

FedST: Federated Style Transfer Learning for Non-IID Image Segmentation

Boyuan Ma¹, Xiang Yin¹, Jing Tan¹, Yongfeng Chen¹, Haiyou Huang¹, Hao Wang¹, Weihua Xue²,
Xiaojuan Ban^{1,3,*}

¹School of Intelligence Science and Technology, Beijing Advanced Innovation Center for Materials Genome Engineering, Shunde Innovation School, Key Laboratory of Intelligent Bionic Unmanned Systems, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China

²School of Materials Science and Technology, Liaoning Technical University, Liaoning 114051, China

³Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang 110004, China
banxj@ustb.edu.cn, mbytony@ustb.edu.cn

Abstract

Federated learning collaboratively trains machine learning models among different clients while keeping data privacy and has become the mainstream for breaking data silos. However, the non-independently and identically distribution (i.e., Non-IID) characteristic of different image domains among different clients reduces the benefits of federated learning and has become a bottleneck problem restricting the accuracy and generalization of federated models. In this work, we propose a novel federated image segmentation method based on style transfer, FedST, by using a denoising diffusion probabilistic model to achieve feature disentanglement and image synthesis of cross-domain image data between multiple clients. Thus it can share style features among clients while protecting structure features of image data, which effectively alleviates the influence of the Non-IID phenomenon. Experiments prove that our method achieves superior segmentation performance compared to state-of-art methods among four different Non-IID datasets in objective and subjective assessment. The code is available at <https://github.com/YoferChen/FedST>.

Introduction

Federated learning (FL) has become one of the mainstream learning paradigms to take advantage of multiple data sources while protecting data privacy (Li et al. 2019; Voigt and Von dem Bussche 2017). In FL, clients collaboratively train a model while keeping the training data decentralized (Kairouz et al. 2021). It provides an opportunity for collaborative learning across multiple clients without infringing privacy (Jiang et al. 2020), which enables it to be used in many fields where data security is of importance, such as medical data analysis (Brisimi et al. 2018; Liu et al. 2021) and wireless communication (Niknam, Dhillon, and Reed 2020).

Despite the rich opportunities offered by FL, non-independent and identically distributed (Non-IID) data among different clients often bring great challenges to FL, resulting in worse performance and slower convergence (Kopparapu and Lin 2020; Kopparapu, Lin, and Zhao 2020). Especially, the Non-IID phenomenon, caused by cross-domain data on image segmentation tasks, is a common issue in practical FL applications, as shown in Figure 1. For instance, in materials science, researchers usually analyze dif-

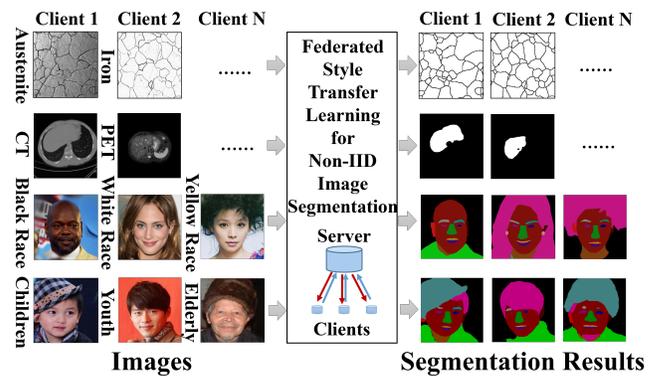


Figure 1: The Non-IID phenomenon, caused by cross-domain data in different clients, is a common issue in practical FL-based segmentation tasks, such as the segmentation of material microscopic images, medical images, and face images. While the images of different clients have similar underlying structures, the different stylistic appearance information among clients will result in worse performance and slower convergence of an FL model.

ferent microscopic images of different materials. Austenitic alloys are composed of iron and many other elements. It renders a different appearance compared to pure iron in microscopic images, even though they have similar grain structures. On the other hand, in clinical practice, doctors often use different image modalities to observe the same human tissue and organ, such as the CT and MRI image modality of the liver. In natural images, cross-domain image segmentation under an FL framework is also common, such as face segmentation of different races or different ages. In these scenarios, while the images in all clients have similar underlying structures, the Non-IID problem caused by different stylistic appearances still hinders the performance of FL model.

To solve the Non-IID problem in image segmentation tasks in FL, our insight is that most participated clients generally have similar underlying structure information and different style information, as shown in Figure 1 and Figure 2. And the goal of image segmentation is to remove the style information in the image and get the segmentation re-

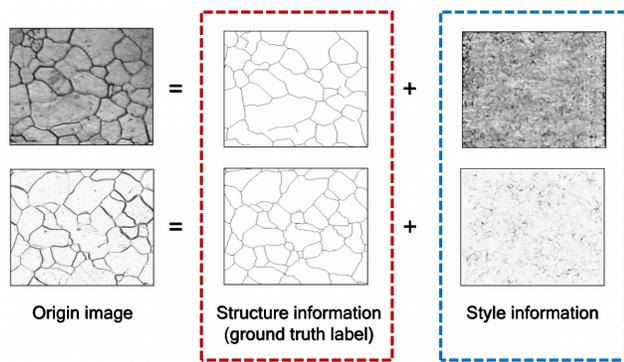


Figure 2: An example of a material microscopic image decomposition.

