VISA-FSS: A Volume-Informed Self Supervised Approach for Few-Shot 3D Segmentation

Mohammad Mozafari*[1], Adeleh Bitarafan*[1,2], Mohammad Farid Azampour[2], Azade Farshad[2], Mahdieh Soleymani Baghshah[1], and Nassir Navab[2]

[1] Sharif University of Technology, Tehran, Iran
[2] Computer Aided Medical Procedures, Technical University of Munich, Germany

Abstract. Few-shot segmentation (FSS) models have gained popularity in medical imaging analysis due to their ability to generalize well to unseen classes with only a small amount of annotated data. A key requirement for the success of FSS models is a diverse set of annotated classes as the base training tasks. This is a difficult condition to meet in the medical domain due to the lack of annotations, especially in volumetric images. To tackle this problem, self-supervised FSS methods for 3D images have been introduced. However, existing methods often ignore intra-volume information in 3D image segmentation, which can limit their performance. To address this issue, we propose a novel self-supervised volume-aware FSS framework for 3D medical images, termed VISA-FSS. In general, VISA-FSS aims to learn continuous shape changes that exist among consecutive slices within a volumetric image to improve the performance of 3D medical segmentation. To achieve this goal, we introduce a volume-aware task generation method that utilizes consecutive slices within a 3D image to construct more varied and realistic self-supervised FSS tasks during training. In addition, to provide pseudo-labels for consecutive slices, a novel strategy is proposed that propagates pseudo-labels of a slice to its adjacent slices using flow field vectors to preserve anatomical shape continuity. In the inference time, we then introduce a volumetric segmentation strategy to fully exploit the inter-slice information within volumetric images. Comprehensive experiments on two common medical benchmarks, including abdomen CT and MRI, demonstrate the effectiveness of our model over state-of-the-art methods. Code is available at https://github.com/sharif-ml-lab/visa-fss

Keywords: Medical image segmentation · Few-shot learning · Few-shot semantic segmentation · Self-supervised learning · Supervoxels.

1 Introduction

Automated image segmentation is a fundamental task in many medical imaging applications, such as diagnosis [24], treatment planning [6], radiation therapy planning, and tumor resection surgeries [12, 7]. In the current literature, numerous fully-supervised deep learning (DL) methods have become dominant in the

* Equal Contribution

medical image segmentation task [5, 16, 18]. They can achieve their full potential when trained on large amounts of fully annotated data, which is often unavailable in the medical domain. Medical data annotation requires expert knowledge, and exhaustive labor, especially for volumetric images [17]. Moreover, supervised DL-based methods are not sufficiently generalizable to previously unseen classes. To address these limitations, few-shot segmentation (FSS) methods have been proposed [21–23, 25], that segment an unseen class based on just a few annotated samples. The main FSS approaches use the idea of meta-learning [11, 13, 9] and apply supervised learning to train a few-shot model. However, to avoid overfitting and improve the generalization capability of FSS models, they rely on a large number of related tasks or classes. This can be challenging as it may require a large amount of annotated data, which may not always be available. Although some works on FSS techniques focus on training with fewer data [4, 26, 20], they require re-training before applying to unseen classes. To eliminate the need for annotated data during training and re-training on unseen classes, some recent works have proposed self-supervised FSS methods for 3D medical images which use superpixel-based pseudo-labels as supervision during training [8, 19]. These methods design their self-supervised tasks (support-query pairs) by applying a predefined transformation (e.g., geometric and intensity transformation) on a support image (i.e., a random slice of a volume) to synthetically form a query one. Thus, these methods do not take into account intra-volume information and context that may be important for the accurate segmentation of volumetric images during inference.

