Supplementary Material

Jinglin Liang[1], Jin Zhong[1], Hanlin Gu[2], Zhongqi Lu[3], Xingxing Tang[2], Gang Dai[1], Shuangping Huang[1,5⋆], Lixin Fan[4], and Qiang Yang[2,4]

[1]South China University of Technology,
[2]The Hong Kong University of Science and Technology,
[3]China University of Petroleum, [4]WeBank, [5]Pazhou Laboratory
eeljl@mail.scut.edu.cn, eehsp@scut.edu.cn

We organize the supplementary material as follows.

- In Section A, we analyze the privacy protection capabilities of our proposed DDDR framework.
- In Section B, we discuss the time and transmission efficiency of DDDR.
- In Section C, we present additional generated samples.
- In Section D, we analyze the generalization capabilities of our proposed Federated Class Inversion.
- In Section E, we present the experimental results of local testing on each client.

A Privacy concerns

A.1 Integration of privacy protection methods

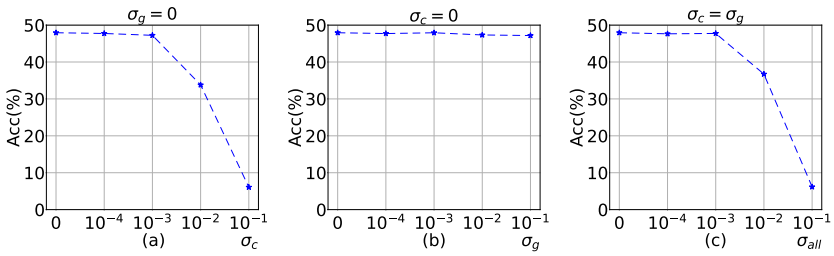

Fig. 1: Variations in the average accuracy of DDDR across different noise intensities, on the Cifar-100 dataset with 5 tasks and non-IID data distribution. σ_c and σ_g denote the standard deviations of Gaussian noise introduced to classifier parameters and class embeddings, respectively. (a) With σ_g set to 0, observing the effect of σ_c on average accuracy. (b) With σ_c set to 0, examining the impact of σ_g on average accuracy. (c) Introducing noise to both class embeddings and classifier parameters to assess their collective influence on average accuracy.

⋆ Corresponding Author

To assess the efficacy of privacy protection strategies within the DDDR framework, we incorporate the widely used randomization privacy protection strategy [6, 12] into DDDR. Specifically, during each round of communication, clients first augment their class embeddings and classifier parameters with Gaussian noise before uploading to the server. This approach significantly lowers the success rate of gradient inversion attacks [12], thus preventing the server or any other federated participants from deducing private data.

Figure 1 illustrates the variation in the model's average accuracy with the introduction of noise intensity. As expected, an increase in the noise intensity added to the classifier parameters leads to a reduction in classifier performance, due to the trade-off between privacy protection and model performance [6]. Unexpectedly, the intensity of noise added to the class embeddings has a minimal impact on model performance.

To explore the reasons behind this, we generate images using class embeddings trained under different noise intensities, which are presented in Figure 2. We observe that the generative quality of class embeddings trained under various noise intensities remains similar. Even at a noise intensity with a standard deviation of 0.1, it is still able to achieve desirable generative outcomes. This may be attributed to the training objective of Federated Class Inversion, which involves searching for an optimal embedding within the input space of a pre-trained conditional diffusion model. Given that this model has been pre-trained on a vast amount of data, its input embedding space is relatively smooth, meaning that perturbations to the embedding do not significantly alter the generative results.

In summary, the randomization privacy protection strategy can be applied to the DDDR framework to enhance privacy protection. Furthermore, our proposed Federated Class Inversion method's generative quality is insensitive to the intensity of noise added, implying that this method can enhance privacy protection without noticeably compromising performance, thereby achieving a effective balance between model performance and privacy protection.

