



3D Arterial Segmentation via Single 2D Projections and Depth Supervision in Contrast-Enhanced CT Images

Alina F. Dima^{1,2(✉)}, Veronika A. Zimmer^{1,2}, Martin J. Menten^{1,4},
Hongwei Bran Li^{1,3}, Markus Graf², Tristan Lemke², Philipp Raffler²,
Robert Graf^{1,2}, Jan S. Kirschke², Rickmer Braren², and Daniel Rueckert^{1,2,4}

¹ School of Computation, Information and Technology,
Technical University of Munich, Munich, Germany
alina.dima@tum.de

² School of Medicine, Klinikum Rechts der Isar,
Technical University of Munich, Munich, Germany

³ Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

⁴ Department of Computing, Imperial College London, London, UK

Abstract. Automated segmentation of the blood vessels in 3D volumes is an essential step for the quantitative diagnosis and treatment of many vascular diseases. 3D vessel segmentation is being actively investigated in existing works, mostly in deep learning approaches. However, training 3D deep networks requires large amounts of manual 3D annotations from experts, which are laborious to obtain. This is especially the case for 3D vessel segmentation, as vessels are sparse yet spread out over many slices and disconnected when visualized in 2D slices. In this work, we propose a novel method to segment the 3D peripancreatic arteries **solely from one annotated 2D projection per training image** with depth supervision. We perform extensive experiments on the segmentation of peripancreatic arteries on 3D contrast-enhanced CT images and demonstrate how well we capture the rich depth information from 2D projections. We demonstrate that by annotating a single, randomly chosen projection for each training sample, we obtain comparable performance to annotating multiple 2D projections, thereby reducing the annotation effort. Furthermore, by mapping the 2D labels to the 3D space using depth information and incorporating this into training, we almost close the performance gap between 3D supervision and 2D supervision. Our code is available at: <https://github.com/alinafdima/3Dseg-mip-depth>.

Keywords: vessel segmentation · 3D segmentation · weakly supervised segmentation · curvilinear structures · 2D projections

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_14.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 141–151, 2023.
https://doi.org/10.1007/978-3-031-43907-0_14

1 Introduction

Automated segmentation of blood vessels in 3D medical images is a crucial step for the diagnosis and treatment of many diseases, where the segmentation can aid in visualization, help with surgery planning, be used to compute biomarkers, and further downstream tasks. Automatic vessel segmentation has been extensively studied, both using classical computer vision algorithms [16] such as vesselness filters [8], or more recently with deep learning [3, 5, 6, 11, 19, 21], where state-of-the-art performance has been achieved for various vessel structures. Supervised deep learning typically requires large, well-curated training sets, which are often laborious to obtain. This is especially the case for 3D vessel segmentation.

Manually delineating 3D vessels typically involves visualizing and annotating a 3D volume through a sequence of 2D cross-sectional slices, which is not a good medium for visualizing 3D vessels. This is because often only the cross-section of a vessel is visible in a 2D slice. In order to segment a vessel, the annotator has to track the cross-section of that vessel through several adjacent slices, which is especially tedious for curved or branching vessel trees. Projecting 3D vessels to a 2D plane allows for the entire vessel tree to be visible within a single 2D image, providing a more robust representation and potentially alleviating the burden of manual annotation. Kozinski et al. [13] propose to annotate up to three maximum intensity projections (MIP) for the task of centerline segmentation [13], obtaining results comparable to full 3D supervision. Compared to centerline segmentation, where the vessel diameter is disregarded, training a 3D vessel segmentation model from 2D annotations poses additional segmentation-specific challenges, as 2D projections only capture the outline of the vessels, providing no information about their interior. Furthermore, the axes of projection are crucial for the model’s success, given the sparsity of information in 2D annotations.

