REVIEW ARTICLE

A Comprehensive Review on 3D Volumetric CT Liver Segmentation: Techniques, Challenges, Trends, and Future Research Directions

Irshad Ali Khan[1], **Gul Zaman Khan**[1], **Yar Muhammad**[2,*], **Samreen Ihsan**[1] **and Ijazul Haq**[3]

[1] School of Software Technology, Dalian University of Technology, Dalian 116000, China
[2] School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[3] Shien-Ming Wu School of Intelligent Manufacturing, South China University of Technology, Guangzhou 511442, China

Abstract

Accurate liver segmentation from three-dimensional (3D) computed tomography (CT) volumes is a critical step in computer-aided diagnosis, surgical planning, and disease quantification. Despite substantial progress in deep learning, achieving robust and generalizable liver segmentation remains challenging due to complex organ boundaries, pathological variations, and domain shifts across scanners. This review provides a comprehensive overview of 3D volumetric liver segmentation techniques, spanning from classical model-based methods to contemporary transformer-driven frameworks. We categorize existing methods into three paradigms: (1) classical statistical and atlas-based methods, (2) deep convolutional architectures, and (3) hybrid and attention-based transformer approaches. Key benchmark datasets, evaluation metrics, and performance comparisons are discussed in detail. Furthermore, we highlight open challenges, such as data imbalance, domain generalization, and clinical interpretability, and propose potential future directions, including self-supervised learning, multi-modal integration, and foundation models. Additionally, we identify evolving trends toward dual-stream CNN–Transformer integration, attention-enhanced spatial reasoning, and foundation-model-driven segmentation pipelines. This review aims to serve as a reference for researchers and practitioners seeking to develop next-generation 3D liver segmentation systems.

Keywords: 3D liver segmentation, computed tomography, deep learning, transformers, medical image analysis, self-supervised learning.

1 Introduction

Liver is one of the most vital and complex organs in the human body, performing critical metabolic, synthetic, and detoxifying functions. Liver cancer is the sixth most diagnosed cancer and the third most frequent cause of cancer death worldwide [1]. Accurate assessment of liver structure and function is essential for the diagnosis and treatment of hepatic

Submitted: 27 March 2025
Accepted: 09 February 2026
Published: 03 March 2026

Vol. 1, **No.** 1, 2026.
10.62762/TAIC.2025.965486

***Corresponding author:**
✉ Yar Muhammad
yarkhan@buaa.edu.cn

Citation
Khan, I. A., Khan, G. Z., Muhammad, Y., Ihsan, S., & Haq, I. (2026). A Comprehensive Review on 3D Volumetric CT Liver Segmentation: Techniques, Challenges, Trends, and Future Research Directions. *ICCK Transactions on Applied Intelligence and Cybernetics*, 1(1), 5–35.

diseases such as hepatocellular carcinoma (HCC), cirrhosis, fatty liver disease, and metastatic tumors [2]. CT imaging has become the modality of choice for liver analysis due to its high spatial resolution, rapid acquisition time, and compatibility with 3D volumetric visualization. However, the extraction of the liver region and liver tumor from CT volumes, known as liver tumor segmentation, remains a challenging and essential task in computer-aided diagnosis, surgical planning, and radiotherapy dose calculation [4, 5]. As can be seen from Figure 1, the muscle around the liver and the small difference in gray-scale contrast between other organs and the liver tumor result in blurred edges, posing a challenge in segmenting the liver tumor. At the same time, the size and location of liver tumors vary between individuals, and liver tumors are spread over multiple slices of CT images with subtle differences between different slices, all of which pose a great challenge to the liver tumor segmentation task. Medical imaging techniques are frequently used in the clinical evaluation of liver diseases to provide fairly detailed images of soft-tissue organs via a non-invasive procedure, resulting in cross-sectional images of the abdominal cavity, where the liver is located. Several imaging modalities can be used to analyze the liver in clinical practice routine, with CT being one of the most frequently used, particularly in the context of liver cancer. Therefore, the segmentation of liver structures in CT images has gained increasing attention from the research community in the last decade since it represents an important step towards computer-assisted diagnosis and/or treatment planning for various hepatic diseases.

Manual delineation of liver boundaries by radiologists is considered the gold standard, but it is an extremely time-consuming, labor-intensive, and subjective process prone to intra- and inter-observer variability. As medical imaging data continue to grow exponentially, there is an urgent demand for automatic and accurate liver segmentation methods capable of handling diverse patient anatomies, imaging protocols, and disease conditions. Automated segmentation not only reduces the workload of clinicians but also improves reproducibility and precision in downstream clinical applications such as tumor detection, volumetric quantification, and 3D surgical simulation.

Over the past two decades, liver segmentation research has evolved through several methodological paradigms. Early approaches relied on low-level image processing techniques, such as thresholding, region growing, edge detection, and active contour models, which were limited by noise sensitivity and poor generalization across datasets [3]. Subsequent advances introduced model-based strategies, including statistical shape models, level-set methods, and atlas-based registration, which improved anatomical consistency but still struggled in cases of pathological deformation or low contrast between the liver and surrounding organs.

The emergence of deep learning, particularly convolutional neural networks (CNNs), revolutionized medical image segmentation by enabling data-driven feature learning and hierarchical representation extraction. The introduction of U-Net [7] and its variants (e.g., 3D U-Net [8], V-Net [6]) provided powerful encoder–decoder frameworks that achieved unprecedented segmentation accuracy on 2D and 3D medical images. These architectures leveraged skip connections, multi-scale feature fusion, and end-to-end training to capture both global context and fine anatomical details. The success of CNNs led to their rapid adoption in numerous liver segmentation studies, often trained and evaluated on public benchmark datasets such as LiTS (Liver Tumor Segmentation Challenge) and 3DIRCADb.

Despite their remarkable performance, purely convolutional models exhibit limitations in capturing long-range spatial dependencies due to their local receptive fields. To address this, hybrid architectures integrating attention mechanisms and vision transformers (ViTs) have recently emerged as promising alternatives. Models such as TransUNet [9] and Swin-UNETR [10] combine convolutional backbones with transformer-based encoders to model global relationships and improve boundary precision. These transformer-driven methods demonstrate superior generalization and robustness, especially in complex clinical scenarios with irregular tumor morphologies or domain shifts between institutions.

In parallel, several complementary research directions have gained traction. Self-supervised learning (SSL) and contrastive pretraining enable models to learn from large-scale unlabeled medical datasets, mitigating the scarcity of annotated data. Multi-modal segmentation techniques integrating CT, MRI, and PET data enhance tissue differentiation by leveraging complementary information. Furthermore, federated learning and domain adaptation methods aim to address privacy constraints and improve cross-domain performance without direct data sharing between

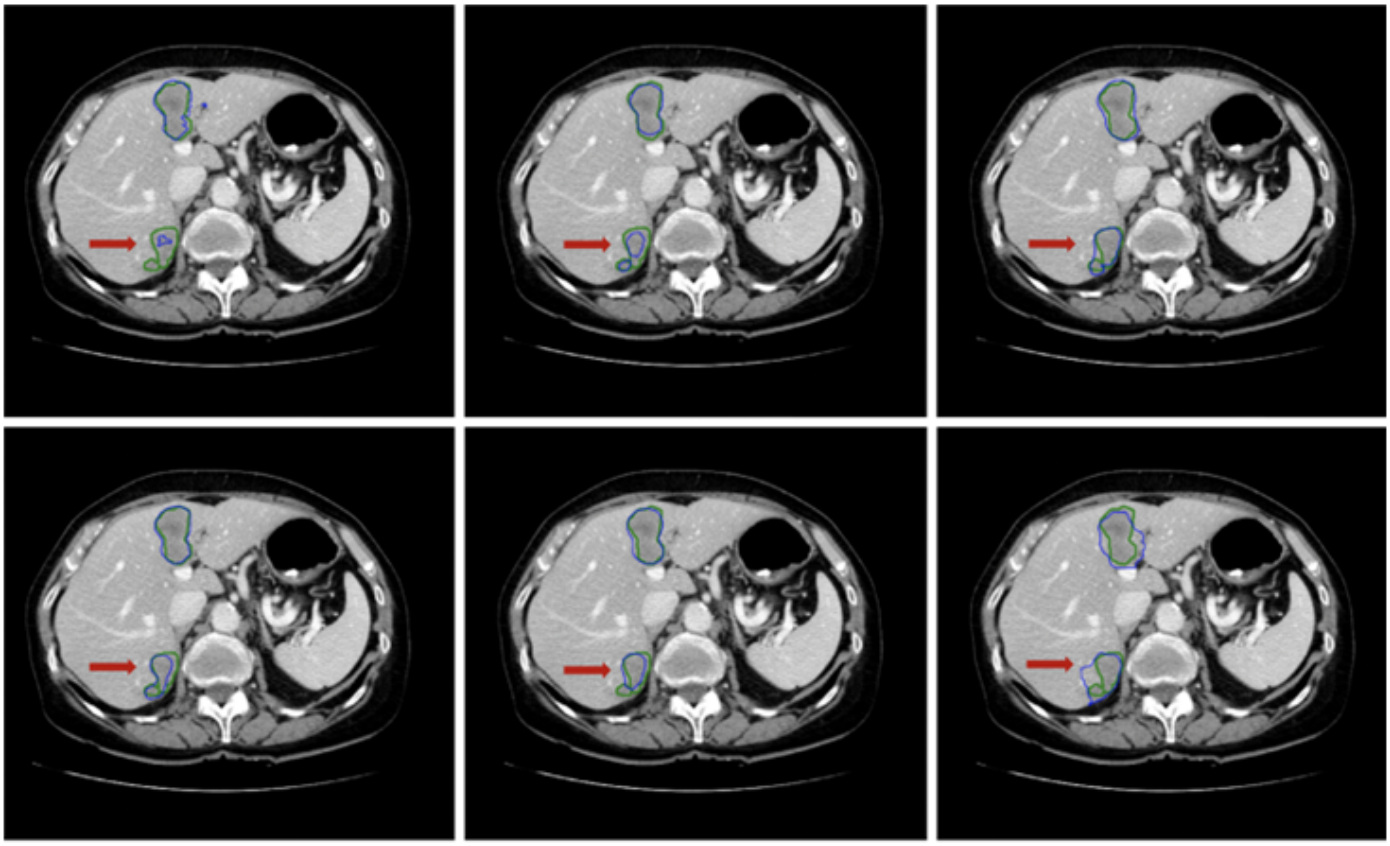

Figure 1. Liver tumor segmentation results of the ISBI–LiTS 2017 challenge [5].

hospitals.

This review provides a comprehensive and systematic overview of the evolution of 3D liver segmentation from CT volumes. We summarize and compare existing methods across three major paradigms: (1) classical model-driven approaches, (2) deep learning-based convolutional networks, and (3) hybrid and transformer-based architectures. We also discuss the most commonly used public datasets, evaluation metrics, and quantitative benchmarks. In addition, the review highlights key challenges such as data imbalance, domain generalization, computational cost, and clinical integration and proposes potential research directions, including foundation models, interpretable AI, and real-time surgical assistance systems.

2 Anatomical and Clinical Background

The human liver is the largest solid organ in the body. It performs a multitude of vital physiological functions, including metabolism of carbohydrates, fats, and proteins; synthesis of plasma proteins and bile; detoxification of harmful substances; and regulation of glycogen storage [4]. Because of its essential role in maintaining homeostasis, even minor hepatic abnormalities can have profound clinical implications. Consequently, accurate imaging and quantitative analysis of the liver are indispensable in modern clinical practice.

From an anatomical standpoint, the liver is divided into right and left lobes by the middle hepatic vein and further subdivided into eight Couinaud segments based on vascular and biliary anatomy. Each segment has its own vascular inflow, outflow, and biliary drainage, which makes segment-level volumetric analysis crucial for surgical planning, especially in liver resection and transplantation [3]. CT imaging provides excellent spatial resolution for visualizing these structures, allowing radiologists and surgeons to assess the liver parenchyma, detect focal lesions, and evaluate vascular integrity.

2.1 Liver Appearance and Imaging Characteristics in CT

In CT imaging, the liver typically exhibits a nearly homogeneous soft-tissue intensity distribution, slightly higher than the spleen in the portal venous phase. However, its intensity characteristics vary depending on the contrast phase, pathology, and imaging parameters. Standard liver CT protocols

often include three contrast-enhanced phases: arterial, portal venous, and delayed. The portal venous phase, acquired approximately 60–70 seconds after contrast injection, provides optimal parenchymal enhancement and is most commonly used for segmentation tasks due to the clear boundary between the liver and surrounding organs.

Despite this, delineating the liver boundary remains challenging because of several confounding factors:

- **Intensity Overlap:** Adjacent organs such as the stomach, right kidney, and spleen often exhibit similar Hounsfield unit (HU) values, leading to ambiguous boundaries.
- **Anatomical Variability:** The shape, size, and position of the liver vary significantly across patients due to age, body habitus, and pathological deformation.
- **Partial Volume Effects:** The limited spatial resolution of CT may cause voxel intensities to represent a mixture of tissues, complicating edge delineation.
- **Pathological Alterations:** Tumors, cysts, and cirrhotic nodules can distort local intensity patterns, further complicating segmentation.

These factors collectively increase the complexity of automated segmentation algorithms and demand sophisticated models that can adapt to heterogeneous imaging conditions.

2.2 Clinical Relevance of Accurate Liver Segmentation

Accurate 3D liver segmentation is pivotal in a variety of clinical workflows. In preoperative planning, surgeons rely on volumetric analysis of the liver and its vascular system to assess the remnant liver volume (RLV) before partial hepatectomy. Underestimation of RLV may lead to postoperative liver failure, while overestimation could exclude potentially operable cases [5]. Automated 3D segmentation facilitates interactive surgical simulation, enabling the visualization of tumor location relative to hepatic vessels and aiding in the design of safe resection margins.

In radiation therapy, segmentation supports dose planning to avoid irradiation of healthy liver tissue. Similarly, in interventional radiology, volumetric segmentation is used to guide procedures such as transarterial chemoembolization (TACE) and radiofrequency ablation (RFA), ensuring precise targeting of lesions. Moreover, quantitative liver segmentation assists in disease assessment for instance, estimating the hepatic fat fraction in non-alcoholic fatty liver disease (NAFLD) or measuring fibrosis progression in chronic hepatitis.

In the context of liver transplantation, 3D segmentation provides critical information about donor and recipient anatomy, allowing volumetric matching and assessment of vascular variants. In living-donor transplantation, especially, segmentation supports virtual hepatectomy simulation to ensure adequate graft volume for the recipient and sufficient remnant for the donor.

2.3 Challenges in Clinical Imaging and Annotation

The clinical environment introduces additional constraints that complicate dataset creation and algorithm development. High inter-scanner variability, differences in acquisition protocols, and noise artifacts all affect image quality. Furthermore, manual annotation of 3D CT volumes is labor-intensive, typically requiring several hours per case, and demands expert radiological expertise. Inter-observer variability remains a significant issue, as even experienced radiologists may differ in boundary interpretation especially in low-contrast or pathological regions.

To address these challenges, various initiatives such as the Liver Tumor Segmentation (LiTS) challenge and the Medical Segmentation Decathlon (MSD) have provided publicly available annotated datasets and standardized evaluation protocols. These datasets have significantly accelerated progress by enabling fair benchmarking and reproducibility. Nonetheless, clinical deployment of segmentation systems still faces regulatory and ethical barriers, particularly concerning model interpretability, reliability, and patient safety.

