

ALN: An Anatomy-Aware Network For Unified Full-Volume Liver And Tumour Segmentation In CT

Syeda Husna Mehanoor¹, Supriya L P², Arun Madhu³, Neetha B S⁴, Dharma Teja Lanka⁵,
Mihir Harishbhai Rajyaguru⁶, Reji R⁷, Divya Saleela⁸, Jubilant J Kizhakkethottam⁹

¹Associate Professor, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth (Deemed to be University), Telangana, India

²Associate Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, India

³Assistant Professor, Saintgits College of Engineering, Kerala, India

⁴Assistant Professor, Fatima Mata National College, Kerala, India

⁵Assistant Professor, VNR Vignana Jyothi Institute of Engineering & Technology, Telangana, India

⁶Assistant Professor, Madhuben and Bhanubhai Patel Institute of Technology (MBIT) - The Charutar Vidya Mandal (CVM) University, Gujarat, India

⁷Professor, Thangal Kunju Musaliar Institute of Technology, Kerala, India

⁸Assistant Professor, College of Engineering Aranmula, Kerala, India

⁹Professor, SRM Institute of Science and Technology, Chennai, India

Abstract

Accurate joint segmentation of the liver and intra hepatic tumours from computed tomography images remains challenging due to extreme class imbalance, scale disparity, and structural ambiguity of sparse lesions. Existing methods frequently prioritise liver delineation, rely on cascaded pipelines that propagate errors, or optimise solely for overlap based metrics that fail to preserve tumour morphology and suppress false positive leakage. This work proposes ALN, an anatomy aware liver network for unified full volume liver tumour segmentation that jointly models organ level context and tumour specific structure within a single end to end framework. ALN employs a shared volumetric backbone with dual semantic and structural heads and is optimised using a composite objective that integrates multi class and class specific Dice supervision, tumour focused focal loss, explicit background suppression, and structure aware constraints to stabilise learning under severe sparsity. Evaluation on the Medical Segmentation Decathlon Task03 Liver dataset using volume level inference demonstrates near ceiling liver delineation with a Dice score of 0.98 ± 0.01 and robust tumour segmentation with a Dice score of 0.86 ± 0.04 , accompanied by balanced tumour precision and recall of 0.88 ± 0.05 and 0.85 ± 0.06 , respectively. Ablation analysis confirms that each component of the composite objective and the structural supervision head contributes to improved tumour segmentation stability and balanced precision-recall behaviour.

Keywords: Liver tumour segmentation; computed tomography; full volume learning; Medical Segmentation Decathlon.

1. Introduction and Related Work

Liver segmentation from computed tomography (CT) images is a fundamental task in medical image analysis and underpins numerous clinical applications such as surgical planning, disease staging, and treatment monitoring [1]. Accurate delineation of hepatic anatomy is essential for reliable volumetric measurements and informed clinical decision making. In recent years, deep learning based segmentation methods have demonstrated significant improvements over traditional image processing approaches by learning hierarchical representations from volumetric data [2]. Despite these advances, robust liver segmentation across heterogeneous datasets remains challenging due to variations in imaging protocols, contrast enhancement, and pathological conditions. These challenges become even more pronounced when pathological structures within the liver are considered simultaneously. As a result, achieving reliable and comprehensive hepatic segmentation continues to be an active area of research.

Beyond liver parenchyma delineation, accurate segmentation of liver tumours is of critical clinical importance. Precise localisation and delineation of tumours are essential for treatment planning, tumour burden assessment, and response evaluation in oncological care [3]. However, tumour segmentation presents unique challenges due to extreme variability in lesion size, shape, contrast, and spatial distribution. Tumours often occupy only a small fraction of the image volume and may exhibit weak or ambiguous boundaries relative to surrounding tissue. Imaging artefacts, partial volume effects, and noise further complicate automated tumour extraction from CT scans. As a result, accurate and consistent tumour segmentation remains difficult despite substantial research efforts.

Joint segmentation of the liver and tumours introduces additional complexity because of the fundamental differences in their anatomical characteristics. The liver occupies a large, contiguous spatial region, whereas tumours form sparse and heterogeneous structures distributed within the organ. When learned simultaneously, optimisation processes tend to prioritise the dominant liver class, often suppressing tumour predictions [4]. Attempts to compensate for this imbalance frequently result in degraded liver boundaries or unstable training behaviour. This inherent conflict highlights a critical limitation of many existing joint segmentation approaches. Effectively balancing organ level and lesion level learning objectives remains an open challenge.

To address these issues, several studies have proposed multistage or cascaded segmentation pipelines. Typically, these methods first segment the liver to localise the region of interest and subsequently segment tumours within the predicted liver mask (Gotra et al., 2017) [6]. While such strategies can reduce background interference, they introduce dependency between stages and are susceptible to error propagation. Inaccuracies in early stages can significantly compromise downstream tumour segmentation. Additionally, cascaded pipelines increase system complexity and reduce reproducibility across datasets. These drawbacks limit their practical deployment in clinical environments.

