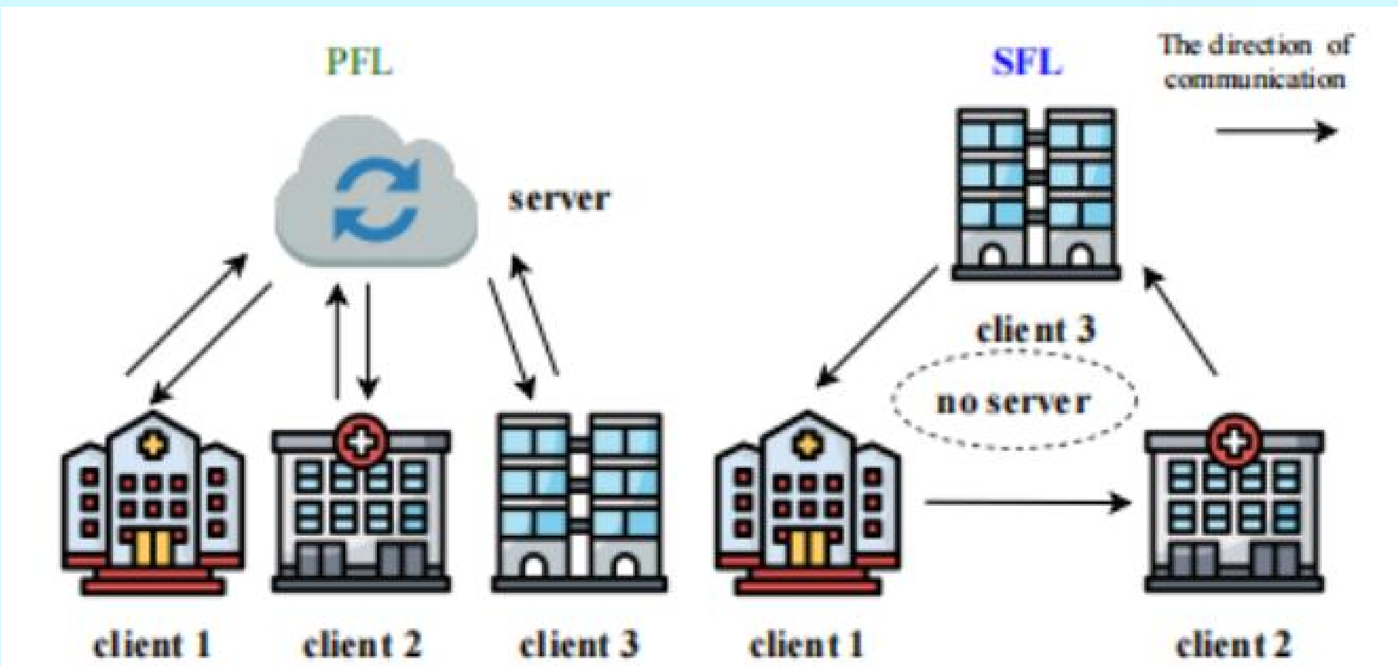
SPFL: Sequential Updates with Parallel Aggregation for Enhanced Federated Learning Under Category and Domain Shifts

Haoyuan Liang* , Shilei Cao1* ,Guowen Li , Zhiyu Ye , Haohuan Fu , Juepeng Zheng†

