This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Enhancing Single-Slice Segmentation with 3D-to-2D Unpaired Scan Distillation

Xin Yu[1], Qi Yang[1], Han Liu[1], Ho Hin Lee[1], Yucheng Tang[2,3], Lucas W. Remedios[1], Michael E. Kim[1], Rendong Zhang[1], Shunxing Bao[2], Yuankai Huo[1,2], Ann Zenobia Moore[4], Luigi Ferrucci[4], and Bennett A. Landman[1,2]

[1] Computer Science, Vanderbilt University, Nashville, TN, USA
[2] Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA
[3] Nivida Corporation
[4] National Institute on Aging, Baltimore, MD, USA

Abstract. 2D single-slice abdominal computed tomography (CT) enables the assessment of body habitus and organ health with low radiation exposure. However, single-slice data necessitates the use of 2D networks for segmentation, but these networks often struggle to capture contextual information effectively. Consequently, even when trained on identical datasets, 3D networks typically achieve superior segmentation results. In this work, we propose a novel 3D-to-2D distillation framework, leveraging pre-trained 3D models to enhance 2D single-slice segmentation. Specifically, we extract the prediction distribution centroid from the 3D representations, to guide the 2D student by learning intra- and inter-class correlation. Unlike traditional knowledge distillation methods that require the same data input, our approach employs unpaired 3D CT scans with any contrast to guide the 2D student model. Experiments conducted on 707 participants from the single-slice Baltimore Longitudinal Study of Aging (BLSA) dataset demonstrate that state-of-the-art 2D multi-organ segmentation methods can benefit from the 3D teacher model, achieving enhanced performance in single-slice multi-organ segmentation. Notably, our approach demonstrates considerable efficacy in low-data regimes, outperforming the model trained with all available training subjects even when utilizing only 200 training subjects. Thus, this work underscores the potential to alleviate manual annotation burdens.

Keywords: 3D-to-2D · Unpaired data distillation · Single-slice.

1 Introduction

Abdominal computed tomography (CT) provides detailed tissue maps with high resolution that enables quantification of body composition. To better assess the internal structures and organ health, many deep learning-based multi-organ segmentation methods have been proposed and demonstrated their effectiveness [8, 16, 15, 7, 1]. These methods can typically be categorized into two types: those trained with 2D slices (2D models) [21, 1, 20, 7, 22, 23] and those trained with 3D

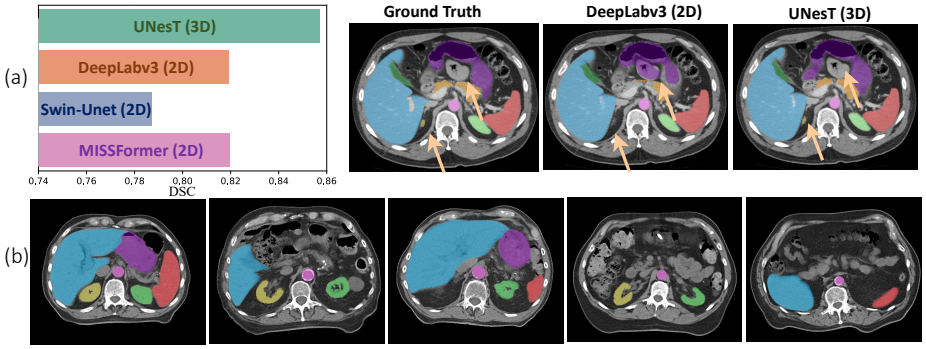

Fig. 1. (a) 3D and 2D model performance comparison. All the models are trained and tested on the MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge (BTCV) dataset. UNesT [28], a 3D model, has superior performance (DSC) compared to other 2D models. (b) Example abdominal slices from the single-slice CT dataset with manual annotation on right/left kidneys, liver, spleen, stomach and aorta. The scans are in a wide range of vertebral levels.

volumes (3D models) [8, 16, 15, 28, 24]. 2D models require lower machine memory costs while 3D methods benefit from more contextual information and usually lead to better segmentation results [10, 9], even when trained on identical data, as depicted in Fig. 1(a). However, in routine monitoring and screening, 2D Non-contrast axial abdominal single-slice CT is preferred over 3D CT to minimize unnecessary radiation exposure [27], necessitating the use of 2D networks. As the example single-slice shown in Fig. 1(b), the heterogeneity in levels of image acquisition across subjects, along with the absence of three-dimensional context and contrast, makes the segmentation of single-slice with 2D models a challenge.

