Learning to Jointly Segment the Liver, Lesions and Vessels from Partially Annotated Datasets

Omar Ali, Alexandre Bone, Marc-Michel Rohe, Eric Vibert, Irene Vignon-Clementel

LEARNING TO JOINTLY SEGMENT THE LIVER, LESIONS AND VESSELS FROM PARTIALLY ANNOTATED DATASETS

Omar Ali[1,2,3], Alexandre Bone[1], Marc-Michel Rohe[1], Eric Vibert[2], Irene Vignon-Clementel[3]

[1] Guerbet Research, France
[2] Paul Brousse Hospital – APHP, Inserm U1193, Paris Saclay University, France
[3] Inria, France

ABSTRACT

The segmentation of the liver, lesions, and vessels from pre-operative CT scans is of major importance in hepatic surgery planning. However, large databases with reference segmentations for these regions of interest remain unavailable, a challenge often encountered in medical image segmentation. In this work, we propose the FuSe loss, a novel loss function for multi-task learning on datasets with partial annotations. By employing the nnU-Net's 3D self-configuring pipeline to calibrate and train a deep network for the joint segmentation of the liver, lesions, and vessels, we show how the FuSe loss allows to learn from the differently annotated IRCAD and LiTS datasets, improving the overall baseline segmentation performance. With the FuSe loss, the dice scores reached up to 95.9%, 70.6% and 60.0% for the liver, lesions, and vessels respectively.

Index Terms— Semantic segmentation, multi-task learning, partially labeled data, weighted loss function

1. INTRODUCTION

Liver cancers are notable contributors to the global burden of cancer lethality, ranking fourth in the most common causes of cancer related deaths [1]. Of the available curative treatments, liver resection continues to be the most frequent, with a five-year survival rate exceeding 50% [2]. In planning safe and successful liver resections, pre-operative segmentations are often required, particularly for the liver, the lesions, and the hepatic vessels, to better localize the tumor in the liver region. Ultimately, similar pre-surgical practices can be expanded to support diagnostic and prognostic decisions, spurring the development of automatic tools for different medical imaging segmentation tasks.

While manual segmentations remain cumbersome and tedious, newly developed deep learning methods have achieved remarkable results for various automatic segmentation tasks [3]. Typically, these methods rely heavily on UNet based architectures, which were first introduced in 2015 by [4]. Thereafter, several architectural improvements to the original model have been proposed such as the ResNets, and the DenseNets [5], [6]. Most recently, having won several segmentation challenges, the nnU-Net has been introduced as a new state of the art deep learning-based method for medical imaging segmentation tasks [7].

In the literature, automatic *liver* segmentation with deep learning methods has been profoundly investigated with overall solid performances [8], [9]. However, *liver lesion* segmentation remains a challenge to tackle in medical imaging, despite the recent methodological advancements proposing cascaded approaches for parallel or sequential liver and lesion learning, and feature fusion encoder-decoder networks [10-12]. Moreover, *hepatic vessel* segmentation has recently gained some ground with methods proposing 2D filtered multi-view inputs, some utilizing variants of the 3D UNets with custom loss functions, and others training on noisy and low-quality vessel labels [13-15]. Similar to lesion segmentation, automatic liver vessel segmentation remains challenging to achieve.

To the best of our knowledge, multi-class models performing joint liver, lesion, and vessel segmentations are not yet established, likely due to the lack of largely annotated datasets with the three labels jointly available. To overcome this challenge, we propose the FuSe loss, a novel multi-class loss function that allows the joint segmentation of potentially overlapping classes like the liver and its vessels, while concurrently handling the union of fully and partially labeled images, thereby granting an increase in the size of the training dataset. Note that although previous works such as [16] have proposed multi-organ segmentation models from the combination of fully and partially labeled datasets, the considered organs occupied distinct and non-overlapping volumes.

The suggested FuSe loss is leveraged to optimize a segmentation model that jointly segments the liver, its lesion(s), and its vessels from the IRCAD [17] and LiTS [18] datasets, respectively considered as *"fully"* and *"partially"* annotated in this chosen task. A weighted version of the FuSe loss is also evaluated, with the objective to correct the imbalance between the different label frequencies. In the subsequent methods section, the FuSe loss function is detailed, followed by a description of the datasets and experiments. Then, a qualitative and quantitative assessment of the liver, lesion(s), and vessel segmentations is portrayed. Lastly, the paper concludes with a short synopsis, and some future perspectives.