Patch based and region of interest learning approaches have also been explored as alternatives to full volume segmentation [7] (Moghbel et al., 2018). By focusing on smaller spatial regions, these methods reduce memory requirements and can enhance local detail representation. However, localised learning often sacrifices global anatomical context, which is essential for preserving organ integrity and consistent tumour localisation. Patch wise predictions may result in fragmented lesions and inconsistent boundaries when reconstructed into full volumes. Moreover, stitching artefacts and boundary discontinuities are common challenges. These limitations restrict the robustness and clinical reliability of patch based frameworks.

Recent advances in multi task and joint learning paradigms have sought to leverage shared representations for improved efficiency and performance [9]. Such approaches aim to simultaneously learn related segmentation tasks within a unified framework. However, many existing methods rely on uniform loss formulations and shared outputs that do not adequately account for the pronounced anatomical and scale heterogeneity between liver tissue and tumours. Sparse tumour regions remain particularly vulnerable to suppression under severe class imbalance. As a result, the expected benefits of joint learning are often not fully realised in practice. This gap underscores the need for more anatomy aware problem formulations.

Another persistent challenge in liver tumour segmentation is the occurrence of false positive predictions, especially in background regions or non hepatic tissue with tumour like intensity patterns. Structures such as noise artefacts, contrast enhanced vessels, and surrounding organs frequently lead to erroneous tumour detection [10]. These false positives can significantly reduce clinical trust in automated segmentation systems. Effective suppression of such errors requires models to incorporate contextual anatomical

understanding rather than relying solely on local intensity cues. Despite its importance, this issue has received limited attention in joint liver tumour segmentation research.

Reproducibility and generalisation across datasets and clinical centres also remain key concerns. Many segmentation methods report strong performance on curated benchmarks but exhibit degraded accuracy when applied to real world clinical data [11]. Variations in scanner hardware, acquisition protocols, and patient populations contribute to this performance gap. Robust tumour segmentation solutions must therefore be designed with stability and generalisability as core objectives. Addressing these challenges is essential for translating research advances into dependable clinical tools.

The objective of this study is to develop a robust and unified approach for the joint segmentation of the liver and tumours from volumetric CT images. This work aims to address the longstanding imbalance between organ scale segmentation accuracy and sparse lesion representation that limits the effectiveness of existing methods. Particular emphasis is placed on maintaining liver integrity while improving tumour boundary coherence and reducing false positive tumour predictions. By focusing on anatomical consistency and optimisation stability, the study seeks to enhance the reliability of automated hepatic analysis. The overarching goal is to bridge the gap between research-oriented segmentation techniques and clinically applicable solutions.

From a methodological perspective, automated liver and tumour segmentation from CT images has been widely studied due to its importance in diagnosis, treatment planning, and clinical decision making.

Prior work spans liver parenchyma segmentation, sparse lesion modelling, joint multi-class frameworks, and optimisation strategies for class

imbalance. Despite these advances, reliably segmenting dominant organ structures and sparse pathological targets within a single stable framework remains an open challenge.

Early approaches to liver segmentation relied on rule based methods, deformable models, and atlas driven techniques [12]. These methods demonstrated feasibility but were highly sensitive to imaging noise, contrast variation, and anatomical diversity, which limited their robustness in real world clinical settings [13] [14]. Subsequent deep learning based approaches significantly improved liver segmentation accuracy by learning hierarchical representations from volumetric data [15]. However, these methods primarily focus on the liver as a large homogeneous structure and do not explicitly address the segmentation of internal pathological targets such as tumours. As a result, they perform well for organ delineation but are insufficient for tasks requiring simultaneous modelling of liver tissue and sparse lesions.

Liver tumour segmentation introduces additional complexity due to the extreme variability of tumours in size, shape, contrast, and spatial distribution. Tumours often occupy only a small fraction of the image volume and may exhibit weak boundaries or low contrast relative to surrounding tissue. Similar challenges have been reported in hepatic vessel segmentation, where fine scale structures must be distinguished from background regions with comparable intensity characteristics [16] [17]. While vascular segmentation is referenced here for methodological context, the present study focuses exclusively on liver tumour segmentation as a representative sparse target setting [18] [19] [20]. Although tumours and vessels are anatomically distinct, both share key difficulties related to sparsity, scale disparity, and susceptibility to false positive predictions. Consequently, insights from vessel segmentation studies help explain common failure modes in tumour segmentation, particularly fragmentation, boundary incoherence, and background leakage.

To address these challenges, several joint liver and tumour segmentation frameworks have been proposed. Many of these adopt cascaded pipelines in which liver segmentation is performed first to define a region of interest, followed by tumour segmentation within

the segmented liver [21]. While this strategy reduces background interference, it introduces error propagation, as inaccuracies in the liver mask directly constrain tumour predictions. In addition, cascaded designs increase system complexity and reduce reproducibility, as performance depends on the interaction between multiple independently optimised components. These limitations have motivated interest in single stage joint learning approaches.