Motivation

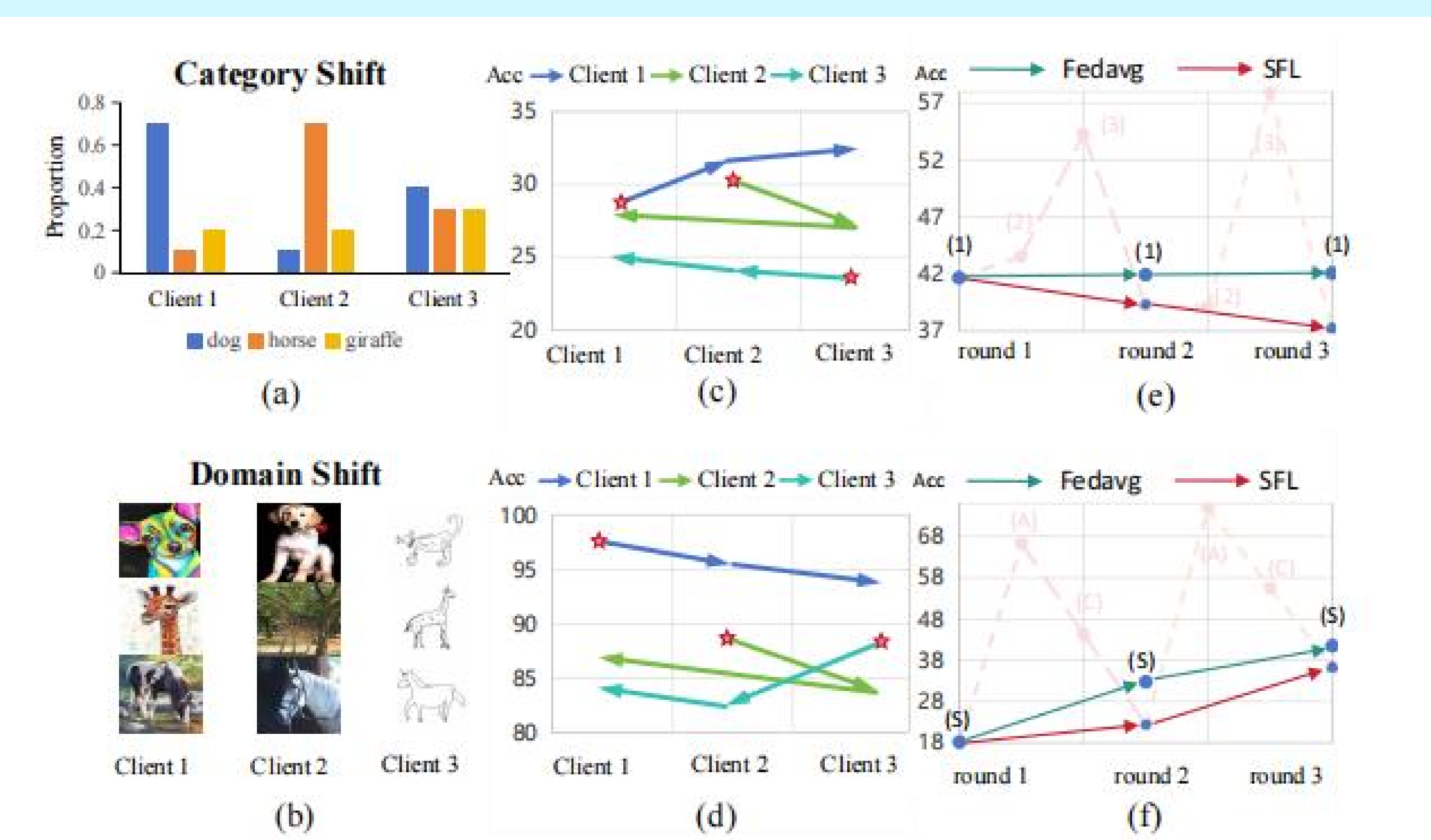
➢ Why should we use sequential federated learning (SFL)?

Federated Learning (FL) has recently emerged as the primary approach to overcoming data silos, enabling collaborative model training without sharing sensitive or proprietary data. Parallel Federated Learning (PFL) aggregates models trained independently on each client's local data, which could prevent the model from converging to the optimal solution due to limited data exposure. In contrast, Sequential Federated Learning (SFL) allows models to traverse client datasets sequentially, enhancing data utilization.



➢ The challenges with SFL in real-world scenarios

However, SFL effectiveness is limited in real-world Non-IID scenarios characterized by category shift (inconsistent class distributions) and domain shift (distribution discrepancies). These shifts cause two critical issues: update order sensitivity, where model performance varies significantly with the sequence of client updates; and catastrophic forgetting, where the model forgets previously learned features when trained on new client data.