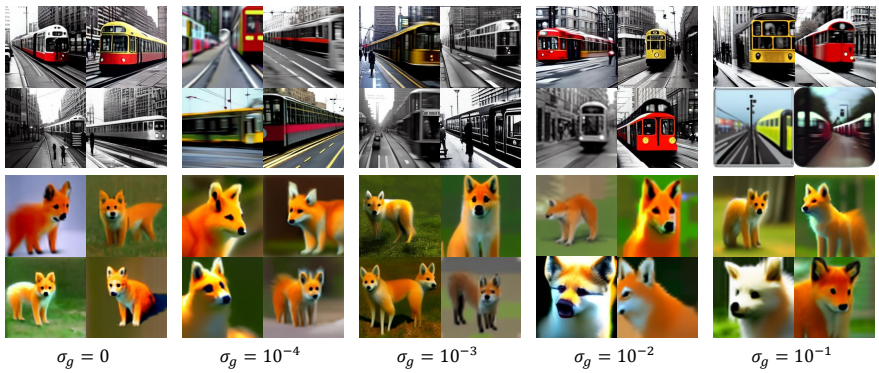

Fig. 2: Showcase of DDDR-generated samples under different noise intensities. σ_g denotes the standard deviation of noise added to the class embeddings uploaded by clients.

A.2 Gradient inversion attacks

Transmitting only class embeddings in Federated Class Inversion is secure. We attempted to reconstruct training images from gradients of class embeddings using gradient inversion attacks [12] but were unsuccessful, as shown in the Figure 3.

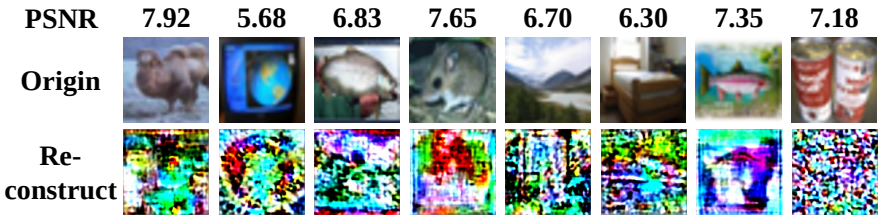

Fig. 3: Results of applying gradient inversion attacks on Federated Class Inversion.

A.3 The likelihood of generating the original data

It is unlikely to generate images that are identical to the original data. We randomly selected 5 classes from CIFAR-100 and presented the most similar real-generated image pairs with the highest PSNR or SSIM in the Figure 4. It can be seen that there are noticeable differences between them.

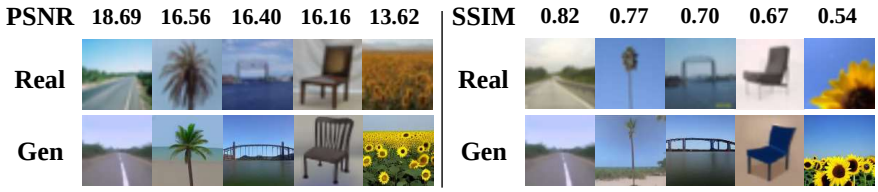

Fig. 4: The most similar real-generated image pairs.

B Time and transmission efficiency

To assess whether DDDR's performance improvement comes at the cost of training efficiency, we conduct an analysis of its training time. In learning a new task, DDDR operates in two stages: the Federated Class Inversion phase, during which a class embedding is optimized for each new category, followed by the Replay-Augmented Training phase, which involves image generation before

classifier training. Image generation allows for server-side execution without utilizing client computational resources, and the generated images can be stored for repeated use. Consequently, Federated Class Inversion and Classifier Training are the two primary factors affecting training time. As shown in Table 1, for the learning of each new task, the time consumed by Federated Class Inversion is significantly less than that required for classifier training, accounting for only about 12% of their combined total. Comparatively, the training duration for classifiers in DDDR and other baseline methods is similar, given the identical training steps among them, with the primary difference being in the loss function used, which does not significantly impact training time. Thus, the additional time incurred by our method compared to other baselines is attributed to the Federated Class Inversion phase, which is significantly shorter than the time for classifier training and does not substantially affect the overall runtime.

Table 1: Training Time Analysis of DDDR on the Cifar-100 Dataset with Five Tasks. FCI denotes the Federated Class Inversion Phase, CT represents the Classifer Training, and IG stands for Image Generation. The local training duration for one client is reported in minutes for both the FCI and CT phases. For IG, the time required to generate 200 images for a single class is reported. All experiments were conducted on a single 3090 GPU.

	FCI	CT	IG
training time (min)	4.8	34.2	3.63
communication rounds	10	100	-

Moreover, the Federated Class Inversion in DDDR is transmission-efficient, as it only transmits low-dimensional class embeddings. For instance, the transmitting parameter of FCI is at most $128K$ for the diffusion model ($1.5B$) on CIFAR-100.