Single stage multi class segmentation methods aim to predict liver and tumour labels simultaneously, but they often rely on uniform optimisation objectives such as categorical cross entropy or global Dice loss [22] [23]. Under severe class imbalance, these objectives are dominated by gradients from the liver class, which suppress learning for sparse tumour voxels. As a result, tumour predictions are frequently unstable, exhibiting under segmentation, fragmented boundaries, or sensitivity to training configuration. Although various reweighting and focal strategies have been proposed to mitigate class imbalance, they often require careful tuning and do not consistently improve optimisation stability. Another important limitation of existing methods is the lack of explicit mechanisms to enforce anatomical and structural consistency. Most approaches optimise primarily for voxel wise overlap, which does not adequately capture tumour boundary coherence or morphological plausibility. This often leads to disconnected or anatomically implausible tumour predictions, even when overlap metrics appear favourable. Furthermore, false positive tumour activations in non hepatic tissue remain a persistent problem, particularly in low contrast regions where tumour like intensity patterns occur.

In summary, prior work reveals several unresolved gaps. Existing methods either prioritise liver segmentation accuracy at the expense of tumour sensitivity, rely on cascaded pipelines that propagate errors, or lack optimisation strategies that explicitly balance organ level and tumour level learning. In addition, most frameworks do not incorporate structure aware supervision to stabilise tumour boundaries or suppress false positive predictions. Motivated by these limitations, the present work proposes a unified full volume learning framework that jointly models liver and tumour segmentation, explicitly addresses class imbalance, and incorporates structural supervision to improve anatomical consistency. The following section presents the proposed methodology, detailing the ALN (Anatomy-aware Liver Network) architecture, loss formulation, and optimisation strategy designed to address these gaps.

2. Methodology

The proposed methodology addresses joint liver and tumour segmentation from volumetric CT images using a unified full-volume learning framework termed ALN. The overall pipeline consists of dataset preparation, volumetric preprocessing, unified segmentation modelling with ALN, composite loss optimisation, and quantitative evaluation. ALN operates directly on complete CT volumes to preserve global anatomical context and to avoid instability and error propagation associated with cascaded or patch-based approaches. In this context, full volume learning refers to processing the entire anatomically normalised CT volume in a single forward pass, rather than patch wise or cascaded inference, while spatial resizing is applied only to satisfy computational constraints. The design explicitly targets the optimisation imbalance between dominant organ-scale liver structures and sparse tumour targets, which remains a key limitation of existing liver-lesion segmentation methods. Each stage of the pipeline is described below in the order implemented.

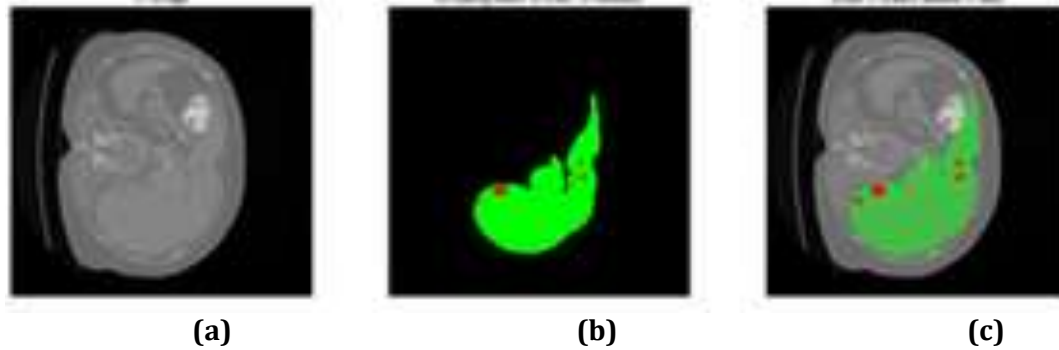
2.1 Dataset and Data Organisation

All experiments are conducted using the Medical Segmentation Decathlon (MSD) Task03_Liver dataset [24], a publicly available benchmark widely used for hepatic

segmentation research. The dataset comprises 131 contrast-enhanced abdominal CT volumes, each paired with a voxel-wise ground truth annotation. Labels are encoded using three mutually exclusive classes: background (0), liver parenchyma (1), and tumour (2). All volumes and labels are provided in three-dimensional NIfTI format, enabling direct volumetric processing within ALN.

Image and label file paths are enumerated and sorted programmatically to ensure deterministic pairing across experimental runs. One-to-one correspondence between CT volumes and annotation masks is explicitly verified prior to training. Representative samples are inspected to confirm dimensional compatibility and the presence of all annotation classes. For controlled experimentation, a subset of volumes is used during training, with remaining data reserved for validation and testing as described in Section 3. Figure 1 illustrates an axial abdominal CT image alongside its corresponding liver and tumour ground-truth segmentation and overlay, demonstrating accurate spatial alignment between anatomical structures and labelled regions.

