Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ADEST-U-Net: A Dual-Encoder Architecture for Cross-Organ and Cross-Modality Transfer Learning in Data-Scarce Medical Image Segmentation

Riham Jeeballah
Sultan Qaboos Univerity
Hamza ZIDOUM
zidoum@squ.edu.om
Sultan Qaboos Univerity
Adhari Al zaabi
Sultan Qaboos University
Abdelhamid Abdessalem
Sultan Qaboos Univerity

Research Article

Keywords: liver tumor segmentation, transfer learning, cross-organ transfer, cross-modality adaptation, dual-encoder architecture, computed tomography, magnetic resonance imaging

Posted Date: July 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9700605/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ADEST-U-Net: A Dual-Encoder Architecture for Cross-Organ and Cross-Modality Transfer Learning in Data-Scarce Medical Image Segmentation

Riham Jeeballah[1], Hamza Zidoum[1*], Adhari Al Zaabi[2], Abdelhamid Abdesselam[1]

[1]Department of Computer Science, Sultan Qaboos University, Muscat, Oman.

[2]Department of Clinical and Human Anatomy, Sultan Qaboos University, Muscat, Oman.

*Corresponding author(s). E-mail(s): zidoum@squ.edu.om;

Abstract

Medical image segmentation often relies on large volumes of annotated data, which are frequently unavailable in clinical practice. This limitation is particularly evident in liver tumor segmentation from computed tomography (CT), where task-aligned pretrained models are scarce. This study investigates cross-organ and cross-modality transfer learning by adapting representations learned from brain tumor segmentation in magnetic resonance imaging (MRI) to liver tumor segmentation in CT under data-scarce conditions.

We propose the Adapted Dual-Encoder Shared-Tail U-Net (ADEST-U-Net), a segmentation architecture that integrates an MRI-pretrained encoder with a CT-specific encoder trained from scratch. Feature representations from both encoders are fused through a shared refinement module before decoding, enabling controlled interaction between transferable and domain-specific features.

Experiments were conducted on the publicly available 3D-IRCADb-01 dataset. Compared with supervised UNet++ trained from scratch and naïve transfer configurations, the proposed model achieved a Dice coefficient of 0.7189 and a precision of 0.7727. Notably, false positives were reduced by 67.6% relative to training from scratch and by 40.4% compared with frozen-weight transfer. Qualitative evaluation further demonstrated improved spatial coherence, with predictions largely confined to liver parenchyma and fewer extrahepatic false detections.

Keywords: liver tumor segmentation, transfer learning, cross-organ transfer, cross-modality adaptation, dual-encoder architecture, computed tomography, magnetic resonance imaging

1 Introduction

Medical image segmentation plays a critical role in computer-assisted clinical workflows, supporting diagnosis, treatment planning, and longitudinal disease monitoring. In liver oncology, accurate tumor delineation from computed tomography (CT) scans is essential for surgical planning, therapy assessment, and outcome evaluation [1]. Despite substantial progress in convolutional neural network (CNN)-based segmentation, robust performance remains heavily dependent on large volumes of annotated training data.

Encoder–decoder architectures such as U-Net [2] and UNet++ [3] have become standard baselines for biomedical segmentation. However, these models typically assume access to sufficiently large, task-specific datasets. In clinical practice, the acquisition of expert-annotated medical images is costly and time-consuming, particularly for tumor segmentation tasks requiring precise boundary delineation [4]. This data scarcity motivates the exploration of transfer learning strategies aimed at reusing knowledge from related domains.

Transfer learning has been widely adopted in medical imaging, primarily focusing on same-organ or same-modality adaptation [5, 6]. Cross-modality adaptation (e.g., CT to MRI) has shown moderate success when anatomical structures are preserved across domains [7, 8]. More recently, cross-organ and cross-modality adaptation has begun to attract attention as a strategy for improving generalization across heterogeneous medical imaging domains [9]. However, systematic investigations of transfer under simultaneous anatomical and modality shifts remain limited.

Prior investigations suggest that naïve fine-tuning under severe anatomical and modality shifts yields limited improvements [10]. These observations raise a fundamental methodological question: under extreme domain mismatch, does single-encoder transfer effectively reuse high-level representations, or does domain-specific interference degrade performance? Addressing this question requires architectural mechanisms that explicitly distinguish transferable features from domain-specific information, rather than relying solely on conventional fine-tuning.

In this study, we propose the Adapted Dual-Encoder Shared-Tail U-Net (ADEST-U-Net), a transfer learning architecture designed to mitigate cross-organ and cross-modality representation mismatch under data-scarce conditions. The model integrates an encoder pretrained on brain tumor MRI with a CT-specific encoder trained from scratch. By fusing features through a shared refinement module prior to decoding, the architecture aims to balance transferable representations with modality-specific adaptations.

The primary contribution of this work is an architectural and empirical investigation of cross-organ and cross-modality transfer learning for medical image segmentation. We systematically analyze the limitations of single-encoder transfer and

demonstrate how dual-encoder feature integration can improve segmentation stability and reduce false detections in liver CT tumor segmentation.

2 Literature Review

Recent years have seen substantial progress in liver and tumor segmentation, primarily benchmarked on publicly available datasets such as the Liver Tumor Segmentation Challenge (LiTS-2017) and the 3DIRCADb-01 dataset. These datasets have become widely adopted reference standards for evaluating segmentation performance, typically reported using the Dice Similarity Coefficient (DSC) for liver and tumor regions.

A wide range of deep learning architectures have been proposed to improve segmentation accuracy on these benchmarks. HANS-Net [11] integrates hybrid attention mechanisms, hierarchical uncertainty estimation, and anatomical priors to enhance segmentation precision and anatomical plausibility. Similarly, CAFCT-Net [12] combines convolutional and transformer blocks within a cascaded framework, while Jiang et al. [13] incorporate lightweight context-aware modules into a SwinUNet backbone to strengthen global context modeling. Yashaswini et al. [14] explored modified UNet and ResUNet architectures on 3DIRCADb, demonstrating the continued relevance of encoder–decoder designs.

Additional architectural refinements include spatial-wise transformer integration (SWTR-UNet) [15], dual-path contextual encoding (DPC-Net) [16], and multi-scale channel attention with deep residual learning (ELANRes-MSCA-UNet) [17]. Earlier foundational work by Chlebus et al. [18] established early deep learning baselines for LiTS using a two-stage fully convolutional framework.

Across these studies, a consistent observation emerges: liver segmentation typically achieves high accuracy due to the organ's relatively consistent structure, whereas liver tumor segmentation remains significantly more challenging. Tumor variability in size, morphology, and contrast continues to limit performance despite increasingly sophisticated architectural components such as attention mechanisms, transformers, and multi-scale fusion strategies.

While architectural innovation has steadily improved benchmark performance, a more fundamental challenge persists: the limited availability of high-quality annotated medical datasets. Manual tumor delineation requires expert radiological interpretation and remains labor-intensive and time-consuming. To address annotation scarcity, complementary strategies have been explored, including synthetic data generation, unsupervised domain adaptation, and transfer learning.

GAN-based synthesis methods attempt to reduce annotation dependence by generating realistic artificial images for data augmentation. Rafiq et al. [19] proposed an enhanced CycleGAN framework to synthesize MRI from CT while preserving anatomical consistency, reporting modest segmentation gains when combining synthetic and real data. Although promising, such approaches are often evaluated within constrained institutional settings and may introduce modality translation artifacts.

Unsupervised domain adaptation (UDA) methods aim to transfer knowledge across modalities without requiring paired data or labeled target images. Yang et al. [20]

proposed a feature alignment framework with cross-pseudo supervision for whole-heart MRI segmentation, while Chen et al. [21] introduced ARL-GAN to embed anatomy-regularized representation learning into CT–MRI translation. Although effective within specific domains, UDA approaches are frequently computationally demanding, sensitive to optimization instability, and typically tailored to particular organ–modality pairs.

Transfer learning has been investigated as a more direct strategy for improving data efficiency. Kieselmann et al. [7] demonstrated cross-modality training using CT-derived labels and synthetic MRI data, while Li et al. [8] proposed a two-stream mutual knowledge distillation framework for cross-modality segmentation. These approaches reported measurable improvements but were generally limited to same-organ adaptation or closely related anatomical structures.

Cross-organ and cross-modality transfer learning remains comparatively under-explored. Lee and Nishikawa [10] introduced a cross-organ, cross-modality transfer learning (XTL) pipeline leveraging intermediate fine-tuning before adapting pretrained models to new anatomical targets. While improvements were observed, the approach highlighted the difficulty of transferring representations across large anatomical and modality discrepancies.

Taken together, existing research demonstrates substantial architectural advances in segmentation performance and several promising strategies for mitigating annotation scarcity. However, most transfer learning approaches rely on single-encoder fine-tuning or indirect multi-stage adaptation, which may be insufficient when source and target domains differ simultaneously in both anatomy and imaging modality. The challenge of effectively reusing pretrained representations under such conditions remains open. Motivated by this gap, the present study investigates an architectural framework that explicitly separates and fuses transferable and domain-specific features for cross-organ and cross-modality liver tumor segmentation.

3 Materials and Method

This section describes the experimental design and methodology used to study cross-organ and cross-modality transfer learning for liver tumor segmentation. The goal of this work is not to present a state-of-the-art liver tumor segmentation benchmark, but rather to analyze the limits of transferring pretrained representations from a brain tumor MRI segmentation model to a liver tumor CT task, and to evaluate an architectural mitigation strategy (ADEST-U-Net) designed to improve feature reuse under a large organ and modality shift.

Importantly, the source domain model (brain MRI) was not trained from scratch in this study. Instead, publicly available pretrained UNet++ weights were used as the only source of transferable knowledge. This setup reflects a realistic scenario in which pretrained models may exist for unrelated organs/modalities, while no pretrained model is available for the target setting.

