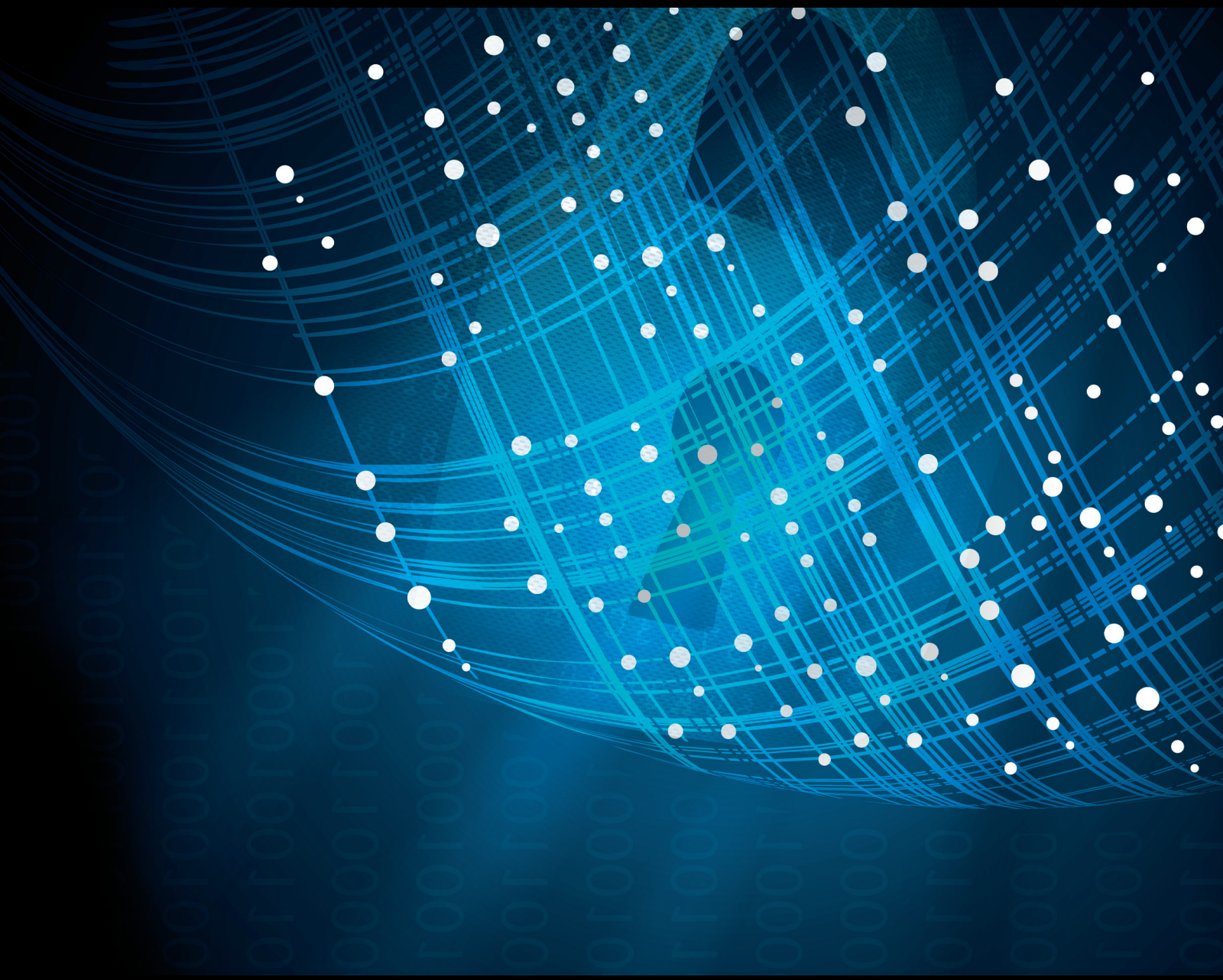
Security and Communication Networks

Deep Learning for Multimedia Security in Communication and Mobile Networks

Lead Guest Editor: Xiaoxian Yang
Guest Editors: Muddesar Iqbal and Yuyu Yin



Deep Learning for Multimedia Security in Communication and Mobile Networks

Security and Communication Networks

Deep Learning for Multimedia Security in Communication and Mobile Networks

Lead Guest Editor: Xiaoxian Yang
Guest Editors: Muddesar Iqbal and Yuyu Yin

Contents

Privacy Preservation Learning with Deep Cooperative Method for Multimedia Data Analysis
Wen Si and Cong Liu
Research Article (10 pages), Article ID 8449987, Volume 2022 (2022)

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8449987, 10 pages
https://doi.org/10.1155/2022/8449987

Research Article

Privacy Preservation Learning with Deep Cooperative Method for Multimedia Data Analysis

Wen Si [1,2] **and Cong Liu** [1]

[1]*Faculty of Business Information, Shanghai Business School, Shanghai 200235, China*
[2]*Huashan Hospital, Fudan University, Shanghai 200025, China*