We aim to transfer the learned contextual knowledge from a pretrained 3D model to a 2D model to improve the performance on multi-organ segmentation. Knowledge distillation has been widely explored in the medical imaging domain. Most of the work has been focused on model compression by transferring knowledge from complex models to light-weight ones [30, 4, 13]. Some other works focus on using knowledge distillation to solve missing modality problem by distilling multi-modality knowledge to a single domain [18, 17]. However, these works typically require the same or paired inputs for the teacher and student models since they require logit match of student and teacher output, which is not applicable in the single-slice dataset where paired 3D volumetric data is not available. Qou *et al.* [3] proposed an unpaired knowledge distillation framework for multi-modal segmentation tasks. However, in this work, the modality-specific knowledge is learned on a batch basis rather than a dataset basis, and the method is designed to handle different inputs of the same dimension. In the domain of natural image processing, many works have been proposed for 3D-to-2D knowledge transfer for indoor scene segmentation [12, 19, 5]. However, it is non-trivial to adapt these

methods to our case due to the differences in image format as well as the requirement of paired 3D2D input, which is not applicable in our scenario.

In this work, we aim to address two key challenges. Firstly, we aim to enable the 3D model to learn an effective embedding without requiring point-to-point matching when passing it to the 2D model. Secondly, we aim to facilitate the 2D model in learning from the embedding extracted from unpaired 3D data. To address the first challenge, we posit to extract the distribution centroid of each class using the 3D model and encourage the 2D centroid to be close to the 3D centroid during distillation. Inspired by [11, 29], we compute the class-wise feature centroids (prototype) over the whole 3D dataset as the class centroid for the 3D model output and compute the sample-wise prototype on-the-fly for 2D samples. To address the second challenge, we extend the DIST loss introduced in [6] to maximize the linear inter- and intra-class correlation between 3D and 2D models instead of point-wise exact match used in conventional knowledge distillation. To this end, we design a novel unpaired 3D-to-2D knowledge distillation framework to improve 2D single-slice segmentation performance. We conduct extensive experiments on three representative 2D segmentation methods on 707 subjects from the single-slice Baltimore Longitudinal Study of Aging (BLSA) dataset with 3D model trained with CT data of different contrasts. The results demonstrate that our 3D-to-2D distillation framework can help improve 2D model segmentation.

Our contributions are three-fold: (1) to the best of our knowledge, we propose the first 3D-to-2D distillation framework in the medical imaging domain that leverages pretrained 3D models to improve 2D model segmentation; (2) our designed approach learns the 3D prototype during 3D pretraining, eliminating the need for 3D input during the inference phase; (3) we demonstrate that our proposed method consistently learns from the unpaired 3D prototype and enhances single-slice segmentation.

2 Method

The proposed 3D-to-2D distillation framework consists of a pretrained 3D segmentation network (teacher) and an active 2D segmentation network (student). The overall pipeline of the framework is shown in Fig. 2. Before the distillation, both the 3D and 2D models are trained using training data before distillation. The 2D model is pretrained to obtain an initialization, whereas the 3D model is trained to acquire 3D features centroids (prototype).

2.1 Prototype Extraction from 3D Model

In our scenario, the aim is to enhance the segmentation performance of single-slice data using existing 3D datasets. Since the input 3D data and 2D data are unpaired, common pixel-wise distillation methods are inapplicable. We propose to align the 2D student model output to the distribution of the 3D teacher model output to preserve rich contextual information learned by 3D networks.

