26. Rivera, M.P., Mehta, A.C., Wahidi, M.M.: Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. Chest **143**(5 Suppl), e142S-e165S (2013)
27. Schütze, H., Manning, C.D., Raghavan, P.: Introduction to information retrieval. vol. 39. Cambridge University Press, Cambridge (2008)
28. Team, N.L.S.T.R.: Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. **365**(5), 395–409 (2011)
29. Tipirneni, S., Reddy, C.K.: Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. ACM Trans. Knowl. Discov. Data (TKDD) **16**(6), 1–17 (2022)
30. Vanguri, R.S., et al.: Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer. Nat. Cancer **3**(10), 1151–1164 (2022)
31. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
32. Wu, C., Wu, F., Huang, Y.: Da-transformer: Distance-aware transformer. arXiv preprint arXiv:2010.06925 (2020)
33. Zhang, X., Zeman, M., Tsiligkaridis, T., Zitnik, M.: Graph-guided network for irregularly sampled multivariate time series. arXiv preprint arXiv:2110.05357 (2021)

# FedContrast-GPA: Heterogeneous Federated Optimization via Local Contrastive Learning and Global Process-Aware Aggregation

Qin Zhou and Guoyan Zheng(✉)

Institute of Medical Robotics, School of Biomedical Engineering, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai 200240, China
guoyan.zheng@sjtu.edu.cn

**Abstract.** Federated learning is a promising strategy for performing privacy-preserving, distributed learning for medical image segmentation. However, the data-level heterogeneity as well as system-level heterogeneity makes it challenging to optimize. In this paper, we propose to improve **Fed**erated optimization via local **Contrast**ive learning and **G**lobal **P**rocess-aware **A**ggregation (referred as FedContrast-GPA), aiming to jointly address both data-level and system-level heterogeneity issues. In specific, To address data-level heterogeneity, we propose to learn a unified latent feature space via an intra-client and inter-client local prototype based contrastive learning scheme. Among which, intra-client contrastive learning is adopted to improve the discriminative ability of learned feature embedding at each client, while inter-client contrastive learning is introduced to achieve cross-client distribution perception and alignment in a privacy preserving manner. To address system-level heterogeneity, we further propose a simple yet effective process-aware aggregation scheme to achieve effective straggler mitigation. Experimental results on six prostate segmentation datasets demonstrate large performance boost over existing state-of-the-art methods.

**Keywords:** Heterogeneous federated learning · Process-aware aggregation · Local prototype learning · Contrastive learning

## 1 Introduction

Recently, federated learning has emerged as a promising strategy for performing privacy-preserving, distributed learning for medical image segmentation. Among various methods, FedAvg [1] has been the de facto approach for federated learning, where the server maintains a global model which is dispatched to each client for updating locally on their own private data. After that, the updated local models

This study was partially supported by the Natural Science Foundation of China via project U20A20199 and 62201341.


H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14221, pp. 660–670, 2023.
https://doi.org/10.1007/978-3-031-43895-0_62

are collected and averaged to produce a global model for the training of the next round. For FedAvg and its variants, a well-known issue is “client drift” caused by non-IID data distribution across different clients (i.e., data-level heterogeneity). To address this issue, a group of methods resort to designing proximal terms or re-parametrization strategies [2,3] to restrain the client drift from the global model. However, these regularization terms inherently limit the local convergence potential, Other methods try to improve the local models’ generalization ability without strict proximal restrictions on model parameters [4,5]. In [27], the authors proposed to learn compact local representations on each device and a global model across all devices, reducing both the intra-client and inter-client data variance. However, these methods perform local updates blindly, totally ignoring the feature distributions of other clients. In medical image segmentation, FedDG [6] was proposed to improve the local models’ generalization ability via exchanging amplitude spectrum to transmit the distribution information across clients. However, the distribution perception step was processed offline, which was fixed upon finished, limiting its potential adaptability to various subsequent tasks.