We propose a novel volume-informed self-supervised approach for Few-Shot 3D Segmentation (VISA-FSS). Generally, VISA-FSS aims to exploit information beyond 2D image slices by learning inter-slice information and continuous shape changes that intrinsically exists among consecutive slices within a 3D image. To this end, we introduce a novel type of self-supervised tasks (see Section 2.2) that builds more varied and realistic self-supervised FSS tasks during training. Besides of generating synthetic queries (like [19] by applying geometric or intensity transformation on the support images), we also utilize consecutive slices within a 3D volume as support and query images. This novel type of task generation (in addition to diversifying the tasks) allows us to present a 2.5D loss function that enforces mask continuity between the prediction of adjacent queries. In addition, to provide pseudo-labels for consecutive slices, we propose the superpixel propagation strategy (SPPS). It propagates the superpixel of a support slice into query ones by using flow field vectors that exist between adjacent slices within a 3D image. We then introduce a novel strategy for volumetric segmentation during inference that also exploits inter-slice information within query volumes. It propagates a segmentation mask among consecutive slices using the few-shot segmenter trained by VISA-FSS. Comprehensive experiments demonstrate the superiority of our method against state-of-the-art FSS approaches.

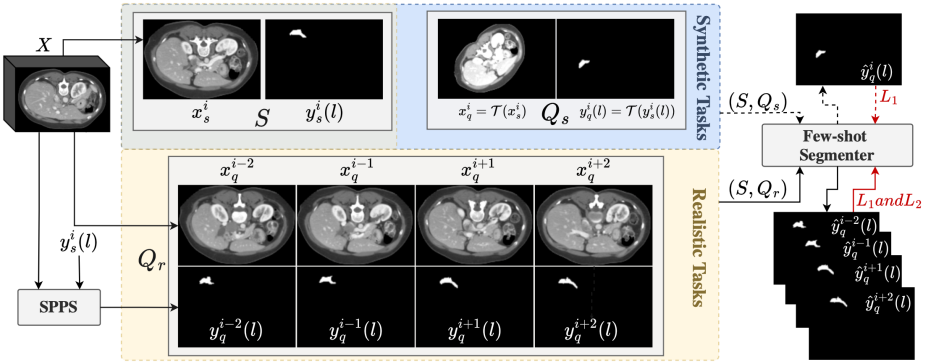

Fig. 1. An overview of the proposed VISA-FSS framework during training, where four adjacent slices of the support image, x_s^i, are taken as query images (i.e. $m = 2$). SPPS is a pseudo-label generation module for consecutive slices.

2 Methodology

In this section, we introduce our proposed VISA-FSS for 3D medical image segmentation. Our method goes beyond 2D image slices and exploits intra-volume information during training. To this end, VISA-FSS designs more varied and realistic self-supervised FSS tasks (support-query pairs) based on two types of transformations: 1) applying a predefined transformation (e.g., geometric and intensity transformation as used in [8, 19]) on a random slice as support image to synthetically make a query one, 2) taking consecutive slices in a 3D volume as support and query images to learn continuous shape transformation that exists intrinsically between consecutive slices within a volumetric image (see Section 2.2). Moreover, the volumetric view of task generation in the second type of tasks allows us to go beyond 2D loss functions. Thus, in Section 2.2, we present a 2.5D loss function that enforces mask continuity between the prediction of adjacent queries during training the few-shot segmenter. In this way, the trained few-shot segmenter is able to effectively segment a new class in a query slice given a support slice, regardless of whether it is in a different volume (due to learning the first type of tasks) or in the same query volume (due to learning the second type of tasks). Finally, we propose a volumetric segmentation strategy for inference time which is elaborated upon in Section 2.3.

