
Open Set Recognition Federated Learning



DMQA Open Seminar (2025. 10. 17)

Data Mining & Quality Analytics Lab.

김다빈



발표자 소개



- **김다빈 (dabin Kim)**
 - 고려대학교 산업경영공학과 대학원 재학
 - Data Mining & Quality Analytics Lab. (김성범 교수님)
 - M.S Student (2025.09 ~ Present)
- **Research Interest**
 - Federated Learning
 - Open Set Recognition
 - Semi-supervised Learning
- **Contact**
 - heydabins@korea.ac.kr



Contents

Open Set Recognition Federated Learning

❖ Introduction

❖ Related Works

- FedOS: using open-set learning to stabilize training in federated learning
- Federated Open Set Recognition with Parameter Disentanglement
- Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

❖ Conclusion



Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning



Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning



Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning

데이터 프라이버시 문제를 해결하기 위해서 여러 클라이언트가 로컬 데이터를 중앙 서버로 직접 전송하지 않고, 각자의 데이터로 모델을 학습한 후 모델 파라미터만을 공유하여 전역 모델을 공동으로 학습하는 방식



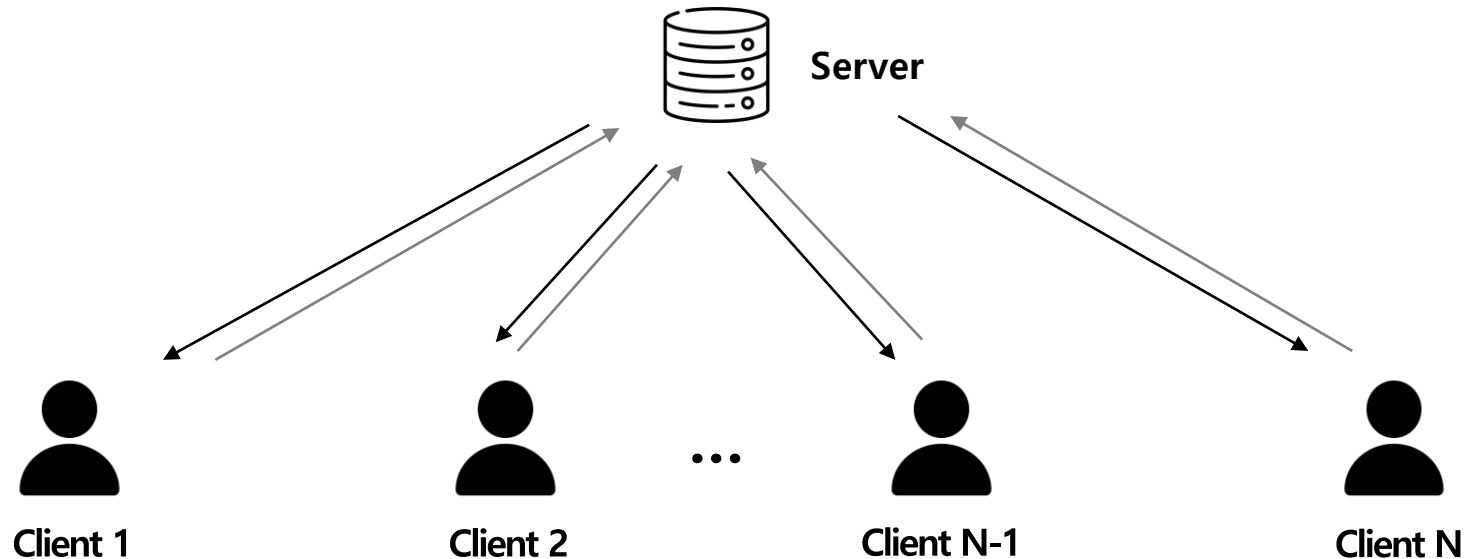
Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning

데이터 프라이버시 문제를 해결하기 위해서 여러 클라이언트가 로컬 데이터를 중앙 서버로 직접 전송하지 않고, 각자의 데이터로 모델을 학습한 후 모델 파라미터만을 공유하여 전역 모델을 공동으로 학습하는 방식



Introduction

Open Set Recognition Federated Learning

Open Set

데이터 프라이버시 문제를 해결하기
고, 각자의 데이터로 모델을 학습

Federated Learning

데이터를 중앙 서버로 직접 전송하지 않
전역 모델을 공동으로 학습하는 방식



Client 1

종료

DMQA Open Seminar

Federated Learning

2025. 03. 07
Korea University
Data Mining & Quality Analytics Lab.
최지형

Federated Learning

발표자:  **최지형**

 2025년 3월 7일

 오후 12시 ~

 온라인 비디오 시청 (YouTube)

 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

Client 2

Client N-1



Client N



Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning



Introduction

Open Set Recognition Federated Learning

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning

여러 클라이언트가 가진 데이터가 서로 다른 클래스 분포를 가지고 있을 때,
알려진 클래스(known)와 알려지지 않은 클래스(unknown)를 동시에 고려하는 연합 학습 방식



Introduction

Open Set Recognition Federated Learning

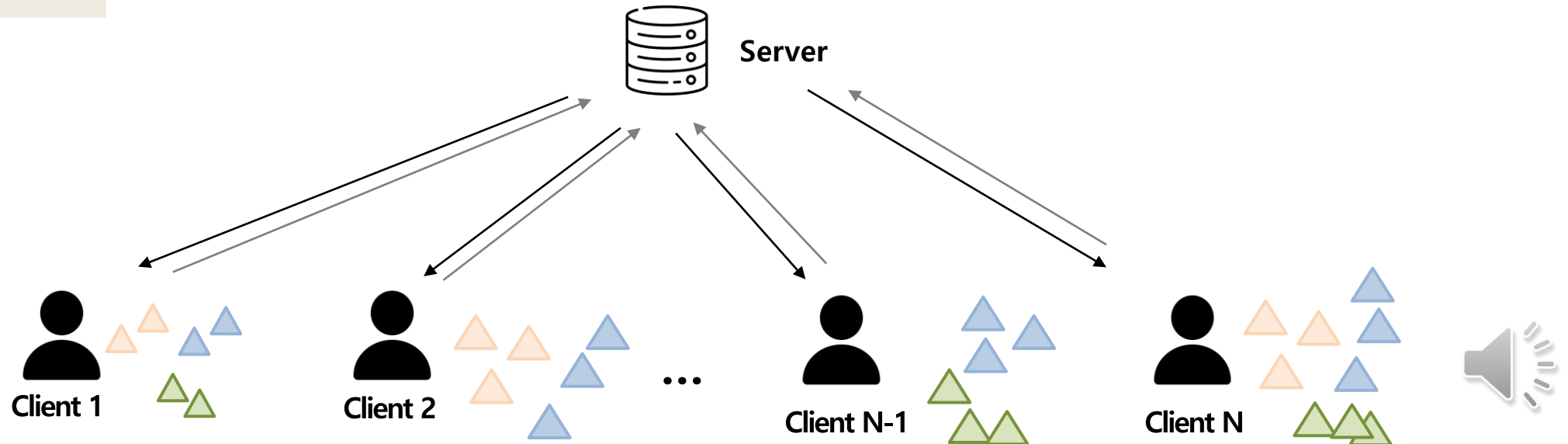
△ △ △ : Known classes
△ △ △ : Unknown class

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning

여러 클라이언트가 가진 데이터가 서로 다른 클래스 분포를 가지고 있을 때,
알려진 클래스(known)와 알려지지 않은 클래스(unknown)를 동시에 고려하는 연합 학습 방식

Train



Introduction

Open Set Recognition Federated Learning

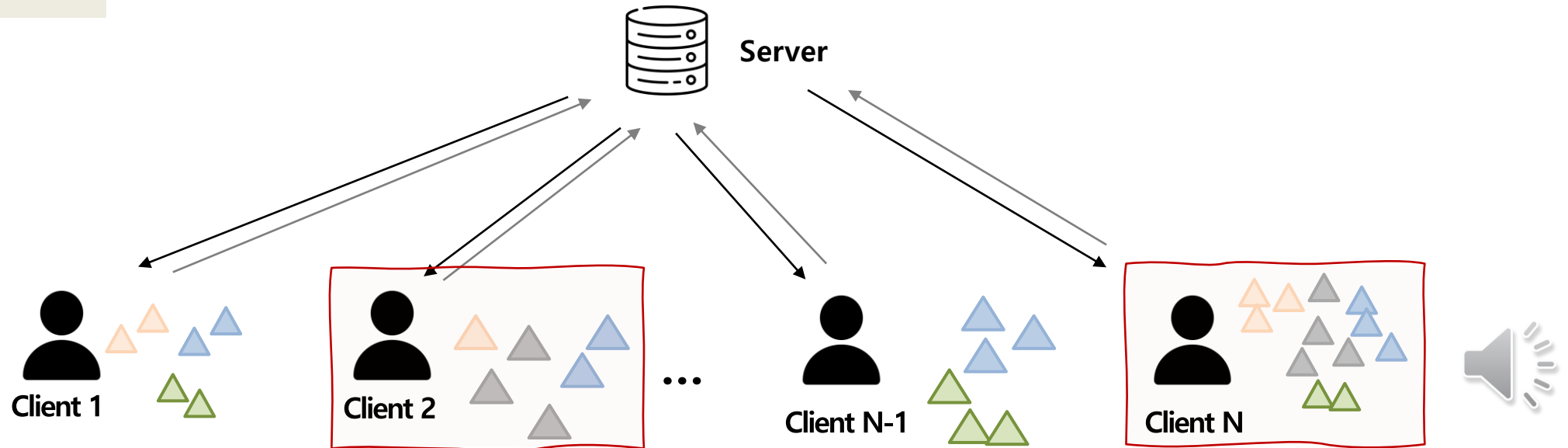
△ △ △ : Known classes
△ △ △ : Unknown class

Open-Set Semi-Supervised Learning + Federated Learning

Open Set Recognition Federated Learning

여러 클라이언트가 가진 데이터가 서로 다른 클래스 분포를 가지고 있을 때,
알려진 클래스(known)와 알려지지 않은 클래스(unknown)를 동시에 고려하는 연합 학습 방식

Test



Related Works

Open Set Recognition Federated Learning

FedOS [arXiv 2022]

FedOS: using open-set learning to stabilize training in federated learning

Mohamad Mohamad Julian Neubert Juan Segundo Argayo
mohamad_m.2@live.com julian.neubert@gmx.de jargayo@gmail.com

Abstract

Federated Learning is a recent approach to train statistical models on distributed datasets without violating privacy constraints. The data locality principle is preserved by sharing the model instead of the data between clients and the server. This brings many advantages but also poses new challenges. In this report, we explore this new research area and perform several experiments to deepen our understanding of what these challenges are and how different problem settings affect the performance of the final model. Finally, we present a novel approach to one of these challenges and compare it to other methods found in literature. The code is available on our github ¹.

ditional distributed optimization and requires dealing with heterogeneous data

2. Related Work

FedProx. There are two key challenges in Federated Learning that differentiate it from traditional distributed optimization: (1) significant variability in terms of the system's characteristics on each device in the network (system heterogeneity), and (2) non-identically distributed data across the network (statistical heterogeneity). FedProx (Tian Li et al. [12]) can be viewed as a generalization and re-parameterization of FedAvg. In this work, the authors have theoretically proven convergence guarantees when learning over data from non-identical distributions (statistical heterogeneity), while also adhering to the constraints each device-level system has by allowing each participating device to perform a variable amount of work (system heterogeneity). They demonstrate that in highly heterogeneous settings, FedProx has significantly more stable and accurate convergence behavior relative to FedAvg, improving absolute test accuracy by 22% on average.

FedIR. As well as [12], the authors of this paper (Hou et al. [6]) recognized that one of the challenges in terms of data diversity relies on the fact that data at the source is far from independent and identically distributed (IID). In

FedPD [ICCV 2022]

FedPD: Federated Open Set Recognition with Parameter Disentanglement

Chen Yang¹ Meihu Zhu¹ Yifan Liu² Yixuan Yuan^{1,2*}
¹City University of Hong Kong ²The Chinese University of Hong Kong
{cyang53, meihuizhu}@cityu.edu.hk 115519505081@hk.cuhk.edu.hk yxyuan@cuhk.edu.hk

Abstract

Existing federated learning (FL) approaches are deployed under the unrealistic closed-set setting, with both training and testing classes belong to the same set, which makes the global model fail to identify the unseen classes as 'unknown'. To this end, we aim to study a novel problem of federated open-set recognition (FedOSR), which learns an open-set recognition (OSR) model under federated paradigm such that it classifies seen classes while at the same time detects unknown classes. In this work, we propose a parameter disentanglement guided federated open-set recognition (FedPD) algorithm to address two core challenges of FedOSR: cross-client inter-set interference between learning closed-set and open-set knowledge and cross-client inter-set inconsistency by data heterogeneity. The proposed FedPD framework mainly leverages two modules, i.e., local parameter disentanglement (LPD) and global divide-and-conquer aggregation (GDCA), to first disentangle client OSR model into different subnetworks, then align the corresponding parts across clients for matched model aggregation. Specifically, on the client side, LPD decomposes an OSR model into a closed-set subnetwork and an open-set subnetwork by the task-related importance, thus preventing inter-set interference. On the server side, GDCA first partitions the two subnetworks into specific and shared parts, and subsequently aligns the corresponding parts through optimal transport to eliminate parameter misalignment. Extensive experiments on various datasets demonstrate the superior performance of our proposed method.



