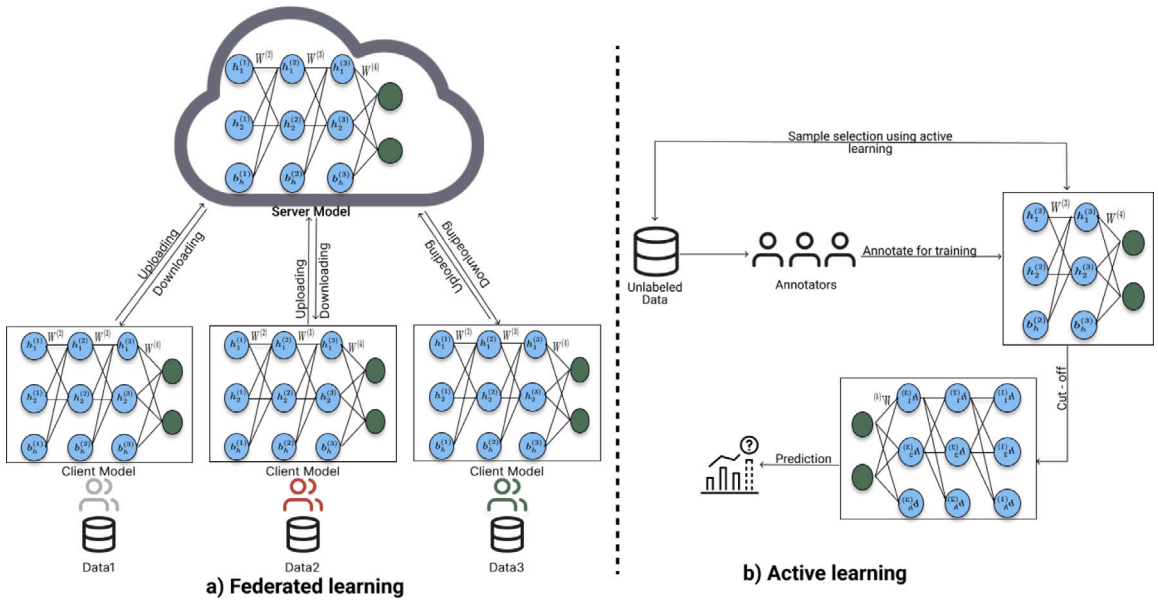

Fig. 28. General framework for (a) federated learning and (b) Active Learning [14].

Training different models on the same data and then averaging their predictions is a feasible but computationally expensive strategy; compressing their combined knowledge into a single model through knowledge distillation is much easier to implement [148]. The idea is to transfer information from a well-trained teacher network to a lightweight and compact student network to improve the performance of the student model. Extensions of this framework include the multiple teacher's single student model (MTSS) [149], which attempted to preserve patient privacy by constructing a multi-organ segmentation student network that learns from multiple pre-trained single-organ segmentation models.

Multi-task learning strives to utilize information shared across two or more auxiliary tasks to improve the handling of each task [150]; viewed as an inductive transfer process, the introduction of an inductive transfer bias allows the prioritization of specific hypotheses over others [151] toward more generalizable solutions. It can leverage a variety of heterogeneous forms of annotations to solve several image-processing tasks at once, using a cascade of task-specific sub-networks or networks with shared encoders and task-specific decoders [152] to benefit from partial parameter sharing between tasks.

Federated learning enables the distributed training of deep models without sharing data between multiple institutions to protect patient privacy. Each institution maintains an individual model that focuses on the local data and requests a global model from a central server to download the global model weights. During training, local model weights are sent to the central server for updating. The central server aggregates the feedback from individual institutions, and the global weights are then updated according to predefined rules based on the varying feedback quality from individual institutions. The model is shown in Fig. 28(a). This is a promising research direction in that it allows for the generation of larger training data sets while maintaining patient privacy [14].

6.6. Active learning

Active learning frameworks offer a promising solution to the challenges of annotating large medical datasets by iteratively selecting the most valuable samples for training deep learning models. This approach not only reduces the annotation burden but also enhances model performance, making it a valuable tool in the development of efficient and accurate medical AI systems [153–155]. Active learning also allows the model to focus on the most challenging samples, thus improving its performance. However, this choice process is challenging and involves uncertainty and representative criteria [156]. The framework is presented in Fig. 28(b).

6.7. Segmentation uncertainty

Various sources of uncertainty impacting network performance can be grouped into epistemic and aleatoric uncertainty. Epistemic (Model) Uncertainty describes parameter uncertainties in the model due to insufficient training. It is reduced by providing more training time and data. Monte-Carlo dropout (or Test Time Dropout) is a stochastic technique that yields an epistemic uncertainty map dependent on dissimilarity in predictions. However, this dropout may negatively affect the training performance itself. Other epistemic uncertainty quantification techniques include deep ensembling, where independent networks are trained, and their predictions are averaged together to obtain uncertainty maps [14].

Aleatoric Uncertainty relates to the inherent uncertainty of the data itself, which can be further subcategorized into homoscedastic uncertainty (constant for all samples in a given set) and heteroscedastic uncertainty (varying among samples). Homoscedastic uncertainty stems from physical properties of the imaging modality, like the variation in positron range and Compton scattering. In contrast, heteroscedastic uncertainty may be due to heterogeneity in annotation quality [157]. Aleatoric uncertainty can be quantified through Test-Time Data Augmentations (TTA) in which multiple forward passes are performed to augmented inputs [158]. This approach enhances model performance and provides a robust measure of uncertainty, making it valuable for various MIS tasks.

Applying CNNs or Transformers for 3D MIS is a computationally expensive process involving many parameters to train [159]. The computational resources available on specific devices may be inadequate for executing complex 3D segmentation tasks. Consequently, there has been a growing interest within the research community in developing lightweight deep learning models capable of performing efficient 3D segmentation under limited computational conditions. Some solutions include depth-separable convolutions [160], which involve replacing a 3D convolution kernel $3 \times 3 \times 3$ with $1 \times 3 \times 3$ intra-slice and $3 \times 1 \times 1$ interslice convolutions, combinations between pointwise, group-wise, and dilated convolutions [161], or channel reduction [162]. It is worth nothing that model size is particularly

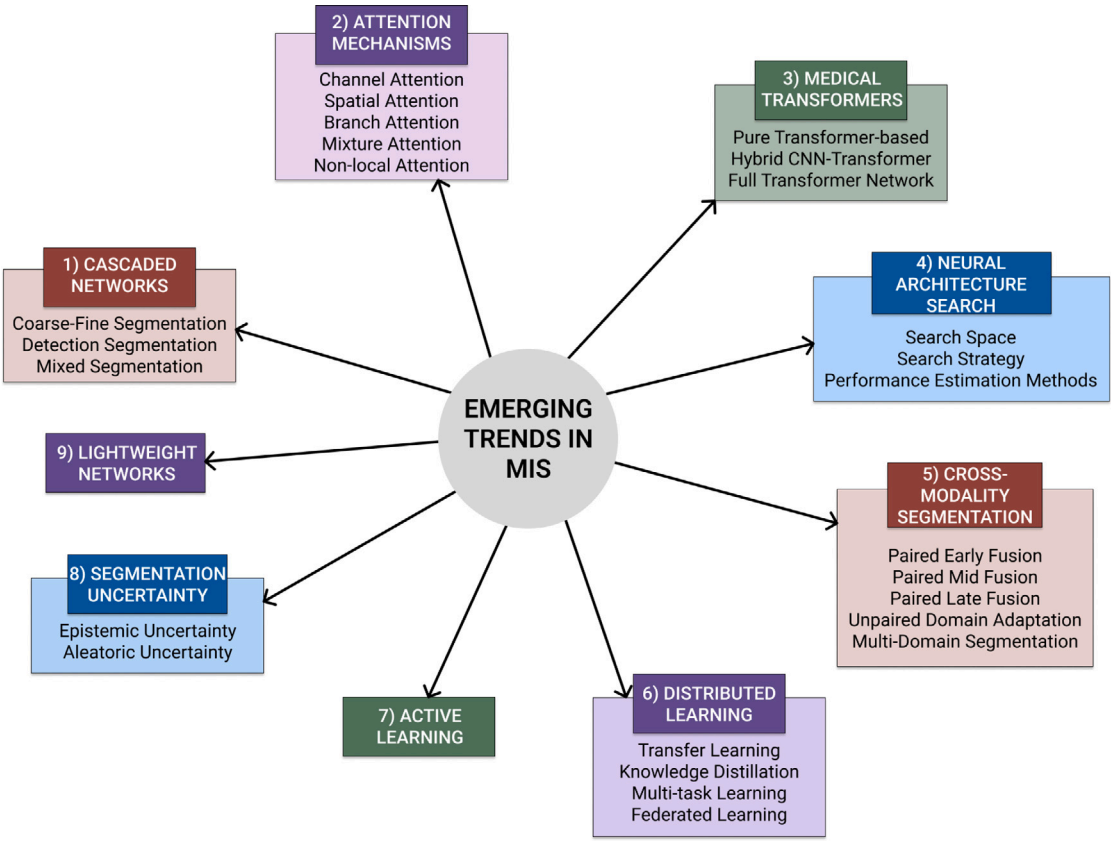

Fig. 29. Summary of Emerging Trends.

relevant in uncertainty estimation because they are often deployed in resource-constrained environments, where uncertainty-aware prediction are crucial for safe decision-making. While heavy-weight models generally offer higher performance due to their extensive parameter counts and computational resources, lightweight models have significantly improved competitive accuracy and efficiency. The trade-off between performance and computational complexity can be effectively managed through innovative architectural designs, attention mechanisms, and efficient feature-processing techniques. Lightweight models like MediLite3DNet, SegFormer3D, and HL-UNet demonstrate that it is possible to achieve high performance in 3D medical image reconstruction without the heavy computational burden typically associated with larger models [163–165]. Fig. 29 presents a summary of emerging trends.

7. Case study: Lumbar spine segmentation

Lumbar spine segmentation has seen significant advancements through various methodologies, particularly deep learning and hybrid approaches, which have demonstrated high accuracy and robustness across different imaging modalities. These advancements are crucial for enhancing clinical diagnostics and treatment planning. Lumbar spine segmentation applications include: detection of spinal anomalies, surgical planning, and treatment monitoring [166–168]. Another application is diagnosis of lumbar deformities and fractures [169]. Lumbar spine segmentation has common challenges as overlapping shadows in X-ray images, unclear boundaries, and inter-patient variability [170]. High similarity between vertebral bones and intervertebral discs in MRI images [171].

7.1. Overview

The vertebral column is essential in supporting the human body and protecting the spinal cord and, subsequently, the central nervous system. The spinal region, located therein, comprises the cervical, thoracic, and lumbar spine [172] and plays a primary role in mobility in the musculoskeletal system, in addition to sustaining organ structure and protecting the body from shock [173]. Defects within this region result in Vertebral Misalignment, the disease responsible for chronic back pain, which is construed as one of the most prevalent medical problems in contemporary society [172]. The slightest damage to this region risks causing immense pain and bodily malfunction to the afflicted patient due to the large concentration of nerves therein [172]. Different types of abnormalities in this region [174], some of which require emergency treatment, like osteoporotic fractures. In contrast, others, like disc degeneration and scoliosis, can be therapeutically treated, though equally painful in their effects on the human body. Biomechanical changes can result in disability and severe discomfort in the short term. Still, they can have far worse long-term complications, such as an eightfold higher mortality rate due to osteoporosis [173]. Yet, it remains underdiagnosed despite its critical nature. To emphasize the impending criticality of the medical problem at hand, we summarize in Table 3 the various diseases that manifest in the human body of a patient suffering from spinal malalignments [172].

