CVPR
SEATTLE, WA JUNE 17-21, 2024

上海科技大学
ShanghaiTech University

Global and Local Prompts Cooperation via Optimal Transport for Federated Learning

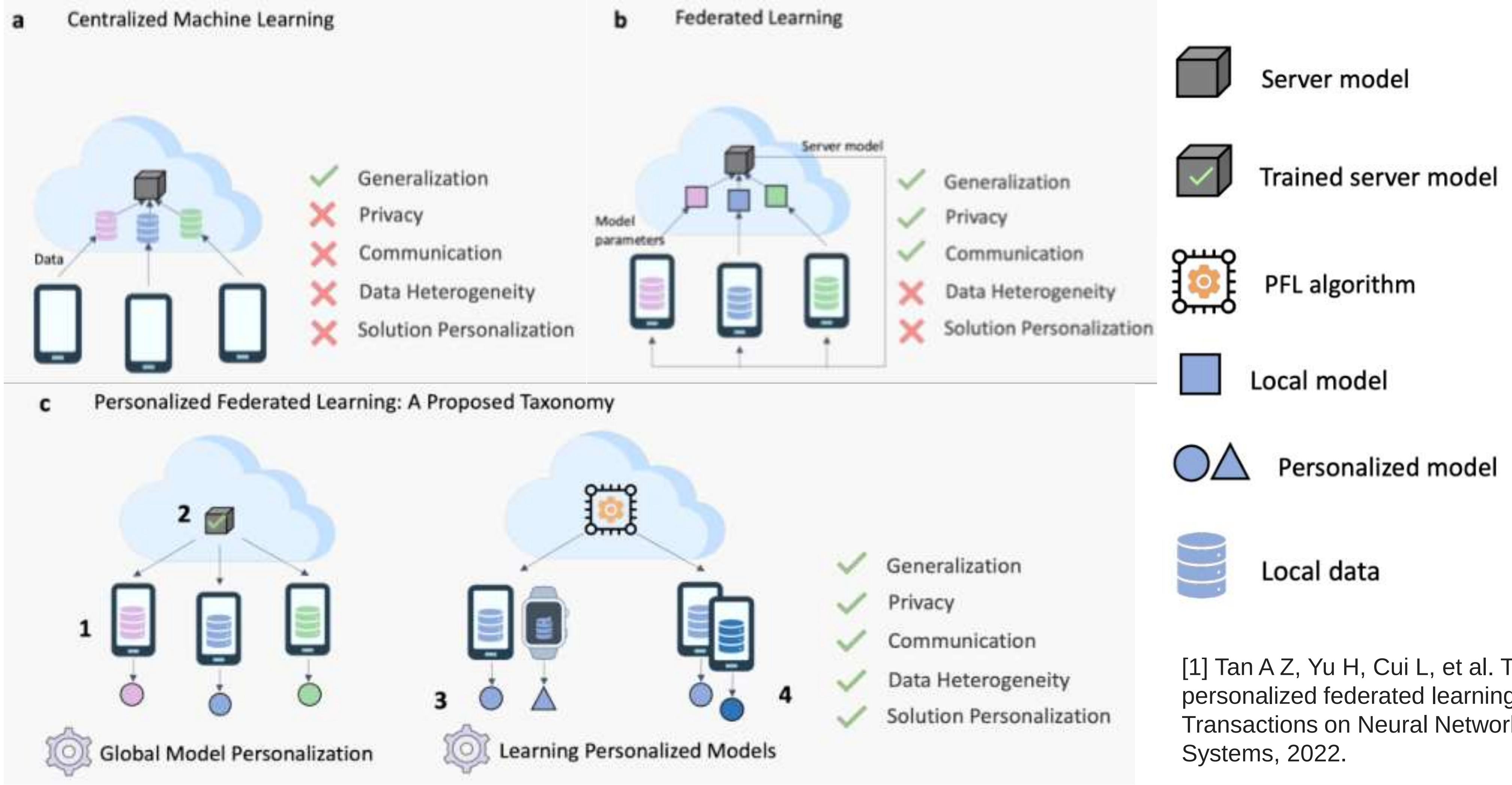
Hongxia Li[1] Wei Huang[2] Jingya Wang[1] Ye Shi[1,*]
[1]ShanghaiTech University, Shanghai, China
[2]RIKEN Center for Advanced Intelligence Project, Japan
{lihx2,wangjingya,shiye}@shanghaitech.edu.cn, wei.huang.vr@riken.jp
https://github.com/HongxiaLee/FedOTP

Paper: https://arxiv.org/pdf/2403.00041.pdf

Code: https://github.com/HongxiaLee/FedOTP

Federated Learning:

Federated learning is an emerging learning paradigm where multiple clients collaboratively train a machine learning model in a privacy-preserving manner. **Personalized federated learning** extends this paradigm to **overcome heterogeneity across clients** by learning personalized models.



