Privacy-Preserving Federated Meta-Learning for Neural Fields

Junhyeog Yun, Minui Hong, Gunhee Kim

Vision and Learning Lab, Seoul National University

junhyeog@snu.ac.kr

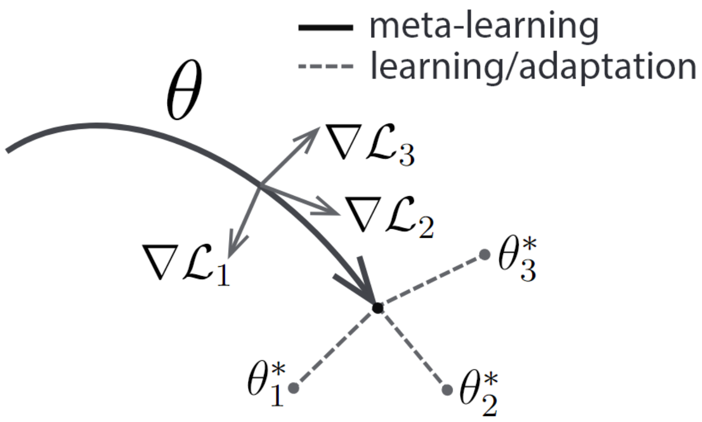
ICCV 2025 Poster

Outline

- Background
  - Federated Learning (FL)
  - Meta-Learning
  - Neural Fields (NFs)
- Motivation
  - Scenario
  - Privacy Leakage in FML for NFs
- Approach (FedMeNF)
  - Privacy Metrics
  - Privacy-Preserving Loss Function
- Experiments
- Summary

Background – Meta-Learning

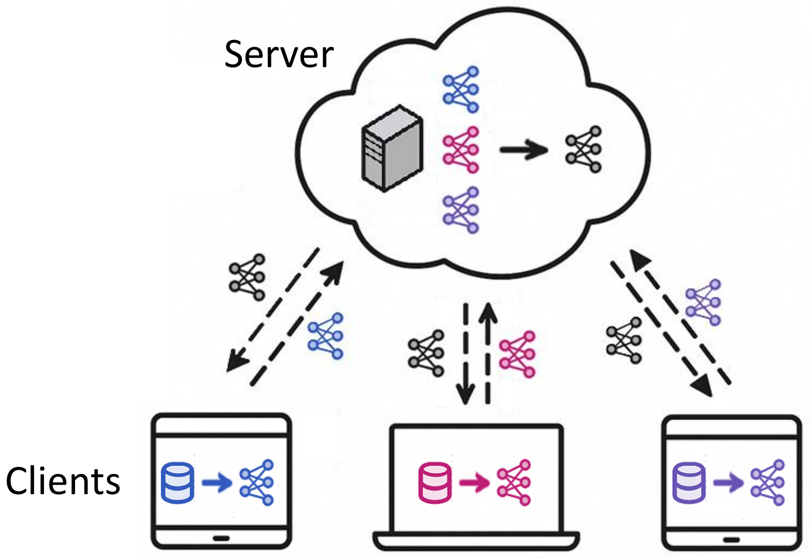
- Traditional machine learning approach
  - one separate model per task
- Meta-learning approach
  - learns learning strategy that generalizes across various tasks **(Learn to Learn)**
  - trains a meta-learner that can **adapt quickly to a new task**, even with only few data samples **(few-shot)**
- Example: quickly transfers know-how from **skateboarding → snowboarding** or **surfing**



C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in ICML, 2017

Background – Federated Learning (FL)

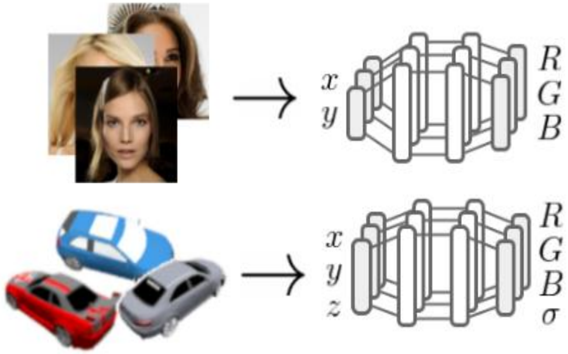
- Privacy-Preserving Collaborative Learning
- Multiple devices or institutions **train together without ever sharing raw data**
- Each client trains locally, then **only model updates (parameters/gradients) are sent to a central server**
- Server **aggregates the updates into a global model** and broadcasts it back to clients for the next round