Figure 1. Parameter disentanglement on FedOSR models and comparison of various aggregation strategies on FedOSR setting. (a) FedAvg simply aggregates multiple OSR models, leading to model collapse. (b) Our FedPD first aligns corresponding close-specific, open-specific and shared parts by optimal transport (OT), then aggregates aligned parameters. The curves inside the network box represent different parameter distribution of each party.

tions and difficult to integrate into a centralized dataset, owing to increasing privacy and ethical concerns, especially for those sensitive data such as location-based services or health information [21]. To break this dilemma, federated learning (FL) [2, 22, 20] provides a privacy-preserving paradigm that allows local clients to collaboratively train a shared global model without data sharing.

FedOSS [IEEE 2024]

FedOSS: Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

Meihu Zhu¹, Jing Liao¹, Member, IEEE, Jun Liu¹, Member, IEEE, and Yixuan Yuan¹, Member, IEEE

Abstract—Open set recognition (OSR) aims to accurately classify known diseases and recognize unseen diseases as the unknown class in medical scenarios. However, in existing OSR approaches, gathering data from distributed sites to construct large-scale centralized training datasets usually leads to high privacy and security risk, which could be alleviated elegantly via the popular cross-site training paradigm, federated learning (FL). To this end, we represent the first effort to formulate federated open set recognition (FedOSR), and meanwhile propose a novel Federated Open Set Synthesis (FedOSS) framework to address the core challenge of FedOSR: the unavailability of unknown samples for all anticipated clients during the training phase. The proposed FedOSS framework mainly leverages two modules, i.e., Discrete Unknown Sample Synthesis (DUSS) and Federated Open Space Sampling (FOSS), to generate virtual unknown samples for learning decision boundaries between known and unknown classes. Specifically, DUSS exploits inter-client knowledge inconsistency to recognize known samples near decision boundaries and then pushes them beyond decision boundaries to synthesize discrete virtual unknown samples. FOSS unites these generated unknown samples from different clients to estimate the class-conditional distributions of open data space near decision boundaries and further samples open data, thereby improving the diversity of virtual unknown samples. Additionally, we conduct comprehensive ablation experiments to verify the effectiveness of DUSS and FOSS. FedOSS shows superior performance on public medical datasets in comparison with state-of-the-art approaches. The source code is available at <https://github.com/CHU-AM-Group/FedOSS>.

achieved remarkable performance on various medical classification tasks [11], [2], [5]. However, these methods are usually evaluated in a closed-set setting, in which categories in test set are known and same as training set [4]. The closed-set setting is typically unreasonable in real-world medical scenarios, since rare or unknown diseases probably arise without any warning and would be misclassified into one of the known diseases [4], [15], [6], [17], [6], [9]. This problem poses an enormous risk to the public health and impedes the application of intelligent healthcare systems. Open-set recognition (OSR) [10], [11] is thus proposed to solve this issue, which aims to accurately classify known classes while at the same time identify open-set data as the unknown class.

Despite the impressive performance of existing OSR methods [4], [15], [6], [17], [6], [9] in open-set setting, they heavily depend on the availability of large-scale centralized datasets. Since the data in a single medical institution are usually limited, one common solution is to gather patient information from different hospitals [12]. Yet, due to growing privacy concerns or legal restrictions [13], distributed patient data are not able to be directly shared across institutions. Federated learning (FL) [14], [15] is a promising paradigm of decentralized machine learning to provide a feasible solution to this dilemma, which learns a global model by sharing the model parameters of clients (hospitals) instead of their raw data, under the orchestration of a trustworthy cloud server. Nevertheless, implementing OSR in such a decentralized FL framework is challenging and has not been investigated so far. Driven by above realistic issues, in this paper, we represent



