# Interactive click-prompt in CT Liver Lesion Segmentation using Response Fusion Attention

**Jessica C. Delmoral**
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial
Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465
PORTO, PORTUGAL
up201200524@edu.fe.up.pt

**João Manuel R.S. Tavares**
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial
Faculdade de Engenharia, Departamento de Engenharia Mecânica, Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-465
PORTO, PORTUGAL
tavares@fe.up.pt

## Abstract

Accurate detection of liver cancer lesions on multiphase computed tomography (CT) remains a critical step in diagnosis and treatment planning. However, manual lesion delineation is time-consuming, subjective, and limited in its exploitation of available imaging data. The TriALS challenge focuses on the segmentation quality in a setting where the anatomical image consists in multi-phase CT with no-contrast. Furthermore, the challenge aims to validate the contribution of a human-in-the-loop segmentation prompt providing varying levels of simulated foreground and background clicks, to the final segmentation quality. In this work, we present a deep learning model, trained within the nnUNet framework, fusing Residual U-Net encoding/ decoding, Response Fusion Attention at skip connections to adaptively guide decoder refinements and 5 fold model ensembling.

***K*eywords** TriALS · CT · Attention · Deep Learning

## 1 Introduction

Hepatocellular carcinoma (HCC) remains one of the most prevalent and lethal malignancies worldwide. Accurate detection and segmentation of liver lesions are crucial for diagnosis, staging, and treatment planning. Multiphase contrast-enhanced computed tomography (CT), encompassing arterial, portal venous, and delayed phases, provides complementary information on lesion vascularization patterns. However, manual assessment is time-consuming and subject to inter-observer variability. Recent advances in deep learning have enabled automated lesion detection and segmentation, with transformer-inspired attention mechanisms offering further improvements in leveraging multiphase data. However, models still often struggle to distinguish between liver, vasculature, bone and tumoral regions. One approach to address this issue is to include a human expert in the segmentation loop. By minimizing an expert's intervention to providing simple clicks to indicate whether a given pixel/region corresponds to a lesion or merely reflects other anatomical structures, this information can guide the segmentation model. Improving the reproducibility and efficiency of liver lesion detection remains a challenge in the following aspects:

- Multi-lesion variability: lesions present in multiple sizes, number and liver segment location, often with distinct texture patterns as well, complicating segmentation [1].
- Multi-acquisition infrastructure: Variations in scanner hardware, acquisition protocols, and reconstruction parameters limit model generalizability [2].

- False positives from other liver structures such as arteries, veins and cysts: Due to their similarity in shape and representation in CT model often result in a greater ammount of false positives hindering overall performance.
- Lack of human-in-the-loop integration: The attempts to produce fully automatic solutions are pertinent, however, the possibility to introduce expert feedback to guide inference is more robust framework, especially when facing out-of-distribution new image examples.

The TriALS challenge addresses these gaps by simulating an interactive segmentation scenario in which models receive incremental foreground and background click annotations. This setup allows for the investigation of how minimal expert input can guide and refine model predictions in real time. In this work, we focused in exploring the performance contribution of adding click-prompt inputs to the segmentation framework through Gaussian heatmap encoding. We hypothesize that such conditioning enables the model to rapidly correct false positives and recover missed lesions, even with sparse clicks.

## 2 Related Work

The topic of liver lesion segmentation has been explored in MICCAI public competitions for several years, with a variety of datasets, encompassing different acquisition conditions. The most famous instance however, is the Liver Tumor Segmentation (LiTS) Challenge dataset which has become a central resource for evaluating algorithms in liver lesion detection and segmentation [3]. Containing expert-annotated CT scans, the dataset captures the heterogeneity of lesion appearances and has enabled the comparison of diverse approaches under standardized conditions. The availability of such challenge datasets has accelerated the adoption of deep learning models and set performance benchmarks for clinical computer-aided diagnosis. Recent advances in architecture design, such as attention-mechanisms and ConvNeXt-inspired networks [4], are very often benchmarked using this dataset, since it is one of the biggest publicly available datasets containing liver lesion segmentation masks.

## 3 Methods

Our segmentation framework is built upon the Response Fusion Attention (RFA) mechanism [5] incorporated into a Res-UNet model, and trained for liver and lesion segmentation in CT image and interactive click-conditioned segmentation. The original RFA block was designed for optic disc/cup segmentation; here, we extend it to process volumetric CT data, at the high- and low- dimensional fusion sites of skip connections of the U-Net architecture to adaptively weight and integrate encoder–decoder features.

### 3.1 Data Pre-processing

The cohort of 80 venous phase CT images provided in the challenge, coupled with the 131 CT images of the Lits dataset, were combined and used as the training data. Volumes of CT and corresponding click maps are resampled to isotropic voxels, Z-score normalized to range [-1, 1] and clipped to [-300, 300 HU], respectively. Data augmentation operations including fliping, gaussian processing and shifting were applied randomly during training.