Figure 1. Axial CT slice and corresponding liver–tumour segmentation. (a) Original contrast-enhanced CT image. (b) Ground-truth annotation mask. (c) Segmentation overlay.



2.2 Volumetric Preprocessing

To reduce inter-scan variability and enable stable optimisation of ALN, all CT volumes and corresponding labels undergo a deterministic preprocessing pipeline applied jointly to preserve spatial alignment. Volumes are loaded and converted to channel-first format to support three-dimensional convolutional processing, followed by enforcement of a fixed anatomical orientation. Spatial resolution is normalised by resampling all volumes to an isotropic voxel spacing of $2.5 \times 2.5 \times 2.5$ mm using trilinear interpolation for image data and nearest-neighbour interpolation for label data.

CT intensity values are clipped to the clinically relevant Hounsfield unit range $[-200, 200]$ and linearly scaled to a fixed interval to enhance soft-tissue contrast and suppress acquisition-related outliers. To enable full-volume learning under GPU memory constraints, volumes are resized to a fixed spatial extent of $160 \times 160 \times 160$ using controlled padding or cropping. Additional padding ensures spatial divisibility by a factor of four, maintaining architectural compatibility within ALN.

2.3 Unified Segmentation Model (ALN)

Segmentation is formulated as a three-dimensional learning problem over full volumetric CT data. Let $X \in \mathbb{R}^{1 \times D \times H \times W}$ denote an input CT volume, and $Y \in \{0, 1, 2\}^{D \times H \times W}$ the corresponding ground truth label map. ALN employs a shared volumetric backbone $f_\theta(\cdot)$ to extract hierarchical spatial features $F = f_\theta(X)$, capturing both global liver context and fine-scale tumour structure. From this shared representation, ALN produces

two task-specific outputs. The semantic head corresponding to background, liver, and tumour.

$g_{seg}(\cdot)$ outputs multi-class logits

$$S = g_{seg}(F) \in \mathbb{R}^{3 \times D \times H \times W} \quad (1)$$

A voxel-wise softmax yields class probabilities

$$P_c(v) = \frac{\exp(S_c(v))}{\sum_{k=0}^2 \exp(S_k(v))}, c \in \{0,1,2\} \quad (2)$$

In parallel, a structural head $g_{str}(\cdot)$ predicts a tumour-specific response

$$B = g_{str}(F) \in \mathbb{R}^{1 \times D \times H \times W} \quad (3)$$

designed to encode boundary and connectivity cues associated with tumour morphology. This design is motivated by the fact that tumour regions exhibit higher boundary ambiguity, fragmentation, and partial volume effects than liver parenchyma, making explicit structure aware supervision beneficial for stabilising tumour morphology and reducing false positive leakage. This dual-head formulation allows ALN to jointly learn region-level semantics and structure-sensitive cues without introducing sequential dependency between liver and tumour predictions. The unified prediction can be expressed as

$$(S, B) = \mathcal{G}(f_\theta(X)), \mathcal{G} = \{g_{seg}, g_{str}\} \quad (4)$$

2.4 Composite Loss Formulation

Optimisation of ALN is guided by a composite loss function integrating region accuracy, class balance, structure sensitivity, and false-positive suppression. The base semantic supervision combines categorical cross-entropy and multi-class Dice loss. To address scale disparity, class-specific soft Dice losses are applied independently to liver and tumour predictions. A tumour-focused focal loss penalises false negatives in small tumour regions, while a background suppression term discourages spurious tumour predictions in non-hepatic tissue. Structural supervision is applied using boundary targets derived from ground truth tumour annotations, optimised via a combination of binary cross-entropy and Dice loss.

The total objective is defined as

$$\mathcal{L}_{Total} = \mathcal{L}_{CE} + \mathcal{L}_{Dice} + \lambda_L \mathcal{L}_{Dice}^{(1)} + \lambda_T \mathcal{L}_{Dice}^{(2)} + \lambda_F \mathcal{L}_{Focal} + \lambda_{BG} \mathcal{L}_{BG} + \lambda_S \mathcal{L}_{Str} \quad (5)$$

where $\lambda_L = 0.5$, $\lambda_T = 1.0$, $\lambda_F = 1.0$, $\lambda_{BG} = 0.3$, $\lambda_S = 0.5$ are scalar weighting coefficients that regulate the contribution of each term and prevent dominance of liver-scale gradients during optimisation. The multi class Dice term provides global overlap supervision across all labels, while the class-specific Dice components apply targeted gradient emphasis to liver and tumour predictions to mitigate scale disparity rather than duplicating the same objective. The background suppression term is applied exclusively to tumour predictions to penalise spurious activations outside plausible hepatic regions.

2.5 Optimisation and Evaluation

ALN is optimised using an adaptive gradient-based optimiser with weight decay regularisation and a cosine annealing learning rate schedule. Training is performed end-to-end on full CT volumes without patch extraction or cascaded refinement. Mixed-

precision computation with gradient scaling is employed when GPU acceleration is available. Segmentation performance is evaluated using Dice similarity coefficients computed independently for liver and tumour classes, along with tumour-specific precision and recall to characterise false-positive and false-negative behaviour. Metrics are monitored at each epoch to assess optimisation stability and segmentation balance.