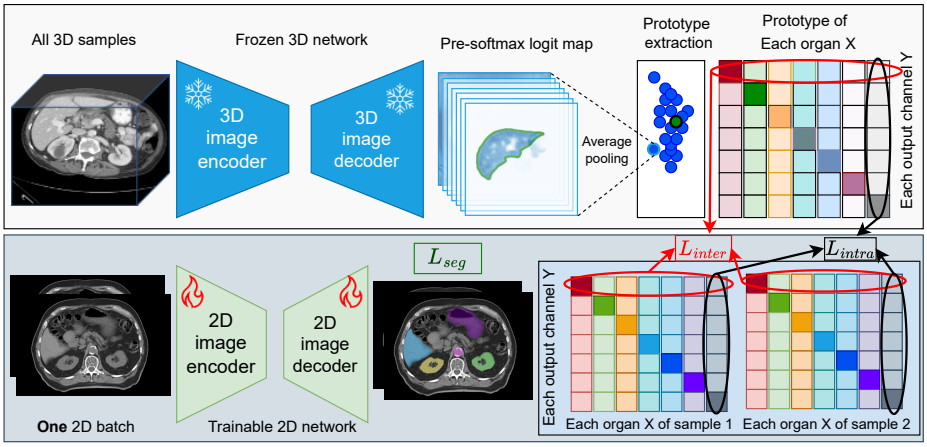

Fig. 2. The overview of the proposed method. The whole pipeline can be divided into a frozen 3D model (teacher) and a trainable 2D model (student). 3D model is pretrained to compute dataset class-wise feature centroid (prototype). 2D prototype is computed on-the-fly during distillation, and encouraged to be close to the 3D prototype via optimizing inter- and intra-class correlation. Each class prototype has the maximum response in the channel corresponding to its class index.

Inspired by [11, 29], we propose computing the class-wise feature centroids vector (prototype) as the class distribution of the 3D model output. The underlying assumption for this approach is that the computed centroid is positioned closer to the true centroid of the underlying cluster [29]. For each slice s in the 3D volume, we take the final feature map of the 3D network $\mathcal{X} \in \mathbb{R}^{H \times W \times CH}$ to calculate the prototype of each class, where H, W and CH represents height, width, and channel, respectively. The centroid of each organ k in each slice z_k is obtained by computing the average feature representation within the corresponding organ mask:

$$z_k = \frac{\sum_i \mathcal{X}^i * \mathbb{1}\left(y(i,k) == 1\right)}{\sum_i \mathbb{1}\left(y(i,k) == 1\right)}, \tag{1}$$

where $\mathbb{1}$ is the indicator function, and i, y represent the ith pixel in the given slice and ground truth, respectively.

As the single-slice data primarily covers limited lumbar regions, we apply the body part regression method proposed in [14] to crop the 3D volume and compute the prototype based on slices within this region. The final prototype of all classes has the size of $C \times CH \times 1 \times 1$, where C and CH represent class and channel, respectively.

2.2 3D-to-2D distillation

Given the prototype of 3D output, we calculate the 2D prototype of 2D slices input on-the-fly during distillation. To distill the learned prototype from the 3D

model to the 2D student, we extend the DIST loss introduced in [6]. This loss function aims to preserve both inter-class and intra-class correlation between the student and teacher models. This is in contrast to conventional knowledge distillation methods that aim to align the student's outputs with the teacher's output on a point-wise basis, which is not applicable in our unpaired cases. In our case, each class is represented by a vector as opposed to a value in the original form. We encourage the class-wise prototype from the student X^{s} to be close to the teachers' X^{t}, as shown in Fig. 2 by optimizing the inter-class loss, written as:

$$\mathcal{L}_{\text{inter}} = \frac{1}{B}\sum_{i=1}^{B} d\left(\boldsymbol{X}_i^{\mathrm{s}}, \boldsymbol{X}^{\mathrm{t}}\right), \tag{2}$$

where B represents batch size, $d(x, y) = 1 - \rho(x, y)$, and $\rho(x, y)$ represents the output of Pearson's correlation coefficient.

Since we only have one teacher prototype per dataset rather than a batch, instead of calculating the intra-class loss based on the batch as done in the original form, we compute the intra-class loss based on the channel vector of each class, as shown in Fig. 2. This is because we use the final feature map to calculate the class prototype; we believe that the channel vector contains responses from other classes regarding this class. We extend the loss to encourage the student channel Y^s to be close to the teacher channel distribution Y^t by optimizing the intra-class loss written as:

$$\mathcal{L}_{\text{intra}} = \frac{1}{B}\sum_{i=1}^{B} d\left(\boldsymbol{Y}_i^{\mathrm{s}}, \boldsymbol{Y}^{\mathrm{t}}\right) \tag{3}$$

The total extended DIST loss with inter- and intra-loss can be written as:

$$\mathcal{L}_{\text{DIST}} = \mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{intra}} \tag{4}$$

Together with the original segmentation loss, the final loss function for the 3D-to-2D distillation is written as:

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \beta\mathcal{L}_{\text{DIST}} \tag{5}$$

where β is a weighting factor to balance different losses.