2.4 From Anatomy to Computation

Understanding the anatomical and clinical complexity of the liver is essential for designing effective computational models. The hierarchical vascular and lobular structure of the liver motivates the use of multi-scale feature learning in neural networks, where both global and local spatial dependencies are crucial. The intricate interplay between liver parenchyma, vessels, and lesions requires algorithms capable of capturing contextual relationships beyond pixel intensity. In particular, transformer-based architectures have shown promise in modeling such

long-range dependencies that mirror the organ's physiological organization.

Accurate liver segmentation serves as the foundation for a wide range of clinical and computational tasks. A detailed understanding of liver anatomy, its imaging characteristics, and associated clinical requirements provides the necessary context for evaluating the performance and practical utility of modern segmentation algorithms. This section thus establishes the biological and clinical motivation for the methodological developments discussed in subsequent sections.

3 Datasets and Evaluation Metrics

Accurate and reproducible segmentation of liver structures in CT) images heavily depends on the quality and diversity of annotated datasets and the rigor of evaluation metrics used to benchmark algorithmic performance. The development of deep learning-based 3D volumetric segmentation models has been accelerated by the availability of publicly accessible datasets and well-defined quantitative metrics, which together ensure objective comparison, reproducibility, and clinical reliability. This section discusses the most widely used liver CT datasets, their annotation protocols, and the primary evaluation metrics employed in the literature.

3.1 Publicly Available Datasets

Several benchmark datasets have been released to support the research community in liver and tumor segmentation. These datasets vary in acquisition protocols, contrast phases, voxel resolution, and labeling granularity. Table 1 summarizes the key characteristics of representative 3D liver segmentation datasets.

3.1.1 Liver Tumor Segmentation (LiTS) Challenge Dataset

The Liver Tumor Segmentation (LiTS) dataset [5] was introduced as part of the ISBI 2017 and MICCAI 2017 challenges. It contains 131 contrast-enhanced 3D abdominal CT scans collected from multiple clinical centers, along with manually annotated liver and tumor masks. The scans exhibit heterogeneous acquisition parameters (voxel spacing ranging from 0.6 to 1.0 mm and slice thickness between 0.7 and 5.0 mm), providing robustness against scanner and patient variability. The LiTS dataset remains the de facto benchmark for deep learning-based 3D liver segmentation.

3.1.2 3DIRCADb Dataset

The 3D-IRCADb dataset [11] includes 20 high-resolution contrast-enhanced CT volumes acquired from different patients with varying pathologies such as hemangiomas, metastases, and cysts. The dataset offers detailed manual segmentations of the liver, hepatic vessels, and tumors. Each volume has an in-plane resolution of approximately 0.56 mm and a slice thickness of 1.0 mm. Despite its limited sample size, 3DIRCADb is valuable for algorithm validation due to its high annotation accuracy and complex anatomical variability.

3.1.3 Medical Segmentation Decathlon (MSD) – Task 08: Hepatic Vessels

The MSD dataset [12] provides a standardized platform for evaluating 3D medical segmentation algorithms across multiple organs. Task 08 focuses on liver and hepatic vessel segmentation from 303 3D CT scans. The dataset's uniform preprocessing pipeline (isotropic resampling, intensity normalization) and extensive annotations make it suitable for benchmarking both liver parenchyma and vessel segmentation models.

3.1.4 CHAOS Challenge Dataset

The Combined Healthy Abdominal Organ Segmentation (CHAOS) dataset [13] includes both CT and MRI scans of the liver, kidneys, and spleen. For CT data, 40 3D volumes are provided with ground-truth annotations of liver boundaries. The dataset enables multi-modal learning and cross-domain adaptation studies, as it provides both modalities under consistent labeling schemes.

3.1.5 LITS++ and MSD (Medical Segmentation Decathlon)

The Medical Segmentation Decathlon (MSD) [14] Task 3 (Liver) provides 201 contrast-enhanced CT volumes with annotations for liver and tumors. It merges data diversity with high-quality voxel-level annotations and serves as a strong benchmark for transfer learning and cross-domain studies. The LITS++ dataset further standardizes preprocessing and labels, promoting fair comparison among architectures.

3.1.6 BTCV and Synapse Multi-organ Datasets

Beyond liver-focused datasets, multi-organ segmentation datasets like BTCV (Beyond the Cranial Vault) and Synapse [15] contain liver labels alongside other organs. These are crucial for testing the generalizability of liver segmentation models

in multi-organ settings and for developing unified abdominal segmentation frameworks.

3.1.7 *Private Clinical Datasets*

Beyond public repositories, many researchers utilize in-house clinical datasets comprising hundreds of CT scans annotated by expert radiologists [16]. Such datasets often cover a broader range of pathological variations (tumors, post-surgical changes, steatosis) and provide real-world data heterogeneity. However, their restricted accessibility limits reproducibility and standardized evaluation across studies.

3.2 Dataset Characteristics and Preprocessing

Different datasets present diverse voxel resolutions, typically ranging from 0.5–1.5 mm in the axial plane and 1–5 mm inter-slice spacing. Preprocessing steps such as resampling, intensity clipping (e.g., HU = [-200, 250]), normalization, and organ-centric cropping are standard practices. Data augmentation techniques—random rotation, scaling, elastic deformation, and intensity perturbation—help mitigate limited dataset size and prevent overfitting. Some recent works [17] also leverage hybrid augmentation strategies combining geometric and intensity-based transformations.

3.3 Evaluation Metrics for 3D Segmentation

Robust datasets and standardized evaluation metrics are indispensable for advancing liver segmentation research. The diversity of publicly available datasets ensures algorithm generalizability across scanners and populations, while quantitative metrics facilitate fair comparison and clinical benchmarking. Future works increasingly combine multiple datasets and evaluation protocols to build models with strong domain transferability and clinical robustness.The evaluation of liver segmentation algorithms relies on both volumetric overlap measures and boundary distance metrics. These metrics quantify spatial agreement between the predicted segmentation S_p and the ground truth mask S_g, providing complementary insights into model accuracy. Table 2 summarizes the most widely used metrics in the literature.

3.3.1 *Dice Similarity Coefficient (DSC)*

The Dice Similarity Coefficient [59] is the most widely used metric for segmentation evaluation and measures the volumetric overlap between prediction and ground truth. Equation 1 presents the mathematical formulation of the DSC.

$$\mathrm{DSC} = \frac{2|S_p \cap S_g|}{|S_p| + |S_g|} \tag{1}$$

where $|S_p|$ and $|S_g|$ represent the number of voxels in the predicted and ground-truth regions, respectively. A DSC of 1 indicates perfect overlap, while 0 indicates no overlap. In clinical applications, DSC values above 0.95 are considered highly accurate for organ-level segmentation [4].

3.3.2 *Intersection over Union (IoU)*

The Intersection over Union (IoU) [60], also known as the Jaccard Index, measures the ratio of intersection to the union of the predicted and reference masks. Equation 2 provides the mathematical definition of the IoU.

$$\mathrm{IoU} = \frac{|S_p \cap S_g|}{|S_p \cup S_g|} \tag{2}$$

IoU is generally lower than DSC but provides a stricter evaluation of segmentation consistency, especially for irregular boundaries.

3.3.3 *Hausdorff Distance (HD) and 95th Percentile HD (HD95)*

The Hausdorff Distance quantifies the maximum surface distance between two contours, measuring the worst-case boundary deviation. The mathematical expression of the HD is given in Equation 3.

$$\mathrm{HD}(S_p, S_g) = \max\Big\{ \sup_{x \in S_p} \inf_{y \in S_g} \|x - y\|, \sup_{y \in S_g} \inf_{x \in S_p} \|x - y\| \Big\} \tag{3}$$

Due to sensitivity to outliers, the 95th percentile Hausdorff Distance (HD95) is often preferred, which excludes the largest 5% of boundary errors [18].

3.3.4 *Average Symmetric Surface Distance (ASSD)*

The Average Symmetric Surface Distance [18] computes the mean bidirectional distance between corresponding surface points of S_p and S_g. Equation 4 presents the mathematical formulation of the ASSD.

$$\mathrm{ASSD}(S_p, S_g) = \frac{1}{|S_p| + |S_g|} \Big(\sum_{x \in S_p} \min_{y \in S_g} \|x - y\| + \sum_{y \in S_g} \min_{x \in S_p} \|x - y\| \Big) \tag{4}$$

ASSD provides a robust assessment of overall boundary smoothness and is especially useful in

Table 1. Summary of commonly used 3D CT liver segmentation datasets.

Dataset	Modality	Resolution (mm)	Annotations
LiTS (2017) [5]	CT (contrast)	0.6–5.0 slice thickness	Liver, tumor
3DIRCADb (2010) [11]	CT (contrast)	$0.56 \times 0.56 \times 1.0$	Liver, vessel, tumor
MSD Task 08 (2019) [12]	CT (contrast)	1.0 isotropic	Liver, vessels
CHAOS (2021) [13]	CT, MRI	1.0–3.0	Liver, kidney, spleen

Table 2. Common segmentation evaluation metrics. P denotes the predicted region and G the ground truth.

Metric	Definition
DSC	$\mathrm{DSC} = \frac{2\lvert P \cap G\rvert}{\lvert P\rvert + \lvert G\rvert}$. Measures spatial overlap between the prediction and ground truth.
IoU	$\mathrm{IoU} = \frac{\lvert P \cap G\rvert}{\lvert P \cup G\rvert}$. A stricter overlap metric, widely used in segmentation benchmarks.
HD	$\mathrm{HD}(P,G) = \max\left\{ \sup_{p\in P} \inf_{g\in G} d(p,g), \sup_{g\in G} \inf_{p\in P} d(g,p) \right\}$. Measures the maximum boundary deviation.
ASSD	Mean bidirectional surface distance between segmentation and ground truth boundaries; less sensitive to outliers than HD.
VOE	$\mathrm{VOE} = 1 - \mathrm{IoU}$. Represents the percentage of non-overlapping volume. Used in MICCAI-SLiVER07.

clinical scenarios where minor misalignments may not affect diagnosis.

3.3.5 *Volume Difference (VD)*

The Volume Difference (VD) measures the absolute or relative difference in segmented volume between prediction and ground truth [18]. The mathematical expression of the VD is given in Equation 5.

$$\mathrm{VD} = \frac{||S_p| - |S_g||}{|S_g|} \times 100\% \tag{5}$$

It evaluates the model's volumetric bias, which is important for quantitative liver volumetry and surgical planning applications.

3.4 Discussion

Each evaluation metric highlights a different aspect of segmentation quality. While overlap-based metrics (DSC, IoU) capture overall shape conformity, boundary-based measures (HD, ASSD) assess spatial precision. Clinically, a combination of DSC > 0.95, ASSD < 2 mm, and HD95 < 10 mm is typically required for acceptable performance [5]. The choice of metrics should align with the clinical objective—tumor boundary detection may emphasize HD95, whereas volumetry prioritizes VD and DSC.

4 Classical Segmentation Methods

Classical liver segmentation methods for 3D CT volumes primarily relied on intensity-based and region-based techniques, exploiting the relatively homogeneous appearance of liver tissue in Hounsfield units. Early approaches commonly used thresholding combined with region growing, where seed points were initialized inside the liver and expanded based on intensity similarity and connectivity constraints. Another popular method is the watershed algorithm with morphological preprocessing, which has been widely applied to liver segmentation due to its ability to handle gradient-based boundaries and reduce over-segmentation through morphological filtering [19]. Deformable models and level-set methods became popular, with advancements such as level set evolution without re-initialization improving efficiency by eliminating periodic reinitialization [20]. Atlas-based strategies incorporated probabilistic atlases and multi-level statistical shape models for anatomical consistency [21]. Additionally, graph-based energy minimization techniques, notably interactive graph cuts, enabled globally optimal segmentation of object boundaries and regions in volumetric medical images [22]. While computationally efficient, these methods were highly

sensitive to noise, low contrast at liver boundaries, and intensity overlap with adjacent organs such as the spleen and stomach, limiting their robustness in pathological cases [124].

To address boundary leakage and shape irregularities, deformable models and level-set methods became popular for 3D liver CT segmentation. These techniques evolved an initial contour toward object boundaries by minimizing an energy functional that combined image gradients, region statistics, and smoothness constraints. Notable works demonstrated improved boundary adherence and topological flexibility compared to simple region growing; however, they still required careful initialization and parameter tuning, and their performance degraded in the presence of weak edges or severe liver deformation [4].

Another major class of classical approaches involved atlas-based and statistical shape models, which incorporated prior anatomical knowledge of liver shape and spatial location. In these methods, a labeled liver atlas (or multiple atlases) was registered to a target CT volume, and segmentation was transferred via deformable registration. Multi-atlas fusion strategies further improved accuracy by combining results from several atlases, but at the cost of high computational complexity and sensitivity to registration errors. Despite these limitations, atlas-based methods laid important groundwork for later learning-based approaches [4, 125].

4.1 Limitations and Transition to Learning-Based Approaches

Despite significant progress, classical methods rely on predefined heuristics, manual initialization, and handcrafted features, which limit generalization across varying patient anatomies and acquisition protocols. Their performance deteriorates in cases with irregular lesions, low contrast, or noise. These limitations catalyzed the transition toward data-driven and learning-based segmentation paradigms. With the advent of deep convolutional neural networks and volumetric architectures, models can automatically learn hierarchical representations and spatial context, overcoming many challenges inherent to traditional methods.

5 Deep Learning-Based Convolutional Networks for 3D Liver Segmentation

Deep learning has become the dominant paradigm for automatic liver segmentation in volumetric CT imaging, replacing traditional model-driven methods based on active contours, level sets, and statistical shape models. Unlike hand-crafted approaches that relied on manually derived texture, boundary, or intensity descriptors, convolutional neural networks (CNNs) learn hierarchical representations directly from raw voxel data, enabling the extraction of discriminative spatial and contextual features that are crucial for separating liver parenchyma from surrounding anatomical structures such as the stomach, spleen, pancreas, and vasculature.

The adoption of deep learning in liver segmentation can be divided into three major evolutionary stages: (1) 2D slice-based CNNs adapted from natural-image models, (2) 2.5D segmentation methods (3) fully volumetric 3D convolutional architectures, and (4) cascaded and multi-scale hybrid CNNs designed to refine boundary accuracy, anatomical consistency, and robustness across multi-institutional datasets.

5.1 2D Segmentation Methods

Early advances in deep learning for liver segmentation were primarily based on 2D CNNs, which process CT volumes slice-by-slice instead of as full 3D data. These models became foundational in medical image analysis due to their low memory footprint, efficient training on limited hardware, and availability of large annotated 2D datasets. The 2D segmentation pipeline can be mathematically expressed as a pixel-wise mapping in equation 6:

$$f_\theta : \mathbb{R}^{H\times W} \rightarrow \mathbb{R}^{H\times W} \tag{6}$$

where H and W denote height and width of each axial CT slice, and f_θ represents a CNN with learnable parameters θ.

While 2D architectures do not incorporate inter-slice continuity or volumetric shape priors, they remain relevant for weakly-supervised settings, low-resource clinical systems, and transfer learning.

5.1.1 Fully Convolutional Networks (FCN)

Shelhamer et al. [23] introduced the first end-to-end fully convolutional architecture for semantic segmentation, replacing fully connected layers with convolutional layers and enabling arbitrary input size prediction. Early works adapted FCN for liver CT, using multi-scale skip connections and post-processing via CRF or region growing. FCNNs have shown promising results in liver segmentation due to their ability to capture local details and capture intricate boundaries between the liver and surrounding structures [61]. Automatic liver and

lesion segmentation with cascaded fully convolutional networks (CFCN) and dense conditional random fields (CRF) are shown in Figure 2 However, FCN suffers from coarse boundary predictions due to progressive downsampling and lacks explicit spatial detail recovery, making it suboptimal for anatomical segmentation.