[1] Tan A Z, Yu H, Cui L, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022.

The main issues of current personalized federated learning:

- Due to the large number of parameters in models, training local models from scratch will **consume a significant amount of time**.
- There will be **substantial communication costs** incurred between clients and servers when sharing and updating model parameters.
- **Overfitting** may occur when training large-scale models with limited client data.

➡ These limitations often restrict the use of complex model architectures, leading to reduced feature capacity.

We need pre-trained models and Prompt Learning → **Federated Prompt Learning** 😃

Smaller trained parameters ✓ **Fast training speed ✓** **Lower communication costs ✓**

Prompt Learning

$$q(y=k|\boldsymbol{x}) = \frac{\exp(sim(g(x), h(P_k))/\tau)}{\sum_{c=1}^{K} \exp(sim(g(x), h(P_c))/\tau)}$$

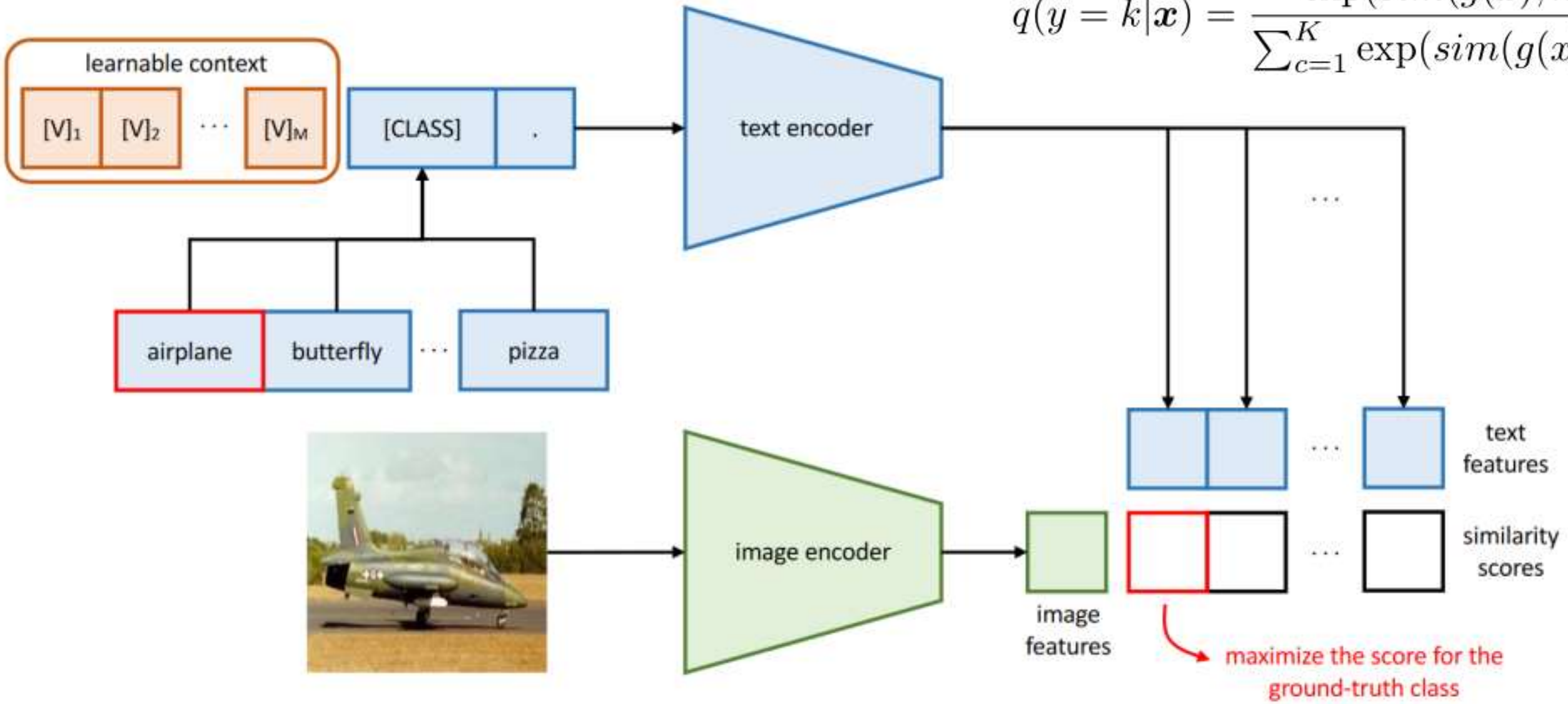
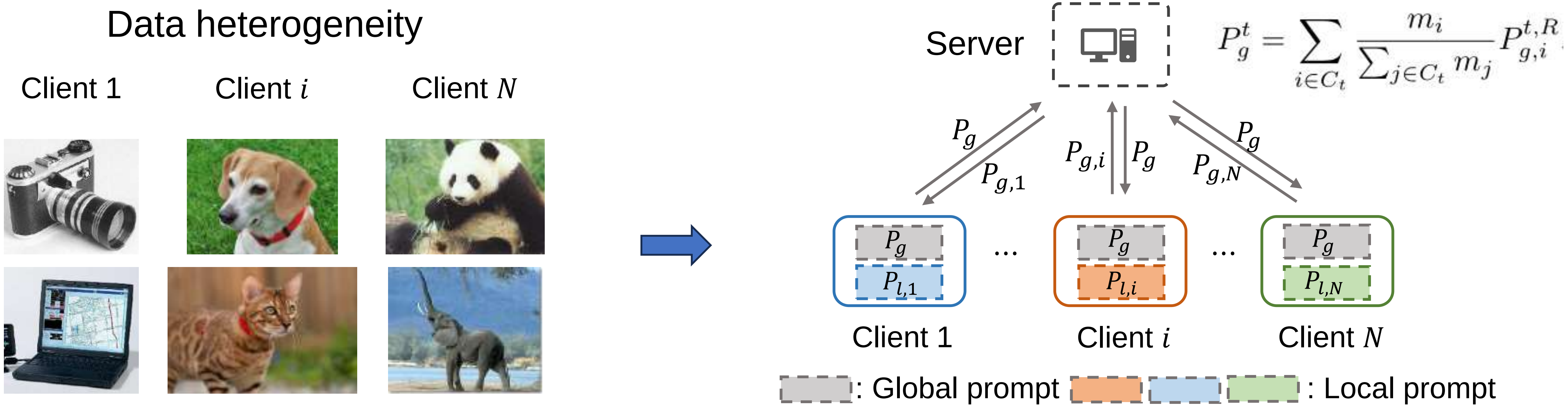