3.1 Dataset Description

All experiments were conducted using the publicly available 3D-IRCADb-01 dataset, which is a benchmark for liver and tumor segmentation from contrast-enhanced abdominal CT scans. It contains 20 patients (10 male and 10 female), with hepatic tumors present in approximately 75% of cases. The dataset provides DICOM CT volumes and expert annotations in mask and mesh formats. Tumor annotations follow Couinaud's liver segmentation system, where tumors may appear in multiple anatomical segments.

3.1.1 3D-IRCADb-01 Processing and Slice Extraction

Each 3D CT volume was decomposed into axial 2D slices to enable controlled experiments using 2D convolutional segmentation models and to increase the effective number of supervised samples. Patients without tumor annotations were excluded. Let the CT volume and tumor mask for patient k be:

$$X^{(k)} \in \mathbb{R}^{512\times512\times D}, \quad Y^{(k)} \in \{0,1\}^{512\times512\times D}.$$

A patient was retained only if:

$$T(k) = \sum_{d=1}^{D}\sum_{i=1}^{512}\sum_{j=1}^{512} Y^{(k)}[i,j,d] > 0.$$

Each retained volume was decomposed into paired slices:

$$x_d^{(k)} = X^{(k)}[:,:,d], \quad y_d^{(k)} = Y^{(k)}[:,:,d].$$

Mask aggregation across liver segments.

Since tumor masks may be provided per Couinaud segment, all region-specific tumor masks were merged into a single binary tumor mask per slice using a pixel-wise logical OR:

$$y_d^{(k)}[i,j] = \bigvee_{s=1}^{S_d} y_d^{(k,s)}[i,j].$$

Tumor-slice filtering (training only).

The 3D-IRCADb-01 dataset comprised a total of 2,063 axial image–mask pairs, of which 1,495 slices contained no visible tumor regions (i.e., entirely black masks) and only 568 slices included annotated tumor areas. This strong imbalance reflects the inherent sparsity of tumor appearances in volumetric liver CT scans.

To mitigate class imbalance during optimization, model training was performed using tumor-containing slices only, consistent with prior observations that dominant non-tumor slices can bias learning toward background prediction and degrade tumor

sensitivity [4]. The tumor-slice training set is defined as:

$$\mathcal{D}_{\text{IRCAD}} = \left\{ \left(x_d^{(k)}, y_d^{(k)} \right) \;\middle|\; \sum_{i=1}^{512} \sum_{j=1}^{512} y_d^{(k)}[i,j] > 0 \right\}.$$

To assess false-positive behavior under more realistic conditions, 10 non-tumor slices were added to both the validation and test sets (Section 3.4).

3.2 Preprocessing and Data Augmentation

All CT slices were used at a fixed input resolution of 512×512 pixels and formatted as single-channel inputs $(512, 512, 1)$.

3.2.1 Intensity Normalization and Mask Binarization

CT intensities were normalized using min–max scaling:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

Tumor masks were binarized:

$$\hat{y}[i,j] = \begin{cases} 1, & y[i,j] > 0, \\ 0, & \text{otherwise.} \end{cases}$$

3.3 Data Augmentation

3.3.1 Data Augmentation (Training Only)

Given the limited number of tumor-containing slices in the 3D-IRCADb-01 dataset and the inherent sparsity of tumor regions in volumetric liver CT scans, data augmentation was applied exclusively to the training split to expand the effective dataset size and improve robustness under limited labeled data.

Three transformation strategies were implemented:

Elastic deformation.

To simulate non-rigid anatomical variability commonly observed in abdominal imaging, elastic deformation was applied following the formulation introduced by Ronneberger et al. [2]. Random displacement fields were generated using Gaussian-smoothed noise with fixed parameters $\alpha = 50$ and $\sigma = 6$:

$$\delta x = \alpha \cdot G_\sigma * R_x, \quad \delta y = \alpha \cdot G_\sigma * R_y,$$

where R_x and R_y denote random fields sampled from a uniform distribution and G_σ represents a Gaussian kernel with standard deviation σ.

Affine translation.

To model minor patient positioning variations and scanner-induced shifts, random affine translations were applied along both spatial axes. Each image–mask pair was translated according to:

$$(i', j') = (i + \Delta i,\ j + \Delta j), \quad \Delta i, \Delta j \in [-10, 10],$$

while preserving image scale and orientation, consistent with prior medical imaging augmentation strategies [10].

Horizontal flipping.

Horizontal flipping was applied to both images and corresponding masks to increase spatial diversity. Although abdominal anatomy is not perfectly symmetric, left–right mirroring preserves tumor morphology and relative intrahepatic positioning while expanding spatial variability, as commonly adopted in segmentation studies [22].

All augmentation operations were applied identically to the image and mask to maintain spatial correspondence.

3.4 Dataset Split

Tumor-containing slices were split into training/validation/test using an 80%/10%/10% ratio. For evaluation of false-positive behavior, 10 non-tumor slices were added to both validation and test sets.

3.5 Evaluation Metrics

Performance was evaluated using Dice Similarity Coefficient (DSC), Intersection over Union (IoU), precision, and recall.

3.5.1 Dice Similarity Coefficient (DSC)

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|}.$$

3.5.2 Intersection over Union (IoU)

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}.$$

3.5.3 Precision and Recall

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}.$$

3.6 Statistical Analysis

To ensure a comprehensive evaluation of segmentation performance, we employed a multi-metric assessment strategy combining overlap-based, detection-based, and boundary-based metrics.

Dice Similarity Coefficient (DSC).

Per-slice Dice was computed to quantify volumetric overlap between predicted masks and ground truth annotations. Given the known instability of Dice in small-lesion scenarios, we report the mean ± standard deviation, median, and interquartile range (IQR).

To compare the two models, we performed a paired bootstrap analysis (10,000 resamples) on the per-slice Dice differences (ADEST − Baseline). The bootstrap provided 95% confidence intervals and two-sided empirical p-values. This non-parametric approach was selected to avoid normality assumptions and to better reflect the skewed Dice distribution observed in small lesion segmentation.

Slice-wise Detection Sensitivity.

Because Dice collapses to zero when lesions are completely missed, we additionally evaluated slice-level detection performance. A slice was considered correctly detected if the predicted mask contained at least one positive pixel on a tumor-positive ground-truth slice. Detection rate was computed as:

$$\text{Detection Rate} = \frac{N_{\text{detected slices}}}{N_{\text{tumor-positive slices}}} \times 100$$

This metric complements Dice by evaluating sensitivity independent of segmentation overlap quality.

Boundary-based Metrics.

To assess contour agreement and mask smoothness, we computed the 95th percentile Hausdorff Distance (HD95) and the Average Symmetric Surface Distance (ASSD). Both metrics were calculated per slice using symmetric boundary distances derived from Euclidean distance transforms. Boundary metrics are particularly important in medical image segmentation as they reflect contour smoothness, fragmentation, and the presence of spurious small components.

3.7 Model Selection and Baseline Architecture

Several widely used segmentation models were considered, including U-Net [2], 3D U-Net [23], V-Net [24], UNet++ [3], DeepLab [25], HRNet [26], Mask R-CNN [27], and nnU-Net [28]. UNet++ [3] was selected as the baseline because its nested skip connections reduce the semantic gap between encoder and decoder features while remaining computationally practical for controlled ablation studies in data-scarce settings.

3.8 Method Overview and Experimental Tracks

The study is organized into three experimental tracks: (1) supervised training from scratch (baseline), (2) transfer learning using an MRI-pretrained UNet++ with progressive fine-tuning (with/without input modality adaptation), and (3) the proposed ADEST-U-Net dual-encoder architecture.

Table 1: Progressive freezing strategies applied to the UNet++ architecture. Block indices follow the UNet++ hierarchical notation.

Model	Frozen Blocks
Model 1	Encoder block [00]
Model 2	Encoder blocks [00], [10]
Model 3	Encoder blocks [00], [10], [20]
Model 4	Encoder blocks [00], [10], [20], [30]
Model 5	Encoder blocks [00], [10], [20], [30] and bottleneck [40]
Model 6	Encoder blocks [00], [10], [20], [30], bottleneck [40], and nested decoder blocks [01], [11], [21], [31]
Model 7	Encoder blocks [00], [10], [20], [30], bottleneck [40], and nested decoder blocks [01], [11], [21], [31], [02], [12], [22]
Model 8	Encoder blocks [00], [10], [20], [30], bottleneck [40], and nested decoder blocks [01], [11], [21], [31], [02], [12], [22], [03], [13]
Model 9	Encoder blocks [00], [10] and dense block [01]
Model 10	Encoder blocks [00], [10], dense blocks [01], [11], [02]
Model 11	Encoder blocks [00], [10], [20], [30], bottleneck [40], and dense blocks [01], [11], [21], [02], [12], [03]

3.8.1 Track 1: Baseline Supervised Learning from Scratch

A UNet++ model was trained from scratch on 3D-IRCADb-01 tumor-containing slices using batch size 4, Adam optimizer, and learning rate 1×10^{-4}. Training was performed for up to 50 epochs with early stopping (patience = 20).

3.8.2 Track 2: Cross-Organ and Cross-Modality Transfer Learning

A UNet++ model pretrained on brain tumor MRI segmentation was adapted to liver CT under two strategies.

Track 2.1: Progressive fine-tuning.

Eleven freezing configurations were evaluated, ranging from freezing early encoder blocks only to freezing the full encoder, bottleneck, and nested decoder blocks (Table 1). All configurations used the same training schedule as Track 1 to ensure comparability.

Track 2.2: Progressive fine-tuning with input modality adaptation.

To reduce MRI–CT mismatch, an Input Modality Adapter (IMA) was introduced:

$$\tilde{x} = \mathrm{ReLU}\left(\mathrm{BN}(W_{1\times1} * x_{\mathrm{CT}})\right).$$

The same 11 freezing configurations were reused to isolate the effect of input-level adaptation.