Correspondence should be addressed to Cong Liu; tjcongliu@gmail.com

Received 10 June 2022; Accepted 11 July 2022; Published 12 August 2022

Academic Editor: Xiaoxian Yang

In parallel with the rapid adoption of deep learning to multimedia data analysis, there has been growing awareness and concerns about data security and privacy. The recent advancement of federated learning enables many network clients to collaboratively train a model under the orchestration of a central server while preserving clients' privacy. However, the standard assumption of independent and identical distribution (IID) may be broken under the federated learning because data label preferences may vary across clients. Recent efforts address this issue either by adapting a strong global model for each local model, respectively, or by training individual local models for similar clients together. However, both strategies degrade in highly non-IID scenarios. This work introduces a novel method, deep cooperative learning (DCL), to address this problem. It leverages the reciprocal structure between deep learning tasks in different clients to obtain effective feedback signals to enhance the learning process of personalized local models. To the best of our knowledge, this is the first time the non-IID is addressed under the principle of task interactions. We demonstrated the effectiveness of DCL on the two tasks of medical multimedia data analysis. The results show that our method presents a significant performance improvement compared with the standard federated learning method. In conclusion, this work developed a method for addressing non-IID problems in deep-learning-based privacy preservation learning. It allows the highly non-IID data to be used to improve the local model performance.

1. Introduction

In recent times, with the widely available medical imaging and computing devices, convolutional neural networks (CNNs) [1] have proven to be powerful tools for medical image segmentation task [2–6] and registration task [7, 8]. Segmentation is considered the most essential medical image process as it divides an image into the regions of interest based on anatomical structures or pathological tumors. The registration is the process of identifying a spatial transformation that maps two imaging modalities, such as CT (computed tomography) and MRI (magnetic resonance imaging), to common coordination such that corresponding anatomical structures are optimally aligned. The resulting pixelwise correspondence is fundamental for multimodality image analysis applications. Typically, training a CNN model requires patient scans to be transferred to a centralized data server where comprehensive analyses could be performed by using the parallel computing ability of the center. Given the increasing volumes of imaging data, the massive data collection and processing may be infeasible in a realistic scene because of the high throughput demands and the growing data privacy concerns. Federated learning [9] trains a global model collaboratively among a set of hospitals under the orchestration of a central server, without sharing their private raw data, so that a global model such as CNN-based segmentation can achieve better training performance than individually working alone. Also, since the data never leave the owner, the concerns about disclosing sensitive patient privacy and legal regulations are mitigated.

While federated learning works well on independent and identically distributed (IID) data, it experiences performance degradation on non-IID data [10, 11]. That is, the

data distribution of individual hospitals may be totally different from each other. The heterogeneous data distributions prevent the global model from convergence because of the conflicting updating directions that these distributions support. Unfortunately, non-IID happens often in real-world applications [12, 13]. For example, consider the cases of image segmentation, where there are two hospitals with different label preferences, as illustrated in Figure 1. The two hospitals annotate different labels for the temporal lobe due to individuals' preferences, although the underlying CT scan is the same. Global model aggregation becomes extremely hard in this case since a correct prediction for hospital 1 is incorrect for hospital 2. Having a single global model is insufficient for this case. It is more appropriate to train a personalized model for each hospital.

Recent efforts to address the non-IID issue can be classified into two strategies. The first strategy attempts to personalize a trained global model for each hospital with different label preferences. Personalization techniques for this category are classified into data-based and model-based approaches. Data-based approaches seek to reduce the local distribution divergences by balancing the distributions with a small amount of public [14, 15] or synthetic [16] data. These methods generally need to modify the local data distributions, which will disturb the local label bias, and thereby are not suited for our case. Instead of changing data, the model-based approaches learn a general global model for future personalization in individual hospitals by domain adaption learning that reduces the domain discrepancy between the global and local models [17–20] or meta-learning that enables the global model to adapt the private data quickly and effectively [21–24]. However, these methods presume the accessibility of a public proxy dataset that a global model will train on, which is unavailable in our case. In contrast to the first strategy that trains a single global model, the second strategy trains personalized models individually. Personalization techniques are classified into architecture-based and similarity-based approaches. The former achieves personalization by decoupling the local private model parameters from the shared global parameters [25], while the latter improves personalized model performance by enforcing stronger pairwise collaboration among hospitals with similar data distributions [26–28]. Both methods exploit pairwise data similarities between hospitals for improving local model performance, but other pairwise relations, such as the reciprocal structure between tasks, remain unexplored in current works.

We propose a new non-IID federated learning paradigm, deep cooperative learning (DCL), which leverages the reciprocal structure between federated learning tasks to obtain effective feedback signals to enhance the learning process of personalized models. We use the medical image segmentation and registration tasks with inherent complementary structures to build the cooperative learning loop. The principle of DCL is simple. Consider two hospitals tasked with the two tasks, respectively. If the two task models work well, the segmentation results, i.e., anatomical structures, could be combined with the input of the registration model to boost its performance since the extra anatomical information helps the registration model find the right alignment of the anatomies. Similarly, since some anatomies are only visible on MRI, the aligned MRI produced from the registration model could be combined with CT to provide extra modality for the segmentation model. More importantly, DCL shares the models rather than model outputs among hospitals during the cooperative training loop, thereby achieving personalization and preserving data privacy simultaneously.

