
UNDERSTANDING THE MAIN CHALLENGES OF FEDERATED LEARNING

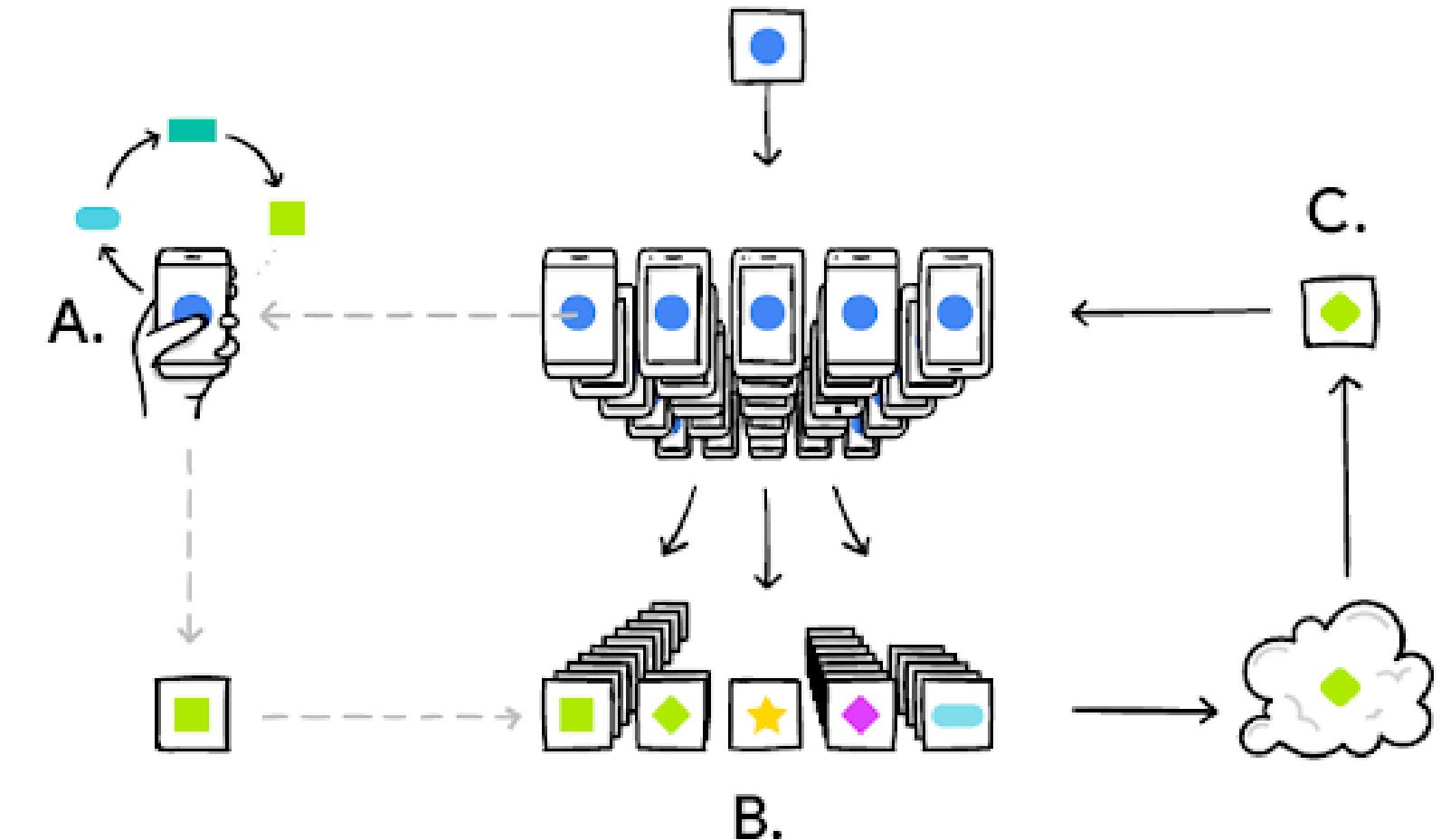


**Politecnico
di Torino**

WHAT IS FEDERATED LEARNING?

FL is a collaborative training of a model made by *multiple* parties *without* data sharing.

User devices perform most of the computation.



FEDERATED LEARNING CHALLENGES



Dealing with many devices



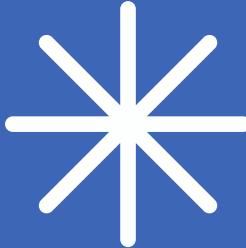
Managing stragglers



Handling non-iid data



Ensuring the privacy



Experiments setting

- **ResNet-50:** Convolutional Neural Network architecture
- **FedAVG, FairAVG, FedProx, FedGKT, CCVR:** algorithms
- **CIFAR-10:** dataset
- **Gradient Inversion Attack:** participation of a malicious user

PRELIMINARIES

- Stragglers
- Effect of non-iid data
- Data distribution
- Normalization



STRAGGLERS

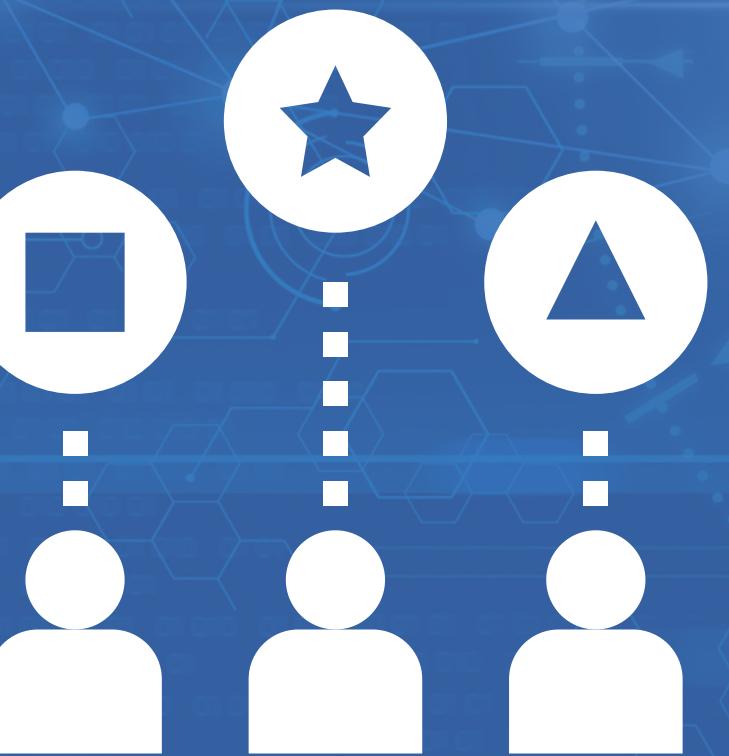
Non-responding/inactive devices

- They *slow down* the training
- In real-world applications, full participation is *infeasible*
- How to manage stragglers?
 - *Dropping* them (**FedAVG**)
 - Using only their *partially updated model* (**FedProx**)