1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

   : Known classes
   : Unknown class

FedOS

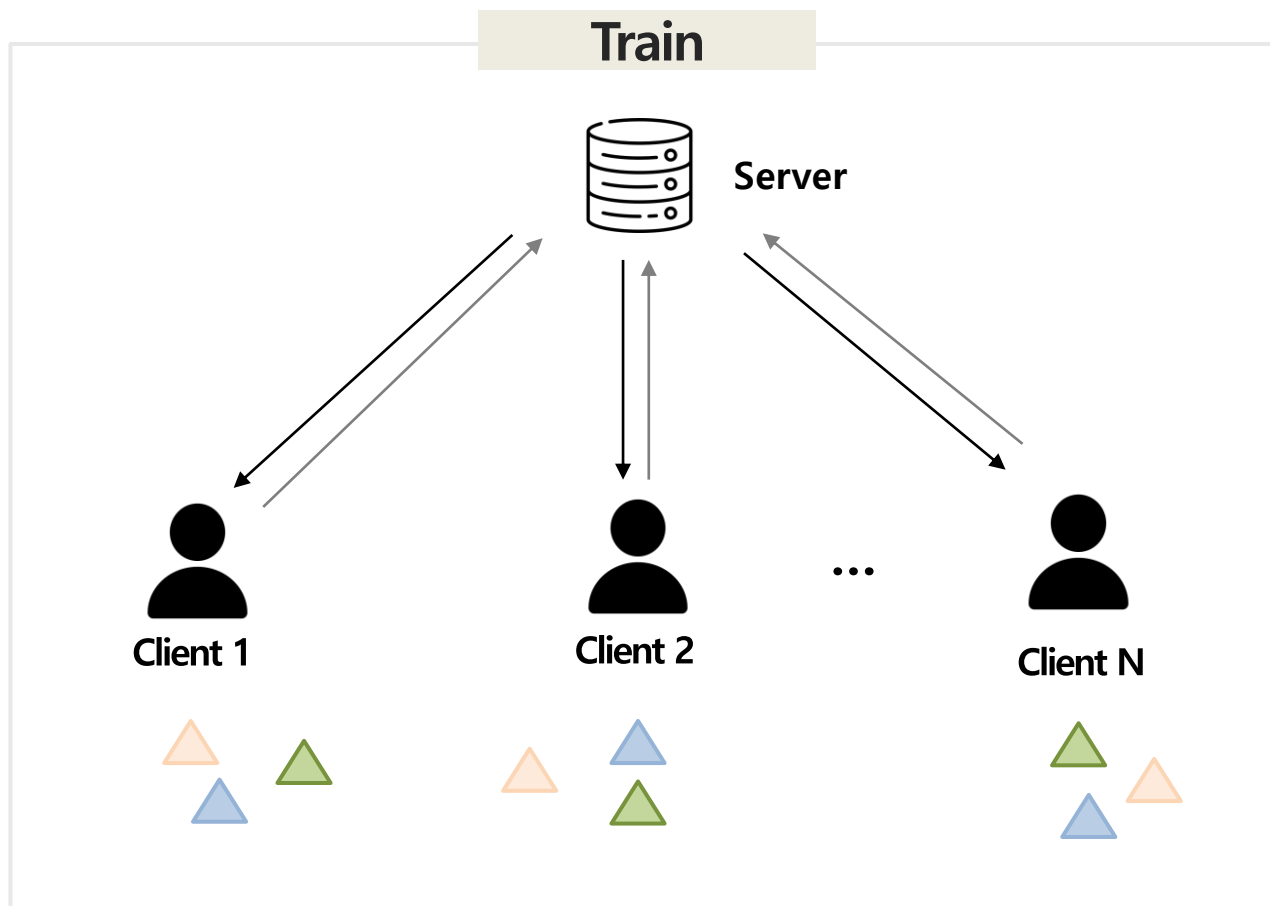


1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

△ △ △ : Known classes
△ △ △ : Unknown class

FedOS

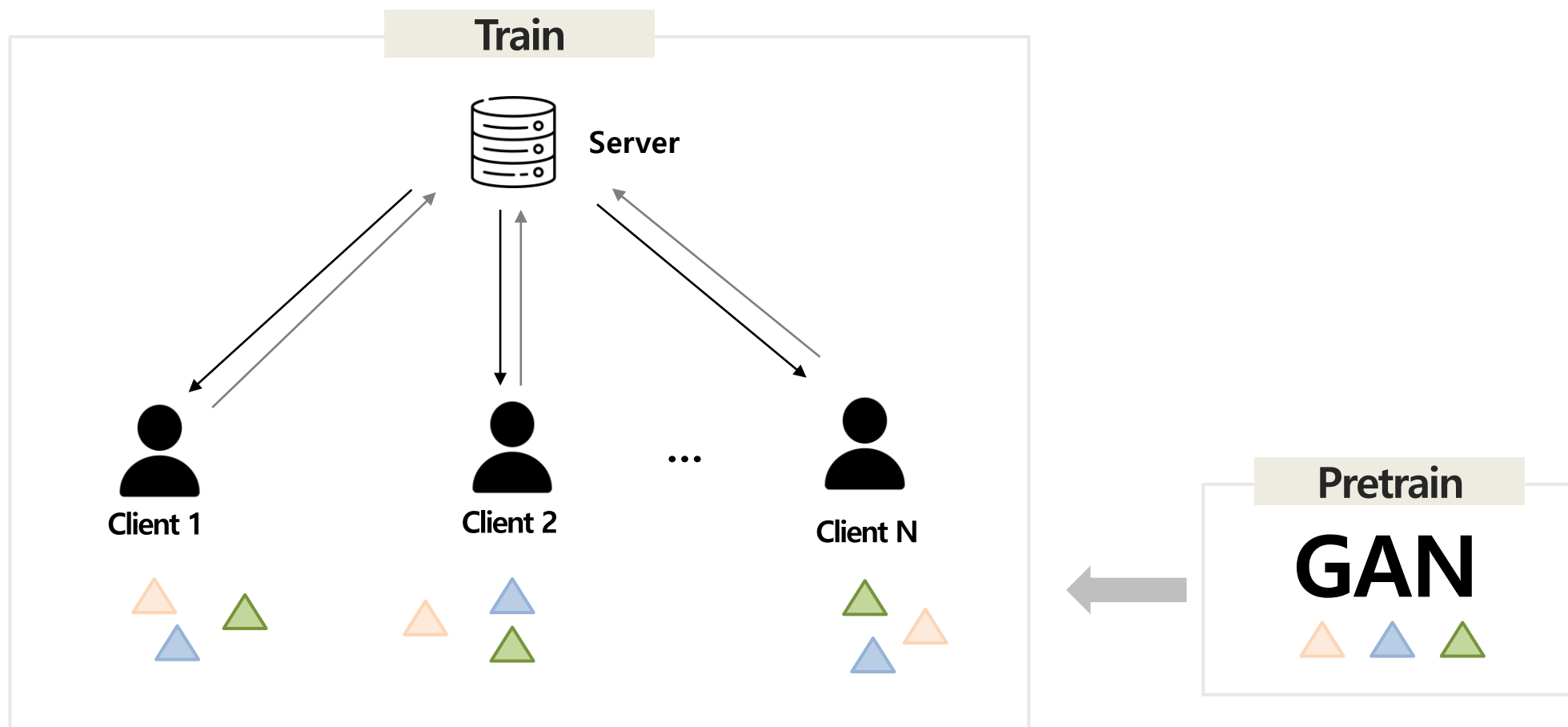


1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

△ △ △ : Known classes
△ △ △ : Unknown class

FedOS

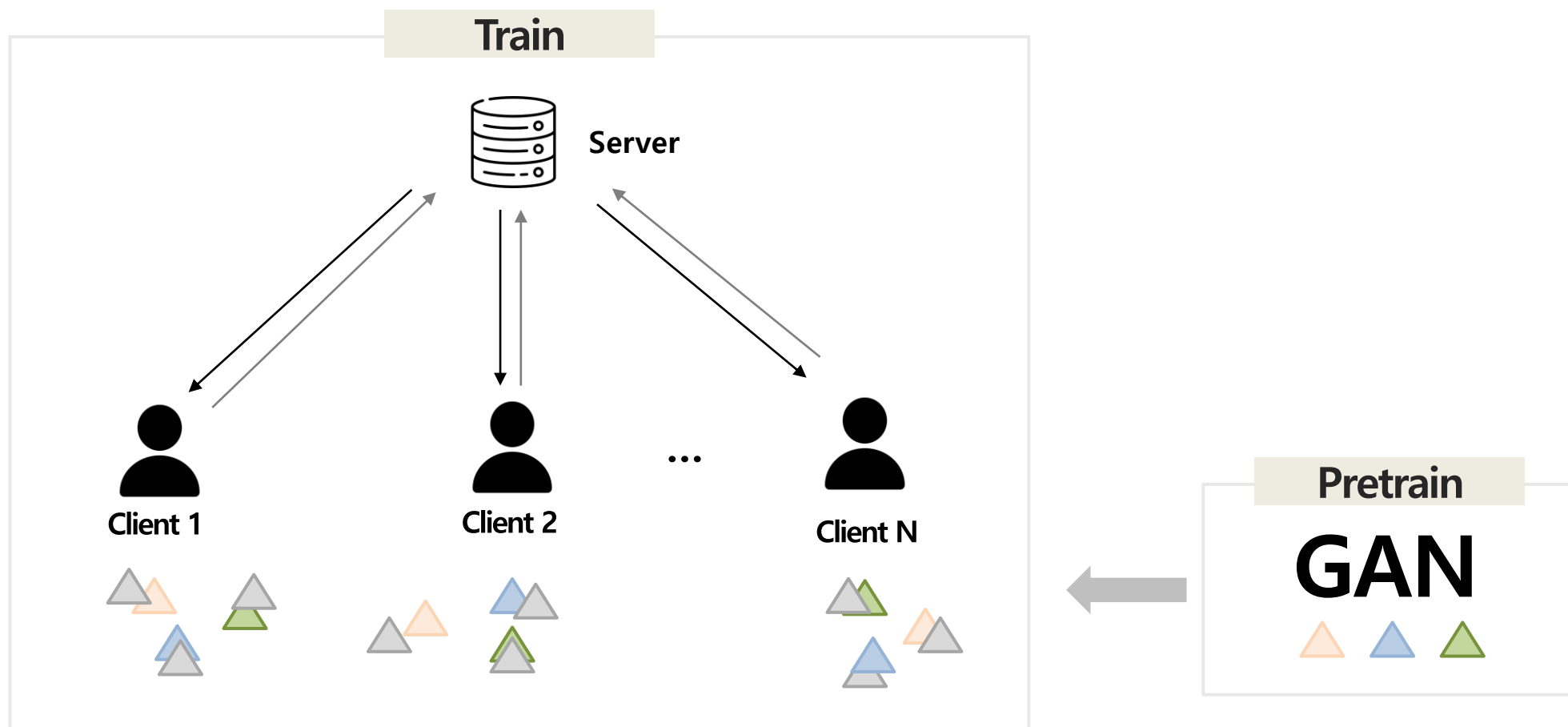


1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

△ △ △ : Known classes
△ △ △ : Unknown class

FedOS

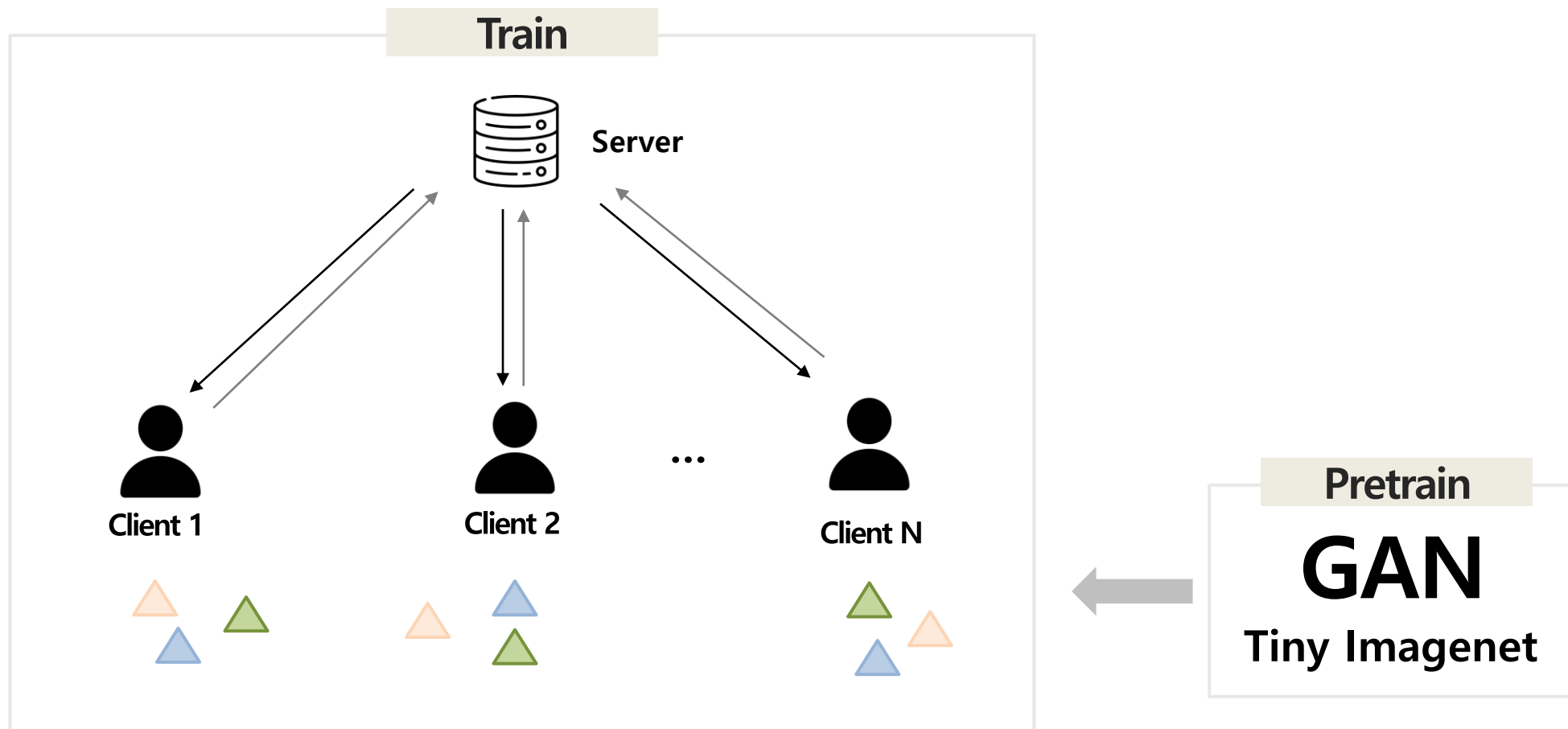


1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

△ △ △ : Known classes
△ △ △ : Unknown class

FedOS

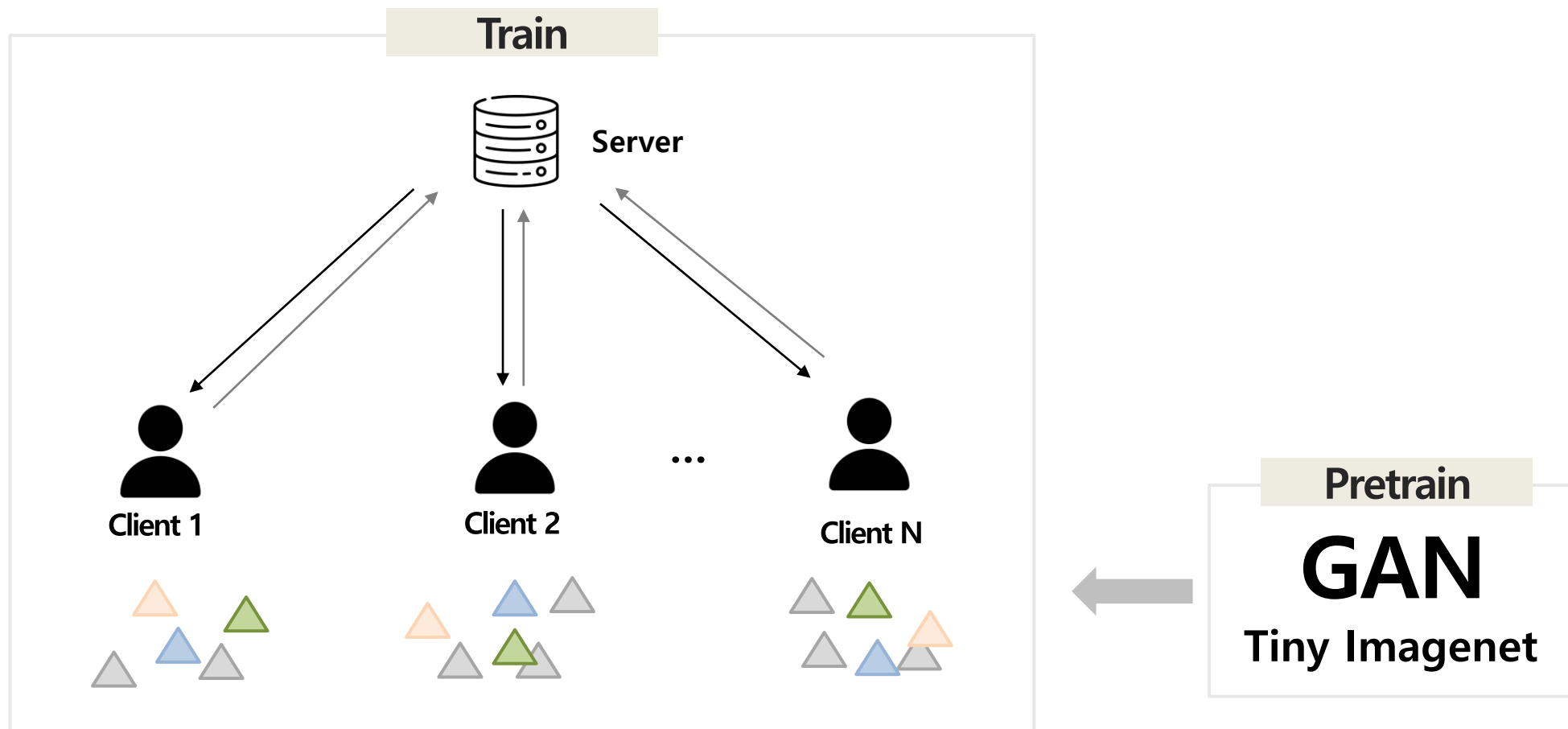


1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

△ △ △ : Known classes
△ △ △ : Unknown class

FedOS



1. FedOS

FedOS: using open-set learning to stabilize training in federated learning

Experiments & Results

- 데이터셋: CIFAR-10
- 정량 평가: Accuracy
- Known class의 개수: 6 / Unknown class의 개수: 4

Relative accuracy @N rounds								
Methods	Alpha=1				Alpha=0.01			
	@125	@250	@375	@500	@125	@250	@375	@500
FedAvg	65.94	82.17	83.03	88.34	40.41	41.92	51.19	60.75
FedIR	77.66	82.98	89.03	89.36	51.75	59.56	64	69.03
FedProx	63.85	82.07	86.51	91.4	28.81	37.64	48.58	48.58
FedOS	77.92	84.09	94.98	95.85	46.6	56.13	64.7	69.11