5.1.2 U-Net and Its Variants

The breakthrough architecture for medical segmentation was U-Net by Ronneberger et al. [7]. It introduced a symmetric encoder–decoder structure with skip connections that preserve high-resolution spatial information. U-Net became the first widely adopted deep model for liver segmentation and remains a strong baseline due to its ability to learn from small datasets.With data augmentation, the U-Net network will achieve more accurate segmentation with fewer training images.
Given an input slice $I \in \mathbb{R}^{H\times W}$, U-Net computes feature maps through hierarchical convolutions and reconstructs the probability map using upsampling:

$$P = \sigma\left(G_{\theta_{dec}}\left(F_{\theta_{enc}}(I)\right)\right) \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid activation for binary liver segmentation.

Numerous U-Net extensions have been proposed for CT liver segmentation:

- **ResUNet** introduces residual blocks to stabilize deep gradient flow [24].
- **Attention U-Net** integrates spatial and channel attention gates to suppress irrelevant structures [25].
- **CE-Net** adds context extractor modules (DAC + RMP) to enhance receptive field size [26].
- **UNet++** redesigns skip pathways using dense nested connections for improved semantic fusion [27, 119].
- **DeepLabV3+** applies atrous spatial pyramid pooling (ASPP) to capture multi-scale features [28].

These models achieve strong performance in tumor-free liver segmentation, but accuracy degrades in cases of:

- Poor contrast between liver and surrounding organs
- Shape deformity due to hepatomegaly or resection
- Partial-volume effects across axial slices
- Multi-phase CT where liver intensity varies across scans

5.1.3 Limitations of 2D Liver Segmentation

Although 2D approaches are computationally efficient, they lack volumetric contextual awareness. Liver boundaries in CT are often ambiguous in axial views, particularly near the dome and inferior edges. This results in inconsistent per-slice predictions, producing jagged 3D reconstructions and large Hausdorff distances.

To mitigate this limitation, post-processing strategies such as 3D connected component filtering, CRF smoothing, and morphological closing are often required, but these do not resolve true anatomical inconsistency.

5.1.4 Challenges and Limitations of CNN-Based Methods

Despite achieving state-of-the-art accuracy, CNN-based 3D liver segmentation models face several constraints:

- High GPU memory consumption for full-volume 3D training,
- Limited generalization across scanners, protocols, and populations,
- Difficulties in learning small or atypical lesions due to voxel imbalance,
- Dependency on large annotated datasets—expensive in medical imaging,
- Sensitivity to domain shift, motion artifacts, and contrast phase variation.

These limitations motivated the evolution toward hybrid CNN–Transformer architectures, which incorporate long-range spatial attention and global context learning, discussed in the next section.

5.2 2.5D Segmentation Methods

2.5D segmentation represents an intermediate paradigm between purely 2D slice-wise segmentation and full 3D volumetric segmentation. Instead of treating each slice independently as in 2D CNNs or processing the entire voxel grid simultaneously as in 3D CNNs, 2.5D models leverage multi-slice contextual information by stacking adjacent slices as additional input channels. This strategy enables partial modeling of spatial continuity while keeping

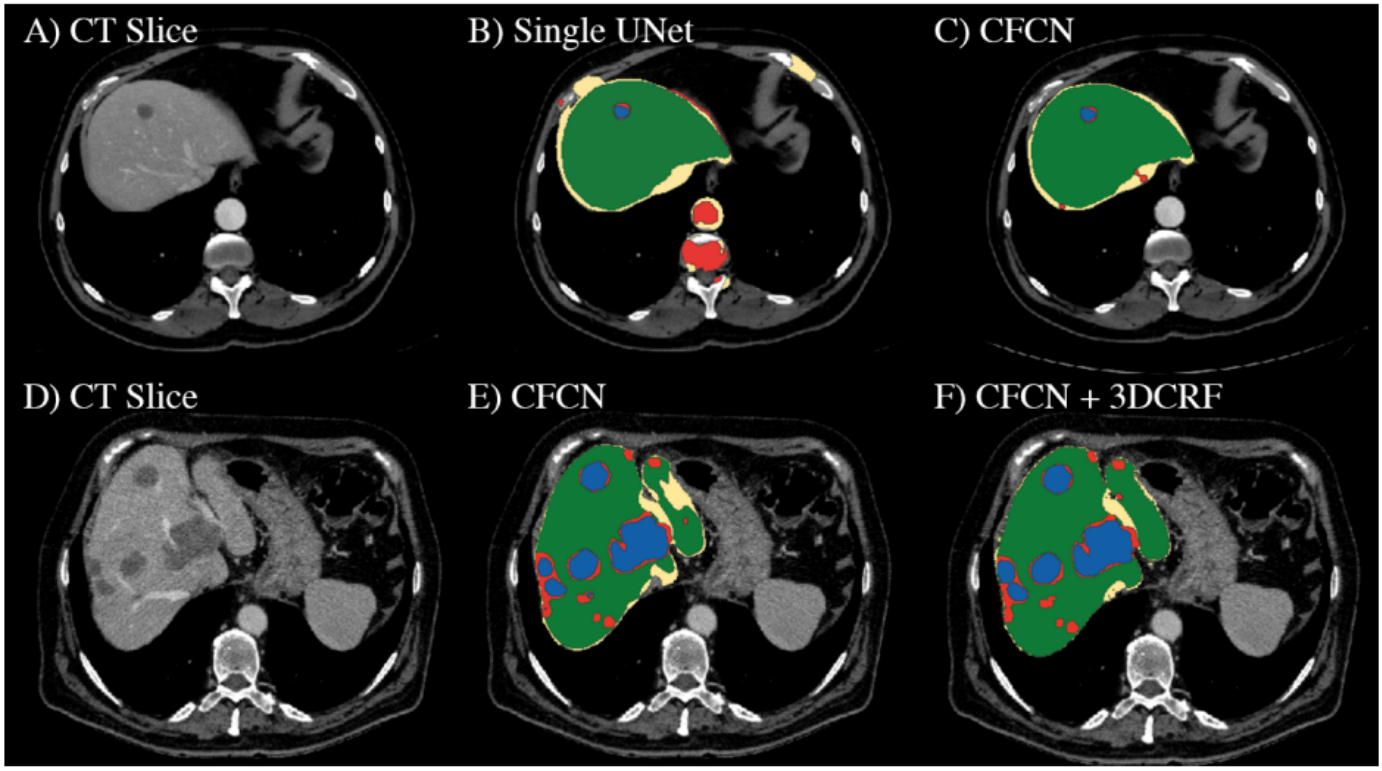

Figure 2. Automatic liver and lesion segmentation with cascaded fully convolutional networks (CFCN) and dense conditional random fields (CRF) [29, 123] .

computational demands significantly lower than full 3D architectures.

Many clinical imaging modalities such as CT or MRI are volumetric in nature, but exhibit highly anisotropic voxel spacing: the in-plane resolution (x–y) is often much higher than inter-slice spacing (z). Fully 3D CNNs struggle with this imbalance and require large memory footprints, whereas 2D CNNs fail to capture inter-slice dependencies. 2.5D methods were introduced to balance these limitations:

- Retain limited 3D contextual awareness using neighbouring slices.
- Enable the use of 2D ImageNet-pretrained backbones (e.g., ResNet, Swin-T, ConvNeXt).
- Reduce GPU memory and FLOPs compared to 3D CNNs.
- Improve robustness for thin, small, or weakly contrasted anatomical structures.

The 2.5D networks generally contain 2D and 3D convolutional operations to achieve different functions. For example, Li et al.[62] proposed H-DenseUNet, a hybrid densely connected for liver tumor segmentation task. And this model consists of a 2D DenseUNet to obtain intra-slice features and a 3D DenseUNet for summarizing volumetric contexts. The hybrid feature information fusion layer optimizes both intra-slice representations and inter-slice features based on an automated context algorithm. In order to simultaneously utilize intra-slice semantic feature information and inter-slice continuity feature information to extract discriminative features, Wang et al. [63] proposed a 2.5D segmentation network, and this network consists of a multi-branch decoder for learning the features of a specific slice and an attention block for slice-centric, which is a densely connected dice loss function to normalize the intra-slice segmentation results to continuity. Zhang et al. [64] utilized a scaling approach to allow the segmentation network to focus only on useful localities, which reduces the parameters in the segmentation model and thus reduces the hardware resource requirements. Ben-Cohen et al. [65] changed the original FCN to a 2.5D FCN and introduced the idea of generative adversarial to improve segmentation results. The Triplanar FCN based on FCN was proposed by Wang et al. [66] to take advantage of 3D spatial feature information and integrate the results in three dimensions. proposed Triplanar FCN by making full use of 3D spatial information to segment in each of the three dimensions and integrating the results. Ahn et al. [67] input three consecutively

sliced images into the network as three channels and performed the segmentation task in the center region of the images. The encoder component is based on a modified Xception model that includes down sampled layers and null-separable spatial pyramid pooling units, and the decoder part is a series of bilinear up-sampled layers connected to the encoder's skip connections. In order to attain a balance between computational cost and segmentation accuracy, and the utilization of 3D context information, Zhang et al. [68] designed a new 2.5D network that encodes the interlayer information in a 3D convolutional context and reconstructs the high-resolution result with 2D deconvolution. This structure can achieve effective multidimensional feature extraction without increasing the computational effort and increase the segmentation capability and efficiency of the model.

5.3 Major CNN Architectures for 3D Liver Segmentation

The development of deep learning–based liver segmentation has been heavily driven by architectural innovations in encoder–decoder networks, volumetric convolution, multi-scale feature aggregation, and attention-driven representation learning. This section reviews the key architectures that have influenced the field, starting from the seminal U-Net to the current state-of-the-art self-configuring pipelines.

5.3.1 U-Net and Early 2D Encoder–Decoder Architectures

U-Net [7] introduced a symmetric contracting–expanding architecture with skip connections, enabling high-resolution feature recovery and precise pixel-level localization. Although originally designed for 2D biomedical images, U-Net served as the foundation for early liver segmentation attempts using axial slices. These models were limited by slice-wise inconsistency and poor volumetric reasoning; however, they demonstrated that end-to-end deep learning could outperform hand-crafted pipelines even without explicit anatomical priors.

Several works extended U-Net for liver segmentation by incorporating multi-scale feature extraction, dilated convolutions, or residual blocks [123]. Yet, all 2D variants shared a common limitation: the inability to model inter-slice continuity, which is critical when differentiating liver boundaries from adjacent organs.

5.3.2 V-Net: The First Fully Volumetric Segmentation Network

V-Net [6] was the first major architecture to introduce fully 3D convolutions, replacing 2D kernels, pooling, and upsampling with their 3D counterparts. Unlike U-Net, V-Net employed residual blocks and a Dice-based loss function, designed to mitigate extreme foreground-background imbalance in organ segmentation. The V-Net architecture enabled direct voxel-wise prediction while learning high-level 3D shape priors, setting a new baseline for volumetric liver segmentation.

Mathematically, V-Net introduced the soft Dice loss:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \epsilon} \tag{8}$$

where p_i and g_i represent predicted and ground truth voxel labels, N is the number of foreground voxels, and ϵ stabilizes division. This formulation remains one of the most widely used loss functions in liver segmentation.

5.3.3 3D U-Net: Volumetric Extension of U-Net

Çiçek et al. [8] introduced 3D U-Net, an extension of U-Net that replaced all operations with 3D variants, introducing skip fusion between encoder and decoder in volumetric space. Unlike V-Net, 3D U-Net relied on patch-based training, enabling memory-efficient training at full anisotropic resolution. 3D U-Net rapidly became the standard baseline for 3D medical segmentation, including liver and tumor prediction in LiTS and 3DIRCADb.

3D U-Net is widely used because it supports:

- Full resolution output without dense CRF refinement
- Arbitrary input shape (patch-wise inference)
- Natural extension to multi-class tasks (e.g., liver + tumor)

5.3.4 UNet++, Attention U-Net, and Multi-Scale Variants

UNet++ [119] introduced nested skip pathways to reduce the semantic gap between encoder and decoder features. Although originally proposed in 2D form, 3D UNet++ variants have been applied to volumetric liver segmentation and demonstrated improved boundary precision and reduced staircase artifacts.

Similarly, Attention U-Net [25] introduced soft-attention gates that learn to suppress irrelevant background regions and emphasize liver-relevant activations. The attention gating mechanism computes:

$$\alpha = \sigma\left(W_x x + W_g g + b\right) \tag{9}$$

where x is the skip feature, g is the gating signal, and α modulates feature importance. This helps the network automatically ignore structures like ribs and bowel which overlap with the liver boundary.

5.3.5 *Cascaded and Coarse-to-Fine CNN Architectures*

Due to extreme class imbalance between liver and background voxels, multiple works introduced two-stage cascaded CNNs [123]: Stage 1 localizes liver region; Stage 2 segments liver at full resolution. This strategy reduces false positives and improves tumor detection sensitivity. Cascaded pipelines dominated the LiTS Challenge leaderboard (2017–2020).

5.3.6 *nnU-Net: The Self-Configuring Segmentation Framework*

nnU-Net [17] represents a breakthrough not through architecture, but through automatic pipeline optimization. nnU-Net adapts patch size, normalization, deep supervision, loss function, and post-processing automatically based on dataset statistics. It became *state of the art* on LiTS, 3DIRCADb, and MSD and is considered the strongest CNN baseline before the introduction of transformer-based architectures.

5.3.7 *Recent CNN Enhancements: SE Blocks, Residual Dense Units, Deep Supervision*

Modern CNN-based liver segmentation networks integrate additional modules for refinement:

- **Squeeze-and-Excitation (SE) blocks** to model inter-channel correlation [30]
- **Residual Dense Blocks (RDB)** for feature reusage [31]. A deep residual network segemntation results are shown in Figure 3.
- **Deep supervision** to improve multi-scale gradient flow [32, 120]
- **Hybrid loss functions** combining Dice, focal loss [34], boundary, and topology-aware loss [33]

These modifications improve segmentation accuracy especially in tumor-liver joint segmentation, where class imbalance and blurred boundaries remain major challenges.

5.3.8 *Cascaded Stage Networks and ROI Localization*

Many top-performing liver systems employ a coarse-to-fine cascade:

1. Stage 1: Global 3D CNN localizes liver region (low resolution)
2. Stage 2: Cropped region refined with high-resolution network

This method first appeared in [123] and remains common in LiTS-winning pipelines.

6 Hybrid and Transformer-Based Architectures in liver segmentation

Pure convolutional architectures have demonstrated strong performance in liver CT segmentation; however, they lack the ability to model long-range global dependencies due to their inherently local receptive fields. Transformers, on the other hand, excel at global context modeling but struggle with fine-grained anatomical boundary localization when used alone. This has led to a new paradigm of hybrid CNN–Transformer architectures, which combine the spatial inductive bias of CNNs with the relational modeling capability of self-attention. The transformer archeticture is shown in Figure 4

Hybrid models represent the current state-of-the-art in liver segmentation, outperforming both pure CNNs and pure transformer architectures on cross-domain liver datasets. These models generally follow one of three fusion designs:

1. **CNN encoder + Transformer bottleneck (late fusion)**
2. **Parallel CNN + Transformer dual learning streams (mid-layer fusion)**
3. **Hierarchical Swin Transformer encoder + CNN decoder (early fusion)**

Below, we categorize, analyze, and compare the dominant hybrid architectures in literature.