To achieve 3D vessel segmentation with only 2D supervision from projections, we first investigate which viewpoints to annotate in order to maximize segmentation performance. We show that it is feasible to segment the full extent of vessels in 3D images with high accuracy by annotating only a single randomly-selected 2D projection per training image. This approach substantially reduces the annotation effort, even compared to works training only on 2D projections. Secondly, by mapping the 2D annotations to the 3D space using the depth of the MIPs, we obtain a partially segmented 3D volume that can be used as an additional supervision signal. We demonstrate the utility of our method on the challenging task of peripancreatic arterial segmentation on contrast-enhanced arterial-phase computed tomography (CT) images, which feature large variance in vessel diameter. Our contribution to 3D vessel segmentation is three-fold:

- Our work shows that highly accurate automatic segmentation of 3D vessels can be learned by annotating single MIPs.
- Based on extensive experimental results, we determine that the best annotation strategy is to label randomly selected viewpoints, while also substantially reducing the annotation cost.

- By incorporating additional depth information obtained from 2D annotations at no extra cost to the annotator, we almost close the gap between 3D supervision and 2D supervision.

2 Related Work

Learning from Weak Annotations. Weak annotations have been used in deep learning segmentation to reduce the annotation effort through cheaper, less accurate, or sparser labeling [20]. Bai et al. [1] learn to perform aortic image segmentation by sparsely annotating only a subset of the input slices. Multiple instance learning approaches bin pixels together by only providing labels at the bin level. Jia et al. [12] use this approach to segment cancer on histopathology images successfully. Annotating 2D projections for 3D data is another approach to using weak segmentation labels, which has garnered popularity recently in the medical domain. Bayat et al. [2] propose to learn the spine posture from 2D radiographs, while Zhou et al. [22] use multi-planar MIPs for multi-organ segmentation of the abdomen. Kozinski et al. [13] propose to segment vessel centerlines using as few as 2-3 annotated MIPs. Chen et al. [4] train a vessel segmentation model from unsupervised 2D labels transferred from a publicly available dataset, however, there is still a gap to be closed between unsupervised and supervised model performance. Our work uses weak annotations in the form of annotations of 2D MIPs for the task of peripancreatic vessel segmentation, where we attempt to reduce the annotation cost to a minimum by only annotating a single projection per training input without sacrificing performance.

Incorporating Depth Information. Depth is one of the properties of the 3D world. Loss of depth information occurs whenever 3D data is projected onto a lower dimensional space. In natural images, depth loss is inherent through image acquisition, therefore attempts to recover or model depth have been employed for 3D natural data. For instance, Fu et al. [9] use neural implicit fields to semantically segment images by transferring labels from 3D primitives to 2D images. Lawin et al. [14] propose to segment 3D point clouds by projecting them onto 2D and training a 2D segmentation network. At inference time, the predicted 2D segmentation labels are remapped back to the original 3D space using the depth information. In the medical domain, depth information has been used in volume rendering techniques [7] to aid with visualization, but it has so far not been employed when working with 2D projections of 3D volumes to recover information loss. We propose to do the conceptually opposite approach from Lawin et al. [14], by projecting 3D volumes onto 2D to facilitate and reduce annotation. We use depth information to map the 2D annotations to the original 3D space at annotation time and generate partial 3D segmentation volumes, which we incorporate in training as an additional loss term.

3 Methodology

Overview. The maximum intensity projection (MIP) of a 3D volume $I \in \mathbb{R}^{N_x \times N_y \times N_z}$ is defined as the highest intensity along a given axis:

$$\text{mip}(x, y) = \max_z I(x, y, z) \in \mathbb{R}^{N_x \times N_y}. \quad (1)$$

For simplicity, we only describe MIPs along the z-axis, but they can be performed on any image axis.