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

closed-set과 open-set에 관여하는 파라미터는 분리되어있다

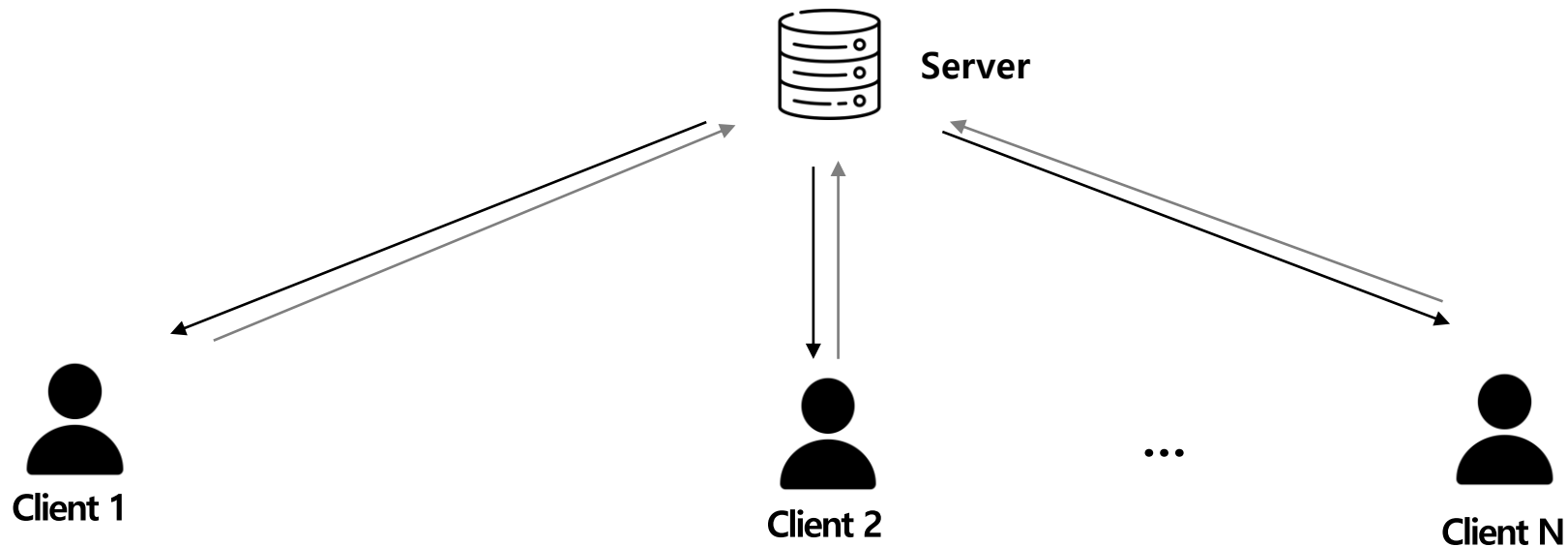


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

closed-set과 open-set에 관여하는 파라미터는 분리되어있다

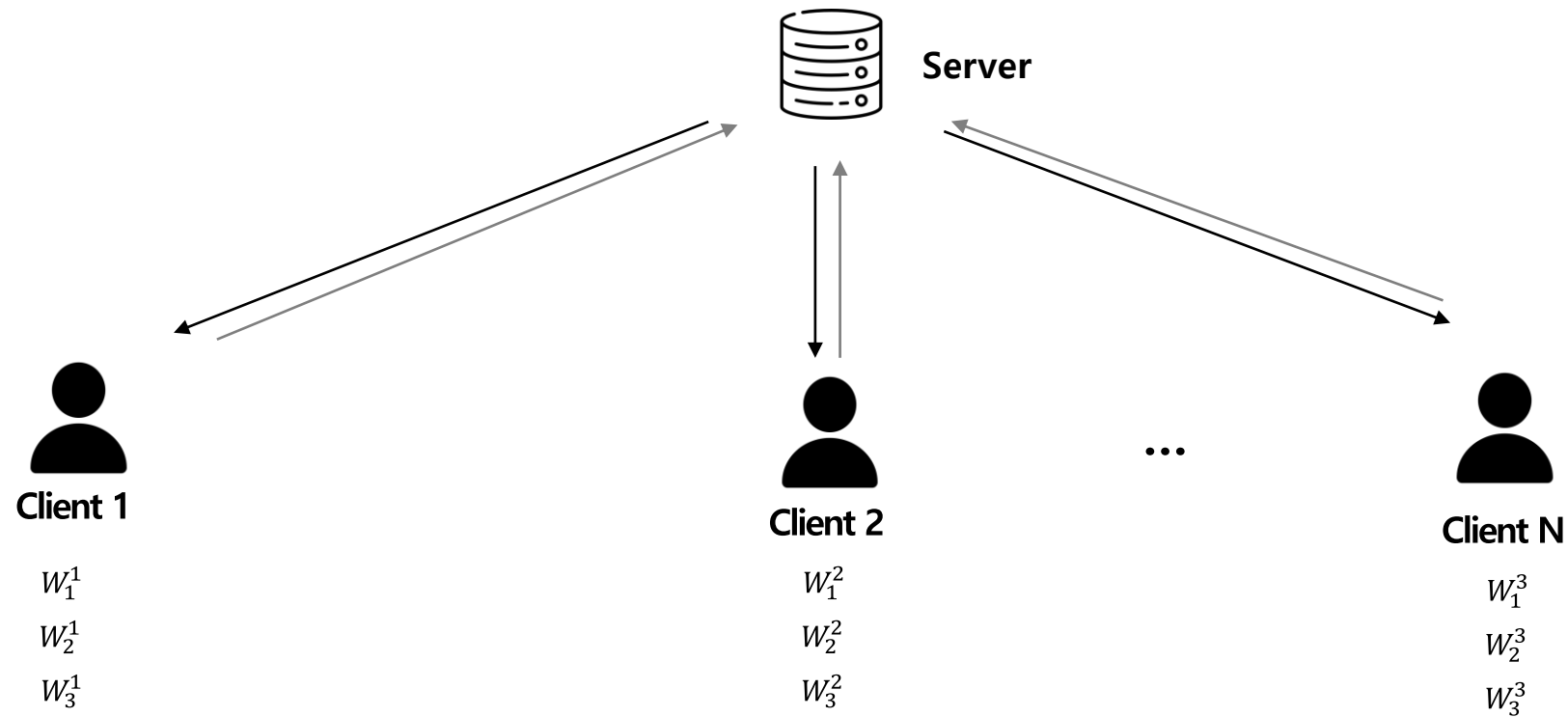


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

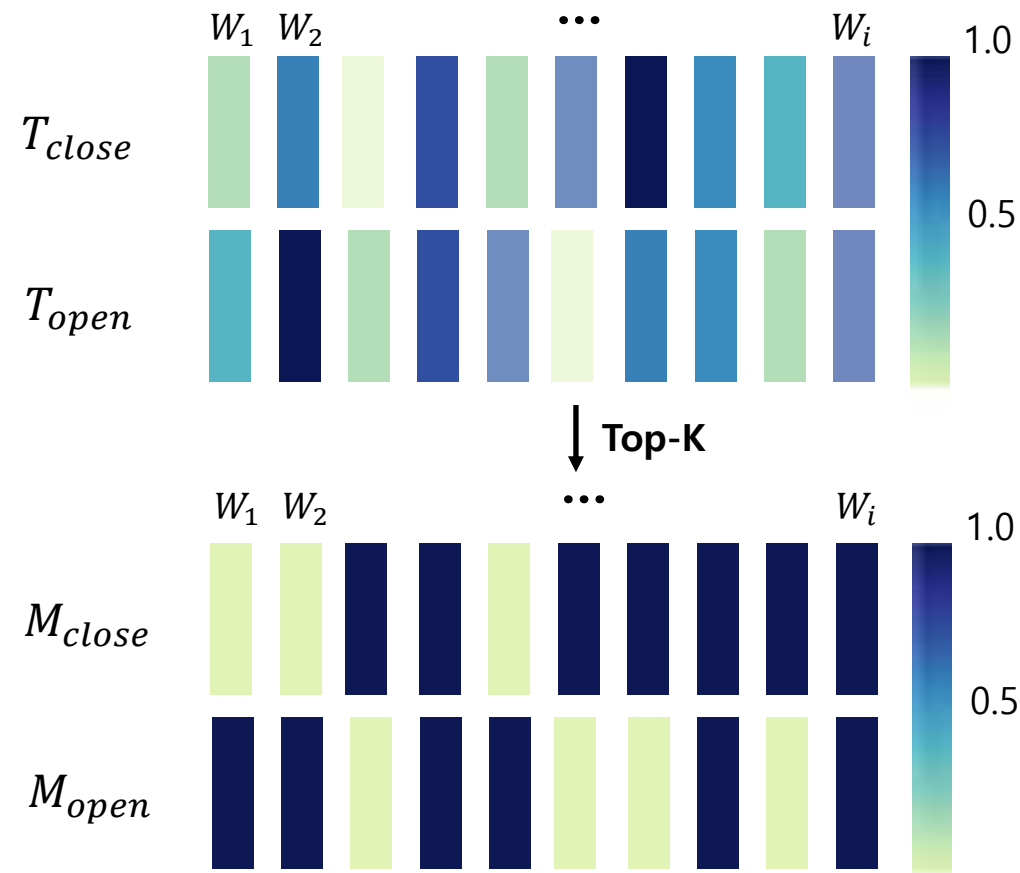
closed-set과 open-set에 관여하는 파라미터는 분리되어있다



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD closed-set과 open-set에 관여하는 파라미터를 분리하자