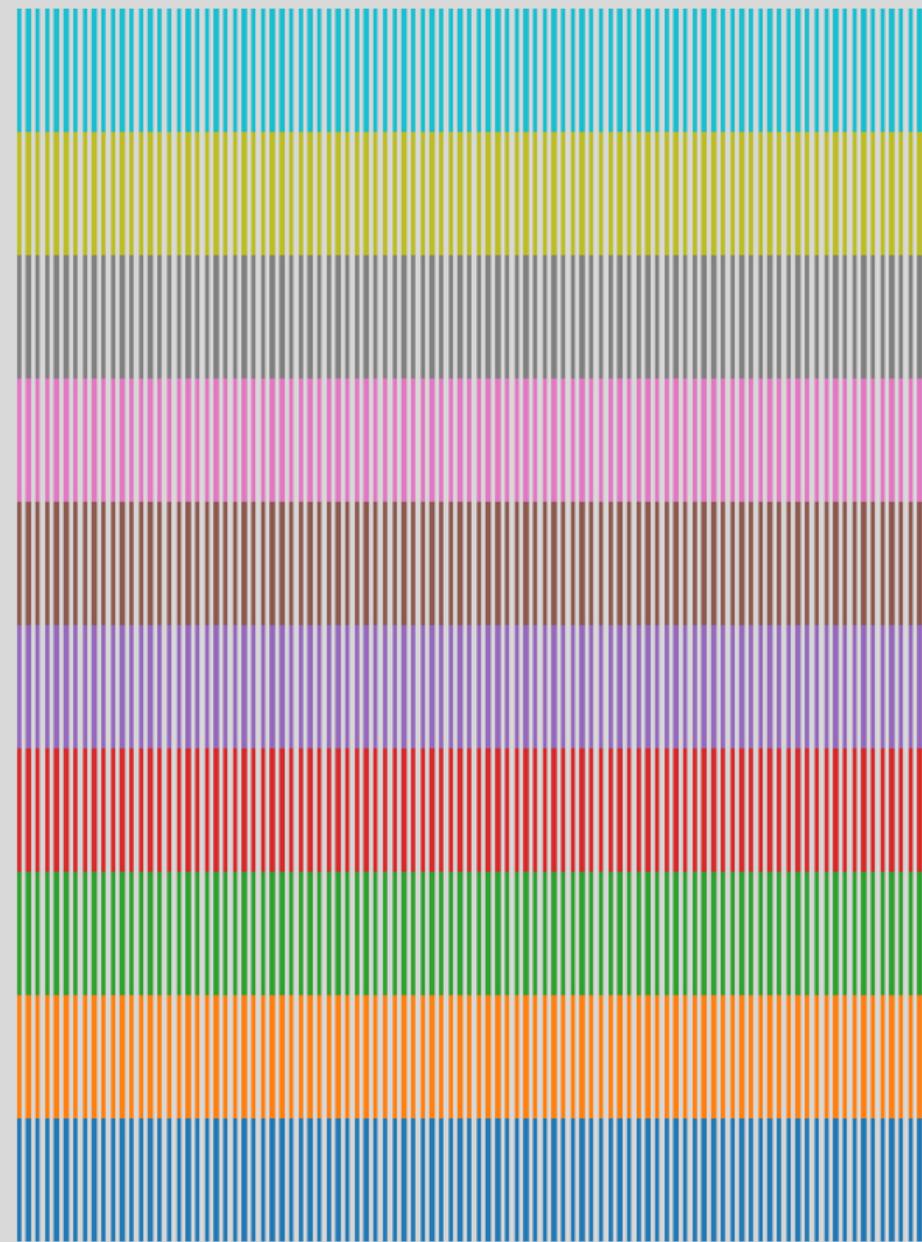


EFFECTS OF NON-IID DATA

- In real-world cases, clients datasets may not follow the population distribution
- *Client drifting* caused by the different distributions among clients (**CCVR**)
- The bigger the local updates, the farther to the global optimum the averaged model is (**FedAVG**)



DATA DISTRIBUTION



IID

mean = 10

variance = 0



non-ID balanced

mean = 1.94

variance = 0.24



non-ID unbalanced

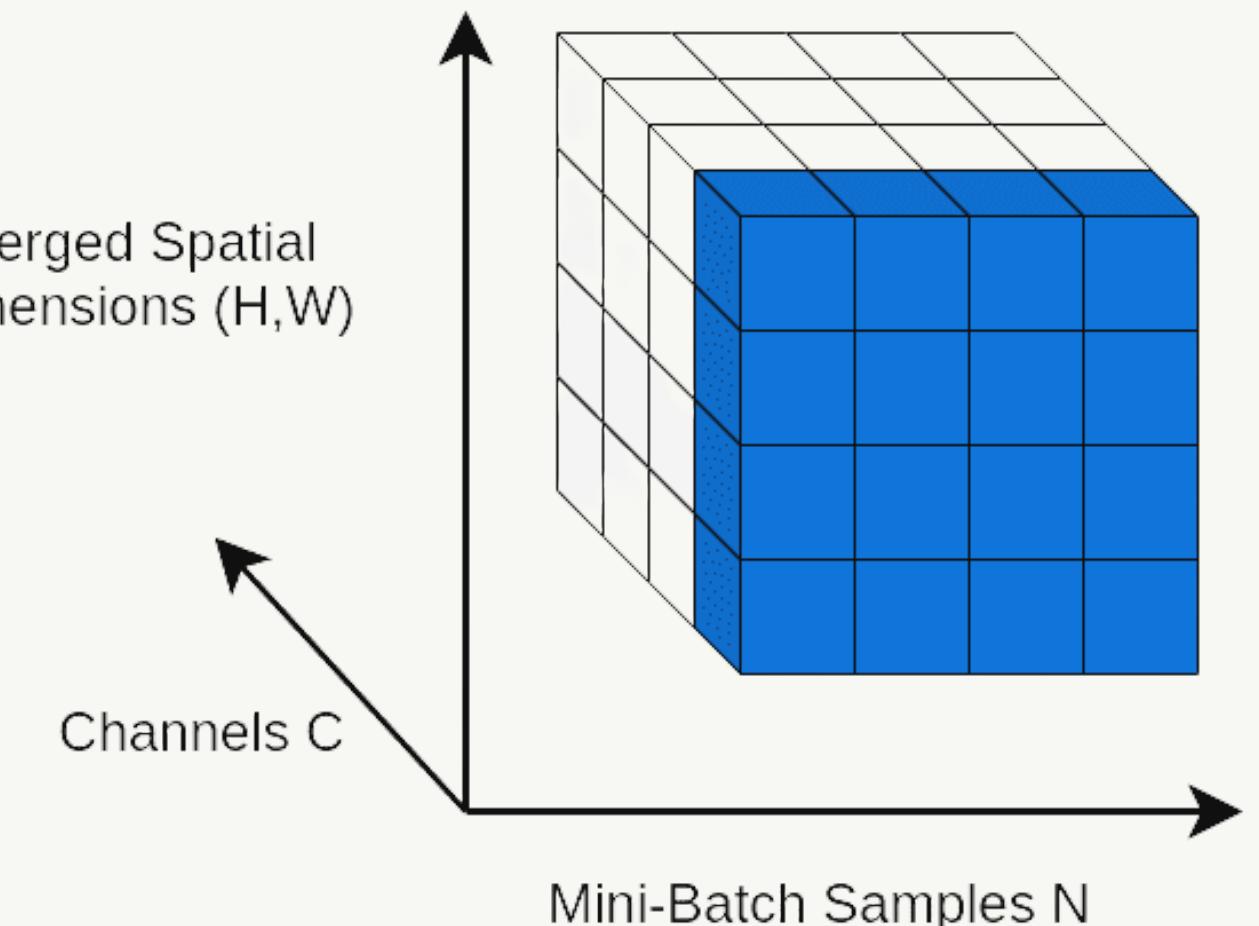
mean = 5.72

variance = 2.63

BATCH NORMALIZATION

- BN prevents the output distribution of each intermediate activation layer from *shifting*.
- It normalizes the output of the previous layer *across the batch*: data follow a Gaussian distribution.