sult (structure information), which is required by the downstream application. Thus, in most cases, structure information is generally useful to users, and style information is not. Take grain structure information as an example, the size and shape distribution is significant information to build the intrinsic relationship between composition, structure, process, and properties, which is fundamental to material design (Hu et al. 2017). The style information, such as color and brightness in the material microscopic images, is caused by different element compositions and preparation processes. In this regard, the structure information in the material microscopic images is user-concerned, while the style information is not. This observation is also applicable to medical image segmentation: researchers generally want to analyze the shape of organs maintained in the structure information (or ground truth label), while the style information caused by different modalities is not user-concerned information.

Based on this observation, we develop a novel network, dubbed as federated style transfer network (FedST). It uses a denoising diffusion probabilistic model (DDPM) to extract and exchange style information among clients, and keep structural information private. Thus each client can access different data distributions within other clients, which alleviates the influence of the Non-IID phenomenon.

Our method has two variations, FedST-separate and FedST-join, depending on whether the diffusion model participates in federation aggregation. The FedST-separate trains a style transfer generator in each client separately and trains a segmentation module across all clients by federation aggregation. The FedST-join trains a unified controllable generator and a segmentation module by federation aggregation. These two variations both enable feature disentanglement and image synthesis of cross-domain image data between multiple clients. In addition, while our method shares the style information among clients, it is maintained in the network parameters and is inexplicable to users. And even if being attacked, it can't restore the structural information, thus it has little impact on users.

In general, our main contributions are highlighted as follows:

- We propose a novel network to tackle the non-IID prob-

lem in FL segmentation, FedST, which uses a denoising diffusion probabilistic model to achieve feature disentanglement and image synthesis of cross-domain image data between multiple clients. Thus alleviating the influence of the Non-IID phenomenon.

- We discuss the two variants of the proposed FedST, including FedST-separate and FedST-join, depending on how to train a style transformation network.
- We conduct extensive experiments among four different Non-IID tasks, including material microscopic image segmentation, medical image segmentation, face segmentation with different races, and face segmentation with different ages. Our method achieves superior performance compared to SOTA methods in objective and subjective assessments. And even if structure information between clients is not similar, sharing style information will also lead to performance improvement.

Related Work

Federated learning (Konečný et al. 2016; Yang et al. 2019; McMahan et al. 2017) which provides a promising privacy-preserving solution to train machine learning models on private data across massively distributed devices has emerged many research interests. (McMahan et al. 2017) first proposed the federated averaging algorithm (FedAvg) for communication-efficient federated training of deep networks which has become a basic paradigm for federated learning.

There have been quite some studies trying to improve FedAvg on non-IID data. Those studies can be divided into four categories: optimization-based, aggregation-based, architecture-based, and data distribution based.

As for the optimization-based methods, FedProx (Li et al. 2020) introduces a proximal term into the loss function during local training. The proximal term constrains the L2-norm distance of the local model and global model which reduce the parameters divergence between different clients caused by statistical heterogeneity. SCAFFOLD (Karimireddy et al. 2019) corrects the local updates by introducing control variates. Each client has a local control variate while sharing a global control variate, the control variates are updated by each client during local training. The difference between the local control variate and the global control variate is used to correct the model gradients in local training. FedDyn (Acar et al. 2021) adds penalty terms to the local loss function and performs global updates to guide the model of each client converges to the global optimal.

As for the aggregation-based methods, FedDF (Lin et al. 2020) uses model distillation to perform federated aggregation, it can not only aggregate heterogeneous models but also enhance the model performance while facing data heterogeneity. FedAMP (Huang et al. 2021) encourages collaboration between users with similar data distribution and inhibits collaboration between users with large data distribution differences to get the client's optimal model. FedAvgM (Hsu, Qi, and Brown 2019) applies additional global updates on the server side to perform a robust aggregation.

As for the architecture-based methods, (Qu et al. 2022)

find self-attention-based architectures are more robust to distribution shifts and hence improve federated learning over heterogeneous data. By switching to the transformer, federated learning gains in multiple image classification tasks.

As for the data distribution-based methods, a data-sharing strategy is used to address this problem (Zhao et al. 2018). They distribute a global sharing dataset, which consists of a public dataset of specific task, to every client in order to slightly reduce the EMD (Rubner, Tomasi, and Guibas 2000) between different client’s local models. Experiment shows that little globally shared data brings obviously performance improvement, but this kind of data-sharing method is not suitable for some researchers due to the lack of large-scale public datasets. (Wu et al. 2022) shares the mean and covariance of local data to perform contrastive learning, which is hard to model cross-domain data. (Dong and Voiculescu 2021) used InstaHide to encrypt the raw images and used the encrypted image to perform federated training which reduces the risk of data privacy leakage but leads to drop in performance (Huang et al. 2020). (Zhu, Hong, and Zhou 2021) used a generator to learn feature representations of global data by distilling aggregated global models, but it needs to gather raw image and label which may lead to information leakage.