In an attempt to diagnose the patient's spine, traditional methods involved inspection tools like magnetic resonance imaging (MRIs), X-rays, computed tomography (CT-scans), and positron imaging tomography (PET) scans [188,189]. However, these modalities heavily depend on radiologists with years of experience and are subject to diagnostic inaccuracies [190]. The challenges to manual segmentation and analysis of medical images of the spine include the time-consuming nature of manual annotation [172] as well as the inaccuracies exacerbated by the small-sized nature of the region of interest and the consequent difficulty faced in bounding it [191]. Furthermore, the inconsistent standards of different radiologists lead to significant differences in segmentation results [166], and the traditional methods of region-based and threshold-based segmentation are limited by the varying influences

Table 3
Manifestations of spinal malalignments.

Disease	Description	Cause	References
Lumbar Spine Stenosis	Narrowing of spine canal	Collisions of tissues located between vertebrae causing inflammation and pain in spinal nerves	Webb et al. [175]; Ross, Jr. and Braunwald [176]; Ghosh et al. [177]
Scoliosis	Spine assumes an S shape or C shape	Exaggerated neurological activity, birth defects	Kumar et al. [178]; Aebi [179]; Horng et al. [180]
Osteoporotic Fractures	Degradation of bone density	Bones weaken over time, most commonly in elderly people aged 80+	Khan et al. [181]; Johnell and Kanis [182]
Thoracolumbar Fractures	Injuries in thoracic and lumbar vertebrae	Hard fall or accident that induces intense impact on the back	Azizi et al. [183]; Raghavendra et al. [174]; McAfee et al. [184]
Degeneration	Weakening of bones	Aging	Fine et al. [185]; Lim et al. [186]; Jebri et al. [187]

of imaging equipment and principles, and the consequent complexity of the shape and content of medical images [192]. Thus, an imminent need for automated segmentation arises [173,193] to facilitate the prompt morphological analysis and effective clinical treatment of these pathologies [166].

7.2. Vertebral segmentation literature review

Automatic vertebral segmentation can be perceived as a pixel-level classification method [166] with diagnostic significance in estimating spinal curvature and recognizing spinal deformities [194] as well as facilitating finite element modeling analysis, biomechanical modeling, and surgical planning for metal implantations. Traditionally, automatic segmentation was achieved using prior-shape models like statistical shape models [195–197], geometric models [198,199], Markov Random Fields (MRF) [200,201] and active contours [202]; all of which essentially revolved around fitting a shape before the spine and distorting it into conformity to the spinal shape [173]. Other models include a priori variational intensity models [203], level sets [204] as well as landmark-framework-based automatic segmentation models [205].

The proliferation of machine learning techniques led to numerous contributions in lumbar spine image segmentation. In [206], vertebral structures were identified using a multi-layer perceptron (MLP) and segmented by deformable registration. In [207], vertebrae were located using random forest regression and then subsequently segmented at the voxel level utilizing a random forest classifier. Authors of Sneath et al. [189] proposed a two-stage decision forest coupled with a morphological image processing technique that facilitates the automatic detection and identification of vertebral bodies in arbitrary field-of-view volume CT scans. In [208], the deformation model was combined with convolutional neural networks (CNNs) to learn spine features and output the probability map of the spine through the CNN, thereby guiding the deformation model to create a boundary for the spine and, therefore, realize the objective of spine segmentation. In [209], the authors combined SVM with the histogram of oriented gradients (HOG) methodology to create tight bounding boxes that encapsulate the region of interest in the upper vertebrae in a mid-sagittal MRI image. Discs are initially segmented based on corresponding axial MRIs calculated from the intersection of the axial slices with the Sagittal, and then leftover discs are segmented using a two-stage classifier. However, in recent years, the explicit modeling of vertebral shape gave way to data-driven learning techniques facilitated by the emergence of more sophisticated spine image datasets that have become publicly available [166]. The most prominent deep-learning-based image segmentation techniques depend on convolutional neural networks (CNNs) or specifically stacked sparse auto-encoders (SSAEs).

CNN-based segmentation architectures include fully convolutional neural networks (FCN) [62], SegNet [59], UNet [70] and 3D Unet [210]. In [211], patch-based segmentation is employed to differentiate between the anterior and posterior parts of the spine. The image input layer receives each pixel in the patch and applies data normalization. The output is then fed into convolutional layers with equal-sized kernels equipped with a trained classification feature. Segmentation training is sped up using batch normalization, the ReLU function, and fully connected layers, which extract the final features. In [212], Segnet is employed, and an encoder–decoder architecture encapsulates a sequence of convolution layers. The encoder selects image features at varying resolutions; the final output is a boundary detail. The convolution layers are equipped with filter banks, and the output is fed into batch normalization followed by a ReLU activation function. Furthermore, max-pooling is used to implement subsampling. The decoder is responsible for upsampling the output signal and restoring it using the max pool layer. The result is then fed into a filter-equipped convolutional layer, followed by batch normalization and ReLU activation again. The final output is applied to a softmax function, and the result is the segmented image. In [213], a CNN was utilized for spine localization and another CNN for spine segmentation. The input to the localization CNN is a 2-D slice of the spine, and they downsample the real spine mask to obtain the localization network's ground truth, using patches to segment the vertebrae one at a time. Authors in [214] used two 3D FCNs for spine localization and segmentation; the spine is localized using a bounding box obtained by regression in the localization network, and voxel-level multiclassification is performed in the segmentation network. In [215], only one 3D FCN is used for spine segmentation; an iterative method is employed to segment the vertebrae sequentially according to the prior rules of their appearance. A memory component is added to the network to ascertain that vertebrae in the current block have been segmented. Finally, a classification component is utilized to label the vertebrae that have been segmented. However, this method struggles to solve the problem of overlapping vertebrae [166].

In [229], a two-stage iterative technique was employed: lower-resolution vertebrae were first sequentially identified and segmented, then low-resolution masks were refined through a CNN. This technique led to a single phase fully convolutional network in [215]. In [230] the authors propose a pixel based segmentation for 2D patch based pixel classifications. In their work, a deep CNN model based on the Active Appearance Model (AAM) is used to aid the initial level of the pipeline by providing a rough bounding area, which is then passed through a modification process from the Atlas-based AAM at the secondary level. Finally, in [231], purposely built fully convolutional networks were employed in a coarse-to-fine segmentation technique that comprised vertebra labeling, spine localization, and vertebrae segmentation. Unet models typically adopt an asymmetrical U-shaped structure [166] that achieves good results in medical segmentation tasks, however, they are limited by their skip layer connections which constrain the encoder and decoder to perform feature fusion only within layers of the same depth. As a result, specific details are lost due to the inability to fuse semantic information with different scales. The structure of the basic Unet is optimized in [232] (Unet ++), which utilizes an efficient collection of Unets of different depths to alleviate the issue of the unknown network depth. Furthermore, Unet++ recreates skip connections to aggregate features of different semantic scales at the decoder subnetworks and offers a pruning method to accelerate inference speed [166]. This is

Table 4
Summary of recent methods for vertebra segmentation and analysis.

Method	Description	Datasets	Applications	Performance metrics
Patch-based Deep Learning (SSAE)	Divides 2D CT slices into patches, uses SSAE for feature extraction, and RUS for data balancing.	VerSe, CSI-Seg, Lumbar CT	Clinical applications	Precision: 89.9%, Recall: 90.2%, Accuracy: 98.9%, F-score: 90.4%, IoU: 82.6%, DC: 90.2% [216]
Coarse-to-Fine Method with HMC	Initial coarse shape followed by HMC segmentation without shape prior.	Standard and non-standard vertebrae datasets	Orthopedics, interventional procedures	Robust to shape and luminance changes [217]
Semi-automatic (Boundary Classification and Mesh Inflation)	Uses boundary classification and mesh inflation for vertebra segmentation.	11 lumbar datasets	Diagnosis of spine pathologies	Detection rate: 93%, DSC: 78% [218]
Machine Learning (SVM)	Uses gradient orientation histograms and SVM for cervical vertebra detection in MR images.	21 T2-weighted MR images	Diagnosis and therapy	Accurate despite severe noise [219]
Deformable Model-based Segmentation	Focuses on thoracic and lumbar vertebrae using a deformable model.	16 patient datasets	Virtual spine straightening	Applicable to scoliosis and bone metastases [220]
Probabilistic Energy Functions	Models intensity, spatial interaction, and shape for segmentation.	Clinical CT images, phantom datasets	BMD measurements	Robust under various noise levels [221]
Edge-mounted Willmore Energy and Probabilistic Model	Combines intensity and shape information for segmentation.	40 clinical CT images, phantom datasets	BMD measurements	Higher accuracy than alternatives [222]
Statistical Shape Decomposition and Conditional Models	Part-based statistical decomposition for detailed segmentation.	30 healthy CT scans, 10 pathological scans	Image-guided interventions, musculoskeletal modeling	Point-to-surface error improvement: 20% (healthy), 17% (pathological) [223]
Iterative Instance Segmentation (FCN)	Uses FCN with memory component for iterative vertebra segmentation and labeling.	Multiple datasets (CT and MR)	Spine analysis, abnormality detection	DSC: 94.9%, Anatomical identification accuracy: 93% [215]
Statistical Shape Models (SSM)	Uses SSM for lumbar vertebra segmentation with B-spline relaxation.	5 patient datasets	Not specified	Global mean DSI: 93.4% [224]
Deep Learning (Res U-Net)	Combines Res U-Net with Otsu's method for feature extraction.	VerSe'20	Preoperative planning	DSC: 87.10% [173]
Interactive Segmentation (Grow-cut)	Semi-automated method using user-defined seed pixels.	23 subject datasets	Diagnosis of herniated discs	Accuracy: 97.22%, Sensitivity: 83.33% [225]
Attention Gate-based Dual-pathway Network (AGNet)	Coarse-to-fine framework with context and edge pathways.	Spine X-ray images, vertebrae dataset	Spinal diagnosis systems	Superior performance compared to state-of-the-art methods [226]
Selective Binary Gaussian Filtering Regularized Level Set	Fully automatic segmentation with morphological operations.	10 trauma patient datasets	Diagnosis, therapy, surgical intervention	DSC: 90.86% (whole spine), 86.08% (thoracic), 95.61% (lumbar) [227]
2D Centroid-detection Guidance Segmentation Network (CD-VerTransUNet)	Utilizes a global information and multitasking approach.	VerSe'20, scoliotic dataset	Diagnosis, treatment planning	DSC: 75.15% (sagittal), 71.16% (coronal) [195]
Atlas-based Segmentation	Registers multiple atlases to target data and use label fusion.	Training dataset from MICCAI workshop	Not specified	Average DICE score: 0.94 [228]

further improved in Unet 3+ [233], which employs deep supervision and full-scale skip connections. To overcome the confusion sparked by similarities between MRI scans of different disks, the authors in [234] used bounding boxes to crop the estimated center, employed zero padding and fed them into the convolutional network. ReLU activation function is used in the 3D convolutional network with the addition of a dropout operation. In [231], a 3D Unet is first used to roughly localize the spine, followed by a heat map regression for spine localization and identification using the Spatial Configuration Net [235], and then finally, a 3D Unet is used for binary segmentation on the identified vertebrae. In [213], patch-based binary segmentation was executed, then heatmaps of vertebrae masks were denoised. In [236] a simple multi-layer perceptron was first trained to regress the localization of the region of the lumbar spine, then a Unet was trained towards multiclass segmentation. The multi-layer perceptron in [213] was then substituted in [214] with two sequential CNNs for multiclass segmentation. A recent study demonstrated that nnUnet can be embedded within a universal model to enable the segmentation of 33 different anatomical structures including vertebrae, pelvic bones, and abdominal organs from CT images [237]. Experimental results using VerSe'19 and VerSe'20 datasets demonstrate high performance segmentation in terms of Dice coefficient and Hausdorff distance as shown in [237] (Table 5).

