CLIP-Guided Federated Learning on Heterogeneous and Long-Tailed Data

Jiangming Shi[1], **Shanshan Zheng**[2], **Xiangbo Yin**[2], **Yang Lu**[2], **Yuan Xie**[3, 4*], **Yanyun Qu**[1, 2*]

[1] Institute of Artificial Intelligence, Xiamen University
[2] School of Informatics, Xiamen University
[3] East China Normal University
[4] Chongqing Institute of East China Normal University
jiangming.shi@outlook.com, shanshanzheng@stu.xmu.edu.cn, S_yinxb@163.com, luyang@xmu.edu.cn, yxie@cs.ecnu.edu.cn, yyqu@xmu.edu.cn

AAAI 2024

Motivation

- CReFF (IJCAI 2022)
- Core idea : Implement the **decoupling strategy** to generate the **class-distribution balanced federated features** for the server model and to **retrain the classifer** by the federated features.

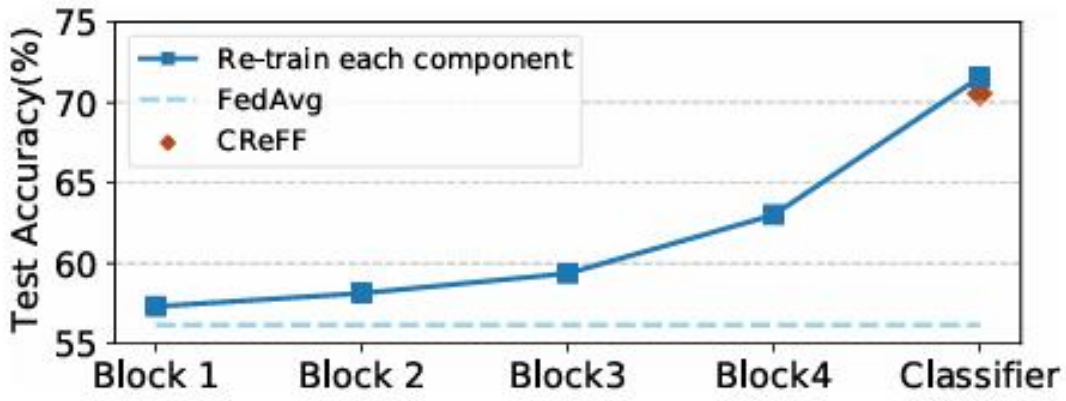


Figure 2: The performance of re-training each component of a FedAvg model pre-trained on CIFAR-10-LT with IF = 100 and $\alpha = 0.5$.

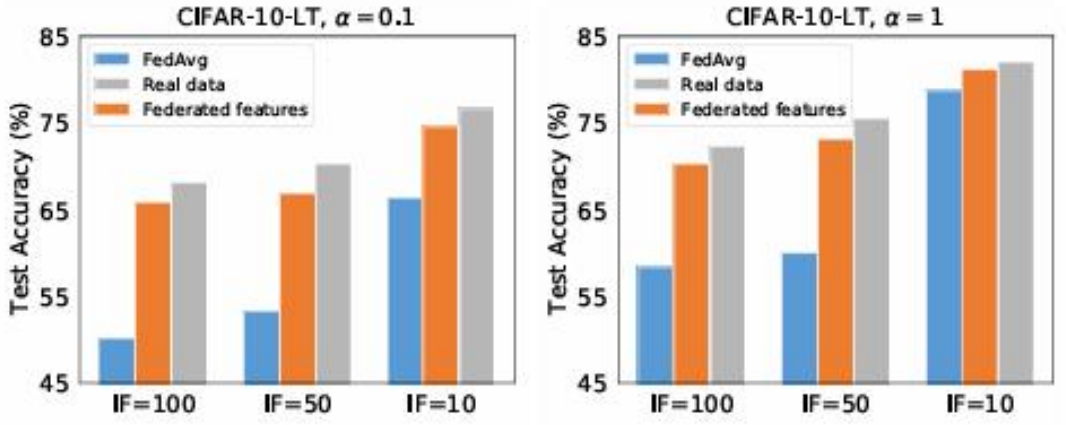


Figure 1: The performance of re-training the classifier of a FedAvg model using the same amount of data but from different sources. Real data is directly collected from clients, and federated features are synthesized by the proposed CReFF.