In this work, we study from the perspective of reducing data distribution variance by performing image-to-image translation between cross-domain image data to achieve sharing of style information. The image-to-image translation is a technology to translate an input image from one domain to another domain. Many researchers use adversarial learning to perform image-to-image translation due to the automatic adaption ability in different domains (Mirza and Osindero 2014; Sasaki, Willcocks, and Breckon 2021). Pix2pix (Isola et al. 2017) uses image-conditional GANs (Mirza and Osindero 2014) for image-to-image translation. Diffusion Models (Ho, Jain, and Abbeel 2020) has recently emerged with a lot of impressive research on image generation (Austin et al. 2021; Kadkhodaie and Simoncelli 2020), which have recently been shown to outperform GAN in image synthesis (Dhariwal and Nichol 2021). Thus the great ability of image synthesis gives the opportunity to reduce the Non-IID problem caused by cross-domain data in FL architecture.

Method

We propose FedST from the perspective of information-sharing. Since the Non-IID phenomenon limits the performance of federated learning, FedST aims to decrease the data distribution difference around clients by sharing style features and keeping user-concerned structure information private. The style information is shared across clients by the image-to-image translation model which learns the ability to translate the client’s ground truth segmentation label (structure information) to different image domain data, the generated data then act as the data augmentation to train the image segmentation model. Our method comes in two variants, FedST-separate and FedST-join, based on whether the image-to-image translation model participates in federation

aggregation. In the following, we present the network architecture, the difference between the two variants, the local learning objective, and the communication cost analysis.

Network Architecture

As shown in Figure 3, the proposed network has two components: a style transfer generator (G), and an image segmentation model (S).

The style transfer generator is used to extract the style information from data by training an image-to-image diffusion model (Saharia et al. 2022a). Diffusion models (Ho, Jain, and Abbeel 2020) denoise samples of the Gaussian distribution over a series of time steps to fit samples of the real data distribution. Conditional diffusion models (Saharia et al. 2022b) permit the conditional signal as an input to the denoising process. The style transfer generator is a kind of conditional diffusion model that uses the label as a conditional input. The network architecture of the style transfer generator is 256×256 class-conditional U-Net model (Saharia et al. 2022a).

In addition, We use U-Net as image segmentation architecture.

FedST-separate The difference between FedST-separate and FedST-join is whether the style transfer model participates in the federation aggregation process. In FedST-separate, each client separately trains their own style transfer model using their local label and local domain image, uploads it to the server, and saves it to the style store. Then, the style store is shared with all clients. Thus each client can generate synthetic cross-domain data using their local label. During each round of federated training of the segmentation model, each client randomly selects a style transfer model from another domain and generates synthetic cross-domain images to participate in FL as data augmentation. It achieves the sharing of style information around clients and alleviates the influence of the Non-IID phenomenon.

FedST-join FedST-separate needs to train a unique style transfer model for each client, which increase the overall parameters of the whole structure. Thus we further propose FedST-join method. It uses federated aggregation strategy to jointly train a unified style transfer model and segmentation model. And in order to perform cross-domain style information sharing by one global style transfer model, we add a controllable generate module after each convolution of the generator, which uses a one-hot encoding domain vector to identify clients, as shown in Figure 3. During forward propagation, a data pair that consist of a domain vector and a segmentation label is fed into the style transfer model. The domain vector controls the domain of generated image through the controllable generate module, which is mapped to a weight vector and a bias vector and performs channel-wise multiply and channel-wise addition respectively with input feature to get the stylistic feature. Thus, the input segmentation label produces the local structure information while the domain vector controls the cross-domain style generation. During each round of federated training, each client randomly generates synthetic cross-domain data as data aug-