SSAE is a dual-stage encoder–decoder architecture in which pixel intensities are first encoded in the encoder through low-dimensional features, which are later used to reconstruct the original intensities in the decoder. SSAE is a fully connected network that represents weight features in a single global weight matrix, unlike CNN, which is more oriented towards partial connections and better emphasizes the significance of locality.

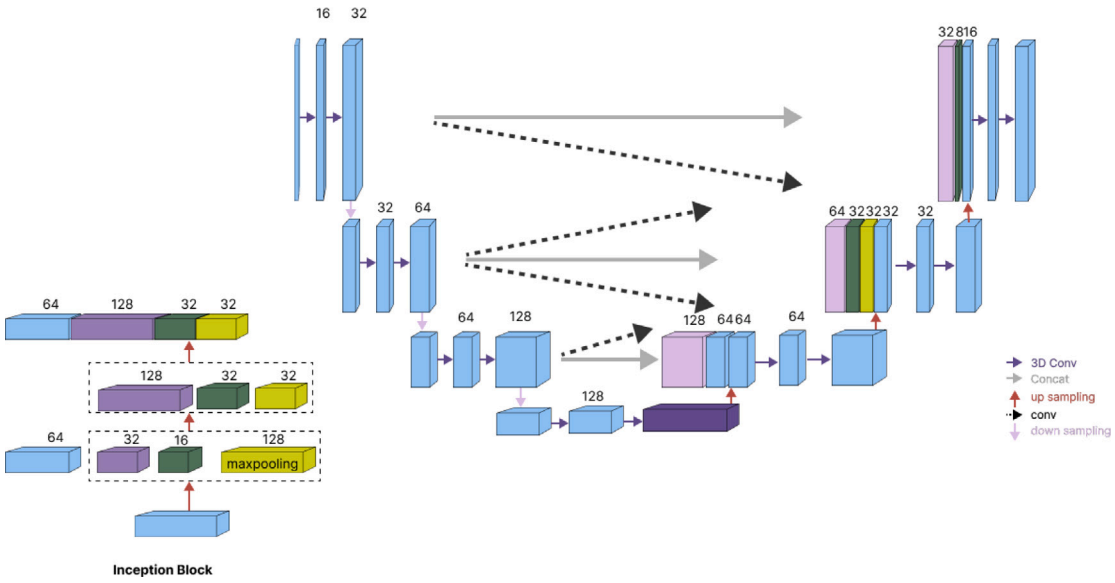

Fig. 30. 3D X Unet architecture in [166].

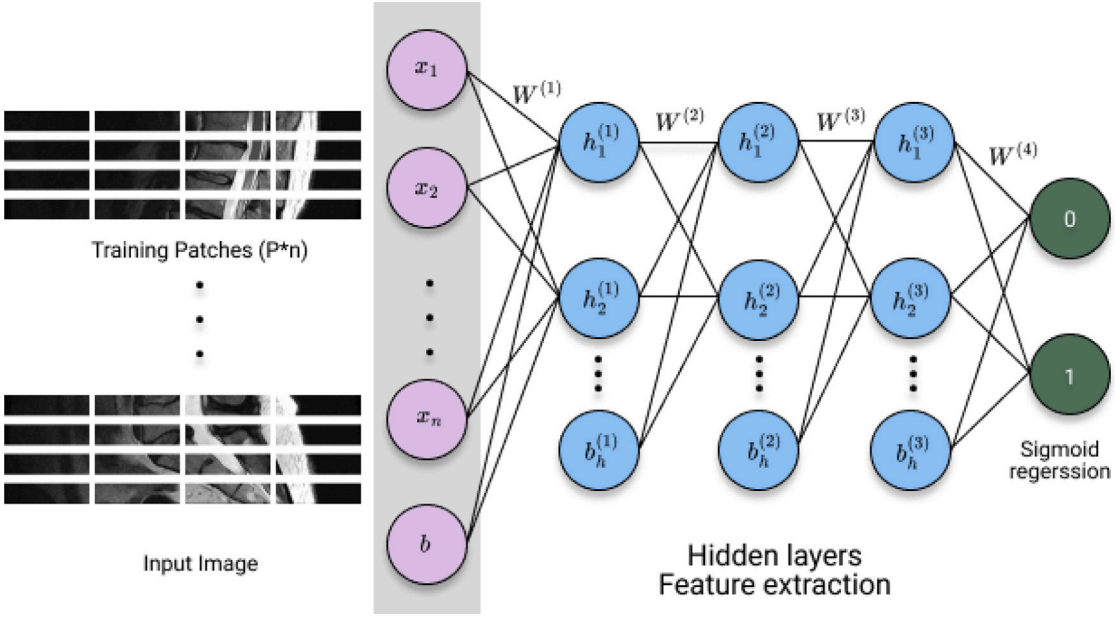

Fig. 31. SSAE architecture in [173].

The first instance of SSAE employment in MIS can be traced back to [238], in which the first deep autoencoder network was developed. In [239], stacked autoencoders were employed in MRIs to identify organs and in [240] a CAD system employed SSAE to identify gastric cancer from breath samples. In [241] an SVM classifier was employed to receive the output from SAE layers that segment stroke lesions, and in [242] SSAE was used to create a model for vertebral segmentation. In [243] the authors achieved automatic vertebrae localization and identification using SSAE and a structured regression forest, and in [244] a SSAE framework was proposed to diagnose Parkinson's disease. SSAE was favored over CNN in [173] owing to its unsupervised extraction of high-level features from the bottom up, leading to more efficient representations, precise patch classifications, and a more robust CT vertebral segmentation. They used a PE module to extract overlapping image patches and their labeling with predefined pixel ratios, whereas a RUS module was employed to address the class imbalance problem. Figs. 30 and 31 represent sample Unet-based [166] and SSAE-based [173] architectures, respectively. A summary of these methods is listed in Table 4.

8. Conclusion

Medical image segmentation remains a profoundly active and vital research area, driven by its direct impact on improving patient diagnostics and therapeutic outcomes. As we have explored throughout this survey, the journey of the transformation of advanced computational tools. This paper has served as a comprehensive guide, thoroughly outlining the evolution of methodologies, from traditional approaches like thresholding and region-based methods to the latest deep learning paradigms such as U-Nets and Transformers. We aimed to equip future researchers with a robust toolkit for developing more sophisticated

and clinically relevant models that adapt to the unique characteristics of diverse imaging modalities like X-ray, CT, and MRI. The ongoing challenges in medical image segmentation, including data limitations, inherent noise and artifacts, and the exploration for universally applicable methods, underscore the extended potential for further innovation. Our survey particularly highlights the importance of distributed learning, especially federated learning, as a critical avenue for future exploration, given the increasing emphasis on patient data privacy and secure multi-institutional collaboration. Furthermore, the rigorous assessment of uncertainty quantification methods alongside lightweight models ensures that deep learning solutions are not only accurate but also trustworthy, interpretable, and deployable in the varied, resource-constrained clinical settings where they are most needed.

CRediT authorship contribution statement

Ahmed Kabil: Writing – original draft, Resources, Formal analysis. **Ghada Khoriba:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Mina Yousef:** Writing – review & editing, Visualization, Data curation. **Essam A. Rashed:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by JST, Japan, PRESTO Grant Number JPMJPR23P7, Japan.

References

[1] S. Hussain, I. Mubeen, N. Ullah, S.S.U.D. Shah, B.A. Khan, M. Zahoor, R. Ullah, F.A. Khan, M.A. Sultan, Modern diagnostic imaging technique applications and risk factors in the medical field: A review, BioMed Res. Int. 2022 (1) (2022) 5164970, http://dx.doi.org/10.1155/2022/5164970.

[2] S. Jardim, J. António, C. Mora, Image thresholding approaches for medical image segmentation-short literature review, Procedia Comput. Sci. 219 (2023) 1485–1492, http://dx.doi.org/10.1016/j.procs.2023.01.439.

[3] L. Li, V.A. Zimmer, J.A. Schnabel, X. Zhuang, Medical image analysis on left atrial LGE MRI for atrial fibrillation studies: A review, Med. Image Anal. 77 (2022) 102360, http://dx.doi.org/10.1016/j.media.2022.102360.

[4] B. Sistaninejhad, H. Rasi, P. Nayeri, A review paper about deep learning for medical image analysis, Comput. Math. Methods Med. 2023 (1) (2023) 7091301, http://dx.doi.org/10.1155/2023/7091301.

[5] G. Du, X. Cao, J. Liang, X. Chen, Y. Zhan, Medical image segmentation based on U-Net: A review, J. Imaging Sci. Technol. 64 (2) (2020) http://dx.doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508.

[6] S. Babu, S. Pandey, K. Palle, P.V. Prasad, B. Mallala, S. Pund, Adaptive medical image segmentation using deep convolutional neural networks, 2023, pp. 15–21, http://dx.doi.org/10.1109/ICPSITIAGS59213.2023.10527488.

[7] K. Ramesh, G.K. Kumar, K. Swapna, D. Datta, S.S. Rajest, A review of medical image segmentation algorithms, EAI Endorsed Trans. Pervasive Heal. Technol. 7 (27) (2021) http://dx.doi.org/10.4108/eai.12-4-2021.169184, e6–e6.

[8] S.K.M.S. Islam, M.A.A. Nasim, I. Hossain, D.M.A. Ullah, D.K.D. Gupta, M.M.H. Bhuiyan, Introduction of medical imaging modalities, in: B. Zheng, S. Andrei, M.K. Sarker, K.D. Gupta (Eds.), Data Driven Approaches on Medical Imaging, Springer Nature Switzerland, Cham, 2023, pp. 1–25, http://dx.doi.org/10.1007/978-3-031-47772-0_1.

[9] K.K. Shung, M. Smith, B.M. Tsui, Principles of Medical Imaging, Academic Press, 2012.

[10] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, A.K. Nandi, Medical image segmentation using deep learning: A survey, IET Image Process. 16 (5) (2022) 1243–1267, http://dx.doi.org/10.1049/ipr2.12419.

[11] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, F. Liu, Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches, Bioengineering 11 (10) (2024) http://dx.doi.org/10.3390/bioengineering11101034.

[12] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, C. Jin, Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation, Comput. Biol. Med. (2023) 107840, http://dx.doi.org/10.1016/j.compbiomed.2023.107840.