Motivation

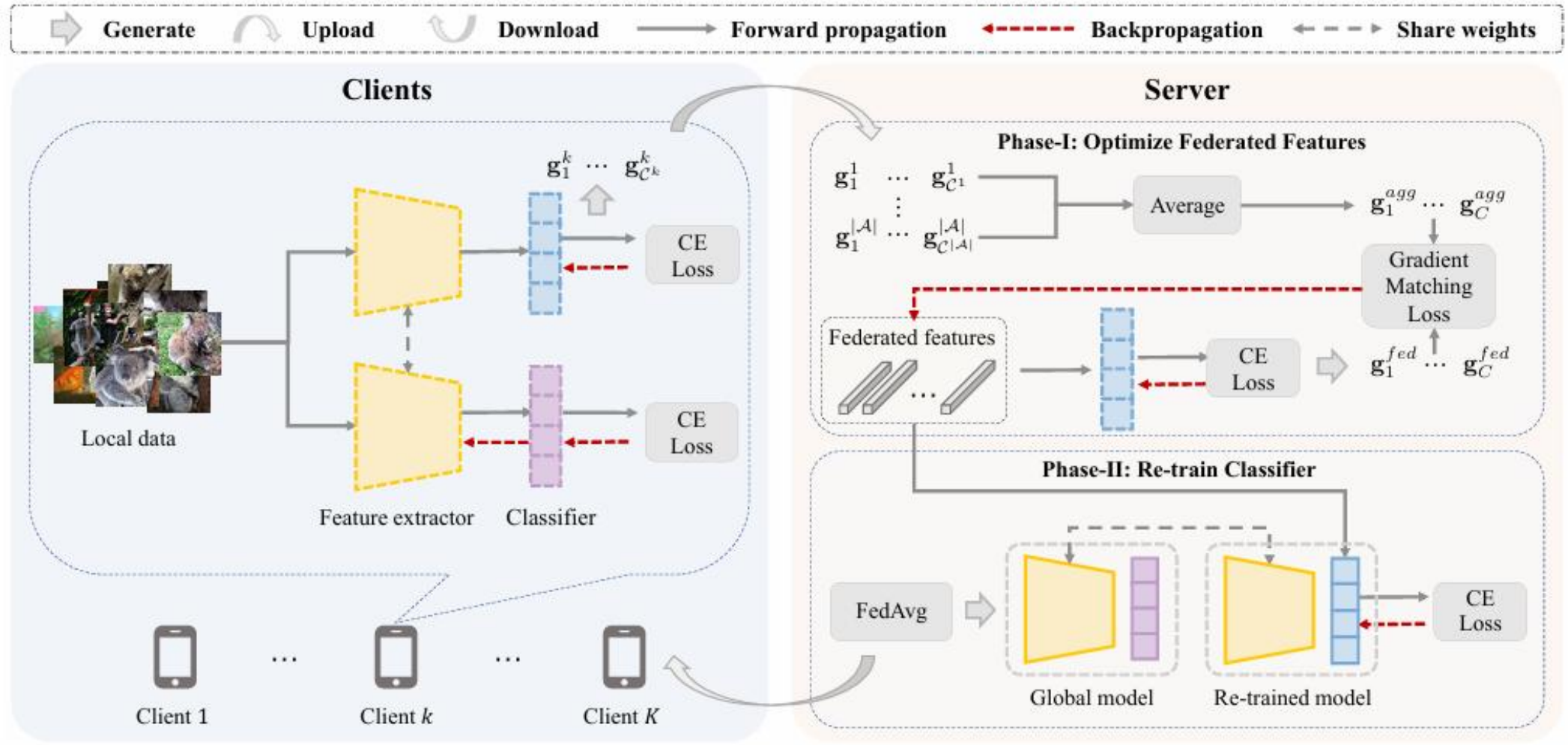


Figure 3: The framework of CReFF. In each round, clients send updated local models and real feature gradients to the server, and the server sends the aggregated global model and the re-trained model to clients.