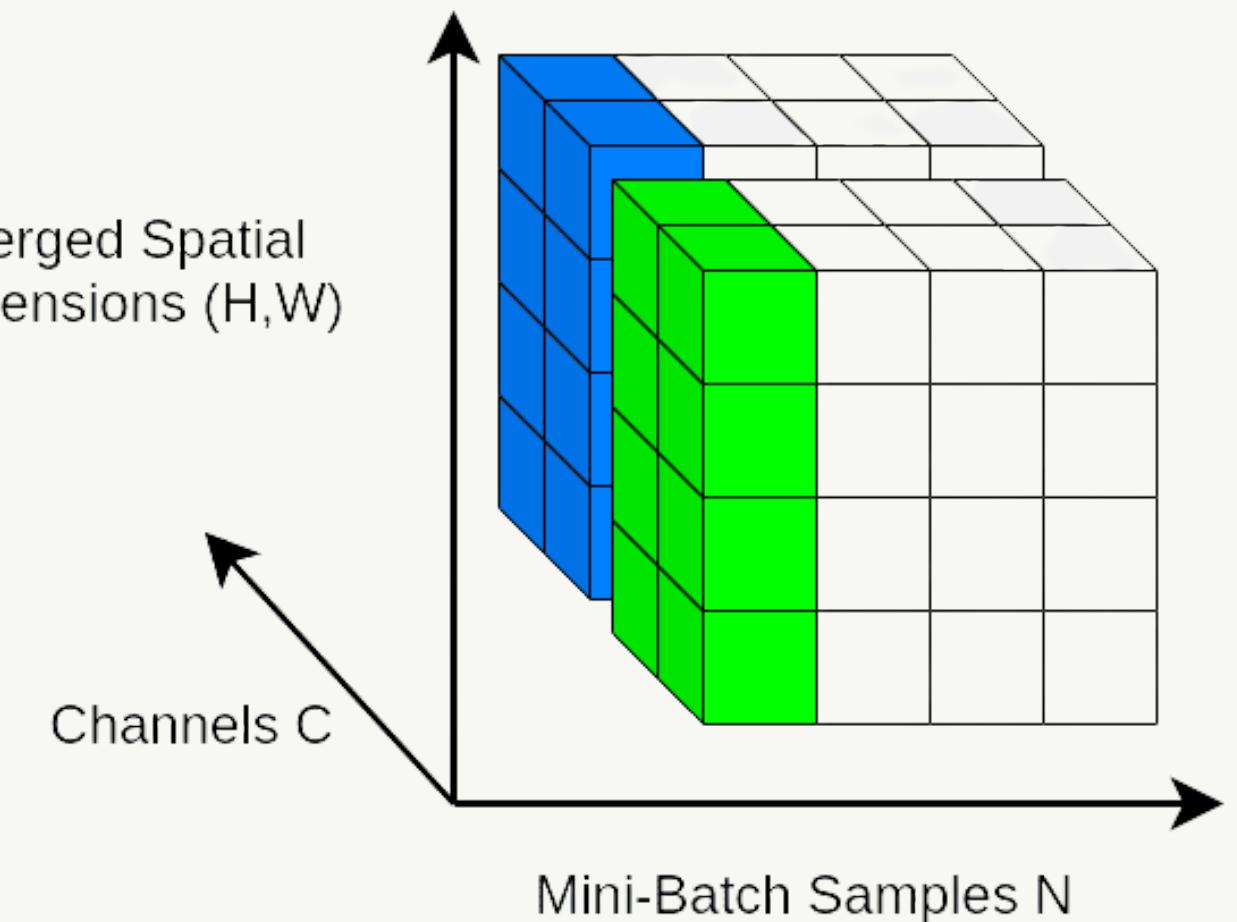
- **Benefits:** faster convergence, less importance given to initial weights.
- **Drawbacks:** instability with small batch sizes, increased training time.



GROUP NORMALIZATION

- GN divides channels into a *group*, taking away the dependence on batch size.
- It normalizes the features *within each group*.

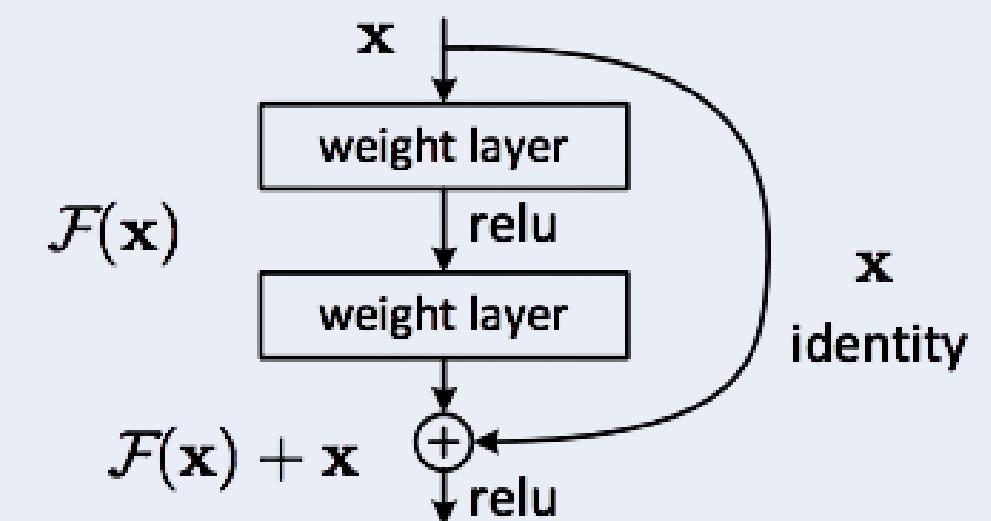
- **Benefit:** mitigation of the problem related to small batches.



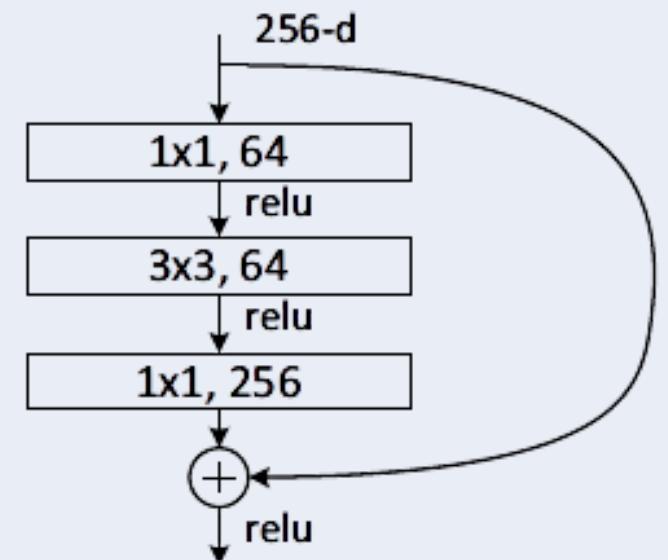
RESNET-50

Residual Neural Network with 50 layers.

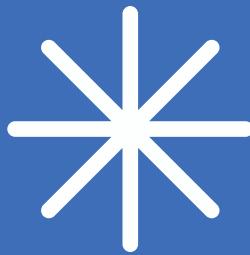
- The *Residual Block* skips connections between layers, alleviating the problem of training deep NN.
- *Bottleneck architecture* for very deep NN (50 layers).



Residual learning: building block



Bottleneck building block



Algorithms

- FedAVG
- FairAVG
- FedProx
- FedGKT
- CCVR

FEDAVG

For each global round:

- Select k devices among all the clients;
- Train each local model for E local epochs;
- Send the updated models to the central server;

The central server then conducts a weighted average.