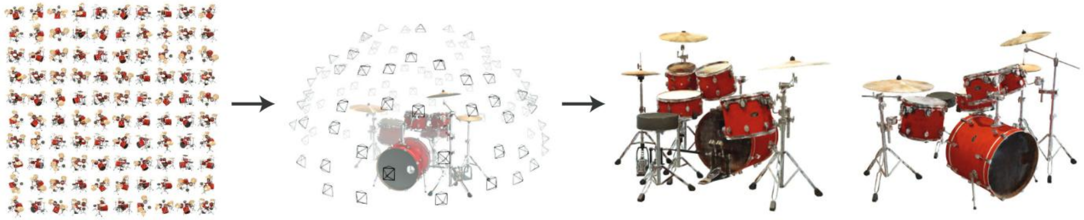
- Global model
- Local models
- Local data

Background – Neural Fields (NFs)

- **Coordinate-based Neural Fields**
  - A deep neural network to approximate **continuous signals**
  - represent continuous functions that **map spatial coordinates to signal values such as color or density**
  - Delivers infinite resolution and **high memory-efficiency** compared with traditional pixel- or point-grid representations



- **NeRF** (Neural Radiance Fields)
  - Learns a neural field from multiple 2D images of a 3D object or scene
  - We can render new views of the same object/scene from arbitrary camera poses



E. Dupont, H. Kim, S. Eslami, D. Rezende, and D. Rosenbaum, "From data to functa: Your data point is a function and you can treat it like one," arXiv:2201.12204, 2022.
B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," Communications of the ACM, 2021.

Motivation – Scenario (*Local* method)

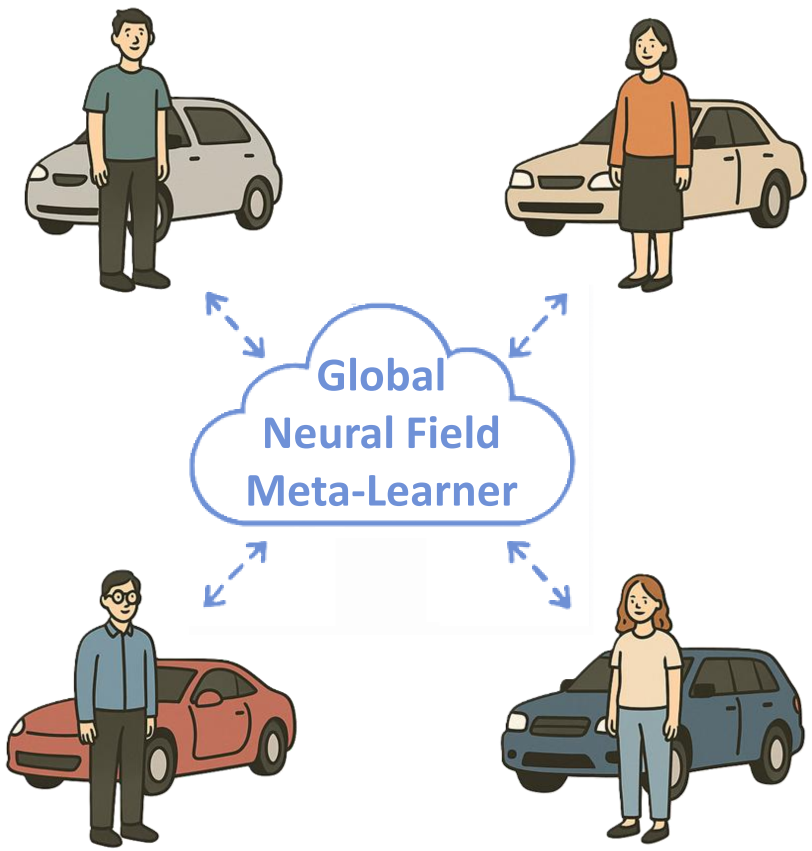
We want to train a **Neural Field Meta-Learner**

which achieves **Fast Optimization** and **Robust Reconstruction Performance**,

even **with Few-Shot**

Motivation – Scenario (FML)

