

# A Client-server Deep Federated Learning for Cross-domain Surgical Image Segmentation

Ronast Subedi<sup>5</sup>, Rebati Raman Gaire<sup>5</sup>, Sharib Ali<sup>4</sup>, Anh Nguyen<sup>3</sup>, Danail Stoyanov<sup>2</sup>, and Binod Bhattarai<sup>1\*</sup>

<sup>1</sup> University of Aberdeen, UK

<sup>2</sup> University College London, UK

<sup>3</sup> University of Liverpool, UK

<sup>4</sup> University of Leeds, UK

<sup>5</sup> NepAI Applied Mathematics and Informatics Institute (NAAMII), Nepal  
{binod.bhattarai}@abdn.ac.uk

**Abstract.** This paper presents a solution to the cross-domain adaptation problem for 2D surgical image segmentation, explicitly considering the privacy protection of distributed datasets belonging to different centers. Deep learning architectures in medical image analysis necessitate extensive training data for better generalization. However, obtaining sufficient diagnostic and surgical data is still challenging, mainly due to the inherent cost of data curation and the need of experts for data annotation. Moreover, increased privacy and legal compliance concerns can make data sharing across clinical sites or regions difficult. Another ubiquitous challenge the medical datasets face is inevitable domain shifts among the collected data at the different centers. To this end, we propose a Client-server deep federated architecture for cross-domain adaptation. A server hosts a set of immutable parameters common to both the source and target domains. The clients consist of the respective domain-specific parameters and make requests to the server while learning their parameters and inferencing. We evaluate our framework in two benchmark datasets, demonstrating applicability in computer-assisted interventions for endoscopic polyp segmentation and diagnostic skin lesion detection and analysis. Our extensive quantitative and qualitative experiments demonstrate the superiority of the proposed method compared to competitive baseline and state-of-the-art methods. Codes are available at: <https://github.com/thetna/distributed-da>.

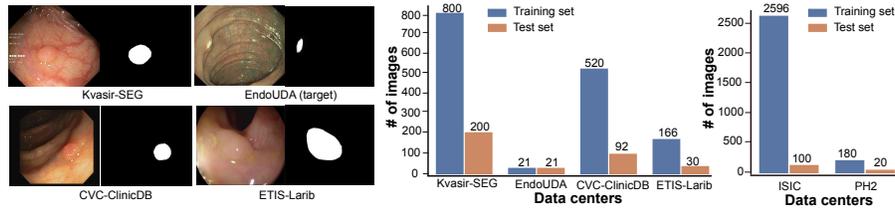
**Keywords:** Domain Adaptation · Federated Learning · Decentralised Storage · Privacy

## 1 Introduction

The deployment of artificial intelligence (AI) technology in medical image analysis is rapidly growing, and training robust deep network architectures demands

---

\* Corresponding author



**Fig. 1.** Sample training examples collected from various centres for polyp segmentation (left); Sizes of training and test set at different centres for polyp segmentation (middle) and skin lesion segmentation (right).

millions of annotated examples. Despite significant progress in establishing large-scale medical datasets, these are still limited in some clinical indications, especially in surgical data science and computer-assisted interventions [21]. Scaling training data needs multi-site collaboration and data sharing [1], which can be complex due to regulatory requirements (e.g. the EU General Data Protection Regulation [31], and China’s cyber power [13]), privacy, and legal concerns. Additionally, even after training, practical AI model deployment in the clinic will require fine-tuning or optimization to local conditions and updates [11]. Therefore, architectures trained in federated and distributed ways to tackle cross-domain adaptation problems are critical. Yet, developing such architectures has challenges [25].

