Using Personalized Federated Learning To Train Diffusion Models

Marios Hadjigeorgiou

CSE3000 Research Project

Responsible Professor: Jérémie Decouchant | Supervisor: Bart Cox

TU Delft Delft University of Technology

Diffusion Models

Forward propagation:updates the state of pixels over time by adding noise

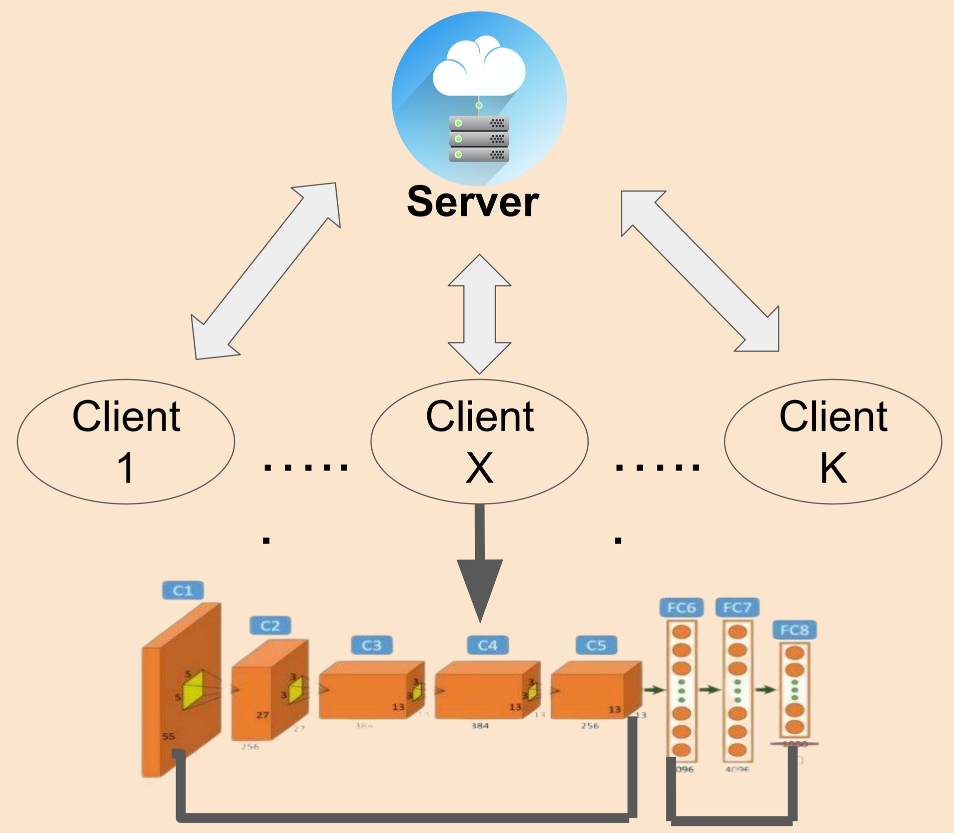
Backward Propagation: reconstructs the original input from the final diffused state

- The training process of TL begins with each client retrieving images and labels from their own dataset.
- Instead of comparing the predicted output with the respectable labels, in diffusion models noise images and predicting noise images are produced.
- The goal of this process is to acquire the loss value by measuring the difference between the two sets of images and produce samples that resemble the input data..

Research Questions

- Using Personalized Federated Learning (PFL) to train Diffusion Models more efficiently.
- Tuning specific hyperparameters to observe the difference in personalization scores for Transfer Learning (TL) .
- Comparing the evaluation results with those of other personalization techniques.

Federated Learning using Transfer Learning



Base Layers + *Personalization Layers*

- The base layers are frozen and each client trains the personalized layers of the model locally.
- The aggregated step involves the averaging of the personalized weights of the clients while the base layers are left intact.
- By using a pre-trained model, the computational cost of training the new personalized model is reduced.