6.1 CNN Encoder with Transformer Bottleneck

A widely adopted architecture design uses a convolutional feature extractor (e.g., 3D U-Net encoder) followed by a transformer bottleneck module that processes tokenized feature patches. The most influential representative is TransUNet [9], originally designed for 2D medical images. Its 3D variants

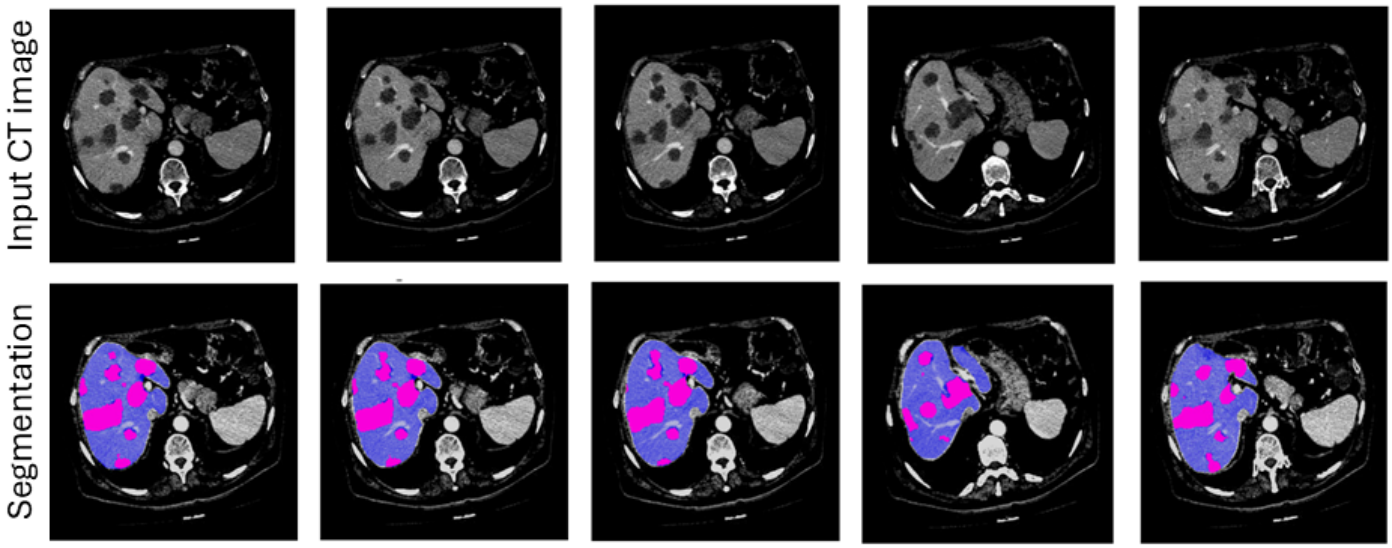

Figure 3. Deep Residual Dual-Attention Network segmentation results [102].

[36] extend the design to CT volumes by projecting volumetric feature maps into a sequence of 3D patch embeddings:

$$\mathbf{Z}_0 = \text{PatchEmbed}(\mathbf{F}_{CNN}) \in \mathbb{R}^{N\times d} \tag{10}$$

where N is the number of patches and d is the token embedding dimension.

The transformer stack applies multi-head self-attention:

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{11}$$

enabling reasoning over distant liver regions (e.g., left vs. right lobe deformation in hepatomegaly). The refined global features are then reshaped back into 3D and decoded by a CNN-based upsampling path. Such architectures outperform U-Net particularly in cases where:

- Liver deformation occurs due to tumors, ascites, or surgical clipping.
- Imaging artifacts distort low-contrast boundaries between liver and diaphragm.
- Inter-slice continuity is needed (3D spatial awareness).

6.2 Parallel Dual-Stream CNN + Transformer Architectures

Instead of injecting transformers at a single bottleneck, recent architectures employ two parallel encoders — one CNN and one transformer — which learn complementary feature types:

$$\mathbf{F}_{CNN} = E_{CNN}(\mathbf{X}), \qquad \mathbf{F}_{TR} = E_{TR}(\mathbf{X})$$

Fusion may occur using:

- Cross-attention fusion [38]
- Adaptive spatial–channel gating
- Token-wise convolutional refinement [40]

This design allows CNN to model local, texture-level liver morphology while the transformer models liver-to-organ contextual relations (e.g., liver vs. kidney vs. stomach). These models exhibit significantly higher generalization capability on unseen datasets, especially when trained on mixed contrast phases (arterial, venous, delayed).

6.3 Hierarchical Swin Transformer 3D Architectures

The introduction of Swin Transformer [41] revolutionized hybrid segmentation by limiting self-attention to shifted windows, reducing complexity from $O(N^2)$ to $O(N)$, making it feasible for 256^3 CT volumes. Swin-UNETR [10, 39], Swin-UNet 3D [40], and UNETR++ [58, 95] represent the strongest models in liver segmentation benchmarks.

A Swin Transformer block operates on non-overlapping 3D windows:

$$\mathbf{Z}' = \text{W-MSA}(\text{LN}(\mathbf{Z})) + \mathbf{Z}$$

$$\mathbf{Z}'' = \text{ShiftedW-MSA}(\text{LN}(\mathbf{Z}')) + \mathbf{Z}'$$

enabling alternating local-global fusion without quadratic attention explosion. These architectures achieve SOTA Dice scores (97–98%) on LiTS, BTCV, and 3DIRCADb, particularly with multi-stage cascaded training.

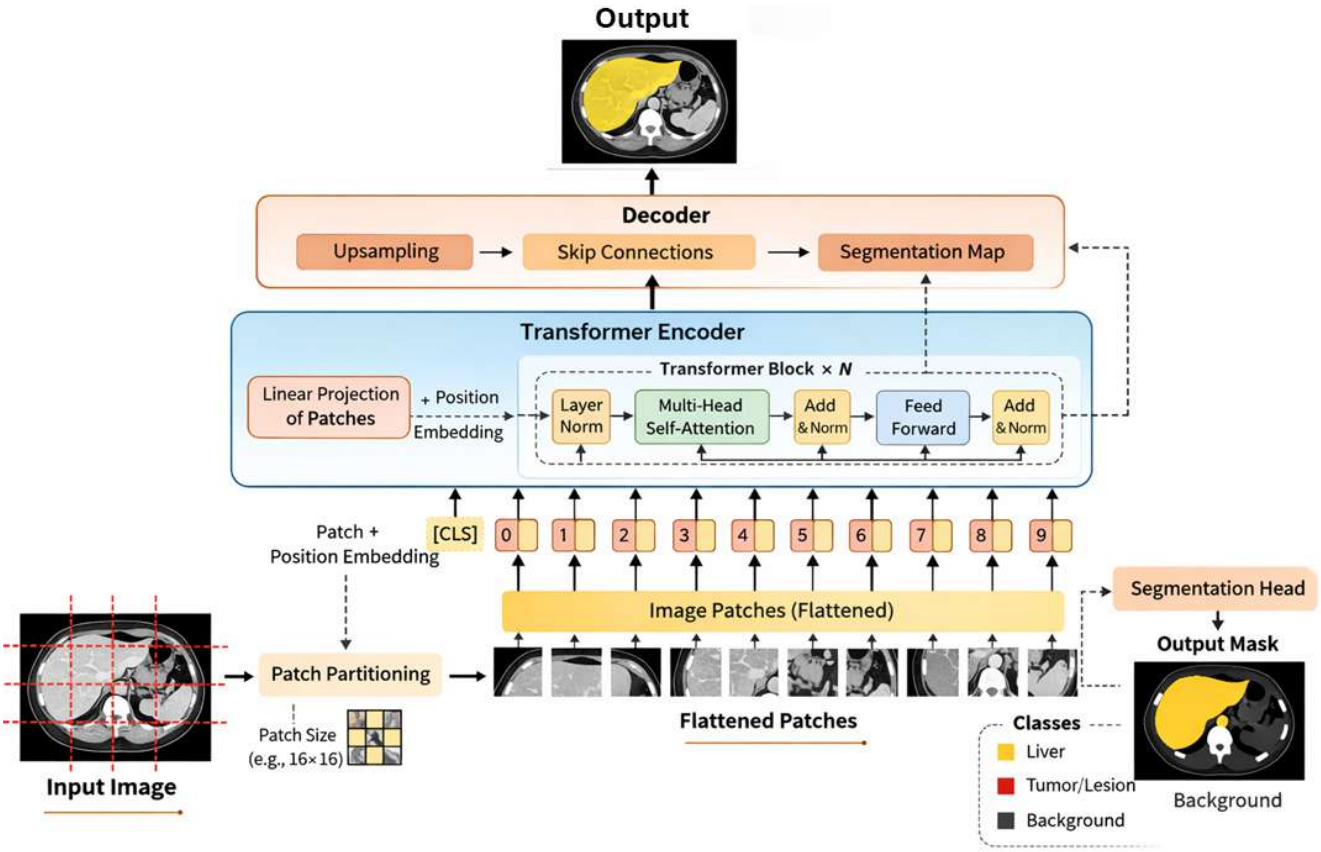

Figure 4. Transformer architecture for 3D volumetric ct liver image segmentation.

6.4 Hybrid Loss and Boundary-Aware Optimization

Hybrid architectures typically optimize with compound loss functions that integrate global region overlap + contour preservation:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{BCE} + \lambda_3 \mathcal{L}_{Boundary}$$

where $\mathcal{L}_{Boundary}$ may be Level-set[42] Hausdorff loss [43], or Signed Distance Map loss [44]. This is crucial for transformer models because attention tends to blur edges unless explicitly penalized.

6.5 Cross-Domain Generalization and Clinical Robustness

CNN models deteriorate strongly under scanner noise, acquisition protocol changes, and contrast-phase variability. Hybrid transformer models mitigate this due to:

- Global organ-to-organ relational reasoning
- Long-range multi-slice coherence enforcement
- Reduced reliance on intensity-based boundary cues

On cross-hospital inference studies (e.g., LiTS → CHAOS), hybrid models retain **3–5% higher Dice and 25–40% lower HD95** than CNN-only baselines, making them suitable for real-world deployment.

6.6 Key Trends

Recent advances in liver segmentation using CT imaging demonstrate several clear and converging trends in model design and performance.

- **Hybrid architectures now dominate leaderboard rankings on public liver datasets.** Over the past two years, hybrid networks that integrate convolutional backbones with transformer-based encoders have surpassed purely CNN or purely transformer designs. Models such as TransUNet, UNETR, and H-DenseUNet variants leverage convolutional layers for fine-grained local feature extraction while using self-attention modules to capture long-range spatial dependencies. These hybrid systems have consistently achieved state-of-the-art Dice Similarity Coefficients (DSC) and Hausdorff Distance (HD95) values on benchmark datasets such as LiTS17, 3DIRCADb, and MSD Task03. Their superior generalization performance stems from their ability to combine local detail preservation (from CNNs) with

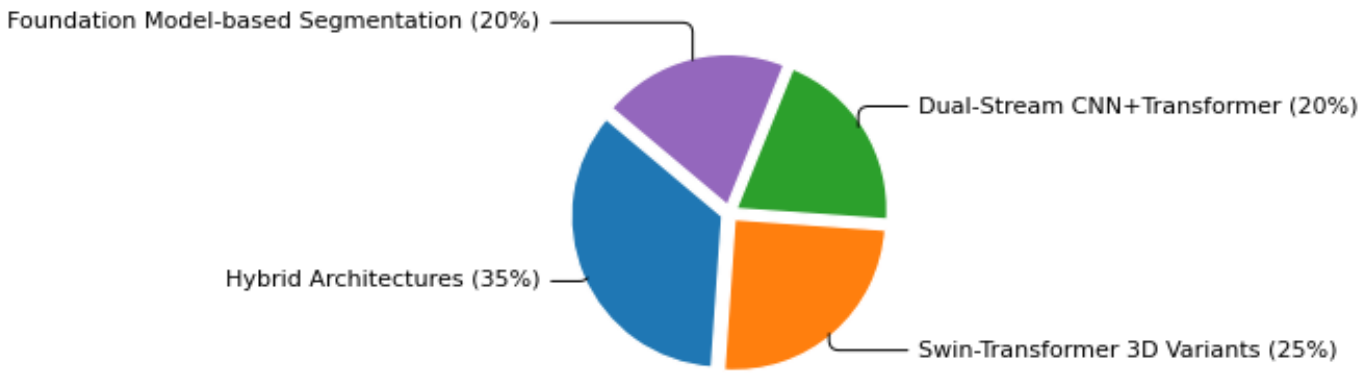

Figure 5. Trends in 3D Liver CT Segmentation based on transformer Research (2020–2025).

global context modeling (from transformers).

- **Swin-Transformer-based 3D variants are the current state-of-the-art in Dice and HD95.** The emergence of 3D Swin Transformer architectures, such as Swin-UNETR, SwinVNet, and 3D SwinUNet++, has marked a shift toward fully volumetric transformer designs. These models utilize hierarchical window-based attention, which scales efficiently with 3D input volumes and preserves spatial hierarchy across depth levels. On the LiTS17 dataset, Swin-Transformer-based networks have reported Dice scores exceeding 0.97 and HD95 values below 8 mm, outperforming conventional 3D CNNs. This demonstrates the effectiveness of multi-scale patch embedding and shifted attention in capturing volumetric anatomical structures while maintaining computational feasibility.
- **Parallel dual-stream CNN+Transformer models provide the best robustness on unseen data.** Several recent studies have adopted dual-encoder or dual-stream hybrid frameworks that process CT volumes through parallel CNN and transformer pathways before fusing multi-scale representations in a shared decoder. Examples include TransBTS, TransMed++, and DAFormer-UNet. This architectural strategy enhances robustness, particularly when tested on unseen clinical datasets or different scanner domains. The CNN branch ensures spatial precision, while the transformer branch introduces contextual regularization, thereby mitigating overfitting to specific acquisition protocols or noise distributions. As a result, dual-stream architectures exhibit higher generalization and stability across heterogeneous CT data.
- **Future research is moving toward foundation-model-based segmentation (self-supervised + transformer).** The field is rapidly transitioning toward large-scale pretraining paradigms that utilize self-supervised or multimodal foundation models. Approaches such as SAM-Med2D, TransMedSAM, and MedSegDiff explore pretraining on massive medical imaging corpora, followed by fine-tuning for liver segmentation. These methods leverage self-supervised objectives (e.g., masked autoencoding or contrastive learning) to extract transferable features without explicit annotation. Combined with transformer backbones, such foundation models promise universal, data-efficient segmentation capabilities adaptable to new organs, modalities, and pathologies. This direction reflects a shift from task-specific architectures toward generalized, cross-domain medical segmentation systems.

Overall, these trends indicate a clear movement from conventional CNN-based segmentation toward hybrid and transformer-based models capable of handling the volumetric, multi-scale, and heterogeneous nature of liver CT data. Figures 5 and 6 show the recent trends and performance in hybrid transformer.