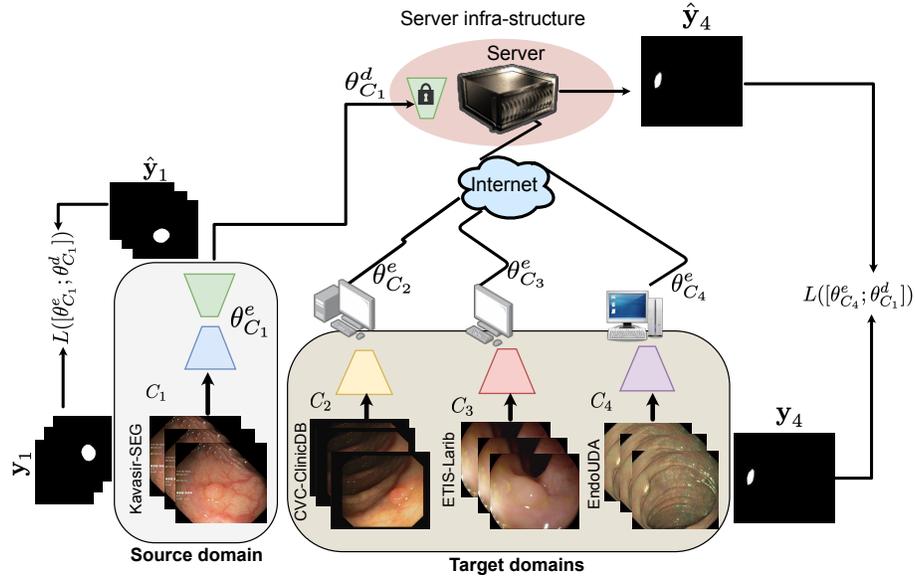
Several works [29,7,8] have been proposed to tackle the problem of cross-domain adaptation in medical imaging. However, these methods require raw source and target domain data and cannot address the ever-increasing privacy concerns in sharing medical data. To circumvent the problem of privacy protection, there is a lot of research interest growing in Federated Learning (FL) in the medical domain [25,18,27,30,28,15,24]. Some methods even rely on synthetic data [10] to avoid sharing real data. For more details, we refer readers to a survey [23] on federated learning for smart health care. The common drawback of most existing methods [25,18,15,24] is that these methods are not designed for the domain shift problem. The most common topology in the FL workflow is averaging the local gradients (FedAvg) at the center and peer-to-peer gradient (FedP2P) sharing. These architectures are effective when data are independent and identically distributed (IID) in every client. In reality, domain shift is quite prevalent as data collected at different centers tend to be center specific. In Fig. 1, we can see the training examples for polyp segmentation collected at different centres. These examples show the discrepancy in lighting, camera pose and modalities in different centers. Some recent works, such as by Guo et al. [9] and FedDG by Liu et al. [19], address cross-domain problems in FL. However, [9] limits to a source-target pair at a time. Also, they employed adversarial loss to align the parameters, which is difficult to optimize. Similarly, FedDG [19] shares the information between the sources in the form of amplitudes of images. Their evaluation is limited to fundus images and MRI.

To tackle the problems of cross-domain adaptation and privacy protection in surgical image segmentation, we propose a simple yet effective Client-server FL architecture consisting of a server and multiple clients, as shown in Fig. 2. A server hosts a set of *immutable* task-specific parameters common to all the clients. Whereas every client requests the server to learn their domain-specific parameters locally and make the inference. In particular, every client learns an encoder’s parameters to obtain an image’s latent representation. These latent representations and ground truth masks are sent to the server. The decoder deployed on the server makes the predictions and computes the loss. The gradients are computed and updated only on the encoder to align the client’s features with task-specific parameters hosted on the server. Aligning domain-specific parameters to common parameters helps diminish the gap between the source and target domains. We can draw an analogy between our framework with public-key cryptography. A client’s network parameter is equivalent to a private key, and the decoder’s parameters shared on the server are equivalent to the public key. Thus a client only with access to its private key can transfer its latent vector to the server containing the public key to obtain the semantic mask. Distributed storage of the parameters diminishes the risk of model parameter theft and adversarial attacks [20]. Moreover, each client communicates to the server only via a latent image representation, which prevents exposing the information of the raw data collected on the client side. It is possible to encrypt data transfer between the server and clients to make it more secure. Finally, the server receives only fixed latent dimension representations, making it agnostic to the client’s architecture. This enables clients to communicate with the server concurrently, improving efficiency. Likewise, none of the centres can modify the parameters deployed on the server; this would prevent the memorisation of client-specific information and parameter poisoning on the server [17].