C Visualization of generated results

To more comprehensively demonstrate the generative capabilities of DDDR, we conduct training for Federated Class Inversion on both the Cifar-100 [7] and Tiny-ImageNet [8] datasets. Utilizing the resultant class embeddings, we generate images, with the outcomes presented in Figure 5 and 6. From the generated results, two observations can be made: 1) DDDR is capable of producing high-quality images, closely matching the distribution of real images in both diversity and fidelity. For instance, the generated images of categories such as bowls, chairs, and tables in Figure 5 are highly realistic and exhibit a wide variety of styles and poses. 2) Despite the high quality of generation, a slight domain discrepancy between the generated and real data is observable [4,9]. For example, in Figure 5, categories such as buses and houses are more frequently

Fig. 5: Visualization of generated outcomes from DDDR and the real data from the CIFAR-100 dataset.

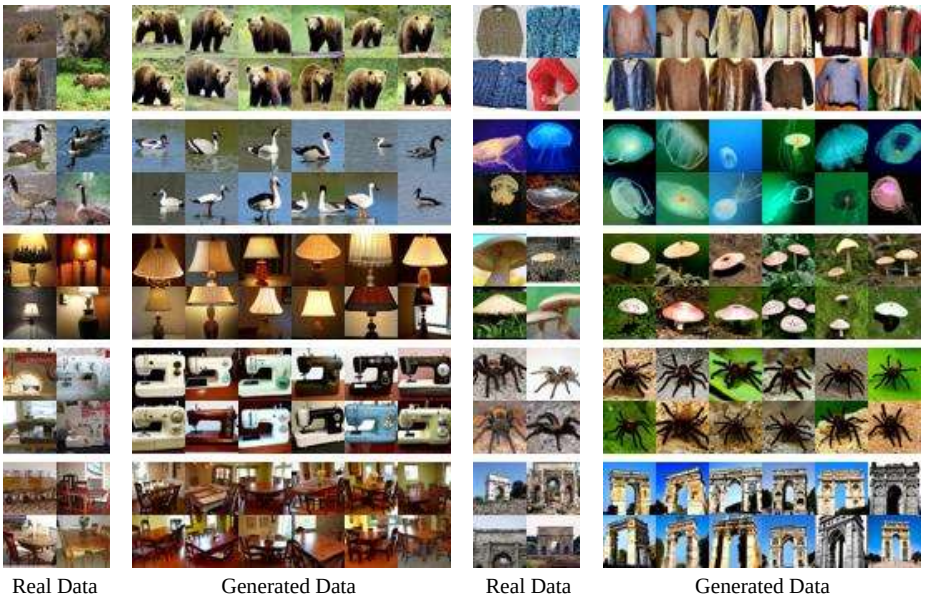

Fig. 6: Visualization of generated outcomes from DDDR and the real data from the Tiny-ImageNet dataset.

depicted in nighttime scenes in the real data, whereas the generated data tends to favor daytime scenes. This underscores the importance of enhancing the classifier's generalization capability across both the generated and real domains. The cause of this domain discrepancy may be attributed to the limited optimization parameters. In DDDR, to enhance training efficiency, the optimization was conducted solely on the class embeddings without fine-tuning the pre-trained diffusion model.

D Generalizability

We demonstrate the generalization capability of DDDR in the following two points: 1) We validated FCI's generative ability on widely used medical image datasets (LiTS [3] and MSD [1]) and fine-grained classification datasets (Stanford Dogs [5]). The results in Figure 7 show that FCI can effectively generate data even when there are significant differences from the pretraining data. 2) The CIFAR-100 and TinyImageNet datasets we used were not used for pretraining the diffusion model [10].

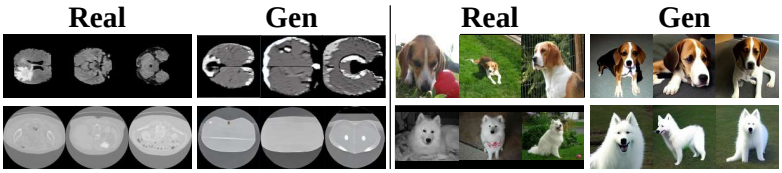

Fig. 7: The most similar real-generated image pairs.

E Local Test Result

Table 2: Results of the comparative experiments on the Cifar-100 dataset. 'T' indicates the task number. 'Acc' denotes average accuracy, with higher values indicating better performance, and 'FM' represents the forgetting measure, where lower values signify lesser forgetting of historical tasks. The best results are highlighted in bold.