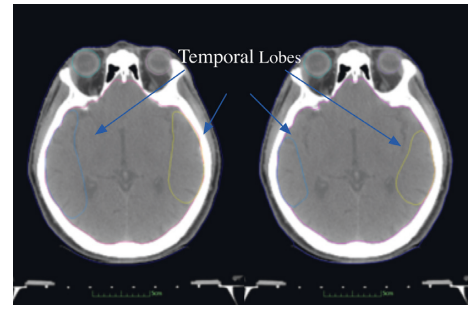


Figure 1: Label (temporal lobe) preferences of two hospitals involved in federated learning.

2. Methods

The deep cooperative learning consists of two steps. First, a reward mechanism is designed to promote mutual benefits between two tasks from different hospitals. The gain produced by one task to another is regarded as reward and fed back to the task model for adjusting its subsequent behavior for better performance. The non-IID labels are shared in this way among hospitals to share task experience and improve the model generalization. Second, a cooperative training mechanism between task models is created which treats individual model as a parameterized agent to maximize its long-term reward. Below, we provide detailed explanations for the two steps.

2.1. Reward. We design the reward mechanism via the deep discriminator networks. Let c and m be unlabeled CT and MRI images, respectively, and $\mathrm{Reg}(.)$ and $\mathrm{Seg}(.)$ be trained registration and segmentation networks/models, respectively. The circulation between task models can be summarized as follows (see Figure 2). $\mathrm{Reg}(c, m)$ registers CT and MRI. $\mathrm{Seg}(\mathrm{Reg}(c, m))$ segments the output of registration model. $\mathrm{Seg}(c)$ and $\mathrm{Seg}(m)$ segment CT and MRI, respectively, and then $\mathrm{Reg}(\mathrm{Seg}(c), \mathrm{Seg}(m))$ registers the results of segmentation models. Suppose $P_{\mathrm{reg}}(.)$ and $P_{\mathrm{seg}}(.)$ are discriminator networks after the adversarial training measuring the confidence of the outputs of the segmentation and registration networks. We define reward 1 as

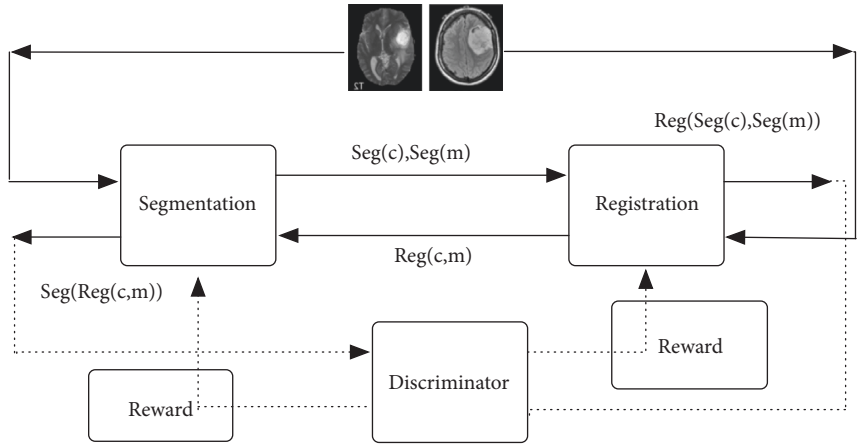


Figure 2: Schematic of the proposed DCL method.

$$r_1 = \log\left(P_{\text{reg}}\left(\text{Reg}\left(\text{Seg}\left(c\right), \text{Seg}\left(m\right)\right)\right) - \log\left(P_{\text{reg}}\left(\text{Reg}\left(c, m\right)\right)\right)\right), \tag{1}$$

where the subtraction measures the difference between direct registration and the segmentation-then-registration, i.e., the promotion derived from the segmentation results. Similarly, we define reward 2 as

$$r_2 = \log\left(P_{\text{seg}}\left(\text{Seg}\left(\text{Reg}\left(c, m\right)\right)\right)\right) - \frac{1}{2}\left(\log\left(P_{\text{seg}}\left(\text{Seg}\left(c\right)\right)\right) + \log\left(P_{\text{seg}}\text{Seg}\left(m\right)\right)\right). \tag{2}$$

to measure the difference between direct segmentation and the registration-then-segmentation, i.e., the promotion derived from the registration results.

2.2. Cooperative Training. With the defined rewards 1 and 2, we next design the cooperative training mechanism. The segmentation and registration networks are treated as parametric representations of the policy and we use a policy gradient algorithm [29] to update these parameters alternatively through federated learning to achieve cooperative learning. If a large or positive reward is observed after performing an action (a parameter update), its gradient is added to the parameters of the current policy function to increase the probability of performing this action at this state. On the contrary, if a small or negative reward is observed after performing an action, its gradient is subtracted from the parameters of the current policy function to decrease the probability of performing this kind of action under this state. Formally, letting the parameters of segmentation [2] and registration networks be θ_{reg} and θ_{seg}, respectively, and the number samples of a mini-batch be K, then the stochastic gradient can be written as

$$\begin{aligned} \nabla_{\theta_{\text{seg}}} E\left[r_1\right] &= \sum_{K=1}^{K} r_{1k}, \\ \nabla_{\theta_{\text{reg}}} E\left[r_2\right] &= \sum_{K=1}^{K} r_{2k}, \end{aligned} \tag{3}$$

where the parameters θ_{reg} and θ_{reg} could be updated according to the policy gradient. The cooperative training algorithm could be summarized as Algorithm 1 in Figure 3.