Fig. 2 Overview of Context Optimization (CoOp). The main idea is to model a prompt's context using a set of learnable vectors, which can be optimized through minimizing the classification loss. Two designs are proposed: one is unified context, which shares the same context vectors with all classes; and the other is class-specific context, which learns for each class a specific set of context vectors.

[1] Zhou K, Yang J, Loy C C, et al. Learning to prompt for vision-language models[J]. International Journal of Computer Vision, 2022, 130(9): 2337-2348.

Global Prompt and Local Prompt



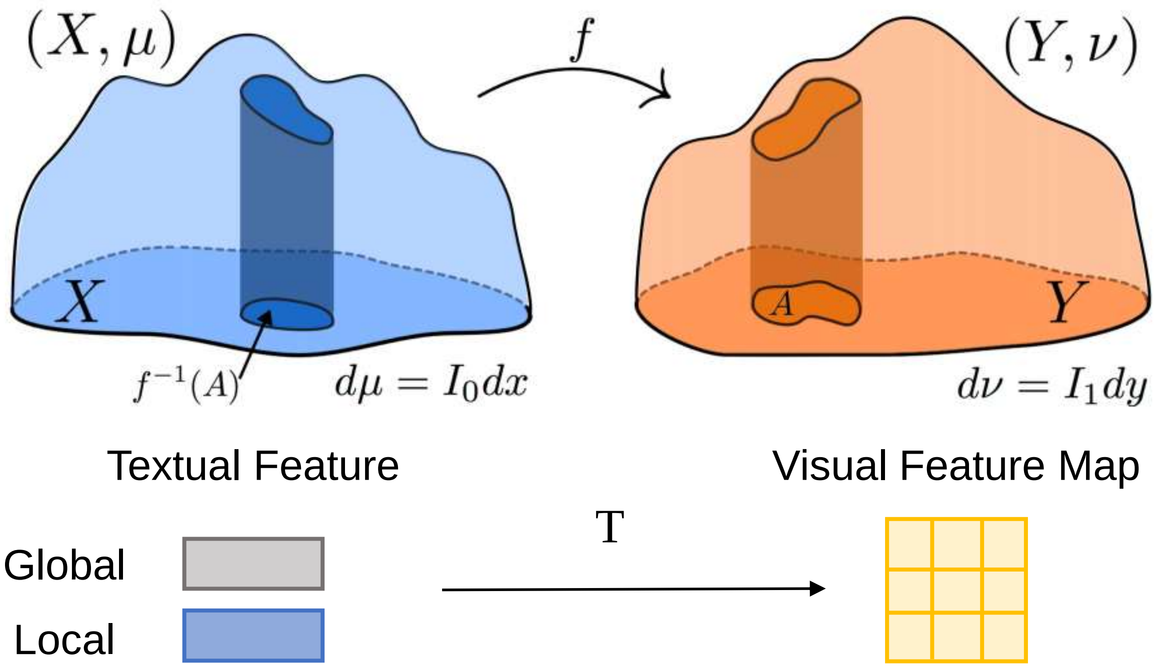
Directly calculate the similarity between two text features and image feature:

× similar to computing the mean of text features and image feature

× **make the two prompts close to the same point**, leading to learn similar features

Optimal Transport (OT)

OT is a promising optimization problem to seek an efficient solution for transporting one distribution to another.



