Deep learning for vascular segmentation and tissue characterization in CT images

Zhang, X.

4

Top-k maximum intensity projection priors for 3D liver vessel segmentation

This chapter was adapted from:

Zhang, X., Broersen, A., van Erp, G., Pintea, S.L. and Dijkstra, J., **Top-K Maximum Intensity Projection Priors for 3D Liver Vessel Segmentation.** (2025) In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pp. 1-5. IEEE.

Abstract

Liver-vessel segmentation is an essential task in the pre-operative planning of liver resection. State-of-the-art $2D$ or $3D$ convolution-based methods focus on liver vessel segmentation on $2D$ CT cross-sectional views, which do not take into account the global liver-vessel topology. To maintain this global vessel topology, we rely on the underlying physics used in the CT reconstruction process, and apply this to liver-vessel segmentation. Concretely, we introduce the concept of *top-k maximum intensity projections*, which mimics the CT reconstruction by replacing the integral along each projection direction, with keeping the top-k maxima along each projection direction. We use these top-k maximum projections to condition a diffusion model and generate $3D$ liver-vessel trees. We evaluate our $3D$ liver-vessel segmentation on the *3D-ircadb-01* dataset, and achieve the highest *Dice* coefficient, intersection-over-union (*IoU*), and *Sensitivity* scores compared to prior work.

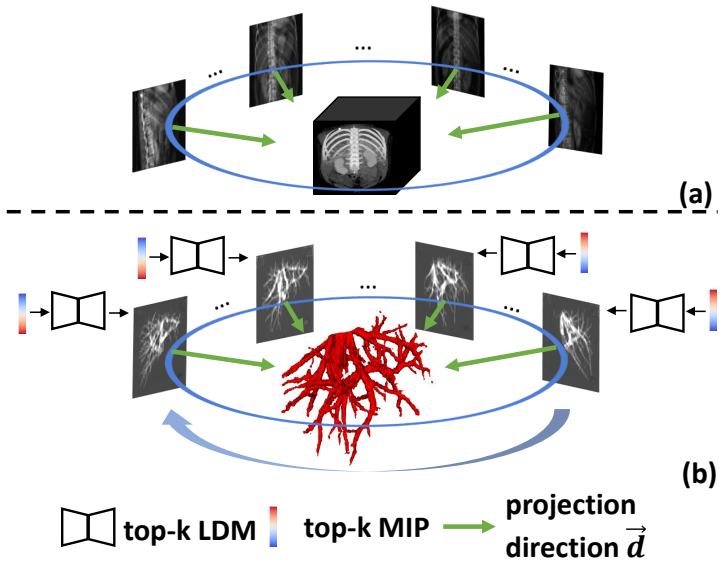

Figure 4.1: **(a) Standard CT reconstruction:** Given the integral projections and projections directions $\vec{d}$, it reconstructs the underlying $3D$ object by back-projecting the integral along each direction; **(b) Our proposed *top-k MIP*:** Given the projections directions $\vec{d}$, our model reconstructs the $3D$ liver-vessel tree by computing the top-k maximum value of the CT scan along each direction and inputting this into a latent diffusion model.

4.1 Introduction

Liver resection is the standard surgical treatment option for primary and secondary liver cancer [1]. In the preoperative planning of the tumor resection, it is essential to minimize perioperative blood loss [1]. Due to the complex anatomy of liver vessels, it is challenging to obtain accurate liver-vessel segmentations automatically.

Liver-vessel segmentation is currently done with convolutional neural networks (CNN). Prior work relies on $2D$ CNN [2, 3] or $3D$ CNN [4, 5] methods to perform segmentation on CT images. However, these methods use sub-volumes (cropped in three dimensions) of the CT images. These sub-volumes approaches do not take the global topology of the liver-vessel tree into account which could lead to discontinuous and incomplete vessel tree predictions.