FAIRAVG

It aggregates the clients models doing a *simple average* instead of a weighted one.

It was empirically proven that a fair aggregation could improve both accuracy and convergence rate.

FEDPROX

The algorithm is a modified version of FedAVG.

It allows to incorporate partial information from stragglers:

- *statistical heterogeneity increases*
- *convergence rate decreases*

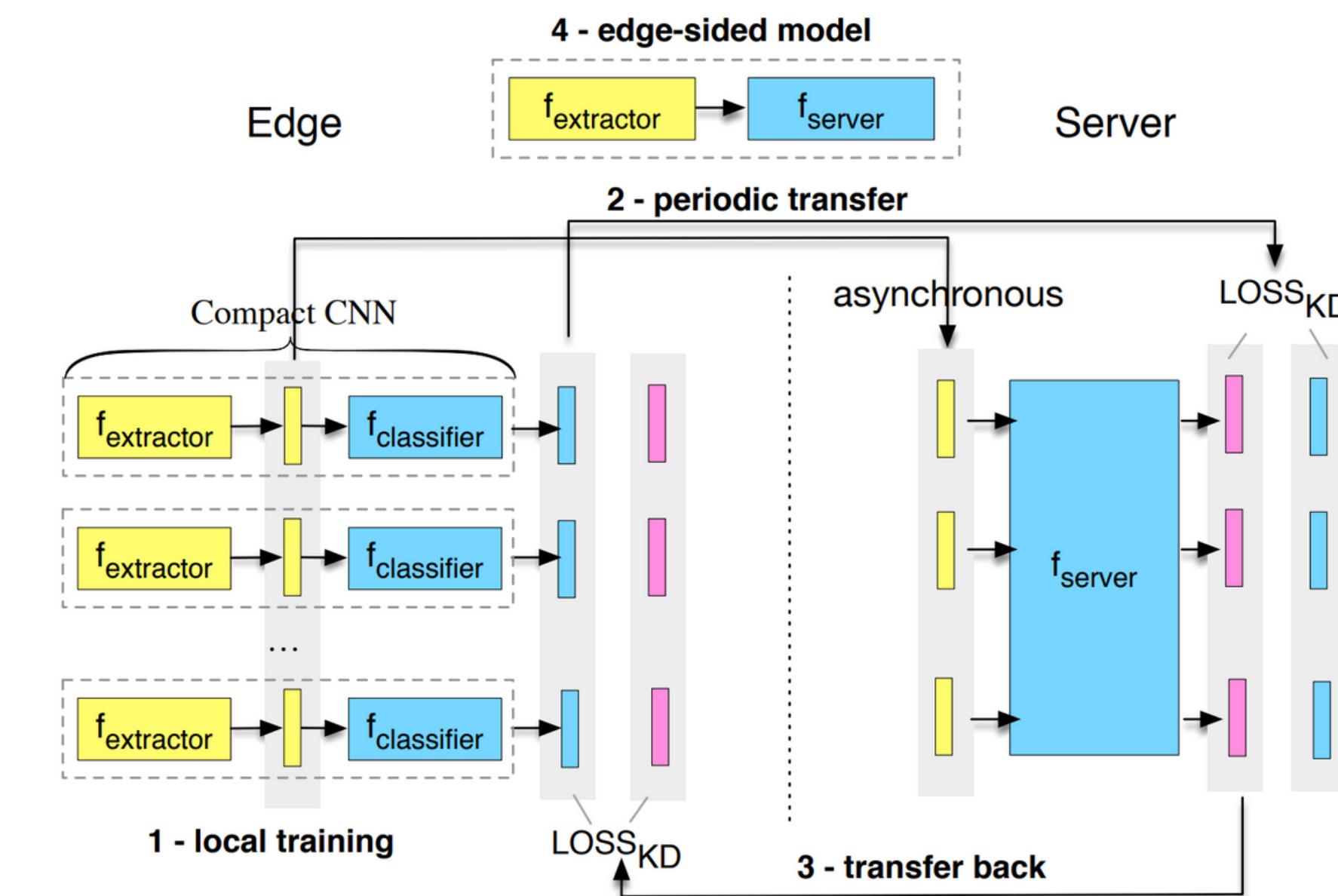
FEDPROX

Parameters:

- Proximal term to limit the impact of variable local updates.
- Hyperparameter μ to control the weight of the regularization term.

FEDGKT

- Addresses the resource-constrained reality of edge devices.
- Alternating minimization approach



FEDGKT

Configuration:

- edge CNN: ResNet8
- server model: ResNet49

FedGKT demands:

- *less computational power on edge devices*
- *fewer parameters*

Classifier Calibration with Virtual Representation

It is a technique consisting on adjusting the classifier using virtual representations sampled from an approximated Gaussian Mixture Model in the feature space with the learned feature extractor.

CCVR tries to mitigate the Client Drift phenomenon.

GRADIENT INVERSION ATTACK

Reconstructs original images from the client gradients.

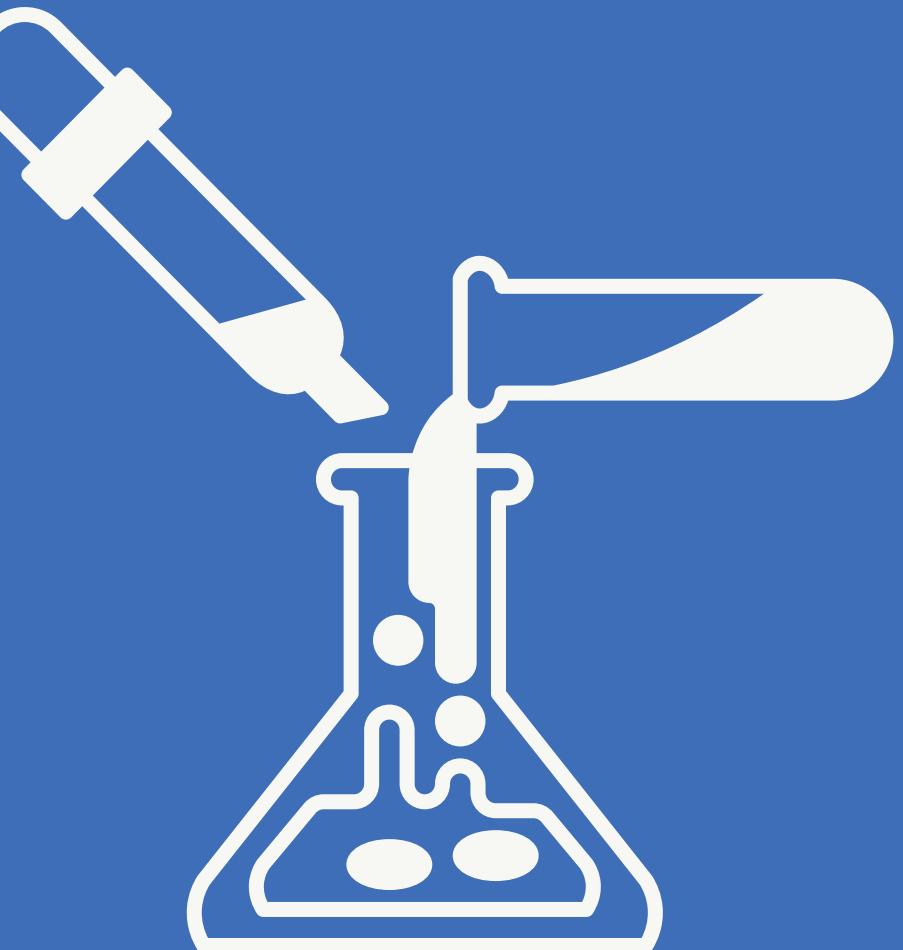
It's easy perform the attack under the assumptions that the attacker knows:

- *batch norm statistics*
- *private labels*

Relaxing them, strongly limits the efficacy of the attack.