3.8.3 Track 3: Proposed ADEST-U-Net Architecture

ADEST-U-Net is designed to mitigate the limitations observed in Track 2 by explicitly separating and fusing pretrained MRI-derived features and CT-specific features, as illustrated in Figure 1. The architecture consists of: (i) Encoder A initialized from the

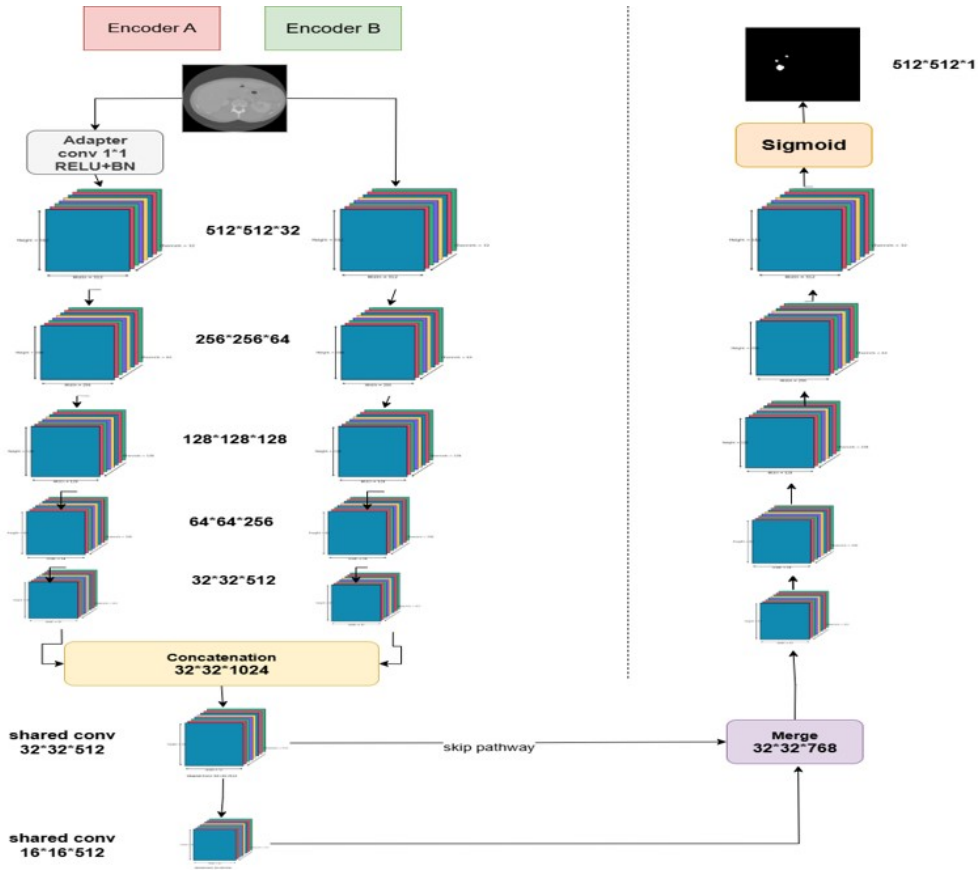


Fig. 1: Proposed architecture of ADEST-U-Net. The model employs two parallel encoders: an MRI-pretrained encoder with modality adaptation and a CT-specific encoder trained from scratch. Bottleneck features from both encoders are concatenated and refined through a shared-tail module before U-Net–style decoding.

MRI-pretrained UNet++ encoder (with input modality adaptation), (ii) Encoder B trained from scratch on CT data, and (iii) a shared-tail refinement module applied to the concatenated bottleneck features prior to decoding.

Importantly, only the weights of the hierarchical encoder blocks were transferred from the MRI-pretrained UNet++ model. The nested dense skip pathways characteristic of UNet++ were not reused. This design choice was motivated by architectural constraints: as shown in Figure 1, directly transferring nested decoder connections would introduce structural incompatibilities in the dual-encoder configuration, particularly during feature fusion at the bottleneck. Restricting transfer to encoder blocks preserves hierarchical feature representations while allowing flexible fusion and refinement in the shared-tail and decoding stages.

Table 2: Architectural specifications of the proposed ADEST-U-Net model.

Component	Specification
Convolution kernel	3×3, stride = 1, padding = 1
Activation	ReLU + Batch Normalization
Downsampling	2×2 max pooling (stride = 2)
Upsampling	Transposed convolution (stride = 2)
Feature maps per stage	$32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$
Bottleneck fusion	Channel-wise concatenation (512+512 = 1024 channels)
Shared-tail refinement	$1024 \rightarrow 512$ channels (Conv + BN + ReLU + Dropout $p = 0.3$)
Dropout	$p = 0.3$
Output layer	1×1 convolution + sigmoid activation

At the bottleneck, feature maps from both encoders are concatenated channel-wise (Figure 1). Given bottleneck feature tensors $F_A \in \mathbb{R}^{H \times W \times 512}$ and $F_B \in \mathbb{R}^{H \times W \times 512}$, the fused representation is:

$$F_{\text{fusion}} = \text{Concat}(F_A, F_B) \in \mathbb{R}^{H \times W \times 1024}.$$

This fused representation is then refined through a shared-tail module (see Figure 1) that reduces the channel dimensionality from 1024 to 512 using convolution, batch normalization, ReLU activation, and dropout ($p = 0.3$), before being passed to the decoder.

Warm-up phase.

For 5 epochs, Encoder A was frozen while Encoder B, the shared-tail module, and the decoder were trained using Dice loss only. The learning rate was set to 1×10^{-4}.

Fine-tuning phase.

All layers were unfrozen and trained end-to-end for up to 80 epochs using a learning rate of 1×10^{-5}, it is adjusted using a ReduceLROnPlateau scheduler, dropout probability $p = 0.3$, and early stopping with patience set to 15 epochs.

The detailed architectural configuration of ADEST-U-Net is summarized in Table 2.

3.9 Loss Function

All models were trained using a composite loss combining Dice loss and shape prior loss [29].

Dice loss.

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2 \sum_{i=1}^{HW} y_i \hat{y}_i}{\sum_{i=1}^{HW} y_i + \sum_{i=1}^{HW} \hat{y}_i}.$$

Shape prior construction.

The shape prior was computed once from the training masks only:

$$\text{prior}[i,j] = \frac{1}{N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} y^{(n)}[i,j].$$

Shape prior visualization.

Figure 2 shows the resulting prior heatmap computed from the aggregated training masks of the 3D-IRCADb-01 dataset. The mild extension toward the right side of the heatmap reflects an artifact introduced by horizontal flipping during augmentation. This behavior was examined in an ablation variant that excluded horizontal flipping to assess its impact on anatomical fidelity and segmentation performance.

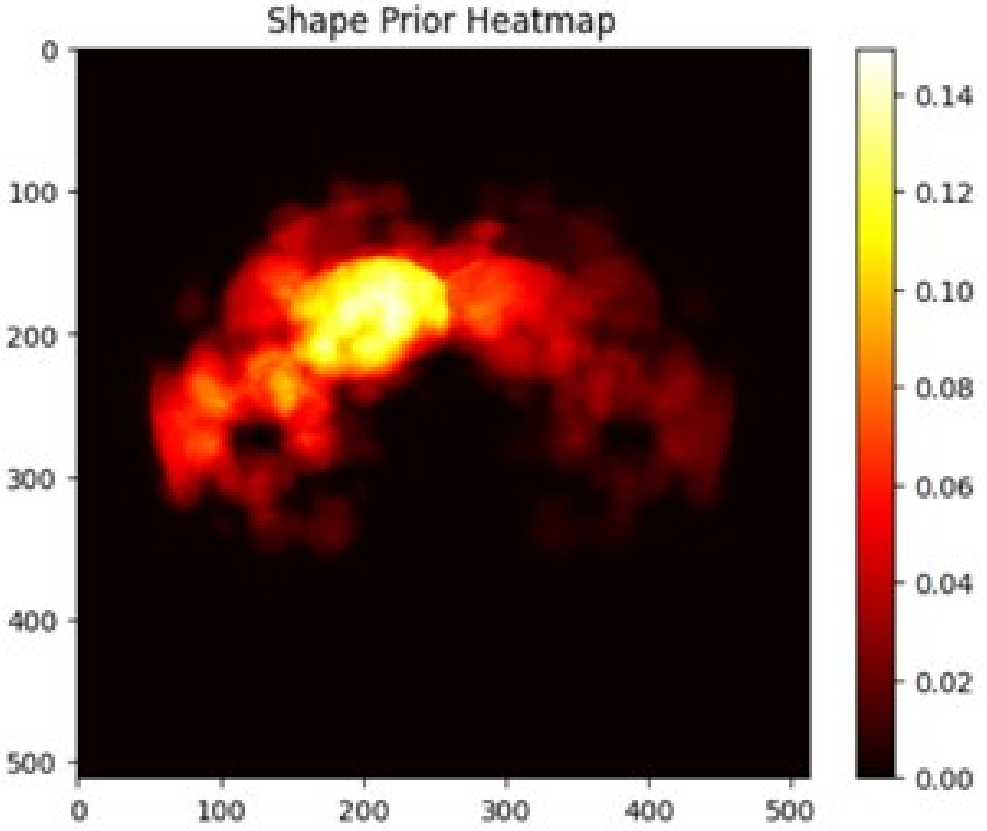


Fig. 2: Shape prior heatmap constructed from the training masks of the 3D-IRCADb-01 dataset. Pixel intensities represent the empirical probability of tumor occurrence across spatial locations, with warmer colors indicating higher likelihood.

Shape prior loss.

$$\mathcal{L}_{\text{prior}} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\hat{y}[i,j] - \text{prior}[i,j]\right)^2 .$$

Combined loss with fixed weights.