To address these shortcomings, prior work [6, 7, 8] makes use of *maximum intensity projection* (MIP) [9]. Given a $3D$ volume and a set of directions, *MIP* computes the maximum voxel value along each direction. Therefore, *MIP* encodes the global vessel topology on $2D$ projections, and is characterized by high signal-to-noise ratio, while enhancing local vessel probability [9].However, existing *MIP*-based methods cannot perform accurately while solely relying on the *MIP* information. Due to the lack of $3D$ information, *MIP* is only an adjunct for segmentation in existing

MIP-based methods, which must be combined with a $3D$ U-Net model [6, 7, 8] or a simulated $3D$ filter [10]. Thus, they need to further encode $3D$ spatial information from the CT cross-sectional view by using $3D$ U-Net models [6] or intersecting 2D filters to simulate 3D convolutions [10]. Therefore, these methods fail to take full advantage of the global topological information of the *MIP*. Moreover, prior *MIP*-based methods focus on the cerebrovascular [10], coronary [6, 8], and retinal vessel [8] segmentation, whereas liver-vessels are characterized by more complex and dense $3D$ structures.

To incorporate the $3D$ liver-vessel structure, we introduce a *top-k maximum intensity projection (top-k MIP)*. Unlike *MIP*, which preserves only the maximum voxel value along any given direction, a *top-k MIP* preserves the top-k maximum values along that direction. This allows us to encode fine-grained information regarding the underlying $3D$ structure of the vessel tree. This information would be missed by simply taking the maximum. Therefore, our model no longer needs to rely on additional $3D$ convolutions over the CT cross-sectional view to encode this $3D$ structure.

Fig. 4.1(a) shows the standard CT reconstruction process: given $2D$ integral projections and their projection directions $\vec{d}$, the CT voxels can be computed by back-projecting the integral along each direction, thus reconstructing the $3D$ object. Similarly, in Fig. 4.1(b) we start from integral projections of liver vessel tree and projection directions $\vec{d}$, the vessel tree projections are computed based on the top-k maximum values along each direction, and we aim to reconstruct the $3D$ liver-vessel tree. To reconstruct the $3D$ liver-vessel tree use these *top-k MIP*s as conditions in a latent diffusion model (LDM) [11]. Overall, we make the following contributions: (1) we propose incorporating the $3D$ topology of the liver structure in a principled way by relying on the underlying physics used in CT reconstruction; (2) to this end, we propose a novel *top-k maximum intensity projection* that encodes the fine-grained $3D$ liver-vessel structure and combine this into a latent diffusion model; (3) finally, we demonstrate improved accuracy of our model on the *3D-ircadb-01* [12] dataset when compared to state-of-the-art methods such *nnUNet* [13], *SwinUNetr* [14], *EnsemDiff* [15] and *MedSegDiff* [16].

4.2 Methods

Our model, displayed in Fig. 4.2, represents the global $3D$ topological structure of the liver via the *top-k MIP* (denoted by $\mathbf{c}$) of the CT cross-sectional view. We further encode *top-k MIP* (via the yellow block) into the latent condition, $\mathbf{c}'$. This $\mathbf{c}'$ is used to condition the latent diffusion model (depicted in blue) which generates the integral projections of ground truth vessel tree by denoising a given noisy ground truth, in the latent space. The ground truth, $\mathbf{x}_0$, and the noisy input $\mathbf{x}_t$ are represented as *integral projections* over directions $\vec{d}$ of the true $3D$ vessel tree segmentation. These *integral*