2.1 Problem Setup

In FSS, a training dataset $D_{tr} = \{(x^i, y^i(l))\}_{i=1}^{N_{tr}}, l \in L_{tr}$, and a testing dataset $D_{te} = \{(x^i, y^i(l))\}_{i=1}^{N_{te}}, l \in L_{te}$ are available, where $(x^i, y^i(l))$ denotes an image-mask pair of the binary class l. L_{tr} and L_{te} are the training and testing classes, respectively, and $L_{tr} \cap L_{te} = \varnothing$. The objective is to train a segmentation model on D_{tr} that is directly applicable to segment an unseen class $l \in L_{te}$ in a query image, $x_q \in D_{te}$, given a few support set $\{(x_s^j, y_s^j(l))\}_{j=1}^{p} \subset D_{te}$. Here, q and

s indicate that an image or mask is from a query or support set. To simplify notation afterwards, we assume $p = 1$, which indicates the number of support images. During training, a few-shot segmenter takes a support-query pair (S, Q) as the input data, where $Q = \{(x_q^i, y_q^i(l))\} \subset D_{tr}$, and $S = \{(x_s^j, y_s^j(l))\} \subset D_{tr}$. Then, the model is trained according to the cross-entropy loss on each support-query pair as follows: $\mathcal{L}(\theta) = -\log p_\theta(y_q|x_q, S)$. In this work, we model $p_\theta(y_q|x_q, S)$ using the prototypical network introduced in [19], called ALPNet. However, the network architecture is not the main focus of this paper, since our VISA-FSS framework can be applied to any FSS network. The main idea of VISA-FSS is to learn a few-shot segmenter according to novel tasks designed in Section 2.2 to be effectively applicable for volumetric segmentation.

2.2 Self-Supervised Task Generation

There is a large level of information in a 3D medical image over its 2D image slices, while prior FSS methods [8, 19] ignore intra-volume information for creating their self-supervised tasks during training, although they are finally applied to segment volumetric images during inference. Previous approaches employ a predefined transformation (e.g., geometric and intensity) to form support-query pairs. We call these predefined transformations as synthetic transformations. On the other hand, there is continuous shape transformation that intrinsically exists between consecutive slices within a volume (we name them realistic transformation). VISA-FSS aims to, besides synthetic transformations, exploit realistic ones to learn more varied and realistic tasks. Figure 1 outlines a graphical overview of the proposed VISA-FSS framework, which involves the use of two types of self-supervised FSS tasks to train the few-shot segmenter. The two types of tasks are synthetic tasks and realistic tasks:

Synthetic Tasks. In the first type, tasks are formed the same as in [8, 19]. For each slice x_s^i, its superpixels are extracted by the unsupervised algorithm [10], and its pseudo-mask is generated by randomly selecting one of its superpixels as a pseudo-organ. Thus, the support is formed as $S = (x_s^i, y_s^i(l)) \subset D_{tr}$, where l denotes the chosen superpixel. Then, after applying a random synthetic transformation $\mathcal{T}$ on S, the synthetic query will be prepared, i.e., $Q_s = (x_q^i, y_q^i(l)) = (\mathcal{T}(x_s^i), \mathcal{T}(y_s^i(l)))$. In this way, the (S, Q_s) pair is taken as the input data of the few-shot segmenter, presenting a 1-way 1-shot segmentation problem. A schematic view of a representative (S, Q_s) pair is given in the blue block of Figure 1.

Realistic Tasks. To make the second type of task, we take $2m$ adjacent slices of the support image x_s^i, as our query images $\{x_q^j\}_{j \in N(i)}$, where $N(i) = \{i-m, ..., i-1, i+1, i+m\}$. These query images can be considered as real deformations of the support image. This encourages the few-shot segmenter to learn intra-volume information contrary to the first type of task. Importantly, pseudo-label generation for consecutive slices is the main challenge. To solve this problem, we introduce a novel strategy called SPPS that propagates the pseudo-label of the support