2.3. Implementation Details. We use the U-Net structure [30] for the segmentation network as illustrated in Figure 4. U-Net is considered one of the standard CNN architectures for image segmentation. The unique skip-connection layers of U-Net can capture the image features at multiple scales while avoiding the loss of the high-frequency details. We further improve the performance of U-Net with two modifications. First, the squeeze-and-excitation block is introduced to adaptively extract image features after each convolution in the U-Net encoder. Second, to avoid the resolution degradation caused by pooling and downsampling, the last pooling layer and downsampling layer of the network are changed to Atrous Spatial Pyramid Pooling (ASPP) [31] block which uses different perceptual field sizes around a single pixel and fuses the convolution results to detect small targets at multiple resolutions.

We use a two-stream regression network for the registration network as illustrated in Figure 5. Each stream takes an imaging modality as input and outputs its feature map. The subsequent regression layers predict the shifts between two images based on their feature maps. We further use the attention mechanism [32, 33] that mimics human attention improving its performance.

Algorithm 1: Deep Cooperative Learning

input :

Trained segmentation and registration models

$Seg(\cdot)$, $Reg\cdot)$, $P_{seg}(\cdot)$, $P_{reg}(\cdot)$

Labels from different hospitals

Learning rate γ

output Improved networks

:

repeat

t = t + 1

for k=1 to K do

Draw samples from ck and mk

$c = c_k$, $m = m_k$

Using $Reg(\cdot)$ register c, m to obtain $Reg(c, m)$

Using $Seg(\cdot)$ segment $Reg(c, m)$ to obtain $Seg(Reg(c, m))$

Using $Seg(\cdot)$ segment c, m to obtain $Seg(c)$, $Seg(m)$

Using $Reg(\cdot)$ register $Seg(c)$, $Seg(m)$ to obtain $Reg(Seg(c), Seg(m))$

Cal. reward $r_1 = log(P_{reg}(Reg(Seg(c), Seg(m)))) - log(P_{reg}(Reg(c,m)))$,

$r_2 = log(P_{Seg}(Seg(Reg(c,m)))) - 1/2log(P_{Seg}(Seg(c))), + log(P_{Seg}Seg(m)))$

$r_{1_k} = r_1, r_{2_k} = r_2$

end

Cal gradient $\nabla_{\theta_{Seg}} E[r_1] = \Sigma_{k=1}^{K} \nabla_{\theta_{Seg}} r_{1_k}$, $\nabla_{\theta_{Reg}} E[r_2] = \Sigma_{k=1}^{K} \nabla_{\theta_{Reg}} r_{2_k}$

Update parameters $\theta_{Seg} = \theta_{Seg} + \gamma \nabla \theta_{Seg} E[r_1]$, $\theta_{Reg} = \theta_{Reg} + \gamma \nabla \theta_{Reg} E[r_2]$

until *Convergence or t > Max iter*.

Figure 3: Deep cooperative learning algorithm.

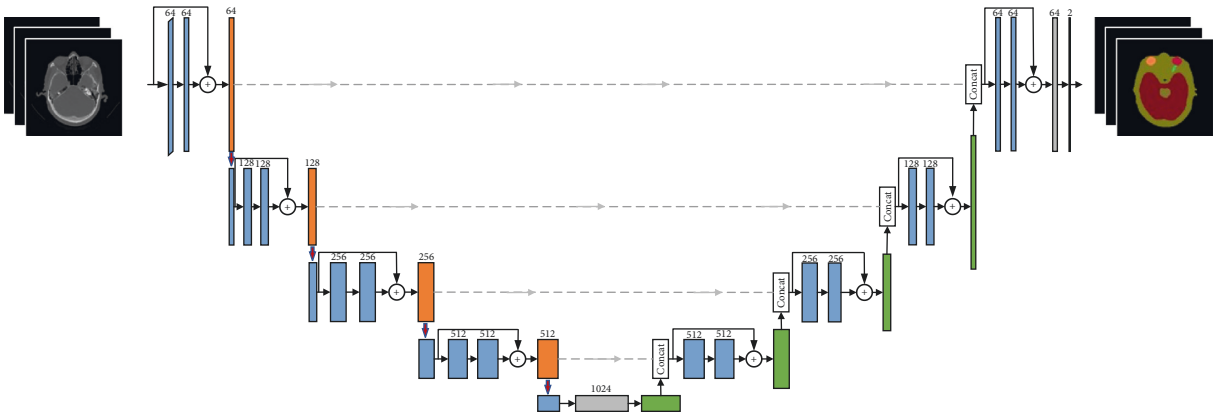


Figure 4: U-Net with SE and ASPP modifications for segmentation task.

The attention mechanism enhances the unaligned parts of two images, thereby concentrating the limited computational resource on them.