The 3D contextual information is implicitly learned during training and no 3D prototype is required during inference.

3 Experiments and Results

3.1 Dataset

The 3D teacher models are trained with 3D volumetric CT datasets in both the Non-contrast phase and the Portal Venous phase, and the 2D student models are trained with BLSA Non-contrast single-slice dataset.

3D dataset The 3D Non-contrast dataset is an in-house 3D dataset collected on 49 splenomegaly subjects with expert-refined annotations for the major organs. We use 30 Portal Venous CT volumes from the MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge dataset (BTCV) to evaluate the case when 3D training data is of different contrast than the single-slice dataset.

Single-slice dataset We use 796 single slices from 707 subjects to train and evaluate the 2D models. The dataset is split into training, validation, and test sets with 295, 117, and 295 different subjects, respectively. All slices are Non-contrast axial slices in the abdominal region containing at least one organ. Due to the lack of context and contrast in the non-contrast single-slice dataset, only major organs including the spleen, right/left kidneys, liver, stomach, and aorta are annotated.

3.2 Implementation Details

We use UNesT [28] as the 3D teacher model, as it proved effective on learning 3D medical representation. We use 2D representative models for both natural image processing and multi-organ segmentation DeepLabv3 [2], Swin-Unet [1], and MISSFormer [7] as our student models. The teacher model is pretrained with the 3D dataset, and the student models are pretrained with the single-slice dataset following the multi-organ segmentation training strategy mentioned in the original paper. For DeepLabv3 however, we adopt the settings outlined in [26, 25], which are more suitable for multi-organ segmentation. Once the 3D model is trained, we store the feature maps from the final layer of the model to compute the prototype offline. After student model has converged, we select the best model on the validation set as the initialization model for distillation, ensuring that performance gains are solely attributed to distillation rather than continued training. During the 3D-2D distillation process, we set the value of β in Equation 5 to 0.5 to balance its contribution with the segmentation loss. The distillation process is trained for 100 epochs.

3.3 Results

We present the quantitative results in Table 1 and Fig. 3, and qualitative segmentation in Fig. 4. We use Dice Similarity Coefficient (DSC) as the metric to evaluate model performance. Comparing results before and after 3D-to-2D distillation, the results after distillation consistently outperform the results without distillation significantly with $p < 0.05$ under the Wilcoxon signed-rank test, especially on DeepLabv3 where distillation improves the results by 2.4%.

As shown in Fig. 3(a), distillation further reduces the performance variation with improved median and tighter quartiles. As shown in Fig 4 and Table 1, most of the models achieve large improvement on right/left kidneys where distillation helps models to distinguish right and left kidneys and improve the under-segmentation problem. The DSC of the spleen shows the largest fluctuation among different models, primarily because only 111 out of 348 slices in the

Table 1. The performance comparison between different methods before and after distillation with the 3D model trained on different contrast CT data. Note: RKidney: right kidney, LKidney: left kidney.

Methods	Spleen	RKidney	LKidney	Liver	Stomach	Aorta	Avg
DeepLabv3	0.769	0.826	0.775	0.878	0.768	0.802	0.806
Swin-Unet	0.779	0.846	0.850	0.877	0.744	0.761	0.811
MISSFormer	0.782	0.865	**0.870**	0.884	0.773	0.832	0.841
distilled with UNesT trained with the BTCV dataset							
DeepLabv3 distill	0.753	0.828	0.828	0.888	0.813	0.799	0.825
Swin-Unet distill	0.755	0.850	0.856	0.875	0.769	0.799	0.824
MISSFormer distill	**0.816**	0.865	0.860	0.886	0.783	0.829	0.842
distilled with UNesT trained with in-house non-contrast dataset							
DeepLabv3 distill	0.752	0.843	0.817	**0.897**	**0.825**	0.778	0.824
Swin-Unet distill	0.779	0.845	0.857	0.879	0.754	0.784	0.821
MISSFormer distill	0.805	**0.870**	**0.870**	0.886	0.769	**0.838**	**0.845**

test set contain the spleen. This may explain why MISSFormer exhibits improvement in most organs, especially in the spleen, but less improvement on average. Comparing the distillation results with the 3D teacher model trained on both the Portal Venous and Non-contrast CT datasets, the models achieve comparable performance across different student models, indicating that the method does not require the same input contrast type.