Different from the above-mentioned methods, in this paper, we aim to tackle the “client drift” problem by exploring a unified latent feature space for different clients in a privacy-preserving manner and by enhancing the feature discriminability of each client. Concretely, we propose to extract local prototypes to represent the feature distribution at each client. Since local prototypes are statistical characteristics, we can share them among different clients without the concern of privacy issues. Then performing cross-client pixel to local prototype matching can help not only to perceive the global feature distribution but also to explicitly align the cross-client features, leading to a more unified latent feature space. Besides, by performing pixel to local prototype matching at each client, we can directly shape and enhance the discriminability of the learned feature space at each client.

Another well-acknowledged concern of federated networks is the “straggler” problem caused by system-level heterogeneity. FedAvg and some of its variants [2,3] directly average the local models weighted by their data amount ratio, which may lead to unexpected deterioration due to asynchronous learning process of local models. Based on the intuition that well-trained local models should contribute more to the global model, in this paper, we propose a simple yet effective process-aware model aggregation scheme, which is demonstrated to effectively suppress the influence of “stragglers”.

The contributions of our method can be summarized as follows:

- We propose a novel FedContrast-GPA framework to simultaneously alleviate both data-level and system-level heterogeneity issues in federated optimization.
- We propose an intra-client and inter-client local prototype based contrastive learning scheme, which not only enhances the feature discriminability of each client, but also explicitly performs cross-client feature distribution perception and alignment in a privacy-preserving manner.
- We introduce a simple yet effective process-aware weighting scheme to suppress the influence of “stragglers” in global model aggregation.

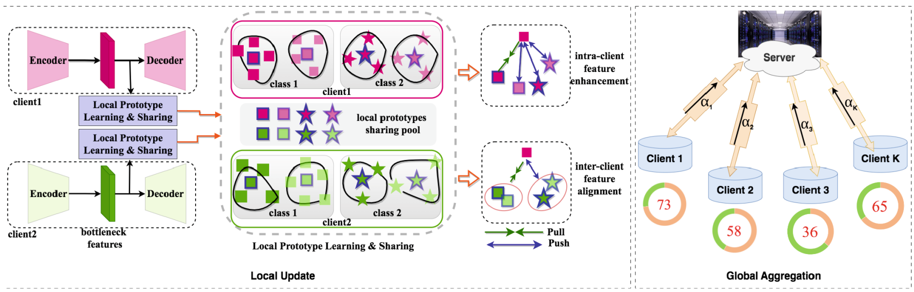


**Fig. 1.** The workflow of the proposed FedContrast-GPA framework. Please note in local update, same shapes mean the features belong to the same semantic class while same shapes of the same color figure online form a sub-cluster of the related semantic class. The local prototypes are marked with border lines. In global aggregation, the numbers indicate the training process. (Color figure online)

## 2 Method

A typical federated learning process consists of two stages: local update at each client and global aggregation at the server side. In this paper, we propose the FedContrast-GPA framework (as shown in Fig. 1), which consists of the intra- and inter-client local prototype based contrastive learning scheme (during local update) and the process-aware aggregation scheme (during global aggregation).

Assuming there exist $K$ clients in the federated network, we denote client $k$ as $S_k$. Then private data set on the $k$-th client can be denoted as $\{\boldsymbol{I}_k^n, \mathbf{Y}_k^n\}, n \in \{1, \cdots, N_k\}$, where $\boldsymbol{I}_k^n, \boldsymbol{Y}_k^n$ are the image and the corresponding segmentation map for the $n$-th instance in $S_k$, and $N_k$ is the number of instances in $S_k$. In this paper, we adopt U-Net as the backbone architecture for segmentation. Denote $\varphi_k = f_k^e \circ f_k^d$ as the mapping function for $S_k$, where $f_k^e$ and $f_k^d$ are the encoder and decoder, and $\circ$ means sequentially executing $f_k^e$ and $f_k^d$. Denote $\boldsymbol{w}_k$ as the parameters of local model on $S_k$, then the goal of federated learning is to find the optimal global model $\boldsymbol{w} = GA(\{\boldsymbol{w}_k\}, k \in \{1, \cdots, K\})$ that generalizes well across different clients, where $GA(\cdot)$ refers to a certain strategy for model aggregation.