	User 1	User 2	User 3	User 4	User 5
IID data	141.13	131.84	126.70	152.20	119.54
non-IID data	189.81	274.24	202.21	184.39	190.41

Per-User FID score using the traditional global model trained with FedAvg

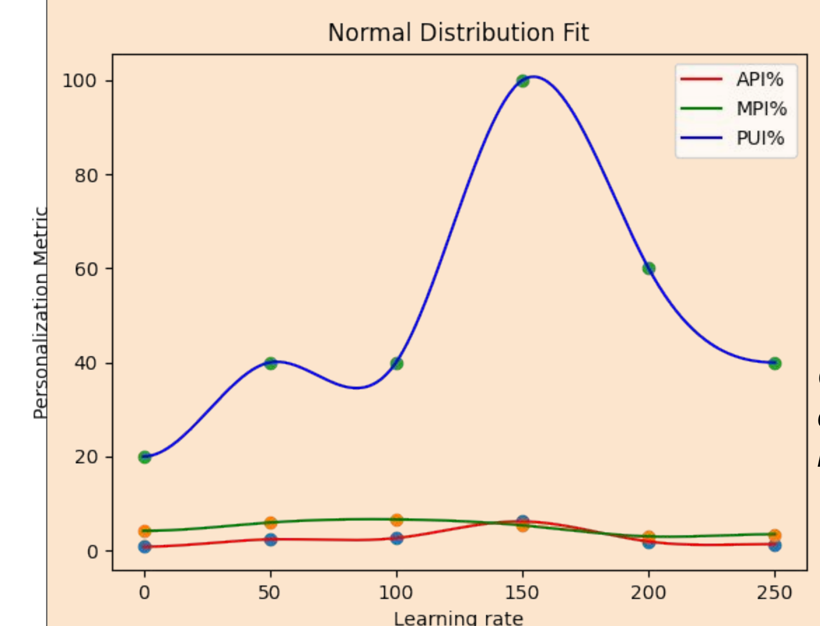
→ The per user score results are significantly lower when the global model is applied to each user's limited dataset on non-IID data compared to IID data.

→ The non-IID data distribution shows clearly that data heterogeneity has a significant role in the performance score of the algorithm.

Results

Personalization Metrics

- **API** : Average Percentage of Improvement.
- **MPI** : Median Percentage of Improvement
- **PUI** : percentage of users with a personalized model that produces better performance

