

Labeling Chaos to Learning Harmony: Federated Learning with Noisy Labels

Vasileios Tsouvalas · Aaqib Saeed · Tanir Ozcelebi · Nirvana Meratnia

Eindhoven University of Technology, The Netherlands

ACM TIST · Vol. 15, No. 2, Article 22 · 2024

01 Motivation

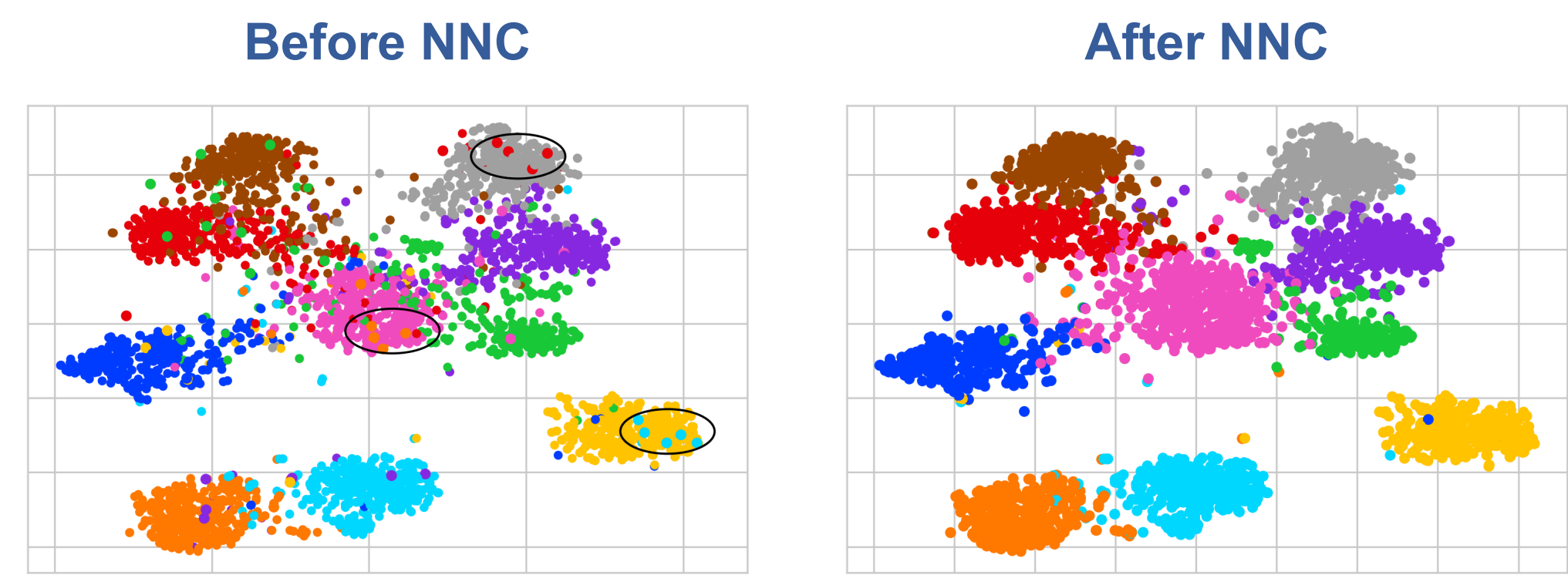
The problem

Federated learning (FL) must learn from unreliable client labels despite uneven noise, limited supervision and unequal device resources.

Contributions

- **Client-specific noise estimates** identify uneven supervision quality across devices.
- **Three intervention points:** *NNC* corrects labels at **initialization**, *AKD* guides **local training**, and *NA-FedAvg* reweights **aggregation**.
- **Broad evaluation** tests synthetic vision/audio noise and human-annotated image noise, showing accuracy gains that depend on the noise pattern and intervention.

Before and after label correction



CIFAR-10; 40% random noise. Fixed contrastive language-image pre-training (CLIP) features.

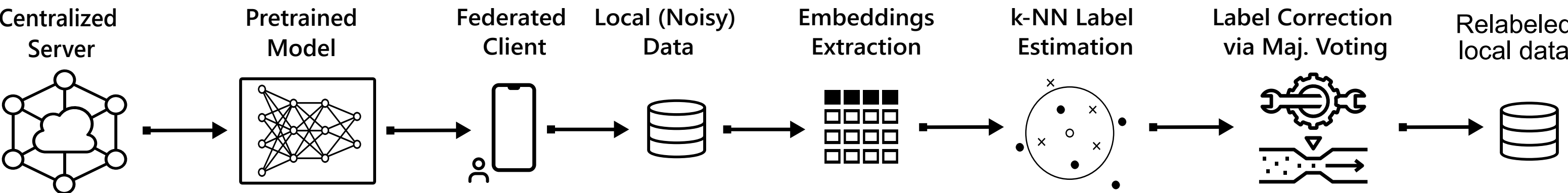
Notation

knowledge distillation (KD) · k-nearest neighbors (kNN) · confident learning (CL) · percentage points (pp) · cross-entropy (CE)