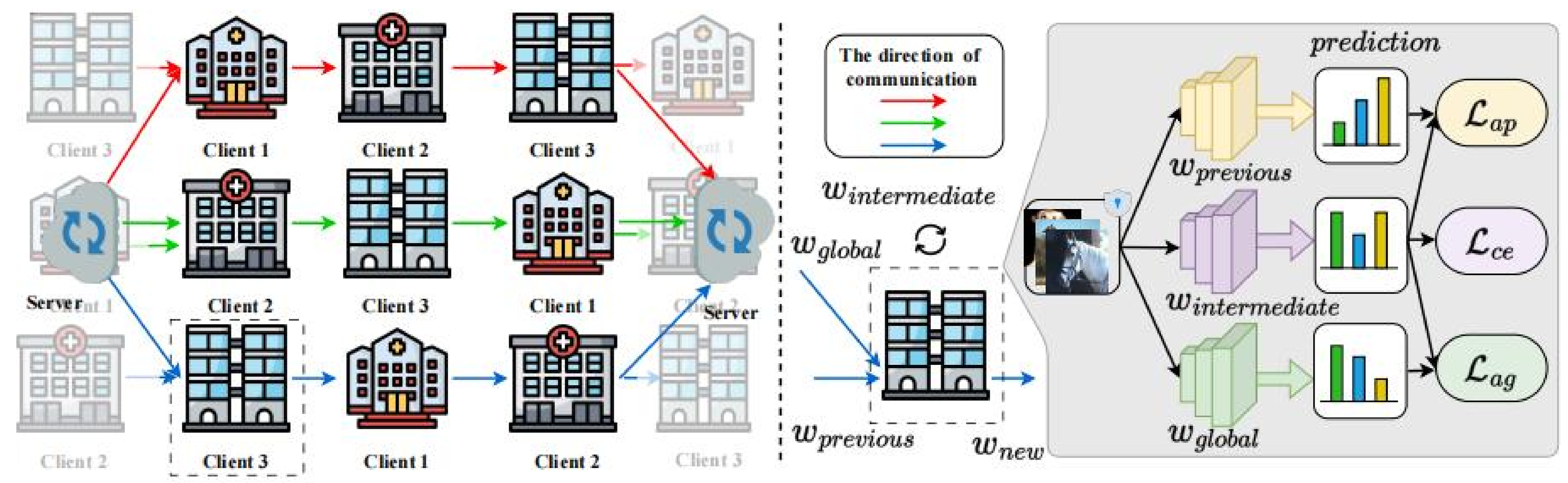


Note: ★ represents the start of updates from client *i*, → represents the direction of transmission

Therefore, based on SFL, we propose a novel updating framework, SPFL (Sequential updates with Parallel aggregation Federated Learning), that can be integrated into existing PFL methods. It integrates sequential updates with parallel aggregation to enhance data utilization and ease update order sensitivity. Meanwhile, we give the convergence analysis of SPFL under strong convex, general convex, and non-convex conditions, proving that this update scheme is significantly better than PFL and SFL. Additionally, we introduce the GLAM (Global-Local Alignment Module) to mitigate catastrophic forgetting by aligning the predictions of the local model with those of previous models and the global model during training.

Our Method

➢ The framework of SPFL (with GLAM)