Employ Optimal Transport (OT) to align two text features and image feature:

✓ **compel two prompts to learn distinct information** due to the constraints of OT

✓ **more fine-grained cross-modal matching**

Unbalanced Optimal Transport

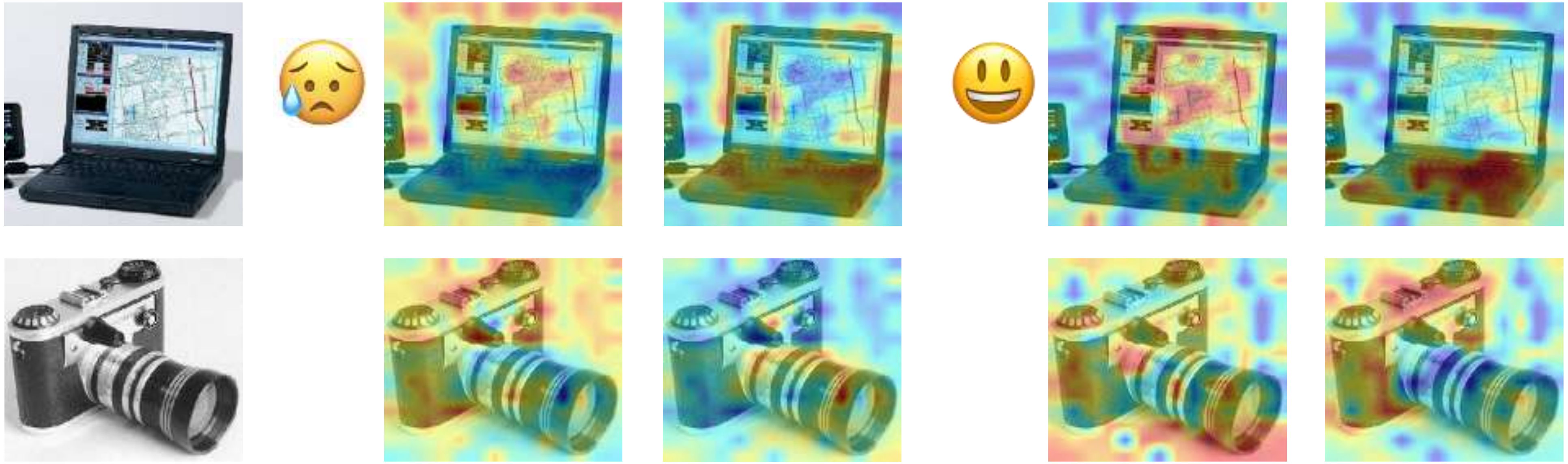
Original Image

Traditional Optimal Transport: Global, Local

Unbalanced Optimal Transport: Global, Local

Laptop

Camera



Traditional Optimal Transport:

$$d_C(\alpha,\beta)=\min_{T\in U(\alpha,\beta)}\langle C,T\rangle,$$

$$U(\alpha,\beta)=\left\{T\in\mathbb{R}_+^{|\alpha|\times|\beta|} \mid T\mathbb{1}_{|\beta|}=\alpha, T^\top\mathbb{1}_{|\alpha|}=\beta\right\}$$

Unbalanced Optimal Transport:

$$d_{C,k}(\alpha,\beta)=\min_{T\in U(\alpha,\beta)}\langle C,T\rangle,$$

$$U(\alpha,\beta)=\{T\in\mathbb{R}_+^{V\times 2} \mid \boxed{T\mathbb{1}_2\le\alpha}, T^\top\mathbb{1}_V=\beta\}$$

$$\|\alpha\|_1\ge\|\beta\|_1=\gamma\ (\gamma\in[0,1])$$

Federated Prompts Cooperation via Optimal Transport (FedOTP)