For Track 1 and Track 2:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{dice}} + 0.1\, \mathcal{L}_{\text{prior}}.$$

For Track 3 (ADEST-U-Net):

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{dice}} + 0.05\, \mathcal{L}_{\text{prior}}.$$

3.10 Hyperparameter Optimization

Hyperparameters were optimized on 3D-IRCADb-01 using Particle Swarm Optimization (PSO) [30]. PSO used a swarm size of 8 particles for 8 iterations (64 evaluations). The search space included input resolution $\{128, 256, 512\}$, optimizer $\{\text{Adam}, \text{RMSProp}, \text{SGD}\}$, loss function variants, and learning rate in $[10^{-5}, 10^{-2}]$. The selected configuration was: resolution 512×512, Dice loss, Adam optimizer, and learning rate 1×10^{-4}.

3.11 Implementation Environment

All experiments were conducted on a workstation equipped with an NVIDIA RTX 6000 Ada GPU (48GB VRAM), 16 vCPUs, and 188GB RAM, using TensorFlow 1.0.3 with mixed precision enabled. Batch size was fixed to 4. Track 1 and Track 2 trained for up to 50 epochs; ADEST-U-Net trained for up to 80 epochs due to staged optimization.

4 Results and Discussion

All models in this section were trained on the 3D-IRCADb-01 training split and evaluated on the 3D-IRCADb-01 validation and test splits using Dice and IoU. The goal is to analyze cross-organ, cross-modality transfer behavior and evaluate architectural mitigation strategies under controlled settings.

4.1 Track 1: Baseline – Supervised Learning with UNet++ Trained from Scratch

This track establishes a supervised reference by training UNet++ from scratch using only labeled liver CT slices from 3D-IRCADb-01. No pretrained weights were used.

The baseline UNet++ achieved a Dice score of 0.5879 and an IoU of 0.4163 on the validation set, and a Dice score of 0.7020 with an IoU of 0.5408 on the test set.

Figure 3 demonstrates that the baseline model detects a substantial portion of tumor regions, but boundary delineation is often coarse and over-segmentation occurs in non-tumor areas. These characteristics provide a quantitative and qualitative reference point for evaluating the impact of transfer learning and architectural modifications.

4.2 Track 2: Cross-Organ Transfer Learning (Brain MRI → Liver CT) Using UNet++

This track investigates whether features learned from brain tumor MRI segmentation can be transferred to liver tumor CT segmentation. A UNet++ model initialized from MRI-pretrained weights was fine-tuned on 3D-IRCADb-01 using (i) progressive fine-tuning and (ii) progressive fine-tuning with an Input Modality Adapter (IMA).

4.2.1 Progressive Fine-Tuning without Modality Adaptation

Table 3 reports performance across the eleven freezing configurations. Model 1 achieved the highest test performance (Dice = 0.3407, IoU = 0.2053), while deeper freezing strategies led to markedly lower Dice scores, indicating limited transferability under strong organ+modality shift.

Table 3: Progressive fine-tuning UNet++ results on 3D-IRCADb-01.

Model	Test Dice	Test IoU	Val Dice	Val IoU
Model 1	**0.3407**	**0.2053**	**0.2512**	**0.1437**
Model 2	0.1706	0.0932	0.1512	0.0818
Model 3	0.1142	0.0606	0.0987	0.0519
Model 4	0.1055	0.1055	0.0926	0.0485
Model 5	0.0071	0.0035	0.0078	0.0039
Model 6	0.0062	0.0031	0.0067	0.0034
Model 7	0.0039	0.0020	0.0052	0.0026
Model 8	0.0025	0.0013	0.0028	0.0014
Model 9	0.0108	0.0054	0.0156	0.0078
Model 10	0.0139	0.0070	0.0203	0.0103
Model 11	0.0083	0.0042	0.0123	0.0062

4.2.2 Progressive Fine-Tuning with Input Modality Adaptation (IMA)

Adding the IMA module improved transfer results. As shown in Table 4, Model 1 achieved the best performance (test Dice = 0.5383, IoU = 0.3683; validation Dice = 0.4794, IoU = 0.3153). Performance again dropped with heavier freezing, indicating that input alignment helps but does not eliminate the need for sufficient trainable capacity.

Table 4: Progressive fine-tuning with IMA results on 3D-IRCADb-01.

Model	Test Dice	Test IoU	Val Dice	Val IoU
Model 1	**0.5383**	**0.3683**	**0.4794**	**0.3153**
Model 2	0.2509	0.1435	0.2231	0.1256
Model 3	0.2182	0.1225	0.1880	0.1037
Model 4	0.1882	0.1039	0.2157	0.1209
Model 5	0.1309	0.0700	0.1093	0.0578
Model 6	0.1119	0.0592	0.1034	0.0545
Model 7	0.1203	0.0640	0.1085	0.0573
Model 8	0.2134	0.1195	0.1525	0.0825
Model 9	0.2266	0.1278	0.1959	0.1086
Model 10	0.1858	0.1024	0.1610	0.0876
Model 11	0.1223	0.0651	0.1311	0.0702

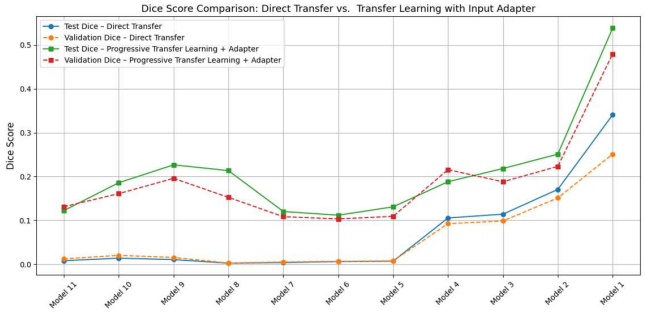


Fig. 4: Comparison of Dice scores for direct transfer versus progressive fine-tuning with the Input Modality Adapter across freezing strategies (Models 1–11) on validation and test sets.

4.3 Track 3: ADEST-U-Net — Dual-Encoder Architecture for Modality-Aware Transfer

This track evaluates the proposed ADEST-U-Net dual-encoder design, which separates CT-specific representation learning from MRI-pretrained feature reuse and fuses both streams at the bottleneck through a shared-tail refinement module.

4.3.1 Evaluation (3D-IRCADb-01)

ADEST-U-Net achieved a Dice score of 0.7189 with an IoU of 0.5611 on the 3D-IRCADb-01 test set. These results indicate that architectural mitigation enables effective reuse of pretrained MRI representations without harming CT-specific learning.

Figure 5 highlights improved boundary continuity and reduced fragmentation relative to direct cross-modality transfer (Track 2). In addition, predicted tumor masks

Table 5: In-domain segmentation results on 3D-IRCADb-01.

Model	Val IoU	Val Dice	Test IoU	Test Dice
Baseline UNet++	0.4163	0.5879	0.5408	0.7020
UNet++ + IMA	0.3153	0.4794	0.3683	0.5383
ADEST-U-Net	**0.5451**	**0.7056**	**0.5611**	**0.7189**

remain confined within liver boundaries, with no spurious activations observed outside the organ region. This behavior is consistent with the intended role of the dual-encoder fusion mechanism in promoting anatomically coherent and modality-aware feature integration.

4.4 Overall Comparison on 3D-IRCADb-01 (In-Domain)

Table 5 summarizes performance for the baseline UNet++ (Track 1), the best-performing fine-tuned UNet++ with Input Modality Adaptation (Track 2, Model 1), and the proposed ADEST-U-Net (Track 3). All models were trained and evaluated on the same validation and test splits of 3D-IRCADb-01.

Two observations are central to the study objective. First, naïve cross-organ transfer (even with input modality adaptation) underperforms supervised training from scratch in this setting. Second, architectural mitigation through ADEST-U-Net restores and slightly improves performance relative to the supervised baseline, indicating controlled feature reuse without negative transfer.

4.5 Contextual Comparison with Prior 3D-IRCADb-01 Methods

To contextualize the reported results within the broader literature, Table 6 summarizes representative liver and liver tumor segmentation Dice scores reported on the 3D-IRCADb-01 dataset.

Across studies, liver segmentation consistently achieves higher Dice scores than tumor segmentation, with several recent architectures exceeding 0.94 for liver segmentation and approaching 0.80 for tumor segmentation. In contrast, tumor Dice values typically range between approximately 0.69 and 0.80, reflecting the greater complexity and variability associated with delineating tumor regions compared to whole-liver structures.

Most existing approaches evaluated on 3D-IRCADb-01 focus primarily on improving segmentation performance through architectural enhancements, including attention mechanisms, hybrid CNN–Transformer designs, and multi-scale feature aggregation strategies. These methods are generally optimized to maximize segmentation accuracy within a single organ and imaging modality.

In contrast, the present study addresses a fundamentally different research question: how pretrained representations behave under extreme cross-organ and cross-modality transfer, and how architectural mitigation strategies can reduce negative transfer. The proposed ADEST-U-Net achieves a tumor Dice score of 0.7189 on

Table 6: Representative liver and liver tumor segmentation Dice scores reported on the 3D-IRCADb-01 dataset.

Model	Liver Dice	Tumor Dice
HANS-Net (2025)	0.9482	0.8008
ResUNet (2025)	0.9144	0.7584
LCAB-SwinUNet (2025)	—	0.7138
SWTR-UNet (2023)	—	0.7900
DPC-Net (2022)	—	0.7640
ELANRes-MSCA-UNet (2025)	—	0.7290
Two-stage 2D FCN (2018)	—	0.6900

3D-IRCADb-01, which falls within the performance range reported by conventional CNN-based architectures.

Accordingly, the purpose of this comparison is to contextualize performance within the established range of the dataset rather than to claim state-of-the-art segmentation accuracy.

4.6 Statistical Analysis

4.7 Quantitative Evaluation on the Test Set

A total of 57 tumor-positive slices were evaluated. Performance was assessed using overlap-based, detection-based, and boundary-based metrics to provide a comprehensive comparison between the baseline UNet++ and ADEST-U-Net models.