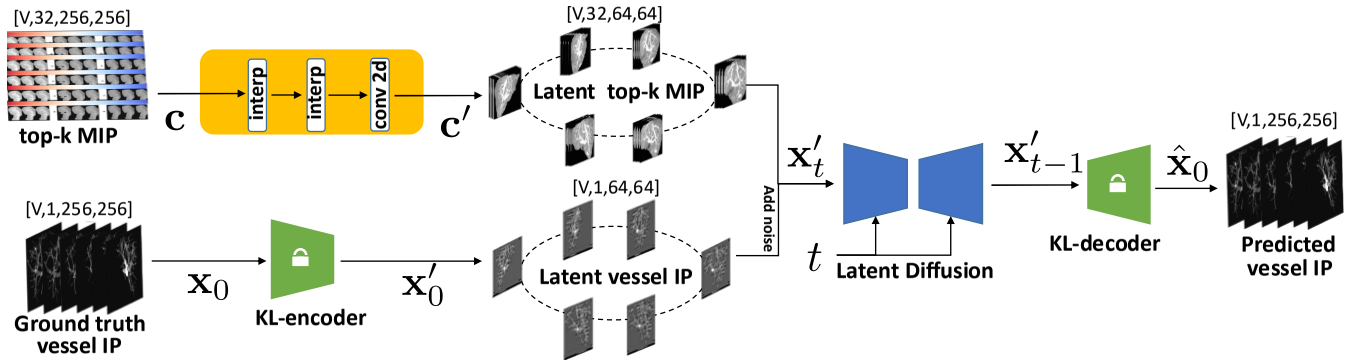

Figure 4.2: **Model outline.** Our model represents the global 3*D* topology of the liver by computing the *top-k MIP* over the CT volume, $\mathbf{c}$. Subsequently, it encodes these *top-k MIP* into a latent condition, $\mathbf{c}'$ (orange). This $\mathbf{c}'$ latent is used to condition the latent diffusion model (blue) which recovers the ground truth vessel tree $\mathbf{x}_0$ from noisy inputs $\mathbf{x}_t$. We represent the ground truth via *integral projections (IP)* of the 3*D* ground truth vessel tree. The ground truth $\mathbf{x}_0$ and the noisy input $\mathbf{x}_t$ are encoded via a KL-autoencoder (green) to be used in the latent diffusion U-net [11]. We denote the different viewing directions by V in the batch size.

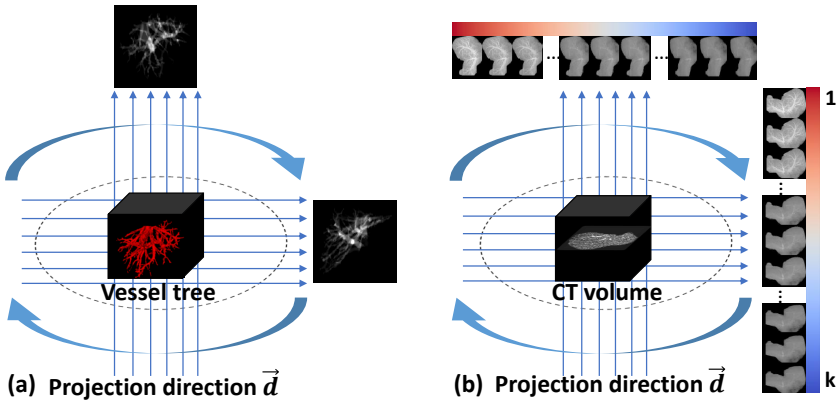

Figure 4.3: **(a) Ground truth IP:** We include all vessel tree slices in the *integral projection (IP)* and accumulate them along projection directions $\vec{d}$. **(b) Condition top-k MIP:** We include all CT slices in the *top-k maximum intensity projection (top-k MIP)* and keep the *top-k* maxima along projection directions $\vec{d}$;

projections (IP) are further encoded/decoded via an autoencoder (depicted in green), to be used in the latent diffusion model.

4.2.1 Ground truth transformation

For obtaining the ground truth *integral projection (IP)*, similar to the CT forward projection [17], we assume a bundle of parallel rays $R^{\vec{d}}$ across different directions $\vec{d}$, penetrating the ground truth binary liver-vessel tree, as in Fig. 4.3(a). The *IP* is the sum of the attenuation coefficient μ (μ=1 for binary liver-vessel trees) along a unit path $\Delta l^{\vec{d}}$, which is equivalent to the logarithm of the ratio between the ray intensity $R^{\vec{d}}$ after the attenuation by the liver vessel tree, and the initial ray intensity $R_0^{\vec{d}}$:

$$\mathbf{x}_0 \overset{\text{def}}{=} \sum_{\Delta l^{\vec{d}}} \mu = -\ln\left(\frac{R^{\vec{d}}}{R_0^{\vec{d}}}\right). \tag{4.1}$$

4.2.2 Top-k MIP transformation

For computing the condition $\mathbf{c}$ to be input to the latent diffusion model we use *top-k MIP*. Different from the integral attenuation in Eq. (4.1), *top-k MIP* keeps the top-k maxima along a ray path $\Delta l^{\vec{d}}$, as shown in Fig. 4.3(b):

$$\mathbf{c} \overset{\text{def}}{=} \text{top-}k\left\{\max\left(\Delta l_i^{\vec{d}}\mu, \Delta l_{i+1}^{\vec{d}}\mu, \ldots, \Delta l_n^{\vec{d}}\mu\right)\right\}, \tag{4.2}$$

where n is the number of CT voxels located along a ray, and $\Delta l_i^{\vec{d}}\mu$ is the value of the i-th CT voxel along the ray.