In the original generative adversarial network [34], the generated images or the real images are alternately fed into the discriminator network, and the generated images are not preprocessed in any way. Considering the characteristics of the segmentation and alignment tasks, we preprocess the generated images based on the attention mechanism to strengthen the relationship between the generated images and the real tokens, emphasizing their higher-order semantic inconsistencies for the evaluation of the generated images by the discriminator.

To prepare the input images for segmentation and registration models, we first resize the CT-MRI slice pairs from 512×512 to 480×480 pixel size and then randomly crop the downsampled images to 384×384 pixel size. We use the random crop to increase the sample size. The

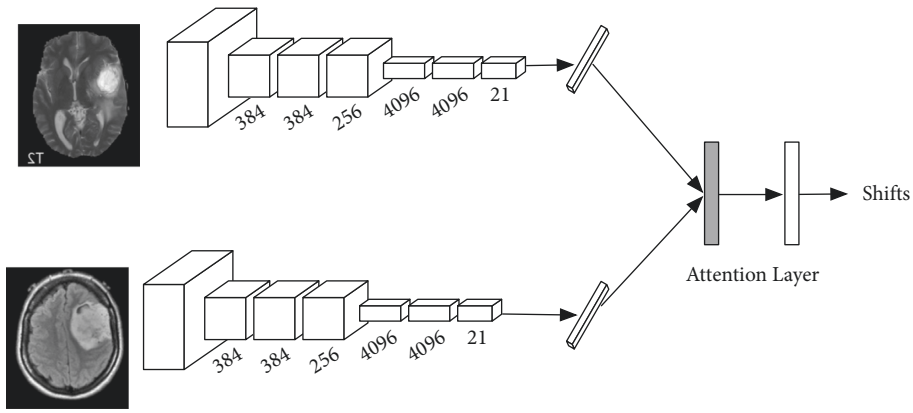


Figure 5: Two-stream regression network with attention layer for registration task.

384×384 pixel size facilitates the downsampling operations in the networks because it could be divided by 2 many times with no remainder. Before feeding the image into the networks, images are normalized to zero mean with unit-variance intensities and augmented with a random horizontal flip.

2.4. Evaluation Metrics. We use Dice coefficient and Hausdorff distance (HD) to evaluate the quality of segmentation. The Dice coefficient is computed as the area of overlap between the prediction (pred) and the ground truth (GT) divided by the total number of pixels in prediction and ground truth:

$$\text{Dice} = \frac{2|\text{Pred} \cap \text{GT}|}{|\text{Pred}| + |\text{GT}|}. \tag{4}$$

The HD measures the boundary distance (D) between predictions and ground truth and is defined as

$$\text{HD} = \max\left(\max_{gt \in GT} \min_{\text{pred} \in \text{Pred}} D(gt, p), \max_{\text{pred} \in \text{Pred}} \min_{\text{gt} \in \text{GT}} D(gt, p)\right). \tag{5}$$

Since the HD metric is sensitive to outliers, we report 95th-percentile HD (HD95) instead.

We regard a registration prediction as successful if the shifts differences are <3 mm in both x and y directions. For I sets of image pairs, we let x_i and y_i denote the shift in the x and y directions, respectively. First, we calculate the number of image pairs that can meet $|x_i| < 3$ and $|y_i| < 3$ and denote that number as J. The registration accuracy is then defined as

$$\text{RegAcc} = \frac{\text{I}}{J} \times 100\%. \tag{6}$$

3. Results

We collected 178 and 81 patients with head and neck cancer from two hospitals, respectively, for this study. The training dataset consists of 142 and 64 patients, respectively, for the two clients and the remaining clients were used to evaluate the DCL performance. We preprocess the images before feeding them into the models. The DCL training protocol was implemented with the TensorFlow federated learning framework [35] on NVIDIA TITAN XP GPU. All networks are initialized with the Xavier initializer and trained with the Adam optimizer, the learning rate of 1e- 4, the batch size of 4, and a total of 25 k updates [36]. An appropriate learning rate is critical in our experiments. We found that learning rate larger than 1e-4 will cause loss oscillation.

We updated the parameters of networks with a stochastic gradient descent method where the initial learning rate was set to 1e-4, and a total of ~25 k updates are used to train the networks.

We first compared the results from the segmentation network for hospital 1 with and without DCL method qualitatively. Twenty-four anatomy structures were used in the study including brain, spinal cord, spinal cord cavity, pituitary, parotid glands, oral cavity, mandible, mandible joint, temporal lobes, and so on. Figure 6 shows the segmentation results of standard federated learning (left) and the results of DCL federal learning (right). Since hospital 2 prefers smaller temporal lobes while hospital 1 prefers larger ones, the label conflict causes the model of hospital 1 to produce an undesired small temporal lobe (arrow). However, with DCL, a consistent temporal lobe (arrow) is predicted by the segmentation model of hospital 1.

We also show the segmentation results from hospital 2 in Figure 7. We find that federated learning with DCL outperforms standard federated learning in small organs such as eyeballs (arrow). The reason could be attributed to the fact that federated learning could increase the

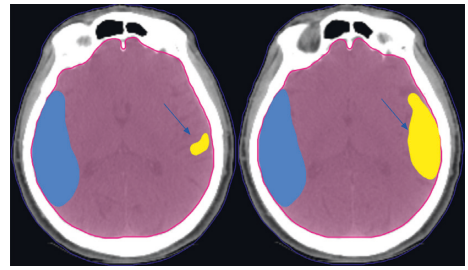


Figure 6: Comparison of segmentation results of hospital 1 on two CT slice images without DCL (left) and with DCL (right).