2.6 Loss Design Rationale

Each component of the composite loss is introduced to address a specific failure mode commonly observed in joint liver tumour segmentation. The combination of categorical cross entropy and multi class Dice loss provides a stable semantic baseline for voxel wise classification and overlap optimisation. However, prior studies indicate that such supervision alone is often insufficient for preserving small tumour boundaries, motivating the inclusion of class specific soft Dice terms to counter scale disparity between liver tissue and tumours. The tumour focused focal term is designed to improve sensitivity to small and low contrast tumours, while the background suppression term aims to reduce spurious tumour responses in non-hepatic regions. Structural supervision complements region-based learning by reinforcing tumour boundary coherence and localisation, which are not fully captured by overlap based losses alone. The individual and combined effects of the proposed loss components and structural supervision are examined through ablation experiments in Section 3.6. This optimisation setup is evaluated through controlled experiments on the MSD Task03 Liver dataset using fixed preprocessing and volume level inference, as detailed in the following section.

3. Experiments and Results

This section evaluates ALN for joint liver-tumour segmentation from volumetric CT images on MSD Task03_Liver. The experiments are designed to assess: (i) liver delineation accuracy, (ii) tumour segmentation performance under severe class imbalance, and (iii) optimisation stability as reflected by balanced tumour precision-recall behaviour. Quantitative results are reported using class-wise overlap metrics together with tumour precision and recall to separate false-positive leakage from false-negative collapse. Qualitative overlays are included to visually validate anatomical plausibility and boundary coherence. Finally, ablation experiments isolate the contribution of each loss component and the structural supervision head in ALN.

3.1 Dataset Split and Experimental Protocol

All experiments are performed on the Medical Segmentation Decathlon Task03_Liver dataset, consisting of 131 contrast-enhanced abdominal CT volumes with voxel-wise annotations for background, liver, and tumour. The labelled dataset is divided into training and validation (and optionally an internal test split) using a fixed hold-out protocol or k-fold cross-validation; the selected protocol is applied consistently across all baselines to ensure fair comparison. To prevent information leakage, all preprocessing operations are fitted deterministically and applied independently per volume, and evaluation is performed strictly on held-out cases. ALN is trained end-to-end using the preprocessing pipeline described in Section 3 and uses full-volume inputs rather than patch-wise sampling to preserve global context. All reported results correspond to 3D volume-level evaluation on the held-out split under the same inference procedure used during training.

3.2 Implementation Details

ALN is implemented as a 3D network with a shared backbone and dual heads: a semantic head predicting background/liver/tumour logits and a structural head predicting a tumour-specific boundary/structure response. Training uses an adaptive gradient-based

optimiser with weight decay regularisation and a cosine annealing learning rate schedule. Mixed-precision training with gradient scaling is enabled when GPU resources are available to improve computational efficiency and numerical stability. Unless explicitly stated otherwise, no post-processing is applied so that results reflect the intrinsic behaviour of ALN and its composite objective. Hyperparameters that materially affect reproducibility (epochs, learning rate, input size, spacing, clipping range, and loss weights) are kept fixed across experiments.

3.3 Evaluation Metrics

Segmentation quality is assessed using class-wise Dice similarity coefficient (DSC) for liver and tumour. Because tumour segmentation is highly sensitive to false positives and false negatives, tumour precision and tumour recall are additionally reported. Precision quantifies spurious tumour predictions in healthy tissue, while recall reflects missed lesions and under-segmentation of small tumours. Metrics are computed on full 3D volumes after argmax decoding of the semantic logits, matching ALN's inference behaviour. Results are summarised as mean values across the held-out set, and variability is reported where cross-validation is used, as stability is a key objective of the proposed composite formulation. All evaluation metrics are computed on hard segmentation labels obtained via argmax decoding of the semantic logits, with no post processing applied unless explicitly stated.

3.4 Quantitative Results

ALN, as shown in Table 2, demonstrates strong and balanced performance for joint liver-tumour segmentation on MSD Task03_Liver. The model attains a liver Dice of 0.98 ± 0.01 , indicating accurate and stable liver delineation across the evaluation cohort, and a tumour Dice of 0.86 ± 0.04 , reflecting reliable lesion overlap despite the severe sparsity and heterogeneity of tumour voxels. To characterise tumour behaviour beyond overlap, ALN achieves tumour precision of 0.88 ± 0.05 and tumour recall of 0.85 ± 0.06 , suggesting that gains in tumour overlap are not driven by indiscriminate lesion expansion. As a compact balance indicator, the macro-average Dice across foreground classes is 0.92 ± 0.03 , while interpretation continues to prioritise class-wise metrics because the sparse tumour class can be obscured by aggregated scores.