EXPERIMENT RESULTS

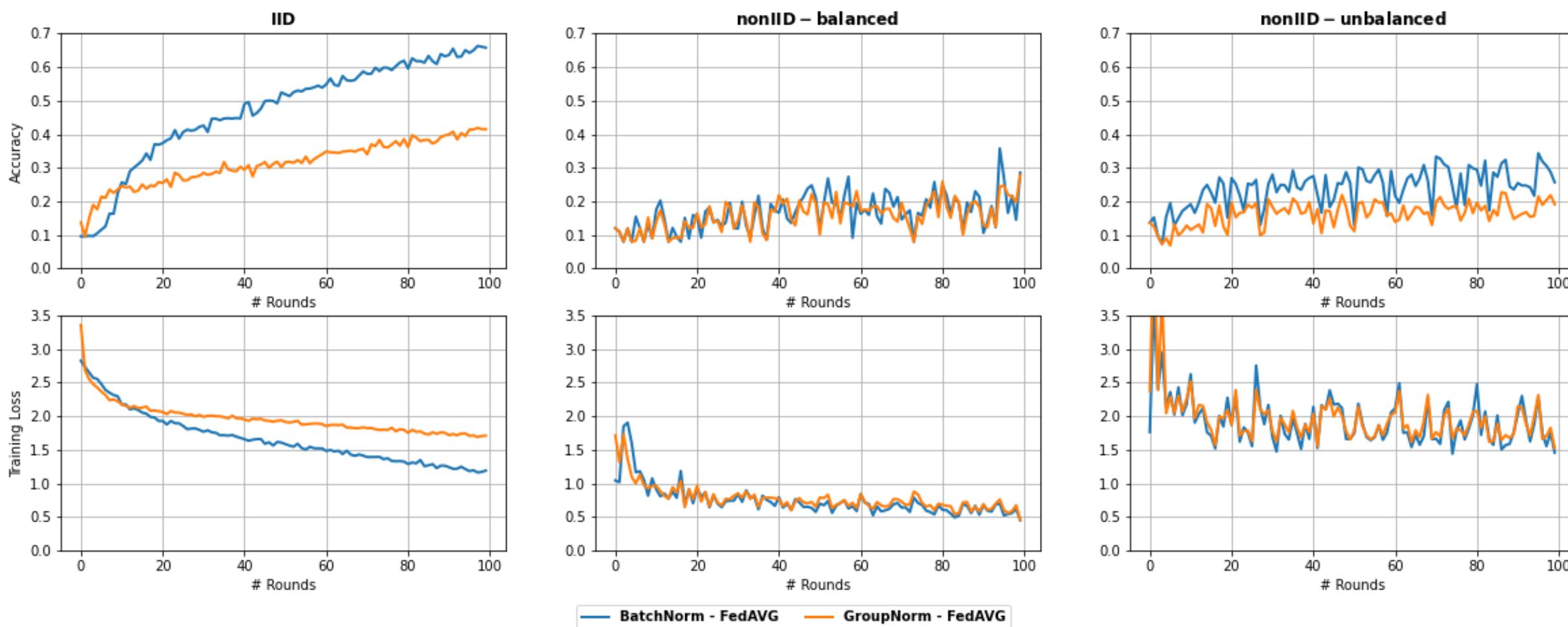
- Statistical heterogeneity
- System heterogeneity
- Privacy



