Federated Multi-Centric Image Segmentation with Uneven Label Distribution

Francesco Galati[1], Rosa Cortese[2], Ferran Prados[3,4,5], Marco Lorenzi[6], and Maria A. Zuluaga[1,7]

[1] Data Science Department, EURECOM, Sophia Antipolis, France
[2] Department of Medicine, Surgery and Neuroscience, University of Siena, Italy
[3] Centre for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, UK
[4] Department of Neuroinflammation, Faculty of Brain Sciences, Queen Square MS Centre, UCL Institute of Neurology, University College London, UK
[5] e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain
[6] Inria Center of Université Côte d'Azur, Epione Team, Sophia Antipolis, France
[7] School of Biomedical Engineering & Imaging Sciences, King's College London, UK
{francesco.galati,maria.zuluaga}@eurecom.fr

Abstract. While federated learning is the state-of-the-art methodology for collaborative learning, its adoption for training segmentation models often relies on the assumption of uniform label distributions across participants, and is generally sensitive to the large variability of multi-centric imaging data. To overcome these issues, we propose a novel federated image segmentation approach adapted to complex non-iid setting typical of real-life conditions. We assume that labeled dataset is not available to all clients, and that clients data exhibit differences in distribution due to three factors: different scanners, imaging modalities and imaged organs. Our proposed framework collaboratively builds a *multimodal data factory* that embeds a shared, disentangled latent representation across participants. In a second asynchronous stage, this setup enables local domain adaptation without exchanging raw data or annotations, facilitating target segmentation. We evaluate our method across three distinct scenarios, including multi-scanner cardiac magnetic resonance segmentation, multi-modality skull stripping, and multi-organ vascular segmentation. The results obtained demonstrate the quality and robustness of our approach as compared to the state-of-the-art methods.

Keywords: Missing Labels · Domain Adaptation · Federated Learning · Image Segmentation.

1 Introduction

In recent years, the use of deep learning has become prevalent in medical image segmentation. Nevertheless, while supervised learning models necessitate large collections of labeled data to prevent overfitting and achieve high quality results, in practice they are often trained on small datasets provided by single

data centers. This limitation is primarily due to the high costs associated with acquiring medical images, and the tedious expertise-demanding effort needed for their annotation. Moreover, while the sharing of medical data is essential to train more robust models, in real-life scenarios it is often complex to gather data from different hospitals in a centralized repository, due to privacy constraints and current regulations. To address this issue, Federated Learning (FL) is a promising collaborative learning approach enabling multiple clients to jointly train a model by sharing partially optimized model parameters instead of private data. In the context of supervised learning for medical image segmentation, current FL schemes are mostly based on the assumption of homogeneous, independent and identically distributed (iid) data across centers. For example, fully labeled data must be available at each site to perform the distributed learning task. This constraint implies that unlabeled data should be discarded, entailing a loss of potential relevant information. Additionally, heterogeneity in data distributions across clients is often neglected, leading to models prone to distribution shifts, i.e., with dropping performance when applied to other clients.

In this work, we propose a novel federated image segmentation approach adapted to complex non-iid setting typical of real-life conditions. We first assume that labeled dataset is not available to all clients, but only to a few ones (e.g. only one) considered as the source domain. Clients data exhibit substantial differences in distribution due to three factors, i.e., different 1) scanners, 2) imaging modalities and 3) imaged organs. To achieve a robust segmentation model in this complex non-iid scenario, we propose a federated learning framework based on two main modeling steps. First, clients collaboratively build a multimodal data factory to create a shared and disentangled data latent representation. Second, individual clients can tailor this common knowledge to adapt their data distribution to the one of the labeled source domain. This second step is performed asynchronously and addresses target segmentation with minimal labeling effort and without the need to exchange images or annotations between clients, thus enhancing efficiency and data governance. We apply our method on multi-scanner cardiac segmentation, multi-modality skull-stripping, and multi-organ vascular segmentation, achieving improved dice scores up to 13.4% as compare to competing segmentation methods from the state-of-the-art.

The paper is organized as follows: Section 2 reviews relevant research on the problems of federated learning and domain shift. Section 3 presents our approach to handle label scarcity and distribution shifts in federated learning. Section 4 evaluates our method using three test scenarios.