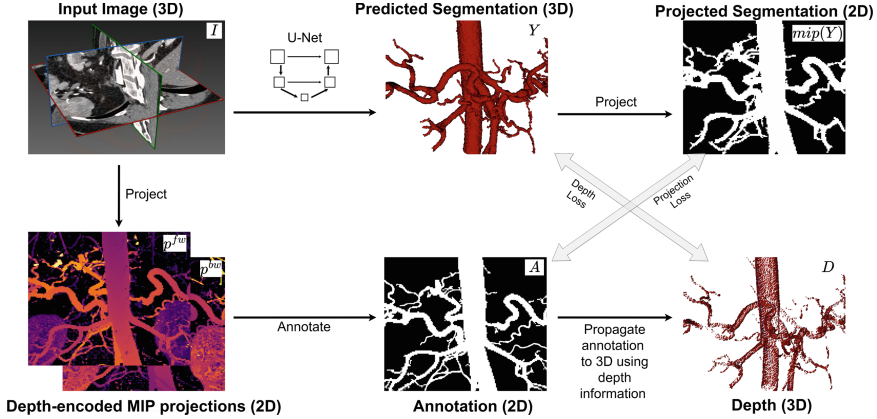


Fig. 1. Method overview. We train a 3D network to segment vessels from 2D annotations. Given an input image I , depth-encoded MIPs p^{fw} , p^{bw} are generated by projecting the input image to 2D. 2D binary labels A are generated by annotating one 2D projection per image. The 2D annotation is mapped to the 3D space using the depth information, resulting in a partially labeled 3D volume D . During training, both 2D annotations and 3D depth maps are used as supervision signals in a combined loss, which uses both predicted 3D segmentation Y and its 2D projection $\text{mip}(Y)$.

Exploiting the fact that arteries are hyperintense in arterial phase CTs, we propose to annotate MIPs of the input volume for binary segmentation. The hyperintensities of the arteries ensures their visibility in the MIP, while additional processing removes most occluding nearby tissue (Sect. 4).

Given a binary 2D annotation of a MIP $A \in \{0, 1\}^{N_x \times N_y}$, we map the foreground pixels in A to the original 3D image space. This is achieved by using the first and last z coordinates where the maximum intensity is observed along any projection ray. Owing to the fact that the vessels in the abdominal cavity are relatively sparse in 2D projections and most of the occluding tissue is removed in postprocessing, this step results in a fairly complete surface of the vessel tree. Furthermore, we can partially fill this surface volume, resulting in a 3D depth map D , which is a partial segmentation of the vessel tree. We use the 2D annotations as well as the depth map to train a 3D segmentation network in a weakly supervised manner.

An overview of our method is presented in Fig. 1. In the following, we describe these components and how they are combined to train a 3D segmentation network in more detail.

Depth Information. We can view MIP as capturing the intensity of the brightest pixel along each ray $r_{xy} \in \mathbb{R}^{N_z}$, where $r_{xy}(z) = I(x, y, z)$. Along each projection ray, we denote the first and last z coordinates which have the same intensity as the MIP to be the forward depth $z^{fw} = \arg \max_z I(x, y, z)$ and backward depth $z^{bw} = \arg \min_z I(x, y, z)$. This information can be utilized for the following: (1) enhancing the MIP visualization, or (2) providing a way to map pixels from the 2D MIP back to the 3D space (depth map). The reason why the maximum intensity is achieved multiple times along a ray is because our images are clipped, which removes a lot of the intensity fluctuations.