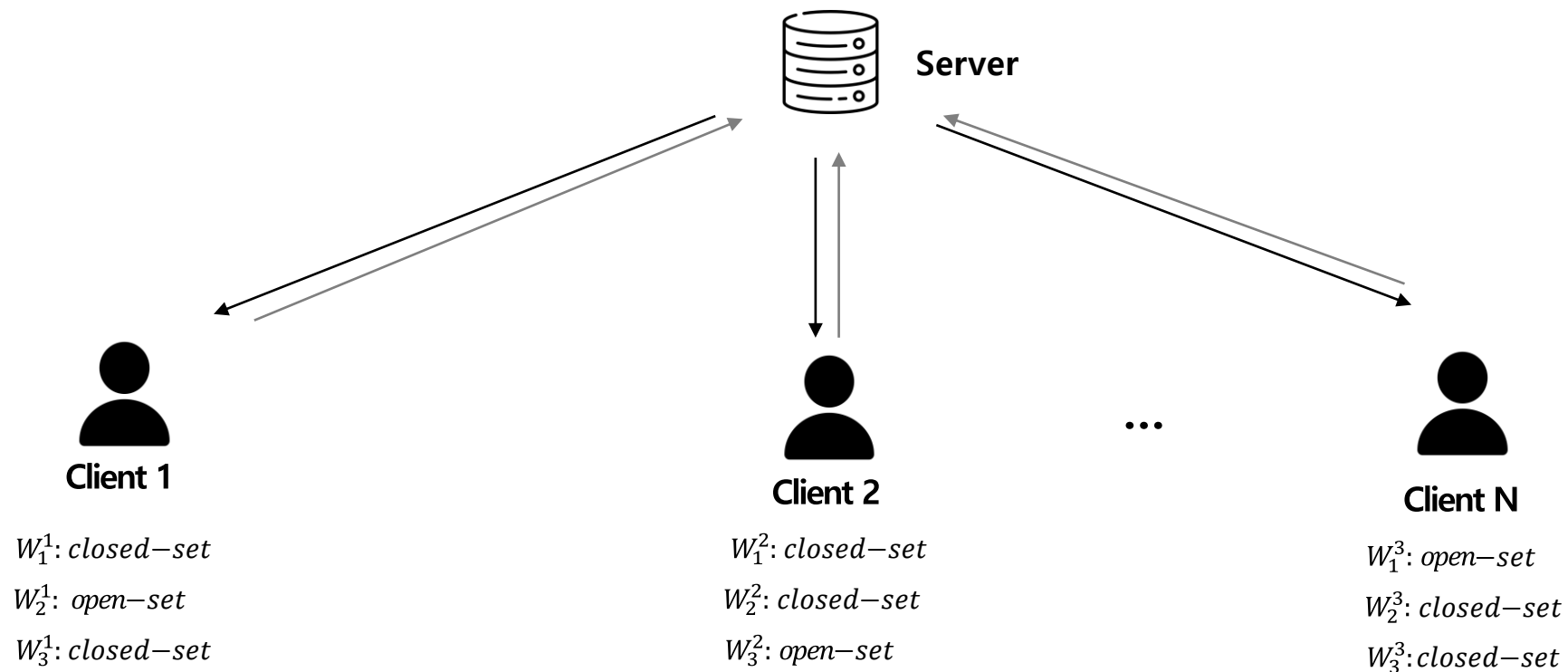


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

closed-set과 open-set에 관여하는 파라미터는 분리되어있다

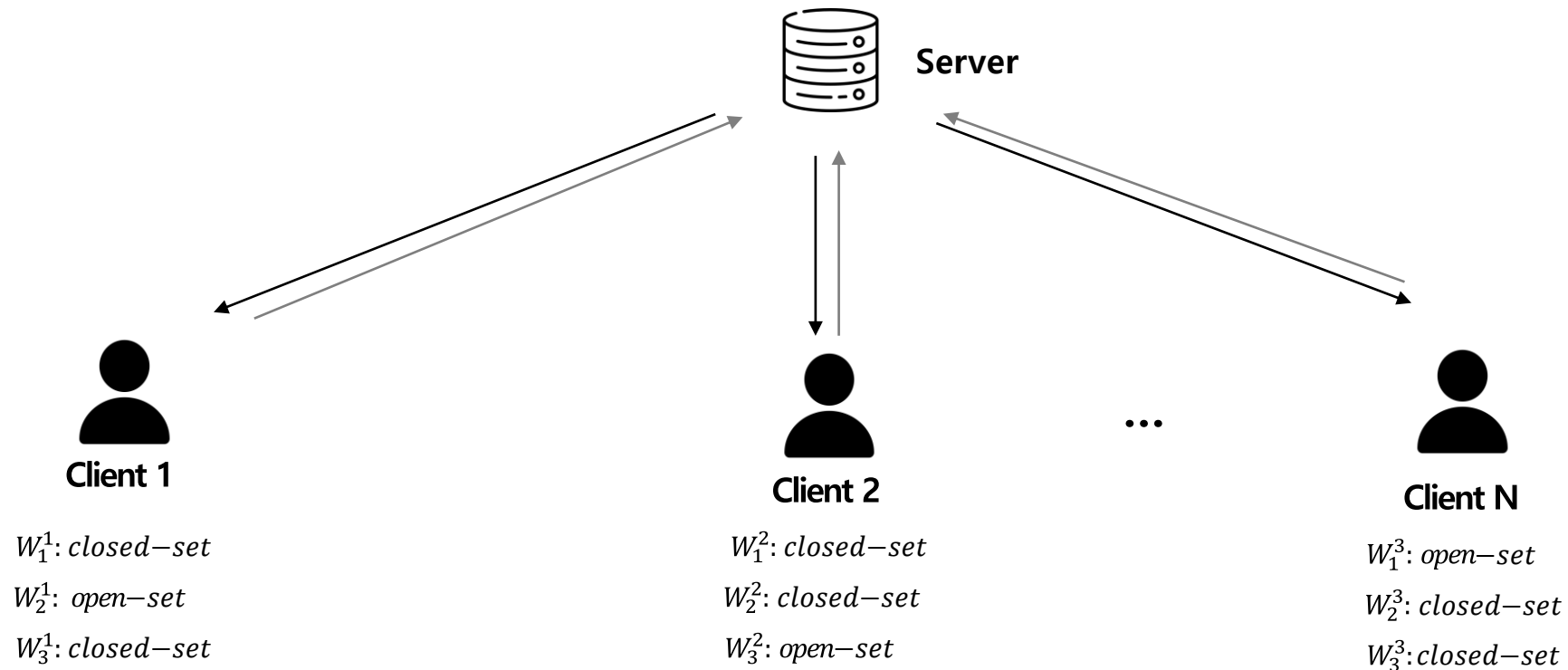


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자

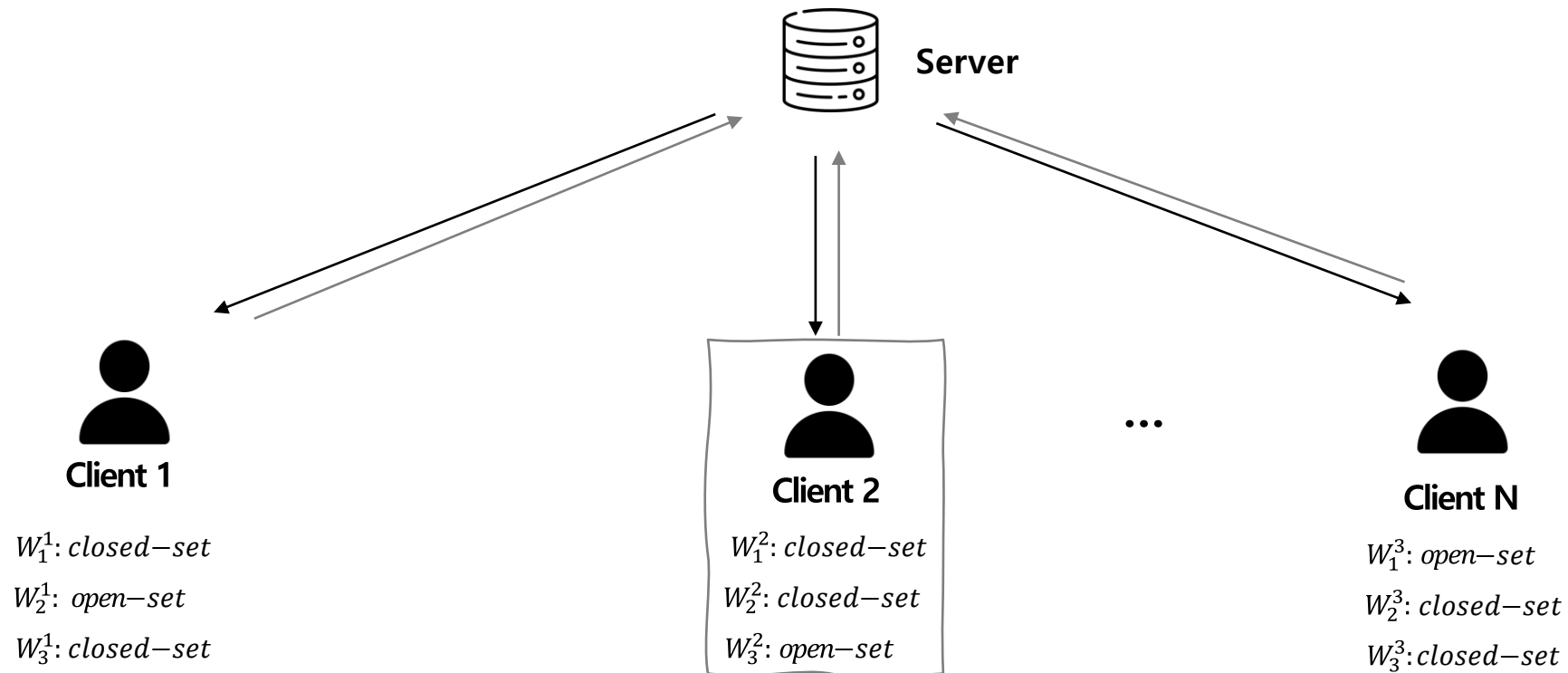


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자

$$\begin{aligned}\tilde{S}_{close}^k &= OT(S^k \odot P_{close}^k, S^2 \odot P_{close}^2) \\ \tilde{S}_{open}^k &= OT(S^k \odot P_{open}^k, S^2 \odot P_{open}^2) \\ \tilde{S}_{share}^k &= OT(S^k \odot P_{share}^k, S^2 \odot P_{share}^2)\end{aligned}$$

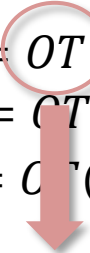


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자

$$\begin{aligned}\tilde{S}_{close}^k &= OT(S^k \odot P_{close}^k, S^2 \odot P_{close}^2) \\ \tilde{S}_{open}^k &= OT(S^k \odot P_{open}^k, S^2 \odot P_{open}^2) \\ \tilde{S}_{share}^k &= OT(S^k \odot P_{share}^k, S^2 \odot P_{share}^2)\end{aligned}$$




2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자

$$\begin{aligned}\tilde{S}_{close}^k &= OT(S^k \odot P_{close}^k, S^2 \odot P_{close}^2) \\ \tilde{S}_{open}^k &= OT(S^k \odot P_{open}^k, S^2 \odot P_{open}^2) \\ \tilde{S}_{share}^k &= OT(S^k \odot P_{share}^k, S^2 \odot P_{share}^2)\end{aligned}$$

의미적으로 가장 가까운 대응 관계

✓ 2번 client의 closed-set 파라미터를 기준으로
k번째 클라이언트의 closed-set 파라미터 중 가장 의미적으로 가까운 것은 무엇인가

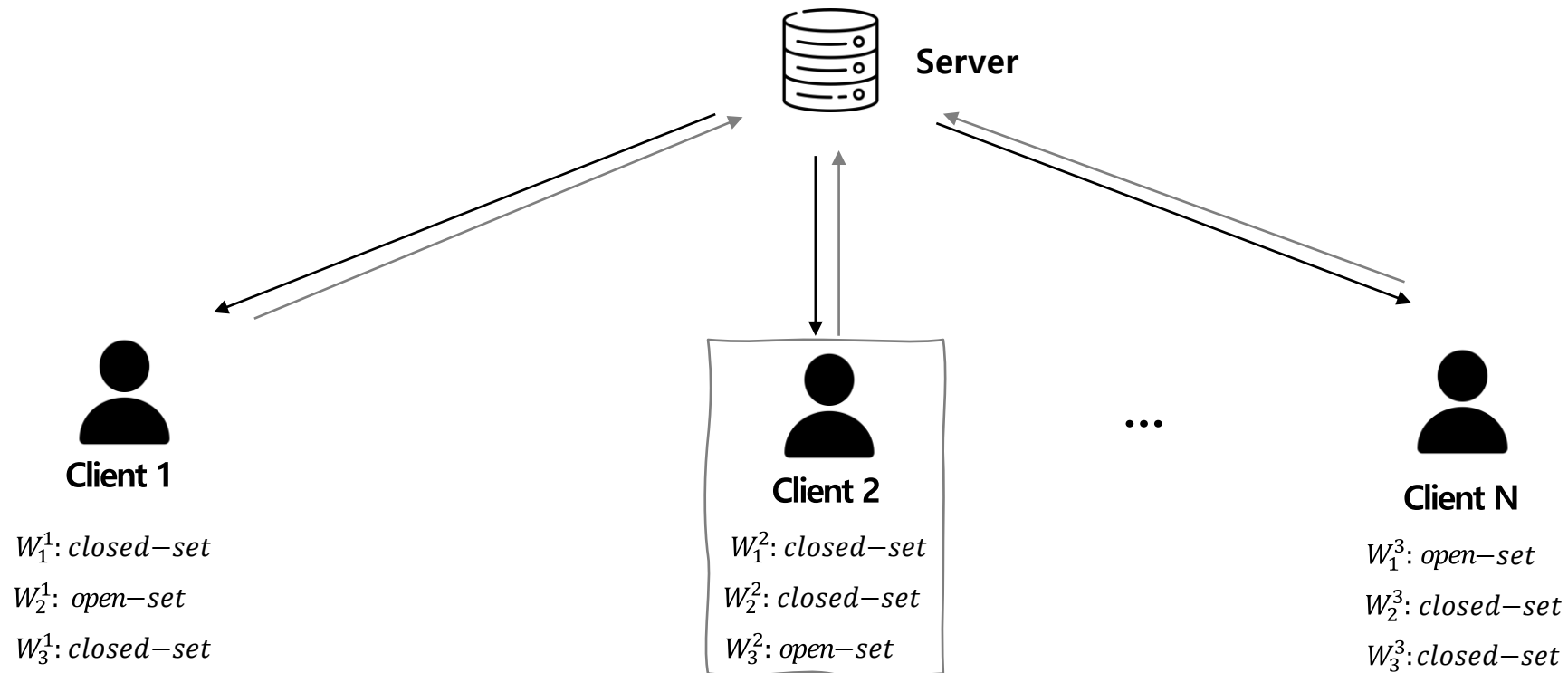


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자

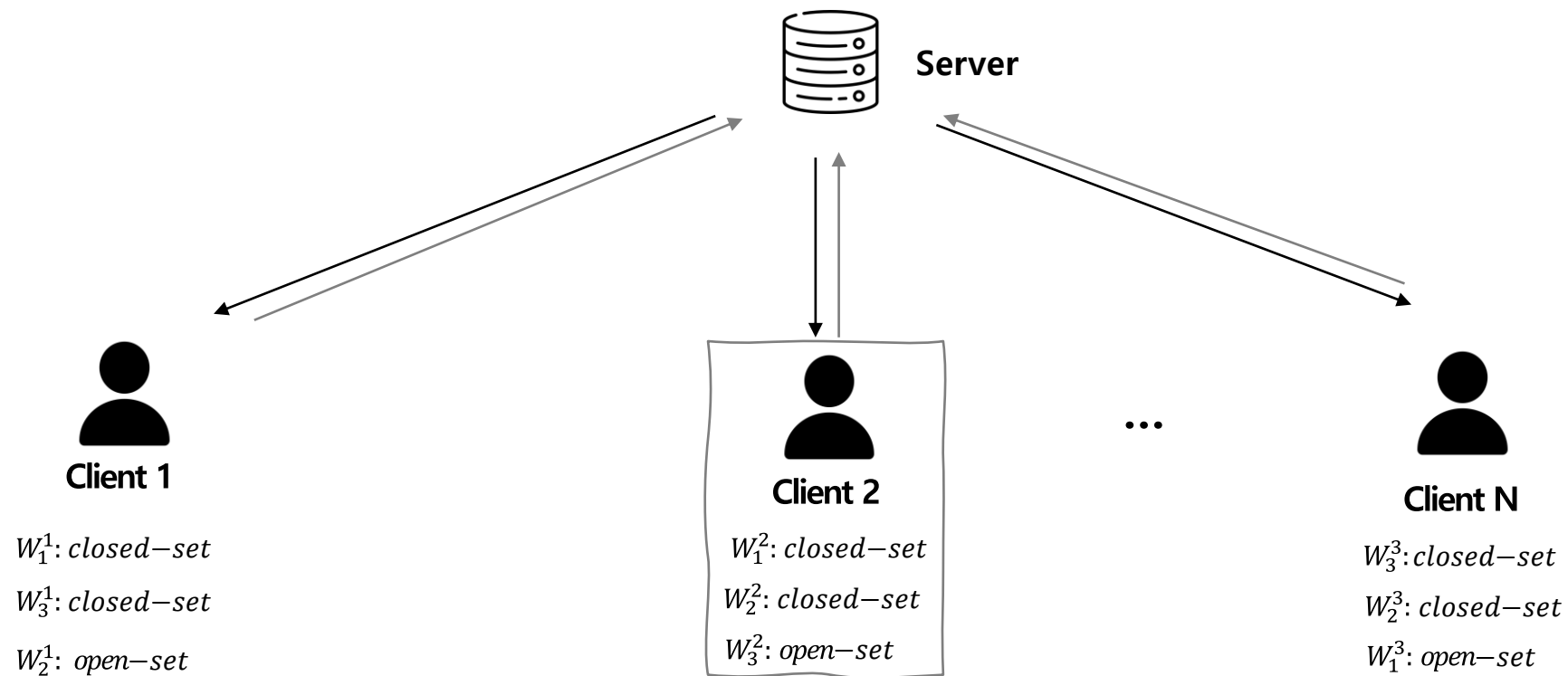


2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

FedPD

클라이언트 간 의미가 같은 파라미터를 정렬해 올바르게 집계하자



2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

Experiments & Results

- 데이터셋: Handwritten digital recognition FL Dataset (HDR-FL)
 - ✓ MNIST, SVHN, USPS, SynthDigits and MNIST-M
- 정량 평가:
 - ✓ Closed-set classification: Accuracy
 - ✓ Open-set detection: AUROC
- Known class의 개수: 6 / Unknown class의 개수: 4