Motivation

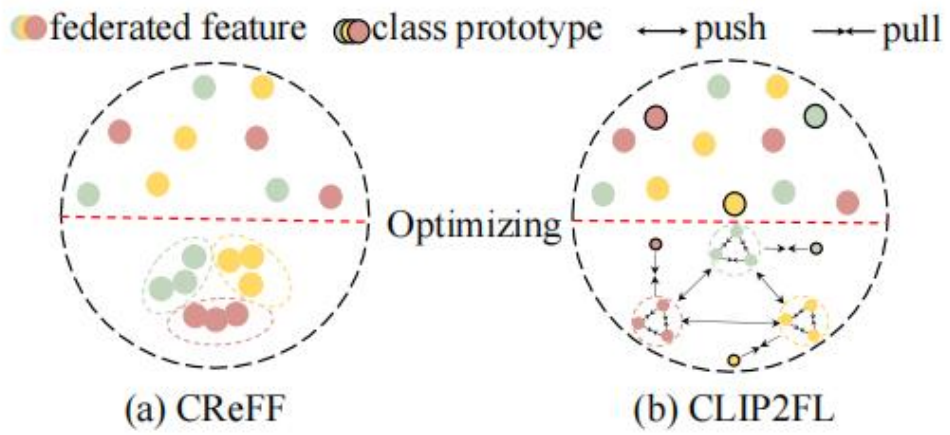


Figure 1: The figure illustrates the differences between CReFF and CLIP2FL: CReFF uses only gradients for supervision, while CLIP2FL introduces the textual features of the CLIP as prototypes for federated feature learning.

- In CReFF, the generation of federated features via client-side gradient information brings two limitations:

1) It is a one-to-many mapping between gradients and samples, which results in the problem becoming ill-posed;

2) It lacks semantic supervision, which could result in the federated features lacking discriminative ability for their respective classes.

Motivation

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- How to use CLIP to solve the FL on heterogeneous and long-tailed data ?

1) How to use CLIP to improve the feature representation of client models ?

- knowledge distillation : CLIP is used as “Teacher” while the client model is treated as “Student”, and knowledge is transferred from Teacher to Student for improving the ability of feature representation.

2) How to use CLIP to mitigate the infuence of heterogeneity and class-distribution imbalance on the server model ?

- use CLIP to constrain the generation of federated features from the client-side gradients under its effective semantic supervision
- retraining strategy : generate the federated features and use them to retrain the balance server model