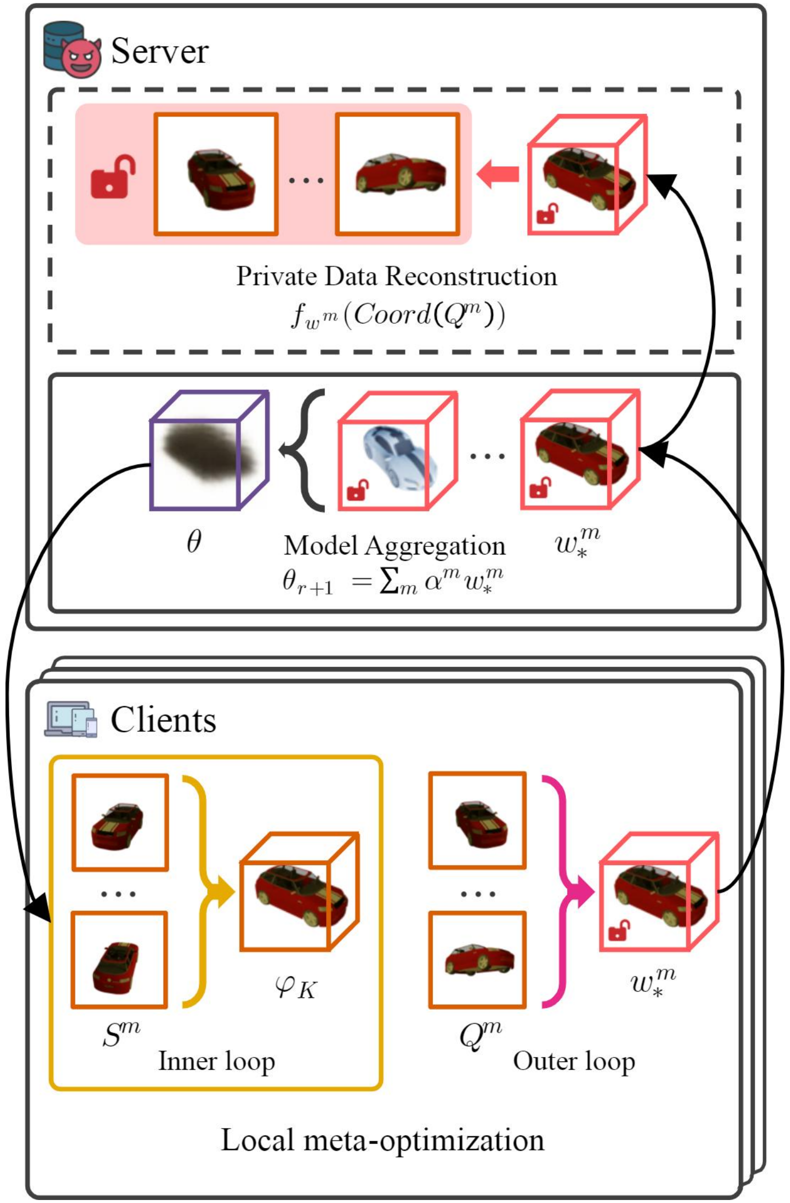
- Meta-learning requires various task data
- However, each client only has data from one car
- *⇒ Federated Meta-Learning (FML)*
  - Multiple clients collaborate
  - Train a global meta-learner
  - Without sharing raw data



Motivation – Privacy Leakage in FML for NFs

Causes

1. Each client only has a single task
   - e.g., car, face, body, …
   - the **meta-learner functions as a neural field** (meta-optimization == 2^{nd}-order optimization)
2. Neural fields inherently encapsulate the data
   - **shared meta-learner can be exploited to infer data**, which violate the client's privacy



Motivation

We propose

*a novel **Fed**erated **Me**ta-Learning approach*

*for **N**eural **F**ields*

that prevents privacy leakage,

*called **FedMeNF***

Method	*Local*	Federated Meta-Learning	**Ours**
Fast optimization	✗	✓	✓
Few-shot adaptation	✗	✓	✓
Privacy preservation	✓	✗	✓

Approach – Privacy Metric

- We need a quantitative metric for "How well did the server reconstruct the client's private data?"
- **Peak Signal-to-Noise Ratio ($PSNR$)**
  - standard image quality metric in reconstruction & novel view synthesis
  - higher $PSNR \Rightarrow$ reconstructed image is closer to ground truth
- $PSNR_p$
  - Ground-truth (GT): client's private image
  - Generated image: server-side reconstruction via shared meta-learner
  - higher $PSNR_p \Rightarrow$ server-reconstructed image is closer to client's private image