Figure 7: Comparison of segmentation results of hospital 2 on two CT slice images without DCL (left) and with DCL (right).

Table 1: Comparison of segmentation Dice values of federated learning with and without the proposed DCL method (large is better).

Organs	Non-IID with DCL method	Non-IID with standard federated learning
Brain	0.985	0.867
Brain stem	0.859	0.881
Spinal cord	0.872	0.856
Spinal cord cavity	0.889	0.73q
Eye L	0.935	0.897
Eye R	0.935	0.919
Len L	0.888	0.793
Len R	0.914	0.746
Optical nerve L	0.894	0.693
Optical nerve R	0.907	0.718
Chiasm	0.883	0.618
Pituitary	0.915	0.871
Parotid L	0.820	0.839
Parotid R	0.826	0.847
Oral cavity	0.918	0.948
Mandible	0.928	0.925
Mandible joint L	0.761	0.824
Mandible joint R	0.794	0.837
Temporal lobe L	0.855	0.848
Temporal lobe R	0.875	0.841
Larynx	0.896	0.933
Pharynx	0.818	0.672
Trachea	0.821	0.812
Thyroid	0.715	0.827
Average	0.871	0.830
p value	<0.05	

relatively insufficient training samples for the small organs.

Table 1 provides the Dice values of segmentation results for hospital 2. It is observed that DCL improved the average Dice value by 5.49% over standard federated learning. We perform Student's t-test on the paired groups of standard federated learning and DCL for all organs. The p value of 0.02 (<0.05) leads to the conclusion that DCL outperforms standard federated learning significantly in terms of the Dice metric. Compared with standard federated learning, DCL helped the segmentation network recognize small organs such as pituitary and optic nerves.

Table 2 reports the comparison of HD95 values of the federated learning with and without the proposed DCL method. As shown in the table, the mean HD95 value is improved by 2.2 mm when federated learning is used with the DCL method. Student's t-test shows that DCL outperforms standard federated learning significantly in terms of the HD95 metric ($p = 0.003$). It is also noted that the HD95 of small volume organs such as crystal, optic chiasma, optical

Table 2: Comparison of average 95th percentile HD values of federated learning with and without the proposed DCL method (small is better; unit: mm).

Organs	Non-IID with DCL method	Non-IID with standard federated learning
Brain	1.65	2.62
Brain stem	3.54	4.89
Spinal cord	4.09	7.53
Spinal cord cavity	3.47	4.38
Eye L	1.82	2.85
Eye R	1.16	2.74
Len L	1.04	2.94
Len R	1.11	2.63
Optical nerve L	1.22	3.96
Optical nerve R	1.53	4.91
Chiasm	1.22	4.58
Pituitary	1.40	2.05
Parotid L	5.01	7.22
Parotid R	4.65	6.96
Oral cavity	2.37	7.67
Mandible	5.34	2.58
Mandible joint L	4.61	3.02
Mandible joint R	2.66	2.95
Temporal lobe L	3.07	12.24
Temporal lobe R	2.24	14.27
Larynx	3.26	6.48
Pharynx	2.57	2.66
Trachea	11.83	21.10
Thyroid	10.52	3.96
Average	3.39	5.61
p value	$0.003 < 0.05$	

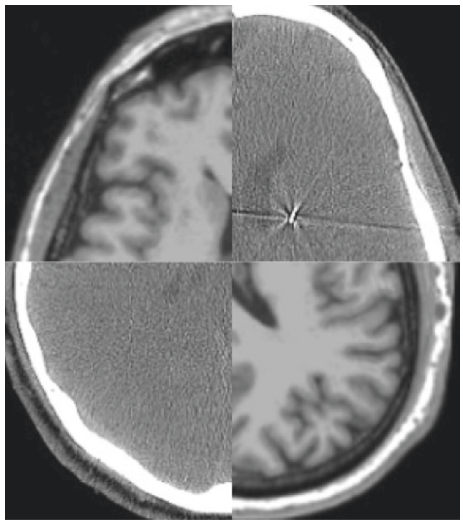


Figure 8: Illustration of the registration results.

nerves, and pituitary is much smaller than that of the large organs. This means that the inconsistency label issue is more significant in small organs and DCL could alleviate it substantially.

We illustrate an example of the registration result for an image pair in Figure 8. We find that the corresponding anatomical structures such as cranium and brain are aligned correctly.

We further provide the numerical comparison in successful registration rates in Table 3. We also performed a chi-square test on RegAcc between the standard federated learning and the federated learning with the DCL method. 53 registration result pairs are involved in the test. It is observed that the registration network with DCL outperforms the standard federated learning.