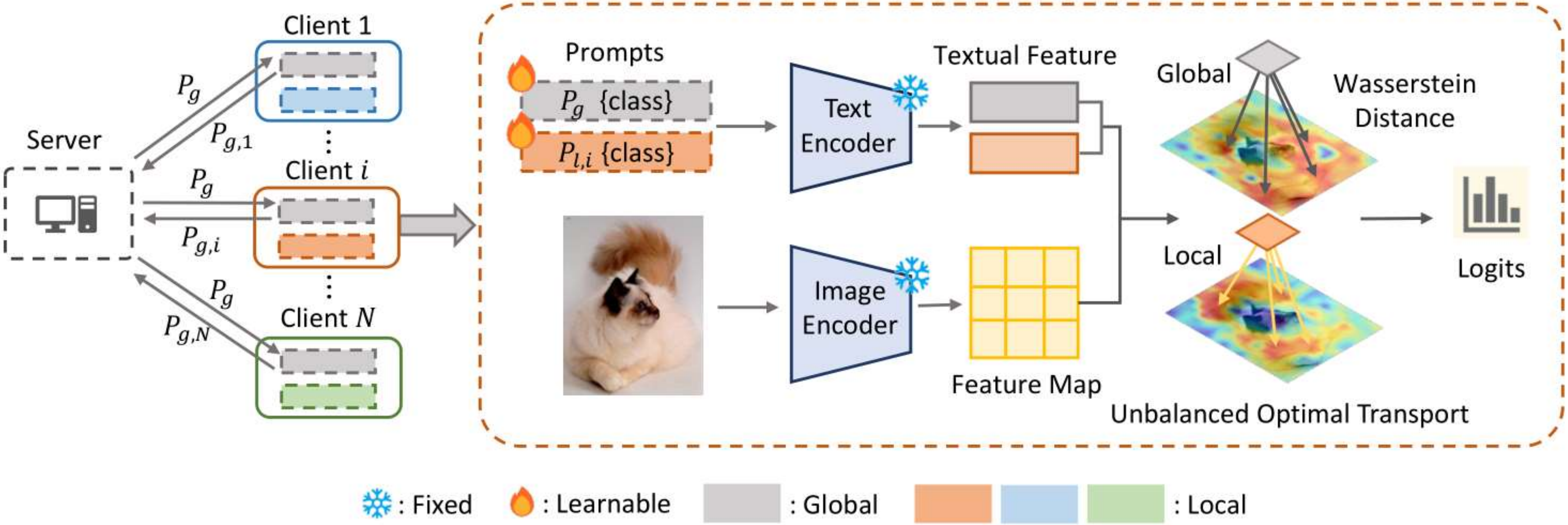


Figure 1. Overview of our FedOTP. On the left, clients transmit global prompts to the server for aggregation while retaining local prompts locally. The right shows the workflow of Global-Local prompt cooperation mechanism, which employs unbalanced Optimal Transport to align visual feature maps with each prompt.

Federated Prompts Cooperation via Optimal Transport (FedOTP)

Matching scores:

Cosine similarity

$$q(y=k|\boldsymbol{x}) = \frac{\exp(sim(g(x), h(P_k))/\tau)}{\sum_{c=1}^{K} \exp(sim(g(x), h(P_c))/\tau)}$$

Optimal transport plan

$$q(y=k|\boldsymbol{x}) = \frac{\exp((1-d_{C,k})/\tau)}{\sum_{c=1}^{K} \exp((1-d_{C,c})/\tau)}$$

Formulation:

Traditional Optimal transport

$$d_C(\alpha, \beta) = \min_{T \in U(\alpha,\beta)} \langle C, T \rangle \,,$$

$$U(\alpha, \beta) = \left\{ T \in \mathbb{R}_+^{|\alpha| \times |\beta|} \mid T\mathbb{1}_{|\beta|} = \alpha, T^\top \mathbb{1}_{|\alpha|} = \beta \right\}$$

Unbalanced Optimal transport

$$d_{C,k}(\alpha, \beta) = \min_{T \in U(\alpha,\beta)} \langle C, T \rangle \,,$$

$$U(\alpha, \beta) = \{ T \in \mathbb{R}_+^{V \times 2} \mid T\mathbb{1}_2 \leq \alpha, T^\top \mathbb{1}_V = \beta \}$$

$$\|\alpha\|_1 \geq \|\beta\|_1 = \gamma \ (\gamma \in [0,1])$$

for fast optimization

$$d_{C,k}(\alpha, \beta) = \min_{T \in U(\alpha,\beta)} \langle C, T \rangle + \underbrace{\lambda \langle T, \log T \rangle}_{\text{entropic regularization term}}$$

Comparison with state-of-the-arts

Table 1. The results of our FedOTP and the benchmark methods on the Pathological Non-IID setting with non-overlapping over 10 clients.