Low-data regime analysis We investigate the 3D-to-2D distillation's effectiveness in low-data regime using DeepLabv3 model with 3D teacher pretrained with Non-contrast dataset. Fig. 3(b) shows the performance comparison between models trained on 100, 150, 200, 250 and all subjects in the training set. The model consistently outperforms the original models after distillation. We observe that the model trained with 200 subjects after distillation achieves better performance than the model trained without distillation, using all available training samples.

4 Discussions and Conclusions

As the first work of leveraging 3D embedding to assist 2D segmentation in the medical imaging domain, our work has many limitations. We only explore our method on the major/large organs where the performance gap between 2D and 3D models is not particularly large, as the single-slice dataset currently only has annotations for major organs. Based on the experimental observation, the method can only enhance the original 2D network performance to a certain extent and cannot bring it close to the performance of 3D models. We exclusively leverage 3D features learned by UNesT in our experiments, lacking additional 3D models for a comprehensive evaluation. Moreover, the generalizability of our

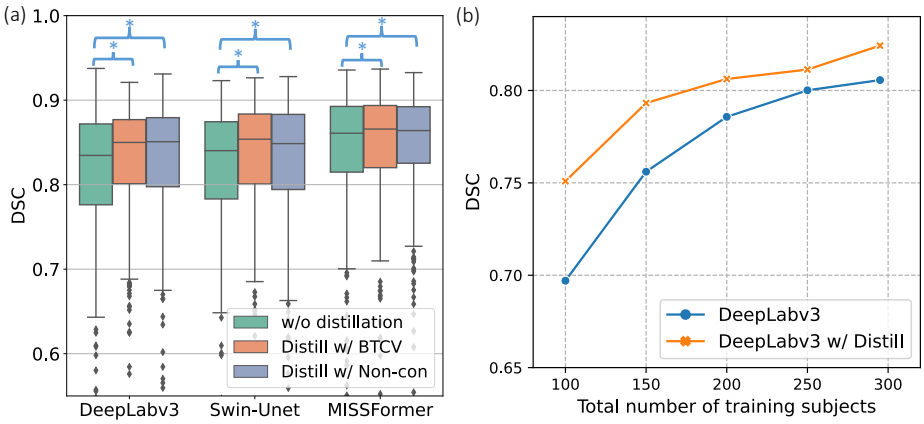

Fig. 3. Comparison of segmentation of models trained with and without distillation. In (a), models with distillation consistently reduce the variation with improved median and quartiles. * indicates statistically significant ($p < 0.05$) by Wilcoxon signed-rank test. (b) shows the results of DeepLabv3 w/ and w/o distillation using different number of training subjects.

method can be further explored by pretraining it with diverse 3D datasets, such as MRI, to assess its performance across different training data modalities.

In this paper, we introduce a novel 3D-to-2D distillation framework designed to enhance 2D single-slice segmentation by integrating embeddings derived from 3D models trained on 3D datasets. The model specifically learns to align the inter-class and intra-class correlation during training and requires no 3D input during inference. Experiments on single-slice dataset with three representative 2D segmentation models demonstrate the effectiveness of our framework. Moving forward, we hope our work could inspire future endeavors aimed at integrating 3D information to enhance 2D segmentation techniques.

Acknowledgements This research is supported by NSF CAREER 1452485, 2040462 and the National Institutes of Health (NIH) under award numbers R01EB017230, R01EB006136, R01NS09529, T32EB001628, 5UL1TR002243-04, 1R01MH121620-01, R01DK135597 (Huo), and T32GM007347; by ViSE/VICTR VR3029; and by the National Center for Research Resources, Grant UL1RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2UL1TR000445-06. This research was conducted with the support from the Intramural Research Program of the National Institute on Aging of the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The identified datasets used for the analysis described were obtained from the Research Derivative (RD), database of clinical and related data. The inhouse imaging dataset(s) used for the analysis described were obtained from ImageVU, a research repository of medical imaging data and image-related metadata. ImageVU and RD are supported by the VICTR CTSA award (ULTR000445 from NCATS/NIH) and Vanderbilt Univer-