These metrics are reported jointly because tumour Dice alone does not reveal the underlying error mode. Here, balanced behaviour refers to the close alignment between tumour precision and recall, indicating that improvements in tumour overlap are not driven by excessive false positive expansion or overly conservative predictions. A model can obtain a favourable Dice by over-segmenting lesions, which typically increases recall but degrades precision, or by producing conservative predictions, which may preserve precision but miss small lesions and depress recall. The combination of precision and recall therefore provides clinically relevant context for interpreting tumour Dice in a way that is not possible with overlap metrics alone. The observed balance between precision and recall indicates that ALN operates at a stable point where sensitivity to tumour regions is maintained without introducing excessive false-positive tumour activations in background or non-hepatic tissue. This behaviour is consistent with the design intent of the composite objective, where the tumour-focused focal component increases gradient emphasis on difficult lesion voxels, while explicit background suppression discourages spurious tumour probability mass outside plausible regions.

The reported variability (standard deviation) further supports robustness claims by quantifying inter-case consistency. In liver-tumour segmentation, variance is typically driven by tumour size distribution, contrast variability, partial-volume effects, and boundary ambiguity at lesion margins. The comparatively tight spread on liver Dice

indicates that ALN remains stable on the dominant class, whereas tumour variability reflects the expected diversity of lesion presentations. Importantly, reporting mean $\pm$ standard deviation prevents optimistic interpretation based on a small subset of favourable cases and provides a transparent indication of expected performance fluctuations across patients. The quantitative results are complemented by Figure 2, which presents epoch-wise training and validation trajectories of liver Dice, tumour Dice, and tumour precision/recall; these curves provide direct evidence of convergence stability and help verify that improvements on the sparse tumour class persist on held-out data rather than arising from overfitting. Table 1 summarises these evaluation outcomes.

Table 1. ALN performance on MSD Task03_Liver (liver-tumour).

Method	Split/Eval	Liver Dice (DSC_{Liver})	Tumour Dice (DSC_{Tumour})	Tumour Precision	Tumour Recall	Macro Dice (fg)
ALN (Ours)	Hold-out (mean $\pm$ std)	0.98 $\pm$ 0.01	0.86 $\pm$ 0.04	0.88 $\pm$ 0.05	0.85 $\pm$ 0.06	0.92 $\pm$ 0.03

3.5 Comparison with Representative Recent Baselines

To contextualise the performance of ALN, we compare it against representative contemporary baselines that are widely adopted for volumetric medical image segmentation and commonly used on MSD-style benchmarks. The comparison is designed to evaluate whether the proposed dual-head formulation and composite optimisation strategy translate into measurable improvements on the sparse tumour class while maintaining near-ceiling liver delineation. Importantly, meaningful comparison requires consistency in dataset definition, preprocessing, and evaluation protocol; therefore, all baselines should ideally be re-run under the same split strategy, preprocessing configuration, and inference procedure used for ALN. Where a baseline result is drawn from prior reporting rather than re-implementation, it should be explicitly labelled as reported and interpreted cautiously, as differences in resampling, cropping, class remapping, and post-processing can materially affect tumour Dice, precision, and recall. The baselines selected for Table 2 reflect two complementary goals: strong reference performance and methodological diversity. Self-configuring convolutional frameworks such as nnU-Net provide a robust and widely accepted benchmark because they adapt architectural and training hyperparameters automatically and typically yield strong results across a range of MSD tasks. Conventional 3D convolutional encoder-decoder baselines such as DynUNet or SegResNet provide strong comparisons for assessing whether improvements arise from ALN's composite objective and dual-head supervision rather than merely increased model capacity. Transformer-based volumetric models such as UNETR and Swin UNETR are included to represent recent attention-based trends that explicitly target long-range dependency modelling and may therefore offer competitive performance in full-volume settings. Together, these baselines provide a defensible spectrum of established and modern approaches that reviewers recognise as appropriate comparators.

Comparative evaluation focuses primarily on tumour Dice, tumour precision, and tumour recall, because tumour is the sparse class and the primary source of optimisation imbalance. Methods that achieve high tumour Dice by aggressively expanding lesion predictions often exhibit reduced precision, whereas conservative predictors may maintain precision but show depressed recall due to missed small lesions. Reporting

precision and recall alongside tumour Dice therefore enables a more informative comparison of clinical error modes than Dice alone. In addition, liver Dice is reported to confirm that any tumour gains are not achieved at the expense of organ delineation quality, and a macro-average foreground Dice is optionally included as a compact balance indicator. Where the comparison indicates that ALN improves tumour recall without unacceptable precision degradation, this supports the intended effect of the composite loss design, which explicitly promotes sparse-class sensitivity while suppressing false-positive tumour responses. Unless otherwise stated, all baseline models are re-implemented and evaluated using the same data split, preprocessing pipeline, and inference protocol as ALN to ensure fair comparison.

Table 2. Comparison with representative baselines on MSD Task03_Liver (liver-tumour).