Methods	Food101	DTD	Caltech101	Flowers102	OxfordPets
Local Training					
Zero-Shot CLIP [55]	75.27±0.05	40.21±0.12	85.14±0.24	62.17±0.12	84.47±0.10
CoOp [73]	82.54±2.42	82.69±0.63	90.41±0.44	88.23±0.76	94.52±1.30
Prompt-based Federated Learning					
PromptFL [26]	74.81±0.64	50.46±0.54	87.90±0.54	73.68±1.58	88.17±1.18
PromptFL+FT [23]	77.16±1.56	53.74±1.36	89.70±0.25	72.31±0.91	91.23±0.50
PromptFL+FedProx [38]	73.96±0.75	50.89±0.71	87.80±1.10	74.14±0.65	87.25±1.48
PromptFL+FedPer [1]	71.29±1.87	50.23±0.82	86.72±1.45	72.11±1.35	89.50±1.62
PromptFL+FedAMP [30]	74.48±1.71	47.16±0.92	87.31±1.60	69.10±0.13	80.21±0.44
pFedPrompt [25]	92.26±1.34	77.14±0.09	96.54±1.31	86.46±0.15	91.84±0.41
FedOTP (Ours)	**92.73±0.15**	**87.67±0.70**	**97.02±0.36**	**96.23±0.44**	**98.82±0.11**

+7%!