### 2.1 Intra- and Inter-client Local-Prototype based Contrastive Learning

**Local Prototype Learning.** Denote the bottleneck features of class $c$ on $S_k$ as $\boldsymbol{F}_k^c = \{\boldsymbol{f}_k^{c,i}, i \in \{1, \cdots, N_k^c\}\}$, where $N_k^c$ represents the number of pixels belonging to class $c$ in the intermediate feature maps. In order to model the feature distribution of $S_k$ from a statistical view, we propose to generate the class-specific local prototypes to capture semantic-aware feature distribution. Considering that the spatial coverage and visual changes may vary dramatically across different classes, we extend the method introduced in [8] to allow learning different number of sub-clusters for different semantic classes. For detailed derivation of local prototype learning, please refer to [8]. Denote the learned local prototypes for class $c$ as $\boldsymbol{P}_k^c = \{\boldsymbol{p}_k^{c,t}, t \in \{1, \cdots, T_c\}\}$ (where $T_c$ refers to the number of sub-clusters for class $c$), and the pixel-to-local-prototype mapping as $\boldsymbol{M_k^c} = \{m_k^{c,i}, i = \{1, \cdots, N_k^c\}\}$, where $m_k^{c,i} \in \{1, \cdots, T_c\}$ represents the assigned sub-cluster index of pixel $i$ in class $c$. Then the learned local prototypes are utilized to perform intra-client feature enhancement and inter-client feature alignment. Please note, $T_c$ may vary for different semantic class to flexibly adapt to its visual characteristics.

**Intra-client Local-Prototype Based Contrastive Learning (Intra-LPCL) for Feature Enhancement.** The motivation of Intra-LPCL is to enhance the discriminability of local features. Specifically, given the bottleneck features $\boldsymbol{F}_k^c$, and the learned local prototypes $\{\boldsymbol{P}_k^c = \{\boldsymbol{p}_k^{c,t}, t \in \{1, \cdots, T_c\}\}, c \in \{1, \cdots, C\}\}$ from all the semantic classes, where $C$ is the total class number. Then contrastive learning is introduced to enforce compactness within a sub-cluster and separation among different sub-clusters. Specifically, the intra-client pixel-to-local-prototype contrastive loss is calculated as,

$$L_k^c = \frac{1}{Z} \sum_{c=1}^{C} \sum_{i=1}^{N_k^c} -log \frac{e^{s_k^{c,m_k^{c,i}}}}{\sum_{c'=1}^{C} \sum_{t=1}^{T_{c'}} e^{s_k^{c',t}}}, \tag{1}$$

where $s_k^{c,m_k^{c,i}}$ denotes the similarity between the $i$-th local feature of class $c$ (i.e., $\boldsymbol{f}_k^{c,i}$) and the local prototype from the sub-cluster that it belongs to (i.e., $\boldsymbol{p}_k^{c,m_k^{c,i}}$), and $s_k^{c',t}$ represents the similarity score between $\boldsymbol{f}_k^{c,i}$ and the local prototype from the $t$-th sub-cluster of class $c'$, where $T_{c'}$ denotes the number of sub-clusters for class $c'$, and $Z$ is the normalization factor to average over all the pixels within a mini-batch. In our method, we adopt cosine similarity to get the similarity score,

$$s_k^{c',t} =< \boldsymbol{f}_k^{c,i}, \boldsymbol{p}_k^{c',t} >, \tag{2}$$

where $<,>$ denotes the cosine similarity function. Please note, visual compactness is only imposed at the sub-cluster granularity, which means the local features should distribute faraway from not only sub-clusters of the other semantic classes, but also other sub-clusters of the same semantic class. Apart from the contrastive loss term, in Intra-LPCL, we also explicitly maximize the feature similarities between local features and their assigned local prototypes as,

$$L_k^d = \frac{1}{Z} \sum_{c=1}^{C} \sum_{i=1}^{N_k^c} 1.0 - < \boldsymbol{f}_k^{c,i}, \boldsymbol{p}_k^{c,m_k^{c,i}} >, \tag{3}$$