2. METHODS

2.1 Loss Function

Let $L(x, y)$ be a generic loss function optimized to learn the discrepancies between x and y, where $x \in [0,1]^{N*K}$ is the predicted probability tensor, $y \in \{0,1\}^{N*K}$ is the corresponding reference segmentation map, K is the total number of available classes, and N is the total number of voxels.

Typically, in medical imaging segmentation tasks, $L(x, y)$ is defined as the summation of the dice and cross entropy losses:

$$\begin{cases} L(x,y) = DSC_{loss}(x,y) + CE_{loss}(x,y) & (1) \\ DSC_{loss}(x,y) = -\frac{2}{K}\sum_{k=0}^{K-1}\frac{\sum_{n=1}^{N} x_n^k \cdot y_n^k}{\sum_{n=1}^{N} x_n^k + \sum_{n=1}^{N} y_n^k} & (1a) \\ CE_{loss}(x,y) = -\frac{1}{N}\sum_{k=0}^{K-1}\sum_{n=1}^{N} y_n^k \log(x_n^k) & (1b) \end{cases}$$

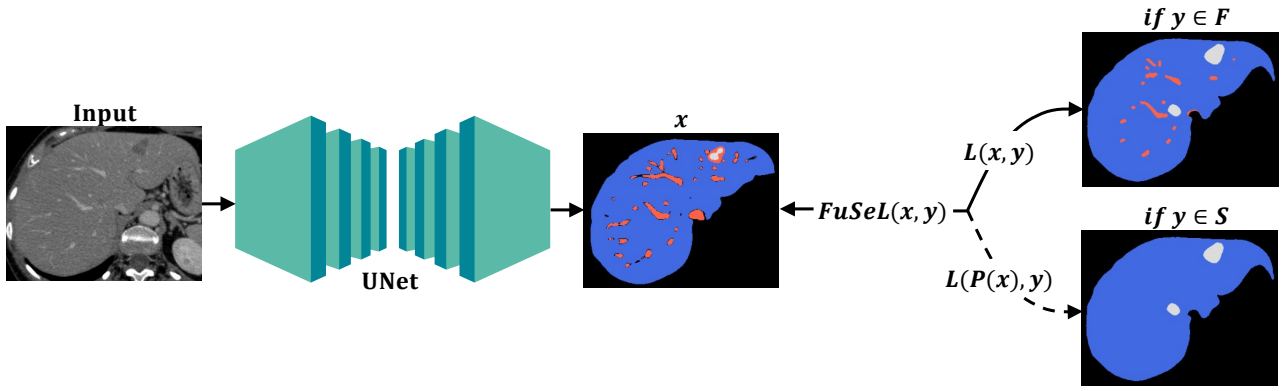

Figure 1: Proposed deep learning framework where the input is a raw CT cropped around the liver, x is the predicted probability tensor, and y is the target segmentation. The light blue, grey, and orange colors represent the liver, lesion(s), and vessels respectively

2.2 FuSe Loss

Partial annotations are often encountered in deep learning applications. With the proposed FuSe loss, the models can learn from varying annotation protocols, regardless of the possible overlap between the classes.

Let F be the finite set of available annotations $y \in \{0,1\}^{N*K}$ with K total classes, and S be a secondary set of annotations $y \in \{0,1\}^{N*(K-m)}$ with $K-m$ total classes, where $0 < m < K$. The FuSe loss is defined as:

$$FuSeL(x,y) = \begin{cases} L(x,y) & if\ y \in F \\ L(P(x),y) & if\ y \in S \end{cases} \quad (2)$$

We denote F as the finite set of *fully*-labeled reference segmentations , S as the finite set of *semi*-labeled reference segmentations, N as the total number of voxels, and P as a projection function that maps $x \in [0,1]^{N*K}$, to $P(x) \in [0,1]^{N*(K-m)}$. Figure 1 illustrates the suggested deep learning framework.

With a consistent supervision of the set of labels in the reference segmentation map, either the baseline loss function is computed for fully labeled targets, or the softmax output probabilities are mapped to match the number of classes in the semi-annotated targets prior to the computation of loss function using $P(x)$.