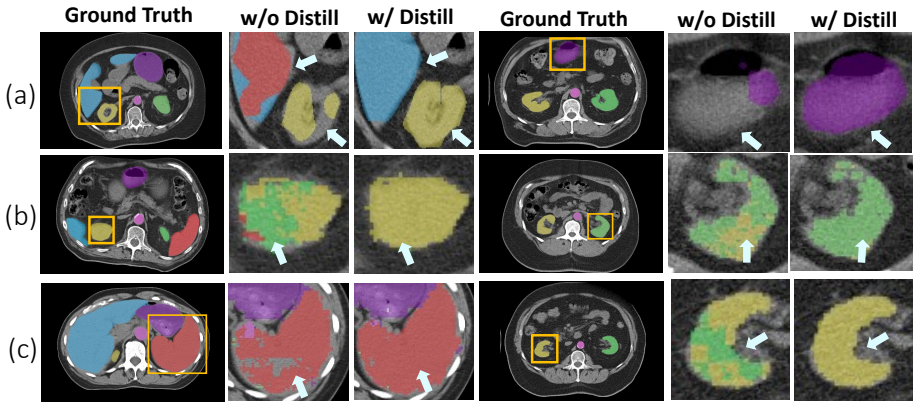

Fig. 4. Qualitative results of models with and without distillation. (a), (b), and (c) represent the results of DeepLabv3, Swin-Unet and MISSFormer, respectively. White arrows emphasized the segmentation improvement on right kidney (yellow), left kidney (green), spleen (red), liver (blue), and stomach (purple).

sity Medical Center institutional funding. ImageVU pilot work was also funded by PCORI (contract CDRN-1306-04869).

References

1. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
2. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
3. Dou, Q., Liu, Q., Heng, P.A., Glocker, B.: Unpaired multi-modal segmentation via knowledge distillation. IEEE Transactions on Medical Imaging **39**, 2415–2425 (2020), https://api.semanticscholar.org/CorpusID:210116515
4. Elbatel, M., Mart'i, R., Li, X.: Fopro-kd: Fourier prompted effective knowledge distillation for long-tailed medical image recognition. IEEE transactions on medical imaging **PP** (2023), https://api.semanticscholar.org/CorpusID:258960485
5. Hou, J., Xie, S., Graham, B., Dai, A., Nießner, M.: Pri3d: Can 3d priors help 2d representation learning? In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5693–5702 (2021)
6. Huang, T., You, S., Wang, F., Qian, C., Xu, C.: Knowledge distillation from a stronger teacher. Advances in Neural Information Processing Systems **35**, 33716–33727 (2022)
7. Huang, X., Deng, Z., Li, D., Yuan, X., Fu, Y.: Missformer: An effective transformer for 2d medical image segmentation. IEEE Transactions on Medical Imaging (2022)
8. Lee, H.H., Liu, Q., Bao, S., Yang, Q., Yu, X., Cai, L.Y., Li, T.Z., Huo, Y., Koutsoukos, X., Landman, B.A.: Scaling up 3d kernels with bayesian frequency reparameterization for medical image segmentation. In: International Conference