Then the final Intra-LPCL loss is calculated as,

$$L_k^{intra} = L_k^c + L_k^d, \tag{4}$$

**Inter-client Local-Prototype Based Contrastive Learning (Inter-LPCL) for Feature Alignment.** The aim of Inter-LPCL is to perform distributed feature alignment across different clients in a privacy-preserving manner, such that the aggregated global model can generalize well across clients. Given the $i$-th local feature of class $c$ from $S_k$ (i.e., $\boldsymbol{f}_k^{c,i}$), and the prototypes pool $\{\boldsymbol{P}_{k'}^{c'} = \{\boldsymbol{p}_{k'}^{c',t}, t \in \{1, \cdots, T_{c'}\}\}, c' \in \{1, \cdots, C\}\}$ from $S_{k'}$ ($k' \neq k$), we don't know the cross-client pixel-to-prototype assignments. Thus, instead of imposing strict restrictions on sub-cluster compactness as done in Intra-LPCL, we loosen the alignment restrictions to category level. Specifically, the local features from $S_k$ are supposed to distribute closer to one of the sub-clusters belonging to the same class in $S_{k'}$, and faraway from sub-clusters of the other semantic classes. Mathematically, the inter-LPCL loss is calcualted as,

$$L_{k,k'}^{inter} = \frac{1}{Z} \sum_{c=1}^{C} \sum_{i=1}^{N_k^c} -log \frac{e^{max(\{s_{k,k'}^{c,t},\, t \in \{1, \cdots, T_c\}\})}}{\sum_{c'=1}^{C} e^{max(\{s_{k,k'}^{c',t}, t \in \{1, \cdots, T_{c'}\}\})}}, \tag{5}$$

where $max(\cdot)$ returns the maximum value in the set, $\{s_{k,k'}^{c',t}, t \in \{1, \cdots, T_{c'}\}\}$ denotes the similarity set calculated between $\boldsymbol{f}_k^{c,i}$ and all the local prototypes from class $c'$ in $S_{k'}$, which is formulated as,

$$s_{k,k'}^{c',t} =< \boldsymbol{f}_k^{c,i}, \boldsymbol{p}_{k'}^{c',t} >, \tag{6}$$

The final Inter-LPCL loss of $S_k$ is then calculated by averaging over $k'$ in $L_{k,k'}^{inter}$, which is formally defined as,

$$L_k^{inter} = \frac{1}{K-1} \sum_{k' \neq k} L_{k,k'}^{inter}, \tag{7}$$

**Overall Objective for Local Update.** The overall loss function for updating local model from $S_k$ is formulated as,

$$L_k = L_k^{seg} + \lambda_1 L_k^{intra} + \lambda_2 L_k^{inter}, \tag{8}$$

where $\lambda_1, \lambda_2$ are the hyper-parameters, $L_k^{seg}$ is the segmentation loss,

$$L_k^{seg} = \frac{1}{Z} \sum_n CE(\varphi_k(\boldsymbol{I}_k^n), \boldsymbol{Y}_k^n), \tag{9}$$

where $CE(\cdot)$ denotes the cross entropy loss.

### 2.2 Process-Aware Global Model Aggregation

During each federated communication, FedAvg updates the global model as weighted average over local models,

$$\boldsymbol{w} = \alpha_k \boldsymbol{w_k}, k \in \{1, \cdots, K\}, \tag{10}$$

**Table 1.** Dice similarity coefficients (DSC) of different settings to demonstrate effectiveness of each component in our method.

| Intra-LPCL | Inter-LPCL | $GA_p$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | Average |
|---|---|---|---|---|---|---|---|---|---|
| | | | 87.5 | 83.0 | 88.9 | 80.1 | 90.0 | 76.3 | 84.3 |
| ✓ | | | 86.3 | 85.5 | 87.1 | 81.3 | 90.8 | 82.5 | 85.6 |
| | ✓ | | 88.8 | 86.0 | 87.8 | 80.0 | 91.4 | 80.2 | 85.7 |
| | | ✓ | 88.2 | **86.3** | 86.9 | 82.8 | 91.1 | 80.5 | 86.0 |
| ✓ | ✓ | ✓ | **89.7** | 85.0 | **89.5** | **83.4** | **91.5** | **87.3** | **87.7** |

where $\alpha_k$ is the aggregation weight for $S_k$, which is commonly set as $\frac{N_k}{\sum_k N_k}$ ($N_k$ is the number of images in $S_k$). Instead of weighting the local models by its data amount ratio, in this paper, we argue that the aggregation weights should reflect the training process of each local model (i.e., well-trained model that generates good segmentation results should contribute more during aggregation). Specifically, denote the mean Dice Similarity Coefficient obtained on the training and validation data of $S_k$ as $DSC_k$, then the normalized weights in our method are calculated as,

$$\alpha_k = \frac{DSC_k}{\sum_k DSC_k}. \tag{11}$$

By introducing the process-aware aggregation scheme, we can effectively detect the straggler, improving the robustness of aggregated global model.