To sum up, we propose a Client-server Federated Learning method for cross-domain surgical image segmentation. We applied our method to two multi-centre datasets for endoscopic polyp and skin lesion segmentation. We compared with multiple baselines, including recent works on cross-domain FL [19,9] and obtained a superior performance.

## 2 Method

**Background:** We consider a scenario where we have  $C_1, C_2, \dots, C_n$  represent  $n$  number of different institution’s centres located at various geographical regions. Each centre collects its data in the form of tuple  $(\mathbf{x}, \mathbf{y})$  where  $\mathbf{x} \in \mathbb{R}^{w \times h \times c}$ ,  $\mathbf{y} \in \mathbb{R}^{w \times h}$ , where,  $w, h, c$  represent the width, height, and number of channels of an image. The annotated examples collected at the different centers are not IID due to variations in the illumination, the instruments used to acquire data, the ethnicity of the patients, the expertise of the clinician who collects the data, etc. We denote the total number of annotated pairs in each centre by  $N_n$ . In this paper, one of the major goals is to address the problem of domain adaptation, avoiding the need for the sharing of raw data to protect privacy.



**Fig. 2.** The schematic diagram of the proposed framework. There are three major components: Source domain, Target domains, and Server infrastructure to share common parameters.

**Learning Source Domain Parameters:** First, we train a semantic segmentation network on the source data. In Figure 2, the Source domain block shows the training of source domain/centre parameters. For us, data collected on Centre 1,  $C_1$  is source data. We employ fully-convolutional encoder-decoder architecture. Such architectures are quite popular for semantic segmentation [2,26]. With the randomly initialised parameters, we minimise the objective of the Equation 1. In Equation 1,  $\theta_{C_1}^e$  and  $\theta_{C_1}^d$  represent the learnable parameters of the encoder and decoder, respectively.

$$L([\theta_{C_1}^e; \theta_{C_1}^d]) = -\frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{j=1}^W \sum_{k=1}^H \mathbf{y}_{ijk} \log \hat{\mathbf{y}}_{ijk} + (1 - \mathbf{y}_{ijk}) \log(1 - \hat{\mathbf{y}}_{ijk}) \quad (1)$$

**Setting-up Server Infra-structure:** Figure 2 Server infra-structure block shows the setting up of the server infrastructure. Once we learn the parameters of the network from the source ( $C_1$ ) data set, we upload the decoder ( $\theta_{C_1}^d$ ) parameters on the server to share with every target client. The decoder module specializes to segment anatomies, given the encoder module’s latent vector representation of the input image. As this segmentation task is common to all the centres, we propose to use a single decoder for all the centres. The previous works on cross-modal [4] and cross-feature [3] representations learning for cross-

domain classification in computer vision employed the idea of sharing Convolutional Neural Networks’ top layers parameters. However, none of these methods were employed in Federated Learning. The idea of sharing top layers parameters is in contrast to the conventional transfer learning [12] where the parameters are initialised with a model pre-trained on Imagenet and fine-tuned only the task-specific fully connected layers. We freeze the shared decoder parameters of the source. This arrangement brings advantages for privacy protection by preventing weight poisoning attacks [17]. Weight poisoning attacks alter the pre-trained model’s parameters to create a backdoor. The parameters of the encoders can be shipped to the target client as per demand. Sometimes, the clients may demand these parameters to initialize their local networks when the training data is very small.