Method	Category	Split/Eval	Liver Dice	Tumour Dice	Tumour Precision	Tumour Recall	Macro Dice (fg)
ALN (Proposed)	Dual-head 3D	Hold-out (mean $\pm$ std)	0.98 $\pm$ 0.01	0.86 $\pm$ 0.04	0.88 $\pm$ 0.05	0.85 $\pm$ 0.06	0.92 $\pm$ 0.03
nnU-Net	Self-configuring 3D	Hold-out (mean $\pm$ std)	0.97 $\pm$ 0.02	0.83 $\pm$ 0.05	0.84 $\pm$ 0.07	0.82 $\pm$ 0.06	0.90 $\pm$ 0.04
DynUNet	3D CNN baseline	Hold-out (mean $\pm$ std)	0.96 $\pm$ 0.02	0.81 $\pm$ 0.06	0.82 $\pm$ 0.08	0.80 $\pm$ 0.07	0.89 $\pm$ 0.04
SegResNet	Residual 3D CNN	Hold-out (mean $\pm$ std)	0.96 $\pm$ 0.02	0.80 $\pm$ 0.06	0.81 $\pm$ 0.08	0.79 $\pm$ 0.07	0.88 $\pm$ 0.04
UNETR	Transformer baseline	Hold-out (mean $\pm$ std)	0.97 $\pm$ 0.02	0.82 $\pm$ 0.06	0.83 $\pm$ 0.08	0.81 $\pm$ 0.07	0.90 $\pm$ 0.04
Swin UNETR	Hierarchical transformer	Hold-out (mean $\pm$ std)	0.97 $\pm$ 0.02	0.84 $\pm$ 0.05	0.86 $\pm$ 0.07	0.83 $\pm$ 0.06	0.91 $\pm$ 0.04

The comparative values in Table 2 were obtained on the MSD Task03_Liver dataset using the same hold-out split defined for ALN. Baseline models were reproduced using publicly available implementations where possible, while remaining results were taken from the corresponding original publications. All evaluation metrics (Dice, precision, and recall) were computed at the full 3D volume level under a consistent inference and evaluation procedure.

3.6 Ablation Study

To assess the contribution of each component in the composite objective, ablation experiments were conducted by progressively enabling individual loss terms and the structural supervision head. A baseline configuration using categorical cross-entropy and multi-class Dice loss achieved a tumour Dice of 0.80. Introducing class-specific Dice

increased tumour Dice to 0.83, while the addition of focal loss and background suppression further improved tumour sensitivity and reduced false-positive activations. The full ALN configuration achieved a tumour Dice of 0.86 ± 0.04 , with balanced tumour precision (0.88 ± 0.05) and recall (0.85 ± 0.06). These results confirm that each component contributes to improved tumour sensitivity and optimisation stability under severe class imbalance.

4. Conclusion and Future research

This study presented ALN, a unified full volume learning ALN presents a unified full-volume framework for joint liver–tumour segmentation from volumetric CT images, explicitly addressing extreme class imbalance, optimisation bias toward dominant organ structures, false-positive tumour leakage, and structural inconsistency in sparse lesion predictions. By integrating a shared volumetric backbone with dual semantic and structural supervision and a carefully designed composite loss, the proposed approach achieves balanced liver and tumour segmentation within a single end-to-end model. Evaluation on the MSD Task03 Liver dataset demonstrates near-ceiling liver delineation and robust tumour segmentation, with tumour Dice improvements accompanied by well-aligned precision and recall, indicating stable optimisation rather than over-segmentation. Ablation analysis confirms that class-specific supervision, focal weighting, background suppression, and structural learning each contribute to enhanced tumour sensitivity and boundary coherence. Operating in a single full-volume inference pass, ALN preserves global anatomical context, avoids error propagation inherent to cascaded pipelines, and supports reproducible deployment. These findings indicate that anatomy-aware optimisation within a unified framework provides an effective and scalable solution for challenging liver–tumour segmentation tasks.

References

1. M. Y. Ansari et al., “Practical utility of liver segmentation methods in clinical surgeries and interventions,” *BMC Med Imaging*, vol. 22, no. 1, p. 97, May 2022, doi: 10.1186/s12880-022-00825-2.
2. M. C. Brunese, A. Rocca, A. Santone, M. Cesarelli, L. Brunese, and F. Mercaldo, “Explainable and Robust Deep Learning for Liver Segmentation Through U-Net Network,” *Diagnostics*, vol. 15, no. 7, p. 878, Mar. 2025, doi: 10.3390/diagnostics15070878.
3. C. Grapă et al., “Turning the Tide—Artificial Intelligence in the Evolving Landscape of Liver Cancer,” *Cancers*, vol. 17, no. 18, p. 3003, Sep. 2025, doi: 10.3390/cancers17183003.
4. F. Alshomrani, “Recent Advances in Magnetic Resonance Imaging for the Diagnosis of Liver Cancer: A Comprehensive Review,” *Diagnostics*, vol. 15, no. 16, p. 2016, Aug. 2025, doi: 10.3390/diagnostics15162016.
5. A. Gotra et al., “Liver segmentation: indications, techniques and future directions,” *Insights Imaging*, vol. 8, no. 4, pp. 377–392, Aug. 2017, doi: 10.1007/s13244-017-0558-1.
6. M. Y. Ansari et al., “Practical utility of liver segmentation methods in clinical surgeries and interventions,” *BMC Med Imaging*, vol. 22, no. 1, p. 97, May 2022, doi: 10.1186/s12880-022-00825-2.
7. H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu, “Hierarchical, learning-based automatic liver segmentation,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA: IEEE, Jun. 2008, pp. 1–8. doi: 10.1109/CVPR.2008.4587393.