## 3 Experiments and Results

**Datasets and Implementation Details.** We validate our method on the challenging task of prostate segmentation from 3D MR images. T2-weighted MRI images used in our study are collected from 6 different data sources [24–26], where each source is treated as a client in our study. We follow [28] to preprocess the data. For data augmentation, both geometric transformations (including elastic deformation, translation, rotation and scaling) and intensity augmentations (including contrast and gaussian noise) are employed in our method. The local model is trained using Adam optimizer with a batch size of 64 and Adam momentum of 0.9 and 0.999. The learning rate is initialzed as 0.001 and multiplied by 0.9 after each round of federated communication. The local epoch in each federated round is empirically set as 1. The hyper-parameters $\lambda_1, \lambda_2$ are empirically set as 0.03 and 0.001 respectively. The number of local prototypes for prostate and background are chosen by grid search, and empirically set as 3 and 6, respectively.

**Ablation Study on the Effectiveness of Each Component:** We denote the process-aware global aggregation as $GA_p$, then the detailed analysis on component effectiveness is presented in Table 1. One can see from this table that

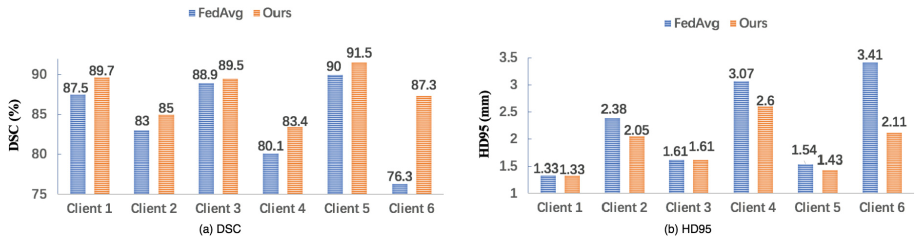


**Fig. 2.** Analysis on the straggler mitigation effect in terms of both DSC and 95% Hausdorff Distance (HD95) metrics.

incorporating the "Intra-LPCL" and "Inter-LPCL" terms can bring +1.3% and +1.4% overall DSC performance gains respectively, validating the effectiveness of intra-client discriminability enhancement and inter-client feature perception and alignment. Albeit its simplicity, our process-aware global aggregation scheme can also boost the DSC segmentation performance by a large margin (i.e., a 1.7% increase over the baseline). Combined together, the proposed FedContrast-GPA framework can witness a gain of 3.4% on the overall DSC performance.

To further demonstrate the advantage of our federated learning strategy, we conduct experiments to analyze the performance of centralized training and separate training, where centralized training is trained by updating the global model sequentially using private data from each client, while separate training is trained by updating each local client with only private data and no global communication. The average Dice performance for each client in centralized training is 73%, 77%, 84%, 72%, 86%, and 76%, respectively, while the average Dice performance for each client in separate training is 85%, 79%, 86%, 73%, 91%, and 27%, respectively. We can see that directly putting data together in centralized training does not bring performance gain due to data heterogeneity. Besides, in separate training, we can see a severe performance drop in some clients without enough learning data and knowledge from others.

**Analysis on the Straggler Mitigation Effect.** To demonstrate the effectiveness of our method in straggler mitigation, we compare the client-specific DSC and HD95 performance between the baseline (FedAvg) and Ours. For clarity, we first define the "stragglers" in a federated learning network as follows: the clients whose performance from the baseline (FedAvg) rank among the worst half of all the clients. Note that the "stragglers" are recognized according to the performance of FedAvg, since our method aims to address the "straggler" problem in FedAvg.