➢ The Alignment scheme for GLAM

$$\mathcal{L}_{ce} = -\frac{1}{B}\sum_{i=1}^{B} \boldsymbol{y}_i{}^T \log(f(\boldsymbol{w}_i; \boldsymbol{x}_i^{\pi_m^n,k})) \quad \mathcal{L}_{ag} = -\frac{1}{B}\sum_{i=1}^{B} f(\boldsymbol{w}_g; x_i^{\pi_m^n,k})^T \log(f(\boldsymbol{w}_i; x_i^{\pi_m^n,k}))$$

$$\mathcal{L}_{ap} = -\frac{1}{2B}\sum_{i=1}^{B} KL(P||Q) + KL(Q||P), \qquad s.t.\ P = f(\boldsymbol{w}_i; x_i^{\pi_m^n,k}), Q = f(\boldsymbol{w}_p; x_i^{\pi_m^n,k})$$

$$\mathcal{L} = \tau\mathcal{L}_{ap} + \rho\mathcal{L}_{ag} + \mathcal{L}_{ce}$$

➢ The Convergence analysis of SPFL

Strongly convex: *Under Assumptions 2, 3, and 5, there exists a constant effective learning rate* $\frac{1}{\mu R} \le \tilde{\eta} \le \frac{1}{6L}$ *and weights* $\theta_r = \left(1 - \frac{\mu\tilde{\eta}}{2}\right)^{-(r+1)}$*, such that the following holds:*

$$\mathbb{E}\left[\mathcal{L}(\bar{\boldsymbol{w}}^{(R)}) - \mathcal{L}(\boldsymbol{w}^*)\right] \le \frac{9}{2}\mu\mathcal{A}^2 \exp\left(-\frac{1}{2}\mu\tilde{\eta}R\right) + \frac{12\tilde{\eta}\sigma^2}{M^2K} + \frac{18L\tilde{\eta}^2\sigma^2}{MK} + \frac{18L\tilde{\eta}^2\zeta_*^2}{MK} \tag{5}$$

General convex: *Under Assumptions 2, 3, and 5, there exists a constant effective learning rate* $\tilde{\eta} \le \frac{1}{6L}$ *and weights* $\theta_r = 1$*, such that the following holds:*

$$\mathbb{E}\left[\mathcal{L}(\bar{\boldsymbol{w}}^{(R)}) - \mathcal{L}(\boldsymbol{w}^*)\right] \le \frac{3\mathcal{A}^2}{\tilde{\eta}R} + \frac{12\tilde{\eta}\sigma^2}{M^2K} + \frac{18L\tilde{\eta}^2\sigma^2}{MK} + \frac{18L\tilde{\eta}^2\zeta_*^2}{MK} \tag{6}$$

Non-convex: *Under Assumptions 2, 3, and 4, there exists a constant effective learning rate* $\tilde{\eta} \le \frac{1}{6L(\beta+1)}$ *and weights* $\theta_r = 1$*, such that the following holds:*

$$\min_{0\le r\le R} \mathbb{E}\left[\left\|\nabla\mathcal{L}(\boldsymbol{w}^{(r)})\right\|^2\right] \le \frac{3\mathcal{B}}{\tilde{\eta}R} + \frac{3L\tilde{\eta}\sigma^2}{M^2K} + \frac{27L^2\tilde{\eta}^2\sigma^2}{8MK} + \frac{27L^2\tilde{\eta}^2\zeta^2}{8M} \tag{7}$$

$\mathcal{A} := \|\boldsymbol{w}^{(0)} - \boldsymbol{w}^*\|$ *for the convex cases and* $\mathcal{B} := \mathcal{L}(\boldsymbol{w}^{(0)}) - \mathcal{L}^*$ *for the non-convex case.*

➢ PFL VS SFL VS SPFL (Ours)