2 Related Work

Federated Learning. Several federated approaches have been proposed to address the setting of multiple domains and unlabeled data centers for medical image synthesis [24,3] and segmentation. While [25] accounts for weak and pixel-wise annotations, this setting does not apply when labels are fully absent in a client. Multiple federated semi-supervised learning models deal with la-

bel scarcity by using pseudo-labeling strategies [21], combined with contrastive learning [26] or knowledge transfer [17]. While effective in certain segmentation tasks, these works do not address the issue of domain shift across sites. FedDG [15] proposes federated domain generalization exploiting episodic frequency learning across multi-source data distributions. Our work considers the more general problem of federated domain adaptation [27], better suited to handle large distributional shifts.

Domain shift and Foundation Models. Domain Adaptation (DA) and Domain Generalization (DG) are the main techniques used to address the problem of domain shift in medical image segmentation. DA techniques transfer knowledge from fully-labeled source domains to a target domain with limited or no labels. In centralized settings, they have been applied to a multitude of organs [4,20,8]. DG typically concentrates on data augmentation [28,7] to mimic changes in intensity and geometry across scanners, protocols, or populations. Overall, DA and DG tend to develop specialized networks that are trained on datasets confined to a single image modality and organ.

Foundation models are trained on massive and diverse datasets, exhibiting remarkable zero-shot generalizability. Their deployment in actual clinical settings is hindered by the need to assemble vast labeled datasets, which is often unfeasible. Furthermore, they require fine-tuning or prompting [16], which is problematic when annotations are scarce or full automation is desired. UniverSeg [1] overcomes these limitations by producing segmentation maps from a query image and an example set of few image-label pairs, but it relies on training with more than 22,000 scans from 53 publicly available datasets, highlighting how data-greedy these methods are. In contrast, our framework involves training with very few labels and without the need for data centralization.

3 Method

We formulate a collaborative learning scheme that involves a group of K clients, each owning a dataset D_k from a unique domain $\mathcal{D}_k$, with $k = 1, ..., K$. Initially, these datasets lack annotations. Operating in an unsupervised manner, we train a multimodal data factory F, which serves multiple functions: 1) performing conditional image synthesis to generate images $\hat{x}$ that resemble those from the clients; 2) providing a disentangled latent space $\mathcal{W}$ which supports the representation and translation of domains with significant differences; 3) allowing for customization through the addition of domain-specific segmentation branches. This design allows the clients to exchange the necessary knowledge to segment data across all domains $\mathcal{D}_i$ without sharing images or annotations, thereby preserving data governance. Figure 1 illustrates the described scenario.

3.1 Multimodal Data Factory via Federated Learning

The first step of our method aims to build the data factory F that integrates domains $\mathcal{D}_k$ from all clients. When fed with a latent code z randomly drawn from

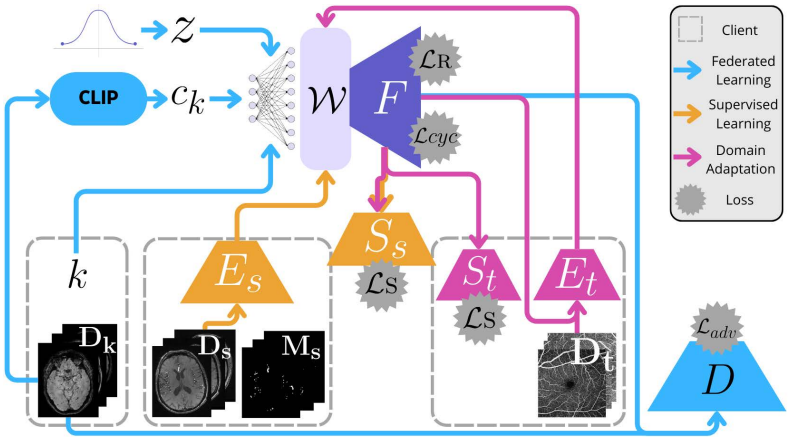

Fig. 1. Using federated learning, K clients collaboratively train a multimodal data factory F (in blue). Afterwards, source clients can contribute by training locally segmentation branches S_s (in orange), while target clients can asynchronously acquire the information required to segment their data D_t via domain adaptation (in pink).