In this work, $F \subset \{0,1\}^{N*4}$ is the set of background, liver, lesion, and vessel labels respectively, $S \subset \{0,1\}^{N*3}$ is the set of background, liver and lesion labels, and $P(x)$ defined below is computed whenever the target y segmentation is partially labeled.

$$P(x) = \begin{cases} x^{background} = x^{k=0} \\ x^{liver} + x^{vessel} = x^{k=1} + x^{k=3} \\ x^{lesion} = x^{k=2} \end{cases} \quad (3)$$

2.3 Weighted FuSe Loss

Learning from different labels in different datasets can trigger imbalances during training particularly if the datasets are considerably different in size. Thus, a weighted coefficient w is defined for every class based on the cardinality of sets F and S such that $\sum_{k=1}^{K} w^k = 1$. In this work, the weights of the liver, lesions, and vessel classes are defined as:

$$w^1 = w^{liver} = \frac{1}{2} \cdot \frac{|S|}{|F| + |S|} \quad (4)$$

$$w^2 = w^{lesion} = w^{liver} \quad (5)$$

$$w^3 = w^{vessels} = \frac{|F|}{|F| + |S|} \quad (6)$$

3. EXPERIMENTS

3.1 Datasets and Preprocessing

IRCAD and LiTS are public datasets with contrast enhanced portal phase computed tomography (CT) abdominal scans [17, 18]. While IRCAD's dataset is fully labeled and consists of 20 patients with liver, lesion, and vessel annotations, LiTS' dataset consists of a total of 201 partially labeled patients with 131 patients having liver and lesion annotations where the hepatic vessels are considered as liver, and the remaining 70 patients are unannotated and only used as a test set to evaluate the liver and lesion segmentations post-training.

In preprocessing, IRCAD's vessel target annotations, which include the segmentations of the entire venous system, are intersected with their corresponding liver masks to disregard any vessel label outside the liver. In addition, the nnU-Net's default preprocessing configuration is employed [7]. It includes a resampling of both the CT scans and the annotations to their median voxel spacing, which corresponds to a spacing of $0.78 \times 0.78 \times 1.5$mm. The input images are configured to a $192 \times 192 \times 60$ patch size.

3.2 Model Architecture

The nnU-Net's 3D full resolution framework is used for the totality of the experiments [7]. The employed model follows the template of the original U-Net architecture, having an encoding path generating incrementally deeper feature maps and a symmetrical decoding path, outputting the predicted segmentations. The feature maps increase from 32 to 320 after each of the 5 pooling layers. The convolutions are either stridden with kernels of size 2^3 or not stridden with kernels of size 3^3. Aside from the sotfmax with 4 channels used in the model's final layer (one channel per class), the entirety of the used activation functions are Leaky-ReLus. Lastly, skip connections and deep supervision layers are used to ease the gradient flow across the network. The setup of the model architecture is detailed in [7].

3.3 Training and Evaluation

Three experiments were conducted to learn the joint segmentation of the liver, the lesions, and the hepatic vessels. The entirety of the