[13] J. Shao, S. Chen, J. Zhou, H. Zhu, Z. Wang, M. Brown, Application of U-Net and optimized clustering in medical image segmentation: A review, CMES Comput. Model. Eng. Sci. 136 (3) (2023) http://dx.doi.org/10.32604/cmes.2023.025499.

[14] P.H. Conze, G. Andrade-Miranda, V.K. Singh, V. Jaouen, D. Visvikis, Current and emerging trends in medical image segmentation with deep learning, IEEE Trans. Radiat. Plasma Med. Sci. 7 (6) (2023) 545–569, http://dx.doi.org/10.1109/TRPMS.2023.3265863.

[15] N. Siddique, S. Paheding, C.P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, IEEE Access 9 (2021) 82031–82057, http://dx.doi.org/10.1109/ACCESS.2021.3086020.

[16] X. Liu, L. Song, S. Liu, Y. Zhang, A review of deep-learning-based medical image segmentation methods, Sustainability 13 (3) (2021) 1224, http://dx.doi.org/10.3390/su13031224.

[17] R. Archana, P.S.E. Jeevaraj, Deep learning models for digital image processing: a review, Artif. Intell. Rev. 57 (1) (2024) 11, http://dx.doi.org/10.1007/s10462-023-10631-z.

[18] E.R. Davies, Machine Vision: Theory, Algorithms, Practicalities, Elsevier, 2004.

[19] N. Otsu, et al., A threshold selection method from gray-level histograms, Automatica 11 (285–296) (1975) 23–27.

[20] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognit. 19 (1) (1986) 41–47, http://dx.doi.org/10.1016/0031-3203(86)90030-0.

[21] J. Kapur, P. Sahoo, A. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Comput. Vis. Graph. Image Process. 29 (3) (1985) 273–285, http://dx.doi.org/10.1016/0734-189X(85)90125-2.

[22] J. Rogowska, Overview and fundamentals of medical image segmentation, in: Handbook of Medical Imaging, Processing and Analysis, Academic Press London, 2000, pp. 69–85, http://dx.doi.org/10.1016/B978-012077790-7/50009-6.

[23] K. Schober, J. Prestin, S.A. Stasyuk, Edge detection with trigonometric polynomial shearlets, Adv. Comput. Math. 47 (2021) 1–41, http://dx.doi.org/10.1007/s10444-020-09838-3.

[24] S. Archa, C.S. Kumar, Segmentation of brain tumor in MRI images using CNN with edge detection, in: 2018 International Conference on Emerging Trends and Innovations in Engineering and Technological Research, ICETIETR, IEEE, 2018, pp. 1–4, http://dx.doi.org/10.1109/ICETIETR.2018.8529081.

[25] X. Hua, J. Qian, H. Zhao, L. Liu, L. Liu, Y. Wu, Automatic intestinal canal Segmentation Based Region growing with multi-scale entropy, in: 2018 IEEE 3rd International Conference on Image, Vision and Computing, ICIVC, 2018, pp. 273–277, http://dx.doi.org/10.1109/ICIVC.2018.8492854.

[26] R. Ranjbarzadeh, S. Jafarzadeh Ghoushchi, N. Tataei Sarshar, E.B. Tirkolaee, S.S. Ali, T. Kumar, M. Bendechache, ME-CCNN: Multi-encoded images and a cascade convolutional neural network for breast tumor segmentation and recognition, Artif. Intell. Rev. 56 (9) (2023) 10099–10136, http://dx.doi.org/10.1007/S10462-023-10426-2.

[27] T. Pavlidis, Algorithms for Graphics and Image Processing, Springer Science & Business Media, 2012.

[28] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, S. Qiu, C.T. Lin, A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation, IEEE/ACM Trans. Comput. Biology Bioinform. 18 (1) (2021) 40–52, http://dx.doi.org/10.1109/TCBB.2019.2963873.

[29] T. Abhiraj, K. Srilakshmi, K. Jayaraman, S. Jayaraman, Enhanced football game optimization-based K-means clustering for multi-level segmentation of medical images, Prog. Artif. Intell. 10 (2021) 517–528, http://dx.doi.org/10.1007/s13748-021-00251-5.

[30] R. Agrawal, M. Sharma, B.K. Singh, Segmentation of brain lesions in MRI and CT scan images: a hybrid approach using k-means clustering and image morphology, J. Inst. Eng. (India): Ser. B 99 (2018) 173–180, http://dx.doi.org/10.1007/s40031-018-0314-z.

[31] A. Husein, M. Harahap, S. Aisyah, W. Purba, A. Muhazir, The implementation of two stages clustering (k-means clustering and adaptive neuro fuzzy inference system) for prediction of medicine need based on medical data, J. Phys.: Conf. Ser. 978 (1) (2018) 012019, http://dx.doi.org/10.1088/1742-6596/978/1/012019.

[32] D.M. Kumar, D. Satyanarayana, M.G. Prasad, An improved gabor wavelet transform and rough K-means clustering algorithm for MRI brain tumor image segmentation, Multimedia Tools Appl. 80 (5) (2021) 6939–6957, http://dx.doi.org/10.1007/s11042-020-09635-6.

[33] I. Mehidi, D.E. Chouaib Belkhiat, D. Jabri, An improved clustering method based on K-Means algorithm for MRI brain tumor segmentation, in: 2019 6th International Conference on Image and Signal Processing and their Applications, ISPA, 2019, pp. 1–6, http://dx.doi.org/10.1109/ISPA48434.2019.8966891.

[34] X. Chen, L. Pan, A survey of graph cuts/graph search based medical image segmentation, IEEE Rev. Biomed. Eng. 11 (2018) 112–124, http://dx.doi.org/10.1109/RBME.2018.2798701.

[35] C. Rother, V. Kolmogorov, A. Blake, "GrabCut" interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (3) (2004) 309–314, http://dx.doi.org/10.1145/1015706.1015720.

[36] L. Fang, X. Liang, C. Xu, Q. Wang, Image segmentation using a novel dual active contour model, Multimedia Tools Appl. 83 (2) (2024) 3707–3724, http://dx.doi.org/10.1007/s11042-023-15472-0.

[37] T.F. Cootes, C.J. Taylor, Active shape models—'smart snakes', in: BMVC92: Proceedings of the British Machine Vision Conference, Organised By the British Machine Vision Association 22–24 September 1992 Leeds, Springer, 1992, pp. 266–275, http://dx.doi.org/10.1007/978-1-4471-3201-1_28.

[38] T. Cootes, G. Edwards, C. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685, http://dx.doi.org/10.1109/34.927467.

[39] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations, J. Comput. Phys. 79 (1) (1988) 12–49, http://dx.doi.org/10.1016/0021-9991(88)90002-2.

[40] K.D. Shah, D.K. Patel, M.P. Thaker, H.A. Patel, M.J. Saikia, B.J. Ranger, EMED-unet: An efficient multi-encoder-decoder based unet for medical image segmentation, IEEE Access 11 (2023) 95253–95266, http://dx.doi.org/10.1109/ACCESS.2023.3309158.

[41] Z. Chen, Medical image segmentation based on U-net, 2547, (1) 2023, http://dx.doi.org/10.1088/1742-6596/2547/1/012010,

[42] N. Kavyasri, V. Akkala, Srivani, S. Pabboju, Brain tumour detection using deep learning: A review, in: 15th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2024, vol. 2, 2024, pp. 2970–2975.

[43] M. Juneja, N. Aggarwal, S.K. Saini, S. Pathak, M. Kaur, M. Jaiswal, A comprehensive review on artificial intelligence-driven preprocessing, segmentation, and classification techniques for precision furcation analysis in radiographic images, Multimedia Tools Appl. (2024) http://dx.doi.org/10.1007/s11042-024-19920-3.

[44] T. Arumaiththurai, B. Mayurathan, The effect of deep learning and machine learning approaches for brain tumor recognition, in: 2021 10th International Conference on Information and Automation for Sustainability, ICIAfS 2021, 2021, pp. 185–190, http://dx.doi.org/10.1109/ICIAfS52090.2021.9605909.

[45] Y. Sneha, Y.M. Roopa, P. Sawant, M.V. Rao, D.L. Padmaja, R. Lalitha, Advancements in brain tumor detection using machine learning applications from MRI image analysis, in: 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2023 - Proceedings, 2023, pp. 809–814, http://dx.doi.org/10.1109/I-SMAC58438.2023.10290231.

[46] G. Chassagnon, M. Vakalopoulou, N. Paragios, M.P. Revel, Artificial intelligence applications for thoracic imaging, Eur. J. Radiol. 123 (2020) http://dx.doi.org/10.1016/j.ejrad.2019.108774.

[47] S. Cai, Y. Xiao, Y. Wang, Two-dimensional medical image segmentation based on U-shaped structure, Int. J. Imaging Syst. Technol. 34 (1) (2024) http://dx.doi.org/10.1002/ima.23023.

[48] F.P. An, J.E. Liu, Medical image segmentation algorithm based on optimized convolutional neural network-adaptive dropout depth calculation, Complexity 2020 (2020) http://dx.doi.org/10.1155/2020/1645479.

[49] A. Lou, S. Guan, M. Loew, DC-UNet: Rethinking the U-net architecture with dual channel efficient CNN for medical image segmentation, 11596, 2021, http://dx.doi.org/10.1117/12.2582338.

[50] E.R. Nafchi, P. Fadavi, S. Amiri, S. Cheraghi, M. Garousi, M. Nabavi, I. Daneshi, M. Gomar, M. Molaie, A. Nouraeinejad, Radiomics model based on computed tomography images for prediction of radiation-induced optic neuropathy following radiotherapy of brain and head and neck tumors, Heliyon 11 (1) (2025).

[51] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al., Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377, http://dx.doi.org/10.1016/j.patcog.2017.10.013.

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

[53] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

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

[55] Z. Qiu, T. Yao, T. Mei, Learning spatio-temporal representation with pseudo-3D residual networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 5534–5542, http://dx.doi.org/10.1109/ICCV.2017.590.

[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 1–9, http://dx.doi.org/10.1109/CVPR.2015.7298594.

[57] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M.S. Nobile, C. Ferretti, D. Besozzi, et al., USE-Net: Incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets, Neurocomputing 365 (2019) 31–43, http://dx.doi.org/10.1016/j.neucom.2019.07.006.

[58] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. 42 (8) (2020) 2011–2023, http://dx.doi.org/10.1109/TPAMI.2019.2913372.

[59] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495, http://dx.doi.org/10.1109/TPAMI.2016.2644615.

[60] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890, http://dx.doi.org/10.1109/CVPR.2017.660.

[61] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. 28 (2015).

[62] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3431–3440, http://dx.doi.org/10.1109/CVPR.2015.7298965.

[63] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected CRFs, 2014, http://dx.doi.org/10.48550/arXiv.1412.7062, arXiv preprint arXiv:1412.7062.

[64] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, http://dx.doi.org/10.1109/TPAMI.2017.2699184.

[65] X. Zhou, R. Takayama, S. Wang, T. Hara, H. Fujita, Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method, Med. Phys. 44 (10) (2017) 5221–5233, http://dx.doi.org/10.1002/mp.12480.