Data partition	IID								non-IID							
Tasks	T=5				T=10				T=5				T=10			
Method	Acc(")		FM(#)		Acc(")		FM(#)		Acc(")		FM(#)		Acc(")		FM(#)	
Finetune	17.33	0.18	0.83	0.01	9.03	0.18	0.88	0.01	16.47	1.12	0.74	0.07	8.58	0.58	0.77	0.05
FedEWC	21.35	0.49	0.69	0.01	11.76	0.50	0.73	0.01	20.94	1.20	0.61	0.05	11.56	1.14	0.67	0.07
Target	34.40	0.97	0.48	0.01	22.95	0.55	0.49	0.01	34.37	2.30	0.48	0.04	21.68	2.27	0.53	0.04
MFCL	42.67	0.82	0.37	0.01	31.35	0.52	0.46	0.01	41.16	2.57	0.33	0.03	28.92	2.14	0.43	0.03
Ours	**51.04**	**0.83**	**0.29**	**0.01**	**43.45**	**0.76**	**0.32**	**0.01**	**48.45**	**3.56**	**0.26**	**0.04**	**41.14**	**4.57**	**0.30**	**0.04**

Table 3: Results of the comparative experiments on the Tiny-ImageNet dataset.

Data partition	IID								non-IID							
Tasks	T=5				T=10				T=5				T=10			
Method	Acc(")		FM(#)		Acc(")		FM(#)		Acc(")		FM(#)		Acc(")		FM(#)	
Finetune	12.29	0.46	0.60	0.01	6.80	0.29	0.67	0.01	11.68	0.61	0.52	0.04	6.57	0.67	0.59	0.03
FedEWC	13.27	0.45	0.49	0.01	8.22	0.30	0.56	0.01	12.55	0.70	0.43	0.03	7.67	0.90	0.50	0.03
Target	17.56	0.49	0.45	0.01	12.53	0.43	0.49	0.01	17.88	0.85	0.43	0.03	11.31	0.90	0.47	0.03
MFCL	15.11	0.47	0.52	0.01	10.13	0.48	0.54	0.01	13.31	1.18	0.45	0.03	8.57	0.45	0.49	0.02
Ours	**25.47**	**0.85**	**0.36**	**0.01**	**19.01**	**0.67**	**0.36**	**0.01**	**23.97**	**1.26**	**0.34**	**0.03**	**16.63**	**0.75**	**0.32**	**0.04**

Our results presentation in the main text follows the mainstream work in the FCCL field [2, 9, 11], calculating metrics on a global test set. However, to demonstrate performance variations across different clients, we also report the mean and standard deviation of metrics from multiple clients' independent tests. From the results in Tables 2 and 3, we can draw the same conclusion as in the main text, namely that our method significantly outperforms the others.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
2. Babakniya, S., Fabian, Z., He, C., Soltanolkotabi, M., Avestimehr, S.: A data-free approach to mitigate catastrophic forgetting in federated class incremental learning for vision tasks. In: NeurIPS. vol. 36 (2023)
3. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., et al.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023)
4. Dai, G., Zhang, Y., Wang, Q., Du, Q., Yu, Z., Liu, Z., Huang, S.: Disentangling writer and character styles for handwriting generation. In: CVPR. pp. 5977–5986 (2023)
5. Dataset, E.: Novel datasets for fine-grained image categorization. In: First Workshop on Fine Grained Visual Categorization, CVPR. Citeseer. Citeseer. Citeseer. vol. 5, p. 2. Citeseer (2011)
6. Kang, Y., Gu, H., Tang, X., He, Y., Zhang, Y., He, J., Han, Y., Fan, L., Yang, Q.: Optimizing privacy, utility and efficiency in constrained multi-objective federated learning. arXiv preprint arXiv:2305.00312 (2023)
7. Krizhevsky, A., et al.: Learning multiple layers of features from tiny images (2009)
8. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015)
9. Qi, D., Zhao, H., Li, S.: Better generative replay for continual federated learning. In: ICLR (2023)
10. Schuhmann, C., Kaczmarczyk, R., Komatsuzaki, A., Katta, A., Vencu, R., Beaumont, R., Jitsev, J., Coombes, T., Mullis, C.: Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. In: NeurIPS Workshop Datacentric AI. No. FZJ-2022-00923 (2021)
11. Zhang, J., Chen, C., Zhuang, W., Lyu, L.: Target: Federated class-continual learning via exemplar-free distillation. In: ICCV. pp. 4782–4793 (2023)
12. Zhu, L., Liu, Z., Han, S.: Deep leakage from gradients. In: NeurIPS. vol. 32 (2019)