a Gaussian distribution, F is trained to produce an image $\hat{x}$ that resembles those from the clients. This is achieved through adversarial learning, which employs an external discriminator D to distinguish between real and fake samples. In response, F aims to fool D by retrieving images that look realistic. Tailoring the generative process more closely to each client's domain $\mathcal{D}_k$, F and D are adapted to be injected with the client identifier $k \in [1, K]$, which is one-hot-encoded, embedded into a 512-dimensional vector, and merged with the feature vector z. To enhance the quality of the images and ensure robust representation across domains with significant gaps, we further condition the generation by introducing a domain-specific key c_k. This is derived locally by computing the average of CLIP [22] encodings of all images x_i within the dataset $\mathcal{D}_k$. The key is processed through a linear layer and averaged with the label condition:

$$\hat{x} = F \leftarrow z \oplus \frac{1}{2}\left(e_k(k) + e_c\left(\frac{1}{|D_k|}\sum_{x_i \in D_k} \text{CLIP}(x_i)\right)\right) \tag{1}$$

where $\oplus$ denotes concatenation, $e_k(\cdot)$ and $e_c(\cdot)$ are the embeddings processed by the additional linear layers for the client identifier k and the average CLIP encoding c_k, respectively. Compatibly with a federated environment, clients participate to the optimization of the training objective of F as follows:

$$\mathcal{L}_{tot} = \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}^k_{adv}(F, D) + \mathcal{L}^k_{R_1}(D) + \mathcal{L}^k_{pl}(F) \tag{2}$$

where $\mathcal{L}^k_{adv}$ is the non-saturating loss [9], $\mathcal{L}^k_{R_1}$ is the R_1 regularization [19], and $\mathcal{L}^k_{pl}$ is the path length regularization [13], with each term computed locally using

only the data from the respective client k. In the process of generating $\hat{x}$, the quantities z, k and c_k are combined into a single, unified latent representation $w \in \mathcal{W}$, a transit latent space which, as detailed in [12], is unwarped by $\mathcal{L}_{pl}^{k}$. This regularization transforms $\mathcal{W}$ into a disentangled latent space where different directions consistently correspond to individual, controllable aspects of variation in the generated images. At the end of the training, $\mathcal{W}$ summarizes the characteristics of all domains $\mathcal{D}_k$, including the domain-specific features differentiating each one. Furthermore, the aforementioned property of disentanglement enables the creation of images that smoothly transition from one domain to another, allowing to find new intermediate domains through latent space morphing.

3.2 Domain Adaptation via Local Training

After training the data factory F, clients independently operate the local adaptation step. We assume that at least one among the K clients, denoted as client s, disposes of annotation masks M_s for the respective dataset D_s, either completely or partially. Client s is thus responsible for the development of a new segmentation branch S_s to be integrated into F [29]. To this end, a local encoder E_s is trained in a fully supervised manner to reverse the generation process detailed in Section 3.1. In particular, given a image x_i^s, E_s aims to find the latent vector $\tilde{w}_i^s$ to be fed into F in order to retrieve the closest reconstruction $\tilde{x}_i^s \approx x_i^s$. In the meanwhile, the feature maps produced by E_s and F are inputted into S_s to produce the corresponding segmentation mask $\tilde{y}_i^s$. This is achieved using mean squared error and LPIPS as reconstruction losses ($\mathcal{L}_{\text{R}}$), while using Dice and cross-entropy as segmentation losses ($\mathcal{L}_{\text{S}}$).