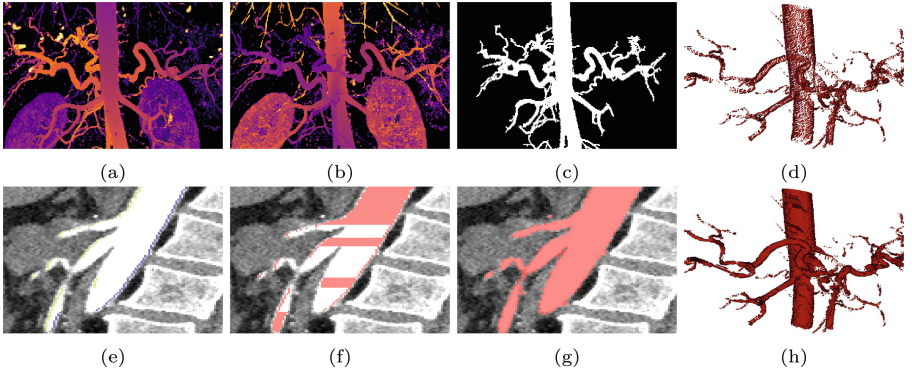


Fig. 2. Example depth-enhanced MIP using (a) forward depth z^{fw} and (b) backward depth z^{bw} visualized in color; (c) binary 2D annotation; a slice view from a 3D volume illustrating: (e) the forward – in green – and backward depth – in blue –, (f) the depth map, (g) 3D ground truth; volume rendering of (h) the depth map and (d) the depth map with only forward and backward depth pixels. The input images are contrast-enhanced.(Color figure online)

Depth-Enhanced MIP. We encode depth information into the MIPs by combining the MIP with the forward and backward depth respectively, in order to achieve better depth perception during annotation: $p^{fw} = \sqrt{mip} \cdot z^{fw}$ defines the forward projection, while $p^{bw} = \sqrt{mip} \cdot z^{bw}$ defines the backward projection. Figure 2 showcases (a) forward and (b) backward depth encoded MIPs.

Depth Map Generation. Foreground pixels from the 2D annotations are mapped to the 3D space by combining a 2D annotation with the forward and backward depth, resulting in a 3D partial vessel segmentation:

-
1. Create an empty 3D volume $D \in \mathbb{R}^{N_x \times N_y \times N_z}$.
 2. For each foreground pixel in the annotation A at location (x, y) , we label (x, y, z^{fw}) and (x, y, z^{bw}) as foreground pixels in D .
 3. If the fluctuation in intensity between z^{fw} and z^{bw} along the ray r_{xy} is below a certain threshold in the source image I , the intermediate pixels are also labeled as foreground in D .
-

Training Loss. We train a 3D segmentation network to predict 3D binary vessel segmentation given a 3D input volume using 2D annotations. Our training set $\mathcal{D}_{tr}(I, A, D)$ consists of 3D volumes I paired with 2D annotations A and their corresponding 3D depth maps D . Given the 3D network output $Y = \theta(I)$, we minimize the following loss during training:

$$\mathcal{L}(Y) = \alpha \cdot \mathcal{CE}(A, \text{mip}(Y)) + (1 - \alpha) \cdot \mathcal{CE}(D, Y) \cdot D, \quad (2)$$

where $\alpha \in [0, 1]$. Our final loss is a convex combination between: **(a)** the cross-entropy(\mathcal{CE}) of the network output projected to 2D and the 2D annotation, as well as **(b)** the cross-entropy between the network output and the depth map, but only applied to positive pixels in the depth map. Notably, the 2D loss constrains the shape of the vessels, while the depth loss promotes the segmentation of the vessel interior.

4 Experimental Design

Dataset. We use an in-house dataset of contrast-enhanced abdominal computed tomography images (CTs) in the arterial phase to segment the peripancreatic arteries [6]. The cohort consists of 141 patients with pancreatic ductal adenocarcinoma, of an equal ratio of male to female patients. Given a 3D arterial CT of the abdominal area, we automatically extract the vertebrae [15, 18] and semi-automatically extract the ribs, which have similar intensities as arteries in arterial CTs and would otherwise occlude the vessels. In order to remove as much of the cluttering surrounding tissue and increase the visibility of the vessels in the projections, the input is windowed so that the vessels appear hyperintense. Details of the exact preprocessing steps can be found in Table 2 of the supplementary material. The dataset contains binary 3D annotations of the peripancreatic arteries carried out by two radiologists, each having annotated half of the dataset. The 2D annotations we use in our experiments are projections of these 3D annotations. For more information about the dataset, see [6].