Methods	Closed-set ACC						Open-set AUC					
	MNIST	SVHN	USPS	Synth	MNIST-M	Avg	MNIST	SVHN	USPS	Synth	MNIST-M	Avg
SoftMax	96.69	76.06	97.97	83.99	83.69	87.68	77.38	65.41	88.80	70.53	66.24	73.67
OpenMax[2] _(CVPR'16)	95.54	63.13	97.97	81.15	76.57	82.87	77.78	57.37	89.21	69.78	60.88	71.00
RPL[4] _(ECCV'20)	98.33	77.22	98.13	84.76	86.32	88.95	77.87	66.70	89.53	73.17	73.23	76.10
PROSER[43] _(CVPR'21)	98.04	75.81	97.20	86.92	84.93	88.58	83.63	65.57	90.28	71.76	69.75	76.20
ARPL[3] _(TPAMI'21)	97.17	69.83	96.44	84.64	84.22	86.46	85.65	59.79	92.53	69.70	68.17	75.16
DIAS[26] _(ECCV'22)	97.50	70.66	97.62	86.11	86.81	87.74	82.90	67.33	90.69	77.48	71.92	78.56
SSB[35] _(ICLR'22)	96.41	60.12	97.46	82.82	74.86	82.33	88.53	57.68	90.45	73.40	70.01	76.61
FedPD_(Ours)	98.73	78.06	98.56	89.32	90.14	90.96	90.98	69.46	93.31	79.43	73.64	81.36

2. FedPD

Federated Open Set Recognition with Parameter Disentanglement

Experiments & Results

- 데이터셋: CIFAR-10
- 정량 평가:
 - ✓ Closed-set classification: Accuracy
 - ✓ Open-set detection: AUROC
- Unknown class의 개수: 4개, 6개, 8개로 실험

Method	Known=4		Known=6		Known=8	
	Closed-set ACC	Open-set AUC	Closed-set ACC	Open-set AUC	Closed-set ACC	Open-set AUC
SoftMax	83.23 $\pm$ 0.28	65.70 $\pm$ 0.14	83.14 $\pm$ 0.38	72.00 $\pm$ 0.44	72.17 $\pm$ 0.41	61.11 $\pm$ 0.45
OpenMax[2] (CVPR'16)	83.24 $\pm$ 0.08	65.96 $\pm$ 0.14	84.56 $\pm$ 0.24	81.69 $\pm$ 0.59	72.72 $\pm$ 0.40	61.52 $\pm$ 0.38
RPL[4] (ECCV'20)	81.23 $\pm$ 0.11	65.10 $\pm$ 0.05	78.78 $\pm$ 0.43	68.72 $\pm$ 0.43	72.98 $\pm$ 0.33	61.36 $\pm$ 0.32
PROSER[43] (CVPR'21)	84.15 $\pm$ 0.13	69.04 $\pm$ 0.07	85.77 $\pm$ 0.48	80.69 $\pm$ 0.27	70.50 $\pm$ 0.21	60.48 $\pm$ 0.41
ARPL[3] (TPAMI'21)	83.91 $\pm$ 0.09	69.02 $\pm$ 0.12	86.54 $\pm$ 0.53	79.83 $\pm$ 0.69	73.63 $\pm$ 0.37	66.78 $\pm$ 0.64
DIAS[26] (ECCV'22)	84.85 $\pm$ 0.04	70.32 $\pm$ 0.09	87.74 $\pm$ 0.16	81.66 $\pm$ 0.25	74.53 $\pm$ 0.37	67.75 $\pm$ 0.34
SSB[35] (ICLR'22)	84.27 $\pm$ 0.06	68.85 $\pm$ 0.11	86.04 $\pm$ 0.70	82.41 $\pm$ 0.33	75.54 $\pm$ 0.18	67.97 $\pm$ 0.23
FedPD (Ours)	86.28 $\pm$ 0.07	71.50 $\pm$ 0.07	89.43 $\pm$ 0.22	85.07 $\pm$ 0.50	75.87 $\pm$ 0.16	69.12 $\pm$ 0.45