Approach – Privacy Metric

- $PSNR_p = 10 \log_{10} \frac{R}{L(w,Q^m)}$
  - $L(w, Q^m)$: MSE loss of the meta-learner on the client's local data
  - **smaller $L(w, Q^m)$ ⇒ larger $PSNR_p$ ⇒ stronger privacy leakage**
- $\Delta L_{i+1} = L(w_{i+1}, B_Q) - L(w_i, B_Q)$
  - change in MSE loss ⇒ change in $PSNR_p$
- The first-order approximation of ΔL_{i+1}

$$\Delta L_{i+1} \approx -\lambda_o \cdot \left(\nabla_{w_i} L(w_i, B_Q)\right)^2 = -\lambda_o \cdot (g_K)^2 \leq 0$$

  - Always ≤ 0 ⇒ MSE loss ↓ ⇒ $PSNR_p$ ↑ each outer step

Approach – Privacy-Preserving Loss Function

- $\Delta L_{i+1} \approx -\lambda_o \cdot \left(\nabla_{w_i} L(w_i, B_Q)\right)^2 = -\lambda_o \cdot (g_K)^2 \leq 0$
- We define a privacy-preserving loss function that constrains the magnitude of g_K

$$L_{pp}(w_i, \varphi_K, B_Q) = L(\varphi_K, B_Q) - \gamma \cdot L(w_i, B_Q)$$

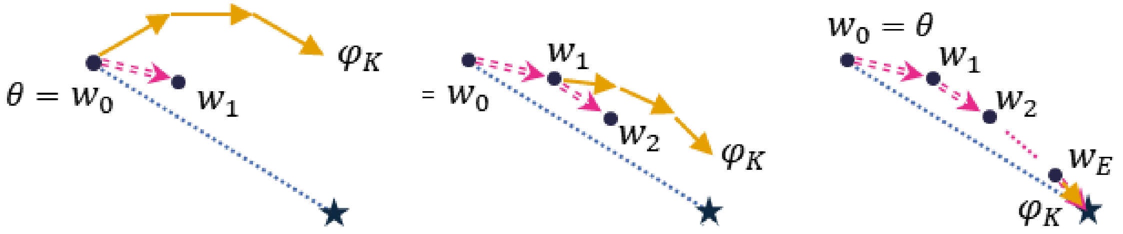
- γ is a regularization coefficient that determines the portion of g_K
- The first-order approximation of ΔL_{i+1} with L_{pp}

$$\Delta L_{i+1} \approx -\lambda_o (1 - \gamma)(g_K)^2 \leq 0$$

- Setting γ closer to 1 ⇒ ΔL_{i+1} closer to 0 ⇒ **keep the rise in $PSNR_p$**

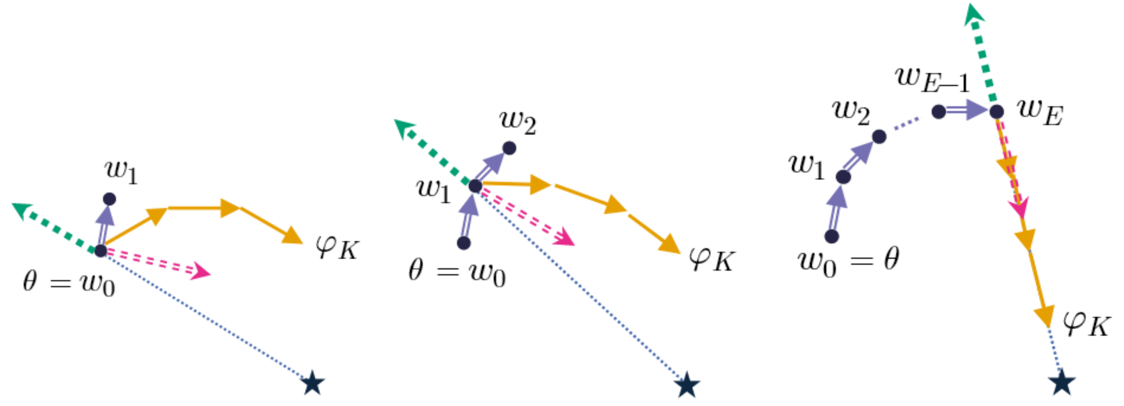
Approach – Privacy-Preserving Loss Function