Comparison with state-of-the-arts

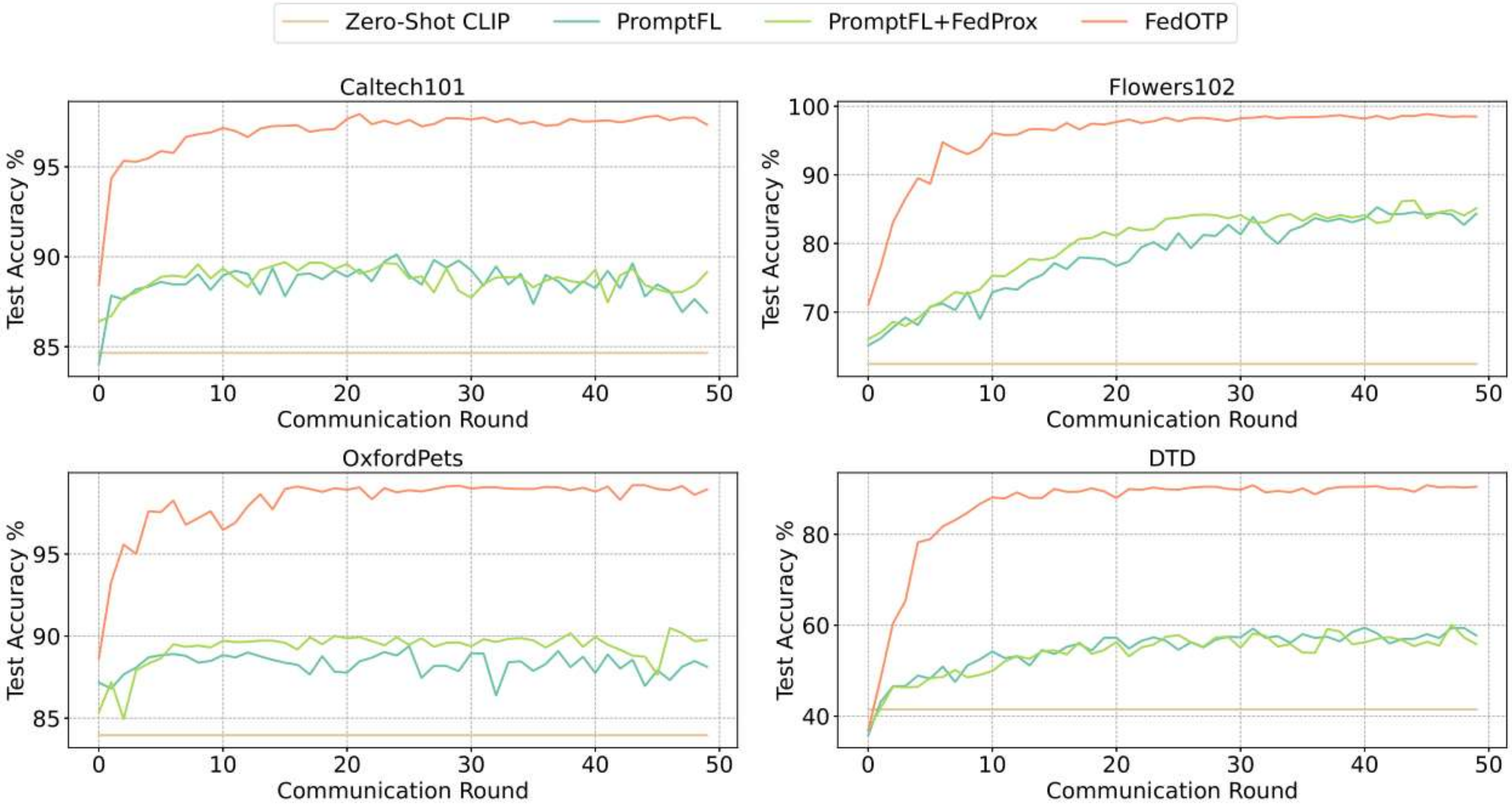


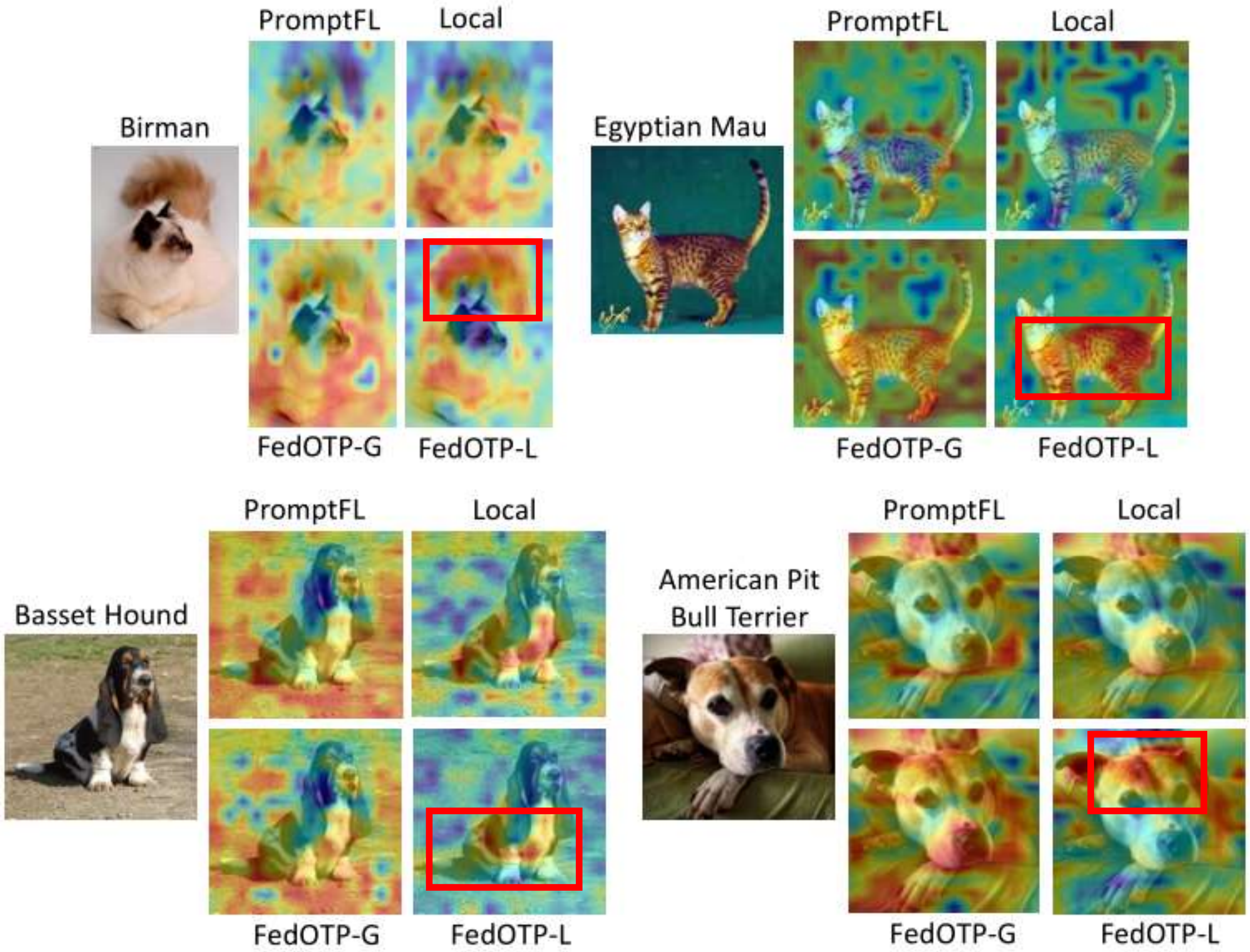
Figure A3. Accuracy curves and convergence behavior of FedOTP and other baselines on four datasets over 10 clients.

Ablation Study

Table A5. Quantitative comparisons on CIFAR-100 dataset with different α of the Dirichlet setting.

Dataset	CIFAR-100					
$\#\alpha$	0.1	0.3	0.5	1	5	10
Local Training						
Zero-Shot CLIP [55]	65.22±0.32	64.92±0.53	65.78±0.41	63.93±0.16	64.01±0.27	65.07±0.35
CoOp [73]	62.01±0.29	74.83±0.45	51.72±0.42	47.03±0.37	41.03±0.23	41.37±0.19
Prompt-based Federated Learning						
PromptFL [26]	72.45±0.64	73.67±0.56	74.37±0.18	73.95±0.14	74.68±0.05	74.43±0.08
PromptFL+FedProx [38]	72.57±0.54	71.11±0.91	74.45±0.19	74.19±0.06	74.23±0.09	74.53±0.07
FedOTP (Similarity Averaging)	78.68±0.17	75.70±0.27	75.28±0.12	74.88±0.16	74.48±0.05	74.31±0.39
FedOTP (Classical OT)	79.93±0.19	77.86±0.09	75.76±0.12	75.38±0.08	75.01±0.05	74.73±0.05
FedOTP (Unbalanced OT)	**80.56±0.12**	**78.03±0.08**	**76.75±0.10**	**76.17±0.13**	**75.75±0.03**	**75.52±0.06**