Once trained, S_s becomes available to any other client t, enabling them to access the combined knowledge from F and S_s for adaptation to their specific dataset D_t. This process is facilitated by the capability of the data factory to generate synthetic, yet realistic samples x_j^s that resemble the characteristics of their native domain $\mathcal{D}_s$. This time, a local encoder E_t is used to derive two distinct latent vectors, $\tilde{w}_j^s$ for reconstruction as in the previous setting, and $\hat{w}_j^s$ for image-to-image translation to optimize a cycle-consistency loss $\mathcal{L}_{cyc}^{t}(E_t) = \mathcal{L}_{\text{R}}(x_j^s, \hat{\hat{x}}_j^s) + \mathcal{L}_{\text{R}}(x_i^t, \hat{\hat{x}}_i^t)$. This ensures that both the synthetic sample x_j^s and the target domain sample x_i^t complete a full cycle of domain transformations to maintain image fidelity: first adapting to the other's domain, then returning to their own, yielding $\hat{\hat{x}}_j^s$ and $\hat{\hat{x}}_i^t$. Furthermore, both E_t and a new segmentation branch S_t receive guidance from S_s to achieve accurate segmentation, primarily on source data x_j^s, with S_s remaining frozen to return masks $\tilde{y}_j^s$. To enforce this supervision, we include a small set M_t of target annotations y_i^t, with $|M_t| \ll |M_s|$. The resulting segmentation loss $\mathcal{L}_{seg}^{t}(E_t, S_t) = \mathcal{L}_{\text{S}}(\hat{y}_j^s, \tilde{y}_j^s) + \mathcal{L}_{\text{S}}(\tilde{y}_i^t, y_i^t)$ forces E_t to perform translation in a label-preserving manner, thanks to the capability of F to disentangle latent vectors within $\mathcal{W}$. This disentanglement enables smooth transitions across distant domains while maintaining control over the specific attributes of the generated image, ensuring that the translation process remains label-consistent. Notably,

this stage does not involve the use of discriminators, as the goal is not to produce exact replicas of the source and target domain images but to assist segmentation.

We highlight that this step operates independently from FL. Inference segmentation on a new image x^t_{new} is performed by averaging predictions $\hat{y}^t_{new}$ and $\tilde{y}^t_{new}$ from both the source and target segmentation branches S_s and S_t.

4 Experiments and Results

The proposed method is demonstrated on several non-iid setup for the segmentation of anatomical structures, presenting increasing heterogeneity: *multi-scanner* (hearth segmentation; same modality, same organ), *multi-modal* (brain segmentation; varying modalities, same organ), and *multi-organ* (vessel segmentation; varying modalities, varying organs).

4.1 Datasets and Tasks

Multi-scanner setup (MS). Cardiac MRI images from the M&Ms Challenge [2] include 345 patients with cardiomyopathies as well as healthy subjects. MR images, taken at both end diastole and end systole, were labeled for the left ventricle (LV), right ventricle (RV), and myocardium (MYO). Data collection was carried out across five centers using scanners from four vendors. The multi-centric setup was simulated by partitioning the data per scanner type, thus obtaining 4 clients. For the FL step, we trained the data factory with clients holding data from scanners Siemens, Philips, and GE. The source domain was selected as the client with scanner type Philips, and the client with scanner type Canon was only used for the DA task (no FL).

Multi-modal setup (MM). The SynthStrip dataset [10], provides a comprehensive collection of head images aggregated from multiple sources and spanning various contrasts, resolutions, and populations ranging from infants to glioblastoma patients. The considered task here is skull-stripping across multiple imaging modalities. Specifically, our study incorporates 20 CT and 20 PET scans from the CERMEP-IDB-MRXFDG dataset [18], as well as 32 PD-weighted (PDw) and 36 T2-weighted (T2w) MRI scans from the FreeSurfer Maintenance (FSM) dataset [6]. The multi-centric setup was simulated by partitioning the data across modalities, thus obtaining 4 clients. The data factory was trained using clients holding CT, PDw and PET data. The source domain is represented by the client with PDw images, and the client with T2 images was used for DA only (no FL).

Multi-organ setup (MO). We selected 49 time-of-flight (TOF) MRA volumes from the OASIS-3 dataset to study brain arteries in 27 cognitively normal adults and 10 patients with cognitive decline, aged between 42 to 95 years. Additionally, we used 28 SWI venographies of adult subjects with no visible lesions, derived from retrospective studies conducted at UCL Queen Square Institute of Neurology, Queen Square MS Centre, University College London. Finally, the OCTA-500 dataset provides optical coherence tomography angiographies (OCTA) in

three different 2D projections, collected from 500 subjects aged 7 to 85 years, with 49.8% of them affected by ophthalmic diseases. Data was partitioned across the 4 datasets. FL training was performed with OASIS, SWI and OCTA. The source domain for this setup was OASIS. The IXI data was used only for DA (no FL).

For all setups, segmentation results are assessed through the Dice coefficient. Performance evaluations are conducted on the hold-out test sets specific to each scenario. For MS, we average the results over three regions LV, RV, and MYO.