4.2.3 Top-k MIP conditioning latent diffusion model

We rely on a latent diffusion model [11] conditioned on the *top-k MIP*, $\mathbf{c}$, for recovering the denoised liver vessel *IPs*, $\hat{\mathbf{x}}_0$. We encode $\mathbf{c}$ into a latent vector via interpolation and convolution (yellow block in Fig. 4.2), into $\mathbf{c}'$. Following [11], we also encode the *IP*, $\mathbf{x}_0$, via a pre-trained Kullback-Leibler (KL)-divergence auto-encoder (green blocks in Fig. 4.2) to obtain $\mathbf{x}'_0$.

Standard in the diffusion forward process, we iteratively add Gaussian noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ to $\mathbf{x}'_0$. We aim to denoise the noisy inputs $\mathbf{x}'_t$, for $t \in \{T, .., 0\}$ by estimating the noise ϵ_θ, modelled as a U-Net [11] (depicted in blue in Fig. 4.2). For this, we optimize the parameters of the model, θ, by minimizing the loss over time steps, t:

$$\mathscr{L}_t(\mathbf{x}'_0, \epsilon, \theta) = \|\epsilon_\theta(\mathbf{x}'_t \mid \mathbf{c}', t) - \epsilon\|_2^2. \tag{4.3}$$

Once the model is trained, given the encoded *top-k MIP* condition $\mathbf{c}'$, and the input Gaussian noise $\mathbf{x}'_T$, we gradually denoise the noisy input $\mathbf{x}'_t, t \in \{T, .., 0\}$ to the estimate of the encoded *IP*, $\hat{\mathbf{x}}'_0$. Given the estimate of the encoded *IP*, $\hat{\mathbf{x}}'_0$, we decode this via the pre-trained KL-divergence auto-encoder (green block in Fig. 4.2) into an estimate of the *IP*, $\hat{\mathbf{x}}_0$.

4.2.4 Post-processing for artifact suppression

Given the estimated *IPs* of the vessel tree, $\hat{\mathbf{x}}_0$, we reconstruct the $3D$ liver vessel tree via filtered back projection (FBP) [17]. Given that, minor inconsistency in the projection domain can cause severe stripe artifacts in the reconstructed images, we employ a simple optimization to suppress these artifacts. Specifically, assume $\mathbf{T}$ initialized by CT image is a reconstructed liver vessel tree without artifact, and given a matrix $\mathbf{A}^{\vec{d}}$ recording the voxel indices penetrated by rays under different projection directions $\vec{d}$, we impose a projection consistency. Concretely, to dilute the irrelevant background introduced by the initialization of $\mathbf{T}$, we enforce that the projection of the reconstructed tree $\mathbf{T}$ on a certain projection direction encoded in $\mathbf{A}_{\vec{d}}$ should be as close as possible to the estimated *IP* under that projection direction, $\hat{\mathbf{x}}_0^{\vec{d}}$:

$$\mathbf{T} = \arg\min_{\mathbf{T}} \sum_{\vec{d}} \|\mathbf{A}^{\vec{d}}\mathbf{T} - \hat{\mathbf{x}}_0^{\vec{d}}\|_2^2. \tag{4.4}$$

In practice, performing this optimization 10 times is sufficient to suppress the stripe artifacts. Thus, we obtain the binary liver-vessel tree $\mathbf{T} \geq \text{percentile}(\mathbf{T}, 95)$. Finally, we perform connected region analysis to cancel small spurious predictions surrounding the segmented vessel tree.