**Federated Cross-domain Adaptation:** Target centres other than the source centre deploy only the encoder network of their choice. In Figure 2, the target domain block depicts it. Every centre feeds its images to its encoder network during training, generating the respective latent representations. The latent representation and the ground truth ( $\mathbf{y}_i$ ) mask from each target centre are pushed to the server where the pre-trained source decoder,  $\theta_{C_1}^d$ , is placed. The decoder feeds the latent representation, which predicts the output segmentation labels ( $\hat{\mathbf{y}}_i$ ). We learn the parameters of the target encoders ( $\theta_{C_i}^e$ ) to minimize the objective given in Equation 2. Since the decoder parameters are frozen and shared with every client, only the target encoder’s parameters are updated on the client side. This helps to align the latent representations to that of the source decoder’s parameters and maximises the benefit from the task-specific discriminative representations learned from the large volume of source data.

$$L([\theta_{C_i}^e; \theta_{C_1}^d]) = -\frac{1}{N_i} \sum_{i=2}^{N_i} \sum_{j=1}^W \sum_{k=1}^H \mathbf{y}_{ijk} \log \hat{\mathbf{y}}_{ijk} + (1 - \mathbf{y}_{ijk}) \log(1 - \hat{\mathbf{y}}_{ijk}) \quad (2)$$

$$\forall i \in 2, \dots, n$$

The only thing that matters for target centres to communicate to the server is the fixed dimension of latent representations of an image. Thus, our architecture gives the flexibility of deploying the various sizes of networks on the client side based on available computing resources. And it is also entirely up to the target centres whether they want to initialize the parameters of the encoder using the parameters of the source domain. If the number of training examples is extremely few, then initialization using the pre-trained model’s weight can prevent overfitting.

### 3 Experiments

**Data sets and Evaluation Protocol:** We applied our method in two benchmark datasets: endoscopic polyp segmentation and skin lesion segmentation. The

**polyp segmentation dataset** contains images collected at four different centres. Kvasir-SEG [14] data set makes the source centre ( $C_1$ ) in our experiment. It has 800 images in the train set and 200 in the test set. These high-resolution images acquired by electromagnetic imaging systems are made available to the public by Simula Lab in Norway. Similarly, the EndoUDA-target data set makes the first target domain ( $C_2$ ) in our experiment, consisting of 21 images in both the training and testing sets [5]. Our experiment’s second target domain centre ( $C_3$ ) consists of images from the CVC-ClinicDB dataset made available to the research community by a research team in Spain. There are 520 images in the train set and 92 in the test set. Finally, the ETIS-Larib dataset released by a laboratory in France makes our third target domain data set ( $C_4$ ). This data set consists of 166 in the train set and 30 images in the test set. These data sets were curated at different time frames in different geographical locations.

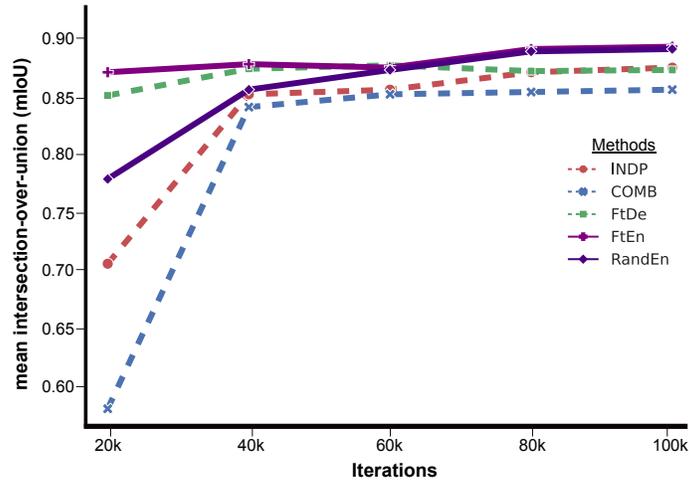
Data	Centres	mIOU							
		INDP	COMB	[16]	[19]	FtDe	[9]	RandEn	FtEn
Endo.	Kvasir-SEG ( $C_1$ , source)	80.3	81.0	82.3	73.5	N/A	80.5	80.3	80.3
	EndoUDA ( $C_2$ )	52.0	57.5	53.1	29.7	59.9	61.7	50.6	62.0
	CVC-ClinicDB ( $C_3$ )	88.3	87.8	86.8	74.5	85.8	83.0	89.1	88.4
	ETIS-Larib ( $C_4$ )	62.1	66.9	61.4	70.8	65.1	71.7	64.3	69.9
Skin	ISIC ( $C_1$ , source)	81.3	75.7	84.9	N/A	N/A	NA	81.3	N/A
	PH2 ( $C_2$ )	88.4	88.3	88.4	N/A	88.0	NA	89.6	89.4

**Table 1.** mIoU scores on Endoscopic Polyp Segmentation Data sets (upper block) and Skin Lesion Segmentation (lower block).