As shown in Fig. 2, for the prostate segmentation task, "Client 2", "Client 4" and "Client 6" are the "stragglers". Compared to the "non-stragglers", we can observe large performance gains on the "stragglers". In specific, the DSC performance gains on "Client 2", "Client 4" and "Client 6" are 2.0%, 3.3% and +11.0% (on average a 5.4% DSC gain), respectively. The HD95 performance

**Table 2.** Comparison with state-of-the-art methods in terms of DSC. Please note that larger DSC numbers indicate better performance. Best results are marked in bold.

| Methods | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | Average |
|---|---|---|---|---|---|---|---|
| FedAvg | 87.5 | 83.0 | 88.9 | 80.1 | 90.0 | 76.3 | 84.3 |
| FedAvg-LG | 83.8 | 82.4 | 85.3 | 75.9 | 89.9 | 56.3 | 78.9 |
| FedDG | 85.7 | 83.4 | 84.2 | 80.6 | 89.4 | 81.4 | 84.1 |
| FedProx | 88.1 | 84.3 | 86.2 | 83.1 | 90.6 | 83.1 | 85.9 |
| MOON | 87.0 | 84.9 | 87.7 | 82.7 | 90.3 | 83.0 | 85.9 |
| Ours | **89.7** | **85.0** | **89.5** | **83.4** | **91.5** | **87.3** | **87.7** |

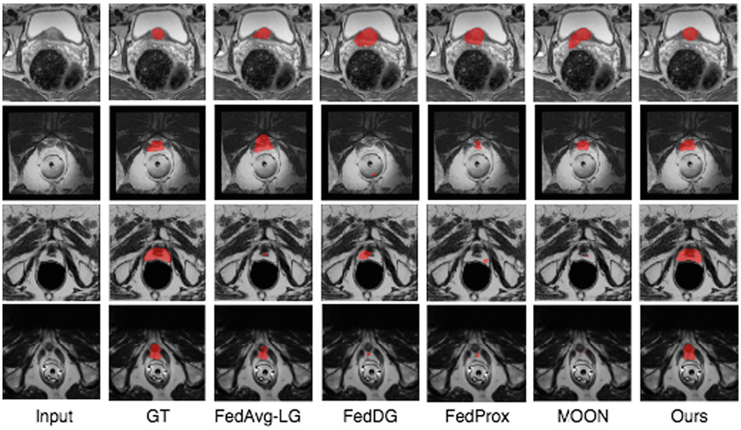


**Fig. 3.** Qualitative comparison results between our method and other state-of-the-arts on prostate segmentation.

gains are –0.33 mm, –0.47 mm, and –1.3 mm, respectively (on average –0.7 mm gain in terms of HD95). Meanwhile, for the "non-stragglers", the improvements are respectively 2.2%, 0.6% and 1.5% (on average 1.4%) in terms of DSC and –0.0 mm, –0.0 mm, and –0.11 mm (on average –0.04 mm) in terms of HD95. From above analysis, we can see that the proposed method can achieve effective straggler mitigation by bringing larger performance gains over stragglers, and slightly boost the performance over the 'non-stragglers".

**Comparison with State-of-the-Art Methods.** We compare the performance of our method with five state-of-the-art (SOTA) methods, including FedAvg [1], FedAvg-LG [27], FedDG [6], FedProx [2] and MOON [4]. For fair comparison, all the SOTA methods were trained/tested on our own dataset splits. The base parameter settings are kept the same as ours, other hyper-parameters are chosen by grid-search (In FedAvg-LG, the number of layers for global aggregation is set as 13, the hyper-parameters in FedDG are the same as the original paper, the weight for proxy term in FedProx is 2.5e-4, and the model contrastive coefficient in MOON is 0.01). In the following, we conduct both quantitative and qualitative comparisons with SOTA methods.

To analyse the performance of our proposed FedContrast-GPA framework, we report the DSCs for all the distributed clients (i.e., $S_1 - S_6$ in Table 2). For a straightforward comparison with the SOTA methods, we also record the average DSC across all the clients. Detailed comparison results are illustrated in Table 2. As shown, overall, our proposed FedContrast-GPA framework achieves superior performance than the listed SOTA methods. Specifically, FedContrast-GPA outperforms the second best method by 1.8% in terms of average DSC, and generates the best DSC performance at each client, demonstrating favorable generalization ability of our method. Figure 3 demonstrates some sampled visualization results from different clients. As shown, listed SOTA methods may fail to obtain good segmentation results on some samples from different clients, while our approach can consistently generate reasonably good results, demonstrating the robustness and generalizability of our method.