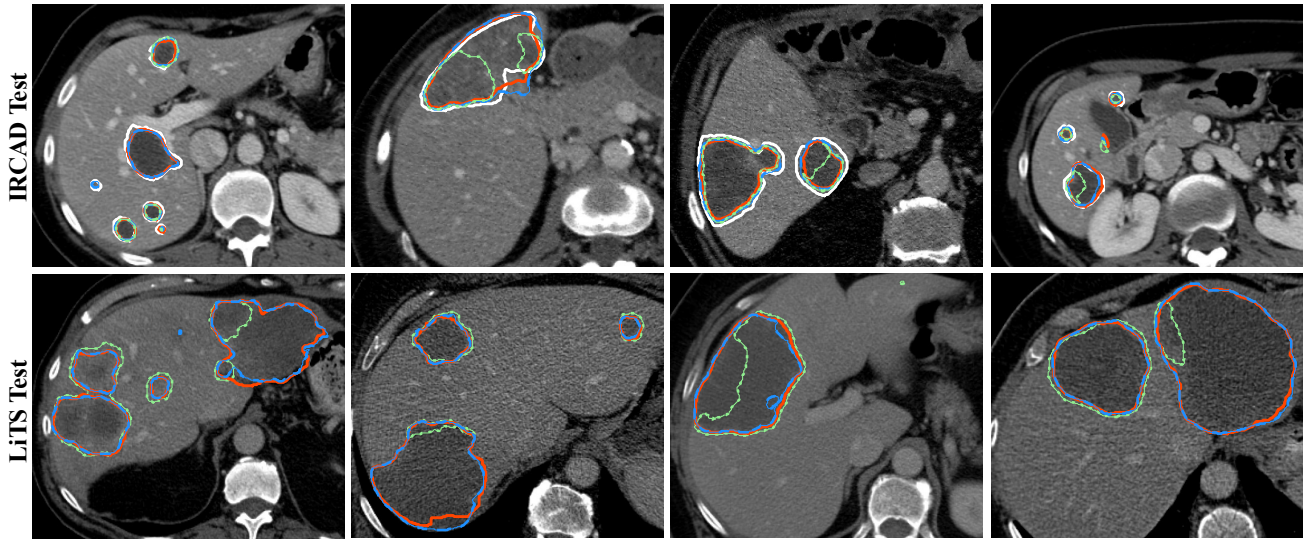

Figure 2: Lesion segmentation on IRCAD and LiTS' test sets. White contours are the ground truth segmentations, which are only available for IRCAD's dataset, red contours are the segmentations with the *non-weighted FuSe* loss, dashed blue contours are the segmentations with the *weighted FuSe* loss, and the dotted green contours are the segmentations with the *baseline* loss.

models in all the experiments are trained for 150 epochs using the stochastic gradient descent algorithm with Nesterov momentum, and a geometrically decaying learning rate.
Data augmentation, including random rotations, flips and elastic deformations is also applied to help avoid overfitting.

The first experiment, referred to as the *baseline* experiment trains and optimizes an nnU-Net on IRCAD's dataset using the *baseline* loss function $L(x, y)$, then evaluates the results using two approaches described below. The second and third experiments referred to as the *FuSe Loss* and *weighted FuSe Loss* experiments respectively, train and optimize a similar nnU-Net using the non-weighted and weighted $FuSeL(x, y)$ functions, on the fusion of IRCAD and LiTS' datasets (where the 131 patients in LiTS' annotated dataset are only used during training, to learn the detection of the liver and the lesions). The training for both experiments is carried out on identical fold partitions as the first experiment.

As for the evaluation, given the availability of the vessel labels solely in the IRCAD dataset, the first evaluation approach assesses the segmentation results of the three classes and consists of performing a five-fold cross validation on the IRCAD dataset to test the trained models on the entire set of patients. Consequently, different splits of 16 training patients and 4 testing patients are created for every fold. Furthermore, with the absence of the vessel labels in LiTS' unannotated test set of 70 patients, the second evaluation approach combines the predicted segmentations of the liver and the vessels using a mapping function similar to $P(x)$, prior to the assessment of the results on LiTS' online evaluation platform.

4. RESULTS

4.1 Qualitative Results

Figure 3 shows the liver, lesion, and vessel segmentation results for the FuSe loss experiment on IRCAD's dataset. It can be seen that the proposed method is able to accurately reproduce the segmentations of the target for the three classes while mainly over-segmenting regions around the liver vessels. In addition, Figure 2 shows a set of lesion segmentations for different patients in IRCAD and LiTS' test sets. The lesion segmentation results with the non-weighted and weighted FuSe loss configurations outperform those obtained with the baseline loss. However, the three experiments display poor performances with patients having a small sized single nodule or nodules located at the extremities of the liver. Those results can be seen in Figure 4 below.

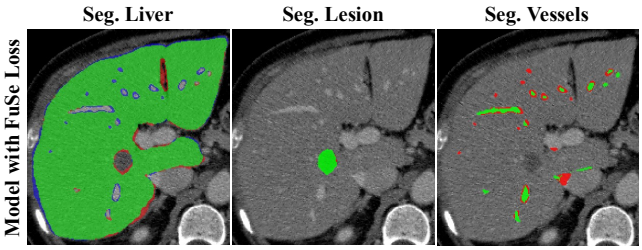

Figure 3: Liver, lesion, and vessel segmentations with the FuSe loss, where blue shows under-segmentations, red shows over segmentations, and green shows the true match with the target.