[66] X.Y. Zhou, M. Shen, C. Riga, G.Z. Yang, S.L. Lee, Focal fcn: Towards small object segmentation with limited training data, 2017, http://dx.doi.org/10.48550/arXiv.1711.01506, arXiv preprint arXiv:1711.01506.

[67] M. Baldeon Calisto, S.K. Lai-Yuen, AdaEn-Net: An ensemble of adaptive 2D–3D fully convolutional networks for medical image segmentation, Neural Netw. 126 (2020) 76–94, http://dx.doi.org/10.1016/j.neunet.2020.03.007.

[68] C. Li, J. Ye, J. He, S. Wang, L. Gu, Y. Qiao, Collaborative multi-view convolutions with gating for accurate and fast volumetric medical image segmentation, in: 2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI, 2021, pp. 571–574, http://dx.doi.org/10.1109/ISBI48211.2021.9433787.

[69] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, UNETR: Transformers for 3D medical image segmentation, in: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2022, pp. 1748–1758, http://dx.doi.org/10.1109/WACV51458.2022.00181.

[70] O. Ronneberger, P. Fischer, T. Brox, 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, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[71] R. Almajalid, J. Shan, M. Zhang, G. Stonis, M. Zhang, Knee bone segmentation on three-dimensional MRI, in: 2019 18th IEEE International Conference on Machine Learning and Applications, ICMLA, 2019, pp. 1725–1730, http://dx.doi.org/10.1109/ICMLA.2019.00280.

[72] S. Jia, A. Despinasse, Z. Wang, H. Delingette, X. Pennec, P. Jaïs, H. Cochet, M. Sermesant, Automatically segmenting the left atrium from cardiac images using successive 3D U-nets and a contour loss, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11395 LNCS, 2019, pp. 221–229, http://dx.doi.org/10.1007/978-3-030-12029-0_24.

[73] Y. Chai, H. Li, Segmentation of liver and its tumor based on U-net, in: 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference, IAEAC, 5, 2021, pp. 208–211, http://dx.doi.org/10.1109/IAEAC50856.2021.9391040.

[74] Z. Yang, A novel brain image segmentation method using an improved 3D U-net model, Sci. Program. 2021 (2021) http://dx.doi.org/10.1155/2021/4801077.

[75] X. Guo, H. Yang, H. Jiang, Improved medical image segmentation method and three-dimensional reconstruction based on 3D-unet, in: 2024 2nd International Conference on Signal Processing and Intelligent Computing, SPIC, IEEE, 2024, pp. 881–885, http://dx.doi.org/10.1109/SPIC62469.2024.10691548.

[76] O. Kemassi, O. Maamri, K. Bouanane, O. Kriker, Dilated convolutions based 3D U-net for multi-modal brain image segmentation, Lecture Notes in Networks and Systems, vol. 413 LNNS (2022) 428–436, http://dx.doi.org/10.1007/978-3-030-96311-8_39.

[77] H. Lu, Y. She, J. Tie, S. Xu, Half-UNet: A simplified U-net architecture for medical image segmentation, Front. Neuroinformatics 16 (2022) http://dx.doi.org/10.3389/fninf.2022.911679.

[78] X. Dong, X. Mao, J. Yao, A novel cardiac image segmentation method using an optimized 3D U-Net model, J. Mech. Med. Biology 23 (09) (2023) 2340102, http://dx.doi.org/10.1142/S0219519423401024.

[79] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention U-Net: Learning where to look for the pancreas, 2018, http://dx.doi.org/10.48550/arXiv.1804.03999, arXiv preprint arXiv:1804.03999.

[80] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207, http://dx.doi.org/10.1016/j.media.2019.01.012.

[81] M. Liang, X. Hu, Recurrent convolutional neural network for object recognition, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3367–3375, http://dx.doi.org/10.1109/CVPR.2015.7298958.

[82] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 2261–2269, http://dx.doi.org/10.1109/CVPR.2017.243.

[83] M.K. Abd-Ellah, A.A. Khalaf, A.I. Awad, H.F. Hamed, TPUAR-Net: Two parallel U-net with asymmetric residual-based deep convolutional neural network for brain tumor segmentation, in: Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, on, Canada, August 27–29, 2019, Proceedings, Part II 16, Springer, 2019, pp. 106–116, http://dx.doi.org/10.1007/978-3-030-27272-2_9.

[84] C. hu, G. Kang, B. Hou, Y. Ma, F. Labeau, Z. Su, Acu-Net: A 3D attention context U-Net for multiple sclerosis lesion segmentation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2020, pp. 1384–1388, http://dx.doi.org/10.1109/ICASSP40776.2020.9054616.

[85] T. Fan, G. Wang, X. Wang, Y. Li, H. Wang, MSN-Net: A multi-scale context nested U-Net for liver segmentation, Signal, Image Video Process. 15 (2021) 1089–1097, http://dx.doi.org/10.1007/s11760-020-01835-9.

[86] S. Leclerc, E. Smistad, T. Grenier, C. Lartizien, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A. Rye Berg, P.M. Jodoin, L. Lovstakken, O. Bernard, RU-Net: A refining segmentation network for 2D echocardiography, in: 2019 IEEE International Ultrasonics Symposium, IUS, 2019, pp. 1160–1163, http://dx.doi.org/10.1109/ULTSYM.2019.8926158.

[87] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211, http://dx.doi.org/10.1038/s41592-020-01008-z.

[88] B. Sabrowsky-Hirsch, S. Thumfart, R. Hofer, W. Fenz, A content-driven architecture for medical image segmentation, in: Proceedings of the 6th International Conference on Communication and Information Processing, 2020, pp. 89–96, http://dx.doi.org/10.1145/3442555.3442570.

[89] Y. Zhao, C. Yang, A. Schweidtmann, Q. Tao, Efficient Bayesian uncertainty estimation for nnu-net, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13438 LNCS, 2022, pp. 535–544, http://dx.doi.org/10.1007/978-3-031-16452-1_51.

[90] C.M. Seibold, S. Reiß, J. Kleesiek, R. Stiefelhagen, Reference-guided pseudo-label generation for medical semantic segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, (2) 2022, pp. 2171–2179, http://dx.doi.org/10.1609/aaai.v36i2.20114.

[91] D.H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation Learning, ICML, vol. 3, (2) Atlanta, 2013, p. 896.

[92] K. Han, L. Liu, Y. Song, Y. Liu, C. Qiu, Y. Tang, Q. Teng, Z. Liu, An effective semi-supervised approach for liver CT image segmentation, IEEE J. Biomed. Heal. Informatics 26 (8) (2022) 3999–4007, http://dx.doi.org/10.1109/JBHI.2022.3167384.

[93] R. Wang, Y. Wu, H. Chen, L. Wang, D. Meng, Neighbor matching for semi-supervised learning, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, Springer, 2021, pp. 439–449, http://dx.doi.org/10.1007/978-3-030-87196-3_41.

[94] M. Sajjadi, M. Javanmardi, T. Tasdizen, Regularization with stochastic transformations and perturbations for deep semi-supervised learning, Adv. Neural Inf. Process. Syst. 29 (2016).

[95] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, 2016, http://dx.doi.org/10.48550/arXiv.1610.02242, arXiv preprint arXiv:1610.02242.

[96] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 29 (2016).

[97] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, 1998, pp. 92–100, http://dx.doi.org/10.1145/279943.279962.

[98] X. Luo, M. Hu, T. Song, G. Wang, S. Zhang, Semi-supervised medical image segmentation via cross teaching between CNN and transformer, in: International Conference on Medical Imaging with Deep Learning, PMLR, 2022, pp. 820–833.

[99] A. Boutillon, B. Borotikar, V. Burdin, P.H. Conze, Multi-structure bone segmentation in pediatric MR images with combined regularization from shape priors and adversarial network, Artif. Intell. Med. 132 (2022) 102364, http://dx.doi.org/10.1016/j.artmed.2022.102364.

[100] H. Ravishankar, R. Venkataramani, S. Thiruvenkadam, P. Sudhakar, V. Vaidya, Learning and incorporating shape models for semantic segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 203–211, http://dx.doi.org/10.1007/978-3-319-66182-7_24.

[101] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S.A. Cook, A. de Marvao, T. Dawes, D.P. O'Regan, B. Kainz, B. Glocker, D. Rueckert, Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation, IEEE Trans. Med. Imaging 37 (2) (2018) 384–395, http://dx.doi.org/10.1109/TMI.2017.2743464.

[102] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger (Eds.), in: Advances in Neural Information Processing Systems, vol. 27, Curran Associates, Inc., 2014, pp. 1–9.

[103] P. Luc, C. Couprie, S. Chintala, J. Verbeek, Semantic segmentation using adversarial networks, 2016, http://dx.doi.org/10.48550/arXiv.1611.08408, arXiv preprint arXiv:1611.08408.

[104] Y. Xue, T. Xu, H. Zhang, L.R. Long, X. Huang, Segan: Adversarial network with multi-scale L_1 loss for medical image segmentation, Neuroinformatics 16 (2018) 383–392, http://dx.doi.org/10.1007/s12021-018-9377-x.

[105] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, E.P. Xing, SCAN: Structure correcting adversarial network for organ segmentation in chest X-rays, in: International Workshop on Deep Learning in Medical Image Analysis, Springer, 2018, pp. 263–273, http://dx.doi.org/10.1007/978-3-030-00889-5_30.

[106] N. Khosravan, A. Mortazi, M. Wallace, U. Bagci, PAN: Projective adversarial network for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22, Springer, 2019, pp. 68–76, http://dx.doi.org/10.1007/978-3-030-32226-7_8.

[107] Q. Chang, H. Qu, Y. Zhang, M. Sabuncu, C. Chen, T. Zhang, D.N. Metaxas, Synthetic learning: Learn from distributed asynchronized discriminator GAN without sharing medical image data, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 13853–13863, http://dx.doi.org/10.1109/CVPR42600.2020.01387.

[108] N. Bayramoglu, M. Kaakinen, L. Eklund, J. Heikkila, Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 64–71, http://dx.doi.org/10.1109/ICCVW.2017.15.

[109] V.K. Singh, H.A. Rashwan, S. Romani, F. Akram, N. Pandey, M.M.K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, et al., Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network, Expert Syst. Appl. 139 (2020) 112855, http://dx.doi.org/10.1016/j.eswa.2019.112855.

[110] J.T. Guibas, T.S. Virdi, P.S. Li, Synthetic medical images from dual generative adversarial networks, 2017, http://dx.doi.org/10.48550/arXiv.1709.01872, arXiv preprint arXiv:1709.01872.

[111] M. Zhao, L. Wang, J. Chen, D. Nie, Y. Cong, S. Ahmad, A. Ho, P. Yuan, S.H. Fung, H.H. Deng, et al., Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11, Springer, 2018, pp. 720–727, http://dx.doi.org/10.1007/978-3-030-00937-3_82.

[112] A.K. Mondal, J. Dolz, C. Desrosiers, Few-shot 3D multi-modal medical image segmentation using generative adversarial learning, 2018, http://dx.doi.org/10.48550/arXiv.1810.12241, arXiv preprint arXiv:1810.12241.

[113] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D.P. Hughes, D.Z. Chen, Deep adversarial networks for biomedical image segmentation utilizing unannotated images, in: Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20, Springer, 2017, pp. 408–416, http://dx.doi.org/10.1007/978-3-319-66179-7_47.