Image Augmentation and Transformation. As the annotations lie on a 2D plane, 3D spatial augmentation cannot be used due to the information sparsity in the ground truth. Instead, we apply an invertible transformation \mathcal{T} to the input volume and apply the inverse transformation \mathcal{T}^{-1} to the network output before applying the loss, such that the ground truth need not be altered. A detailed description of the augmentations and transformations used can be found in Table 1 in the supplementary material.

Training and Evaluation. We use a 3D U-Net [17] with four layers as our backbone, together with Xavier initialization [10]. A diagram of the network architecture can be found in Fig. 2 in the supplementary material. The loss weight α is tuned at 0.5, as this empirically yields the best performance. Our experiments are averaged over 5-fold cross-validation with 80 train samples, 20 validation samples, and a fixed test set of 41 samples. The network initialization is different for each fold but kept consistent across different experiments run on the same fold. This way, both data variance and initialization variance are accounted for through cross-validation. To measure the performance of our models, we use the Dice score, precision, recall, and mean surface distance (MSD). We also compute the skeleton recall as the percentage of the ground truth skeleton pixels which are present in the prediction.

Table 1. Viewpoint ablation. We compare models trained on single random viewpoints (VPs) with (+D) or without (−D) depth against fixed viewpoint baselines without depth and full 3D supervision. We distinguish between model selection based on 2D annotations vs. 3D annotations on the validation set. The best-performing models for each model selection (2D *vs.* 3D) are highlighted in bold.

Experiment	Model Selection	Dice \uparrow	Precision \uparrow	Recall \uparrow	Skeleton Recall \uparrow	MSD \downarrow
3D	3D	92.18 \pm 0.35	93.86 \pm 0.81	90.64 \pm 0.64	76.04 \pm 4.51	1.15 \pm 0.11
fixed 3VP	3D	92.02 \pm 0.52	93.05 \pm 0.61	91.13 \pm 0.79	78.61 \pm 1.52	1.13 \pm 0.11
fixed 2VP	3D	91.29 \pm 0.78	91.46 \pm 2.13	91.37 \pm 1.45	78.51 \pm 2.78	1.13 \pm 0.09
fixed 3VP	2D	90.78 \pm 1.30	90.66 \pm 1.30	91.18 \pm 3.08	81.77 \pm 2.13	1.16 \pm 0.13
fixed 2VP	2D	90.22 \pm 1.19	88.16 \pm 2.86	92.74 \pm 1.63	82.18 \pm 2.47	1.14 \pm 0.09
fixed 1VP	2D	60.76 \pm 24.14	50.47 \pm 23.21	92.52 \pm 3.09	81.19 \pm 2.39	2.96 \pm 3.15
random 1VP−D	2D	91.29 \pm 0.81	91.42 \pm 0.92	91.45 \pm 1.00	80.16 \pm 2.35	1.13 \pm 0.04
random 1VP+D	2D	91.69 \pm 0.48	90.77 \pm 1.76	92.79 \pm 0.95	81.27 \pm 2.02	1.15 \pm 0.11

5 Results

The Effectiveness of 2D Projections and Depth Supervision. We compare training using single random viewpoints with and without depth information against baselines that use more supervision. Models trained on full 3D ground truth represent the upper bound baseline, which is very expensive to annotate. We implement [13] as a baseline on our dataset, training on up to 3 fixed orthogonal projections. We distinguish between models selected according to the 2D performance on the validation set (2D) which is a fair baseline, and models selected according to the 3D performance on the validation set (3D), which is an unfair baseline as it requires 3D annotations on the validation set. With the exception of the single fixed viewpoint baselines where the models have the tendency to diverge towards over- or segmentation, we perform binary hole-filling on the output of all of our other models, as producing hollow objects is a common under-segmentation issue.