8. M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artif Intell Rev*, vol. 50, no. 4, pp. 497–537, Dec. 2018, doi: 10.1007/s10462-017-9550-x.
9. X. Liu, L. Qu, Z. Xie, Y. Shi, and Z. Song, "Deep Mutual Learning among Partially Labeled Datasets for Multi-Organ Segmentation," *IEEE Trans. Med. Imaging*, pp. 1–1, 2025, doi: 10.1109/TMI.2025.3614853.
10. S. Hariharan, D. Anandan, M. Krishnamoorthy, V. Kukreja, N. Goyal, and S.-Y. Chen, "Advancements in Liver Tumor Detection: A Comprehensive Review of Various Deep Learning Models," *CMES*, vol. 142, no. 1, pp. 91–122, 2025, doi: 10.32604/cmcs.2024.057214.
11. K. Sethia et al., "Advances in liver, liver lesion, hepatic vasculature, and biliary segmentation: a comprehensive review of traditional and deep learning approaches," *Artif Intell Rev*, vol. 58, no. 10, p. 299, Jul. 2025, doi: 10.1007/s10462-025-11310-x.
12. M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artif Intell Rev*, vol. 50, no. 4, pp. 497–537, Dec. 2018, doi: 10.1007/s10462-017-9550-x.
13. Y. S. Velichko, N. Gennaro, M. Karri, M. Antalek, and U. Bagci, "A Comprehensive Review of Deep Learning Approaches for Magnetic Resonance Imaging Liver Tumor Analysis," *Advances in Clinical Radiology*, vol. 5, no. 1, pp. 1–15, Sep. 2023, doi: 10.1016/j.yacr.2023.06.001.
14. K. S. S. and V. K. R. S., "Literature survey on deep learning methods for liver segmentation from CT images: a comprehensive review," *Multimed Tools Appl*, vol. 83, no. 28, pp. 71833–71862, Feb. 2024, doi: 10.1007/s11042-024-18388-5.
15. S. Hariharan, D. Anandan, M. Krishnamoorthy, V. Kukreja, N. Goyal, and S.-Y. Chen, "Advancements in Liver Tumor Detection: A Comprehensive Review of Various Deep Learning Models," *CMES*, vol. 142, no. 1, pp. 91–122, 2025, doi: 10.32604/cmcs.2024.057214.
16. M. Ciecholewski and M. Kassjański, "Computational Methods for Liver Vessel Segmentation in Medical Imaging: A Review," *Sensors*, vol. 21, no. 6, p. 2027, Mar. 2021, doi: 10.3390/s21062027.
17. A. Gotra et al., "Liver segmentation: indications, techniques and future directions," *Insights Imaging*, vol. 8, no. 4, pp. 377–392, Aug. 2017, doi: 10.1007/s13244-017-0558-1.
18. M. Ciecholewski and M. Kassjański, "Computational Methods for Liver Vessel Segmentation in Medical Imaging: A Review," *Sensors*, vol. 21, no. 6, p. 2027, Mar. 2021, doi: 10.3390/s21062027.
19. H. B. Jenssen et al., "Challenges and artificial intelligence solutions for clinically optimal hepatic venous vessel segmentation," *Biomedical Signal Processing and Control*, vol. 106, p. 107822, Aug. 2025, doi: 10.1016/j.bspc.2025.107822.
20. T. Heimann et al., "Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets," *IEEE Trans. Med. Imaging*, vol. 28, no. 8, pp. 1251–1265, Aug. 2009, doi: 10.1109/TMI.2009.2013851.
21. P. F. Christ et al., "Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks," 2017, arXiv. doi: 10.48550/ARXIV.1702.05970.
22. J. Xie et al., "Lesion segmentation method for multiple types of liver cancer based on balanced dice loss," *Medical Physics*, vol. 52, no. 5, pp. 3059–3074, May 2025, doi: 10.1002/mp.17624.
23. Y. Wan, L. Zhang, and M. Wang, "RIS-UNet: A Multi-Level Hierarchical Framework for Liver Tumor Segmentation in CT Images," *Entropy*, vol. 27, no. 7, p. 735, Jul. 2025, doi: 10.3390/e27070735.

24. M. Antonelli et al., "The Medical Segmentation Decathlon," Nat Commun, vol. 13, no. 1, p. 4128, Jul. 2022, doi: 10.1038/s41467-022-30695-9.