For **skin lesion segmentation**, we took data set collected at two different centres: ISIC(International Skin Imaging Collaboration) [6] and PH2 [22]. In ISIC, there are 2596 training examples and 102 test examples. The PH2 database is curated through a joint research collaboration between the Universidade do Porto, Tecnico Lisboa, and the Dermatology Service of Hospital Pedro Hispano in Matosinhos, Portugal. In this data set, there are only 180 training examples and 20 testing examples. We consider ISIC and PH2 source and target domain, respectively. We use the mean Intersection of Union (mIoU) and dice score for both datasets for quantitative evaluations. Qualitative comparisons also validate our idea.

**Baselines:** We have compared the performance of our method with several competitive baselines, including both non-federated and federated frameworks. One of the naive baselines is to train a model for each target centre independently (*INDP*). The models of the centres with less training data overfit. Another configuration is creating a data pool by combining the training data (*COMB*) from all the centres and training a single model. However, this method does not address any of the issues regarding privacy and compliance. Another viable option is to adapt a pre-trained model to a new domain by fine-tuning the parameters of the latter layers (*FtDe*). We also compared our method with competitive feder-

ated learning algorithms. FedAvg [16] averages the gradients computed in every center and shares the average gradients with the clients. This method ignores the non-IID nature of data from different centres. FedDG [19] is another Federated Learning method for domain adaption published at CVPR 2021. Finally, we also compared with another recent work by Guo et al. [9] for federated learning for multi-institutional data published at CVPR 2021. Our methods have two variants: initialising clients’ parameters randomly (**RandEn**) and with the source’s parameters (**FtEn**)



**Fig. 3.** Curves show the learning behaviour of different methods on PH2 data set.

**Implementation Details and Learning Behaviour:** We implement our algorithms on PyTorch framework. All the images were resized to the dimension of  $418 \times 418$ . For optimization, we employ Adam optimizer with values of  $\beta_1$  and  $\beta_2$  set to 0.9 and 0.999 respectively. We initialize learning rate to  $2e-4$  and set the decaying of the learning rate every 25k iterations. Taking UNet as a base architecture, we train the networks for 100k iterations and save the best-performing checkpoints on the validation set and report the performance on the test set. Figure 3 summarises the learning behaviour of the different methods for the first 100k iterations on PH2 data set, a target domain for skin lesion segmentation. The solid lines are our methods, and the dashed lines are the compared methods. The smooth curves demonstrate that our methods are easy to optimise the parameters.

**Quantitative Evaluations:** Table 3 shows the quantitative performance comparison. In the table, the last two grey-shaded columns show the performance of our methods. Our method outperforms *INDP* in every target centre. This signifies the importance of domain adaptation by our method. Compared to the other Federated Learning methods, our methods obtain the highest performance on

2/3 of target centres and are competitive on the third one for endoscopic polyp segmentation. On skin lesion segmentation, our method surpassed all the compared baselines and the recent competitive Federated Learning methods. We have compared the performance on the dice score, too and obtained a similar performance (See Supplementary). We have also evaluated the performance with varying sizes of the target domain/centre encoders. Please see the appendix for the details.

**Qualitative Evaluations:** Figure 4 shows the qualitative performance comparisons between the baselines and the proposed methods on the target domains. Rows 2-4 (inclusive) are from endoscopy benchmarks, and the last row is from skin benchmarks. FedAvg, our closest work, fails to generalise well on target domains (see ETIS-Larib and CVC-ClinicDB). Whereas our method is consistent in every target domain. These results further validate that our method is superior to the others. You can find more examples in the appendix section.