4.2 Implementation Details and Competing Methods

Data was preprocessed compatibly with the federated learning scenario. Our federated learning framework is implemented using PyTorch 1.13.1 and FedBiomed 5.0.1 [5], with 350 FL training rounds, 2000 stochastic gradient steps per round, and batch size 2. FL aggregation is performed with FedAvg, by using uniform weights and by sampling 2 clients per round. To enhance convergence and smoother integration across different domains, we run a refinement stage of 35 rounds with 200 iterations each. After the FL step, supervised segmentation is trained locally on all clients hosting labeled datasets D_s for 15k iterations with batches of 8 images. This is followed by domain adaptation, conducted on each target clients t for 20k iterations with batches of 4 images. Once training is finished, the checkpoints with the best validation performance on each client's local validation set are selected for the final evaluation. The model architecture, including components F, D, E_k, and S_k, builds upon preceding works [13,23,29]. We trained, validated and tested our proposed method as well as the state-of-the-art methods on two NVIDIA GeForce RTX 2080 Ti GPUs. All code and experiments can be accessed on github.com/i-vesseg/RobustMedSeg.

We compare our method against four state-of-the-art DA for image segmentation in heterogeneous setting: **1) nnU-Net** [11], a self-configuring method for deep learning-based biomedical image segmentation, validated on a wide range of segmentation tasks with state-of-the-art performance. We combine it with Data Augmentation (nnU-Net+DAug) and Transfer Learning (nnU-Net+TL); **2) FedMed-GAN** [24], a federated image-to-image translation method for unpaired cross-modality image synthesis. This is concatenated with nnU-Net to perform downstream segmentation; **3) FedDG** [15], introducing federated domain generalization to enhance model adaptability to unseen domains. To leverage multiple source domains, the network is trained using all datasets except the target; **4 and 5) SAM/MedSAM** [14,16], a foundation model pretrained over 1.1 billion segmentation masks in its original version and fine-tuned with 1.5 million medical annotated images; **6)UniverSeg** [1], a foundation model leveraging in-context learning to solve unseen segmentation tasks with little to no labeled data.

Table 1, column G, details the methods (including ours) requiring a small set M_t of target annotations to guide the segmentation process. In our experiments, this set always includes three midpoint slices, extracted from three random volumes of D_t. For nnU-Net+TL, SAM and MedSAM, the set is used to fine-tune

Table 1. Segmentation results (Dice score) across setups (MS, MM and MO) for the target clients. Column G indicates methods requiring M_t to guide the segmentation process.

	G	MS			MM			MO		
		Siemens	GE	Canon*	PET	CT	T2w*	SWI	OCTA	IXI*
nnU-Net		44.0±30.1	82.0±9.3	75.9±16.4	17.7±8.6	43.6±4.1	68.8±17.3	0.0±0.0	3.0±1.5	67.0±2.6
+DAug		78.1±18.7	**86.3±7.2**	**85.5±8.2**	0.1±0.2	76.7±7.3	71.7±21.2	0.0±0.0	0.2±0.6	48.4±13.2
+TL	✓	80.8±15.6	85.4±7.7	85.0±9.0	63.7±4.3	70.3±9.4	87.7±3.4	56.9±2.7	69.3±10.5	**71.5±3.5**
FedMed-GAN		12.0±20.5	65.7±26.5	67.6±30.1	0.3±0.5	0.0±0.0	12.2±22.6	0.0±0.0	0.7±0.4	0.5±0.6
FedDG		81.8±12.7	85.9±7.9	81.6±8.4	0.4±0.8	2.9±3.8	62.6±4.2	0.2±0.1	11.0±7.4	66.2±2.1
SAM	✓	1.8±1.8	1.4±1.4	0.6±1.0	12.7±6.7	9.8±1.3	21.8±1.8	32.1±6.2	34.6±12.6	2.2±0.2
MedSAM	✓	4.3±5.7	4.0±3.0	4.3±4.8	34.3±12.6	10.2±3.3	37.7±4.7	3.4±1.1	13.2±6.3	2.0±0.4
UniverSeg	✓	81.8±12.7	85.9±7.9	56.1±35.9	77.5±5.8	36.0±2.2	57.5±3.0	4.0±1.4	21.2±5.2	7.9±3.1
Ours	✓	**82.5**±16.4	83.4±9.1	80.1±9.5	**89.1±13.1**	**90.1±2.2**	**91.8±4.4**	**63.1±2.1**	**71.6±8.0**	67.7±1.9

* Only used for DA, not contributing to FL.