- **Existing Meta-optimization:** memorizes the training data

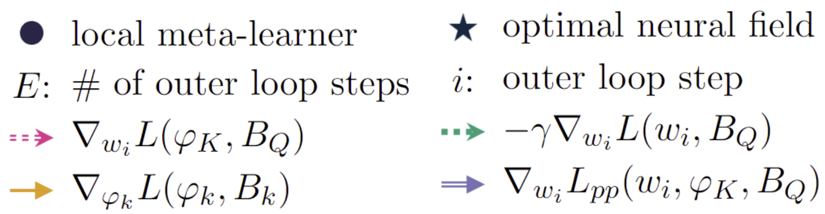


(a) $i = 0$ (b) $i = 1$ (c) $i = E - 1$

- **Privacy-Preserving Meta-optimization:** avoids memorizing the data & learns only the learning procedure



(a) $i = 0$ (b) $i = 1$ (c) $i = E - 1$



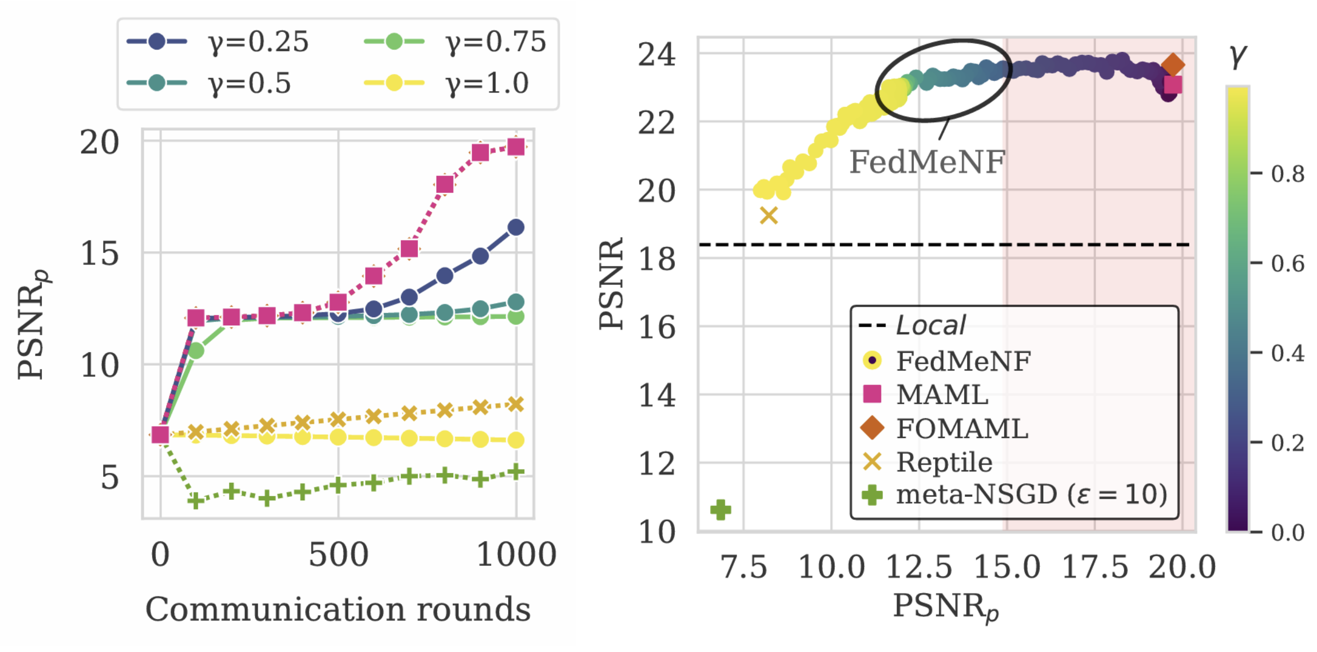
Experiments - Settings

- Baselines
  - Federated Meta-Learning = Federated Learning + Meta-Learning
    - Federated Learning: FedAvg, FedProx, Scaffold, FedNova, FedExP, and FedACG
    - Meta-Learning: MAML, FOMAML, Reptile, and meta-NSGD
- Datasets

Modality	**Dataset**	**Scenario**
3D (NeRF)	ShapeNet	3D Car
	FaceScape	3D Face
Image	PetFace	Cat image
Video	GoldDB	Golf-swing video