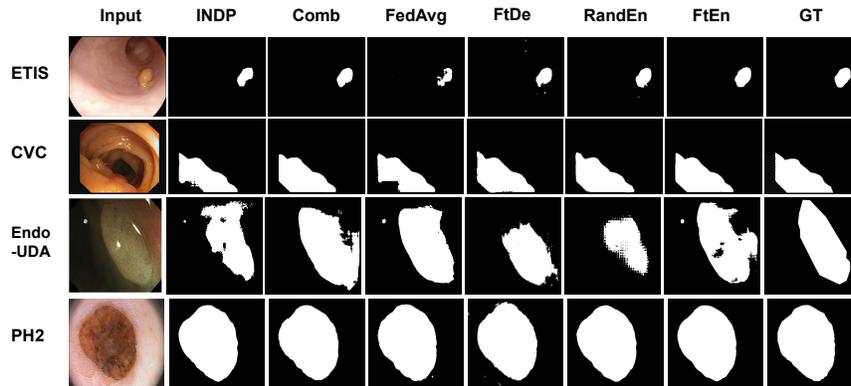


Fig. 4. Qualitative Comparisons

**Computational Complexity:** For INDP and FedAvg, the parameters of both the encoders and decoders grow in  $O(n)$  with the number of centres. Similarly, for FtDe and our method, the parameters of the encoder grow in  $O(n)$ , while the parameters of the decoder are constant, i.e.,  $O(1)$ . Although the growth of the parameters for both the encoder and decoder for COMB is  $O(1)$ , it does not address any privacy concerns. From these, our method is computationally less expensive and has high privacy protection.

## 4 Conclusions

In this paper, we presented a client-server Federated Learning architecture for cross-domain surgical image segmentation. Our architecture addresses the cross-domain adaptation problem without sharing the raw images. Moreover, sharing

only a part of the parameters from the source domain enhances privacy protection. Extensive experiments on two benchmarks from various data centres demonstrated improved cross-domain generalisation and privacy protection over the baselines and the competitive contemporary method.

## 5 Acknowledgements

This work is partly supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145Z/16/Z]; Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1]; The Royal Academy of Engineering Chair in Emerging Technologies scheme; and the EndoMapper project by Horizon 2020 FET (GA 863146).

## References

1. Ali, S., Jha, D., Ghatwary, N., Realdon, S., Cannizzaro, R., Salem, O.E., Lamarque, D., Daul, C., Riegler, M.A., Anonsen, K.V., et al.: a multi-centre polyp detection and segmentation dataset for generalisability assessment. *Scientific Data* **10** (2023)
2. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *T-PAMI* **39** (2017)
3. Bhattarai, B., Sharma, G., Jurie, F.: Deep fusion of visual signatures for client-server facial analysis. In: *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*. pp. 1–8 (2016)
4. Castrejon, L., Aytar, Y., Vondrick, C., Pirsiavash, H., Torralba, A.: Learning aligned cross-modal representations from weakly aligned data. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2940–2949 (2016)
5. Celik, N., Ali, S., Gupta, S., Braden, B., Rittscher, J.: Endouda: A modality independent segmentation approach for endoscopy imaging. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 303–312. Springer (2021)
6. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368* (2019)
7. Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R., de Leeuw, F.E., Tempny, C.M., Van Ginneken, B., et al.: Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*. pp. 516–524. Springer (2017)
8. Gholami, A., Subramanian, S., Shenoy, V., Himthani, N., Yue, X., Zhao, S., Jin, P., Biros, G., Keutzer, K.: A novel domain adaptation framework for medical image segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018*. Springer (2019)
9. Guo, P., Wang, P., Zhou, J., Jiang, S., Patel, V.M.: Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In: *CVPR* (2021)
10. Hu, S., Goetz, J., Malik, K., Zhan, H., Liu, Z., Liu, Y.: Fedsynth: Gradient compression via synthetic data in federated learning. *arXiv preprint arXiv:2204.01273* (2022)
11. Hu, Y., Jacob, J., Parker, G.J., Hawkes, D.J., Hurst, J.R., Stoyanov, D.: The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. *Nature Machine Intelligence* (2020)
12. Huh, M., Agrawal, P., Efros, A.A.: What makes imagenet good for transfer learning? *arXiv preprint arXiv:1608.08614* (2016)
13. Inkster, N.: *China’s cyber power*. Routledge (2018)
14. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., Lange, T.d., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: *International Conference on Multimedia Modeling*. pp. 451–462. Springer (2020)
15. Karargyris, A., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Bala, S., Beutel, D.J., Bittorf, V., Chaudhari, A., Chowdhury, A., et al.: Medperf: Open