FEDAVG

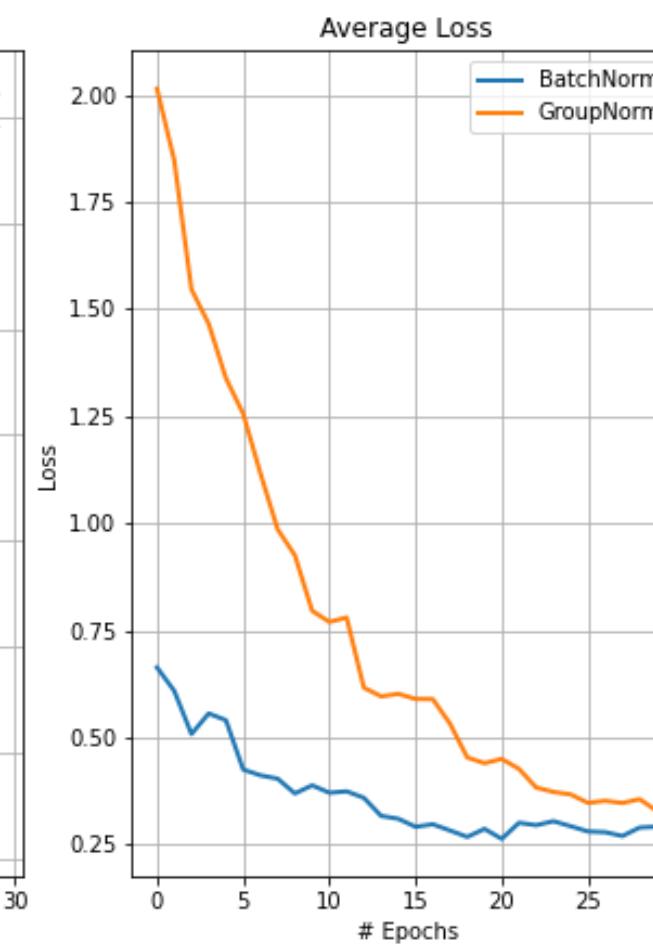
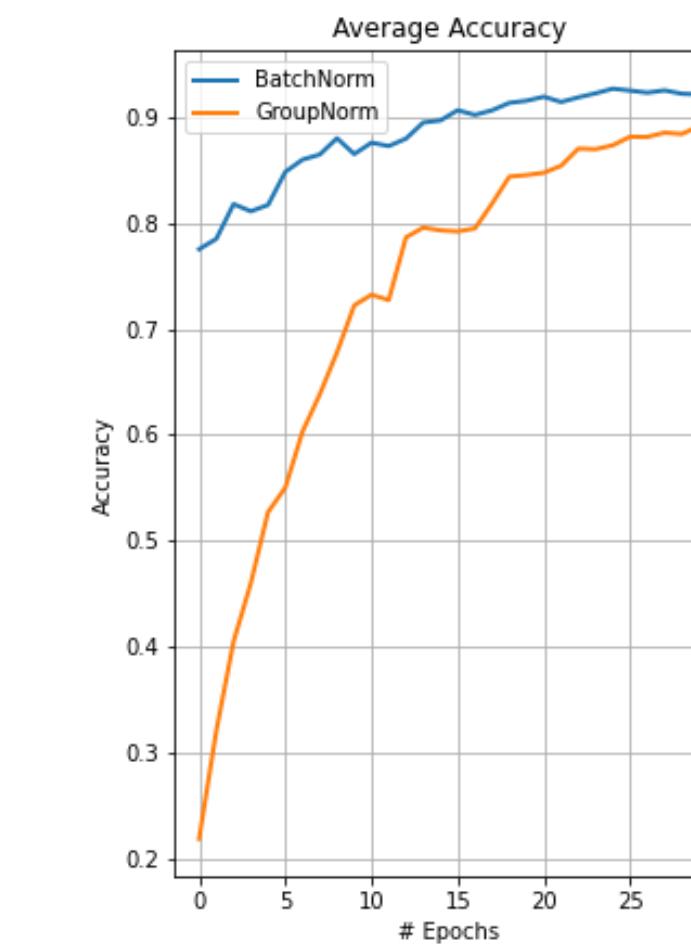
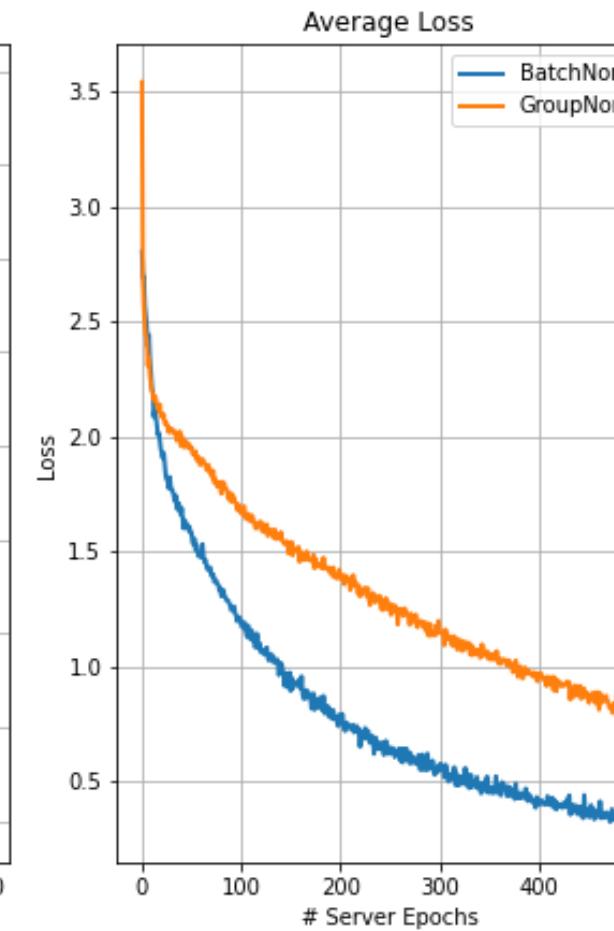
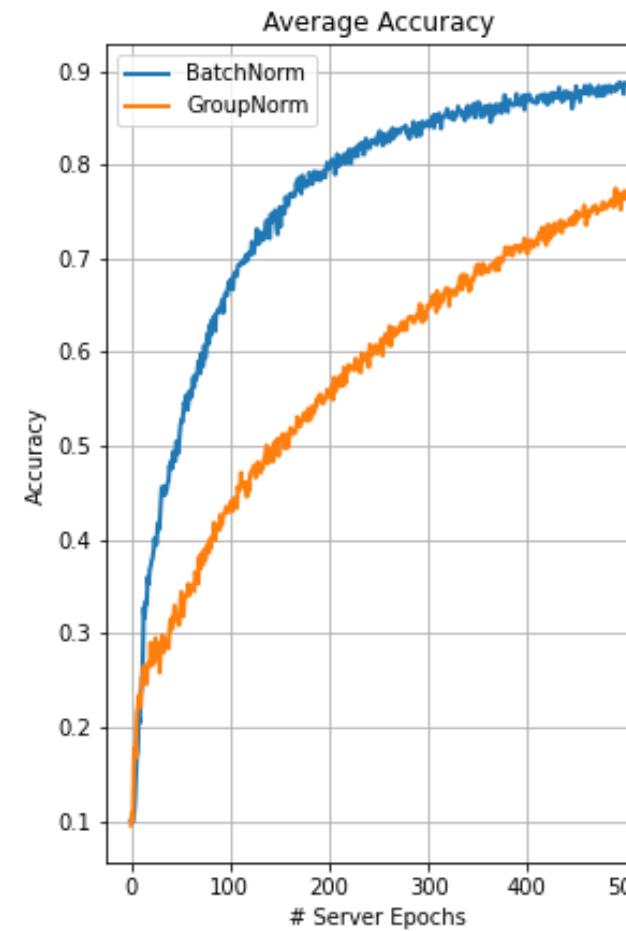


- BatchNorm always better than GroupNorm
- nonIID – balanced:
 - slightly worse on accuracy compared to unbalanced case
 - lower and more stable loss



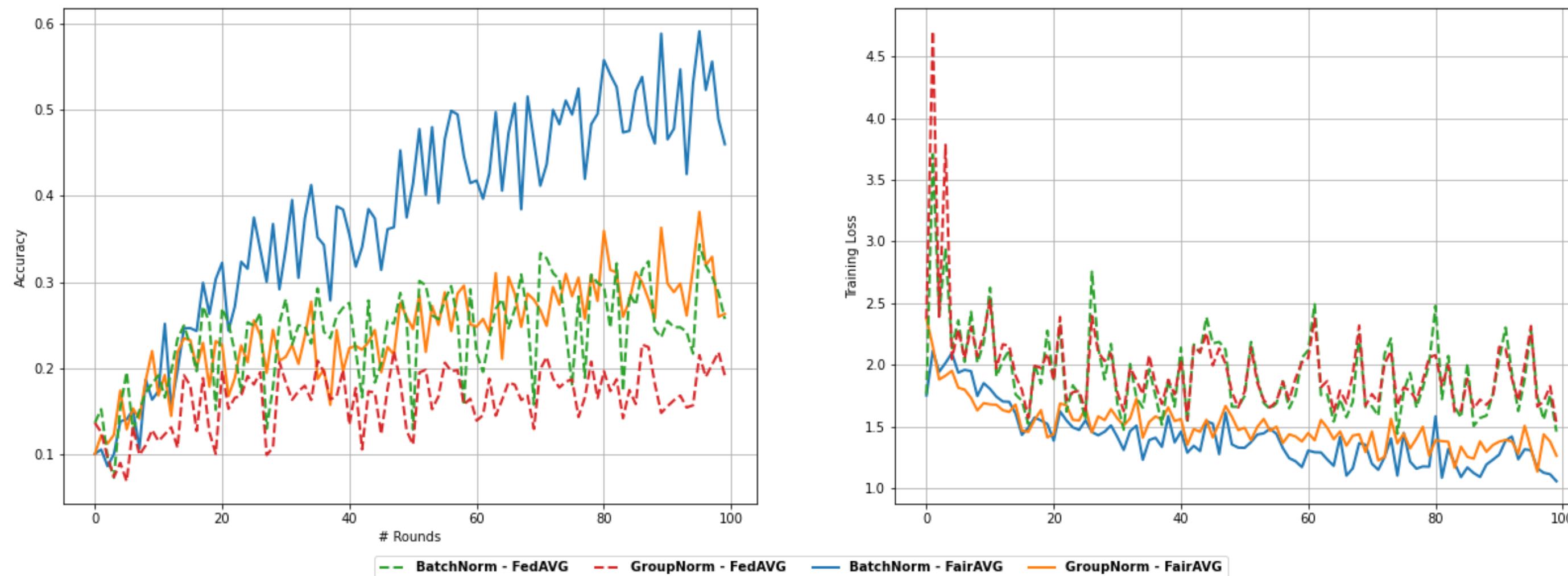
FEDAVG

- with more epochs comparable with centralized model (IID case)