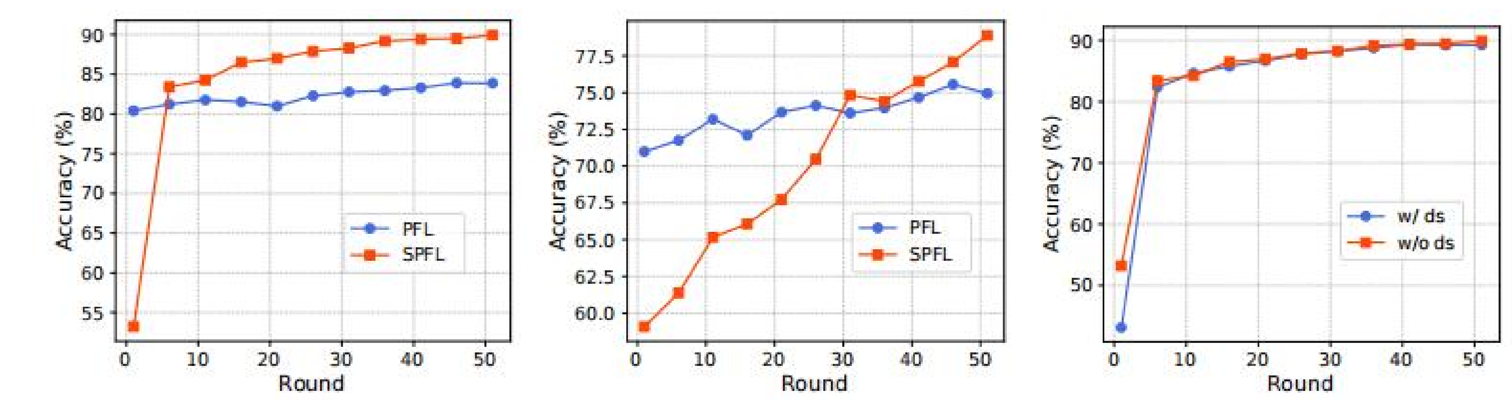
Method	Bound ($\mathcal{A} = \|\boldsymbol{w}^{(0)} - \boldsymbol{w}^*\|$)
PFL (FedAvg) [27]	$\frac{\sigma^2}{\mu MKR} + \frac{L\sigma^2}{\mu^2KR^2} + \frac{L\zeta_*^2}{\mu^2R^2} + \mu\mathcal{A}^2\exp\left(-\frac{\mu R}{L}\right)$
PFL (Scaffold) [16]	$\frac{\sigma^2}{\mu MKR} + \frac{L\sigma^2}{\mu^2KR^2} + \frac{L\zeta_*^2}{\mu^2R^2} + \mu\mathcal{A}^2\exp\left(-\frac{\mu R}{L}\right)$
SFL [27]	$\frac{\sigma^2}{\mu MKR} + \frac{L\sigma^2}{\mu^2MKR^2} + \frac{L\zeta_*^2}{\mu^2MR^2} + \mu\mathcal{A}^2\exp\left(-\frac{\mu R}{L}\right)$
SPFL (ours)	$\frac{\sigma^2}{\mu M^2KR} + \frac{L\sigma^2}{\mu^2MKR^2} + \frac{L\zeta_*^2}{\mu^2MR^2} + \mu\mathcal{A}^2\exp\left(-\frac{\mu R}{L}\right)$