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

FedOSS

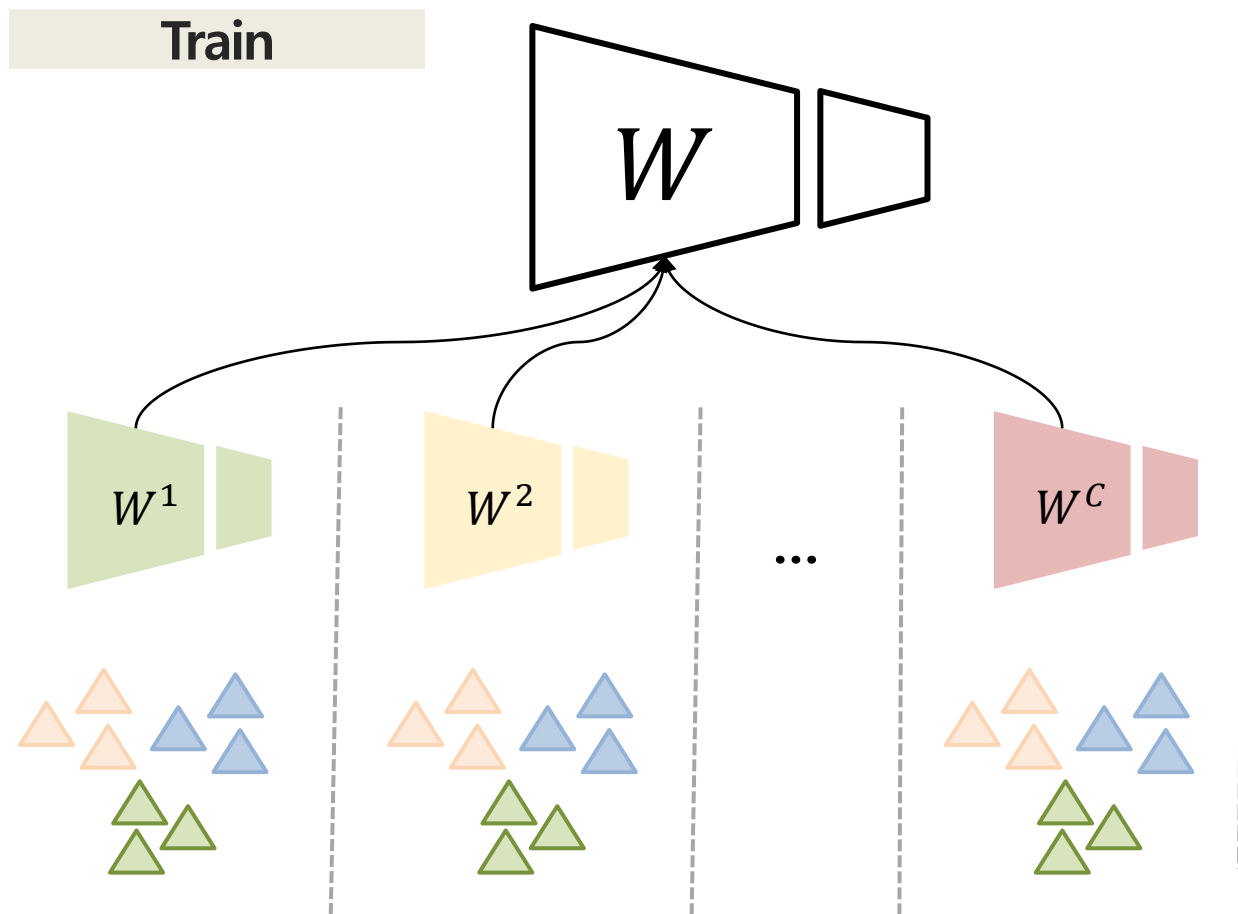


3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS

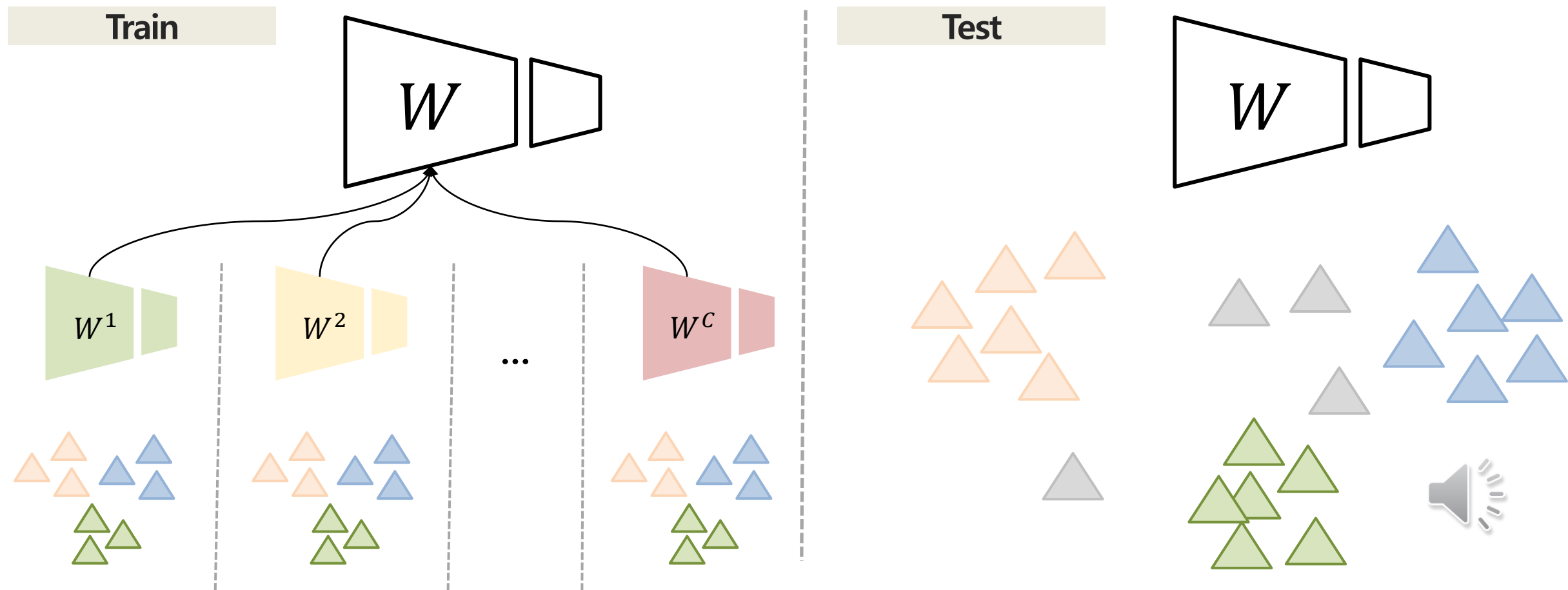


3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS

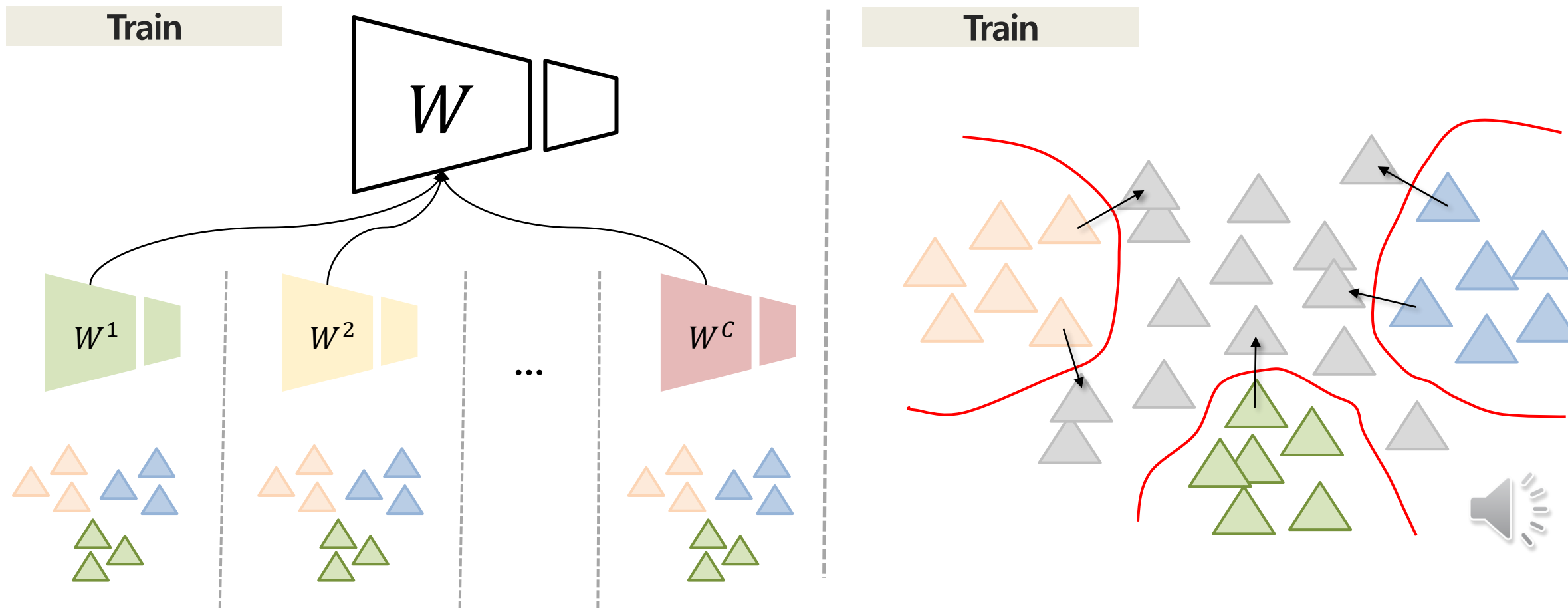


3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS

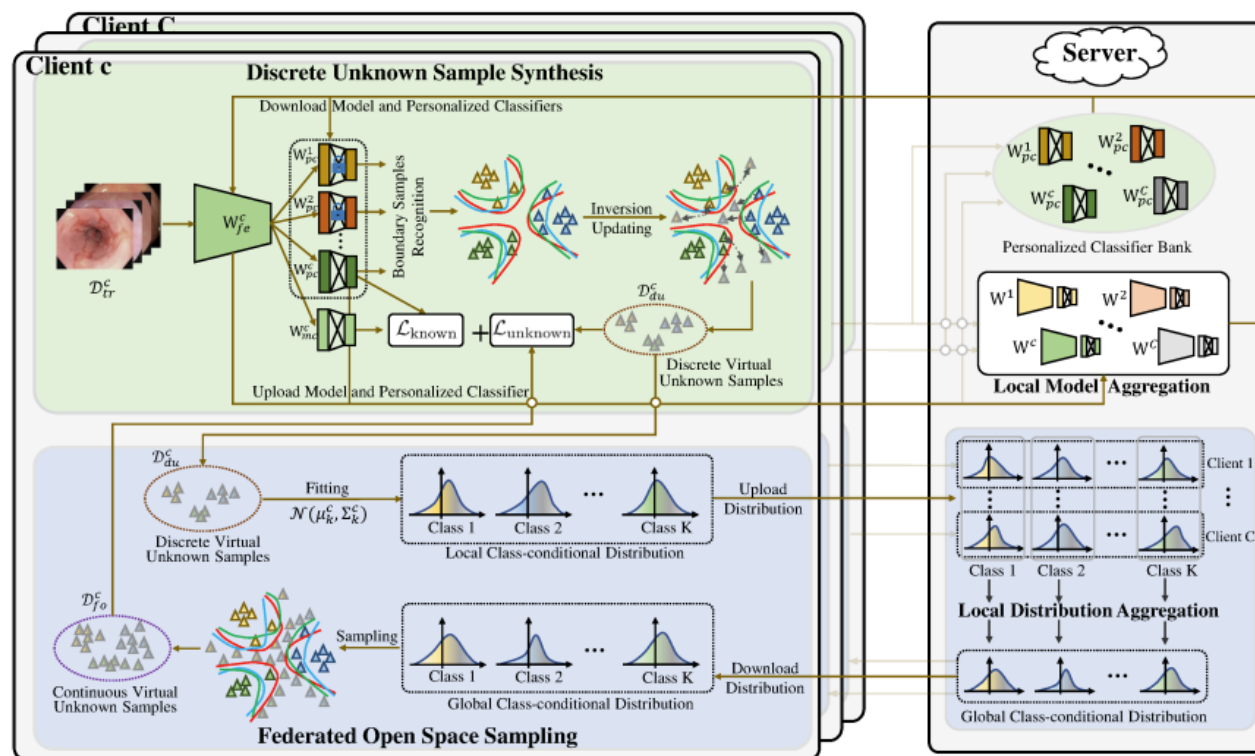


Fig. 2. The overview of the proposed FedOSS framework to federated open set recognition (Best view in color). FedOSS overcomes the challenge in unavailability of open data via discrete unknown sample synthesis (DUSS) and federated open space sampling (FOSS). DUSS utilizes the client discrepancy to recognize boundary samples, which are used to synthesize unknown samples via inversion updating. The generated unknown samples from all clients are inputted into FOSS to estimate the distribution of open space for sampling diverse unknown data.



3. FedOSS

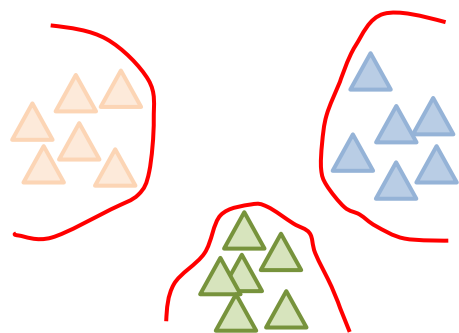
Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

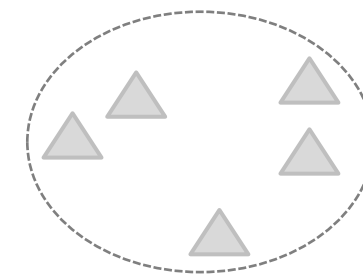
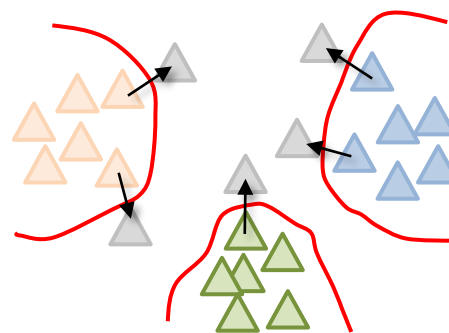
FedOSS

DUSS

Local



Inversion updating



Discrete Virtual
Unknown Samples

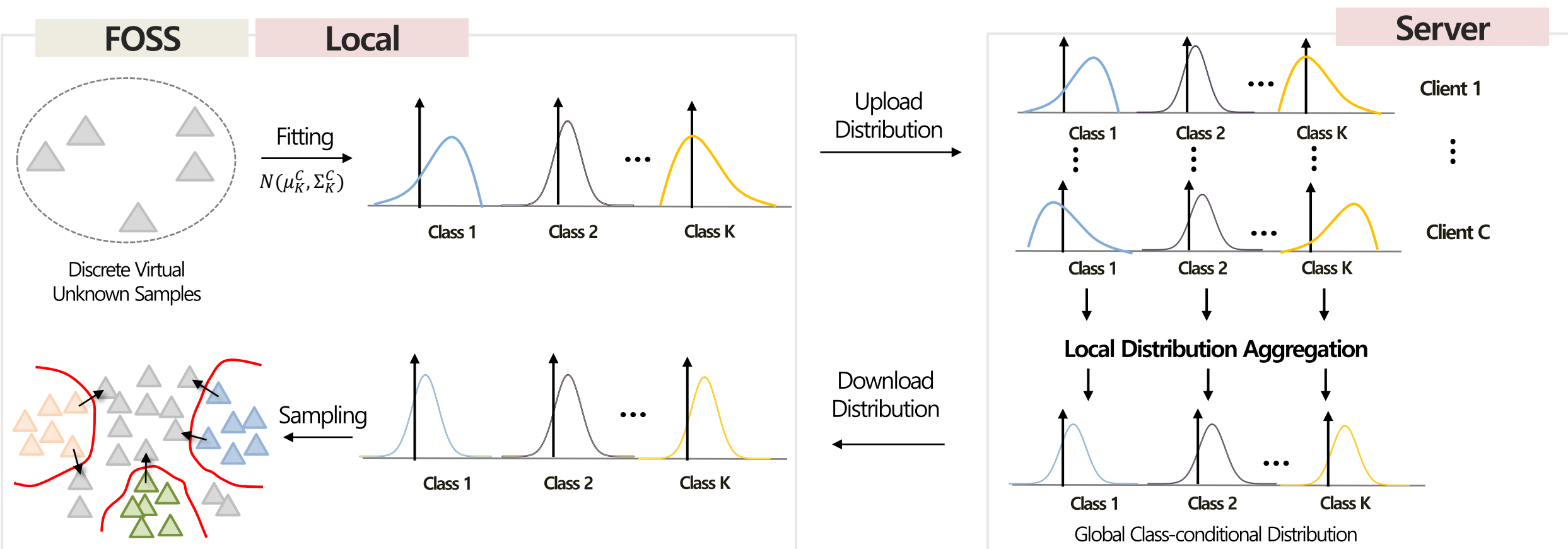


3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

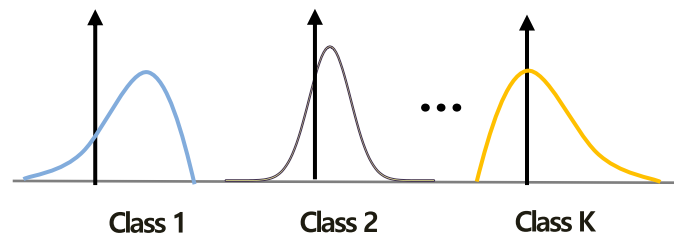
△ △ △ : Known classes
△ △ △ : Unknown class

FedOSS

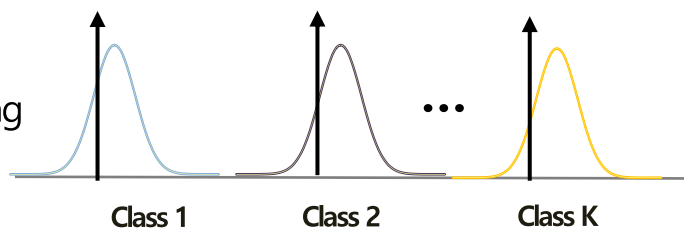
FOSS

Local

Fitting
 $N(\mu_K^C, \Sigma_K^C)$



Sampling



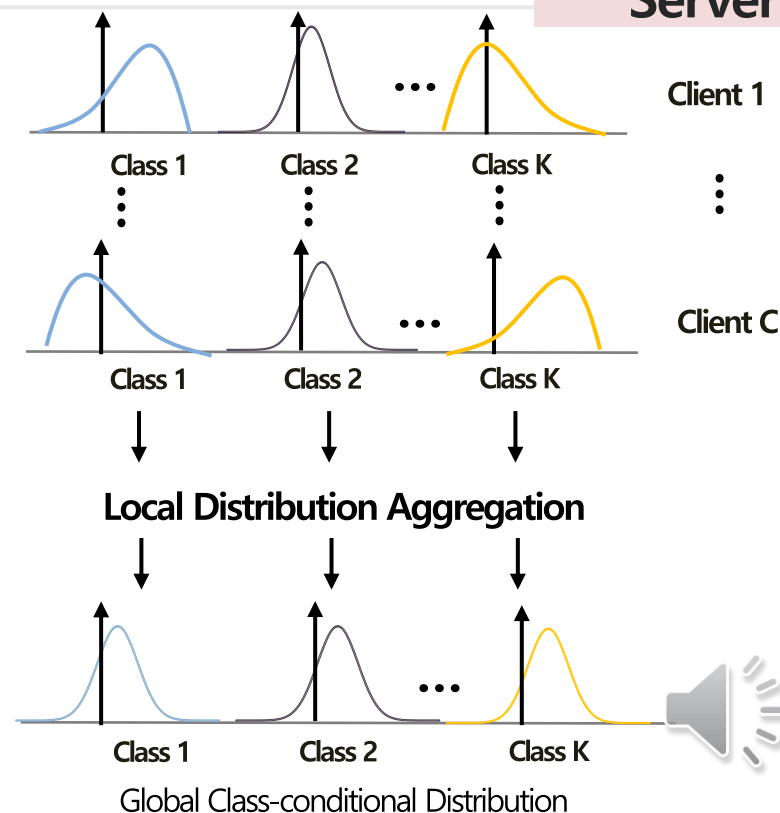
Discrete Virtual Unknown Samples

Continuous Virtual Unknown Samples

Upload Distribution

Download Distribution

Server



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

Experiments & Results

- 데이터셋: Microscopic Peripheral Blood Cell Dataset(17,092 데이터, 8 classes)
- 정량 평가: ACC(%)/ Recall(%)/ Precision(%)/ F1-score(%)
- Unknown class의 개수: 3개, 5개로 실험

Methods	U = 3				U = 5			
	ACC (%)	Recall (%)	Precision (%)	F1-score (%)	ACC (%)	Recall (%)	Precision (%)	F1-score (%)
Softmax	70.42±3.52	81.03±1.19	72.09±2.76	59.56±4.07	44.20±2.11	74.48±0.51	62.48±5.17	48.78±1.99
OpenMax [4]	73.96±7.32	76.72±4.11	74.98±6.35	72.52±6.70	65.78±3.72	71.88±3.63	72.05±2.13	65.04±1.91
CPN [5]	74.72±5.58	82.18±4.18	75.12±4.43	74.81±5.41	70.89±2.01	83.17±1.70	73.77±2.25	73.07±1.10
CAC [6]	75.50±6.42	82.67±4.54	75.52±5.94	75.75±6.23	74.85±2.54	<u>85.86±1.25</u>	74.86±1.58	75.54±1.21
PROSER [7]	77.48±3.39	82.99±4.21	77.56±2.26	77.64±2.99	61.01±7.36	<u>81.42±1.97</u>	69.29±2.19	64.74±4.38
FedMix [60]	72.18±4.83	73.90±7.12	72.82±5.91	71.92±7.50	57.26±10.56	74.74±7.10	65.87±3.64	61.97±6.54
SSB [8]	<u>78.71±7.92</u>	<u>83.38±4.73</u>	<u>79.46±6.56</u>	<u>79.05±7.19</u>	73.46±10.86	84.41±4.97	73.92±6.25	74.28±9.04
OVNRN [9]	77.35±8.02	83.45±5.54	78.87±5.36	77.65±7.50	<u>77.95±4.16</u>	87.49±1.31	<u>76.20±2.08</u>	<u>78.00±2.90</u>
Ours	80.04±5.38	82.46±5.31	80.73±4.02	81.06±4.59	82.91±4.27	84.54±3.80	79.97±1.96	80.56±3.58