Method——CLIP2FL

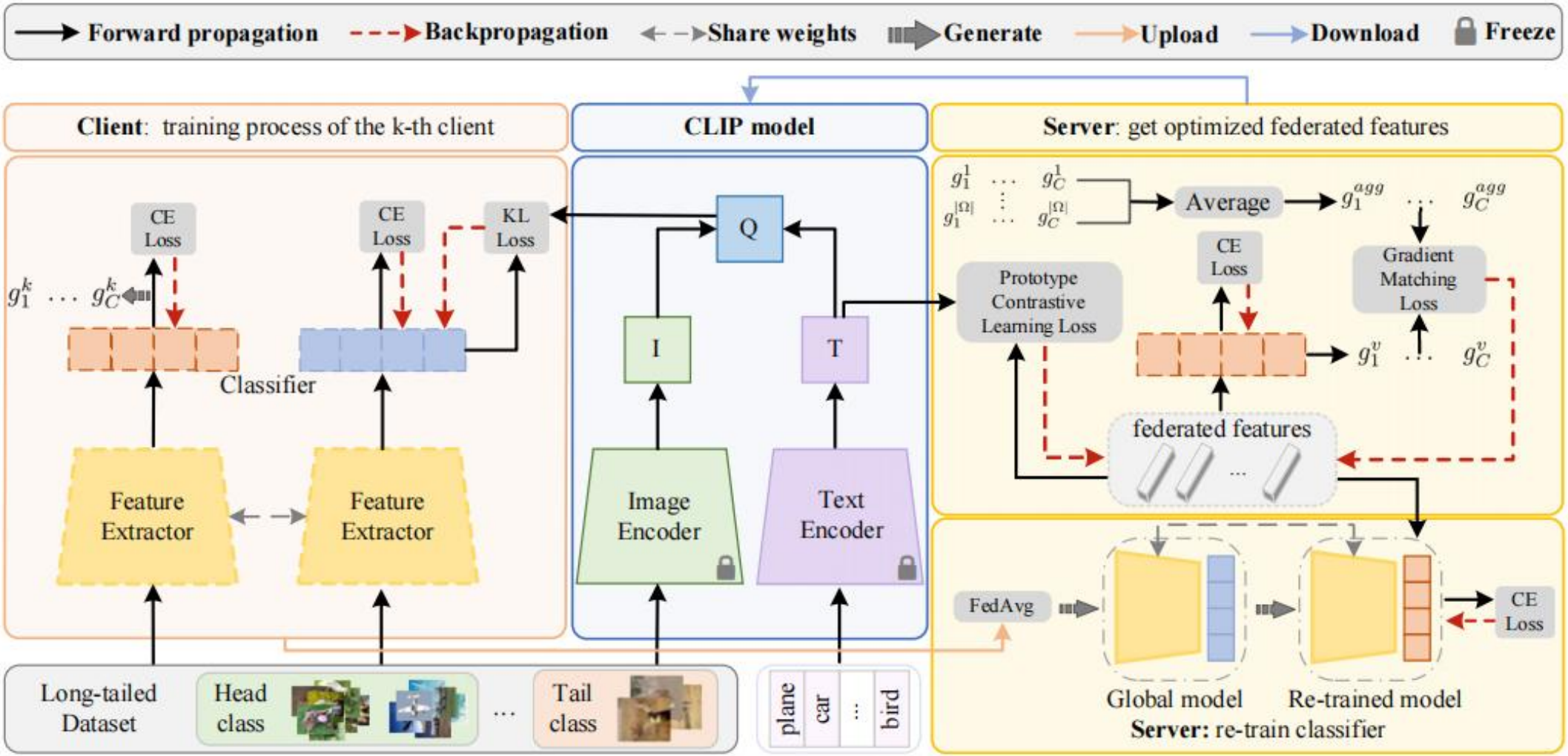


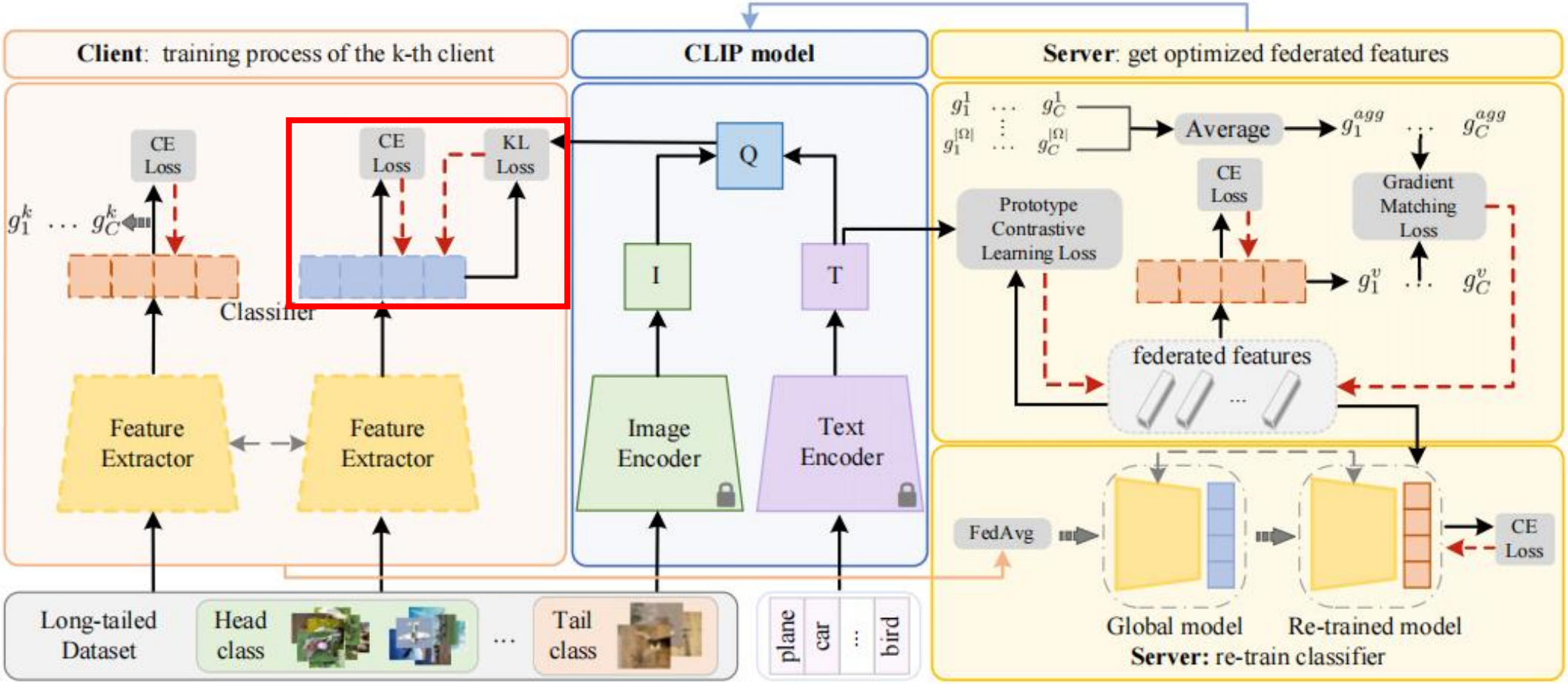
Figure 2: The framework of CLIP2FL. It includes two core components: local training on clients and classifier re-training on the server. A prior knowledge-rich CLIP model acts as a bridge to connect the two components and helps the two core components to learn better.

Method——CLIP2FL

$$L_{loc} = L_{ce}(y, p_k^t) + \beta \cdot KL(q_k^t || p_k^t), \quad (1)$$

- **Client-side Learning**

$$g_c^k = \frac{1}{n_c^k} \sum_{i=1}^{n_c^k} \nabla_{\hat{\varphi}^t} L_{ce}\left(z_{c,i}^k, y_i\right). \quad (3)$$