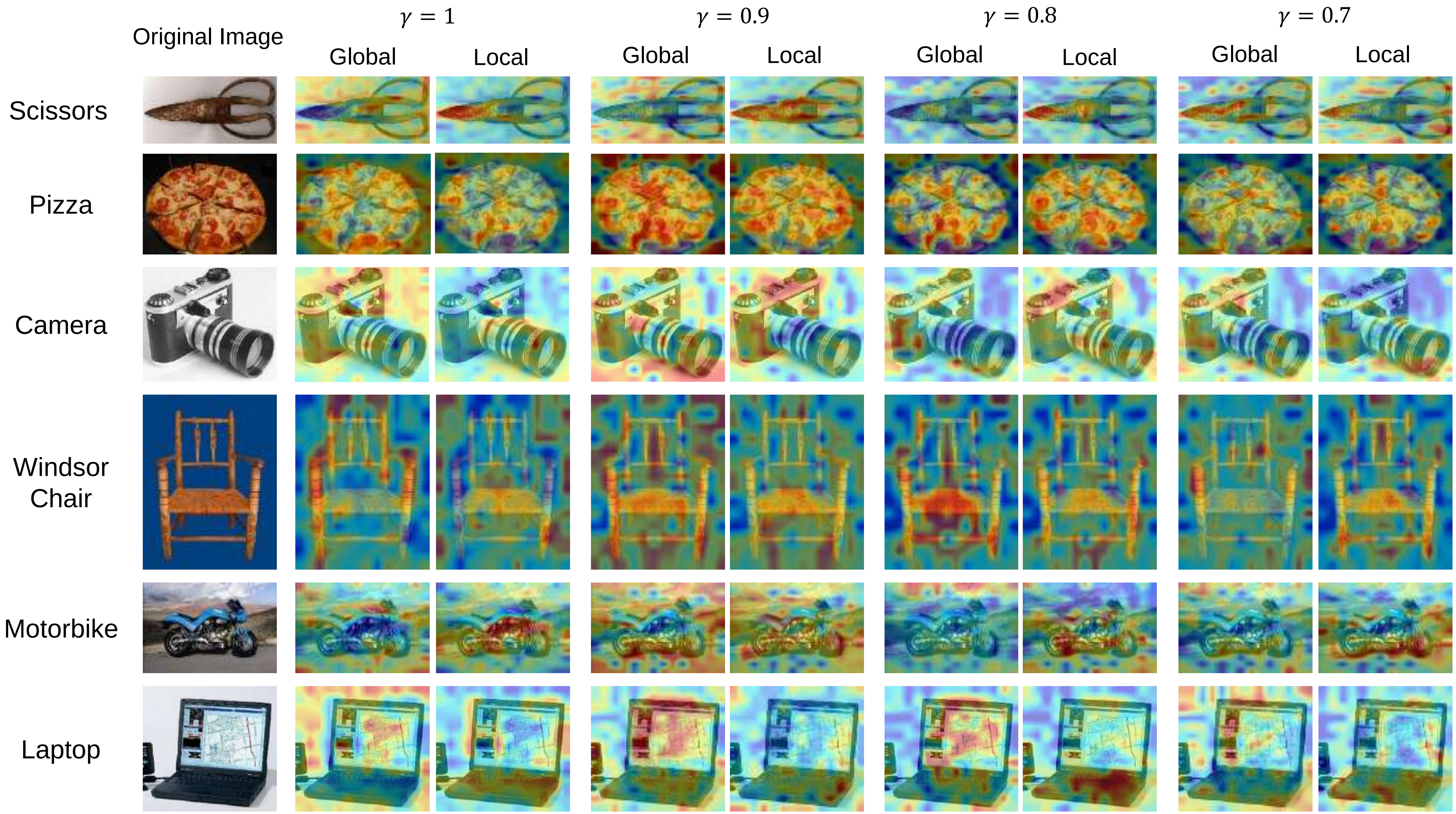
Heatmaps of similarity between text features and image feature maps



FedOTP-G: Global Prompt

FedOTP-L: Local Prompt

Heatmaps of transport plans related to global and local prompts



Conclusion

- We have proposed to use optimal transport to promote cooperation between global and local prompts for federated learning, namely FedOTP.

✓ Fast convergence and lower communication cost
✓ Capture consensus across clients and client specific traits at the same time
✓ Focus on the main object of the image
✓ More robust due to optimal transport

Scan QR code for more details