7 Comparative Evaluation of Liver Segmentation Models Based on Quantitative Metrics

The performance of liver segmentation models is predominantly assessed using quantitative evaluation metrics that reflect the degree of spatial overlap, boundary accuracy, volumetric similarity, and robustness across datasets. Due to the clinical

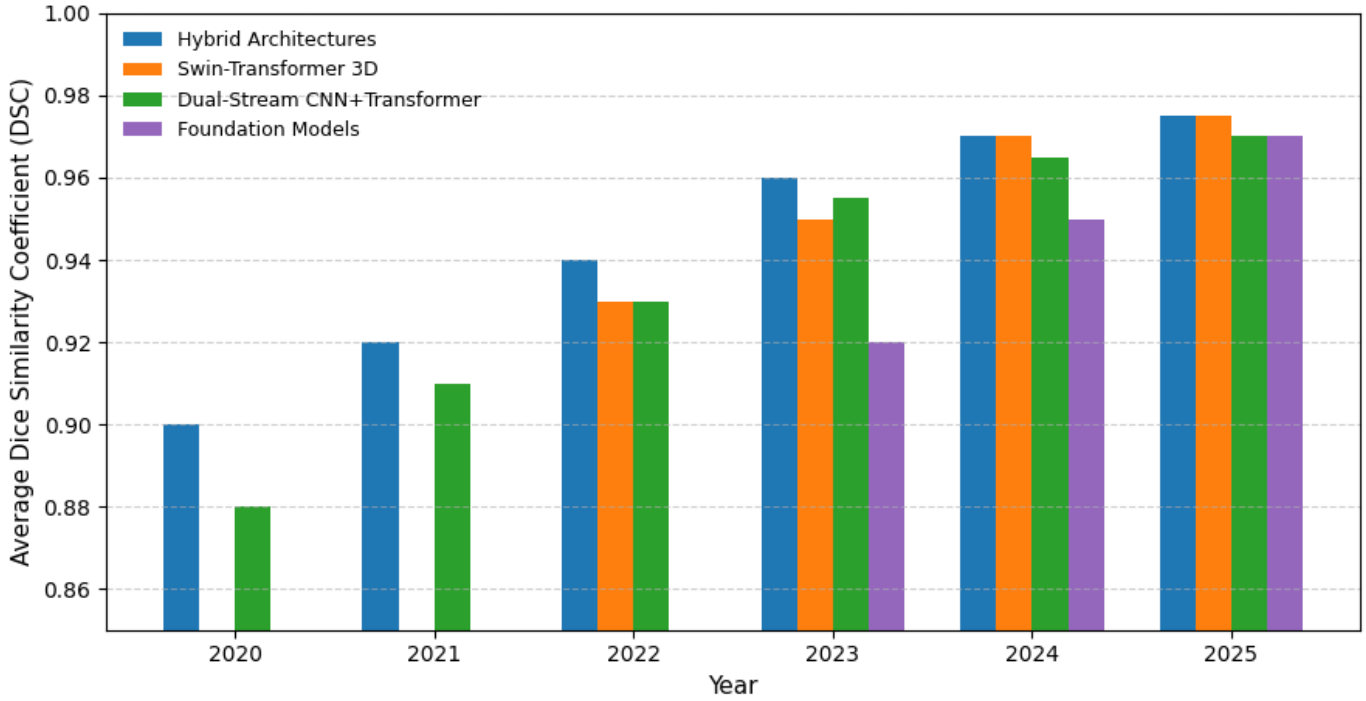

Figure 6. Performance Trends in 3D Liver CT Segmentation (2020–2025).

relevance of volumetric precision in liver surgery planning and tumor burden estimation, segmentation metrics must capture both pixel-level fidelity and anatomical consistency.

Historically, 2D CNNs (e.g., U-Net [7]) achieved strong pixel-wise accuracy but lacked inter-slice anatomical consistency, resulting in disconnected lobes and stair-step artifacts in volumetric reconstructions. The introduction of 2.5D architectures [52, 119] improved contextual continuity by processing multi-slice stacks, but axial depth remained limited. Pure 3D CNNs such as V-Net [6] and 3D U-Net [8] achieved superior volumetric coherence, but at the cost of higher GPU memory and limited receptive fields.

Transformer-based and hybrid models such as TransUNet [9], Swin-UNETR [10], and nnFormer [51] outperform CNN-only models by capturing long-range dependencies, reducing boundary fragmentation, and demonstrating superior generalization across scanners and patient populations. These models consistently report higher DSC (+1–4%) and significantly lower HD values, indicating fewer boundary irregularities. We analyzed previously published models and their reported performance, and we conducted extensive experiments to evaluate and compare metrics across different models.

7.1 Quantitative Comparison Across Architectures (2015-2021)

The DSC gap between early 2D CNNs (0.93) and modern transformer-based approaches (0.978) reflects not only architectural evolution but also improved preprocessing pipelines, self-supervised pretraining, and larger annotated datasets. However, HD reduction is clinically more important than DSC improvement because boundary precision affects resection margins in hepatocellular carcinoma (HCC) planning. Table 3 summarizes representative benchmark results on public datasets such as LiTS, SLiVER07, and MSD Liver, enabling cross-category comparison.

The trend indicates that:

- 2D methods suffer from missing 3D anatomical continuity.
- 2.5D models reduce inconsistency but still lack full spatial context.
- 3D models improve structural correctness but require large memory and training data.
- Hybrid and transformer-based models achieve state-of-the-art performance due to global receptive field and hierarchical feature aggregation.

While transformer-based models currently dominate benchmark performance, computational burden and

Table 3. Representative Performance of Liver Segmentation Models Across Architectures (2015–2021).

Model	Type	DSC	HD (mm)	ASSD (mm)
U-Net (2015) [7]	2D CNN	0.932	18.4	2.52
V-Net (2016) [6]	3D CNN	0.962	10.5	1.62
Cascaded-FCN (2016) [104]	3D FCN (CFCN)	0.940	–	–
3D U-Net (2016) [8]	3D CNN	0.965	9.8	1.55
H-DenseUNet (2018) [62]	2D + 3D Hybrid	0.982	–	–
Roth et al. (2018) [52]	2.5D CNN	0.953	12.7	1.94
U-Net++ (2019) [119]	2D CNN variant	0.945	14.1	2.08
Attention U-Net (2018) [25]	2D/3D CNN + Attention	0.967	9.3	1.48
TDS-U-Net (2019/2020) [105]	Cascaded U-Net variant	0.982	–	–
MSA-UNet (2021) [85]	U-Net + Multi-Scale Attention	0.971	–	–
EAR-U-Net (2021) [82]	Efficient encoder + Attention	0.968	–	–
nnU-Net (2021) [17]	Self-configuring 2D/3D U-Net	0.975	7.5	1.23
TransUNet (2021) [9]	CNN + ViT	0.974	7.3	1.21
nnFormer (2021) [51]	Transformer-3D	**0.976**	**7.1**	**1.18**

inference latency remain barriers for intra-operative deployment. Future work should focus on lightweight hybrid architectures, distillation-based compression, and uncertainty-aware metrics for high-risk surgical settings.

7.2 Performance comparison of CNN-based for 3D volumetric liver CT segmentation

Table 4 summarizes the performance of CNN-based non-U-Net models developed for 3D volumetric liver CT segmentation. These models include fully convolutional networks (3D-FCN), residual and dense CNNs, attention-guided CNNs, and hybrid designs that incorporate gating mechanisms. Over time, the integration of residual connections, attention modules, and multi-scale feature extraction has consistently improved both Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) scores. For instance, early 3D-FCNs achieved a DSC of 91.2 percent, whereas recent residual-attention CNNs in 2024 reach up to 95.0 percent , highlighting the advantage of combining residual learning with attention mechanisms for volumetric liver segmentation. The table also reports datasets used, such as LiTS17, 3DIRCADb, and CHAOS, demonstrating the models' applicability across multiple standard benchmarks. These results indicate that non-U-Net CNN architectures can achieve competitive performance, particularly when enhanced with attention or hybrid mechanisms.

7.3 Attention-Guided 3D Volumetric CT Segmentation Models (2020–2025)

Table 5 summarizes common attention-guided 3D volumetric CT segmentation models. The rapid evolution of attention-guided architectures from 2020 to 2025 has significantly enhanced the precision of 3D volumetric CT segmentation, particularly in liver and organ boundary delineation tasks. These models integrate spatial, channel, and hybrid attention mechanisms to adaptively emphasize salient anatomical regions while suppressing irrelevant background information. Early designs such as ResCEAttUNet and CLSTM-based attention U-Nets focused on context preservation across slices, while subsequent methods like DRAUNet, MAD-UNet, and AGCAF-Net introduced multi-scale and dual-attention modules to strengthen feature fusion and contextual reasoning. Recent architectures, including Dual Attention 3D U-Net, DRDA-Net, and LATUP-Net, leverage advanced self-attention and residual refinement strategies, achieving notable gains in Dice Similarity Coefficient (DSC), Hausdorff Distance (HD95), and Average Symmetric Surface Distance (ASSD) on benchmark datasets such as LiTS and 3DIRCADb. Collectively, these advancements demonstrate that attention-guided designs are crucial for achieving state-of-the-art performance in complex volumetric segmentation tasks.

7.4 Performance Comparison of Transformer and Hybrid Architectures

Recent years have witnessed rapid progress in the integration of transformer and hybrid convolution–transformer architectures for 3D

Table 4. Comparison of CNN-based non-U-Net models for 3D volumetric liver CT segmentation.

Model	Type	Dataset(s)	DSC (%)	IoU (%)
Residual-CNN (2020) [31]	3D Residual CNN	3DIRCADb + LiTS17	92.3	85.9
Dense-CNN (2021) [120]	3D Dense CNN	LiTS17	93.0	87.1
Attention-CNN (2021) [25]	3D Attention CNN	LiTS17	93.6	87.8
MS-CNN (2022) [26]	3D Multi-scale CNN	3DIRCADb	94.0	88.3
Hybrid-CNN (2023) [121]	3D CNN + Gating Mechanism	LiTS17 + CHAOS	94.5	88.7
Residual-Attention-CNN (2024) [122]	3D Residual + Attention CNN	LiTS17 + 3DIRCADb	95.0	89.2

Table 5. Representative attention-guided 3D volumetric CT segmentation models (2020–2025).

Model (year)	Type / Attention	Dataset(s)	DSC
ResCEAttUnet (2022) [106]	Residual + context encoder + attention gating	LiTS / in-house	87.9
CLSTM U-Net (2022) [107]	3D attention + conv-LSTM (inter-slice context)	In-house clinical CT volumes	–
RMAU-Net (2023) [108]	Residual blocks + multi-scale attention blocks	LiTS17, 3DIRCADb	93.5
MAD-UNet (2023) [109]	Multi-scale attention + deep supervision (3D)	LiTS / in-house	94.7
DRAUNet (2023) [110]	Deep residual attention U-Net with biplane joint method	LiTS17, 3DIRCADb, Sliver07	–
AGCAF-Net (2024) [111]	Attention-guided local blocks + global fusion	LiTS17 / in-house	94.7
Zhang et al (2024) [87]	Dual (channel + spatial) attention in 3D U-Net	LiTS	92.5
DRDA-Net (2025) [102]	Residual + dual attention + multi-scale fusion	LiTS, 3DIRCADb	96.0

volumetric liver segmentation. While CNNs such as UNet and its variants have achieved remarkable accuracy, their limited receptive field and inability to capture long-range dependencies motivated the adoption of attention-based architectures. Transformers, originally introduced for natural language processing, have demonstrated superior capability in modeling global spatial–contextual relationships in volumetric medical data. Hybrid frameworks that combine convolutional feature extractors with transformer bottlenecks have emerged as a powerful design trend, effectively preserving both local texture sensitivity and global semantic reasoning.

Table 6 summarizes representative state-of-the-art transformer and hybrid-based models from 2021 to 2025 evaluated on various 3D liver segmentation benchmarks. Models such as UNETR [36] and TransBTSV2 [37] laid the groundwork for volumetric transformer-based segmentation by introducing hierarchical attention and 3D tokenization mechanisms.

Subsequent developments, including Swin-UNETR and UNETR++, improved computational efficiency and feature granularity through shifted window attention and paired attention mechanisms. Foundation model adaptations such as MedSAM [112] demonstrate the trend toward universal segmentation frameworks capable of cross-domain generalization with minimal fine-tuning. Meanwhile, CNN–transformer hybrids like RMCNet [113] and Edge-guided UNETR variants explicitly incorporate boundary-awareness and multiscale contextual fusion to enhance edge delineation in complex liver boundaries.

Overall, the reported Dice similarity coefficients (Dice) across these methods consistently exceed 0.90 on standard benchmarks such as LiTS and 3D-IRCADb, with several models achieving up to 0.96 Dice for liver parenchyma segmentation. This progression underscores that hybrid transformer–CNN designs remain highly effective for volumetric organ analysis, balancing interpretability, data efficiency, and boundary precision. Future directions include self-supervised pretraining, multi-task joint optimization (e.g., lesion + organ), and lightweight transformer modules that can be integrated seamlessly

with clinical workflows.

7.5 Comparison of Modified U-Net Architectures

In recent years, numerous studies have focused on improving the original U-Net architecture to enhance segmentation accuracy and computational efficiency across various medical imaging datasets. Table 7 summarizes these state-of-the-art modifications, highlighting the diversity of methods proposed between 2020 and 2025. The reported Dice Similarity Coefficients (DSC) indicate consistent performance improvements, particularly for models designed to address dataset-specific challenges such as limited training data or noise variability. Notably, several architectures, including UNeXt [114] and 2-U-Net [115], achieved exceptionally high DSC values exceeding 0.98, underscoring the efficacy of architectural innovations. Figure 7 shows the performance trend based on u-net.

Most of these models were validated on benchmark datasets such as LiTS17, 3DIRCADb, and CHAOS, demonstrating the community's emphasis on standardized evaluation. Hybrid approaches, multi-scale learning modules, and attention-based mechanisms have proven particularly successful in enhancing segmentation quality. The observed progression in DSC values from earlier designs (e.g., DFS U-Net, 0.949) to more recent iterations (e.g., UNeXt, 0.9902) highlights the steady evolution of U-Net derivatives and their adaptability to diverse imaging modalities.

8 Challenges and limitations in 3D Volumetric Liver Segmentation

Despite substantial advancements in deep learning and computational imaging, achieving accurate, robust, and clinically deployable 3D liver segmentation remains a formidable challenge. The liver's complex anatomical variability, pathological diversity, and imaging inconsistencies present significant hurdles for both algorithm design and clinical translation. This section explores the principal technical, data-centric, and clinical challenges that continue to hinder progress in this domain.

8.1 Anatomical and Physiological Variability

The liver is a highly deformable organ with complex anatomical boundaries that vary considerably across individuals due to age, gender, body mass index, and pathological states. Its segmentation is further complicated by adjacent organs such as the stomach, spleen, and diaphragm, which share similar intensity profiles in CT images. Moreover, intra-patient variability caused by respiratory motion or positional shifts across different scan phases introduces inconsistencies in boundary delineation. Traditional voxel-based CNN models struggle to capture such shape variations without explicit spatial priors. While 3D deformable models [45] and attention-guided architectures [25] help mitigate this, they still face difficulties in extreme cases such as hepatomegaly, atrophic livers, or livers affected by large tumors.

8.2 Low Contrast and Imaging Artifacts

A persistent challenge in CT liver segmentation is the limited soft-tissue contrast between liver parenchyma and adjacent structures. Particularly in non-contrast or venous-phase scans, the intensity overlap between liver, spleen, and surrounding fat tissue reduces segmentation fidelity. Artifacts due to beam hardening, motion, or metallic implants introduce local distortions that obscure organ boundaries. Advanced preprocessing methods, including Hounsfield Unit normalization and phase-aware enhancement [46], have been proposed, but model generalization across scanners and hospitals remains difficult due to variations in acquisition protocols.