4.7.1 Overlap Performance (Dice)

The baseline UNet++ achieved a median Dice of 0.5666 (IQR 0.0000–0.7659) with a mean of 0.4482 ± 0.3444. ADEST-U-Net obtained a median Dice of 0.4610 (IQR 0.0000–0.7965) and a mean of 0.4201 ± 0.3854.

To assess statistical significance, a paired bootstrap analysis (10,000 resamples) was conducted on the per-slice Dice differences (ADEST − Baseline). The observed mean difference was -0.0280, with a 95% confidence interval of $[-0.0825, 0.0210]$ and a bootstrap p-value of 0.273. As the confidence interval includes zero, no statistically significant difference in Dice was observed between the two models. The large variance across slices reflects the instability of Dice in small-lesion scenarios and the impact of complete misses.

4.7.2 Detection Sensitivity and Boundary Metrics

While Dice provides overlap information, additional metrics revealed distinct behavioral differences between the two architectures. Slice-wise detection sensitivity showed that the baseline model detected tumors in 49 out of 57 slices (85.96%), whereas ADEST-U-Net detected tumors in 34 out of 57 slices (59.65%). The lower detection rate of ADEST indicates a higher frequency of complete misses, particularly in small-lesion cases.

Boundary-based metrics demonstrated an opposite trend. ADEST achieved substantially lower median boundary distances, indicating improved contour agreement and reduced fragmentation. Specifically, the median HD95 decreased from 220.57 pixels (baseline) to 78.41 pixels (ADEST), while the median ASSD decreased from 60.56 pixels to 24.02 pixels.

Table 7 summarizes the quantitative comparison.

Table 7: Test set results on 57 tumor-positive slices. Lower values indicate better performance for HD95 and ASSD.

Metric	Baseline	ADEST
Dice (mean ± std)	0.4482 ± 0.3444	0.4201 ± 0.3854
Dice (median)	0.5666	0.4610
Detection (%)	85.96	59.65
HD95 (median)	220.57	78.41
ASSD (median)	60.56	24.02

4.8 Discussion

4.8.1 Interpretation of Cross-Organ, Cross-Modality Transfer

The statistical analysis presented in Section 4 confirms that direct transfer from brain MRI to liver CT remains fundamentally constrained. Bootstrap analysis of per-slice Dice differences revealed no statistically significant improvement between ADEST-U-Net and the supervised baseline (95% CI: $[-0.0825, 0.0210]$, $p = 0.273$). This indicates that under cross-organ and cross-modality transfer, architectural mitigation alone does not guarantee superiority over fully supervised training.

However, the multi-metric evaluation reveals that Dice alone does not fully capture segmentation behavior. While the baseline model achieved slightly higher mean Dice and substantially higher slice-wise detection sensitivity (85.96% vs. 59.65%), boundary-based metrics demonstrated markedly improved contour agreement for ADEST-U-Net. The median HD95 and ASSD were significantly lower for ADEST, indicating smoother and more spatially coherent predictions.

This divergence between overlap-based and boundary-based metrics suggests that cross-organ transfer introduces a trade-off between detection sensitivity and structural regularity.

4.8.2 Architectural Behavior and Error Patterns

The quantitative error breakdown in Table 8 further clarifies this behavior. The scratch baseline achieved the highest recall (0.8712 on the test set) but produced 119,632 false positives, reflecting substantial over-segmentation. In contrast, ADEST-U-Net reduced false positives to 38,726, a 67.6% reduction relative to the baseline , while maintaining a recall of 0.6720 and achieving the highest precision (0.7727).

These findings align with the boundary metrics: the baseline prioritizes sensitivity, often generating fragmented or scattered tumor predictions, which inflate HD95 and ASSD values. ADEST-U-Net, by contrast, suppresses spurious activations and produces anatomically plausible segmentations confined within liver boundaries. Although this conservative behavior reduces detection sensitivity for small lesions, it significantly improves structural consistency and precision.

Thus, the proposed dual-encoder architecture does not merely aim to maximize Dice, but rather to mitigate negative transfer by separating modality-specific representation learning from pretrained feature reuse. The shared-tail fusion mechanism provides controlled feature interaction, preventing incompatible MRI features from dominating CT adaptation.

4.8.3 Implications for Cross-Domain Segmentation

The absence of statistical significance in Dice, combined with improved boundary regularity and reduced false positives, highlights the limitations of relying solely on overlap-based metrics in small-lesion segmentation tasks. Dice is highly sensitive to complete misses and small object variability, leading to inflated variance across slices. In contrast, boundary metrics and precision–recall analysis reveal meaningful architectural differences that Dice alone obscures.

These results suggest that successful cross-organ transfer may not initially manifest as large Dice improvements, but rather as improved structural plausibility and error suppression. Such improvements may be particularly relevant in clinical settings where false positives can increase radiologist burden and reduce trust in automated systems.

4.8.4 Limitations and Future Directions

Several limitations should be acknowledged.

First, evaluation was conducted on a single public dataset (3D-IRCADb-01), which limits the assessment of robustness under broader acquisition variability. Although the dataset provides well-annotated scans, it does not fully represent real-world heterogeneity in scanner characteristics, reconstruction parameters, contrast phases, and institutional imaging protocols.

Second, performance analysis was primarily slice-based. While Dice, detection rate, and boundary metrics provide complementary insights, lesion-level sensitivity and volumetric consistency were not explicitly quantified. Given the clinical importance of small and subtle tumors, a dedicated lesion-level evaluation framework would provide deeper insight into detection reliability.

Third, although ADEST-U-Net mitigates negative transfer through dual-encoder feature separation and shared-tail fusion, the current fusion strategy may introduce an information bottleneck. Specifically, concatenation of both encoder outputs produces 1024 feature maps at the bottleneck level, which are subsequently reduced to 512 channels through shared convolutional layers. This dimensionality compression may inadvertently discard complementary modality-specific information before full cross-domain integration occurs. The observed reduction in small-lesion sensitivity suggests

that preserving a larger portion of the combined feature space, or introducing adaptive feature weighting mechanisms, may further enhance representational richness and tumor detectability.

Future research should therefore explore architectures that maintain higher bottleneck dimensionality, incorporate learnable feature selection or attention-based fusion mechanisms, and dynamically weight modality-specific representations. In addition, multi-center validation across heterogeneous CT datasets, lesion-level performance analysis, and architectural refinements such as multi-scale attention modules or boundary-aware loss functions may further improve robustness and small-tumor sensitivity. Advanced domain adaptation strategies, including feature-level alignment and large-scale self-supervised pretraining, may als o enhance generalization across modality and anatomical variability.

5 Qualitative and Error Analysis

Beyond aggregate Dice and IoU metrics, we conducted a focused qualitative and error-based analysis to better understand segmentation behavior across the three experimental tracks. Representative models were selected from each track: (i) a baseline UNet++ trained from scratch (Track 1), (ii) UNet++ with input modality adaptation and first-layer freezing (Track 2), and (iii) the proposed ADEST-U-Net (Track 3). All analyses were performed exclusively on held-out test samples from the 3D-IRCADb-01 dataset.

5.1 Baseline UNet++ (Track 1)

The scratch baseline exhibited a recall-dominant behavior. Qualitative inspection revealed frequent over-segmentation, with tumor predictions extending beyond liver boundaries and activating in anatomically implausible regions. As illustrated in Figure 6, false positive regions (red) frequently appear outside the liver area, indicating poor spatial constraint learning. Although large tumor masses were generally captured, boundary leakage and extrahepatic false positives were common. This pattern explains the combination of high recall and comparatively reduced precision observed in the quantitative analysis.

5.2 UNet++ with Input Modality Adaptation (Track 2)

In contrast, the Track 2 configuration demonstrated a conservative but unstable segmentation behavior. While spurious activations were reduced relative to the scratch baseline, the model frequently under-segmented tumor regions. As shown in Figure 7, small and low-contrast lesions were often missed, and tumor coverage was incomplete in several cases. Freezing low-level MRI-pretrained filters likely constrained adaptation to CT-specific intensity distributions and liver texture characteristics, resulting in elevated false negatives and a marked reduction in recall.

5.3 ADEST-U-Net (Track 3)

Qualitative inspection of ADEST-U-Net revealed a more balanced segmentation profile. False positives were substantially reduced compared to the scratch baseline, and predicted tumor masks were largely confined within liver boundaries, indicating improved anatomical plausibility. As shown in Figure 8, segmentation contours appeared smoother and less fragmented, consistent with the improved boundary metrics reported earlier. Although small tumors remained challenging, the overall spatial consistency and suppression of spurious activations were markedly improved relative to Tracks 1 and 2.

5.3.1 Quantitative Error Interpretation

Table 8 provides a detailed breakdown of true positives (TP), false positives (FP), false negatives (FN), precision, recall, and F1 score across validation and test sets.

The scratch baseline (Track 1) achieved the highest recall, reflecting strong tumor sensitivity. However, this performance was accompanied by a large number of false positives, consistent with the qualitative observation of boundary leakage and extrahepatic activations. The resulting precision–recall imbalance highlights an over-segmentation bias.

The input-modality-adapted configuration (Track 2) demonstrated the opposite failure mode. Although false positives were reduced, false negatives increased substantially, leading to a pronounced drop in recall. This pattern suggests that partial reuse of MRI-pretrained filters constrained effective adaptation to CT-specific characteristics, reinforcing the risk of negative transfer in naïve cross-organ fine-tuning.

In contrast, ADEST-U-Net (Track 3) achieved a more balanced error distribution. False positives were markedly reduced relative to the scratch baseline, while recall remained substantially higher than in Track 2. The resulting precision–recall trade-off produced the highest F1 score on the test set. Importantly, this quantitative improvement aligns with qualitative observations of smoother contours, reduced fragmentation, and anatomically plausible tumor localization.