In Table 1 we compare our method against the 3D baseline, as well as baselines trained on multiple viewpoints. We see that by using **depth information**

paired with training using a single random viewpoint per sample performs almost at the level of models trained on 3D labels, at a very small fraction of the annotation cost. The depth information also reduces model variance compared to the same setup without depth information. Even without depth information, training the model on single **randomly** chosen viewpoints offers a robust training signal that the Dice score is on par with training on 2 fixed viewpoints under ideal model selection at only half the annotation cost. Randomly selecting viewpoints for training acts as powerful data augmentation, which is why we are able to obtain performance comparable to using more fixed viewpoints. Under ideal 3D-based model selection, three views would come even closer to full 3D performance; however, with realistic 2D-based model selection, fixed viewpoints are more prone to diverge. This occurs because sometimes 2D-based model selection favors divergent models which only segment hollow objects, which cannot be fixed in postprocessing. Single fixed viewpoints contain so little information on their own that models trained on such input fail to learn how to segment the vessels and generally converge to over-segmenting in the blind spots in the projections. We conclude that using random viewpoints is not only helpful in reducing annotation cost but also decreases model variance.

In terms of other metrics, randomly chosen projection viewpoints with and without depth improve both recall and skeleton recall even compared to fully 3D annotations, while generally reducing precision. We theorize that this is because the dataset itself contains noisy annotations and fully supervised models better overfit to the type of data annotation, whereas our models converge to following the contrast and segmenting more vessels, which are sometimes wrongfully labeled as background in the ground truth. MSD are not very telling in our dataset due to the noisy annotations and the nature of vessels, as an under- or over-segmented vessel branch can quickly translate into a large surface distance.

The Effect of Dataset Size. We vary the size of the training set from $|\mathcal{D}_{tr}| = 80$ to as little as $|\mathcal{D}_{tr}| = 10$ samples, while keeping the size of the validation and test sets constant, and train models on single random viewpoints.

In Table 2, we compare single random projections trained with and without depth information at varying dataset sizes to illustrate the usefulness of the depth information with different amounts of training data. Our depth loss offers consistent improvement across multiple dataset sizes and reduces the overall performance variance. The performance boost is noticeable across the board, the only exception being precision. The smaller the dataset size is, the greater the performance boost from the depth. We perform a Wilcoxon rank-sum statistical test comparing the individual sample predictions of the models trained at various dataset sizes with single random orthogonal viewpoints with or without depth information, obtaining a statistically significant (p-value of < 0.0001). We conclude that the depth information complements the segmentation effectively.

Table 2. Dataset size ablation. We vary the training dataset size $|\mathcal{D}_{tr}|$ and compare models trained on single random viewpoints, with or without depth. Best performing models in each setting are highlighted.

$ \mathcal{D}_{tr} $	Depth	Dice \uparrow	Precision \uparrow	Recall \uparrow	Skeleton Recall \uparrow	MSD \downarrow
10	−D	86.03 ± 2.94	88.23 ± 2.58	84.81 ± 6.42	78.25 ± 2.20	1.92 ± 0.55
10	+D	89.06 ± 1.20	88.55 ± 1.73	89.91 ± 1.29	78.95 ± 3.62	1.80 ± 0.28
20	−D	88.22 ± 3.89	90.26 ± 1.64	86.74 ± 6.56	80.78 ± 1.66	1.44 ± 0.20
20	+D	90.51 ± 0.38	89.84 ± 0.90	91.50 ± 1.23	80.00 ± 1.95	1.33 ± 0.16
40	−D	88.07 ± 2.34	89.09 ± 2.01	87.62 ± 4.43	78.38 ± 2.39	1.38 ± 0.10
40	+D	90.21 ± 0.89	89.08 ± 2.89	91.82 ± 2.11	79.16 ± 2.36	1.24 ± 0.14
80	−D	91.29 ± 0.81	91.42 ± 0.92	91.45 ± 1.00	80.16 ± 2.35	1.13 ± 0.04
80	+D	91.69 ± 0.48	90.77 ± 1.76	92.79 ± 0.95	81.27 ± 2.02	1.15 ± 0.11