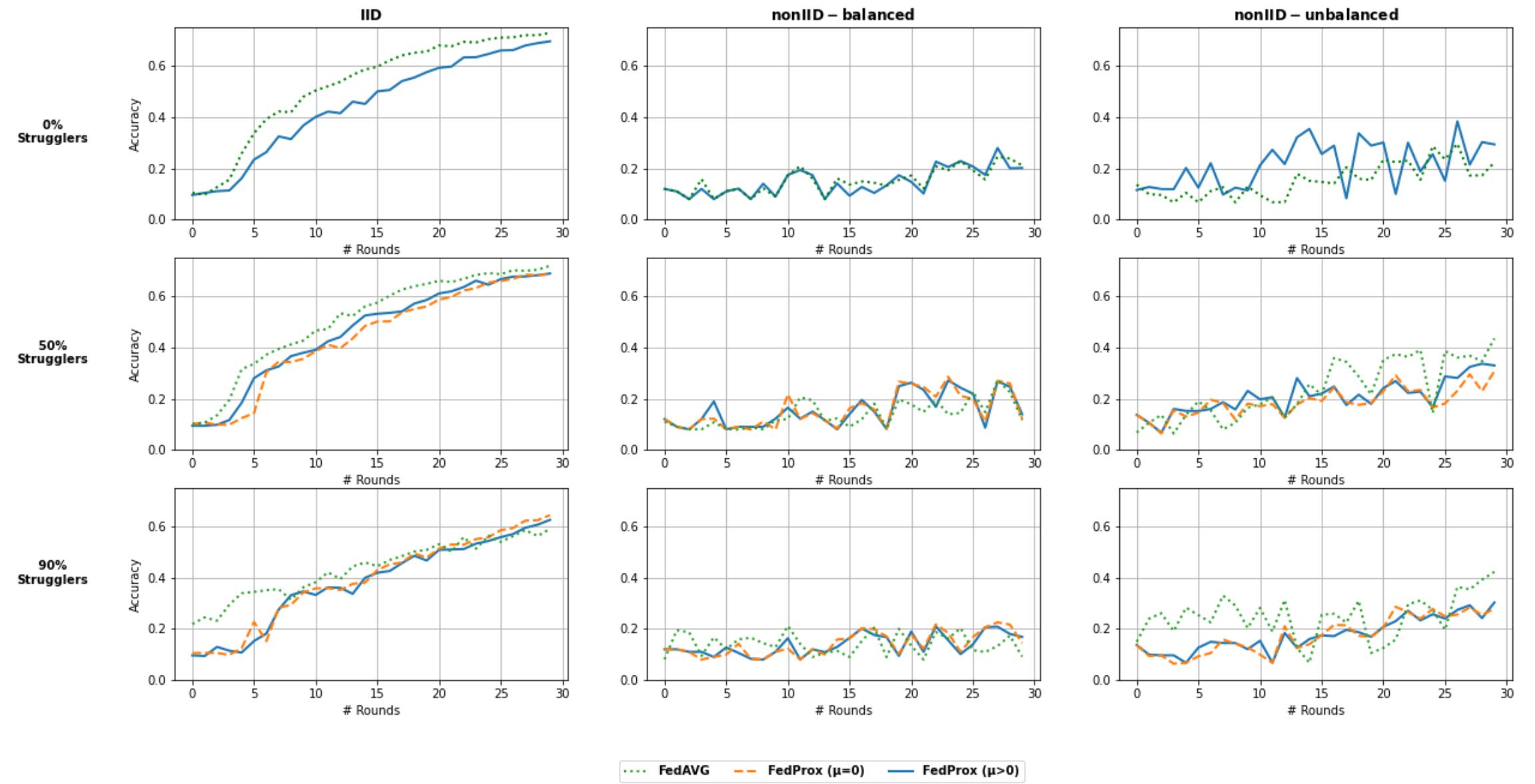


FAIRAVG

- FairAVG increases performance (up to 23%) w.r.t. FedAVG
- Accuracy curve still unstable but loss curve smoother



FEDPROX



- FedAVG generally performs slightly better
- nonIID-unbalanced:
 - FedProx wins when there are no stragglers (proximal term alleviates effect of Statistical Heterogeneity)
 - With stragglers FedAVG wins (less client drift)

- Lower complexity on edge devices with performance similar to FedAVG
- Less time to get same results as FedAVG
- The IID case is the first one where GroupNorm is not worse than BatchNorm
- nonIID - unbalanced case perform better than FedAVG