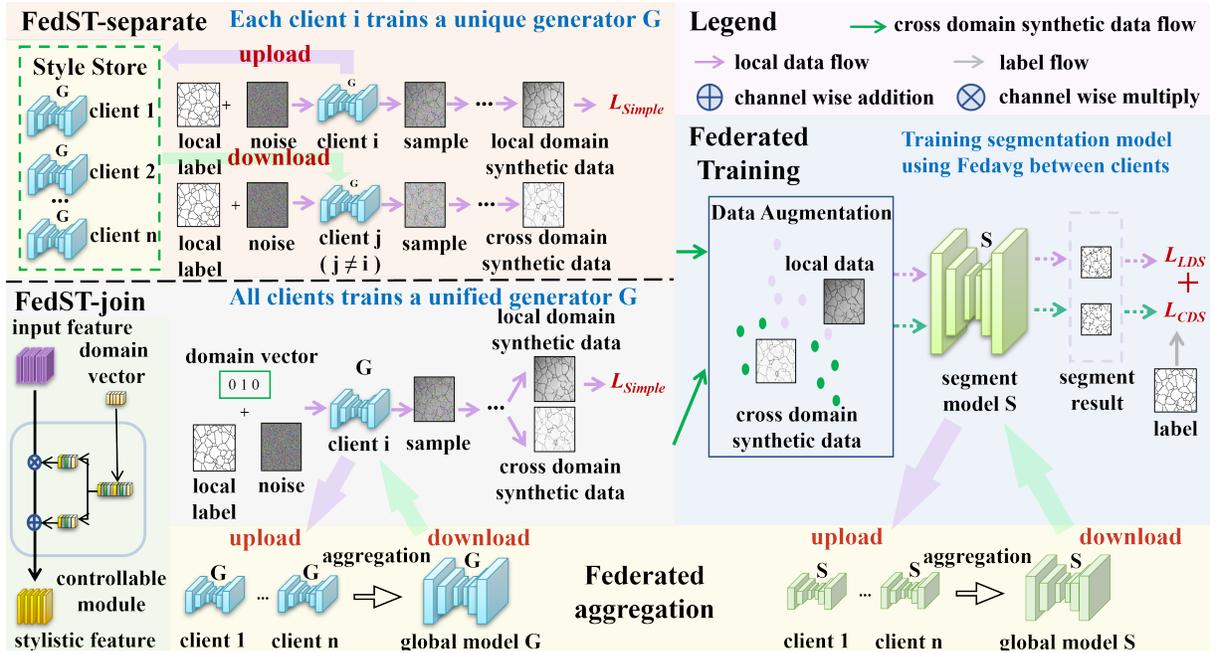


Figure 3: Overview of the proposed federated style transfer. The FedST-separate and FedST-join are two variants. The former lets each client train a unique style transfer generator and constructs a unified style store to save them. And it exchanges generators to let each client generate cross-domain data using their own local label. While the latter is equipped with a global controllable module to train a unified style transfer generator around all clients using the Fedavg method. And each client can modify the domain vector to generate cross-domain data. Finally, both of them use Fedavg to train the target image segmentation model using local and synthetic data.

mentation to decrease the Non-IID influence for the segmentation model.

Local Objective The loss function of the proposed method consists of two parts: the cross-domain style information modeling loss L_{Simple} and the segmentation loss L_{Seg} , as shown in Equation 1.

$$L = L_{Simple} + L_{Seg} \quad (1)$$

The first part is used to train the style transfer model, which is defined as:

$$\mathbb{E}_{(x,y)} \mathbb{E}_{\epsilon \sim N(0,I)} \mathbb{E}_{\gamma} \|f_{\theta}(x, \sqrt{\gamma}y + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon\|^2 \quad (2)$$

The above diffusion model loss L_{Simple} is proposed in (Saharia et al. 2022a). y is the given training image. x is the input condition. γ is the noise level indicator and f_{θ} is the neural network. The second part is used to train the segmentation model, it consists of two components: the local domain segmentation loss and the cross-domain segmentation loss, defined as:

$$L_{Seg} = L_{LDS} + \lambda L_{CDS} \quad (3)$$

where the local domain segmentation loss L_{LDS} is the combination of focal loss and dice loss between the segmentation result of real local training data and the corresponding ground truth segmentation label. λ is a super parameter to control the weight of loss items. The cross-domain segmentation loss L_{CDS} is the combination of focal loss and dice

loss between the segmentation result of the cross-domain synthetic image and the corresponding ground truth segmentation label which let the local segmentation model to learn the cross-domain data distribution.

Communication Cost Analysis

The communication cost of FedST-join will be lower than that of FedST-separate when the following condition is met:

$$m < \frac{(c-1)}{2} \quad (4)$$

Where m refers to the number of federated training communication rounds, and c refers to the number of clients. Thus, the communication cost of FedST-join is lower than that of FedST-separate when the number of communication rounds is smaller than half of the number of clients participating in federated learning.

Experiments

Experimental Setup

To verify the performance of the proposed method, we compare FedST with SOTA approaches dedicated to solving the Non-IID problem in FL, such as FedAvg (McMahan et al. 2017), FedProx (Li et al. 2020), FedDyn (Acar et al. 2021), and FedDc (Gao et al. 2022). We conduct extensive experiments on four cross-domain Non-IID tasks to evaluate the generality of our method, including material microscopic