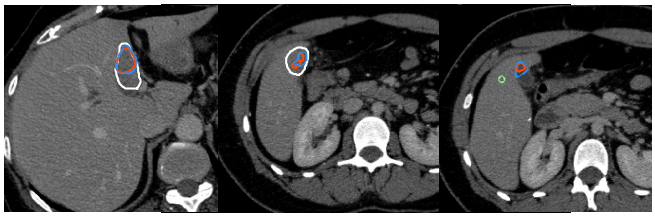

Figure 4: Poor lesion segmentations from IRCAD's test set (same color coding as Figure 3).

4.2 Quantitative Results

The liver, lesion, and vessel dice and Hausdorff scores of the baseline experiment are compared the FuSe loss and the weighted FuSe loss experiments on the IRCAD dataset. Table 1 highlights the 5-fold cross validation results on IRCAD's dataset, and the liver and lesion segmentation results on LiTS' unannotated test set, regarded-

Table 1: Mean and standard deviation of the dice (%) and Hausdorff distance (mm) scores evaluated on the different test sets. Bold and underlined values are the best and second-best results respectively.

Exp.[1] / Conf.[2]	Lesion				Liver				Vessel	
	LiTS		IRCAD		LiTS		IRCAD		IRCAD	
Metrics	Dice	Hausdorff	Dice	Hausdorff	Dice	Hausdorff	Dice	Hausdorff	Dice	Hausdorff
Baseline	58.8 (NA[3])	6.69 (NA)	<u>51.1</u> (±15.6)	70.48 (±76.18)	92.9 (NA)	31.94 (NA)	**94.8** (±1.7)	24.71 (±8.6)	**62.4** (±6.8)	<u>48.67</u> (±48.4)
FuSe Loss	**70.6** (NA)	**6.12** (NA)	**51.9** (±4.7)	**60.39** (±72.62)	**95.9** (NA)	<u>26.99</u> (NA)	<u>94.6</u> (±2.0)	**22.26** (±7.03)	60.0 (±5.1)	50.89 (±50.4)
Weighted FuSe Loss	<u>66.6</u> (NA)	<u>6.54</u> (NA)	48.8 (±7.1)	<u>66.58</u> (±70.86)	<u>95.5</u> (NA)	**26.95** (NA)	94.1 (±1.8)	<u>22.88</u> (±10.15)	<u>60.6</u> (±10.7)	**46.54** (±46.1)

[1.] Experiment, [2.] Configuration, [3.] Not assigned as values are not reported in LiTS

as the benchmark for liver and lesion segmentation evaluation.

The segmentations obtained with the proposed FuSe loss configuration for the liver and lesions on LiTS' test set clearly outperform those obtained with the baseline loss, with an increase in the dice score from 58.8% to 70.6% and a decrease in the Hausdorff distance from 6.69mm to 6.12mm for lesion segmentations. Similarly, for liver segmentation, the dice increases from 92.9% to 95.9% and the Hausdorff distance decreases from 31.94mm to 26.99mm on LiTS for the FuSe loss configuration. Furthermore, the results of the proposed method for liver, lesion, and vessel segmentation on IRCAD's dataset remain within a close margin of the baseline experiment, improving the average lesion segmentation dice and Hausdorff scores from 51.1% to 51.9% and from 70.48mm to 60.39mm respectively. However, the liver dice scores slightly regress by 0.2% and the vessels dice scores by 2.4%, with an increase in the Hausdorff distance from 48.67mm to 50.89mm for the vessels. Moreover, the results show that the added weights in the weighted FuSe loss experiment do not improve the overall dice scores of the three classes. Despite the 0.6% increase in the vessel's dice score compared to the FuSe loss configuration, the liver and lesion's dice scores decreased by 0.5%, 0.4%, 3.1%, and 4% on IRCAD and LiTS's datasets respectively. However, the Hausdorff distances computed with the weighted FuSe loss configuration, outperform those computed with the baseline configuration for the liver, lesion, and vessel classes.