02 FedLN

Three alternative interventions address client label noise

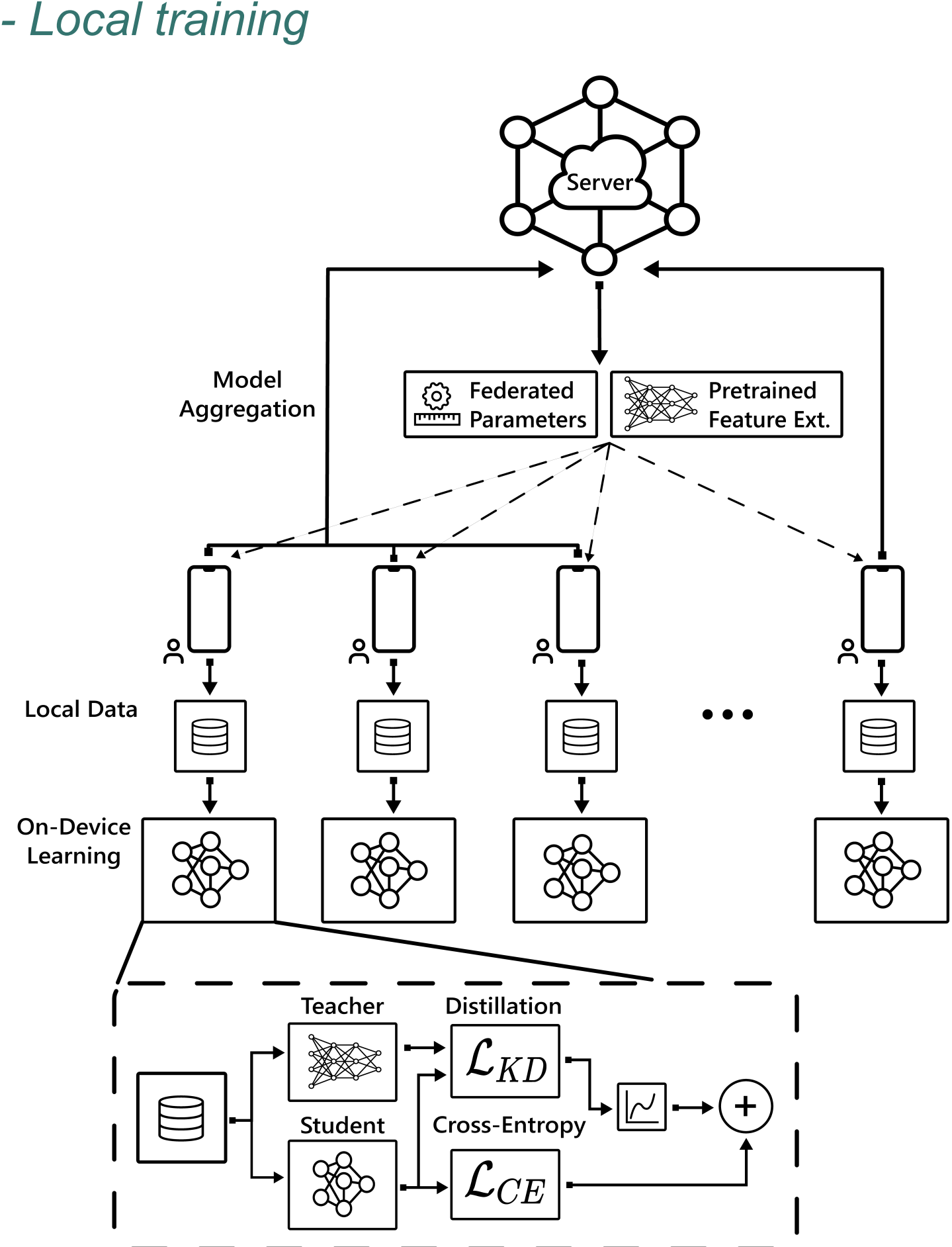
Nearest Neighbor-based Correction (NNC) - Initialization



$$\hat{y}_i = \arg \max_c \sum_{j \in \mathcal{N}_k(g(x_i))} \mathbf{1}[y_j = c]$$

g : pretrained encoder; $\mathcal{N}_k$: indices of the k nearest other local embeddings (Euclidean distance).
Replace each label by the most frequent neighbor label; break ties randomly.