[114] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, http://dx.doi.org/10.48550/arXiv.1411.1784, arXiv preprint arXiv:1411.1784.

[115] C. You, W. Dai, Y. Min, F. Liu, D.A. Clifton, S.K. Zhou, L. Staib, J.S. Duncan, Rethinking semi-supervised medical image segmentation: A variance-reduction perspective, 36, 2023,

[116] H. Fan, J. Cao, X. Chen, S. Lin, K. Polat, J. Zhou, Own-background contrastive learning guided by pseudo-label for semi-supervised medical image segmentation, Appl. Soft Comput. 171 (2025) http://dx.doi.org/10.1016/j.asoc.2025.112749.

[117] X. Hu, D. Zeng, X. Xu, Y. Shi, Semi-supervised contrastive learning for label-efficient medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, Springer, 2021, pp. 481–490, http://dx.doi.org/10.1007/978-3-030-87196-3_45.

[118] S.M. Hooper, S. Wu, R.H. Davies, A. Bhuva, E.B. Schelbert, J.C. Moon, P. Kellman, H. Xue, C. Langlotz, C. Ré, Evaluating semi-supervision methods for medical image segmentation: applications in cardiac magnetic resonance imaging, J. Med. Imaging 10 (2) (2023) http://dx.doi.org/10.1117/1.JMI.10.2.024007, 024007–024007.

[119] J. Hou, X. Ding, J.D. Deng, Semi-supervised semantic segmentation of vessel images using leaking perturbations, in: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2022, pp. 1769–1778, http://dx.doi.org/10.1109/WACV51458.2022.00183.

[120] Z. Ren, R. Yeh, A. Schwing, Not all unlabeled data are equal: Learning to weight data in semi-supervised learning, Adv. Neural Inf. Process. Syst. 33 (2020) 21786–21797.

[121] G. Zhang, Z. Yang, B. Huo, S. Chai, S. Jiang, Automatic segmentation of organs at risk and tumors in CT images of lung cancer from partially labelled datasets with a semi-supervised conditional nnU-Net, Comput. Methods Programs Biomed. 211 (2021) 106419, http://dx.doi.org/10.1016/j.cmpb.2021.106419.

[122] K. Zhang, X. Zhuang, Cyclemix: A holistic strategy for medical image segmentation from scribble supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11656–11665, http://dx.doi.org/10.1109/CVPR52688.2022.01136.

[123] Z. Xu, D. Lu, Y. Wang, J. Luo, J. Jayender, K. Ma, Y. Zheng, X. Li, Noisy labels are treasure: mean-teacher-assisted confident learning for hepatic vessel segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, Springer, 2021, pp. 3–13, http://dx.doi.org/10.1007/978-3-030-87193-2_1.

[124] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.Y. Lo, P. Dollár, R. Girshick, Segment anything, in: 2023 IEEE/CVF International Conference on Computer Vision, ICCV, 2023, pp. 3992–4003, http://dx.doi.org/10.1109/ICCV51070.2023.00371.

[125] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2980–2988, http://dx.doi.org/10.1109/ICCV.2017.322.

[126] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, 2020, http://dx.doi.org/10.48550/arXiv.2004.10934, arXiv preprint arXiv:2004.10934.

[127] A. Iantsen, D. Visvikis, M. Hatt, Squeeze-and-excitation normalization for automated delineation of head and neck primary tumors in combined PET and CT images, in: Head and Neck Tumor Segmentation: First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1, Springer, 2021, pp. 37–43, http://dx.doi.org/10.1007/978-3-030-67194-5_4.

[128] H. Seshimo, E.A. Rashed, Segmentation of low-grade brain tumors using mutual attention multimodal MRI, Sensors 24 (23) (2024) http://dx.doi.org/10.3390/s24237576.

[129] Z. Wang, N. Zou, D. Shen, S. Ji, Non-local u-nets for biomedical image segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (04) 2020, pp. 6315–6322, http://dx.doi.org/10.1609/aaai.v34i04.6100.

[130] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, 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.

[131] F. Yang, F. Wang, P. Dong, B. Wang, HCA-former: Hybrid convolution attention transformer for 3D medical image segmentation, Biomed. Signal Process. Control. 90 (2024) http://dx.doi.org/10.1016/j.bspc.2023.105834.

[132] F. Sui, H. Wang, F. Zhang, Cross-scale informative priors network for medical image segmentation, Digit. Signal Process.: Rev. J. 157 (2025) http://dx.doi.org/10.1016/j.dsp.2024.104883.

[133] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H.R. Roth, D. Xu, Swin UNETR: Swin transformers for semantic segmentation of brain tumors in mri images, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 272–284, http://dx.doi.org/10.1007/978-3-031-08999-2_22.

[134] B. Dong, W. Wang, D.P. Fan, J. Li, H. Fu, L. Shao, Polyp-PVT: Polyp segmentation with pyramid vision transformers, CAAI Artif. Intell. Res. 2 (2023) 9150015, http://dx.doi.org/10.26599/AIR.2023.9150015.

[135] X. Yu, Y. Tang, Y. Zhou, R. Gao, Q. Yang, H.H. Lee, T. Li, S. Bao, Y. Huo, Z. Xu, et al., Characterizing renal structures with 3D block aggregate transformers, 2022, http://dx.doi.org/10.48550/arXiv.2203.02430, arXiv preprint arXiv:2203.02430.

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

[137] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: European Conference on Computer Vision, Springer, 2022, pp. 205–218, http://dx.doi.org/10.1007/978-3-031-25066-8_9.

[138] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, UNETR: Transformers for 3D medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2022, pp. 574–584, http://dx.doi.org/10.48550/arXiv.2103.10504.

[139] Y. Weng, T. Zhou, Y. Li, X. Qiu, NAS-Unet: Neural architecture search for medical image segmentation, IEEE Access 7 (2019) 44247–44257, http://dx.doi.org/10.1109/ACCESS.2019.2908991.

[140] H. Ha, S. Rana, S. Gupta, T. Nguyen, S. Venkatesh, et al., Bayesian optimization with unknown search space, Adv. Neural Inf. Process. Syst. 32 (2019).

[141] J. Vanschoren, Meta-learning: A survey, 2018, http://dx.doi.org/10.48550/arXiv.1810.03548, arXiv preprint arXiv:1810.03548.

[142] Y. Zhang, D. Sidibé, O. Morel, F. Mériaudeau, Deep multimodal fusion for semantic image segmentation: A survey, Image Vis. Comput. 105 (2021) 104042, http://dx.doi.org/10.1016/j.imavis.2020.104042.

[143] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232, http://dx.doi.org/10.1109/ICCV.2017.244.

[144] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al., Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25, Springer, 2017, pp. 597–609, http://dx.doi.org/10.1007/978-3-319-59050-9_47.

[145] N. Karani, K. Chaitanya, C. Baumgartner, E. Konukoglu, A lifelong learning approach to brain MR segmentation across scanners and protocols, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 476–484, http://dx.doi.org/10.1007/978-3-030-00928-1_54.

[146] Q. Dou, Q. Liu, P.A. Heng, B. Glocker, Unpaired multi-modal segmentation via knowledge distillation, IEEE Trans. Med. Imaging 39 (7) (2020) 2415–2425, http://dx.doi.org/10.1109/TMI.2019.2963882.

[147] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/5.726791.

[148] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, 2015, http://dx.doi.org/10.48550/arXiv.1503.02531, arXiv preprint arXiv:1503.02531.

[149] L. Zhang, S. Feng, Y. Wang, Y. Wang, Y. Zhang, X. Chen, Q. Tian, Unsupervised ensemble distillation for multi-organ segmentation, in: 2022 IEEE 19th International Symposium on Biomedical Imaging, ISBI, IEEE, 2022, pp. 1–5, http://dx.doi.org/10.1109/ISBI52829.2022.9761568.

[150] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J.N. Chiang, Z. Wu, X. Ding, Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, Med. Image Anal. 63 (2020) 101693, http://dx.doi.org/10.1016/j.media.2020.101693.

[151] S. Ruder, An overview of multi-task learning in deep neural networks, 2017, http://dx.doi.org/10.48550/arXiv.1706.05098, arXiv preprint arXiv:1706.05098.

[152] C. Playout, R. Duval, F. Cheriet, A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11, Springer, 2018, pp. 101–108, http://dx.doi.org/10.1007/978-3-030-00934-2_12.

[153] V. Nath, D. Yang, B.A. Landman, D. Xu, H.R. Roth, Diminishing uncertainty within the training pool: Active learning for medical image segmentation, IEEE Trans. Med. Imaging 40 (10) (2021) 2534–2547, http://dx.doi.org/10.1109/TMI.2020.3048055.

[154] B. Khanal, B. Bhattarai, B. Khanal, D. Stoyanov, C.A. Linte, M-VAAL: Multimodal variational adversarial active learning for downstream medical image analysis tasks, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 14122 LNCS, 2024, pp. 48–63, http://dx.doi.org/10.1007/978-3-031-48593-0_4.

[155] J. Carse, S. McKenna, Active learning for patch-based digital pathology using convolutional neural networks to reduce annotation costs, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11435 LNCS, 2019, pp. 20–27, http://dx.doi.org/10.1007/978-3-030-23937-4_3.

[156] S. Budd, E.C. Robinson, B. Kainz, A survey on active learning and human-in-the-loop deep learning for medical image analysis, Med. Image Anal. 71 (2021) 102062, http://dx.doi.org/10.1016/j.media.2021.102062.

[157] D.C. Castro, I. Walker, B. Glocker, Causality matters in medical imaging, Nat. Commun. 11 (1) (2020) 3673, http://dx.doi.org/10.1038/s41467-020-17478-w.

[158] P. Conde, C. Premebida, Adaptive-TTA: accuracy-consistent weighted test time augmentation method for the uncertainty calibration of deep learning classifiers, in: BMVC 2022 - 33rd British Machine Vision Conference Proceedings, 2022.

[159] J. Ma, Y. Zhang, S. Gu, X. An, Z. Wang, C. Ge, C. Wang, F. Zhang, Y. Wang, Y. Xu, et al., Fast and low-GPU-memory abdomen CT organ segmentation: the flare challenge, Med. Image Anal. 82 (2022) 102616, http://dx.doi.org/10.1016/j.media.2022.102616.

[160] W. Lei, H. Mei, Z. Sun, S. Ye, R. Gu, H. Wang, R. Huang, S. Zhang, S. Zhang, G. Wang, Automatic segmentation of organs-at-risk from head-and-neck CT using separable convolutional neural network with hard-region-weighted loss, Neurocomputing 442 (2021) 184–199, http://dx.doi.org/10.1016/j.neucom.2021.01.135.

[161] C. Chen, X. Liu, M. Ding, J. Zheng, J. Li, 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, Springer, 2019, pp. 184–192, http://dx.doi.org/10.1007/978-3-030-32248-9_21.

[162] O. Yaniv, O. Portnoy, A. Talmon, N. Kiryati, E. Konen, A. Mayer, V-net light-parameter-efficient 3-D convolutional neural network for prostate MRI segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, IEEE, 2020, pp. 442–445, http://dx.doi.org/10.1109/ISBI45749.2020.9098643.

[163] G. Qin, W. Li, MRI three-dimensional reconstruction of liver and tumor based on deep learning, 2024, pp. 451–456, http://dx.doi.org/10.1145/3672758.3672832.