5. DISCUSSION AND CONCLUSION

Automatic liver, lesion, and vessel segmentation is essential in presurgical planning. With the unavailability of large extensively annotated datasets, the proposed FuSe loss allowed to learn an accurate joint segmentation model from partially labeled training data via a projection function that maps the different predictions to their reference labels. The results on LiTS' test set show that in the FuSe configuration the lesion dice and Hausdorff sores improved by 11.8%, 0.57mm and the liver dice and Hausdorff scores improved by 3% and 2.45mm respectively. The proposed method also achieved solid segmentation results on IRCAD's dataset for liver, lesion, and vessel segmentations with dice scores attaining 94.6%, 51.9%, and 60.0% for each class respectively.

Further investigation on the implications of the size of the training dataset on the vessel segmentation performance is required. In particular, additional consideration is required on the choice of weights since imbalances arise not only due to the dataset size variations but also due to the varying number of voxels representing every class in the reference segmentation map. Finally, the proposed method will be further extended and tested on different segmentations tasks presenting similar challenges.

Acknowledgments This work was supported by Guerbet and the Region Ile-de-France.

6. REFERENCES

[1] A. Villanueva, "Hepatocellular Carcinoma," *The new england journal of medicine,* vol. 380, no. 15, pp. 1450-1462, 2019.

[2] D.H. Lee et al., "Long-term surgical outcomes in patients with hepatocellular carcinoma undergoing laparoscopic vs. open liver resection: A retrospective and propensity score-matched study," *Asian J Surg,* vol. 44, no. 1, pp. 206-212, 2021.

[3] N. Tajbakhsh, L. Jeyaseelan, and Q. Li, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis,* vol. 63, no. 1361-8415, p. 101693, 2020.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segementation," in *The Medical Image Computing and Computer Assisted Intervention Society 2015,* Munich 2015.

[5] K.He et al., "Deep Residual Learning for Image Recognition," *Conference on Computer Vision and Pattern Recognition 2016,* pp. 770-778, 2016.

[6] G. Huang, Z. Liu, and L.v.d. Maaten, "Densely Connected Convolutional Networks," *Conference on Computer Vision and Pattern Recognition 2017,* pp. 2261-2269, 2017.

[7] F. Isensee, Kohl et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods,* vol. 18, pp. 203-211, 2021.

[8] X. Li et al., "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Transactions on Medical Imaging,* vol. 37, no. 12, pp. 2663-2674, 2017.

[9] A.B. Habib et al., "Performance Analysis of Different 2D and 3D CNN Model for Liver Semantic Segmentation: A Review," in *The Molecular Imaging and Contrast Agent Database 2020*, Singapore, 2020.

[10] P.F. Christ et al., "Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks," 23 February 2017. https://arxiv.org/pdf/1702.05970.pdf.

[11] E. Vorontsov et al., "Liver Lesion Segmentation Informed by Joint Liver Segmentation," in *The International Symposium on Biomedical Imaging 2018*, Washington, 2018.

[12] X. Chen, R. Zhang, and P. Yan, "Feature Fusion Encoder Decoder Network for Automatic Liver Lesion Segmentation,"

The International Symposium on Biomedical Imaging 2019, Venice, 2019.

[13] T. Kitrungrotsakul et al., "VesselNet: A deep convolutional neural network with multi pathwaysfor robust hepatic vessel segmentation," *Computerized Medical Imaging and Graphics,* vol. 75, pp. 74-83, 2019.

[14] D. Keshwani et al., "TopNet: Topology Preserving Metric Learning for Vessel Tree Reconstruction and Labelling," in *The Medical Image Computing and Computer Assisted Intervention Society 2020*, Lima, 2020.

[15] M. Xu et al., "Training Liver Vessel Segmentation Deep Neural Networks on Noisy Labels from Contrast CT Imaging," in *The International Symposium on Biomedical Imaging 2020*, Iowa City, 2020.

[16] Y. Zhou et al., "Prior-aware neural network for partially-supervised multi-organ segmentation", *in The International Conference on Comupter Vision* 2019, Seoul.

[17] "3D-IRCADb 01," IRCAD, Available: https://www.ircad.fr/research/3d-ircadb-01/.

[18] P. Bilic et al., "The Liver Tumor Segmentation Benchmark (LiTS)," 2019, Available: https://arxiv.org/abs/1901.04056.