Evaluation Results

➢ The Main comparative experiments

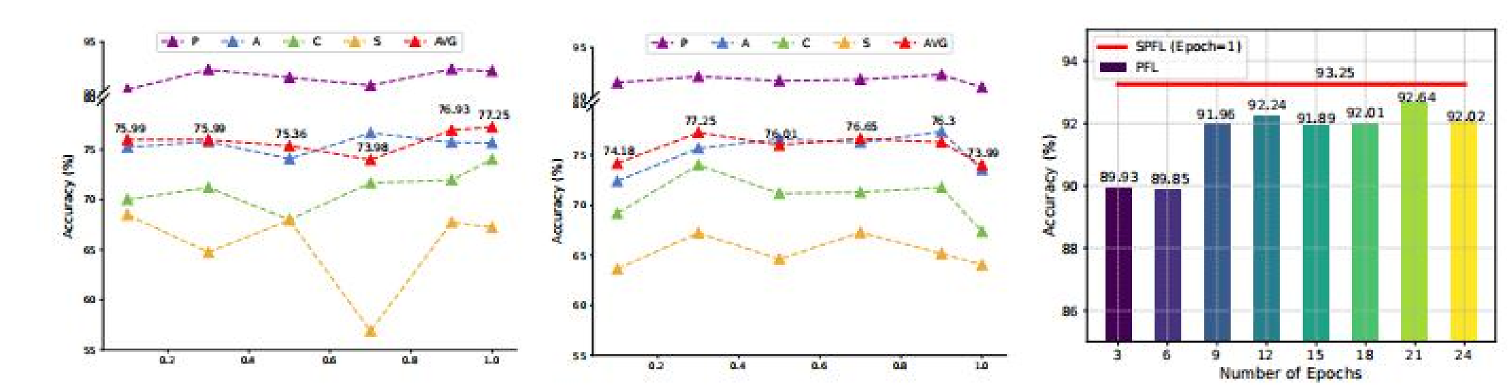
Method	Cifar-10 resnet18 (num=10)	Cifar-100 resnet18 (num=10)	CINIC-10 simple-cnn (num=10)	Fmnist resnet18 (num=10)	Avg
FedAvg [5]	75.00	57.55	39.74	81.13	63.35
+ SPFL	78.88 (**+3.88**)	64.33 (**+6.78**)	41.01 (**+1.27**)	84.75(**+3.62**)	67.24 (**+3.89**)
FedDc [28]	80.52	64.44	40.92	83.44	67.33
+ SPFL	85.90(**+5.38**)	69.88 (**+5.44**)	44.14 (**+3.23**)	87.67(**+4.23**)	71.90 (**+4.57**)
FedDyn [17]	77.91	64.11	40.91	81.35	66.03
+ SPFL	80.90(**+2.99**)	64.24 (**+0.13**)	43.06 (**+2.15**)	83.23(**+1.88**)	67.86 (**+1.83**)
FedNova [29]	76.02	57.64	39.82	81.38	63.46
+ SPFL	79.31(**+3.29**)	64.47 (**+6.83**)	41.32 (**+1.50**)	84.85(**+3.47**)	67.49(**+4.03**)
FedProx [30]	76.04	57.56	39.65	81.23	63.62
+ SPFL	77.25(**+1.21**)	60.46 (**+2.90**)	40.27 (**+0.62**)	83.89(**+2.66**)	65.47 (**+1.85**)
MOON [31]	78.70	58.44	40.11	81.29	63.68
+ SPFL	79.98 (**+1.28**)	63.54 (**+5.20**)	41.11 (**+1.00**)	84.83(**+3.54**)	67.37(**+3.69**)
SCAFFOLD [16]	76.38	56.15	36.00	80.71	62.31
+ SPFL	78.73 (**+2.35**)	65.05 (**+8.90**)	39.52 (**+3.52**)	85.80(**+5.09**)	67.26 (**+4.95**)
FedDisco [24]	76.32	57.50	39.67	81.24	63.68
+ SPFL	78.58(**+2.26**)	64.44 (**+6.94**)	40.79 (**+1.12**)	84.27(**+3.02**)	67.13(**+3.45**)

➢ The Convergence Analysis Experiment



(a) IID (b) Non-IID (c) Without/With weighted average

➢ The Parametric ablation experiment



(a) Sensitivity to ρ (b) Sensitivity to τ (c) The efficiency of SPFL

➢ The Resource consumption comparison

Table 6: Efficiency Comparison on CIFAR-10 (Dirichlet α=0.1, Number of clients=10)

Method	Accuracy(%)	Rounds	Comm. Cost (GB)	Total Time(s)	Compute Cost(GB)
PFL	79.15	500	116.90	3419.91	175.35
SPFL(Ours)	84.17	15	21.042	678.6	87.675

Table 7: Efficiency Comparison on CIFAR-10 (Dirichlet α=0.1, Number of clients=100)

Method	Accuracy(%)	Rounds	Comm. Cost (TB)	Total Time(s)	Compute Cost(TB)
PFL	81.38	2000	18.30	30259.82	27.46
SPFL(Ours)	86.68	10	4.68	14713.13	22.85

If you are interested in our research or would like to discuss potential collaboration, please feel free to contact us:

lianghy68@mail2.sysu.edu.cn

Team Homepage

WeChat