image into query ones. Specifically, we consecutively apply flow field vectors that exist between adjacent image slices on $y_s^i(l)$ to generate pseudo-label $y_q^j(l)$ as follows: $y_q^j(l) = y_s^i(l) \circ \phi(x_s^i, x_q^{i+1}) \circ \phi(x_q^{i+1}, x_q^{i+2}) \circ ... \circ \phi(x_q^{i+m-1}, x_q^{i+m})$ for $j > m$, and $y_q^j(l) = y_s^i(l) \circ \phi(x_s^i, x_q^{i-1}) \circ \phi(x_q^{i-1}, x_q^{i-2}) \circ ... \circ \phi(x_q^{i-m+1}, x_q^{i-m})$ for $j < m$, where $\phi(x^i, x^j)$ is the flow field vector between x^i and x^j, which can be computed by deformably registering the two images using VoxelMorph [2] or Vol2Flow [3]. A schematic illustration of pseudo-label generation using SPPS is depicted in the Supplementary Materials. The yellow block in Figure 1 demonstrates a representative (S, Q_r) pair formed using realistic tasks, where $Q_r = \{(x_q^j, y_q^j(l))\}_{j \in N(i)}$.

Loss Function The network is trained end-to-end in two stages. In the first stage, we train the few-shot segmenter on both types of synthetic and realistic tasks using the segmentation loss employed in [19] and regularization loss defined in [25], which are based on the standard cross-entropy loss. Specifically, in each iteration, the segmentation loss L_{seg} can be followed as: $L_{seg} = \frac{-1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} y_q^i(l)(h,w) \log(\hat{y}_q^i(l)(h,w)) + (1 - y_q^i(l)(h,w)) \log(1 - \hat{y}_q^i(l)(h,w))$, which is applied on a random query x_q^i (formed by synthetic or realistic transformations) to predict the segmentation mask $\hat{y}_q^i(l)$, where $l \in L_{tr}$. The regularization loss L_{reg} is defined to segment the class l in its corresponding support image x_s^i, as follows: $L_{reg} = \frac{-1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} y_s^i(l)(h,w) \log(\hat{y}_s^i(l)(h,w)) + (1 - y_s^i(l)(h,w)) \log(1 - \hat{y}_s^i(l)(h,w))$.

Overall, in each iteration, the loss function during the first-stage training is $L_1 = L_{seg} + L_{reg}$.

In the second stage of training, we aim to exploit information beyond 2D image slices in a volumetric image by employing realistic tasks.

To this end, we define the 2.5D loss function, $L_{2.5D}$, which enforces mask continuity among the prediction of adjacent queries. The proposed $L_{2.5D}$ profits the Dice loss [18] to measure the similarity between the predicted mask of $2m$ adjacent slices of the support image x_s^i as follows:

$$L_{2.5D} = \frac{1}{2m-1} \sum_{j \in N(i)} (1 - Dice(\hat{y}_q^j(l), \hat{y}_q^{j+1}(l))). \tag{1}$$

Specifically, the loss function compares the predicted mask of a query slice with the predicted mask of its adjacent slice and penalizes any discontinuities between them. This helps ensure that the model produces consistent and coherent segmentation masks across multiple slices, improving the overall quality and accuracy of the segmentation. Hence, in the second-stage training, we train the network only on realistic tasks using the loss function: $L_2 = L_{seg} + L_{reg} + \lambda_1 L_{2.5D}$, where λ_1 is linearly increased from 0 to 0.5 every 1000th iteration during training. Finally, after self-supervised learning, the few-shot segmenter can be directly utilized for inference on unseen classes.

2.3 Volumetric Segmentation Strategy

During inference, the goal is to segment query volumes based on a support volume with only a sparse set of human-annotated slices, while the few-shot segmenter is trained with 2D images. To evaluate 2D segmentation on 3D volumetric images, we take inspiration from [21] and propose the volumetric segmentation propagation strategy (VSPS).