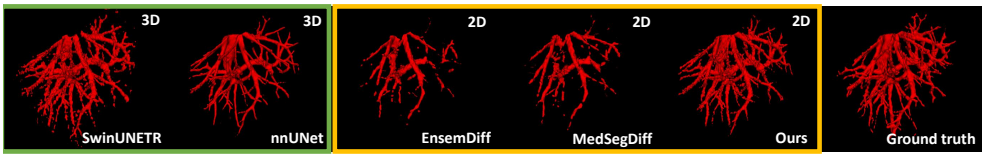

Figure 4.4: **Qualitative comparison on the *3D-ircadb-01* [12] dataset.** We highlight in green the predictions of the $3D$ methods and in yellow the predictions of the $2D$ methods. Our model makes more complete and continuous predictions, resembling the ground truth.

4.3 Experiments and results

4.3.1 Experimental setting

We evaluate our proposed model on the *3D-ircadb-01* [12] dataset, consisting of 20 CT scans. We exclude the vena cava from the region of interest, using liver masks to focus only the vessels within the liver. We resize the liver-masked ROIs to $[256\times256\times256]$ for each CT scan, and project 180 viewing directions from 0 to 180 degrees for each CT scan, totaling 3600 projections. k is 32 when creating the *top-k MIP*.

Small liver vessel annotations are incomplete in the *3D-ircadb-01* dataset [5]. To mitigate this, we asked clinical experts to score the completeness of the annotations based on a 5-point criterion and picked cases $(4, 6, 8, 11, 16)$ whose score ≥ 4 as the test set. We apply leave-one-out cross-validation, such that we have 19 cases (*i.e.* 3420 projections) for training and a single case (*i.e.* 180 projections) for testing, per fold.

We trained *nnUNet* [13] and *SwinUNetr* [14] using $3D$ CT sub-volumes as input, and using their default pre- and post-processing. For *MedSegDiff* [16] and *EnsemDiff* [15] we used $2D$ CT slices, and ensemble segmentation results 5 times. Our method does not require an ensemble. For all the experiments, we used an NVIDIA A100 (40GB) GPU.

4.3.2 Quantitative and qualitative comparison

We report voxel-wise Dice coefficient (*DSC*), Intersection-Over-Union (*IoU*), Sensitivity (*Sen*), Specificity (*Spe*) and centerline Dice coefficient (*clDice*) [18]. Table 4.1 shows the quantitative evaluation, while Fig. 4.4 shows a few prediction examples. Our method achieves the highest *DSC*, *IoU*, and *Sen* scores compared to the other baselines on the *3D-ircadb-01* dataset, demonstrating the added value of the proposed top-k MIP projections. *nnUNet* has a higher *clDice* score than our method. However, the vessel centerline was estimated from ground truth vessel annotations, following [18]. Therefore, incomplete and discontinuous voxel-wise ground truth could bias the centerline extraction in *clDice* [18]. Fig. 4.4 shows that both our method and *nnUNet* predict well-connected vessel structures compared to the ground truth, indicating that *clDice* might not be a sensible choice when comparing segments with inconsistent

Table 4.1: **Quantitative evaluation on the *3D-ircadb-01* [12] dataset**. Our method predicts complete liver-vessel segmentations, while having the highest *IoU*, *Sen* and *DSC* scores. The lower *clDice* scores of our method compared to *nnUNet* could be due to imprecise annotations. Finally, there is a trade-off between complete segmentation (higher *Sen*) or accurate segmentation (higher *Spe*).

	Conv type	View type	*DSC* (%)	*clDice* (%)	*IoU* (%)	*Sen* (%)	*Spe* (%)	*FLOPs* (G)	*Params* (M)
nnUNet[13]	3D	cross section	58.76 ± 9.89	$\mathbf{71.46} \pm 5.67$	42.31 ± 10.19	43.32 ± 11.18	$\mathbf{100} \pm 0$	2.90×10^3	30.7
Swin UNETR[14]	3D	cross section	57.80 ± 9.93	64.16 ± 7.10	41.31 ± 9.51	46.71 ± 13.21	99.96 ± 0.02	6.15×10^2	62.2
EnsemDiff[15]	2D	cross section	54.82 ± 9.64	60.61 ± 9.55	38.37 ± 9.21	40.05 ± 10.45	99.98 ± 0.02	9.96×10^2	113.7
MedSegDiff[16]	2D	cross section	59.59 ± 7.73	66.03 ± 8.05	42.85 ± 7.53	47.38 ± 10.44	99.95 ± 0.05	1.05×10^3	136.8
Ours	2D	projection	$\mathbf{64.25} \pm 7.19$	65.87 ± 8.16	$\mathbf{47.75} \pm 8.00$	$\mathbf{54.03} \pm 11.23$	99.96 ± 0.01	6.83×10^2	84.5