6 Conclusion

In this work, we present an approach for 3D segmentation of peripancreatic arteries using very sparse 2D annotations. Using a labeled dataset consisting of single, randomly selected, orthogonal 2D annotations for each training sample and additional depth information obtained at no extra cost, we obtain accuracy almost on par with fully supervised models trained on 3D data at a mere fraction of the annotation cost. Limitations of our work are that the depth information relies on the assumption that the vessels exhibit minimal intensity fluctuations within local neighborhoods, which might not hold on other datasets, where more sophisticated ray-tracing methods would be more effective in locating the front and back of projected objects. Furthermore, careful preprocessing is performed to eliminate occluders, which would limit its transferability to datasets with many occluding objects of similar intensities. Further investigation is needed to quantify how manual 2D annotations compare to our 3D-derived annotations, where we expect occluders to affect the annotation process.

References

1. Bai, W., et al.: Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 586–594. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_67
2. Bayat, A.: Inferring the 3D standing spine posture from 2D radiographs. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 775–784. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_75
3. Chen, C., Chuah, J.H., Ali, R., Wang, Y.: Retinal vessel segmentation using deep learning: a review. IEEE Access **9**, 111985–112004 (2021)

4. Chen, H., Wang, X., Wang, L.: 3D vessel segmentation with limited guidance of 2D structure-agnostic vessel annotations. arXiv preprint [arXiv:2302.03299](https://arxiv.org/abs/2302.03299) (2023)
5. Ciecholewski, M., Kassjański, M.: Computational methods for liver vessel segmentation in medical imaging: A review. *Sensors* **21**(6), 2027 (2021)
6. Dima, A., et al.: Segmentation of peripancreatic arteries in multispectral computed tomography imaging. In: Lian, C., Cao, X., Rekik, I., Xu, X., Yan, P. (eds.) *MLMI 2021*. LNCS, vol. 12966, pp. 596–605. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87589-3_61
7. Drebin, R.A., Carpenter, L., Hanrahan, P.: Volume rendering. *ACM Siggraph Comput. Graphics* **22**(4), 65–74 (1988)
8. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). <https://doi.org/10.1007/BFb0056195>
9. Fu, X., et al.: Panoptic NeRF: 3D-to-2D label transfer for panoptic urban scene segmentation. In: *International Conference on 3D Vision, 3DV 2022, Prague, Czech Republic, 12–16 September 2022*, pp. 1–11. IEEE (2022)
10. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034 (2015)
11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)
12. Jia, Z., Huang, X., Eric, I., Chang, C., Xu, Y.: Constrained deep weak supervision for histopathology image segmentation. *IEEE Trans. Med. Imaging* **36**(11), 2376–2388 (2017)
13. Koziński, M., Mosinska, A., Salzmann, M., Fua, P.: Tracing in 2D to reduce the annotation effort for 3D deep delineation of linear structures. *Med. Image Anal.* **60**, 101590 (2020)
14. Lawin, F.J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F.S., Felsberg, M.: Deep projective 3D semantic segmentation. In: Felsberg, M., Heyden, A., Krüger, N. (eds.) *CAIP 2017*. LNCS, vol. 10424, pp. 95–107. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64689-3_8
15. Löffler, M.T., et al.: A vertebral segmentation dataset with fracture grading. *Radiol. Artif. Intell.* **2**(4), e190138 (2020)
16. Luboz, V., et al.: A segmentation and reconstruction technique for 3D vascular structures. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3749, pp. 43–50. Springer, Heidelberg (2005). https://doi.org/10.1007/11566465_6
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Sekuboyina, A., et al.: VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images. *Med. Image Anal.* **73**, 102166 (2021)
19. Shi, F., et al.: Intracranial vessel wall segmentation using convolutional neural networks. *IEEE Trans. Biomed. Eng.* **66**(10), 2840–2847 (2019)
20. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **63**, 101693 (2020)

21. Tetteh, G., et al.: Deepvesselnet: vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *Front. Neurosci.*, 1285 (2020)
22. Zhou, Y., et al.: Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 121–140. IEEE (2019)