| Segmentation Datasets Client Numbers Non-IID setting across clients Metrics | Material microscopic image 2 clients | | | Medical image 2 clients | | Face image 3 clients | | Face image 79 clients | |
|--|---|---------------|---------------|----------------------------|---------------|-------------------------|---------------|--------------------------|---------------|
| | Austenite / pure iron image | | | CT / MRI | | Different races | | Different ages | |
| | VI↓ | Dice ↑ | MPA ↑ | Dice ↑ | MPA ↑ | Dice ↑ | MPA ↑ | Dice ↑ | MPA ↑ |
| FedAvg (PMLR 2017) | 0.9785 | 0.6754 | 0.9122 | 0.3622 | 0.8662 | 0.5225 | 0.1145 | 0.7832 | 0.8102 |
| FedProx (MLSys 2020) | 0.7992 | 0.7004 | 0.9205 | 0.4344 | 0.8864 | 0.5363 | 0.1194 | 0.7859 | 0.8130 |
| FedDyn (ICLR 2021) | 0.4915 | 0.7278 | 0.9315 | 0.5230 | 0.9079 | 0.5844 | 0.1242 | 0.8041 | 0.8370 |
| FedDc (CVPR 2022) | 0.4022 | 0.7400 | 0.9367 | 0.5758 | 0.9117 | 0.5936 | 0.1392 | 0.7979 | 0.8362 |
| FedST-separate | 0.2978 | <u>0.7982</u> | 0.9246 | 0.9233 | <u>0.9693</u> | 0.9046 | <u>0.5962</u> | 0.8560 | 0.8763 |
| FedST-join | <u>0.2984</u> | 0.8145 | <u>0.9354</u> | <u>0.9221</u> | 0.9700 | <u>0.8998</u> | 0.6035 | <u>0.8470</u> | <u>0.8685</u> |

Table 1: The objective evaluation of FedST and other SOTA methods on the test datasets of four image segmentation tasks. The bold value and underline value denote the best and the second-best performance in each metric, respectively.

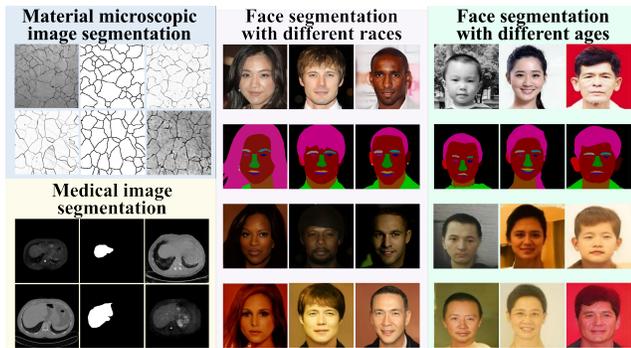


Figure 4: Example of cross-domain style transformation of each image segmentation task.

image segmentation, medical image segmentation, face segmentation with different races, and face segmentation with different ages.

As shown in Figure 4, the left part shows the example of medical image data set and material microscopic data set respectively. And it is shown the origin image and its corresponding segmentation label as well as cross-domain synthetic image from left to right. The middle and right part show the example of the face segmentation data with different races or ages. They are shown the original image, its corresponding segmentation, and cross-domain synthetic images from top to down.

As for the material microscopic image segmentation task, the dataset consists of two image domains: austenite microscopic image data and pure iron microscopic image data, which are independently stored in two clients to simulate the Non-IID phenomenon, respectively. The difference in style features between microscopic images of two metals is caused by different element compositions. The austenite dataset contains a total of 223 optical images of austenite with a resolution of 1024×1024 pixels. The pure iron dataset contains a total of 136 serial section optical images of pure iron with a resolution of 2800×1600 pixels. We pre-processed the original images into amounts of small patches with a size of 384×384 pixels and selected 892 patches from each domain for the experiment. Both two images have 2 semantic classes (grain and grain boundary).

As for the medical image segmentation, the data were sampled from CHAOS challenge (Kavur et al. 2019). We chose two cross-modalities images, CT and MRI (T2-SPIR), independently stored in two clients, to perform the Non-IID liver segmentation task. For each modality, 347 slices were chosen and resized to 384×384 pixels for the experiment.

As for the face segmentation task with different races, the data were sampled from CelebAMask-HQ (Lee et al. 2020). We divided the data into three different clients according to the ethnic tag: yellow race from Asian, white race from European and black race from African. For each client, 750 slices were chosen and resized to 384×384 pixels for the experiment.

As for the face segmentation task with different ages, the data were sampled from All-Age-Faces (AAF) Dataset (Cheng et al. 2019), which consists of face images of different ages from 2 to 80. In the experiment, 20 images were randomly selected from each age as private data owned by each client. These images were scaled to the size of 384×384 pixels and labeled by manual annotation. In other words, 79 participants are constructed for Non-IID federated learning.

We use PyTorch (Paszke et al. 2019) to implement federate style transfer and the other baselines. The federated optimizer of the proposed FedST is Adam. The batch size is set to 6 in material microscopic image segmentation task and medical image segmentation task while 3 in face segmentation task. The number of local epochs is set to 1 for all approaches. The number of communication rounds is set to 50 for all four tasks where all federated learning approaches have little or no accuracy gain with more communications.

The VI (Meilă 2007), Dice Coefficient (Sorensen 1948), and MPA (Long, Shelhamer, and Darrell 2015) are used to evaluate the objective performance of different algorithms. For the VI metric, a lower value indicates a better performance. While for the other metrics, a larger value indicates a better performance.

Result Comparison

We first compare the objective segmentation performance of our method and SOTA approaches in Table 1 on four tasks respectively.