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

Experiments & Results

- 데이터셋: Gastrointestinal Endoscopy Dataset(10,662 데이터, 15 classes)
- 정량 평가: ACC(%)/ Recall(%)/ Precision(%)/ F1-score(%)
- Unknown class의 개수: 3개, 9개로 실험

Methods	U = 3				U = 9			
	ACC (%)	Recall (%)	Precision (%)	F1-score (%)	ACC (%)	Recall (%)	Precision (%)	F1-score (%)
Softmax	77.41±8.19	84.75±0.37	81.64±4.39	76.60±3.75	60.62±3.67	84.43±0.96	66.81±2.50	65.86±1.46
OpenMax [4]	80.27±5.41	81.80±7.74	84.91±1.87	79.79±7.80	65.99±1.88	82.09±3.24	66.89±7.96	69.80±3.39
CPN [5]	80.35±5.19	84.36±4.86	80.30±7.90	80.25±7.06	61.03±9.31	82.21±3.47	64.81±8.68	67.57±7.55
CAC [6]	83.34±4.65	87.52±1.73	84.67±2.59	84.67±2.60	64.73±6.51	83.51±1.11	68.37±4.40	70.11±4.82
PROSER [7]	<u>85.70±2.95</u>	89.05±0.67	<u>86.28±2.65</u>	<u>85.56±1.25</u>	69.29±3.05	85.66±2.79	72.44±2.08	73.16±1.23
FedMix [60]	83.62±3.27	86.27±2.18	80.86±5.71	82.52±4.23	69.92±3.38	82.85±1.46	72.76±1.71	73.33±3.32
SSB [8]	80.89±4.74	84.92±3.13	83.83±2.53	81.89±3.48	66.45±5.59	83.55±3.29	71.70±2.93	72.06±4.73
OVNR [9]	78.49±9.29	84.47±2.63	83.09±2.95	81.58±4.86	<u>71.45±7.78</u>	86.72±2.86	<u>74.84±3.84</u>	<u>76.06±5.35</u>
Ours	88.29±1.80	<u>88.29±3.07</u>	89.07±1.46	87.78±3.23	77.17±2.05	<u>86.23±2.01</u>	77.89±0.94	79.41±2.46



3. FedOSS

Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

Experiments & Results

- 데이터셋: 3D Organ Classification Dataset(1743 데이터, 11 classes)
- 정량 평가: ACC(%)/ Recall(%)/ Precision(%)/ F1-score(%)
- Unknown class의 개수: 4개, 7개로 실험

Methods	U = 4				U = 7			
	ACC (%)	Recall (%)	Precision (%)	F1-score (%)	ACC (%)	Recall (%)	Precision (%)	F1-score (%)
Softmax	67.72±1.29	82.20±2.81	71.45±1.47	72.27±1.73	44.89±3.24	75.57±1.96	56.82±0.62	52.79±2.29
OpenMax [4]	49.95±3.19	56.94±9.58	52.36±1.72	49.44±6.66	66.38±4.80	54.04±6.62	68.20±10.92	52.45±5.71
CPN [5]	59.12±3.12	74.25±1.84	65.84±3.83	65.20±1.70	40.40±7.13	63.81±6.59	51.54±13.07	48.43±6.95
CAC [6]	71.92±5.36	82.90±3.91	75.34±6.58	76.38±5.59	44.40±5.37	73.45±2.80	53.72±3.71	54.99±2.66
PROSER [7]	78.22±1.64	86.76±2.53	79.96±0.89	81.92±1.11	51.77±11.59	69.19±10.41	64.51±7.98	56.60±9.92
FedMix [60]	77.75±2.89	87.06±3.09	80.78±1.86	81.79±2.86	43.84±2.97	72.45±3.52	55.38±6.49	53.05±4.74
SSB [8]	67.81±3.98	81.51±2.04	75.21±5.45	74.19±4.19	46.23±7.03	70.88±5.60	58.16±5.33	52.66±6.69
OVRN [9]	64.18±1.30	78.59±2.64	73.77±4.56	71.17±2.10	57.69±3.52	79.82±2.09	61.59±4.30	62.55±4.25
Ours	80.42±2.92	81.77±1.81	85.36±3.87	82.84±2.28	78.32±3.41	75.44±4.02	80.86±1.59	75.81±3.30



Conclusion

Open Set Recognition Federated Learning

- ✓ FedOSR의 초기 아이디어
- ✓ GAN을 활용해 unknown 샘플을 합성

- ✓ 경계부 샘플을 밀어 unknown 합성
- ✓ 클라이언트들이 생성한 unknown을 연합 샘플링

FedOS [arXiv 2022]

FedOS: using open-set learning to stabilize training in federated learning

Mohamad Mohamad
mohamad_m.28@live.com

Julian Neubert
julian.neubert@mx.de

Juan Segundo Argayo
jargayo@gmail.com

Abstract

Federated Learning is a recent approach to train statistical models on distributed datasets without violating privacy constraints. The data locality principle is preserved by sharing the model instead of the data between clients and the server. This brings many advantages but also poses new challenges. In this report, we explore this new research area and perform several experiments to deepen our understanding of what these challenges are and how different problem settings affect the performance of the final model. Finally, we present a novel approach to one of these challenges and compare it to other methods found in literature. The code is available on our github¹.

1. Introduction

Nowadays, phones are the primary and sometimes the only computing device for many people. As such, the interaction with them and their powerful sensors leave an unprecedented amount of rich data, much of it private in nature. Cameras, microphones, GPS, and tactile sensors are all examples of this. In the field of Big Data and Machine Learning applications, this data regains importance as models learned with it hold the promise of greatly improving

ditional distributed optimization and requires dealing with heterogeneous data

2. Related Work

FedProx. There are two key challenges in Federated Learning that differentiate it from traditional distributed optimization: (1) significant variability in terms of the system's characteristics on each device in the network (system heterogeneity), and (2) non-identically distributed data across the network (statistical heterogeneity). FedProx (Tian Li et al. [12]) can be viewed as a generalization and re-parameterization of FedAvg. In this work, the authors have theoretically proven convergence guarantees when learning over data from non-identical distributions (statistical heterogeneity), while also adhering to the constraints each device-level system has by allowing each participating device to perform a variable amount of work (system heterogeneity). They demonstrate that in highly heterogeneous settings, FedProx has significantly more stable and accurate convergence behavior relative to FedAvg, improving absolute test accuracy by 22% on average.

FedIR. As well as [12], the authors of this paper (Hua et al. [6]) recognized that one of the challenges in terms of data diversity relies on the fact that data at the source is far from independent and identically distributed (IID). In

FedPD [ICCV 2022]

FedPD: Federated Open Set Recognition with Parameter Disentanglement

Chen Yang¹ Meihu Zhu¹ Yifan Liu² Yixuan Yuan^{1,2*}
¹City University of Hong Kong ²The Chinese University of Hong Kong
{cyang53, meihuzhu}@cityu.edu.hk {115519505081@hk.cuhk.edu.hk, yxyuan@cuhk.edu.hk}

Abstract

Existing federated learning (FL) approaches are deployed under the unrealistic closed-set setting, with both training and testing classes belong to the same set, which makes the global model fail to identify the unseen classes as 'unknown'. To this end, we aim to study a novel problem of federated open-set recognition (FedOSR), which learns an open-set recognition (OSR) model under federated paradigm such that it classifies seen classes while at the same time detects unknown classes. In this work, we propose a parameter disentanglement guided federated open-set recognition (FedPD) algorithm to address two core challenges of FedOSR: cross-client inter-set interference between learning closed-set and open-set knowledge and cross-client inter-set inconsistency by data heterogeneity. The proposed FedPD framework mainly leverages two modules, i.e., local parameter disentanglement (LPD) and global divide-and-conquer aggregation (GDCA), to first disentangle client OSR model into different subnetworks, then align the corresponding parts across clients for matched model aggregation. Specifically, on the client side, LPD decomposes an OSR model into a closed-set subnetwork and an open-set subnetwork by the task-related importance, thus preventing inter-set interference. On the server side, GDCA first partitions the two subnetworks into specific and shared parts, and subsequently aligns the corresponding parts through optimal transport to eliminate parameter misalignment. Extensive experiments on various datasets demonstrate the superior performance of our proposed method.

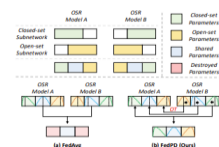


Figure 1. Parameter disentanglement on FedOSR models and comparison of various aggregation strategies on FedOSR setting. (a) FedAvg simply aggregates multiple OSR models, leading to model collapse. (b) Our FedPD first aligns corresponding close-specific, open-specific and shared parts by optimal transport (OT), then aggregates aligned parameters. The curves inside the network box represent different parameter distribution of each party.

tions and difficult to integrate into a centralized dataset, owing to increasing privacy and ethical concerns, especially for those sensitive data such as location-based services or health information [2]. To break this dilemma, federated learning (FL) [2, 23, 20] provides a privacy-preserving paradigm that allows local clients to collaboratively train a shared global model without data sharing.

FedOSS [IEEE 2024]

FedOSS: Federated Open Set Recognition via Inter-Client Discrepancy and Collaboration

Meihu Zhu¹, Jing Liao¹, Member, IEEE, Jun Liu¹, Member, IEEE, and Yixuan Yuan¹, Member, IEEE

Abstract—Open set recognition (OSR) aims to accurately classify known diseases and recognize unseen diseases as the unknown class in medical scenarios. However, in existing OSR approaches, gathering data from distributed sites to construct large-scale centralized training datasets usually leads to high privacy and security risk, which could be alleviated elegantly via the popular cross-site training paradigm, federated learning (FL). To this end, we represent the first effort to formulate federated open set recognition (FedOSR), and meanwhile propose a novel Federated Open Set Synthesis (FedOSS) framework to address the core challenge of FedOSR: the unavailability of unknown samples for all anticipated clients during the training phase. The proposed FedOSS framework mainly leverages two modules, i.e., Discrete Unknown Sample Synthesis (DUS) and Federated Open Space Sampling (FOSS), to generate virtual unknown samples for learning decision boundaries between known and unknown classes. Specifically, DUS exploits inter-client knowledge inconsistency to recognize known samples near decision boundaries and then pushes them beyond decision boundaries to synthesize discrete virtual unknown samples. FOSS unites these generated unknown samples from different clients to estimate the class-conditional distributions of open data space near decision boundaries and further samples open data, thereby improving the diversity of virtual unknown samples. Additionally, we conduct comprehensive ablation experiments to verify the effectiveness of DUS and FOSS. FedOSS shows superior performance on public medical datasets in comparison with state-of-the-art approaches. The source code is available at <https://github.com/CHU-AM-Group/FedOSS>.

Index Terms—Open set recognition, federated learning, medical image classification.

achieved remarkable performance on various medical classification tasks [11], [23], [5]. However, these methods are usually evaluated in a closed-set setting, in which categories in test set are known and same as training set [4]. The closed-set setting is typically unreasonable in real-world medical scenarios, since rare or unknown diseases probably arise without any warning and would be misclassified into one of the known diseases [4], [13], [6], [17], [6], [9]. This problem poses an enormous risk to the public health and impedes the application of intelligent healthcare systems. Open-set recognition (OSR) [10], [11] is thus proposed to solve this issue, which aims to accurately classify known classes while at the same time identify open-set data as the unknown class.

Despite the impressive performance of existing OSR methods [4], [5], [6], [7], [8], [9] in open-set setting, they heavily depend on the availability of large-scale centralized datasets. Since the data in a single medical institution are usually limited, one common solution is to gather patient information from different hospitals [12]. Yet, due to growing privacy concerns or legal restrictions [13], distributed patient data are not able to be directly shared across institutions. Federated learning (FL) [14], [15] is a promising paradigm of decentralized machine learning to provide a feasible solution to this dilemma, which learns a global model by sharing the model parameters of clients (hospitals) instead of their raw data, under the orchestration of a trustworthy cloud server. Nevertheless, implementing OSR in such a decentralized FL framework is challenging and has not been investigated so far. Driven by above realistic issues, in this paper, we represent

- ✓ 클라이언트 모델 파라미터를 Closed-set / Open-set 서브네트워크로 분해
- ✓ 서버에서 Optimal Transport 기반 정렬로 부분별 집계



Thank you