[164] Y. Dai, Q. Wang, S. Cui, Y. Yin, W. Song, MediLite3DNet: A lightweight network for segmentation of nasopharyngeal airways, Med. Biol. Eng. Comput. (2024) http://dx.doi.org/10.1007/s11517-024-03252-3.

[165] J. Liu, J. Wu, R. Jing, H. Yu, J. Liu, L. Song, Lightweight explicit 3D human digitization via normal integration, Sensors 25 (5) (2025) http://dx.doi.org/10.3390/s25051513.

[166] H. Lu, M. Li, K. Yu, Y. Zhang, L. Yu, Lumbar spine segmentation method based on deep learning, J. Appl. Clin. Med. Phys. 24 (6) (2023) e13996, http://dx.doi.org/10.1002/acm2.13996.

[167] P. Basak, R. Sarmun, S. Kabir, I. Al-Hashimi, E.H. Bhuiyan, A. Hasan, M.S. Khan, M.E. Chowdhury, Machine-agnostic automated lumbar MRI segmentation using a cascaded model based on generative neurons, Expert Syst. Appl. 264 (2025) http://dx.doi.org/10.1016/j.eswa.2024.125862.

[168] Y. He, Q. Du, H. Wu, Y. Du, J. Xu, Y. Xi, H. Yang, Multi-head consistent semi-supervised learning for lumbar CT segmentation, Biomed. Signal Process. Control. 90 (2024) http://dx.doi.org/10.1016/j.bspc.2023.105794.

[169] M. Mushtaq, M.U. Akram, N.S. Alghamdi, J. Fatima, R.F. Masood, Localization and edge-based segmentation of lumbar spine vertebrae to identify the deformities using deep learning models, Sensors 22 (4) (2022) http://dx.doi.org/10.3390/s22041547.

[170] K.C. Kim, H.C. Cho, T.J. Jang, J.M. Choi, J.K. Seo, Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation, Comput. Methods Programs Biomed. 200 (2021) http://dx.doi.org/10.1016/j.cmpb.2020.105833.

[171] J. Liu, Y. Zhou, X. Cui, F. Jin, G. Suo, H. Xu, J. Yang, Multi-scale hybrid attention convolutional neural network for automatic segmentation of lumbar vertebrae from MRI, IEEE Access 12 (2024) 77999–78013, http://dx.doi.org/10.1109/ACCESS.2024.3407833.

[172] J. Andrew, M. DivyaVarshini, P. Barjo, I. Tigga, Spine magnetic resonance image segmentation using deep learning techniques, in: 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS, IEEE, 2020, pp. 945–950, http://dx.doi.org/10.1109/ICACCS48705.2020.9074218.

[173] S.F. Qadri, H. Lin, L. Shen, M. Ahmad, S. Qadri, S. Khan, M. Khan, S.S. Zareen, M.A. Akbar, M.B. Bin Heyat, et al., CT-based automatic spine segmentation using patch-based deep learning, Int. J. Intell. Syst. 2023 (1) (2023) 2345835, http://dx.doi.org/10.1155/2023/2345835.

[174] U. Raghavendra, N.S. Bhat, A. Gudigar, U.R. Acharya, Automated system for the detection of thoracolumbar fractures using a CNN architecture, Future Gener. Comput. Syst. 85 (2018) 184–189, http://dx.doi.org/10.1016/j.future.2018.03.023.

[175] C.W. Webb, K. Aguirre, P.H. Seidenberg, Lumbar spinal stenosis: diagnosis and management, Am. Fam. Physician 109 (4) (2024) 350–359.

[176] J. Ross, Jr., E. Braunwald, Aortic stenosis, Circulation 38 (1s5) (1968) V–61.

[177] S. Ghosh, M.R. Malgireddy, V. Chaudhary, G. Dhillon, A new approach to automatic disc localization in clinical lumbar MRI: Combining machine learning with heuristics, in: 2012 9th IEEE International Symposium on Biomedical Imaging, ISBI, 2012, pp. 114–117, http://dx.doi.org/10.1109/ISBI.2012.6235497.

[178] A. Kumar, A. Kumar, P. Sharma, Scoliosis: Review of diagnosis and treatment, Int. J. Converg. Heal. 4 (1) (2024) 23–23.

[179] M. Aebi, The adult scoliosis, Eur. Spine J. 14 (2005) 925–948, http://dx.doi.org/10.1007/s00586-005-1053-9.

[180] M.H. Horng, C.P. Kuok, M.J. Fu, C.J. Lin, Y.N. Sun, Cobb angle measurement of spine from X-ray images using convolutional neural network, Comput. Math. Methods Med. 2019 (1) (2019) 6357171, http://dx.doi.org/10.1155/2019/6357171.

[181] A.A. Khan, R.H. Slart, D.S. Ali, O. Bock, J.J. Carey, P. Camacho, K. Engelke, P.A. Erba, N.C. Harvey, W.F. Lems, et al., Osteoporotic fractures: diagnosis, evaluation, and significance from the international working group on DXA best practices, in: Mayo Clinic Proceedings, vol. 99, (7) Elsevier, 2024, pp. 1127–1141.

[182] O. Johnell, J. Kanis, Epidemiology of osteoporotic fractures, Osteoporos Int. 16 (2005) S3–S7, http://dx.doi.org/10.1007/s00198-004-1702-6.

[183] A. Azizi, A. Azizzadeh, Y. Tavakoli, N. Vahed, T. Mousavi, Thoracolumbar fracture and spinal cord injury in blunt trauma: a systematic review, meta-analysis, and meta-regression, Neurosurg. Rev. 47 (1) (2024) 333.

[184] P.C. McAfee, H. Yuan, B. Fredrickson, J. Lubicky, The value of computed tomography in thoracolumbar fractures. An analysis of one hundred consecutive cases and a new classification, J. Bone Jt. Surg. 65 (4) (1983) 461–473.

[185] N. Fine, S. Lively, C.A. Séguin, A.V. Perruccio, M. Kapoor, R. Rampersaud, Intervertebral disc degeneration and osteoarthritis: a common molecular disease spectrum, Nat. Rev. Rheumatol. 19 (3) (2023) 136–152.

[186] L.S. Lim, P. Mitchell, J.M. Seddon, F.G. Holz, T.Y. Wong, Age-related macular degeneration, Lancet 379 (9827) (2012) 1728–1738, http://dx.doi.org/10.1016/S0140-6736(22)02609-5.

[187] B. Jebri, M. Phillips, K. Knapp, A. Appelboam, A. Reuben, G. Slabaugh, Detection of degenerative change in lateral projection cervical spine X-ray images, in: Medical Imaging 2015: Computer-Aided Diagnosis, vol. 9414, SPIE, 2015, pp. 18–25, http://dx.doi.org/10.1117/12.2082515.

[188] I. Jeon, E. Kong, Application of simultaneous 18F-FDG PET/MRI for evaluating residual lesion in pyogenic spine infection: A case report, Infect. Chemother. 52 (4) (2020) 626, http://dx.doi.org/10.3947/ic.2020.52.4.626.

[189] R.J. Sneath, A. Khan, C. Hutchinson, An objective assessment of lumbar spine degeneration/ageing seen on MRI using an ensemble method—A novel approach to lumbar MRI reporting, Spine 47 (5) (2022) E187–E195, http://dx.doi.org/10.1097/BRS.0000000000004159.

[190] C. Boulay, G. Bollini, J. Legaye, C. Tardieu, D. Prat-Pradal, B. Chabrol, J.L. Jouve, G. Duval-Beaupère, J. Pélissier, Pelvic incidence: A predictive factor for three-dimensional acetabular orientation—A preliminary study, Anat. Res. Int. 2014 (1) (2014) 594650, http://dx.doi.org/10.1155/2014/594650.

[191] D. Forsberg, E. Sjöblom, J.L. Sunshine, Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data, J. Digit. Imaging 30 (4) (2017) 406–412, http://dx.doi.org/10.1007/s10278-017-9945-x.

[192] S. Furqan Qadri, D. Ai, G. Hu, M. Ahmad, Y. Huang, Y. Wang, J. Yang, Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images, Appl. Sci. 9 (1) (2018) 69, http://dx.doi.org/10.3390/app9010069.

[193] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251, http://dx.doi.org/10.1109/TMI.2016.2538465.

[194] S.F. Qadri, M. Ahmad, D. Ai, J. Yang, Y. Wang, Deep belief network based vertebra segmentation for CT images, in: Image and Graphics Technologies and Applications: 13th Conference on Image and Graphics Technologies and Applications, IGTA 2018, Beijing, China, April 8–10, 2018, Revised Selected Papers 13, Springer, 2018, pp. 536–545, http://dx.doi.org/10.1007/978-981-13-1702-6_53.

[195] M. Pereañez, et al., Accurate segmentation of vertebral bodies and processes using statistical shape decomposition and conditional models, IEEE Trans. Med. Imaging 34 (8) (2015) 1627–1639, http://dx.doi.org/10.1109/TMI.2015.2396774.

[196] I. Castro-Mateos, J.M. Pozo, M. Pereañez, K. Lekadir, A. Lazary, A.F. Frangi, Statistical interspace models (SIMs): application to robust 3D spine segmentation, IEEE Trans. Med. Imaging 34 (8) (2015) 1663–1675, http://dx.doi.org/10.1109/TMI.2015.2443912.

[197] A. Rasoulian, R. Rohling, P. Abolmaesumi, Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+pose model, IEEE Trans. Med. Imaging 32 (10) (2013) 1890–1900, http://dx.doi.org/10.1109/TMI.2013.2268424.

[198] B. Ibragimov, R. Korez, B. Likar, F. Pernuš, L. Xing, T. Vrtovec, Segmentation of pathological structures by landmark-assisted deformable models, IEEE Trans. Med. Imaging 36 (7) (2017) 1457–1469, http://dx.doi.org/10.1109/TMI.2017.2667578.

[199] B. Ibragimov, B. Likar, F. Pernuš, T. Vrtovec, Shape representation for efficient landmark-based segmentation in 3-D, IEEE Trans. Med. Imaging 33 (4) (2014) 861–874, http://dx.doi.org/10.1109/TMI.2013.2296976.

[200] S. Kadoury, H. Labelle, N. Paragios, Spine segmentation in medical images using manifold embeddings and higher-order MRFs, IEEE Trans. Med. Imaging 32 (7) (2013) 1227–1238, http://dx.doi.org/10.1109/TMI.2013.2244903.

[201] S. Kadoury, H. Labelle, N. Paragios, Automatic inference of articulated spine models in CT images using high-order Markov random fields, Med. Image Anal. 15 (4) (2011) 426–437, http://dx.doi.org/10.1016/j.media.2011.01.006.

[202] J.S. Athertya, G.S. Kumar, Automatic segmentation of vertebral contours from CT images using fuzzy corners, Comput. Biol. Med. 72 (2016) 75–89, http://dx.doi.org/10.1016/j.compbiomed.2016.03.009.

[203] K. Hammernik, T. Ebner, D. Stern, M. Urschler, T. Pock, Vertebrae segmentation in 3D CT images based on a variational framework, in: Recent Advances in Computational Methods and Clinical Applications for Spine Imaging, Springer, 2015, pp. 227–233, http://dx.doi.org/10.1007/978-3-319-14148-0_20.