on Medical Image Computing and Computer-Assisted Intervention. pp. 632–641. Springer (2023)
9. Li, J., Chen, J., Tang, Y., Wang, C., Landman, B.A., Zhou, S.K.: Transforming medical imaging with transformers? a comparative review of key properties, current progresses, and future perspectives. Medical image analysis p. 102762 (2023)
10. Liao, H., Zhou, S.K., Luo, J.: Deep Network Design for Medical Image Computing: Principles and Applications. Elsevier Science & Technology (2022)
11. Liu, H., Xu, Z., Gao, R., Li, H., Wang, J., Chabin, G., Oguz, I., Grbic, S.: Cosst: Multi-organ segmentation with partially labeled datasets using comprehensive supervisions and self-training. IEEE Transactions on Medical Imaging (2024)
12. Liu, Z., Qi, X., Fu, C.W.: 3d-to-2d distillation for indoor scene parsing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4464–4474 (2021)
13. Qin, D., Bu, J.J., Liu, Z., Shen, X., Zhou, S., Gu, J.J., Wang, Z.H., Wu, L., Dai, H.F.: Efficient medical image segmentation based on knowledge distillation. IEEE Transactions on Medical Imaging **40**(12), 3820–3831 (2021)
14. Tang, Y., Gao, R., Han, S., Chen, Y., Gao, D., Nath, V., Bermudez, C., Savona, M.R., Bao, S., Lyu, I., et al.: Body part regression with self-supervision. IEEE transactions on medical imaging **40**(5), 1499–1507 (2021)
15. Tang, Y., Gao, R., Lee, H.H., Han, S., Chen, Y., Gao, D., Nath, V., Bermudez, C., Savona, M.R., Abramson, R.G., et al.: High-resolution 3d abdominal segmentation with random patch network fusion. Medical image analysis **69**, 101894 (2021)
16. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20730–20740 (2022)
17. Wang, H., Ma, C., Zhang, J., Zhang, Y., Avery, J., Hull, L., Carneiro, G.: Learnable cross-modal knowledge distillation for multi-modal learning with missing modality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2023)
18. Wang, S., Yan, Z., Zhang, D., Wei, H., Li, Z., Li, R.: Prototype knowledge distillation for medical segmentation with missing modality. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
19. Wu, Z., Li, Y., Huang, Y., Gu, L., Harada, T., Sato, H.: 3d segmenter: 3d transformer based semantic segmentation via 2d panoramic distillation. In: The Eleventh International Conference on Learning Representations (2022)
20. Yang, Q., Yu, X., Lee, H.H., Cai, L.Y., Xu, K., Bao, S., Huo, Y., Moore, A.Z., Makrogiannis, S., Ferrucci, L., et al.: Single slice thigh ct muscle group segmentation with domain adaptation and self-training. Journal of Medical Imaging **10**(4), 044001–044001 (2023)
21. Yang, Q., Yu, X., Lee, H.H., Tang, Y., Bao, S., Gravenstein, K.S., Moore, A.Z., Makrogiannis, S., Ferrucci, L., Landman, B.A.: Label efficient segmentation of single slice thigh ct with two-stage pseudo labels. Journal of Medical Imaging **9**(5), 052405–052405 (2022)
22. Yang, Q., Yu, X., Lee, H.H., Tang, Y., Bao, S., Gravenstein, K.S., Moore, A.Z., Makrogiannis, S., Ferrucci, L., Landman, B.A.: Quantification of muscle, bones, and fat on single slice thigh ct. In: Medical Imaging 2022: Image Processing. vol. 12032, pp. 422–429. SPIE (2022)

23. Yu, X., Lou, B., Zhang, D., Winkel, D., Arrahmane, N., Diallo, M., Meng, T., von Busch, H., Grimm, R., Kiefer, B., et al.: Deep attentive panoptic model for prostate cancer detection using biparametric mri scans. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 594–604. Springer (2020)
24. Yu, X., Tang, Y., Yang, Q., Lee, H.H., Bao, S., Huo, Y., Landman, B.A.: Enhancing hierarchical transformers for whole brain segmentation with intracranial measurements integration. In: Medical Imaging 2024: Clinical and Biomedical Imaging. vol. 12930, pp. 134–141. SPIE (2024)
25. Yu, X., Tang, Y., Yang, Q., Lee, H.H., Bao, S., Moore, A.Z., Ferrucci, L., Landman, B.A.: Accelerating 2d abdominal organ segmentation with active learning. In: Proceedings of SPIE–the International Society for Optical Engineering. vol. 12032. NIH Public Access (2022)
26. Yu, X., Tang, Y., Yang, Q., Lee, H.H., Gao, R., Bao, S., Moore, A.Z., Ferrucci, L., Landman, B.A.: Longitudinal variability analysis on low-dose abdominal ct with deep learning-based segmentation. In: Medical Imaging 2023: Image Processing. vol. 12464, pp. 477–483. SPIE (2023)
27. Yu, X., Yang, Q., Tang, Y., Gao, R., Bao, S., Cai, L.Y., Lee, H.H., Huo, Y., Moore, A.Z., Ferrucci, L., et al.: Reducing positional variance in cross-sectional abdominal ct slices with deep conditional generative models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 202–212. Springer (2022)
28. Yu, X., Yang, Q., Zhou, Y., Cai, L.Y., Gao, R., Lee, H.H., Li, T., Bao, S., Xu, Z., Lasko, T.A., et al.: Unest: Local spatial representation learning with hierarchical transformer for efficient medical segmentation. Medical Image Analysis **90**, 102939 (2023)
29. Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12414–12424 (2021)
30. Zhao, Q., Zhong, L., Xiao, J., Zhang, J., Chen, Y., Liao, W., Zhang, S., Wang, G.: Efficient multi-organ segmentation from 3d abdominal ct images with lightweight network and knowledge distillation. IEEE Transactions on Medical Imaging **42**, 2513–2523 (2023), https://api.semanticscholar.org/CorpusID:257824647