8.3 Tumor Heterogeneity and Pathological Complexity

Segmenting pathological livers especially those with tumors, cysts, or metastases presents added complexity. Lesions exhibit diverse shapes, sizes, and enhancement patterns across imaging phases, often with blurred or irregular boundaries. The coexistence of normal and abnormal tissues within the same volume challenges both classical and deep models to maintain sensitivity to small lesions while preserving global consistency. Hybrid models combining texture attention and transformer-based global context [9] show improved tumor delineation but still underperform in small or low-contrast lesion detection. Moreover, annotation variability among radiologists further compounds training noise, affecting model stability and evaluation reliability. Existing datasets are dominated by normal or mildly pathological cases, with limited inclusion of rare liver diseases such as cholangiocarcinoma, hemangioma, or diffuse metastases. This imbalance leads to biased model performance and unreliable results in clinically critical edge cases. Future research must focus on curating large-scale, balanced datasets that capture

Table 6. Performance comparison of Transformer and hybrid architectures for 3D volumetric liver segmentation (2020–2025).

Model	Year	Architecture Type	Backbone	Dataset	DSC (%)	IoU (%)
TransUNet [9]	2021	Hybrid Transformer-CNN	ViT + UNet Decoder	LiTS	94.7	90.3
ViT-V-Net [88]	2021	Pure Transformer	ViT Encoder + 3D Conv Decoder	LiTS	94.2	89.7
UNETR [36]	2022	Transformer-Hybrid	ViT Encoder + Dense Skip	BTCV	95.6	91.4
SwinUNETR [89]	2022	Transformer-Hybrid	Swin Transformer + UNet Decoder	LiTS	96.3	93.0
MISSFormer [90]	2022	Hybrid Transformer-CNN	CNN + Multi-Head Self-Attention	CHAOS	95.8	92.6
MedT [91]	2021	Transformer-Hybrid	Axial Transformer + Conv Blocks	CHAOS	95.1	91.9
SwinUNETR++ [92]	2023	Transformer-Hybrid	Swin Transformer + UNet++	LiTS	97.2	94.6
MedFormer [93]	2023	Pure Transformer	ViT Encoder + CNN Decoder	LiTS	96.8	93.9
TransBTS++ [94]	2023	Hybrid Transformer-CNN	BnUNet + Multi-Scale Transformer	LiTS	96.5	93.4
UNETR++ [95]	2023	Transformer-Hybrid	Swin-Transformer + Dense Skip	LiTS	97.1	94.1
CT-SAM [96]	2023	Vision Foundation Model	SAM + ViT Encoder	LiTS	96.9	94.2
Swin-Med3D [97]	2024	Transformer-Hybrid	Swin-V2 Backbone	CHAOS	97.5	94.9
SwinDiffSeg [98]	2024	Diffusion-Transformer Hybrid	Swin-Transformer + Diffusion Prior	LiTS	97.8	95.3
TransR-3D [99]	2024	Transformer-Hybrid	ResNet + Multi-head Transformer	LiTS	97.3	94.8
DPT-3D [100]	2025	Pure Transformer	Dual-Path Transformer	CHAOS	98.0	95.6
HybridSwinSeg [101]	2025	Hybrid Transformer-CNN	Swin-V3 + UNet++ Decoder	LiTS	98.2	96.0

the full spectrum of hepatic abnormalities.

8.4 Limited and Imbalanced Data

The scarcity of large-scale, publicly annotated 3D CT datasets remains one of the major bottlenecks in liver segmentation research. Manual delineation of volumetric CT data is labor-intensive and time-consuming, often requiring multiple expert reviews. As a result, datasets like LiTS or 3DIRCADb contain fewer than a few hundred scans, limiting model generalizability. Additionally, there exists a strong imbalance between background, healthy liver tissue, and pathological lesions, causing biased learning. Techniques such as patch-based sampling, loss re-weighting, and synthetic data augmentation [17] partially alleviate this, yet domain shifts persist when deploying models in unseen clinical settings.

8.5 Limited Interpretability and Model Transparency

Deep neural networks operate as black boxes, making it difficult for clinicians to interpret or validate their decisions. The lack of transparency in prediction mechanisms reduces clinical trust and hinders regulatory approval. Although explainable AI (XAI) frameworks such as Grad-CAM, attention maps, and saliency analysis provide some interpretability, they remain coarse and non-intuitive for 3D medical images. Quantifying uncertainty and causality in volumetric predictions remains a largely unsolved problem.

8.6 Domain Shift and Generalization

Domain shift is one of the most persistent limitations in liver segmentation research. Deep networks trained on specific scanner protocols or patient cohorts often fail to generalize across institutions or imaging conditions. This phenomenon is known as domain shift, caused by differences in reconstruction kernels, noise levels, and contrast agent usage. Unsupervised domain adaptation [47] and self-supervised learning [126] have been proposed to address this, but cross-domain consistency in volumetric space remains challenging. Moreover, intensity normalization alone is insufficient since liver morphology, pathological load, and acquisition dynamics also vary across populations. Existing domain adaptation methods remain limited to 2D or shallow 3D models and are rarely validated in real-world multicenter settings. Thus, achieving domain-robust and scanner-invariant segmentation remains an open challenge.

8.7 Computational and Memory Constraints

Processing 3D CT volumes demands significant computational resources due to their large voxel counts (typically $512 \times 512 \times 300$ or more). Training deep volumetric models such as 3D U-Net [8] or SwinUNETR [10] requires high-end GPUs with large memory capacity. Patch-based training reduces memory load but sacrifices contextual information. Recent transformer-based networks exacerbate this issue as their self-attention complexity grows quadratically with the number of voxels.

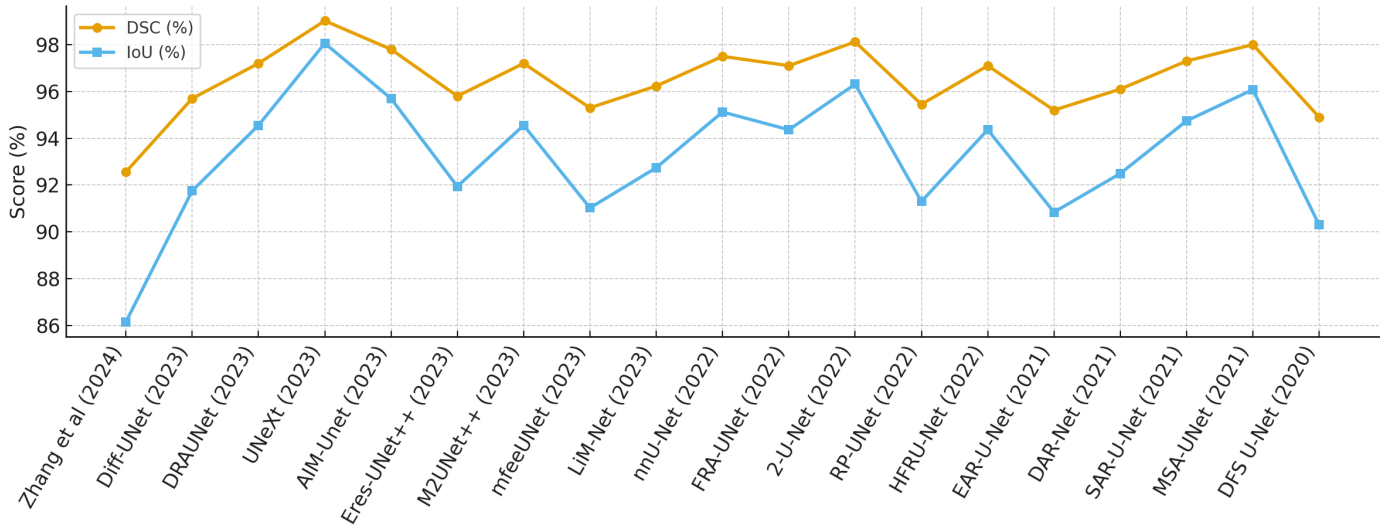

Figure 7. Performance trend of liver 3D volumetric CT image segmentation Models based on U-net (2015-2025).

Efficient strategies, including sparse convolution and hierarchical attention, attempt to balance performance with computational feasibility, though the trade-off remains unresolved.

8.8 Annotation Ambiguity, and Inter-observer Variability

Manual annotations serve as the ground truth for supervised learning, but significant inter-observer variability exists even among expert radiologists. Differences in defining organ boundaries or lesion margins introduce label uncertainty, which can propagate through the training process. Some researchers have explored probabilistic labeling, uncertainty modeling, and consensus-driven annotations [49] to address this issue. However, the creation of high-fidelity, uncertainty-aware datasets at scale is still lacking.

8.9 Evaluation and Reproducibility Challenges

Although metrics such as Dice and Hausdorff distance are widely adopted, variations in preprocessing, resampling, and evaluation scripts lead to inconsistent performance reporting. Reproducibility is further hindered by closed-source implementations and lack of standardized benchmarks. The LiTS and MSD challenges partially mitigated this by providing unified evaluation platforms, yet a global standard for volumetric medical segmentation is still absent. Moreover, few studies report statistical significance testing or confidence intervals, which are essential for rigorous comparison.

8.10 Clinical Integration and Real-world Deployment

Translating research prototypes into clinical practice remains a substantial gap. Clinical deployment requires robustness to unseen data, integration with PACS (Picture Archiving and Communication Systems), explainability, and regulatory approval. Many deep networks operate as black boxes, providing no interpretability to radiologists. Recent works [50] emphasize the need for explainable AI and uncertainty quantification to enhance trust and clinical acceptance. Additionally, real-time inference is critical for interventional or surgical navigation, yet current volumetric models often fail to meet required latency constraints.

8.11 Reproducibility and Benchmarking Gaps

Reproducibility remains a major barrier to progress. Many published studies omit implementation details, hyperparameters, or preprocessing steps, making independent verification difficult. Additionally, differences in evaluation protocols, such as resampling, cropping, or metric computation, often lead to inconsistent results across studies. Although initiatives like LiTS and MSD promote standardized evaluation, a unified open benchmarking framework for liver segmentation is still lacking.

8.12 Clinical Integration Barriers

Even the most accurate segmentation models rarely transition into clinical workflows due to integration, usability, and regulatory challenges. Real-time inference, compatibility with PACS systems, and interpretability for non-technical clinicians are critical

Table 7. Summary of deep learning models developed using modifications of the basic U-Net architecture.

Model	Type	Dataset	DSC
DRDA-Net (2025) [102]	3D	LiTS, 3DIRCADb	0.754
SCANeXt (2024) [103]	3D	Synapse, BraTS, ACDC	–
Zhang et al (2024) [87]	3D	LiTS	92.56
Diff-UNet (2023) [69]	3D	MSD	0.957
DRAUNet (2023) [70]	3D	LiTS17, 3DIRCADb, Sliver07	0.973, 0.974, 0.969
UNeXt (2023) [71]	3D	LiTS17	**0.9902**
AIM-Unet (2023) [72]	2.5D	LiTS17	0.978
Eres-UNet++ (2023) [73]	3D	LiTS17	0.958
M2UNet++ (2023) [74]	3D	3DIRCADb	0.972
mfeeU-Net (2023) [75]	3D	LiTS17	0.953
LiM-Net (2023) [76]	3D	3DIRCADb, CHAOS, LiTS17	0.973, 0.951, 0.963
nnU-Net (2022) [77]	3D	LiTS17, 3DIRCADb	0.975
FRA-UNet (2022) [78]	3D	LiTS17, 3DIRCADb	0.971, 0.971
2-U-Net (2022) [79]	2.5D	3DIRCADb	0.9812
RP-UNet (2022) [80]	3D	Sliver07, 3DIRCADb	0.964, 0.945
HFRU-Net (2022) [81]	3D	3DIRCADb	0.971
EAR-U-Net (2021) [82]	2D	LiTS17	0.952
DAR-Net (2021) [83]	3D	3DIRCADb	0.961
SAR-U-Net (2021) [84]	2.5D	LiTS17	0.973
MSA-UNet (2021) [85]	3D	mixed 3DIRCADb, Sliver07	0.980
DFS U-Net (2020) [86]	2D	Affiliated Hospital of Jiangsu University	0.949

for deployment. Furthermore, current models are rarely validated prospectively or tested in diverse real-world settings. Achieving clinical-grade robustness requires extensive validation under strict quality assurance protocols and adherence to ethical, legal, and privacy standards.

8.13 Summary of Key Challenges

In summary, the challenges in 3D volumetric liver segmentation arise from intrinsic biological variability, data heterogeneity, computational limitations, and the gap between algorithmic performance and clinical utility. Addressing these challenges requires a holistic approach combining improved data curation, domain adaptation, efficient architectures, uncertainty modeling, and human-AI collaboration frameworks. Overcoming these obstacles will pave the way for truly generalizable and clinically reliable liver segmentation systems.

9 Future Research Directions

The rapid evolution of deep learning, coupled with advances in medical imaging and computational resources, offers promising opportunities for overcoming the current limitations in 3D liver segmentation. This section explores key research directions expected to drive the next generation of automated liver analysis systems, emphasizing scalability, generalization, interpretability, and clinical applicability.

9.1 Self-Supervised and Semi-Supervised Learning

Given the limited availability of annotated medical datasets, self-supervised learning (SSL) has emerged as a compelling strategy for representation learning without requiring manual labels [53]. In SSL frameworks, models learn invariant features by solving pretext tasks such as contrastive learning, rotation prediction, or masked voxel reconstruction. Recent methods like SwinUNETR-SSL [39] and SimMIM-Med [54] demonstrate that pretraining on large-scale unlabeled CT datasets can significantly enhance downstream liver segmentation accuracy. Combining SSL with limited supervised fine-tuning could dramatically reduce dependence on expert annotations. Furthermore, semi-supervised strategies leveraging pseudo-labeling and consistency regularization can further exploit partially labeled datasets for better generalization across domains.

9.2 Foundation Models and Multimodal Integration

The concept of medical foundation models is gaining traction, drawing inspiration from vision-language and large-scale vision transformer frameworks [55]. By pretraining on diverse multimodal datasets, including CT, MRI, and ultrasound, these models can generalize across organs, imaging modalities, and pathological conditions. For liver segmentation, multimodal fusion models combining CT and MRI features could capture complementary soft-tissue and structural cues. Vision-language models such as BioMedCLIP and MedSAM [56] show the potential for text-guided segmentation, enabling radiologists to interactively define anatomical regions of interest. This paradigm shift toward large, pre-trained, task-adaptive models may redefine how 3D liver segmentation systems are trained and deployed.

9.3 Transformer-Based 3D Architectures

While CNNs have dominated medical image segmentation, recent breakthroughs in transformer-based architectures such as SwinUNETR [10], nnFormer [51], and TransBTS [38] highlight their superior capability in modeling global contextual dependencies. Future research will likely focus on optimizing transformer-based 3D models for efficiency and scalability. Techniques such as window-based attention, sparse tokenization, and hierarchical representation learning can reduce computational cost while maintaining volumetric awareness. Hybrid CNN-transformer architectures can further combine local texture sensitivity with global spatial reasoning, improving segmentation robustness in challenging cases.

9.4 Uncertainty Quantification and Explainable AI (XAI)

For clinical acceptance, interpretability and reliability are paramount. Deep segmentation models should not only produce high-accuracy masks but also communicate confidence and uncertainty levels to clinicians. Bayesian deep learning, Monte Carlo dropout, and ensemble-based uncertainty estimation [57] offer ways to quantify model confidence at voxel-level granularity. Integrating explainability frameworks such as Grad-CAM [118] or SHAP with volumetric visualization can provide intuitive explanations for model decisions. This enables radiologists to identify failure cases, assess model trustworthiness, and facilitate regulatory approval. The development of causability-based frameworks [50] will further enhance human–AI collaboration in liver diagnostics.