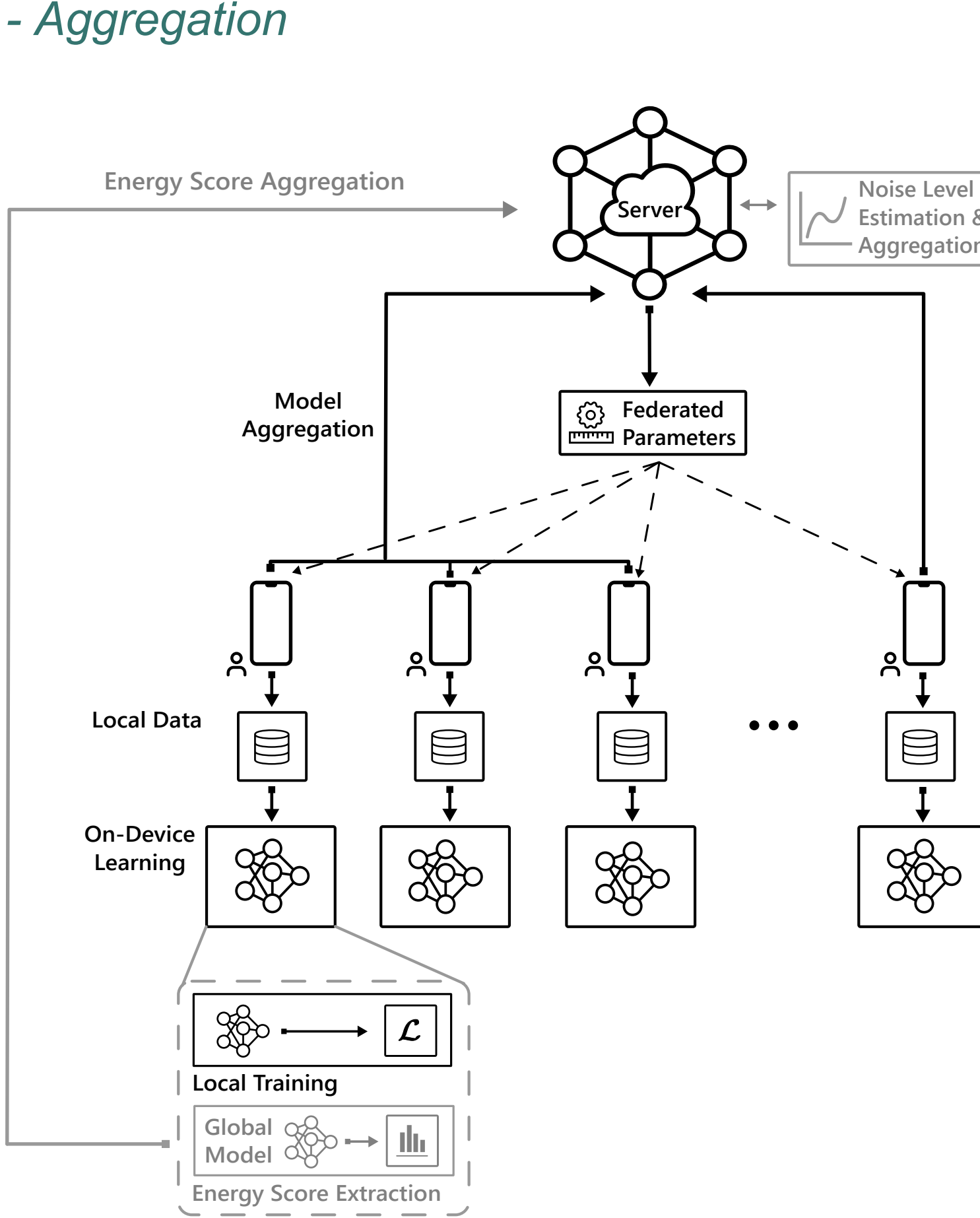
Adaptive Knowledge Distillation (AKD) - Local training



$$\mathcal{L}_{AKD} = \mathcal{L}_{CE} + \beta u_{\epsilon}(n_l^m) \mathcal{L}_{KD}$$

$\mathcal{L}_{CE}$: label loss; $\mathcal{L}_{KD}$: teacher loss.
 u_{ϵ} gates teacher loss by client noise n_l^m ;
 β sets the strength of that guidance.

Noise-aware Federated Averaging (NA-FedAvg) - Aggregation



$$\gamma_m = (1 - n_l^m) \frac{N_m}{N}$$

N_m/N : client data share; n_l^m : estimated noise.
Estimate noise once after warm-up;
reuse it to reweight subsequent updates.