## 4 Conclusions

In this paper, we proposed a novel FedContrast-GPA framework to simultaneously address both the data-level heterogeneity and the system-level heterogeneity issues in federated networks. Extensive ablation studies and comparisons with the SOTA methods demonstrated the effectiveness of the proposed method.

## References


1. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282. PMLR (2017)
2. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proc. Mach. Learn. Syst. **2**, 429–450 (2020)
3. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: SCAFFOLD: stochastic controlled averaging for federated learning. In: International Conference on Machine Learning, pp. 5132–5143. PMLR (2020)
4. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10713–10722) (2021)
5. Mendieta, M., Yang, T., Wang, P., Lee, M., Ding, Z., Chen, C.: Local learning matters: rethinking data-level heterogeneity in federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8397–8406 (2022)
6. 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)

7. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with Superpixels: training few-shot medical image segmentation without annotation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12374, pp. 762–780. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58526-6_45
8. Zhou, T., Wang, W., Konukoglu, E., Van Gool, L.: Rethinking semantic segmentation: a prototype view. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2582–2593 (2022)
9. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
10. Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., Khazaeni, Y.: Federated learning with matched averaging. arXiv preprint arXiv:2002.06440 (2020)
11. Hsu, T.M.H., Qi, H., Brown, M.: Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335 (2019)
12. Wang, J., Liu, Q., Liang, H., Joshi, G., Poor, H.V.: Tackling the objective inconsistency problem in heterogeneous federated optimization. Adv. Neural Inf. Process. Syst. **33**, 7611–7623 (2020)
13. Wu, Y., Zeng, D., Wang, Z., Shi, Y., Hu, J.: Federated contrastive learning for volumetric medical image segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 367–377. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_35
14. Dong, N., Xing, E.P.: Few-shot semantic segmentation with prototype learning. In: British Machine Vision Conference (BMVC). vol. 3, no. 4 (2018)
15. Liu, J., Qin, Y.: Prototype refinement network for few-shot segmentation. arXiv preprint arXiv:2002.03579 (2020)
16. Liu, Y., Zhang, X., Zhang, S., He, X.: Part-aware prototype network for few-shot semantic segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 142–158. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_9
17. Yu, Q., Dang, K., Tajbakhsh, N., Terzopoulos, D., Ding, X.: A location-sensitive local prototype network for few-shot medical image segmentation. In: IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 262–266. IEEE (2021)
18. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. Technologies **9**(1), 2 (2020)
19. Liu, W., Wu, Z., Ding, H., Liu, F., Lin, J. and Lin, G.: Few-shot segmentation with global and local contrastive learning. arXiv preprint arXiv:2108.05293 (2021)
20. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. Adv. Neural Inf. Process. Syst. **33**, 12546–12558 (2020)
21. Zeng, D., et al.: Positional contrastive learning for volumetric medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 221–230. Springer (2021)
22. Wu, Y., Zeng, D., Wang, Z., Shi, Y., Hu, J.: Distributed contrastive learning for medical image segmentation. Med. Image Anal. **81**, 102564 (2022)
23. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
24. Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., et al.: NCI-ISBI 2013 Challenge: Automated Segmentation of Prostate Structures (2015)

25. Lemaitre, G., Marti, R., Freixenet, J., Vilanova. J. C., et al.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. In: Computers in Biology and Medicine. vol. 60, pp. 8–31 (2015)
26. Litjens, G., Toth, R., Ven, W., Hoeks, C., et al.: Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. In: Medical Image Analysis. vol. 18, pp. 359–373 (2014)
27. Liang, P.P., et al.: Think locally, act globally: federated learning with local and global representations. In: Workshop on Federated Learning at Advances in Neural Information Processing Systems. vol. 32 (2019)
28. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans. Med. Imaging **39**(9), 2713–2724 (2020)
29. Patro, S., Sahu, K.K.: Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462 (2015)