Collectively, the qualitative and quantitative analyses demonstrate three distinct behavioral regimes: recall-dominant over-segmentation (Track 1), conservative under-segmentation (Track 2), and structurally balanced mitigation of negative transfer (Track 3). These findings reinforce the central hypothesis that modality-aware architectural separation is critical when transferring pretrained representations across anatomically and radiologically distinct domains.

6 Conclusion

This study investigated cross-organ, cross-modality transfer learning for liver tumor segmentation in CT imaging through a structured three-track experimental framework. By systematically comparing supervised training from scratch, naïve cross-organ transfer with input modality adaptation, and a dual-encoder mitigation strategy, we provide an empirical analysis of how pretrained knowledge behaves when transferred across anatomically and radiologically distinct domains.

Table 8: Segmentation performance comparison on the 3D-IRCADb-01 dataset. Metrics include true positives (TP), false positives (FP), false negatives (FN), precision, recall, and F1 score on validation and test sets.

Model	Set	TP	FP	FN	Prec.	Recall	F1
Baseline UNet++	Val	124,034	147,882	26,017	0.4569	0.8266	0.5887
	Test	170,674	119,632	25,269	0.5870	0.8712	0.7012
UNet++ + IMA	Val	64,432	54,307	85,619	0.5420	0.4292	0.4789
	Test	96,084	64,931	99,859	0.5969	0.4911	0.5386
ADEST-U-Net	Val	106,656	45,624	43,395	0.7005	0.7106	0.7047
	Test	131,678	38,726	64,265	0.7727	0.6720	**0.7189**

The findings demonstrate that direct reuse of MRI-pretrained representations within a single-encoder architecture may introduce negative transfer, particularly when low-level features are not aligned with the target modality. Although input modality adaptation partially alleviates this mismatch, freezing pretrained filters can restrict adaptation and degrade lesion sensitivity.

In contrast, the proposed ADEST-U-Net architecture mitigates negative transfer by explicitly separating MRI-derived and CT-specific representations through a dual-encoder design with shared-tail refinement. This structural decoupling yields a more balanced precision–recall trade-off and substantially reduces false positive activations, resulting in improved anatomical plausibility and segmentation stability.

The contribution of this work lies not in achieving state-of-the-art Dice performance, but in providing architectural insight into when and how cross-organ transfer becomes usable. The results suggest that modality-aware architectural separation is a necessary condition for stable knowledge reuse across heterogeneous medical imaging domains.

Future work should extend this framework through multi-center validation, lesion-level evaluation protocols, and architectural refinements targeting small tumor detectability. Further exploration of adaptive fusion mechanisms and large-scale self-supervised medical pretraining may clarify the upper bounds of cross-organ transfer in clinical segmentation tasks.

Statements and Declarations

Ethics Declaration

This study used a publicly available dataset and did not involve direct patient recruitment, intervention, or identifiable patient information.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data and Code Availability

Data Availability

To ensure transparency and reproducibility, the datasets used in this study are made available where permissible.

The **3D-IRCADb-01** liver CT dataset is publicly accessible through the official IRCAD repository. The processed version used in this work, including extracted 2D slices, predefined training, validation, and test splits, and augmentation outputs—is available via Google Drive: https://drive.google.com/drive/folders/10keDJW1yJ3N22KUXTjkYhdpNmf-dp_ZA

Additional preprocessing steps (e.g., normalization, resizing, and channel formatting) were applied dynamically during data loading and were not stored as separate files.

Code Availability

All source code for preprocessing, training, evaluation, and model development is publicly available at: https://github.com/RihamJeeballah/thesis-segmentation

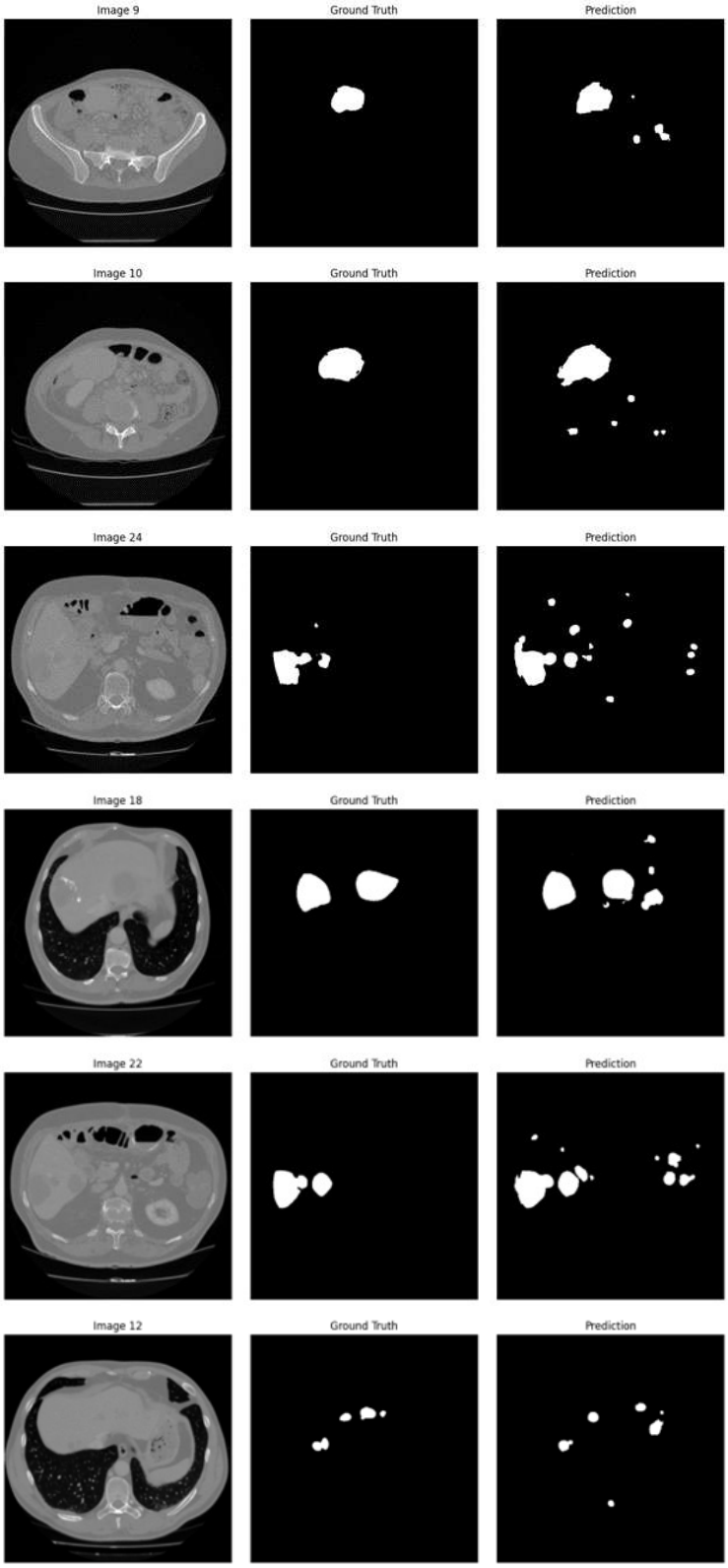


Fig. 3: Examples of baseline UNet++ performance on the 3D-IRCADb-01 test set, showing CT input, expert annotation, and predicted segmentation masks.

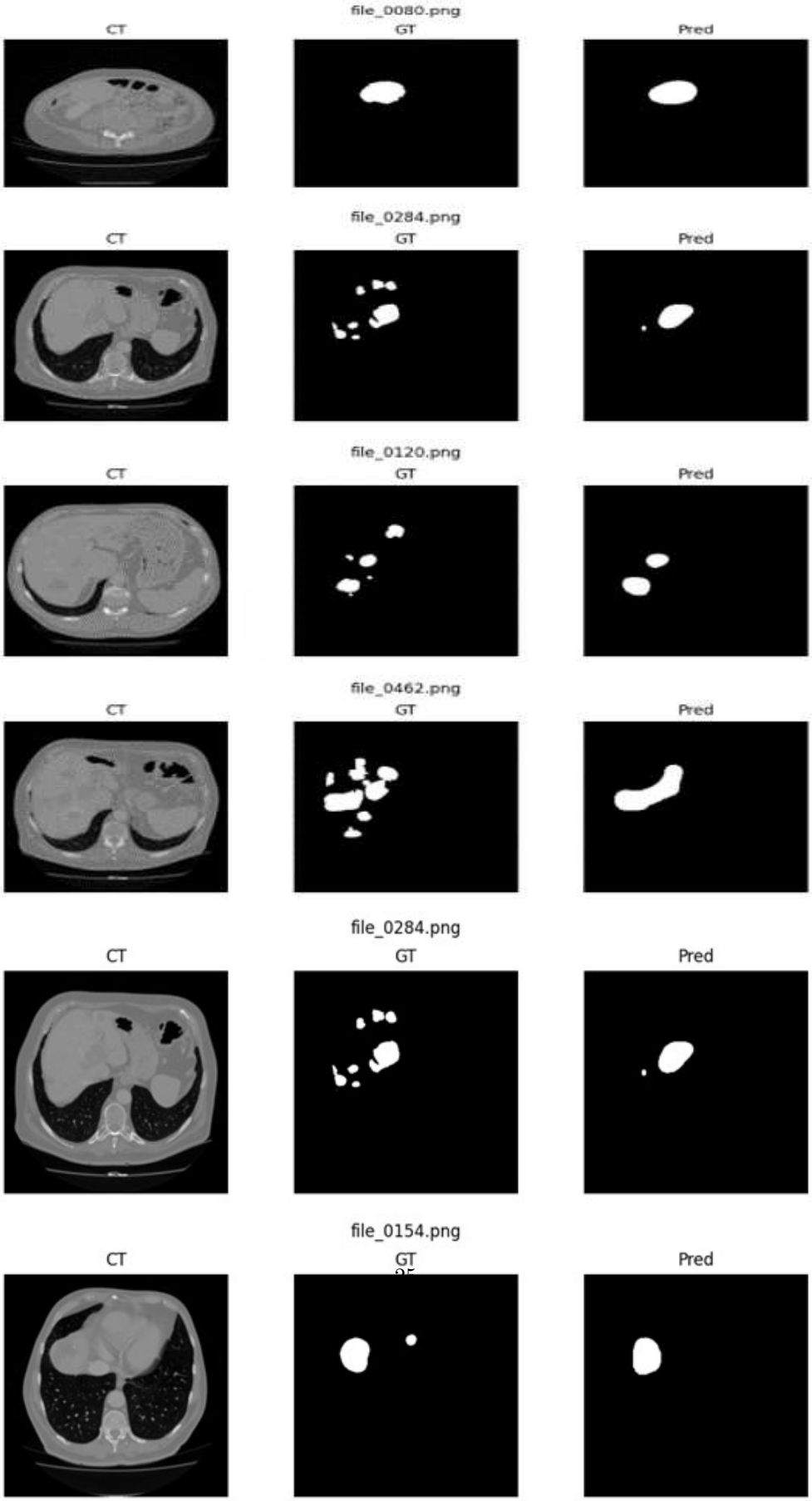


Fig. 5: Evaluation examples for ADEST-U-Net on the 3D-IRCADb-01 test set, showing CT slice, ground-truth mask, and predicted segmentation.