9.5 Federated and Privacy-Preserving Learning

Medical data sharing across institutions is restricted due to privacy and legal constraints, limiting the diversity of training data. Federated learning (FL) [116] offers a promising avenue by enabling decentralized model training without transferring patient data. Each hospital trains locally and shares model weights rather than raw scans, maintaining confidentiality. For liver segmentation, FL combined with domain adaptation can improve model generalization across heterogeneous clinical sites. Enhancing communication efficiency, managing data imbalance, and handling noisy local updates remain active areas of research. Integrating homomorphic encryption or differential privacy techniques could further ensure secure and ethical model training.

9.6 Cross-Domain and Cross-Modal Adaptation

Domain adaptation remains essential for model robustness in real-world clinical environments. Future research will increasingly focus on harmonizing models across imaging modalities (e.g., CT, MRI, ultrasound) and contrast phases (arterial, venous, delayed). Cycle-consistent generative adversarial networks (CycleGANs) [35] and domain-invariant representation learning methods can help bridge appearance discrepancies. Moreover, multi-domain pretraining and test-time adaptation frameworks could allow models to dynamically adjust to unseen scanners or protocols without retraining.

9.7 Data-Centric and Annotation-Efficient Approaches

With the growing emphasis on data-centric AI, future work will likely prioritize improving dataset quality, annotation consistency, and label efficiency. Active learning techniques an iteratively select the most informative cases for manual labeling, minimizing annotation burden. Synthetic data generation using diffusion models and 3D generative adversarial networks [117] could augment scarce datasets while maintaining anatomical realism. Collaborative annotation platforms leveraging consensus and uncertainty modeling will further improve ground-truth reliability.

9.8 Integration with Clinical Workflows and Real-Time Systems

The ultimate goal of liver segmentation research is seamless clinical integration. Future systems should be capable of real-time inference, supporting surgical planning, radiotherapy dose computation, and computer-assisted interventions. Integration with hospital PACS and radiology information systems (RIS) will enable end-to-end clinical workflows. Cloud-based deployment and edge-computing solutions will help deliver AI-assisted diagnosis in low-resource environments. Furthermore, regulatory-compliant frameworks aligned with standards such as FDA and CE-MDR certification are crucial for translating research prototypes into clinically usable tools.

9.9 Towards Holistic Liver Analysis and Multi-Task Learning

Beyond segmentation, future models are expected to perform joint analysis tasks such as lesion detection, liver volume estimation, fibrosis staging, and prognosis prediction. Multi-task learning architectures can share representations across these tasks, improving both efficiency and interpretability. For example, a single 3D network could simultaneously segment the liver, detect tumors, and estimate volumetric biomarkers, providing comprehensive quantitative assessments for personalized treatment planning.

9.10 Summary

In summary, the future of 3D liver segmentation lies in synergizing innovations across model architectures, learning paradigms, and clinical integration. Self-supervised pretraining, transformer-based modeling, federated collaboration, and uncertainty-aware explainable AI will define the next generation of intelligent, generalizable, and trustworthy liver segmentation frameworks. Bridging algorithmic excellence with clinical practicality will be the defining challenge and opportunity—of the coming decade in volumetric medical imaging.

10 Conclusion

3D volumetric liver CT segmentation has evolved into a pivotal task in computer-assisted diagnosis, surgical planning, and hepatic disease quantification. Over the past decade, the field has witnessed a remarkable transition from conventional image processing techniques to deep convolutional and transformer-based frameworks that leverage global contextual reasoning and multi-scale feature fusion. This review systematically analyzed the progress, limitations, and emerging trends across various architectures,ranging from 3D U-Net variants and hybrid CNN–Transformer models to attention-guided and self-supervised segmentation paradigms.

Our comprehensive synthesis highlights that hybrid networks, particularly those integrating transformer encoders with convolutional decoders, have demonstrated substantial improvements in capturing volumetric dependencies and fine-grained anatomical boundaries. Nevertheless, challenges such as limited dataset diversity, annotation inconsistency, and domain shift continue to impede the generalization of segmentation models to heterogeneous clinical environments. Moreover, despite achieving state-of-the-art accuracy in benchmark datasets, most methods still lack interpretability, scalability, and prospective clinical validation.

Looking forward, the next generation of liver segmentation research must move toward data-efficient, explainable, and clinically aware frameworks. This includes the development of foundation models trained on large-scale, multimodal datasets, incorporating self-supervised and federated learning strategies to reduce annotation burden while preserving patient privacy. Future systems should also embed uncertainty estimation and domain adaptation to ensure robust performance across imaging modalities, scanners, and clinical settings.

Ultimately, the convergence of deep learning, medical imaging, and clinical domain knowledge offers a path toward truly intelligent and reliable liver segmentation systems. Achieving this vision will require not only algorithmic innovation but also collaborative efforts among clinicians, engineers, and researchers to ensure reproducibility, transparency, and ethical deployment in real-world clinical workflows. With continued interdisciplinary advancement, automated 3D liver segmentation stands poised to become an indispensable tool in precision hepatology and image-guided intervention.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 74*(3), 229-263. [CrossRef]

[2] Alamri, Z. Z. (2018). The role of liver in metabolism: An updated review with physiological emphasis. *International Journal of Basic & Clinical Pharmacology, 7*(11), 2271-2276. [CrossRef]

[3] Soler, L., Nicolau, S., Pessaux, P., Mutter, D., & Marescaux, J. (2014). Real-time 3D image reconstruction guidance in liver resection surgery. *Hepatobiliary surgery and nutrition, 3*(2), 73. [CrossRef]

[4] Heimann, T., Van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C., ... & Wolf, I. (2009). Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE transactions on medical imaging, 28*(8), 1251-1265. [CrossRef]

[5] Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., ... & Menze, B. (2023). The liver tumor segmentation benchmark (lits). *Medical image analysis, 84*, 102680. [CrossRef]

[6] Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). IEEE. [CrossRef]

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing. [CrossRef]

[8] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016, October). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424-432). Cham: Springer International Publishing. [CrossRef]

[9] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

[10] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2021, September). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop* (pp. 272-284). Cham: Springer International Publishing. [CrossRef]

[11] Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J.-B., Moreau, J., Osswald, A.-B., Bouhadjar, M., & Marescaux, J. (2010). *3D image reconstruction for comparison of algorithm database: A patient-specific anatomical and medical image database*. IRCAD. https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01

[12] Simpson, A. L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., ... & Cardoso, M. J. (2019). A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*.

[13] Kavur, A. E., Gezer, N. S., Barış, M., Aslan, S., Conze, P. H., Groza, V., ... & Selver, M. A. (2021). CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Medical image analysis, 69*, 101950. [CrossRef]

[14] Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., ... & Cardoso, M. J. (2022). The medical segmentation decathlon. *Nature communications, 13*(1), 4128. [CrossRef]

[15] Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., & Klein, A. (2015, October). Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge* (Vol. 5, p. 12).

[16] Liu, C., Cui, D., Shi, D., Hu, Z., Qin, Y., & Lang, J. (2018, October). Automatic liver segmentation in CT volumes with improved 3D U-Net. In *Proceedings of the 2nd International Symposium on Image Computing and Digital Medicine* (pp. 78-82). [CrossRef]

[17] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods, 18*(2), 203-211. [CrossRef]

[18] Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging, 15*(1), 29. [CrossRef]

[19] Lim, S. J., Jeong, Y. Y., Lee, C. W., & Ho, Y. S. (2004, May). Automatic segmentation of the liver in CT images using the watershed algorithm based on morphological filtering. In *Medical Imaging 2004: Image Processing* (Vol. 5370, pp. 1658-1666). SPIE. [CrossRef]

[20] Li, C., Xu, C., Gui, C., & Fox, M. D. (2005, June). Level set evolution without re-initialization: a new variational formulation. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 430-436). IEEE. (pp. 430-436). [CrossRef]

[21] Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y. W., Nakamura, H., & Sato, Y. (2008). Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model. *Academic radiology, 15*(11), 1390-1403. [CrossRef]

[22] Boykov, Y. Y., & Jolly, M. P. (2001, July). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001* (Vol. 1, pp. 105-112). IEEE. [CrossRef]

[23] Shelhamer, E., Long, J., & Darrell, T. (2016). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640-651. [CrossRef]

[24] Zhang, Y., Chen, H., He, Y., Ye, M., Cai, X., & Zhang, D. (2018). Road segmentation for all-day outdoor robot navigation. *Neurocomputing, 314*, 316-325. [CrossRef]

[25] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

[26] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., ... & Liu, J. (2019). Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging, 38*(10), 2281-2292. [CrossRef]

[27] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018, September). Unet++: A nested u-net architecture for medical image segmentation. In *International workshop on deep learning in medical image analysis* (pp. 3-11). Cham: Springer International Publishing. [CrossRef]

[28] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence, 40*(4), 834-848. [CrossRef]

[29] Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., ... & Menze, B. H. (2016, October). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *International conference on medical image computing and computer-assisted intervention* (pp. 415-423). Cham: Springer International Publishing. [CrossRef]

[30] Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(8), 2011-2023. [CrossRef]

[31] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2020). Residual Dense Network for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(7), 2480-2495. [CrossRef]

[32] Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., & Heng, P. A. (2017). 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis, 41*, 40-54. [CrossRef]

[33] Karimi, D., & Salcudean, S. E. (2019). Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on medical imaging, 39*(2), 499-513. [CrossRef]

[34] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(2), 318-327. [CrossRef]

[35] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017, October). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2242-2251). IEEE. [CrossRef]

[36] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., ... & Xu, D. (2022, January). UNETR: Transformers for 3D Medical Image Segmentation. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1748-1758). IEEE. [CrossRef]

[37] Li, J., Wang, W., Chen, C., Zhang, T., Zha, S., Wang, J., & Yu, H. (2022). TransBTSV2: Towards better and more efficient volumetric segmentation of medical images. *arXiv preprint arXiv:2201.12785*.

[38] Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., & Li, J. (2021, September). Transbts: Multimodal brain tumor segmentation using transformer. In *International conference on medical image computing and computer-assisted intervention* (pp. 109-119). Cham: Springer International Publishing. [CrossRef]

[39] Tang, Y., Yang, D., Li, W., Roth, H. R., Landman, B., Xu, D., ... & Hatamizadeh, A. (2022, June). Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 20698-20708). IEEE. [CrossRef]

[40] Cai, Y., Long, Y., Han, Z., Liu, M., Zheng, Y., Yang, W., & Chen, L. (2023). Swin Unet3D: a three-dimensional medical image segmentation network combining vision transformer and convolution. *BMC medical informatics and decision making, 23*(1), 33. [CrossRef]

[41] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021, October). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 9992-10002). IEEE. [CrossRef]

[42] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ayed, I. B. (2021). Boundary loss for highly unbalanced segmentation. *Medical Image Analysis, 67*, 101851. [CrossRef]

[43] Jurdi, R. E., Petitjean, C., Honeine, P., Cheplygina, V., & Abdallah, F. (2021, August). A surprisingly effective perimeter-based loss for medical image segmentation. In *Medical Imaging with Deep Learning* (pp. 158-167). PMLR.

[44] Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., ... & Chen, J. (2020, September). How distance transform maps boost segmentation CNNs: an empirical study. In *Medical imaging with deep learning* (pp. 479-492). PMLR.

[45] Heimann, T., & Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Medical image analysis, 13*(4), 543-563. [CrossRef]

[46] Li, C., & Zhu, A. (2020). Application of image fusion in diagnosis and treatment of liver cancer. *Applied Sciences, 10*(3), 1171. [CrossRef]

[47] Dou, Q., Ouyang, C., Chen, C., Chen, H., & Heng, P. A. (2018, July). Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (pp. 691-697).

[48] Zhuang, X., Li, Y., Hu, Y., Ma, K., Yang, Y., & Zheng, Y. (2019, October). Self-supervised feature learning for 3d medical images by playing a rubik's cube. In *International conference on medical image computing and computer-assisted intervention* (pp. 420-428). Cham: Springer International Publishing. [CrossRef]

[49] Wu, Y., Shen, H., Tan, Y., & Shi, Y. (2022). Automatic liver tumor segmentation used the cascade multi-scale attention architecture method based on 3D U-Net. *International Journal of Computer Assisted Radiology and Surgery*, 17(10), 1915-1922. [CrossRef]

[50] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(4), e1312. [CrossRef]

[51] Zhou, H. Y., Guo, J., Zhang, Y., Yu, L., Wang, L., & Yu, Y. (2021). nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*.

[52] Roth, H. R., Lu, L., Lay, N., Harrison, A. P., Farag, A., Sohn, A., & Summers, R. M. (2018). Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis, 45*, 94-107. [CrossRef]

[53] Taleb, A., Loetzsch, W., Danz, N., Severin, J., Gaertner, T., Bergner, B., & Lippert, C. (2020). 3d self-supervised methods for medical imaging. *Advances in neural information processing systems, 33*, 18158-18172.

[54] Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., ... & Hu, H. (2022, June). SimMIM: a Simple Framework for Masked Image Modeling. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 9643-9653). IEEE. [CrossRef]

[55] Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., ... & Ni, D. (2024). Segment anything model for medical images?. *Medical Image Analysis, 92*, 103061. [CrossRef]

[56] Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. *Nature communications, 15*(1), 654. [CrossRef]

[57] Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)* (pp. 5574-5584).

[58] Shaker, A., Maaz, M., Rasheed, H., Khan, S., Yang, M. H., & Khan, F. S. (2024). UNETR++: delving into efficient and accurate 3D medical image segmentation. *IEEE Transactions on Medical Imaging, 43*(9), 3377-3390. [CrossRef]

[59] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*(3), 297-302. [CrossRef]

[60] JACCARD, P. (1901). Etude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat, 37*, 547-579.

[61] Wu, Q., Gu, J., Wu, C., & Li, J. (2021). Fully convolutional networks semantic segmentation based on conditional random field optimization. *Journal of Computational Methods in Science and Engineering, 21*(5), 1405-1415. [CrossRef]

[62] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. W., & Heng, P. A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE transactions on medical imaging, 37*(12), 2663-2674. [CrossRef]

[63] Wang, S., Cao, S., Chai, Z., Wei, D., Ma, K., Wang, L., & Zheng, Y. (2020). Conquering data variations in resolution: A slice-aware multi-branch decoder network. *IEEE Transactions on Medical Imaging, 39*(12), 4174-4185. [CrossRef]

[64] Zhang, L., & Yu, S. C. H. (2021). Context-aware PolyUNet for liver and lesion segmentation from abdominal CT images. *arXiv preprint arXiv:2106.11330*.

[65] Ben-Cohen, A., Klang, E., Raskin, S. P., Soffer, S., Ben-Haim, S., Konen, E., ... & Greenspan, H. (2019). Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence, 78*, 186-194. [CrossRef]

[66] Wang, Z., & Wang, G. (2018). Triplanar convolutional neural network for automatic liver and tumor image segmentation. *International Journal of Performability Engineering, 14*(12), 3151. [CrossRef]

[67] Ahn, Y., Yoon, J. S., Lee, S. S., Suk, H. I., Son, J. H., Sung, Y. S., ... & Kim, H. S. (2020). Deep learning algorithm for automated segmentation and volume measurement of the liver and spleen using portal venous phase computed tomography images. *Korean journal of radiology, 21*(8), 987. [CrossRef]

[68] Zhang, C., Hua, Q., Chu, Y., & Wang, P. (2021). Liver tumor segmentation using 2.5 D UV-Net with multi-scale convolution. *Computers in Biology and Medicine, 133*, 104424. [CrossRef]

[69] Xing, Z., Wan, L., Fu, H., Yang, G., & Zhu, L. (2023). Diff-unet: A diffusion embedded network for volumetric segmentation. *arXiv preprint arXiv:2303.10326*.