- benchmarking platform for medical artificial intelligence using federated evaluation. arXiv preprint arXiv:2110.01406 (2021)
16. Konečný, J., McMahan, B., Ramage, D.: Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv:1511.03575 (2015)
  17. Kurita, K., Michel, P., Neubig, G.: Weight poisoning attacks on pretrained models. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 2793–2806 (2020)
  18. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine* **37** (2020)
  19. 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. pp. 1013–1023 (2021)
  20. Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., Lu, F.: Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition* **110**, 107332 (2021)
  21. Maier-Hein, L., Eisenmann, M., Sarikaya, D., März, K., Collins, T., Malpani, A., Fallert, J., Feussner, H., Giannarou, S., Mascagni, P., et al.: Surgical data science—from concepts toward clinical translation. *Medical image analysis* (2022)
  22. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R., Rozeira, J.: Ph 2-a dermoscopic image database for research and benchmarking. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp. 5437–5440. IEEE (2013)
  23. Nguyen, D.C., Pham, Q.V., Pathirana, P.N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., Hwang, W.J.: Federated learning for smart healthcare: A survey. *ACM Computing Surveys (CSUR)* **55** (2022)
  24. Parekh, V.S., Lai, S., Braverman, V., Leal, J., Rowe, S., Pillai, J.J., Jacobs, M.A.: Cross-domain federated learning in medical imaging. arXiv preprint arXiv:2112.10001 (2021)
  25. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al.: The future of digital health with federated learning. *NPJ digital medicine* (2020)
  26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
  27. Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R., et al.: Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports* **10**(1), 1–12 (2020)
  28. Shen, Y., Zhou, Y., Yu, L.: Cd2-pfed: Cyclic distillation-guided channel decoupling for model personalization in federated learning. In: CVPR (2022)
  29. Swati, Z.N.K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., Lu, J.: Brain tumor classification for mr images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics* (2019)
  30. Tramel, E.W.: Siloed federated learning for multi-centric histopathology datasets. In: Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning: Second MICCAI Workshop, DART 2020, and First MICCAI Workshop, DCL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru (2020)
  31. Voigt, P., Von dem Bussche, A.: The eu general data protection regulation (gdpr). *A Practical Guide*, 1st Ed., Cham: Springer International Publishing **10**(3152676), 10–5555 (2017)

## A Appendix

**Varying Encoder Size:** To further validate the robustness of our proposed framework, we trained the networks with different numbers of learnable parameters in the encoder module as shown in table 2. We vary the parameters by adding or removing the constituting layers in the encoding blocks of the network. We designate the conventional encoder of the UNet architecture as an encoder with medium size. The learnable parameters in the medium encoder are approximately 17 million. Table 3 depicts the architecture of a general encoder. It consists of three down-sampling layers, represented as *Down Block*. We vary the number of layers in the Down Blocks of the general encoder as per the availability of labeled data and computing resources in the particular centre.

Data	Centres	Size	Trainable Parameters	mIoU	
				INDP	RandEnc(Ours)
Polyp	ETIS-Larib	Small	6,311,616	61.6	62.9
		Medium	17,080,896	62.1	64.3
		Large	27,850,176	62.5	64.7
Skin	PH2	Small	6,311,616	88.4	88.4
		Medium	17,080,896	88.5	89.6
		Large	27,850,176	89.4	89.8

**Table 2.** Result on Endoscopic Polyp Segmentation Data sets (upper block) and Skin Lesion Segmentation (lower block) with different sizes of encoder networks.