In Figure 9, we plot the loss values as the function of training steps. The top subfigure shows the segmentation cross-entropy loss and the bottom subfigure shows the registration mean squared error. The standard federated learning and the DCL are plotted with gray color and red color, respectively. As illustrated in the figure, we find that the extra supervised signal from DCL prevents the segmentation training from overfitting. The segmentation quality is steadily improved after 10 k training steps. In registration training, the overfitting phenomenon is not observed, but the training of standard federated learning is stuck at a high error level.

Table 3: Comparison of registration shift and accuracy of federated learning with and without the proposed DCL method.

Method	Shifts (mm)	RegACC
Non-IID with standard federated learning	2.81	79.4%
Non-IID with DCL method	1.39	81.2%
p value	<0.05	<0.05

4. Discussion

In the previous section, we demonstrated the feasibility of exploiting the reciprocal structure between segmentation and registration task among different hospitals to improve

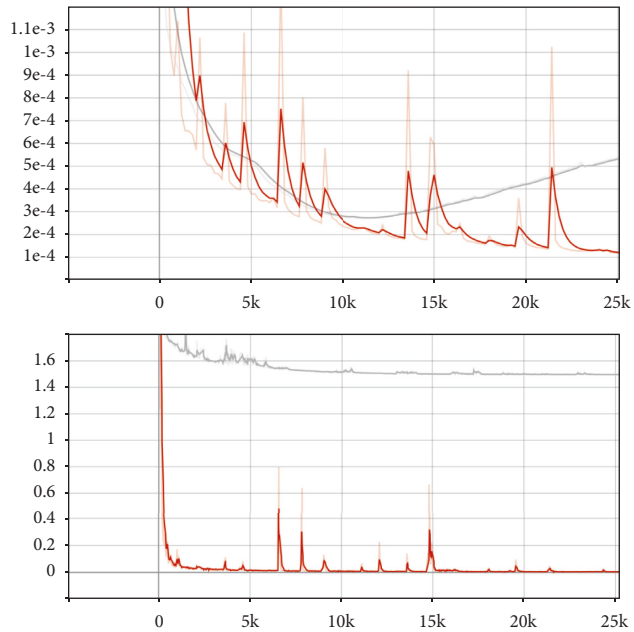


Figure 9: Training curve comparison between standard federated learning (gray) and DCL (red).

the performance of the segmentation and registration model in individual hospitals. The experiment results suggest that the proposed DCL method is necessary and contributes significantly to performance improvement. The proposed DCL method outperforms the standard federated learning by 5.49%, 2.2 mm, and 1.8% in terms of Dice, 95th percentile HD, and registration accuracy. The superior performance of DCL could be attributed to the model cooperation among a set of hospitals under the orchestration of the DCL.

The proposed DCL method is different from the global model personalization methods that personalize a single global model for each client through data or model adaptations that involve additional training on each local dataset [14, 15, 37, 38]. While these methods aim to collaboratively train a shared model without sharing private data, DCL is designed to enhance the local model with the help of other hospitals but still preserve the personalization of the local model. In contrast to global model personalization methods, personalization or preference is never lost for each client.

The proposed DCL method could be classified to the catalog of learning personalized models that build personalized models by modifying the federated learning model aggregation process. Our method is most close to the similarity-based approaches in this catalog which leverage client pairwise similarities to improve personalized model performance where similar personalized models are built for related clients. FedAMP [28] excels in capturing pairwise client relationships to learn similar models for related clients. It may be sensitive to poor data quality, whereas DCL leverages tasks' complementarity to provide extra supervision signals and thereby is not affected by the data quality. Model interpolation methods [39] learn personalized models using a mixture of global and local models. However, they are likely to experience a degradation in performance in highly non-IID scenarios as they use a single global model as a basis for personalization. In contrast to the model interpolation methods, DCL can work under any data distribution. In this study, the data could be totally different for the segmentation task and the registration task. Overall, the major novelty of DCL is the exploitation of the task reciprocal structure, whereas current non-IID approaches mainly leverage the data similarities. The cooperative relationship is exploited to provide mutual rewards or pseudo-labels for the tasks of different hospitals. Since the reward is extracted from the first hospital's A task to the second hospital's B task, the biased data distributions are mitigated during the learning loop. The primary benefit of task cooperation is the robustness of the data distributions.

While the task cooperation improves the local model performance under highly non-IID scenarios, it also prevents the tasks that do not contain reciprocal structures from the DCL protocol. Additionally, since the trained models are shared between hospitals, the label preferences of one hospital may be leaked to others and may increase the risk of privacy exposure. Furthermore, it is still not clear to what extent these methods harm data privacy, and there are no quantitative measures to identify the degree of privacy leakage. Finally, while DCL is an effective method for the non-IID problem, other issues remain as open questions for the future when using federated learning for healthcare including decentralized online optimization [40], unbalanced data [41], limited communication bandwidth [42], and unreliable and limited device availability [43].

5. Conclusions

We developed a method named deep cooperative learning (DCL) to address the non-IID problem in federated learning. Comprehensive experiments have been carried out on CT and MRI segmentation and registration tasks and datasets to demonstrate the effectiveness of DCL. The results obtained from head-neck cancer patients of two hospitals show that the method outperforms the standard federated learning in segmentation and registration tasks. The method is, therefore, a solution for leveraging biased labels across hospitals.

Data Availability

The experiment data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Natural Science Foundation of Shanghai (20ZR1440300).