For the material microscopic image segmentation task, VI is an important metric to evaluate the over-segmentation and under-segmentation of each result, which is also widely used

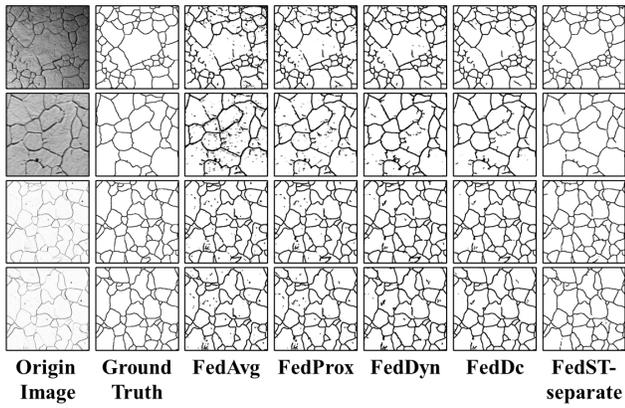


Figure 5: Qualitative comparison on the segmentation results of different methods on material microscopic image segmentation task. The first two row shows the result of austinate data while the last two rows show the result of pure iron data.

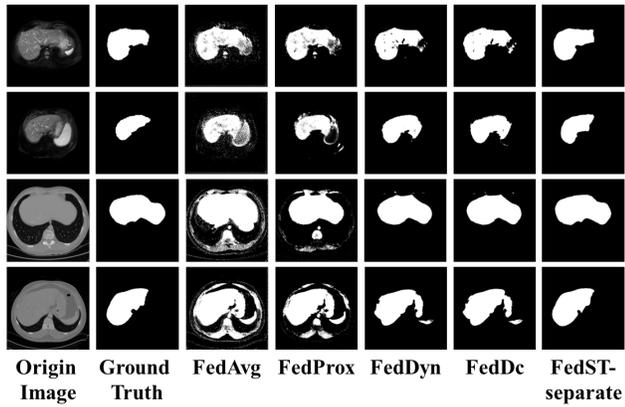


Figure 6: Qualitative comparison on the segmentation results of different methods on medical image segmentation task. The first two rows show the result of MRI data while the last two rows show the result of CT data.



Figure 7: Qualitative comparison on the segmentation results of different methods on face segmentation with different races. The first row shows the result of black race from Africa, the second row shows the white race from Europe, the last row shows the yellow race from Asia.



Figure 8: Qualitative comparison on the segmentation results of different methods on face segmentation with different ages. We select 3 ages to evaluate different methods. The first row shows the result of the children, the second row shows the youth, the last row shows the elderly.

in the field of neural cell image segmentation (Funke et al. 2018). Compared with other different methods, we find that FedST-separate achieves superior performance on VI metric and competitive performance on other metrics. FedST-separate and FedST-join achieve comparable results. Figure 4 shows the cross-domain synthetic image. It can be found that the style transfer generator learns the ability to translate the image segmentation label to different material microscopic data. Figure 5 shows the segmentation result of different methods on microscopic image segmentation task. We find that FedST achieves less over-segmentation and under-segmentation of grain boundary in the segmentation result compared to other methods.

For the medical image segmentation task, compared with different federated learning approaches, we observe that FedST shows a promising performance under non-IID settings. Figure 4 shows the cross-domain synthetic image. It can be found that the style transfer generator roughly learns the ability to translate the image segmentation label to different medical modalities. Figure 6 shows the segmentation result of different methods on the medical image segmentation task. We find that FedAvg and FedProx failed to learn the segmentation ability due to the Non-IID data.

For the face segmentation task with different races, we observe that FedST shows superior performance compared with the other method. It can be found that the style transfer generator roughly learns the ability to translate the image segmentation label to different racial face data. Figure 7 shows the segmentation result of different methods. We can find that FedST achieves promising segmentation results on small organs, such as eyes, and eyebrows.

For the face segmentation task with different ages, in order to accelerate the training process, we pre-train the image segmentation module and style transformation module on CelebAMask-HQ for all methods and only randomly select the other 3 different style transformation modules for each client in the process of exchanging style features of FedST. As shown in Table 1 and Figure 8, The FedST also outperforms other methods in objective and subjective assessment.

In addition, we conduct an experiment to evaluate the

training procedure of different methods. Figure 9(a-b) show the performance curve of different methods during training. It can be found that FedST-separate will achieve better performance compared to other methods on the same communication round.

To sum up, FedST achieves promising performance compared to other methods among four different tasks. FedST-separate and FedST-joint achieve comparable results.