Assume, $X_s = \{x_s^1, x_s^2, ..., x_s^{n_s}\}$ and $X_q = \{x_q^1, x_q^2, ..., x_q^{n_q}\}$ denote support and query volumes, comprising of n_s and n_q consecutive slices, respectively. We follow the same setting as [8, 19, 21] in which slices containing semantic class l are divided into K equally-spaced groups, including $[X_s^1, X_s^2, ..., X_s^K]$ in the support, and $[X_q^1, X_q^2, ..., X_q^K]$ in the query volume, where X^k indicates the set of slices in the k^{th} group. Suppose, in each of the k groups in the support volume, the manual annotation of the middle slice $[(x_s^c)^1, (x_s^c)^2, ..., (x_s^c)^K]$ are available as in [8, 19, 21]. For volumetric segmentation, previous methods [8, 19, 21], for each group $k \in \{1, ..., K\}$, pair the annotated center slice in the support volume with all the unannotated slices of the corresponding group in the query volume. More precisely, $((x_s^c)^k, (y_s^c)^k)$ is considered as the support for all slices in X_q^k, where $(y_s^c)^k$ is annotation of the center slice $(x_s^c)^k$. Finally, they use the 2D few-shot segmenter to find the mask of each of the query slices individually and therefore segment the whole query volume accordingly. In this work, we exploit the VSPS algorithm, which is based on two steps. In the first step, an inter-volume task is constructed to segment the center slice of each group in the query volume. More precisely, the center slice of each query group, $(x_q^c)^k$, is segmented using $((x_s^c)^k, (y_s^c)^k)$ as the support. Then, by employing the volumetric view even in the inference time, we construct intra-volume tasks to segment other slices of each group. Formally, VSPS consecutively segments each $(x_q^j)^k \in X_q^k$, starting $(x_q^c)^k$, with respect to the image-mask pair of its previous slice, i.e., $((x_q^{j-1})^k, (\hat{y}_q^{j-1})^k)$. In fact, we first find the pseudo-mask of $(x_q^c)^k$ using the 2D few-shot segmenter and consequently consider this pseudo-annotated slice as the support for all other slices in X_q^k. It is worth mentioning that our task generation strategy discussed in Section 2.2 is capable of handling such intra-volume tasks. Further details of the VSPS algorithm are brought in the Supplementary Materials.

3 Experiments

3.1 Experimental Setup

To unify experiment results, we follow the evaluation protocol established by [19], such as Hyper-parameters, data preprocessing techniques, evaluation metric (i.e., Dice score), and compared methods. The architecture and implementation of the network are exactly the same as developed in [19]. Moreover, during inference, a support volume with 3 annotated slices (i.e., $K = 3$) is used as a reference to segment each query volume, the same as in [19]. Also, we set $m = 3$, taking 3 adjacent slices of the support image as consecutive query images. However, the effect of this hyper-parameter is investigated in the Supplementary Materials.

Dataset Following [8, 19], we perform experiments on two common medical benchmarks, including abdominal CT image scans from MICCAI 2015 Multi-Atlas Abdomen Labeling challenge [15] and abdominal MRI image scans from ISBI 2019 Combined Healthy Abdominal Organ Segmentation Challenge [14]. In addition, in all experiments, average results are reported according to 5-fold cross-validation on four anatomical structures the same as in [8, 19], including left kidney (LK), right kidney (RK), spleen, and liver.

3.2 Results and Discussion

Comparison with Existing Approaches. Table 1 compares VISA-FSS with state-of-the-art FSS methods in terms of Dice, including: Vanilla PANet [25], SE-Net [21], SSL-RPNet [23], SSL-ALPNet [19], and CRAPNet [8]. Vanilla PANet and SE-Net are baselines on natural and medical images, respectively, which utilize manual annotations for training. SSL-RPNet, SSL-ALPNet, and CRAP-Net are self-supervised methods that construct their FSS tasks using synthetic transformations (e.g., geometric and intensity) in the same way, and are only different in the network architecture. As demonstrated, VISA-FSS outperforms vanilla PANet and SE-Net without using any manual annotation in its training phase. Moreover, the performance gains of VISA-FSS compared with SSL-RPNet, SSL-ALPNet, and CRAPNet highlight the benefit of learning continuous shape transformation among consecutive slices within a 3D image for volumetric segmentation. Also, the performance of VISA-FSS was evaluated using Hausedorff Distance and Surface Dice metrics on CT and MRI datasets. On the CT dataset, VISA-FSS reduced SSLALPNet's Hausedorff Distance from 30.07 to 23.62 and improved Surface Dice from 89.31% to 90.68%. On the MRI dataset, it decreased Hausedorff Distance from 26.49 to 22.14 and improved Surface Dice from 90.16% to 91.08%. Further experiments and qualitative results are given in the Supplementary Materials, demonstrating satisfactory results on different abdominal organs.