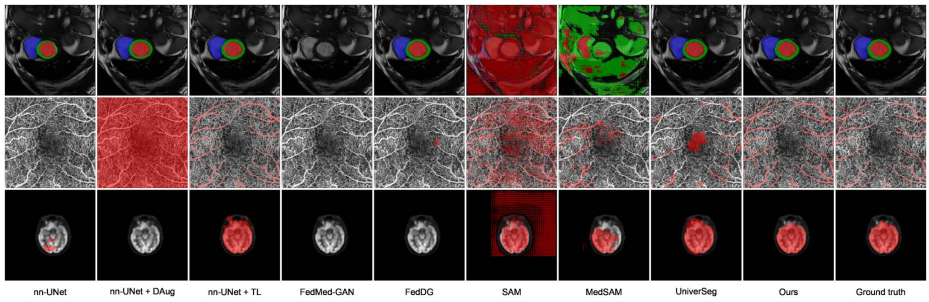

Fig. 2. Comparative analysis of the segmentation results: scanner Siemens (top), OCTA imagery (center), and PET scans (bottom) using different methods.

the initial model, while in UniverSeg it is inputted as a segmentation query support. nnU-Net, nnU-Net+DA, and FedMed-GAN use standard setups, i.e., the former two are supervised with D_s and M_s, the latter is unsupervised with D_s and D_t.

4.3 Results

Table 1 reflects the impact of domain gaps on segmentation performances across methods: overall, the dice score averaged over every domain and method in each scenario drops to 55.7±34.6, 36.5±29.4, and 21.5±26.1 for respectively MS, MM and MO. This underscores the complexity of the tasks under study as many methods show performances below 5%, with a 50% failure rate for MO. In this context, our framework leads to stable systematically high dice score across scenarios, outperforming the competing methods in 6 out of 9 cases (Figure 2). We note that the performance of our approach never drops below 60% dice score. This is different for example from nnU-Net+DAug, which performs positively in MS, but fails in MM. Still concerning MS, our method leads to the best performances for the target Siemens, documented as the most challenging in [7].

5 Conclusion

We propose a novel federated segmentation framework tackling the real-world challenge of missing labeled datasets in multi-centric data, based on the collaborative construction of a multimodal data factory followed by asynchronous domain adaptation. We extensively validated our work on three distinct scenarios of increasing complexity: multi-scanner cardiac MR segmentation, multi-modal skull stripping, and multi-organ vascular segmentation. The results demonstrate the robustness and versatility of our framework, which not only improves reliability across different data domains but also avoids the exchange of raw data or annotations between clients. Our solution leverages labeled and unlabeled data in heterogeneous scenarios, addressing the challenge of data distribution shifts that often hinders the translation of deep learning models into clinical practice.

Acknowledgments. This work has been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project (ANR-19-P3IA-0002), and by the ANR JCJC project I-VESSEG (22-CE45-0015-01).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Universeg: Universal medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21438–21451 (2023)
2. Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreño, M., Albiol, A., Kong, F., Shadden, S.C., Acero, J.C., Sundaresan, V., Saber, M., Elattar, M., Li, H., Menze, B., Khader, F., Haarburger, C., Scannell, C.M., Veta, M., Carscadden, A., Punithakumar, K., Liu, X., Tsaftaris, S.A., Huang, X., Yang, X., Li, L., Zhuang, X., Viladés, D., Descalzo, M.L., Guala, A., Mura, L.L., Friedrich, M.G., Garg, R., Lebel, J., Henriques, F., Karakas, M., Çavuş, E., Petersen, S.E., Escalera, S., Seguí, S., Rodríguez-Palomares, J.F., Lekadir, K.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. IEEE Transactions on Medical Imaging **40**(12), 3543–3554 (2021)
3. Chang, Q., Yan, Z., Zhou, M., Qu, H., He, X., Zhang, H., Baskaran, L., Al'Aref, S., Li, H., Zhang, S., Metaxas, D.N.: Mining multi-center heterogeneous medical data with distributed synthetic learning. Nature Communications **14**(1), 5510 (2023)
4. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. Proceedings of the AAAI Conference on Artificial Intelligence **33**(1), 865–872 (2019)
5. Cremonesi, F., Vesin, M., Cansiz, S., Bouillard, Y., Balelli, I., Innocenti, L., Silva, S., Ayed, S.S., Taiello, R., Kameni, L., et al.: Fed-biomed: Open, transparent and trusted federated learning for real-world healthcare applications. arXiv : 2304.12012 (2023)
6. Fischl, B.: Freesurfer. NeuroImage **62**(2), 774–781 (2012)