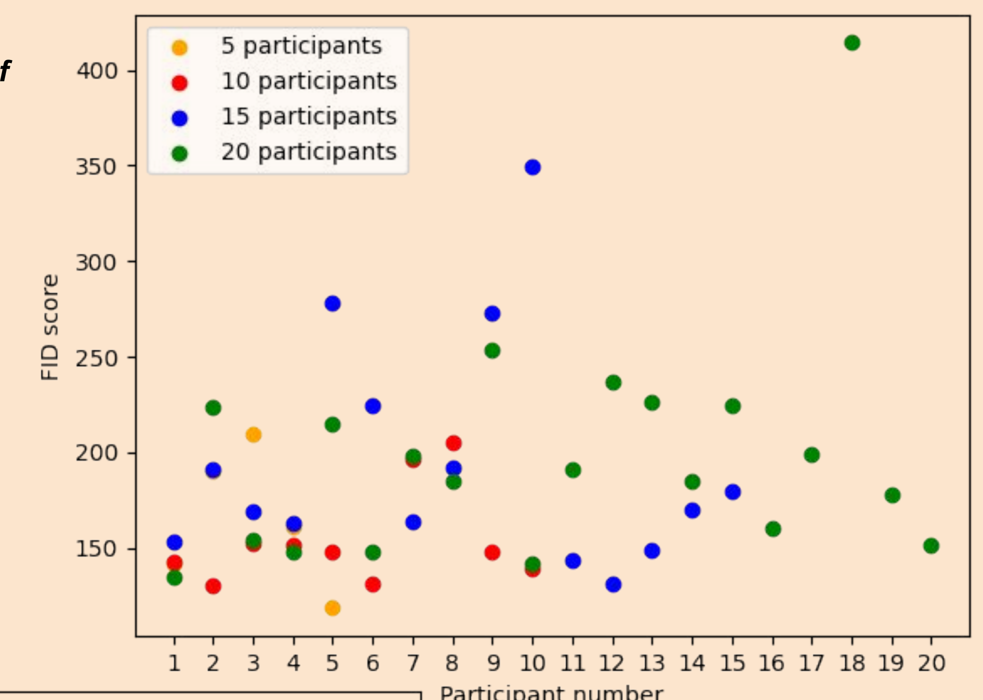


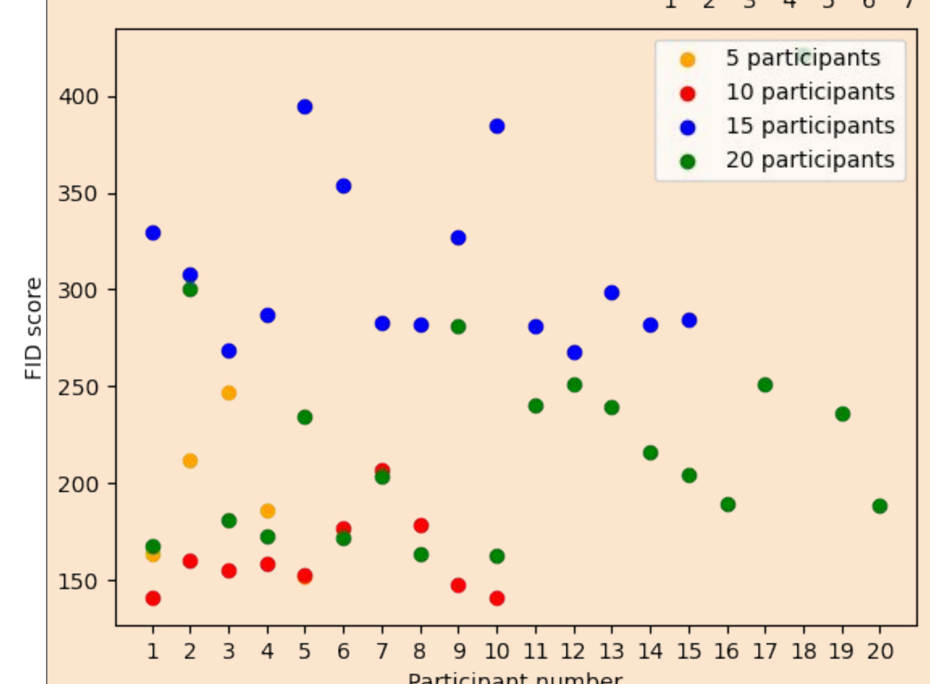
Learning rate	API(%)	MPI(%)	PUI(%)
0.01	0.0	0.0	0
0.001	**13.66**	**20**	**100**
0.0001	6.19	5.36	**100**
0.00001	3.06	15.30	20
SSFD 1	2.24	11.21	20
SSFD 2	0.0	0.0	0
SSFD 3	0.0	0.0	0

Global model personalization scores for different numbers of learning rates on non-IID dataset

Global model personalization scores for different numbers of base layers on non-IID dataset

Local model performance scores for different numbers of participants on non-IID dataset





Global model performance scores for different numbers of participants on non-IID dataset

Number of Users	API(%)	MPI(%)	PUI(%)
5	**13.66**	**20.00**	**100**
10	5.95	4.74	70
15	11.32	9.20	**100**
20	12.53	14.40	90

Conclusions

	API(%)	MPI(%)	PUI(%)
TL in isolation	5.59	7.53	**100**
TL in FL setting	**13.66**	**20.00**	**100**

- Training the diffusion model under a FL setting converges more accurately than training in isolation. Both API and MPI scores are 2X improved.

Metrics	PersFL	FedPer	pFedMe	perFed
PUI(%)	100	100	100	100
MPI(%)	11.23	6.59	10.81	8.55
API(%)	10.83	6.41	10.47	8.85

Comparison with other Personalization methodologies

- Compared to previous research, our results agree that TL improves the personalization results for the new local model compared to the global model
- Our algorithm obtains **100% PUI**, **13.66% API** and **20% MPI** score, outperforming other Personalized Federated Learning methods, in specific non-IID data settings.

	Global Model	Personalized Model
User 1	189.81	137.14
User 2	274.24	187.08
User 3	202.21	136.29
User 4	184.39	127.33
User 5	190.41	124.08

Hyperparameter Tuning results

→ the number of base layers and the learning rate form a normal distribution where any value above or below the optimal option results in overfitting and underfitting respectively and a less optimal personalization score.

→ The number of participants showed unstable performance in terms of both converge and personalization scores.