03 Evaluation

+17.11 pp

NNC vs. FedAvg

CIFAR-10; 70% random noise

+10.14 pp

AKD vs. FedAvg

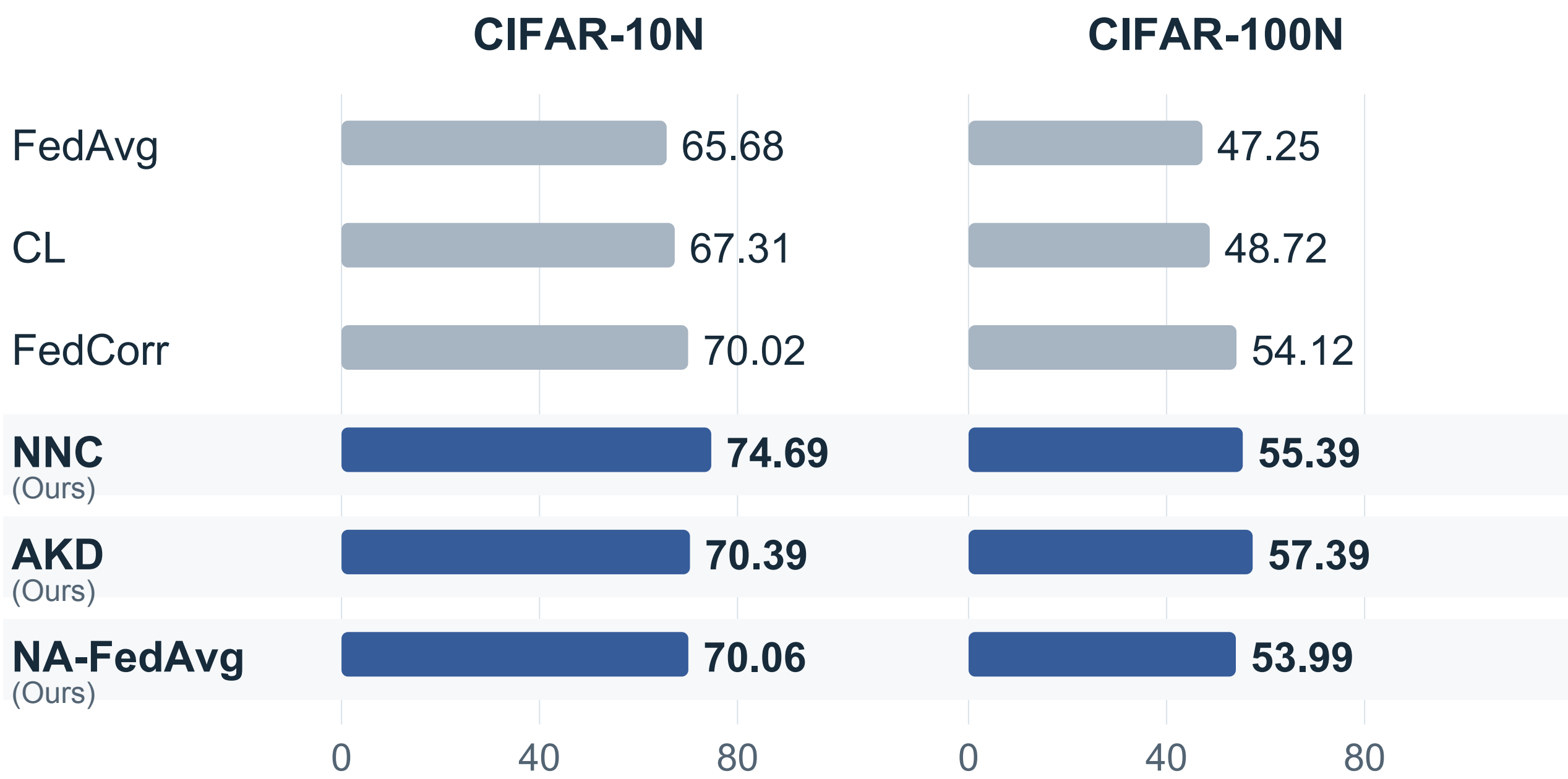
CIFAR-100N; human label noise

Accuracy depends on the noise pattern

Method	Random noise (%)	Class flipping (%)
FedAvg	57.65	63.54
FedCorr	67.27	65.23
NNC (Ours)	74.76	54.59
AKD · embeddings (Ours)	68.43	68.39
NA-FedAvg (Ours)	65.34	67.92

CIFAR-10/ResNet-20; 30 clients, 200 rounds, three-run mean accuracy. 70% label noise on 80% of clients; random versus class-flipping noise.

Accuracy with human annotation noise



Three-run mean test accuracy (%); dispersion unreported. 30 clients, 80% participation; 200/500 rounds for CIFAR-10N/100N. *AKD* uses embeddings.

Takeaway

Match the intervention to the noise: *NNC* gains 17.11 pp under random noise but falls below FedAvg under class flipping. With human annotations, **NNC leads on CIFAR-10N and AKD on CIFAR-100N** among these methods.



Paper



Code