Method——CLIP2FL

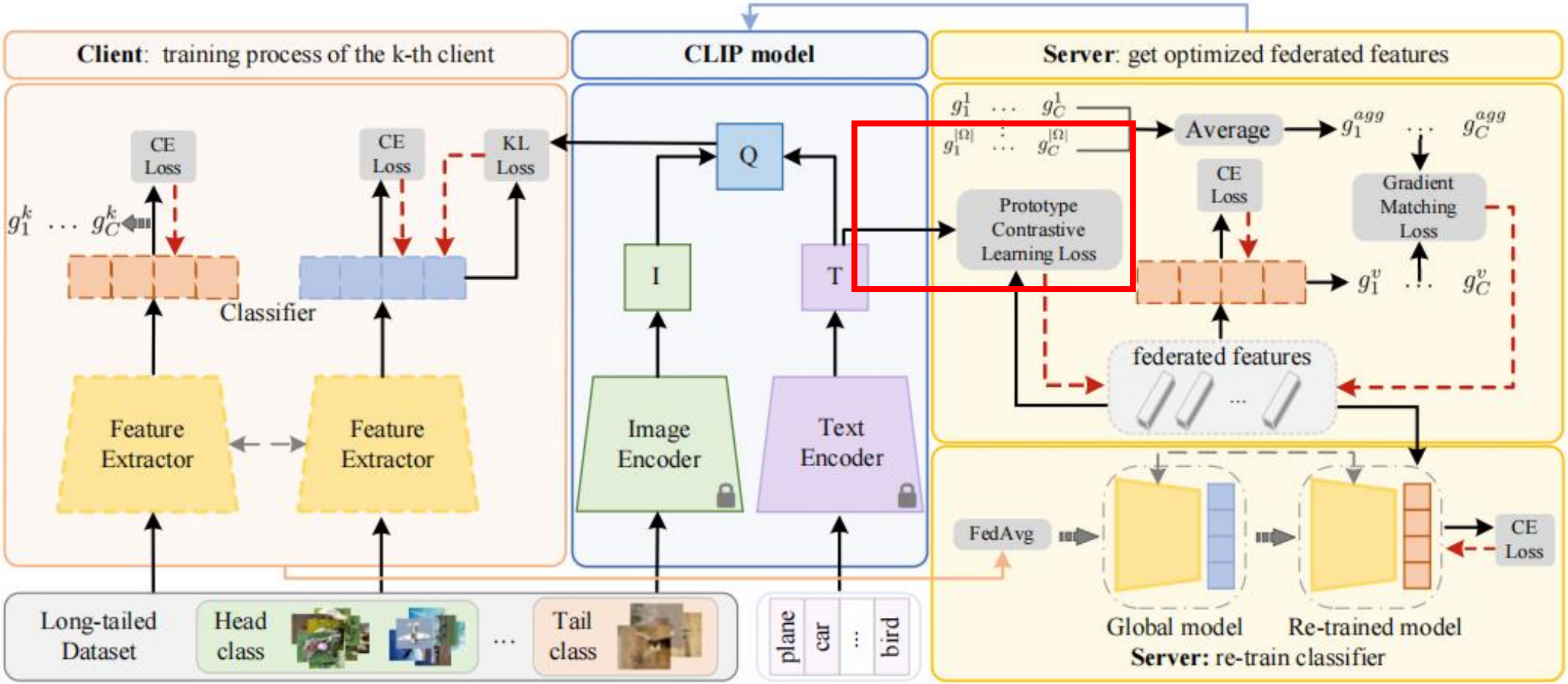
- **Server-side Learning**

$$g_c^{agg} = \frac{1}{|\Omega_c^t|} \sum_{k=1}^{|\Omega_c^t|} g_c^k, \tag{4}$$

$$g_c^v = \frac{1}{m} \sum_{i=1} \nabla_{\hat{\varphi}^t} L_{ce}\left(v_{c,i}^t, y_i\right). \tag{5}$$

$$L_{grad} = D\left(g_c^v, g_c^{agg}\right) = \frac{1}{C} \sum_{j=1}^{C} \left(1 - \frac{g_c^v[j] \cdot g_c^{agg}[j]}{\|g_c^v[j]\| \times \|g_c^{agg}[j]\|}\right), \tag{6}$$

$$L_{pcl} = \sum_{i=1}^{C \times m} -\log \frac{\exp\left(\langle v_{c,i}, f_r^c \rangle / \tau\right)}{\sum_{j=1}^{C \times m} 1_{[j \neq i]} \exp\left(\langle v_{c,i}, v_j \rangle / \tau\right)}, \tag{7}$$