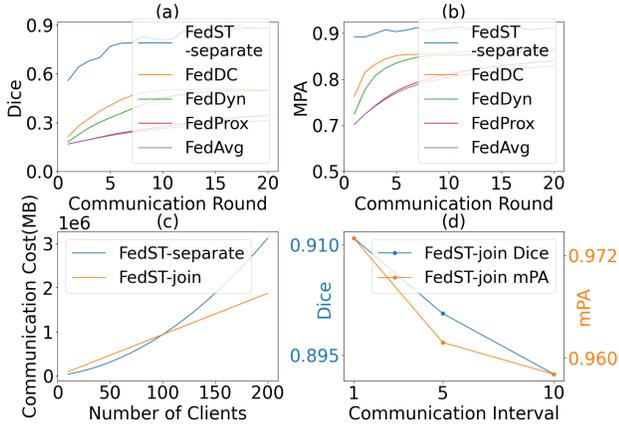


Figure 9: Effectiveness of cross-domain style transformation. (a-b) The performance curve of different methods during training. (c) Comparison of communication costs of FedST-separate and FedST-joint with different numbers of clients. The communication rounds are set to 50. (d) The ablation experiment of FedST-separate and FedST-joint on medical image segmentation task regarding the communication interval.

Ablation Studies of FedST

The proposed FedST consists of FedST-separate and FedST-joint, depending on how to train a style transformation network. Both of them achieve an obvious improvement over the Fedavg baseline by sharing style information. As shown in Eq 4, the communication cost of FedST-joint is lower than that of FedST-separate when the number of communication rounds (50 in this work) is smaller than half of the number of clients. Thus, as shown in Figure 9(c), the FedST-joint has a lower communication cost when the number of clients exceeds 100.

The communication frequency of the style transfer model during the FedST training is an important parameter to determine generate quality. We conduct an experiment of FedST-separate and FedST-joint regarding the communication interval, as shown in Figure 9(d). It is shown that when the total number of epochs for each client is determined, the performance will degrade when the communication interval increases.

Analysis of Information Sharing

FedST can extract and exchange style information among clients, thus effectively alleviating the influence of the Non-IID phenomenon. We further provide the data distribution

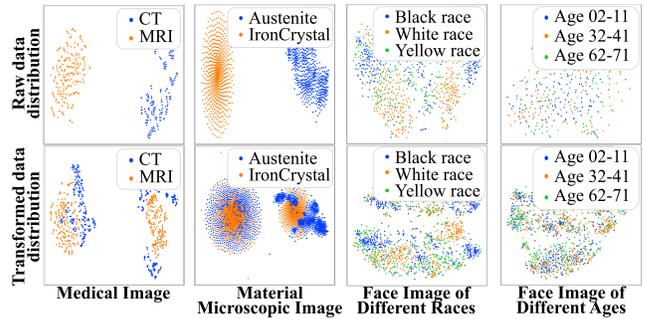


Figure 10: T-SNE embedding visualization of data distribution across clients for raw data and transformed data. Each client has a different domain of raw data to simulate the non-iid setting. In order to analyze AAF data more effectively, we have aggregated it into three age groups for visualization.

of raw data and transformed data after FedST in Figure 10, which indicates that the data distribution of original clients is non-iid, while the data generated in our method can alleviate this phenomenon.

FedST exchange style information among clients. This might introduce the risk of information leakage. However, in the image segmentation task, the structural information is the crucial information, which is generally required by the downstream application. Even though FedST shares the style information among clients, it keeps structural information private for each client. Thus it has little impact on users. In addition, the style information is maintained in the network parameters, which is inexplicable to users. And even if being attacked, the attacker can't restore the structural information of other clients.

Conclusion

We have proposed a novel FL model to solve the Non-IID problem in the image segmentation tasks. Our method realizes information sharing based on transfer learning theory. It can be considered to exchange style information across clients while keeping the structure information private. Extensive experiments among four different domain tasks show that our method achieves promising performance compared to SOTA methods in objective and subjective assessment. We will further improve the quality of synthetic images and alleviate the influence of the Non-IID in future work.

Acknowledgements

This work was supported by the National Key R & D Program of China under Grant 2022ZD0118001, the National Natural Science Foundation of China under Grant U22A2022 and 62106019, Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB under Grant BK22BF010, and Fundamental Research Funds for the Central Universities of China under Grant 00007467, and China Postdoctoral Science Foundation under Grant 2021M700383. The computing work is supported by USTB MatCom of Beijing Advanced Innovation Center for Materials Genome Engineering.