7. Full, P.M., Isensee, F., Jäger, P.F., Maier-Hein, K.: Studying robustness of semantic segmentation under domain shift in cardiac mri. In: Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges - 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020. vol. 12592, pp. 238–249. Springer (2020)
8. Galati, F., Falcetta, D., Cortese, R., Casolla, B., Prados, F., Burgos, N., Zuluaga, M.A.: A2V: A semi-supervised domain adaptation framework for brain vessel segmentation via two-phase training angiography-to-venography translation. In: 34th British Machine Vision Conference. pp. 750–751 (2023)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. vol. 27 (2014)
10. Hoopes, A., Mora, J.S., Dalca, A.V., Fischl, B., Hoffmann, M.: SynthStrip: skull-stripping for any brain image. NeuroImage **260**, 119474 (2022)
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. Nature methods **18**(2), 203–211 (2021)
12. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
13. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv : 2304.02643 (2023)
15. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1013–1023 (June 2021)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
17. Ma, Y., Wang, J., Yang, J., Wang, L.: Model-heterogeneous semi-supervised federated learning for medical image segmentation. IEEE Transactions on Medical Imaging pp. 1–1 (2023)
18. Mérida, I., Jung, J., Bouvard, S., Le Bars, D., Lancelot, S., Lavenne, F., Bouillot, C., Redouté, J., Hammers, A., Costes, N.: Cermep-idb-mrxfdg: a database of 37 normal adult human brain [18f] fdg pet, t1 and flair mri, and ct images available for research. EJNMMI research **11**(1), 1–10 (2021)
19. Mescheder, L., Geiger, A., Nowozin, S.: Which training methods for GANs do actually converge? In: Proceedings of the 35th International Conference on Machine Learning. vol. 80, pp. 3481–3490 (2018)
20. Peng, L., Lin, L., Cheng, P., Huang, Z., Tang, X.: Unsupervised domain adaptation for cross-modality retinal vessel segmentation via disentangling representation style transfer and collaborative consistency learning. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2022)
21. Qiu, L., Cheng, J., Gao, H., Xiong, W., Ren, H.: Federated semi-supervised learning for medical image segmentation via pseudo-label denoising. IEEE Journal of Biomedical and Health Informatics **27**(10), 4672–4683 (2023)

22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning. vol. 139, pp. 8748–8763 (2021)
23. Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., Cohen-Or, D.: Encoding in style: A stylegan encoder for image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2287–2296 (2021)
24. Wang, J., Xie, G., Huang, Y., Lyu, J., Zheng, F., Zheng, Y., Jin, Y.: Fedmed-GAN: Federated domain translation on unsupervised cross-modality brain image synthesis. Neurocomputing **546**, 126282 (2023)
25. Wicaksana, J., Yan, Z., Zhang, D., Huang, X., Wu, H., Yang, X., Cheng, K.T.: Fedmix: Mixed supervised federated learning for medical image segmentation. IEEE Transactions on Medical Imaging **42**(7), 1955–1968 (2023)
26. Wu, H., Zhang, B., Chen, C., Qin, J.: Federated semi-supervised medical image segmentation via prototype-based pseudo-labeling and contrastive learning. IEEE Transactions on Medical Imaging **43**(2), 649–661 (2024)
27. Yao, C.H., Gong, B., Qi, H., Cui, Y., Zhu, Y., Yang, M.H.: Federated multi-target domain adaptation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1424–1433 (2022)
28. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., Xu, Z.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Transactions on Medical Imaging **39**(7), 2531–2540 (2020)
29. Zhang, Y., Ling, H., Gao, J., Yin, K., Lafleche, J.F., Barriuso, A., Torralba, A., Fidler, S.: Datasetgan: Efficient labeled data factory with minimal human effort. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10145–10155 (2021)