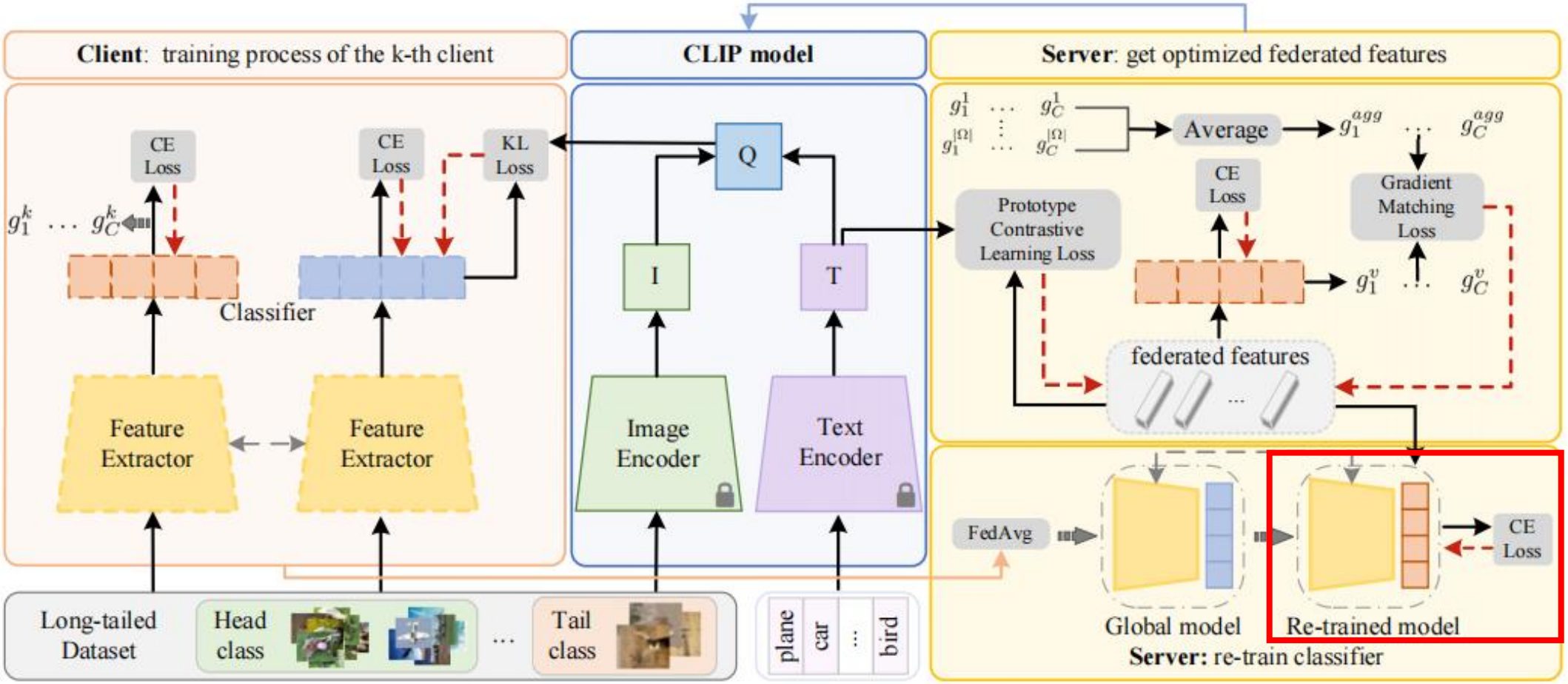


Method——CLIP2FL

- **Classifer Re-Training**

$$w^{t+1} = \sum_{k \in \Omega^t} \frac{|\mathcal{D}^k|}{\sum_{k \in \Omega^t} |\mathcal{D}^k|} w_k^{t+1}. \tag{9}$$

$$\hat{\varphi}^{t+1} \leftarrow \hat{\varphi}^t - \eta \nabla_{\hat{\varphi}} L_{ce}\left(v_i^t, y_i\right). \tag{10}$$



Experiments

Type	Method	CIFAR-10-LT			CIFAR-100-LT		
		IF=100	IF=50	IF=10	IF=100	IF=50	IF=10
Heterogeneity-oriented FL methods	FedAvg	56.17	59.36	77.45	30.34	36.35	45.87
	FedAvgM	52.03	57.11	70.81	30.80	35.33	44.66
	FedProx	56.92	60.89	76.53	31.67	36.30	46.10
	FedDF	55.15	58.74	76.51	31.43	36.22	46.19
	FedBE	55.79	59.55	77.78	31.97	36.39	46.25
	CCVR	69.53	71.89	78.48	33.43	36.98	46.88
	FedNova	57.79	63.91	77.79	32.64	36.62	46.75
Imbalance-oriented FL methods	Fed-Focal Loss	53.83	57.42	73.74	30.67	35.25	45.52
	Ratio Loss	59.75	64.77	78.14	32.95	36.88	46.79
	FedAvg+τ-norm	49.95	51.41	72.08	26.22	33.71	43.65
SOTA	CReFF	70.55	73.08	80.71	34.67	37.64	47.08
Proposed method	CLIP2FL	**73.37** (↑ 2.82)	**75.35** (↑ 2.27)	**81.18** (↑ 0.47)	**37.56** (↑ 2.89)	**41.29** (↑ 3.65)	**48.20** (↑ 1.12)

Table 1: Top-1 classification accuracy(%) on CIFAR-10-LT and CIFAR-100-LT datasets with different FL methods, where the results are referred in (Shang et al. 2022). The best results are marked in bold.