Operations	Output Size
Input Image	$c \times h \times h$
Down Block	$c \times \frac{h}{2} \times \frac{h}{2}$
Down Block	$c \times \frac{h}{4} \times \frac{h}{4}$
Down Block	$c \times \frac{h}{8} \times \frac{h}{8}$

**Table 3.** Architecture of the general encoder.

The proposed federated distributed framework for domain adaptation provides flexibility to choose different network architectures to learn a common latent representation of images. These architectures can be designed based on the data size and available resources at a particular center. The performance evaluation results of various encoder sizes on polyp segmentation and skin lesion segmentation are shown in table 2. From the table, we can observe that segmentation performance for various sizes of the encoder is higher when trained using our framework than training independently for a specific center.

**Equations for Gradient Update:** Equations 3 and 4 show mathematical formulation to update gradients of decoder and encoder modules in the proposed framework. Equation 3 updates the gradient of the decoder in the source center. Equation 4 represents the mechanism of updating gradients of each encoder

module in every center. Here,  $\alpha$  is the center-specific learning rate. Please note the parameters of the decoder remain the same for every target centre.  $L$  denotes the loss function.

$$\theta_{C_1}^d = \theta_{C_1}^d - \alpha_1 \times \frac{\partial L([\theta_{C_1}^e; \theta_{C_1}^d])}{\partial \theta_{C_1}^d}; \quad (3)$$

$$\theta_{C_i}^e = \theta_{C_i}^e - \alpha_i \times \frac{\partial L([\theta_{C_i}^e; \theta_{C_1}^d])}{\partial \theta_{C_i}^e}; \forall i \in 1, \dots, n \quad (4)$$

**Additional Qualitative Results:** Figure 5 depicts the segmentation results on Endoscopic polyp data from the target domain centres. The first four rows are from the ETIS-Larib dataset, the next three are from the CVC-ClinicDB data set, and the last two are from EndoUDA. Columns **RandEn** and **FtEn** show the results of our method.

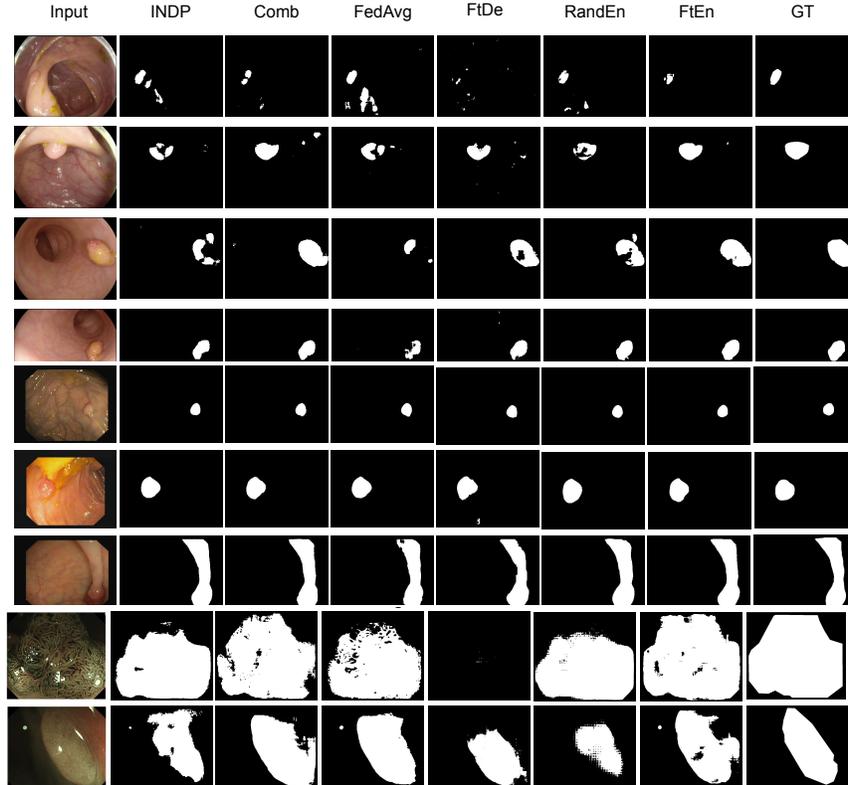


Fig. 5. Additional Qualitative Comparison