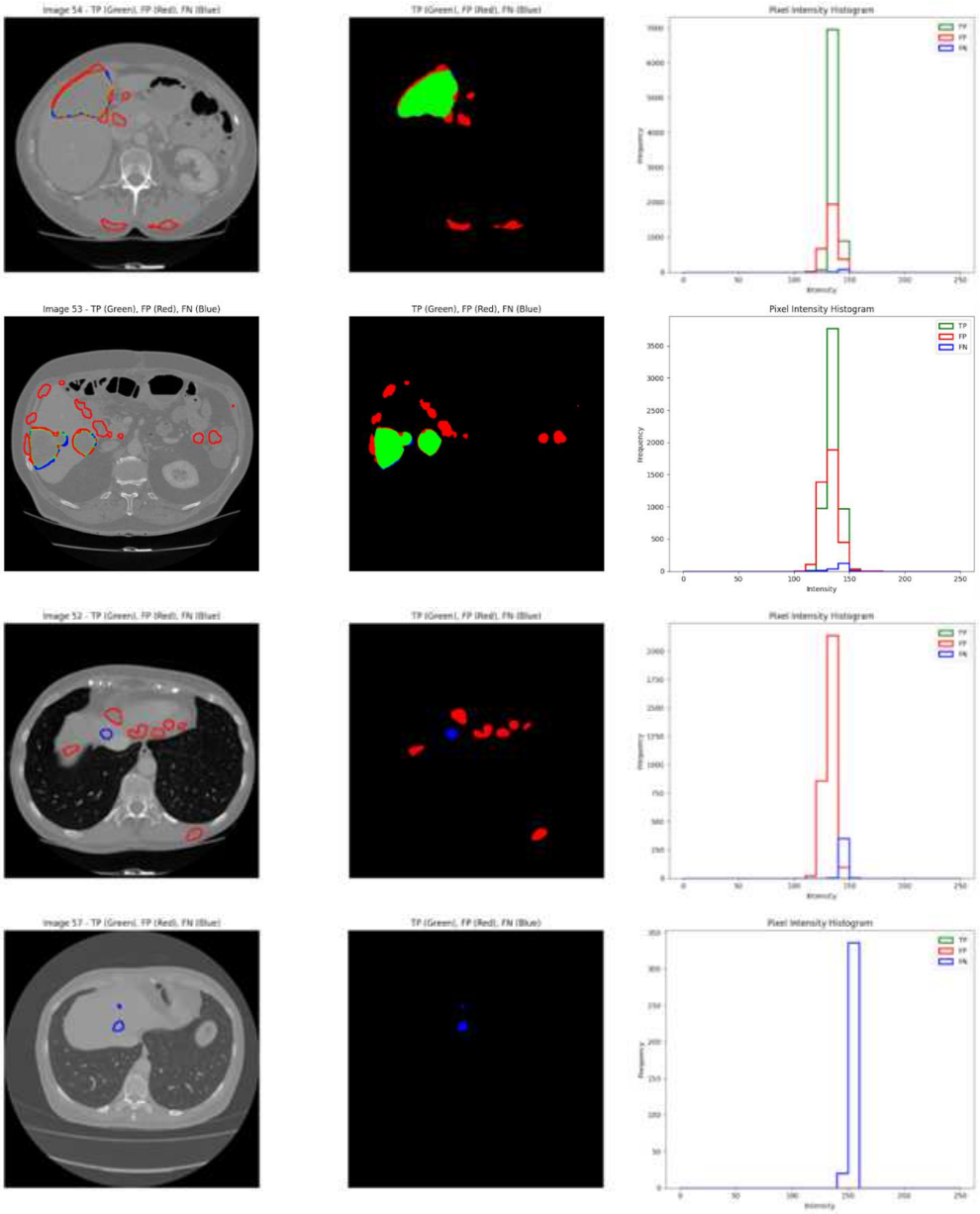


Fig. 6: Qualitative segmentation results for the baseline UNet++ on 3D-IRCADb-01 test samples. True positives are shown in green, false positives in red, and false negatives in blue. The model demonstrates over-segmentation and boundary leakage beyond liver regions.

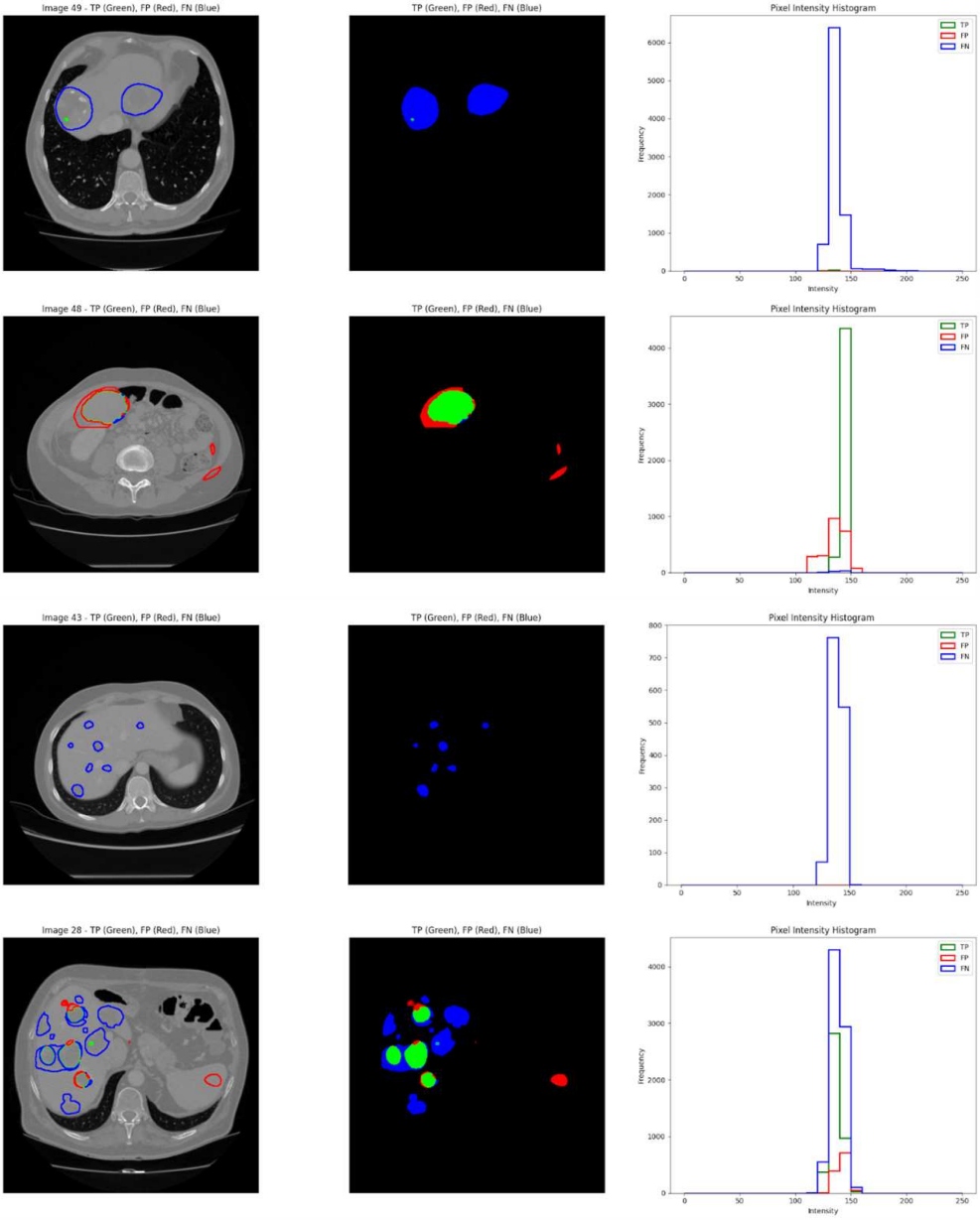


Fig. 7: Qualitative segmentation results for Track 2 (UNet++ with input modality adaptation and first encoder block frozen) on 3D-IRCADb-01 test samples. True positives are shown in green, false positives in red, and false negatives in blue. The model exhibits incomplete tumor coverage and frequent missed small lesions.