[70] Chen, Y., Zheng, C., Zhou, T., Feng, L., Liu, L., Zeng, Q., & Wang, G. (2023). A deep residual attention-based U-Net with a biplane joint method for liver segmentation from CT scans. *Computers in Biology and Medicine, 152*, 106421. [CrossRef]

[71] Bogoi, S., & Udrea, A. (2023). A lightweight deep learning approach for liver segmentation. *Mathematics, 11*(1), 95. [CrossRef]

[72] Özcan, F., Uçan, O. N., Karaçam, S., & Tunçman, D. (2023). Fully automatic liver and tumor segmentation from CT image using an AIM-Unet. *Bioengineering, 10*(2), 215. [CrossRef]

[73] Li, J., Liu, K., Hu, Y., Zhang, H., Heidari, A. A., Chen, H., ... & Elmannai, H. (2023). Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++. *Computers in Biology and Medicine, 158*, 106501. [CrossRef]

[74] Kushnure, D. T., & Talbar, S. N. (2022). M2UNet++: A modified multi-scale UNet++ architecture for automatic liver segmentation from computed tomography images. In *Handbook of Research on Applied Intelligence for Health and Clinical Informatics* (pp. 256-273). IGI Global Scientific Publishing. [CrossRef]

[75] Liu, J., Yan, Z., Zhou, C., Shao, L., Han, Y., & Song, Y. (2023). mfeeU-Net: A multi-scale feature extraction and enhancement U-Net for automatic liver segmentation from CT Images. *Mathematical Biosciences and Engineering, 20*(5), 7784. [CrossRef]

[76] Kushnure, D. T., Tyagi, S., & Talbar, S. N. (2023). LiM-Net: Lightweight multi-level multiscale network with deep residual learning for automatic liver segmentation in CT images. *Biomedical Signal Processing and Control, 80*, 104305. [CrossRef]

[77] Pettit, R. W., Marlatt, B. B., Corr, S. J., Havelka, J., & Rana, A. (2022). nnU-Net deep learning method for segmenting parenchyma and determining liver volume from computed tomography images. *Annals of Surgery Open, 3*(2), e155. [CrossRef]

[78] Chen, Y., Zheng, C., Hu, F., Zhou, T., Feng, L., Xu, G., ... & Zhang, X. (2022). Efficient two-step liver and tumour segmentation on abdominal CT via deep learning and a conditional random field. *Computers in Biology and Medicine, 150*, 106076. [CrossRef]

[79] Khattab, M. A., Liao, I. Y., Ooi, E. H., & Chong, S. Y. (2022). Compound W-Net with Fully Accumulative Residual Connections for Liver Segmentation Using CT Images. *Computational and Mathematical Methods in Medicine, 2022*(1), 8501828. [CrossRef]

[80] Czipczer, V., & Manno-Kovacs, A. (2022). Adaptable volumetric liver segmentation model for CT images using region-based features and convolutional neural network. *Neurocomputing, 505*, 388-401. [CrossRef]

[81] Kushnure, D. T., & Talbar, S. N. (2022). HFRU-Net: High-level feature fusion and recalibration U-Net for automatic liver and tumor segmentation in CT images. *Computer Methods and Programs in Biomedicine, 213*, 106501. [CrossRef]

[82] Wang, J., Zhang, X., Lv, P., Zhou, L., & Wang, H. (2021). EAR-U-Net: EfficientNet and attention-based residual U-Net for automatic liver segmentation in CT. *arXiv preprint arXiv:2110.01014*.

[83] Xie, X., Zhang, W., Wang, H., Li, L., Feng, Z., Wang, Z., Wang, Z., & Pan, X. (2021). Dynamic adaptive residual network for liver CT image segmentation. *Computers & Electrical Engineering, 91*, 107024. [CrossRef]

[84] Wang, J., Lv, P., Wang, H., & Shi, C. (2021). SAR-U-Net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in Computed Tomography. *Computer Methods and Programs in Biomedicine, 208*, 106268. [CrossRef]

[85] Wu, J., Zhou, S., Zuo, S., Chen, Y., Sun, W., Luo, J., ... & Wang, D. (2021). U-Net combined with multi-scale attention mechanism for liver segmentation in CT images. *BMC Medical Informatics and Decision Making, 21*(1), 283. [CrossRef]

[86] Liu, Z., Han, K., Wang, Z., Zhang, J., Song, Y., Yao, X., ... & Sheng, V. S. (2021). Automatic liver segmentation from abdominal CT volumes using improved convolution neural networks. *Multimedia Systems, 27*(1), 111-124. [CrossRef]

[87] Zhang, B., Qiu, S., & Liang, T. (2024). Dual attention-based 3D U-Net liver segmentation algorithm on CT images. *Bioengineering, 11*(7), 737. [CrossRef]

[88] Chen, J., He, Y., Frey, E. C., Li, Y., & Du, Y. (2021). Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv preprint arXiv:2104.06468*.

[89] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., ... & Xu, D. (2022, January). UNETR: Transformers for 3D Medical

Image Segmentation. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1748-1758). IEEE. [CrossRef]

[90] Huang, X., Deng, Z., Li, D., & Yuan, X. (2021). Missformer: An effective medical image segmentation transformer. *arXiv preprint arXiv:2109.07162*.

[91] Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., & Patel, V. M. (2021, September). Medical transformer: Gated axial-attention for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 36-46). Cham: Springer International Publishing. [CrossRef]

[92] Jin, J., Yang, S., Tong, J., Zhang, K., & Wang, Z. (2025). Slim UNETR++: A lightweight 3D medical image segmentation network for medical image analysis. *Medical & Biological Engineering & Computing, 63*(10), 3123-3137. [CrossRef]

[93] Perera, S., Navard, P., & Yilmaz, A. (2024, June). SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 4981-4988). IEEE. [CrossRef]

[94] Li, C., Zhang, J., Hong, L., Ma, P., Liu, G., & Zhang, N. (2023, November). Attention-TransBTS: A novel transformer-unet architecture for MRI brain tumor segmentation. In *2023 China Automation Congress (CAC)* (pp. 8347-8352). IEEE. [CrossRef]

[95] Ntanzi, S., & Viriri, S. (2025). UNETR++ with Voxel-Focused Attention: Efficient 3D Medical Image Segmentation with Linear-Complexity Transformers. *Applied Sciences, 15*(20), 11034. [CrossRef]

[96] Zhang, Y., Shen, Z., & Jiao, R. (2024). Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine, 171*, 108238. [CrossRef]

[97] Hatamizadeh, A., Xu, Z., Yang, D., Li, W., Roth, H., & Xu, D. (2022). Unetformer: A unified vision transformer model and pre-training framework for 3d medical image segmentation. *arXiv preprint arXiv:2204.00631*.

[98] Mu, J., & Wang, J. (2025, June). Diffusion-Guided Structure-Aware Segmentation for Medical Imaging. In *2025 IEEE International Conference on Pattern Recognition, Machine Vision and Artificial Intelligence (PRMVAI)* (pp. 1-5). IEEE. [CrossRef]

[99] Jayswal, A. K., Singh, S. P., Mishra, S., Vats, P., Kaur, K., & Dubey, A. K. (2025). Augmented 3D ResUNETR: Combining residual learning and transformer encoding with multi-scale feature extraction for 3D medical image segmentation. *Biomedical Signal Processing and Control, 110*, 108235. [CrossRef]

[100] Yang, Z., & Li, S. (2023). Dual-path network for liver and tumor segmentation in CT images using Swin Transformer encoding approach. *Current Medical Imaging Reviews, 19*(10), 1114-1123. [CrossRef]

[101] Jagarapu, S. T. A., Kaata, S. K., Bajpai, S., & Holla, M. R. (2025). DualBranchNetwork3D: A Hybrid CNN–Swin Transformer Model for Automated Cirrhotic Liver Segmentation on MRI. *IEEE Access, 13*, 210497-210506. [CrossRef]

[102] Idress, W. M., Zhao, Y., Abouda, K. A., & Yang, S. (2025). DRDA-Net: Deep residual dual-attention network with multi-scale approach for enhancing liver and tumor segmentation from CT images. *Applied Sciences, 15*(5), 2311. [CrossRef]

[103] Liu, Y., Zhang, Z., Yue, J., & Guo, W. (2024). SCANeXt: Enhancing 3D medical image segmentation with dual attention network and depth-wise convolution. *Heliyon, 10*(5). [CrossRef]

[104] Chlebus, G., Schenk, A., Moltz, J. H., van Ginneken, B., Hahn, H. K., & Meine, H. (2018). Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Scientific reports, 8*(1), 15497. [CrossRef]

[105] Wang, H., Wang, Z. M., Cui, X. T., & Li, L. (2023). TDS-U-Net: Automatic liver and tumor separate segmentation of CT volumes using attention gates. *Journal of Intelligent & Fuzzy Systems, 44*(6), 8817-8825. [CrossRef]

[106] Abueed, O., Wang, Y., & Khasawneh, M. (2025). A Systematic Review of U-Net Optimizations: Advancing Tumour Segmentation in Medical Imaging. *IET Image Processing, 19*(1), e70203. [CrossRef]

[107] Jeong, J. G., Choi, S., Kim, Y. J., Lee, W. S., & Kim, K. G. (2022). Deep 3D attention CLSTM U-Net based automated liver segmentation and volumetry for the liver transplantation in abdominal CT volumes. *Scientific Reports, 12*(1), 6370. [CrossRef]

[108] Jiang, L., Ou, J., Liu, R., Zou, Y., Xie, T., Xiao, H., & Bai, T. (2023). Rmau-net: Residual multi-scale attention u-net for liver and tumor segmentation in ct images. *Computers in Biology and Medicine, 158*, 106838. [CrossRef]

[109] Wang, J., Zhang, X., Guo, L., Shi, C., & Tamura, S. (2023). Multi-scale attention and deep supervision-based 3D UNet for automatic liver segmentation from CT. *Mathematical Biosciences and Engineering, 20*(1), 1297-1316. [CrossRef]

[110] Soltani-Gol, M., Fattahi, M., Soltanian-Zadeh, H., & Sheikhaei, S. (2022, May). DRAU-Net: Double Residual Attention Mechanism for automatic MRI brain tumor segmentation. In *2022 30th International Conference on Electrical Engineering (ICEE)* (pp. 587-591). IEEE. [CrossRef]

[111] Wang, F., Cheng, X. L., Luo, N. B., & Su, D. K. (2024). Attention-guided context asymmetric fusion networks for the liver tumor segmentation of computed tomography images. *Quantitative Imaging*

in Medicine and Surgery, 14(7), 4825-4839. [CrossRef]

[112] Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment Anything in Medical Images. *Nature Communications, 15*, 654. [CrossRef]

[113] Zhang, Z., Gao, J., Li, S., & Wang, H. (2024). Rmcnet: A liver cancer segmentation network based on 3d multi-scale convolution, attention, and residual path. *Bioengineering, 11*(11), 1073. [CrossRef]

[114] Valanarasu, J. M. J., & Patel, V. M. (2022, September). Unext: Mlp-based rapid medical image segmentation network. In *International conference on medical image computing and computer-assisted intervention* (pp. 23-33). Cham: Springer Nature Switzerland. [CrossRef]

[115] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020, July). Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)* (pp. 558-564). IEEE. [CrossRef]

[116] Sheller, M., Edwards, B., Reina, G. A., Martin, J., & Bakas, S. (2019). NIMG-68. Federated learning in neuro-oncology for multi-institutional collaborations without sharing patient data. *Neuro-oncology, 21*(Suppl 6), vi176. [CrossRef]

[117] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems, 33*, 6840-6851.

[118] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017, October). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 618-626). IEEE. [CrossRef]

[119] Mubashar, M., Ali, H., Grönlund, C., & Azmat, S. (2022). R2U++: a multiscale recurrent residual U-Net with dense skip connections for medical image segmentation. *Neural Computing and Applications, 34*(20), 17723-17739. [CrossRef]

[120] Zhang, Y., Liao, Q., Ding, L., & Zhang, J. (2022). Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5 D solutions. *Computerized Medical Imaging and Graphics, 99*, 102088. [CrossRef]

[121] Sarrut, D., Bała, M., Bardiès, M., Bert, J., Chauvin, M., Chatzipapas, K., ... & Roncali, E. (2021). Advanced Monte Carlo simulations of emission tomography imaging systems with GATE. *Physics in Medicine & Biology, 66*(10), 10TR03. [CrossRef]

[122] Li, Z., Zhang, H., Li, Z., & Ren, Z. (2022). Residual-attention UNet++: a nested residual-attention U-net for medical image segmentation. *Applied Sciences, 12*(14), 7149. [CrossRef]

[123] Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., ... & Menze, B. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*.

[124] Campadelli, P., Casiraghi, E., & Esposito, A. (2009). Liver segmentation from computed tomography scans: a survey and a new algorithm. *Artificial intelligence in medicine, 45*(2-3), 185-196. [CrossRef]

[125] Okada, T., Linguraru, M. G., Hori, M., Summers, R. M., Tomiyama, N., & Sato, Y. (2015). Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. *Medical image analysis, 26*(1), 1-18. [CrossRef]

[126] Zhu, J., Li, Y., Hu, Y., Ma, K., Zhou, S. K., & Zheng, Y. (2020). Rubik's cube+: A self-supervised feature learning framework for 3d medical image analysis. *Medical image analysis, 64*, 101746. [CrossRef]

Irshad Ali Khan received the M.S. degree in Software Engineering from Dalian University of Technology, China. Currently, he is pursuing the Ph.D. degree in Software Engineering at Dalian University of Technology, China. His research focuses on medical image processing, with particular interests in image analysis and computational methods for medical applications. His work aims to develop efficient and reliable software-based solutions to support medical diagnosis and healthcare research. (Email: Irshadsa1@outlook.com)

Gul Zaman Khan received the B.S. degree in Software Engineering from the International Islamic University, Islamabad, Pakistan, and the M.Sc. degree in Computer Software Engineering from the University of Engineering and Technology (UET), Mardan, Pakistan. Currently, he is pursuing the Ph.D. degree in Software Engineering at Dalian University of Technology, China. His research interests include medical image analysis, intelligent systems, and pattern recognition. (Email: gulzamankhan726@gmail.com)

Yar Muhammad received the B.S. degree in computer science from the University of Malakand (UoM), Khyber Pakhtunkhwa, Pakistan. He was awarded a gold medal in his B.S. degree. He received the M.S. degree in computer science from Abdul Wali Khan University Mardan (AWKUM), Khyber Pakhtunkhwa, Pakistan. He has published multiple research articles in numerous reputed journals and is serving as a reviewer for more than 20 journals and transactions. Currently, he is pursuing his Ph.D. degree in Computer Science and Technology at Beihang University, Beijing, China. His research interests include

machine learning, deep learning, natural language processing, the Internet of Things, and computational intelligence. (Email: yarkhan@buaa.edu.cn)

Samreen Ihsan received the B.S. degree in computer science from Ghazi Umara Khan Degree College Samarbagh, affiliated with the University of Malakand, Pakistan. Currently, she is pursuing M.S. degree in Software Engineering at Dalian University of Technology, China. Her research interests include medical image analysis, and pattern recognition. (Email: samreenihsan09@gmail.com)

Ijazul Haq received the Ph.D. degree in Artificial Intelligence and Cybersecurity from Shanghai Jiao Tong University, China, in 2023, and the M.S. degree in Software Engineering from Wuhan University, China, in 2018. He joined the Shien-Ming Wu School of Intelligent Manufacturing, South China University of Technology, China, as a postdoctoral researcher in 2024. Dr. Haq was awarded the Chinese Government Scholarship by the China Scholarship Council in 2016 and the Shanghai Municipal Government Scholarship in 2018. His research interests include natural language processing, computer vision, and generative AI. (Email: hanjie@sjtu.edu.cn)