References

- Acar, D. A. E.; Zhao, Y.; Matas, R.; Mattina, M.; Whatmough, P.; and Saligrama, V. 2021. Federated Learning Based on Dynamic Regularization. In *International Conference on Learning Representations*.
- Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and van den Berg, R. 2021. Structured Denoising Diffusion Models in Discrete State-Spaces. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 17981–17993. Curran Associates, Inc.
- Brisimi, T. S.; Chen, R.; Mela, T.; Olshevsky, A.; Paschalidis, I. C.; and Shi, W. 2018. Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112: 59–67.
- Cheng, J.; Li, Y.; Wang, J.; Yu, L.; and Wang, S. 2019. Exploiting effective facial patches for robust gender recognition. *Tsinghua Science and Technology*, 24(3): 333–345.
- Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34: 8780–8794.
- Dong, N.; and Voiculescu, I. 2021. Federated contrastive learning for decentralized unlabeled medical images. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, 378–387. Springer.
- Funke, J.; Tschoop, F.; Grisaitis, W.; Sheridan, A.; Singh, C.; Saalfeld, S.; and Turaga, S. C. 2018. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE transactions on pattern analysis and machine intelligence*, 41(7): 1669–1680.
- Gao, L.; Fu, H.; Li, L.; Chen, Y.; Xu, M.; and Xu, C.-Z. 2022. Feddc: Federated learning with non-iid data via local drift decoupling and correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10112–10121.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33: 6840–6851.
- Hsu, T.-M. H.; Qi, H.; and Brown, M. 2019. Measuring the effects of non-identical data distribution for federated visual classification. arXiv:1909.06335.
- Hu, J.; Shi, Y. N.; Sauvage, X.; Sha, G.; and Lu, K. 2017. Grain boundary stability governs hardening and softening in extremely fine nanograined metals. *Science*, 355(6331): 1292–1296.
- Huang, Y.; Chu, L.; Zhou, Z.; Wang, L.; Liu, J.; Pei, J.; and Zhang, Y. 2021. Personalized Cross-Silo Federated Learning on Non-IID Data. In *AAAI*, 7865–7873.
- Huang, Y.; Song, Z.; Li, K.; and Arora, S. 2020. Instahide: Instance-hiding schemes for private distributed learning. In *International conference on machine learning*, 4507–4518. PMLR.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.
- Jiang, J. C.; Kantarci, B.; Oktug, S.; and Soyata, T. 2020. Federated learning in smart city sensing: Challenges and opportunities. *Sensors*, 20(21): 6230.
- Kadkhodaie, Z.; and Simoncelli, E. P. 2020. Solving linear inverse problems using the prior implicit in a denoiser. arXiv:2007.13640.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2): 1–210.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S. J.; Stich, S. U.; and Suresh, A. T. 2019. SCAFFOLD: Stochastic Controlled Averaging for On-Device Federated Learning. arXiv:1910.06378.
- Kavur, A. E.; Selver, M. A.; Dicle, O.; Barış, M.; and Gezer, N. S. 2019. CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data.
- Konečný, J.; McMahan, H. B.; Yu, F. X.; Richtárik, P.; Suresh, A. T.; and Bacon, D. 2016. Federated learning: Strategies for improving communication efficiency. arXiv:1610.05492.
- Kopparapu, K.; and Lin, E. 2020. Fedfmc: Sequential efficient federated learning on non-iid data. arXiv:2006.10937.
- Kopparapu, K.; Lin, E.; and Zhao, J. 2020. Fedcd: Improving performance in non-iid federated learning. arXiv:2006.09637.
- Lee, C.-H.; Liu, Z.; Wu, L.; and Luo, P. 2020. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2: 429–450.
- Li, Y.; Yang, Y.; Zhou, W.; and Hospedales, T. 2019. Feature-critic networks for heterogeneous domain generalization. In *International Conference on Machine Learning*, 3915–3924. PMLR.
- Lin, T.; Kong, L.; Stich, S. U.; and Jaggi, M. 2020. Ensemble distillation for robust model fusion in federated learning. *Advances in Neural Information Processing Systems*, 33: 2351–2363.
- Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021. 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*, 1013–1023.
- Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of

- deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Meilă, M. 2007. Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 98(5): 873–895.
- Mirza, M.; and Osindero, S. 2014. Conditional generative adversarial nets. arXiv:1411.1784.
- Niknam, S.; Dhillon, H. S.; and Reed, J. H. 2020. Federated learning for wireless communications: Motivation, opportunities, and challenges. *IEEE Communications Magazine*, 58(6): 46–51.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Qu, L.; Zhou, Y.; Liang, P. P.; Xia, Y.; Wang, F.; Adeli, E.; Fei-Fei, L.; and Rubin, D. 2022. Rethinking architecture design for tackling data heterogeneity in federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10061–10071.
- Rubner, Y.; Tomasi, C.; and Guibas, L. J. 2000. The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40(2): 99–121.
- Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; and Norouzi, M. 2022a. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, 1–10.
- Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D. J.; and Norouzi, M. 2022b. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Sasaki, H.; Willcocks, C. G.; and Breckon, T. P. 2021. Unitddpm: Unpaired image translation with denoising diffusion probabilistic models. arXiv:2104.05358.
- Sorensen, T. A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.*, 5: 1–34.
- Voigt, P.; and Von dem Bussche, A. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed.*, Cham: Springer International Publishing, 10(3152676): 10–5555.
- Wu, Y.; Wang, Z.; Zeng, D.; Li, M.; Shi, Y.; and Hu, J. 2022. Decentralized unsupervised learning of visual representations. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, 2326–2333.
- Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2): 1–19.
- Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; and Chandra, V. 2018. Federated learning with non-iid data. arXiv:1806.00582.
- Zhu, Z.; Hong, J.; and Zhou, J. 2021. Data-free knowledge distillation for heterogeneous federated learning. In *International Conference on Machine Learning*, 12878–12889. PMLR.