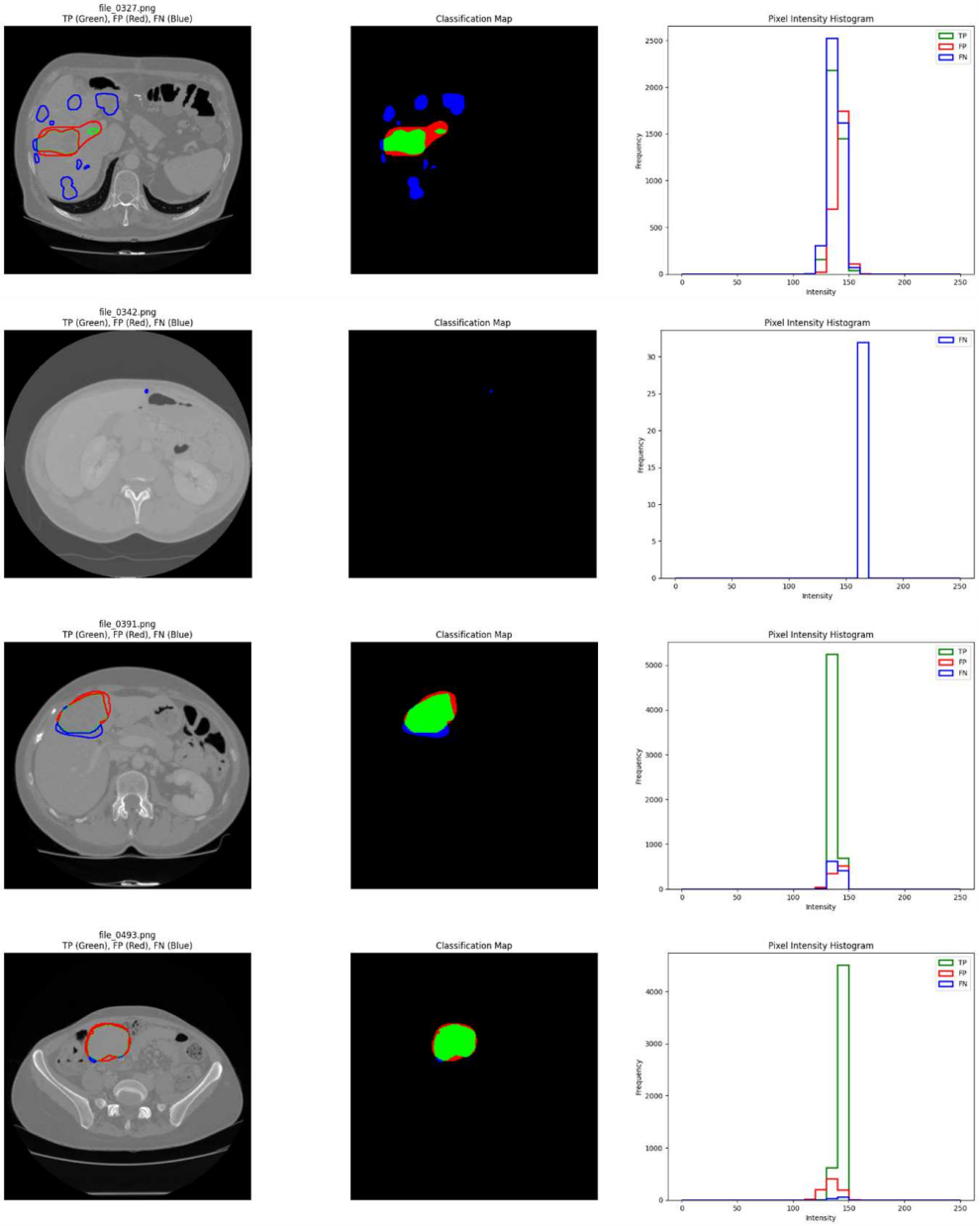


Fig. 8: Qualitative segmentation results for Track 3 using the proposed ADEST-U-Net on 3D-IRCADb-01 test samples. True positives are shown in green, false positives in red, and false negatives in blue. Compared to previous tracks, the model demonstrates reduced false positives, improved anatomical plausibility, and more spatially consistent tumor delineation.

References

[1] Mienye, I.D., Swart, T.G., Obaido, G., Jordan, M., Ilono, P.: Deep Convolutional Neural Networks in Medical Image Analysis: A Review. Multidisciplinary Digital Publishing Institute (MDPI) (2025). https://doi.org/10.3390/info16030195

[2] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)

[3] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation (2018)

[4] Ayalew, Y.A., Fante, K.A., Mohammed, M.A.: Modified u-net for liver cancer segmentation from computed tomography images with a new class balancing method. BMC Biomedical Engineering **3** (2021) https://doi.org/10.1186/s42490-021-00050-y

[5] Karimi, D., Warfield, S.K., Gholipour, A.: Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. Artificial Intelligence in Medicine **116** (2021) https://doi.org/10.1016/j.artmed.2021.102078

[6] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A.J., Al-Shamma, O., Fadhel, M.A., Zhang, J., Santamaría, J., Duan, Y.: Novel transfer learning approach for medical imaging with limited labeled data. Cancers **13** (2021) https://doi.org/10.3390/cancers13071590

[7] Kieselmann, J.P., Fuller, C.D., Gurney-Champion, O.J., Oelfke, U.: Cross-modality deep learning: Contouring of mri data from annotated ct data only. Medical Physics **48**, 1673–1684 (2021) https://doi.org/10.1002/mp.14619

[8] Li, K., Yu, L., Wang, S., Heng, P.-A.: Towards cross-modality medical image segmentation with online mutual knowledge distillation (2020)

[9] Mithunraj, Preethika, Mishra, S., Satapathy, S.K., Mohanty, S.N., Ali, A.B.M., Batool, N., Shomurotova, S.: Optimizing cnn performance in medical imaging with cross-modality pre-training: A study using mobilenetv3. Intelligence-Based Medicine **12** (2025) https://doi.org/10.1016/j.ibmed.2025.100281

[10] Lee, J., Nishikawa, R.M.: Cross-organ, cross-modality transfer learning: Feasibility study for segmentation and classification. IEEE Access **8**, 210194–210205 (2020) https://doi.org/10.1109/ACCESS.2020.3038909

[11] Abian, A.I., Debnath, R.K., Rahman, M.A., Raiaan, M.A.K., Islam, M.R., Karim, A., Mohamed, R.E., Azam, S.: Hans-net: Hyperbolic convolution and adaptive temporal attention for accurate and generalizable liver and tumor segmentation in ct imaging (2025)

[12] Kang, M., Ting, C.M., Ting, F.F., Phan, R.C.W.: Cafct-net: A cnn-transformer hybrid network with contextual and attentional feature fusion for liver tumor segmentation. In: Proceedings - International Conference on Image Processing, ICIP, pp. 2970–2974. IEEE Computer Society, ??? (2024). https://doi.org/10.1109/ICIP51287.2024.10647768

[13] Jiang, L., Hu, J., Huang, T.: Improved swinunet with fusion transformer and large kernel convolutional attention for liver and tumor segmentation in ct images. Scientific Reports **15** (2025) https://doi.org/10.1038/s41598-025-98938-5

[14] Yashaswini, G.N., Manjunath, R.V., Shubha, B., Prabha, P., Aishwarya, N., Manu, H.M.: Deep learning technique for automatic liver and liver tumor segmentation in ct images. Journal of Liver Transplantation **17** (2025) https://doi.org/10.1016/j.liver.2024.100251

[15] Hille, G., Agrawal, S., Tummala, P., Wybranski, C., Pech, M., Surov, A., Saalfeld, S.: Joint liver and hepatic lesion segmentation in mri using a hybrid cnn with transformer layers (2023)

[16] Zhang, Y., Yang, J., Liu, Y., Tian, J., Wang, S., Zhong, C., Shi, Z., Zhang, Y., He, Z.: Decoupled pyramid correlation network for liver tumor segmentation from ct images (2022) https://doi.org/10.1002/mp.15723

[17] Song, Z., Wu, W., Wu, S.: Multi-scale convolutional attention and structural re-parameterized residual-based 3d u-net for liver and liver tumor segmentation from ct. Sensors **25** (2025) https://doi.org/10.3390/s25061814

[18] Chlebus, G., Schenk, A., Moltz, J.H., Ginneken, B., Hahn, H.K., Meine, H.: Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based postprocessing. Scientific Reports **8** (2018) https://doi.org/10.1038/s41598-018-33860-7

[19] Rafiq, M., Ali, H., Mujtaba, G., Shah, Z., Azmat, S.: Cross modality medical image synthesis for improving liver segmentation (2025)

[20] Yang, M., Wu, Z., Zheng, H., Huang, L., Ding, W., Pan, L., Yin, L.: Cross-modality medical image segmentation via enhanced feature alignment and cross pseudo supervision learning. Diagnostics **14** (2024) https://doi.org/10.3390/diagnostics14161751

[21] Chen, X., Lian, C., Wang, L., Deng, H., Kuang, T., Fung, S., Gateno, J., Yap, P.T., Xia, J.J., Shen, D.: Anatomy-regularized representation learning for cross-modality medical image segmentation. IEEE Transactions on Medical Imaging **40**, 274–285 (2021) https://doi.org/10.1109/TMI.2020.3025133

[22] Fan, T., Wang, G., Li, Y., Wang, H.: Ma-net: A multi-scale attention network for liver and tumor segmentation. IEEE Access **8**, 179656–179665 (2020) https:

//doi.org/10.1109/ACCESS.2020.3025372

[23] Çiçek, Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation (2016)

[24] Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation (2016)

[25] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**, 834–848 (2018) https://doi.org/10.1109/TPAMI.2017.2699184

[26] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B.: Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**, 3349–3364 (2021) https://doi.org/10.1109/TPAMI.2020.2983686

[27] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn (2018)

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

[29] Jurdi, R.E., Petitjean, C., Cheplygina, V., Honeine, P., Abdallah, F.: Effect of prior-based losses on segmentation performance: A benchmark (2022)

[30] Pereira, G.: Particle swarm optimization. Technical report (2014). https://www.researchgate.net/publication/228518470