Experiments – Privacy-Performance Trade-off

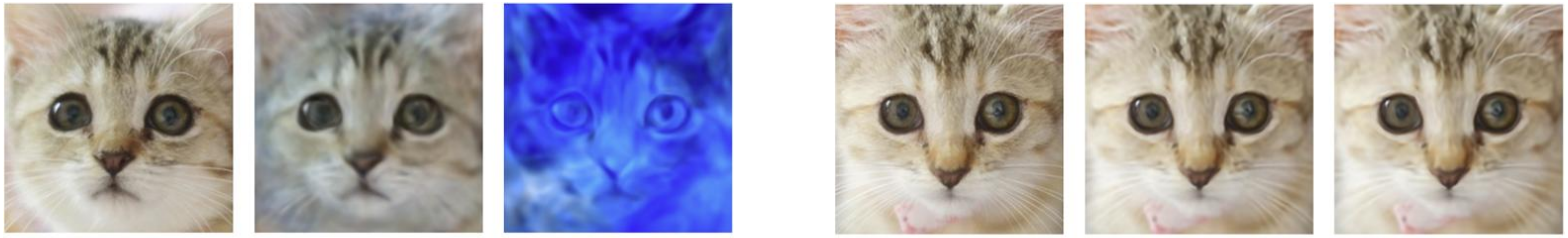
- Our FedMeNF establishes an efficient frontier that balances privacy protection and reconstruction performance



(a) PSNR$_p$ / Rounds (b) PSNR / PSNR$_p$

Experiments – Privacy-Performance Trade-off

- [Left] Reconstruction results of the client's private image on the server: (b) using MAML and (c) using FedMeNF
- [Right] Reconstruction results of a new private image on the client: (e) using MAML and (f) using FedMeNF



(a) GT ($PSNR_p$) (b) MAML (17.33) (c) FedMeNF (**5.84**) (d) GT (PSNR) (e) MAML (**35.74**) (f) FedMeNF (35.16)

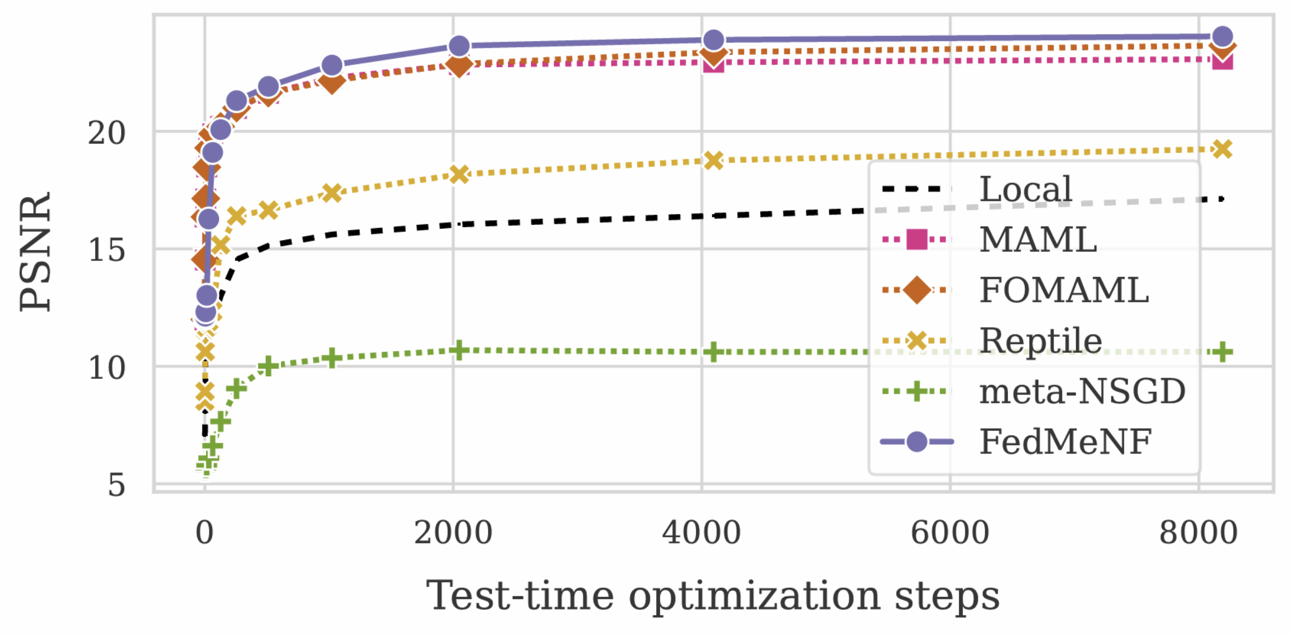
- [Upper] Reconstruction results of the client's private video on the server: (a) using MAML and (b) using FedMeNF
- [Lower] Reconstruction results of a new private video on the client: (c) using MAML and (d) using FedMeNF