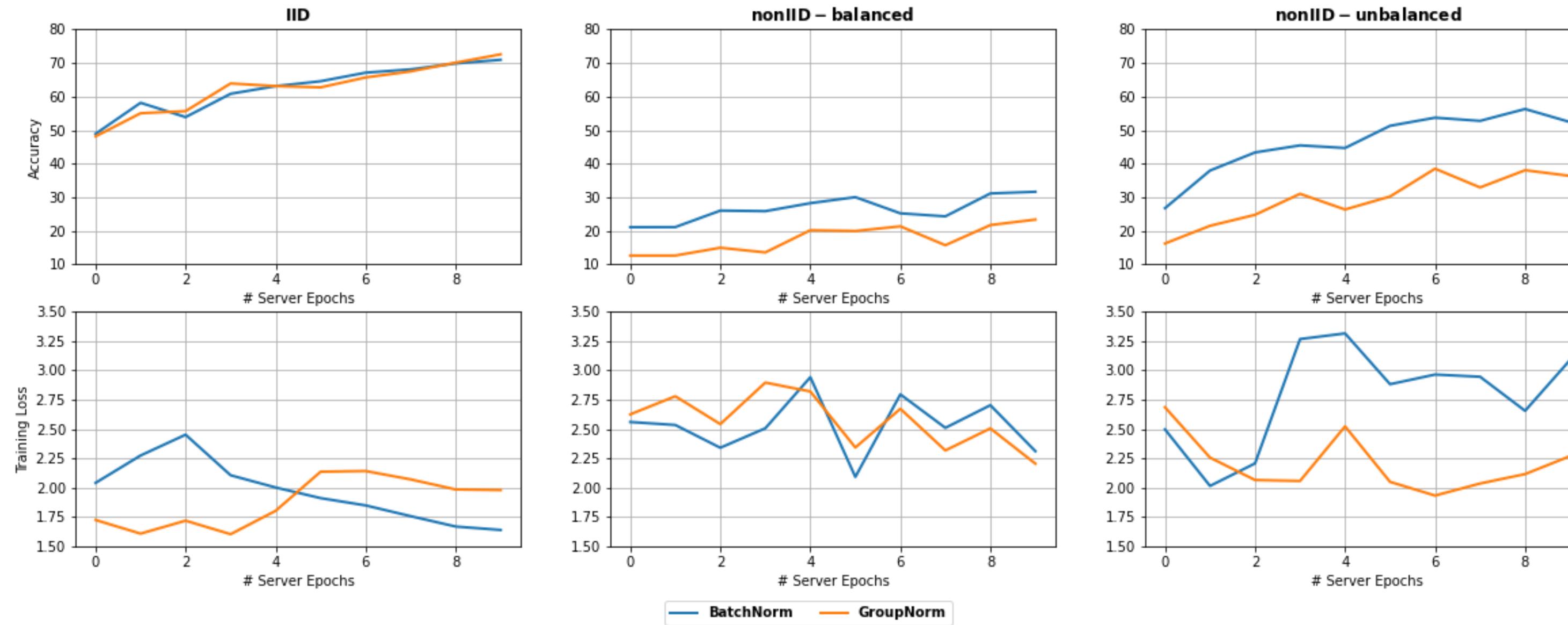
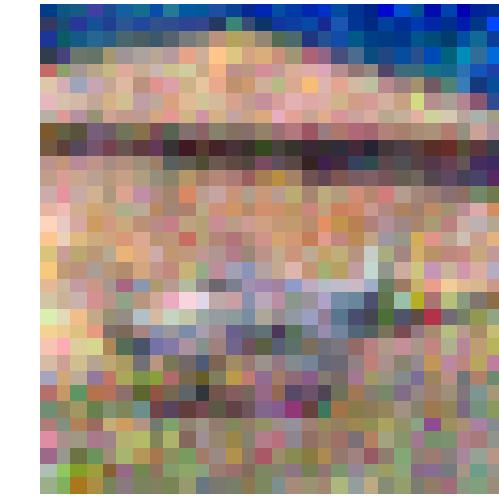
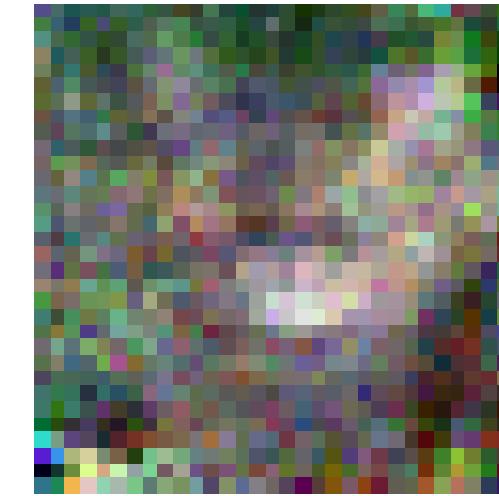
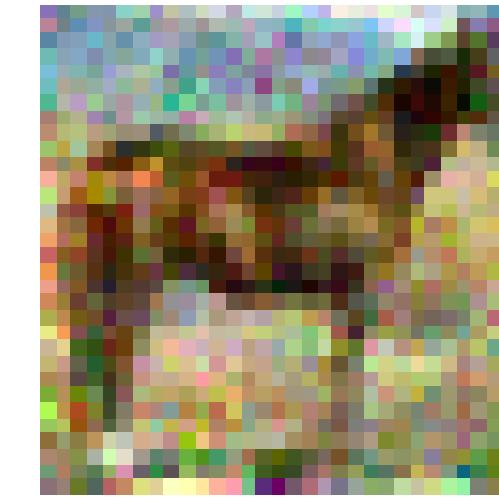
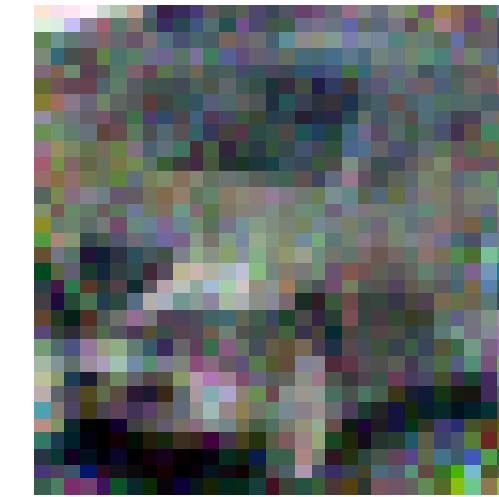
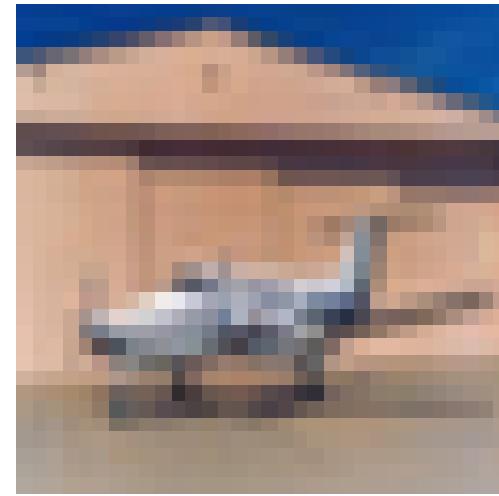
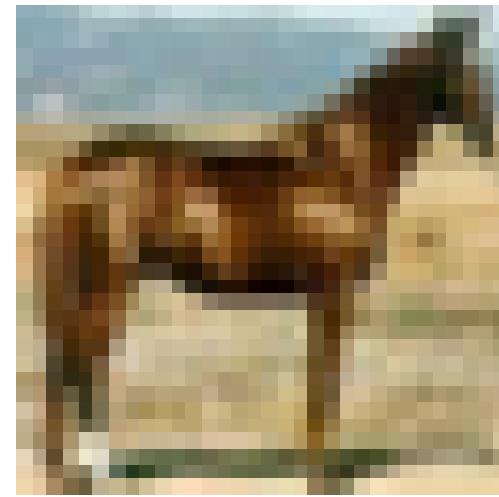


TABLE VII: CCVR

algorithm	scheme	Mc	accuracy
FedAVG + CCVR	IID	100	66.43 (-1.6)
		500	66.39 (-1.6)
		1000	66.38 (-1.6)
	non-IID balanced	100	42.52 (+10.5)
		500	42.55 (+10.5)
		1000	42.45 (+10.4)
	non-IID unbalanced	100	37.17 (+8.2)
		500	37.10 (+8.1)
		1000	37.07 (+8.0)
FairAVG + CCVR	non-IID unbalanced	100	60.92 (+8.9)
		500	61.52 (+9.5)
		1000	61.54 (+9.5)

- Performance gains in all cases
- Small drop in accuracy in the IID case, due to the fact that there is no client drift

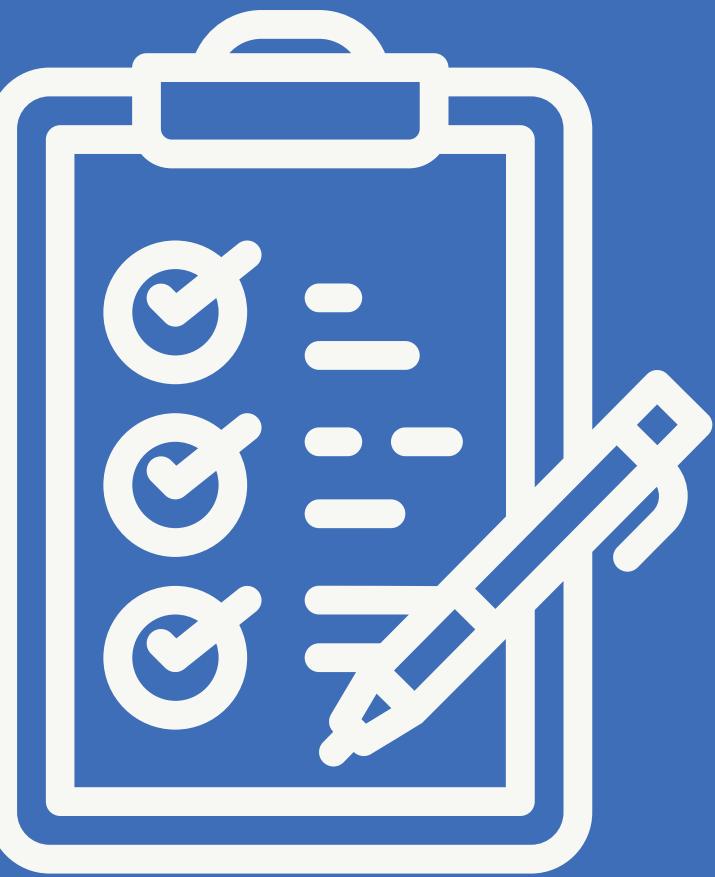
GRADIENT INVERSION ATTACK



ORIGINAL IMAGES

RECONSTRUCTED IMAGES

CONCLUSIONS





PROJECT 6
GROUP 1

ALESSANDRO CASELLA

S306081

ANGELICA MARRONE

S291261

DAVIDE DI MAURO

S306089