Table 1. Comparison results of different methods (in Dice score) on abdominal images.

Method	Abdominal-CT				Mean	Abdominal-MRI				Mean
	RK	LK	Spleen	Liver		RK	LK	Spleen	Liver	
Vanilla PANet [25]	21.19	20.67	36.04	49.55	31.86	32.19	30.99	40.58	50.40	38.53
SE-Net [21]	12.51	24.42	43.66	35.42	29.00	47.96	45.78	47.30	29.02	42.51
SSL-RPNet [23]	66.73	65.14	64.01	72.99	67.22	81.96	71.46	73.55	75.99	75.74
CRAPNet [8]	74.18	74.69	70.37	75.41	73.66	86.42	81.95	74.32	76.46	79.79
SSL-ALPNet [19]	71.81	72.36	70.96	78.29	73.35	85.18	81.92	72.18	76.10	78.84
VISA-FSS (Ours)	**76.17**	**77.05**	**76.51**	**78.70**	**77.11**	**89.55**	**87.90**	**78.05**	**77.00**	**83.12**

Effect of Task Generation. To investigate the effect of realistic tasks in self-supervised FSS models, we perform an ablation study on the absence of this

type of task. The experiment results are given in rows (a) and (b) of Table 2. As expected, performance gains can be observed when both synthetic and realistic tasks are employed during training. This can highlight that the use of more and diverse tasks improves the performance of FSS models.

Of note, to generate pseudo-label for consecutive slices, instead of SPPS, we can also employ supervoxel generation strategies like the popular SLIC algorithm [1]. However, we observed that by doing so the performance is 66.83 in the term of mean Dice score, under-performing SPPS (row (b) in Table 2) by about 8%. It can be inferred that contrary to SLIC, SPPS implicitly takes pseudo-label shape continuity into account due to its propagation process, which can help construct effective realistic tasks. To intuitively illustrate this issue, visual comparison of some pseudo-labels generated by SLIC and SPPS is depicted in the Supplementary Materials. In addition, to demonstrate the importance of the 2.5D loss function defined in Equation 1 during training, we report the performance with and without $L_{2.5D}$ in Table 2 (see row (d) and (e)). We observe over 1% increase in the average Dice due to applying the 2.5D loss function.

Table 2. Ablation studies on task generation and different types of volumetric segmentation strategies on abdominal CT dataset (Results are based on Dice score).

	Training		Inference	Organs				Mean
	Tasks	Loss	Vol. Seg. Str.	RK	LK	Spleen	Liver	
(a)	Syn.	w.o. $L_{2.5D}$	VSS [21]	71.81	72.36	70.96	78.29	73.35
(b)	Syn. + Re.	w.o. $L_{2.5D}$	VSS [21]	71.59	72.02	73.85	78.57	74.01
(c)	Syn. + Re.	w.o. $L_{2.5D}$	RPS	71.13	71.72	72.68	77.69	73.31
(d)	Syn. + Re.	w.o. $L_{2.5D}$	VSPS	74.67	75.14	75.00	78.74	75.88
(e)	Syn. + Re.	w. $L_{2.5D}$	VSPS	76.17	77.05	76.51	78.70	77.11

Importance of the Volumetric Segmentation Strategy. To verify the influence of our proposed volumetric segmentation strategy, we compare VSPS against two different strategies: VSS and RPS. VSS (volumetric segmentation strategy) is exactly the same protocol established by [21] (explained in detail in Section 2.3). In addition, RPS (registration-based propagation strategy) is a ablated version of VSPS which propagates the annotation of the center slice in each query volume group into unannotated slices in the same group using registration-based models like [3] instead of using the trained few-shot segmenter. Comparison results are given in rows (b) to (d) of Table 2, demonstrating the superiority of VSPS compared with other strategies. In fact, due to learning synthetic transformations (e.g., geometric and intensity transformation) during training, VSPS, during inference, can successfully segment a new class in a query slice given a support slice from a different volume. Also, due to learning realistic transformations (e.g., intra-volume transformations), each query slice can be effectively segmented with respect to its neighbour slice.