[204] P.H. Lim, U. Bagci, L. Bai, A robust segmentation framework for spine trauma diagnosis, in: Computational Methods and Clinical Applications for Spine Imaging: Proceedings of the Workshop Held At the 16th International Conference on Medical Image Computing and Computer Assisted Intervention, September 22-26, 2013, Nagoya, Japan, Springer, 2014, pp. 25–33, http://dx.doi.org/10.1007/978-3-319-07269-2_3.

[205] R. Korez, B. Ibragimov, B. Likar, F. Pernuš, T. Vrtovec, A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation, IEEE Trans. Med. Imaging 34 (8) (2015) 1649–1662, http://dx.doi.org/10.1109/TMI.2015.2389334.

[206] A. Suzani, A. Rasoulian, A. Seitel, S. Fels, R.N. Rohling, P. Abolmaesumi, Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images, in: Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 9415, SPIE, 2015, pp. 269–275, http://dx.doi.org/10.1117/12.2081542.

[207] C. Chu, D.L. Belavỳ, G. Armbrecht, M. Bansmann, D. Felsenberg, G. Zheng, Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method, PLoS One 10 (11) (2015) e0143327, http://dx.doi.org/10.1371/journal.pone.0143327.

[208] R. Korez, B. Likar, F. Pernuš, T. Vrtovec, Model-based segmentation of vertebral bodies from MR images with 3D CNNs, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 433–441, http://dx.doi.org/10.1007/978-3-319-46723-8_50.

[209] N. Lang, Y. Zhang, E. Zhang, J. Zhang, D. Chow, P. Chang, J.Y. Hon, H. Yuan, M.Y. Su, Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI, Magn. Reson. Imaging 64 (2019) 4–12, http://dx.doi.org/10.1016/j.mri.2019.02.013.

[210] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, Springer, 2016, pp. 424–432, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[211] A.S. Al Kafri, S. Sudirman, A.J. Hussain, D. Al-Jumeily, P. Fergus, F. Natalia, H. Meidia, N. Afriliana, A. Sophian, M. Al-Jumaily, et al., Segmentation of lumbar spine MRI images for stenosis detection using patch-based pixel classification neural network, in: 2018 IEEE Congress on Evolutionary Computation, CEC, IEEE, 2018, pp. 1–8, http://dx.doi.org/10.1109/CEC.2018.8477893.

[212] A.S. Al-Kafri, S. Sudirman, A. Hussain, D. Al-Jumeily, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi, M. Al-Jumaily, Boundary delineation of MRI images for lumbar spinal stenosis detection through semantic segmentation using deep neural networks, IEEE Access 7 (2019) 43487–43501, http://dx.doi.org/10.1109/ACCESS.2019.2908002.

[213] A. Sekuboyina, J. Kukačka, J.S. Kirschke, B.H. Menze, A. Valentinitsch, Attention-driven deep learning for pathological spine segmentation, in: International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer, 2017, pp. 108–119, http://dx.doi.org/10.1007/978-3-319-74113-0_10.

[214] R. Janssens, G. Zeng, G. Zheng, Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, 2018, pp. 893–897, http://dx.doi.org/10.1109/ISBI.2018.8363715.

[215] N. Lessmann, B. Van Ginneken, P.A. De Jong, I. Išgum, Iterative fully convolutional neural networks for automatic vertebra segmentation and identification, Med. Image Anal. 53 (2019) 142–155, http://dx.doi.org/10.1016/j.media.2019.02.005.

[216] I. Castro-Mateos, J.M. Pozo, A. Lazary, A. Frangi, 3D vertebra segmentation by feature selection active shape model, 20, 2015, pp. 241–245, http://dx.doi.org/10.1007/978-3-319-14148-0_22.

[217] S. Aydodgu, K.W. Yung, D. Stoyanov, D. Kalaskar, E. Mazomenos, Improving vertebrae segmentation using a centroid detection-guided transformer-based network, 2024, http://dx.doi.org/10.1109/ISBI56570.2024.10635318.

[218] D. Zukić, A. Vlasák, T. Dukatz, J. Egger, D. Hořínek, C. Nimsky, A. Kolb, Segmentation of vertebral bodies in MR images, 2012, pp. 135–142, http://dx.doi.org/10.2312/PE/VMV/VMV12/135-142.

[219] M. Sajeer, M. Mallikarjunaswamy, Segmentation, diagnosis and analysis of intervertebral discs using lumbar spine MRI, Int. J. Appl. Eng. Res. 10 (44) (2015) 30631–30636.

[220] Z. Yang, G. Jia, S. Wang, Y. Wang, G. Bai, S. Tian, S. Tang, Research on segmentation algorithm for vertebral CT images based on spatial configuration-net and U-net deep learning model, 2023, pp. 236–241, http://dx.doi.org/10.1145/3644116.3644159.

[221] L. Boneta, A. Shalaby, A. Refaey, K. Loukhaoukha, G. Giakos, 3D simultaneous segmentation and registration of vertebral bodies for accurate BMD measurements, in: IST 2017 - IEEE International Conference on Imaging Systems and Techniques, Proceedings, 2018-January, 2017, pp. 1–5, http://dx.doi.org/10.1109/IST.2017.8261450.

[222] Sushmitha, M. Kanthi, V. Kedlaya K, T. Parupudi, S.N. Bhat, S.G. Nayak, Identification of vertebrae in CT scans for improved clinical outcomes using advanced image segmentation, Signals 5 (4) (2024) 869–882, http://dx.doi.org/10.3390/signals5040047.

[223] S. Ruiz-Espana, A. Diaz-Parra, E. Arana, D. Moratal, A fully automated level-set based segmentation method of thoracic and lumbar vertebral bodies in computed tomography images, 2015-November, 2015, pp. 3049–3052, http://dx.doi.org/10.1109/EMBC.2015.7319035.

[224] D. Forsberg, Atlas-based segmentation of the thoracic and lumbar vertebrae, 20, 2015, pp. 215–220, http://dx.doi.org/10.1007/978-3-319-14148-0_18.

[225] J.B. Courbot, E. Rust, E. Monfrini, C. Collet, 2-step robust vertebra segmentation, in: 5th International Conference on Image Processing, Theory, Tools and Applications 2015, IPTA 2015, 2015, pp. 157–162, http://dx.doi.org/10.1109/IPTA.2015.7367118.

[226] M. Aslan, A. Shalaby, A. Ali, A.A. Farag, Model-based segmentation, reconstruction and analysis of the vertebral body from spinal CT, Lect. Notes Comput. Vis. Biomech. 18 (2015) 381–438, http://dx.doi.org/10.1007/978-3-319-12508-4_13.

[227] W. Shi, T. Xu, H. Yang, Y. Xi, Y. Du, J. Li, J. Li, Attention gate based dual-pathway network for vertebra segmentation of X-Ray spine images, IEEE J. Biomed. Heal. Informatics 26 (8) (2022) 3976–3987, http://dx.doi.org/10.1109/JBHI.2022.3158968.

[228] S. Hanaoka, Y. Nomura, M. Nemoto, Y. Masutani, E. Maeda, T. Yoshikawa, N. Hayashi, N. Yoshioka, K. Ohtomo, Automated segmentation method for spinal column based on a dual elliptic column model and its application for virtual spinal straightening, J. Comput. Assist. Tomogr. 34 (1) (2010) 156–162, http://dx.doi.org/10.1097/RCT.0b013e3181b12242.

[229] N. Lessmann, B. van Ginneken, I. Išgum, Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images, in: Medical Imaging 2018: Image Processing, vol. 10574, SPIE, 2018, pp. 39–44, http://dx.doi.org/10.1117/12.2292731.

[230] R. Cheng, H.R. Roth, N. Lay, L. Lu, B. Turkbey, W. Gandler, E.S. McCreedy, T. Pohida, P.A. Pinto, P. Choyke, et al., Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks, J. Med. Imaging 4 (4) (2017) http://dx.doi.org/10.1117/1.JMI.4.4.041302, 041302–041302.

[231] C. Payer, D. Stern, H. Bischof, M. Urschler, Coarse to fine vertebrae localization and segmentation with SpatialConfiguration-net and U-net, in: VISIGRAPP (5: VISAPP), 2020, pp. 124–133, http://dx.doi.org/10.5220/0008975201240133.

[232] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018, pp. 3–11, http://dx.doi.org/10.1007/978-3-030-00889-5_1.

[233] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.W. Chen, J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2020, pp. 1055–1059, http://dx.doi.org/10.1109/ICASSP40776.2020.9053405.

[234] G. Hille, S. Saalfeld, S. Serowy, K. Tönnies, Vertebral body segmentation in wide range clinical routine spine MRI data, Comput. Methods Programs Biomed. 155 (2018) 93–99, http://dx.doi.org/10.1016/j.cmpb.2017.12.013.

[235] C. Payer, D. Štern, H. Bischof, M. Urschler, Integrating spatial configuration into heatmap regression based CNNs for landmark localization, Med. Image Anal. 54 (2019) 207–219, http://dx.doi.org/10.1016/j.media.2019.03.007.

[236] A. Sekuboyina, A. Valentinitsch, J.S. Kirschke, B.H. Menze, A localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets, 2017, http://dx.doi.org/10.48550/arXiv.1703.04347, arXiv preprint arXiv:1703.04347.

[237] P. Liu, Y. Deng, C. Wang, Y. Hui, Q. Li, J. Li, S. Luo, M. Sun, Q. Quan, S. Yang, et al., Universal segmentation of 33 anatomies, 2022, http://dx.doi.org/10.48550/arXiv.2203.02098, arXiv preprint arXiv:2203.02098.

[238] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507, http://dx.doi.org/10.1126/science.1127647.

[239] H.C. Shin, M.R. Orton, D.J. Collins, S.J. Doran, M.O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1930–1943, http://dx.doi.org/10.1109/TPAMI.2012.277.

[240] M.A. Aslam, C. Xue, Y. Chen, A. Zhang, M. Liu, K. Wang, D. Cui, Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network, Sci. Rep. 11 (1) (2021) 4014, http://dx.doi.org/10.1038/s41598-021-83184-2.

[241] G. Praveen, A. Agrawal, P. Sundaram, S. Sardesai, Ischemic stroke lesion segmentation using stacked sparse autoencoder, Comput. Biol. Med. 99 (2018) 38–52, http://dx.doi.org/10.1016/j.compbiomed.2018.05.027.

[242] S.F. Qadri, Z. Zhao, D. Ai, M. Ahmad, Y. Wang, Vertebrae segmentation via stacked sparse autoencoder from computed tomography images, in: Eleventh International Conference on Digital Image Processing (ICDIP 2019), vol. 11179, SPIE, 2019, pp. 1206–1211, http://dx.doi.org/10.1117/12.2540176.

[243] X. Wang, S. Zhai, Y. Niu, Automatic vertebrae localization and identification by combining deep SSAE contextual features and structured regression forest, J. Digit. Imaging 32 (2019) 336–348, http://dx.doi.org/10.1007/s10278-018-0140-5.

[244] S. Li, H. Lei, F. Zhou, J. Gardezi, B. Lei, Longitudinal and multi-modal data learning for parkinson's disease diagnosis via stacked sparse auto-encoder, in: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, 2019, pp. 384–387, http://dx.doi.org/10.1109/ISBI.2019.8759385.