Table 4.2: **Liver vessel *IPs* estimation with (w/) and without (w/o) artifact suppression (opt)**. Artifact suppression makes the vessel *IPs* more similar to the ground truth vessel *IPs*.

	PSNR(↑)	SSIM(↑)
w/o opt	13.51 ± 0.78	0.45 ± 0.07
w/ opt	$\mathbf{15.15} \pm 1.32$	$\mathbf{0.66} \pm 0.05$

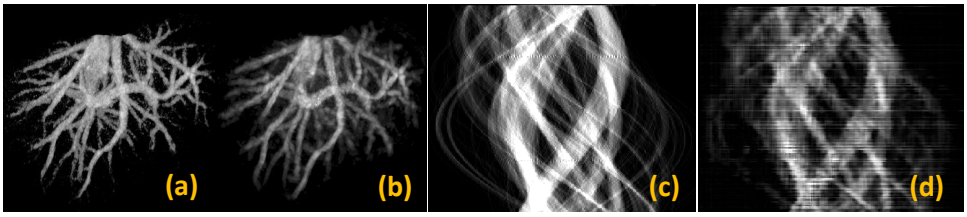

Figure 4.5: **Vessel reconstruction with/without artifacts suppression.** (a) reconstructed vessel tree with artifact suppression; (b) reconstructed vessel tree without artifact suppression; (c) full-view projection of a slice of the reconstructed vessel trees with artifacts suppression; (d) full-view projection without artifacts suppression. Artifact suppression improves the consistency between different projection views. The vertical axis in (c) and (d) is the projection view.

connectivity. Additionally, all methods in Table 4.1 have *Spe* scores close to 1, indicating that there are few false positives in liver vessel segmentation.

4.4 Discussion

4.4.1 Effect of artifact suppression

Our model generates liver vessel *IPs* based on the intensity of the *top-k MIP*. However, due to memory constraints, our current design limits the model from learning correlations between different projection directions. The optimization in Eq. (4.4) compensates for this shortcoming. This optimization suppresses the reconstruction artifact caused by projection inconsistency.

Table 4.2 tests the added value of this artifact suppression optimization. The *IPs* of the reconstructed vessel tree with artifact suppression have higher *PSNR* [19] and *SSIM* [19], indicating that the optimized *IPs* have a similar appearance to the ground truth. Full-view projections in Fig. 4.5(c)-(d) also show that the vessel *IPs* consistency between projections is improved by the artifact suppression.

4.4.2 Limitations and possible improvements

Model limitations. Our proposed model can handle most cases, where the contrast is within normal ranges, in the liver vessel area. However, segmenting extremely low-contrast vascular regions, as shown in the yellow box in Fig. 4.6 remains challenging for our model.

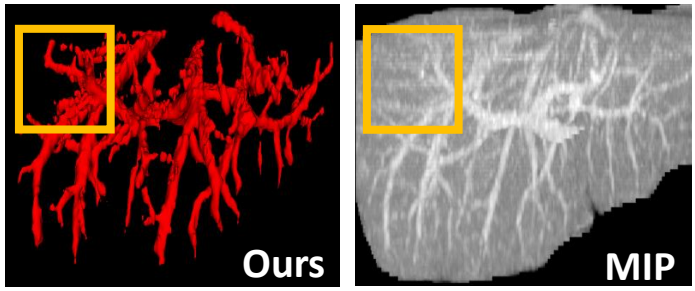

Figure 4.6: **Failure analysis.** Correctly segmenting low-contrast CT images is challenging for our method.