4 Conclusion

This work introduces a novel framework called VISA-FSS, which aims to perform few-shot 3D segmentation without requiring any manual annotations during training. VISA-FSS leverages inter-slice information and continuous shape changes that exist across consecutive slices within a 3D image. During training, it uses consecutive slices within a 3D volume as support and query images, as well as support-query pairs generated by applying geometric and intensity transformations. This allows us to exploit intra-volume information and introduce a 2.5D loss function that penalizes the model for making predictions that are discontinuous among adjacent slices. Finally, during inference, a novel strategy for volumetric segmentation is introduced to employ the volumetric view even during the testing time.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels. Tech. rep. (2010)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
3. Bitarafan, A., Azampour, M.F., Bakhtari, K., Soleymani Baghshah, M., Keicher, M., Navab, N.: Vol2flow: Segment 3d volumes using a sequence of registration flows. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Proceedings, Part IV. pp. 609–618. Springer (2022)
4. Bitarafan, A., Nikdan, M., Baghshah, M.S.: 3d image segmentation with sparse annotation by self-training and internal registration. IEEE Journal of Biomedical and Health Informatics **25**(7), 2665–2672 (2020)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
6. Chen, X., Sun, S., Bai, N., Han, K., Liu, Q., Yao, S., Tang, H., Zhang, C., Lu, Z., Huang, Q., et al.: A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. Radiotherapy and Oncology **160**, 175–184 (2021)
7. Denner, S., Khakzar, A., Sajid, M., Saleh, M., Spiclin, Z., Kim, S.T., Navab, N.: Spatio-temporal learning from longitudinal data for multiple sclerosis lesion segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6. pp. 111–121. Springer (2021)
8. Ding, H., Sun, C., Tang, H., Cai, D., Yan, Y.: Few-shot medical image segmentation with cycle-resemblance attention. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2488–2497 (2023)
9. Farshad, A., Makarevich, A., Belagiannis, V., Navab, N.: Metamedseg: Volumetric meta-learning for few-shot organ segmentation. In: Domain Adaptation and Representation Transfer 2022. pp. 45–55. Springer (2022)

10. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International journal of computer vision **59**, 167–181 (2004)
11. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017)
12. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging **32**, 582–596 (2019)
13. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(9), 5149–5169 (2021)
14. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
15. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge. vol. 5, p. 12 (2015)
16. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE transactions on medical imaging **37**(12), 2663–2674 (2018)
17. Lutnick, B., Ginley, B., Govind, D., McGarry, S.D., LaViolette, P.S., Yacoub, R., Jain, S., Tomaszewski, J.E., Jen, K.Y., Sarder, P.: An integrated iterative annotation technique for easing neural network training in medical image analysis. Nature machine intelligence **1**(2), 112–119 (2019)
18. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
19. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. pp. 762–780. Springer (2020)
20. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D.: Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 669–677. Springer (2019)
21. Roy, A.G., Siddiqui, S., Pölsterl, S., Navab, N., Wachinger, C.: 'squeeze & excite'guided few-shot segmentation of volumetric images. Medical image analysis **59**, 101587 (2020)
22. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems **30** (2017)
23. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3918–3928 (2021)
24. Tsochatzidis, L., Koutla, P., Costaridou, L., Pratikakis, I.: Integrating segmentation information into cnn for breast cancer diagnosis of mammographic masses. Computer Methods and Programs in Biomedicine **200**, 105913 (2021)

25. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: proceedings of the IEEE/CVF international conference on computer vision. pp. 9197–9206 (2019)
26. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8543–8553 (2019)