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[3] W. Huang, Y. Zeng, L. Chen et al., "AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Medical Physics*, vol. 46, no. 2, pp. 576–589, 2019.

[4] H. Gao, J. Xiao, Y. Yin, T. Liu, and J. Shi, "A mutually supervised graph attention network for few-shot segmentation: the perspective of fully utilizing limited samples," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2022.

[5] J. Xiao, H. Xu, H. Gao, M. Bian, and Y. Li, "A weakly supervised semantic segmentation network by aggregating seed cues: the multi-object proposal generation perspective," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 1s, pp. 1–19, 2021.

[6] T. Liu, X. Feng, R. Wang et al., "Discriminative cervical lesion detection in colposcopic images with global class activation and local bin excitation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1411–1421, 2022.

[7] C. Liu, L. Ma, Z. Lu, X. Jin, and J. Xu, "Multimodal medical image registration via common representations learning and differentiable geometric constraints," *Electronics Letters*, vol. 55, no. 6, pp. 316–318, 2019.

[8] F. P. Tavares and J. M. R. Tavares, "Medical image registration: a review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 2, pp. 73–93, 2014.

[9] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[10] Z. Zhang, "Semi-supervised federated learning with non-IID data: algorithm and system design," in *2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application*vol. 12, pp. 157–164, HPCC/DSS/SmartCity/DependSys, 2021.

[11] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–17, 2022.

[12] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, *On the Convergence of Fedavg on Non-iid Data*, arXiv preprint arXiv: 1907.02189, 2019.

[13] Q. Li, Y. Diao, Q. Chen, and B. He, *Federated Learning on Non-iid Data Silos: An Experimental Study*, arXiv preprint arXiv:2102.02079, 2021.

[14] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, *Mixup: Beyond Empirical Risk Minimization*, arXiv preprint arXiv: 1710.09412, 2017.

[15] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, *Federated Learning with Non-iid Data*, arXiv preprint arXiv: 1806.00582, 2018.

[16] F. Sattler, S. Wiedemann, K.-R. Muller, and W. Samek, "Robust and communication-efficient federated learning from non-i.i.d. Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3400–3413, 2020.

[17] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2351–2363, 2020.

[18] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "Fedhealth: FedHealth: a federated transfer learning framework for wearable healthcare," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.

[19] H. Yang, H. He, W. Zhang, and X. Cao, "FedSteg: a federated transfer learning framework for secure image steganalysis," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1084–1094, 2021.

[20] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[21] R. Drissi and Y. Drissi, "A perspective view and survey of meta-learning," *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.

[22] Y. Jiang, J. Konečnỳ, K. Rush, and S. Kannan, *Improving Federated Learning Personalization via Model Agnostic Meta Learning*, arXiv preprint arXiv:1909.12488, 2019.

[23] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3557–3568, 2020.

[24] C. T Dinh, N. Tran, and J. Nguyen, "Personalized federated learning with moreau envelopes," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21394–21405, 2020.

[25] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, *Federated Learning with Personalization Layers*, arXiv preprint arXiv:1912.00818, 2019.

[26] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[27] L. Corinzia, A. Beuret, and J. M. Buhmann, *Variational Federated Multi-Task Learning*, arXiv preprint arXiv: 1906.06268, 2019.

[28] Y. Huang, "Personalized cross-silo federated learning on non-iid data," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, pp. 7865–7873, 2021.

[29] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *International conference on machine learning*, vol. 24, pp. 387–395, 2014.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation, lecture notes in computer science," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*pp. 234–241, Munich, Germany, 2015.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[32] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[33] Q. Wang, J. Zhang, S. Song, and Z. Zhang, "Attentional neural network: feature selection using cognitive feedback," *Advances in Neural Information Processing Systems*, vol. 74, pp. 2033–2041, 2014.

[34] I. Goodfellow, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*,pp. 2672–2680, Montréal, Canada, 2014.

[35] M. Abadi, "TensorFlow: learning functions at scale, ACM SIGPLAN Notices," *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming*, vol. 51, no. 9, p. 1, 2016.

[36] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.

[37] H. Eichner, T. Koren, B. McMahan, N. Srebro, and K. Talwar, "Semi-cyclic stochastic gradient descent," in *International Conference on Machine Learning*, pp. 1764–1773, 2019.

[38] T. Wang, J.-Y. Zhu, A. Torralba, and A. A. Efros, *Dataset Distillation*, arXiv preprint arXiv:1811.10959, 2018.

[39] F. Hanzely and P. Richtárik, *Federated Learning of a Mixture of Global and Local Models*, arXiv preprint arXiv:2002.05516, 2020.

[40] R. Luo, F. Tian, T. Qin, E. Chen, and T.-Y. Liu, "Neural architecture optimization," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[41] K. Balachandar, N. Lam, C. Yi et al., "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 945–954, 2018.

[42] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 317–333, 2020.

[43] V. Tolpegin, S. Truex, M. E. Gursoy, and L. Liu, "Data poisoning attacks against federated learning systems, Computer Security - ESORICS 2020," *European Symposium on Research in Computer Security*, pp. 480–501, 2020.