Experiments

Type	Method	ImageNet-LT			
		All	Many	Medium	Few
Heterogeneity-oriented FL methods	FedAvg	23.85	34.92	19.18	7.10
	FedAvgM	22.57	33.93	18.55	6.73
	FedProx	22.99	34.25	17.06	6.37
	FedDF	21.63	31.78	15.52	4.48
	CCVR	25.49	36.72	20.24	9.26
Imbalance-oriented FL methods	Fed-Focal Loss	21.60	31.74	15.77	5.52
	Ratio Loss	24.31	36.33	18.14	7.41
	FedAvg+τ-norm	21.58	31.66	15.76	4.33
SOTA	CReFF	26.31	**37.44**	21.87	10.29
Proposed method	CLIP2FL	**27.72** (↑ 1.41)	35.06 (↓ 2.38)	**27.55** (↑ 5.68)	**24.80** (↑ 14.51)

Table 2: Top-1 classification accuracy(%) on ImageNet-LT dataset with different FL methods, where the results are referred in (Shang et al. 2022). The best results are marked in bold.

Experiments

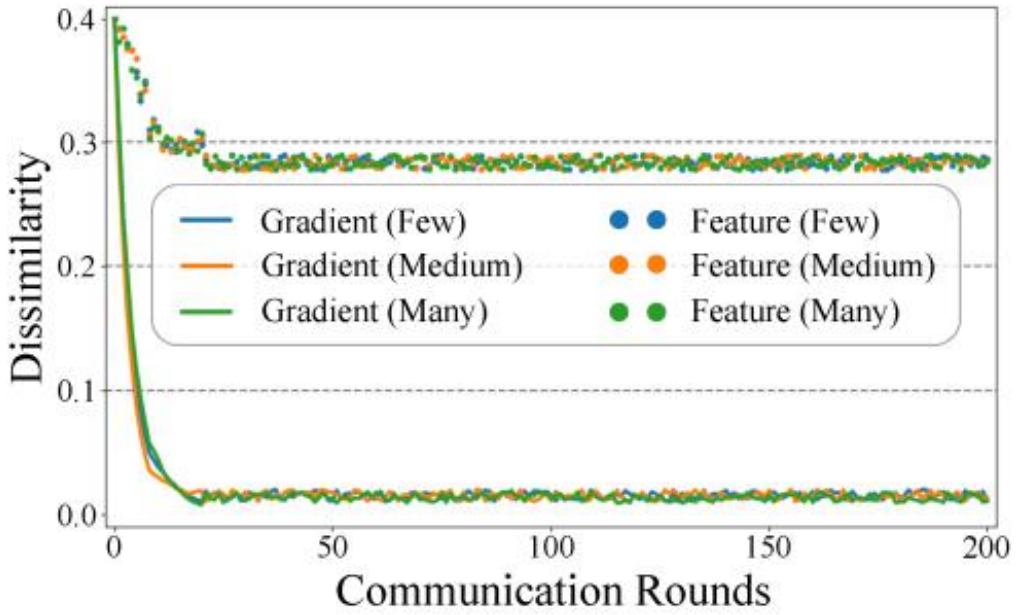


Figure 4: Analysis of two dissimilarities on CIFAR-10-LT with IF = 100. The dotted lines denote the dissimilarities between federated features and real features, and the solid lines denote the dissimilarities between federated feature gradients and real feature gradients.

Method	KL	L_{pcl}	CIFAR-10-LT	CIFAR-100-LT
Baseline			70.55	34.67
w/o L_{pcl}	✓		72.20	35.88
CLIP2FL	✓	✓	**73.37**	**37.56**

Table 3: Ablation study to investigate the effectiveness of CLIP2FL on CIFAR-10-LT and CIFAR-100-LT with IF = 100.

Thanks