A combination of generated Foreground click map and a background click maps is used as the final click heatmap. The Foreground map generation starts with a zero-initialized array $heatmap_{tumor}$ with the same dimensions as the CT volume. For each tumor click, the corresponding voxel was set to 1 in a local array, smoothed with a Gaussian filter, and added to $heatmap_{tumor}$. The Background map generation starts with a zero-initialized array $heatmap_{background}$. Each background click was set to 1, followed by Gaussian filtering. Voxels overlapping with $heatmap_{tumor} > 0$ were reset to zero.

The final heatmap can be obtained as:

$$heatmap = heatmap_{tumor} - heatmap_{background}$$

Positive values indicate regions likely belonging to lesions, while negative values represent voxels influenced by background clicks, guiding the network to avoid false positives.

### 3.2 Training and Evaluation

The nnUNet framework was used to train the proposed RFA Res-UNet models. More precisely, an extension of the models present in the package dynamic network architectures was developed. The proposed model consists of the classical Residual U-Net encoder and decoders, which was combined with an Attention Fusion module, combining the encoding and decoding paths features. NnuNet performs 3D inference via cropped patches of fixed size, which in our case was set to [128, 128, 128]. The encoder consisted in 6 convolutional stages, and 32, 64, 128, 256, 320, 320 feature maps extracted in each stage. A set of five model with the described architecture were trained with a training/validation dataset split generated in 5-fold cross-validation. The models were trained with 250 iterations per epochs, during 200 epochs. Models were trained using an equally weighted Dice + cross-entropy hybrid loss and optimized with Stochastic Gradient Descent algorithm. Evaluation included Dice similarity coefficient (DSC) for segmentation accuracy, sensitivity and recall for lesion detection, and average symmetric surface distance (ASSD) for boundary precision.

## 4 Preliminary Results

Preliminary 5-fold cross validation results performed over the provided dataset (the 80 venous phase CTs), are presented in Table 1, where a comparison between the lesion segmentation results with, and without the usage of expert click inputs.

Table 1: Five-fold cross-validation performance of RFA Res-UNet models on the challenge dataset: Task 1 (CT-only inputs) and Task 2 (CT combined with click-based heatmap inputs).

| Algorithm | Dice | ASSD | Precision | Recall |
|---|---|---|---|---|
| CT only | 43.66 | 19.76 | 77 | 40.8 |
| CT + click heatmap | 77.35 | 2.2 | 81.40 | 79.35 |

## 5 Discussion and Conclusion

The proposed Attention Fusion Res-UNet model, trained within the nnU-Net framework, leverages CT and click encoded feature maps to aid in the correct liver cancer lesion detection. The preliminary performance evaluation results suggest that the click encoding maps generated, significantly improve overall performance. Leveraging the nnU-Net framework ensures standardized and reproducible training, facilitating broader adoption in clinical research. This approach represents a step toward precision radiology, where automated multiphase integration can augment radiologist decision-making and improve patient outcomes.

**Github Repository** Link to source code Github repository: https://github.com/Jess-Co-Del/TriALS2025

**Registered Team Name: SlicenDice** Do the members of the team listed above agree to make their submission public as part of the challenge archive? Yes

## References

[1] Jessica C Delmoral and João Manuel RS Tavares. Semantic segmentation of ct liver structures: a systematic review of recent trends and bibliometric analysis: neural network-based methods for liver semantic segmentation. *Journal of Medical Systems*, 48(1):97, 2024.

[2] Akash Nayak, Esha Baidya Kayal, Manish Arya, Jayanth Culli, Sonal Krishan, Sumeet Agarwal, and Amit Mehndiratta. Computer-aided diagnosis of cirrhosis and hepatocellular carcinoma using multi-phase abdomen ct. *International journal of computer assisted radiology and surgery*, 14(8):1341–1352, 2019.

[3] Patrick Bilic, Patrick Christ, Hongwei Bran Li, Eugene Vorontsov, Avi Ben-Cohen, Georgios Kaissis, Adi Szeskin, Colin Jacobs, Gabriel Efrain Humpire Mamani, Gabriel Chartrand, et al. The liver tumor segmentation benchmark (lits). *Medical image analysis*, 84:102680, 2023.

[4] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F. Jaeger, and Klaus Maier-Hein. Mednext: Transformer-driven scaling of convnets for medical image segmentation, 2024.

[5] Siddhartha Mallick, Jayanta Paul, and Jaya Sil. Response fusion attention u-convnext for accurate segmentation of optic disc and optic cup. *Neurocomputing*, 559:126798, 2023.