(a) MAML ($PSNR_p = 23.43$) (b) FedMeNF ($PSNR_p =$ **8.38**)

(c) MAML ($PSNR =$ **29.65**) (d) FedMeNF ($PSNR = 29.15$)

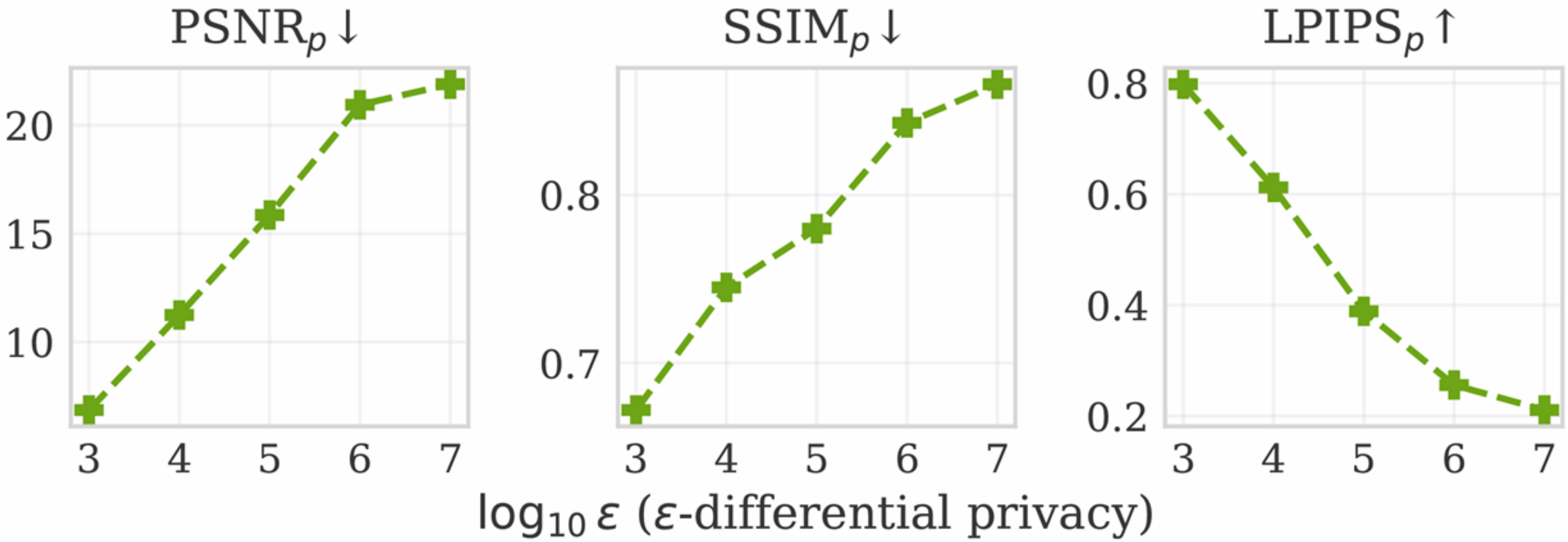
Experiments – Test-Time Optimization

- Competitive optimization speed and reconstruction quality



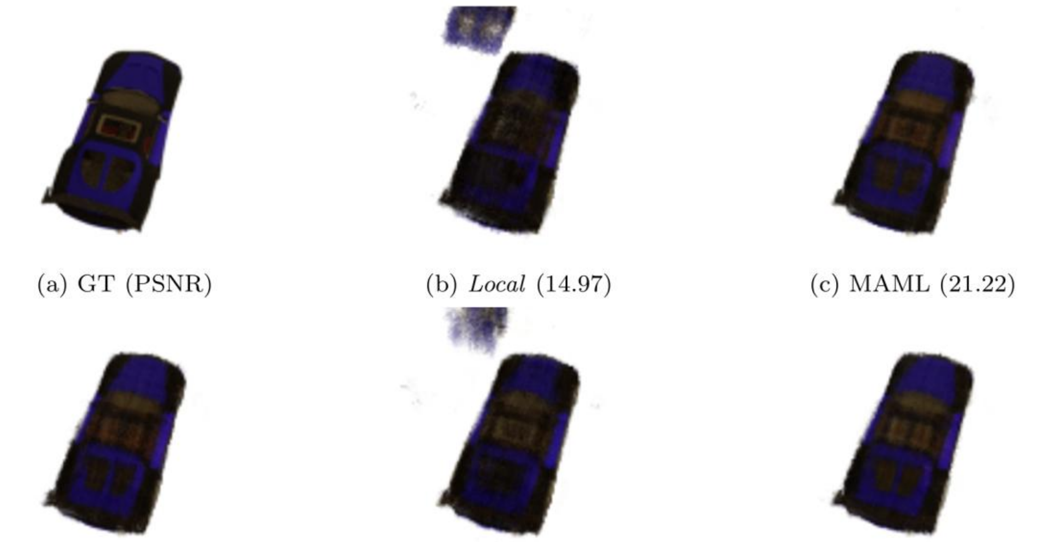
Experiments – Correlation between ϵ and Privacy Metrics

- We examine the correlation between the privacy metrics and ϵ of the differential privacy framework using meta-NSGD.
- The privacy metrics degrade as ϵ increases, supporting their generalizability as a measure of privacy leakage.



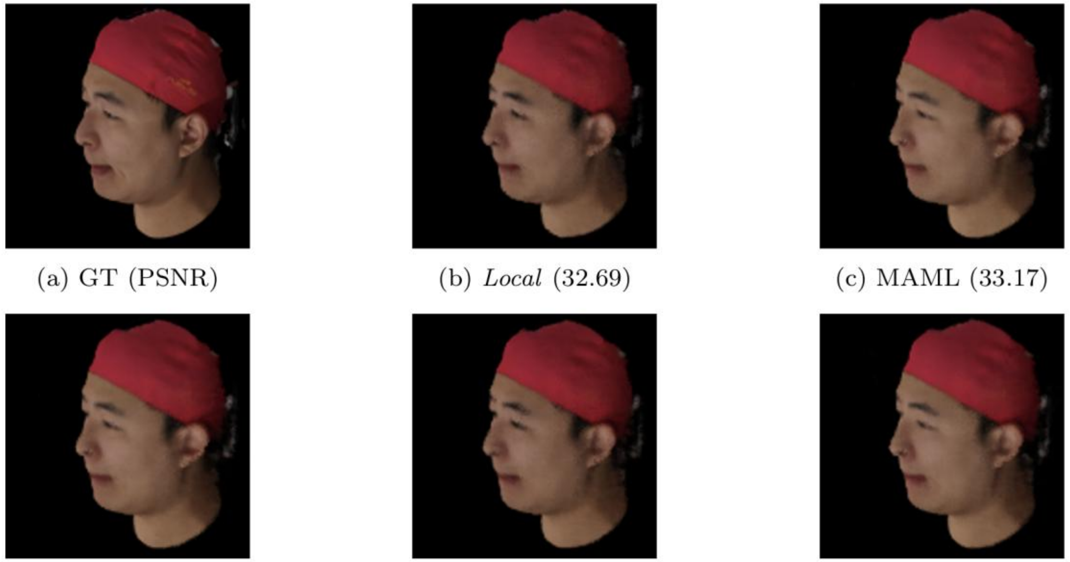
Experiments – Qualitative Results

- Competitive optimization speed and reconstruction quality



(a) GT (PSNR) (b) *Local* (14.97) (c) MAML (21.22)

(d) FOMAML (21.44) (e) Reptile (17.57) (f) FedMeNF (**21.92**)



(a) GT (PSNR) (b) *Local* (32.69) (c) MAML (33.17)

(d) FOMAML (33.26) (e) Reptile (32.72) (f) FedMeNF (**33.54**)

Summary

- The **first study** to address **federated learning for neural fields on private data**
- We **theoretically and empirically show how privacy leakage occurs** during the federated meta-learning for neural fields
- We propose FedMeNF that **preserves the privacy of local data with minimal impact on optimization speed and reconstruction quality**
- **Comprehensive experiments** on FedMeNF across various data modalities, private data sizes, and levels of data diversity, **outperforming baseline methods**