Possible improvements. Strengthening the correlation between projection views and learning a dynamic projection for the *top-k* MIP are the two improvement directions. Currently, each view is processed independently, and thus we cannot test using the standard MIP. In the standard MIP the depth information is collapsed to a single maximum value, and cannot be recovered. A model that considers the correlation between projection views, will make it possible to use of the standard MIP, instead of the *top-k* MIP. But now, the standard MIP, whose depth information is completely collapsed to a single maximum, is not applicable to the current method.

4.5 Conclusion

We propose a principled way of incorporating $3D$ liver-vessel topology in $2D$ diffusion models for liver-vessel segmentation. Accordingly, we draw inspiration from the physics involved in the $3D$ CT reconstruction. Concretely, we propose to condition a latent diffusion model on *top-k maximum intensity projections* of the CT cross-sectional view. Our proposed model achieves competitive results compared to the existing baselines, validating our approach.

References

[1] O. Alirr and A. Rahni. “Survey on liver tumour resection planning system: steps, techniques, and parameters”. In: *Journal of Digital Imaging* 33.2 (2020), pages 304–323.

[2] T. Kitrungrotsakul, X. Han, Y. Iwamoto, et al. “VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation”. In: *Computerized Medical Imaging and Graphics* 75 (2019), pages 74–83.

[3] Z. Gao, Q. Zong, Y. Wang, et al. “Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation”. In: *IEEE Transactions on Medical Imaging* 42.10 (2023), pages 3059–3068.

[4] Q. Huang, J. Sun, H. Ding, et al. “Robust liver vessel extraction using 3D U-Net with variant dice loss function”. In: *Computers in biology and medicine* 101 (2018), pages 153–162.

[5] Q. Yan, B. Wang, W. Zhang, et al. “Attention-guided deep neural network with multi-scale feature fusion for liver vessel segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 25.7 (2020), pages 2629–2642.

[6] Z. Guo, Z. Tan, J. Feng, and J. Zhou. “3D vascular segmentation supervised by 2D annotation of maximum intensity projection”. In: *IEEE Transactions on Medical Imaging* (2024).

[7] Y. Liu, H. Kwak, and I. Oh. “Cerebrovascular segmentation model based on spatial attention-guided 3D inception U-Net with multi-directional MIPs”. In: *Applied Sciences* 12.5 (2022), page 2288.

[8] H. Chen, X. Wang, H. Li, and L. Wang. “3D vessel segmentation with limited guidance of 2D structure-agnostic vessel annotations”. In: *IEEE Journal of Biomedical and Health Informatics* (2024).

[9] S. Napel, M. Marks, G. Rubin, et al. “CT angiography with spiral CT and maximum intensity projection”. In: *Radiology* 185.2 (1992), pages 607–610.

[10] G. Tetteh, V. Efremov, N. Forkert, et al. “Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes”. In: *Frontiers in Neuroscience* 14 (2020), page 592352.

[11] R. Rombach, A. Blattmann, D. Lorenz, et al. “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pages 10684–10695.

[12] L. Soler, A. Hostettler, V. Agnus, et al. “3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database”. In: ().

[13] F. Isensee, P. F. Jaeger, S. A. Kohl, et al. “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pages 203–211.

[14] A. Hatamizadeh, V. Nath, Y. Tang, et al. “Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images”. In: *International MICCAI Brainlesion Workshop*. Sept. 2021, pages 272–284.

[15] J. Wolleb, R. Sandkühler, F. Bieder, et al. “Diffusion models for implicit image segmentation ensembles”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR, Dec. 2022, pages 1336–1348.

[16] J. Wu, R. FU, H. Fang, et al. “MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model”. In: *Medical Imaging with Deep Learning*. 2023.

[17] R. Schofield, L. King, U. Tayal, et al. “Image reconstruction: Part 1–understanding filtered back projection, noise and image acquisition”. In: *Journal of cardiovascular computed tomography* 14.3 (2020), pages 219–225.

[18] S. Shit, J. Paetzold, A. Sekuboyina, et al. “clDice-a novel topology-preserving loss function for tubular structure segmentation”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pages 16560–16569.

[19] A. Hore and D. Ziou. “Image quality metrics: PSNR vs. SSIM”. In: *2010 20th international conference on pattern recognition*. IEEE. 2010, pages 2366–2369.