Federated Monocular 3D Object Detection for Autonomous Driving

Department of Electrical and Computer Engineering
University of British Columbia

Fangyuan Chi(fangchi@ece.ubc.ca), Yixiao Wang, Panos Nasiopoulos, Victor C.M. Leung, Mahsa T. Pourazad

Presenter Introduction

Fangyuan Chi

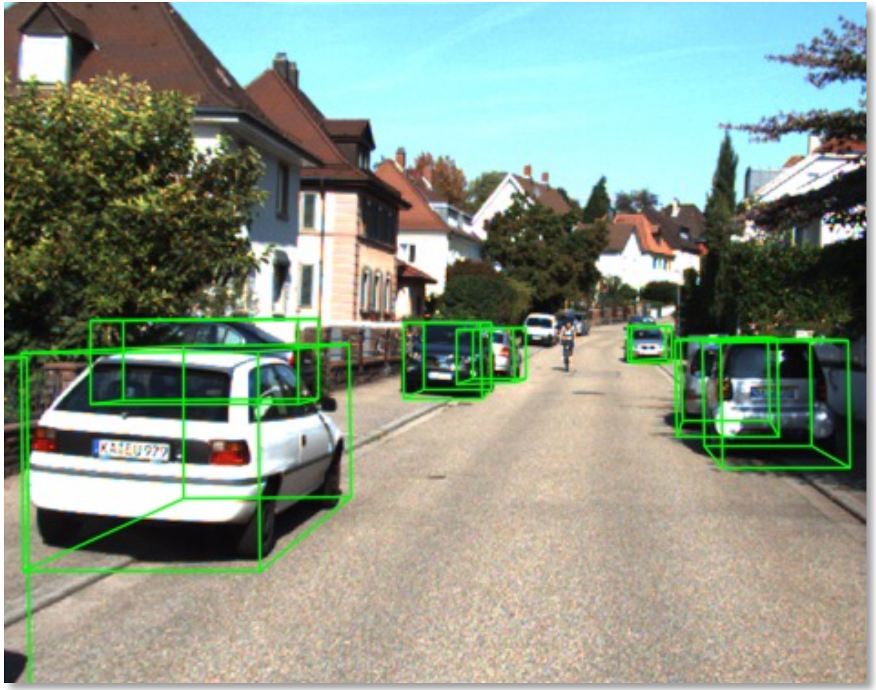
She received the B.A.Sc and M.A.Sc. degree in computer engineering from the University of British Columbia in 2012 and 2015 respectively.

In 2020, she joined, as a Ph.D student, the Digital Multimedia Lab research group at the Electrical and Computer Engineering department, University of British Columbia, Canada.

Her research area mainly focuses on computer vision, vehicle-to-everything communication, and decision making in cooperative autonomous driving

Background

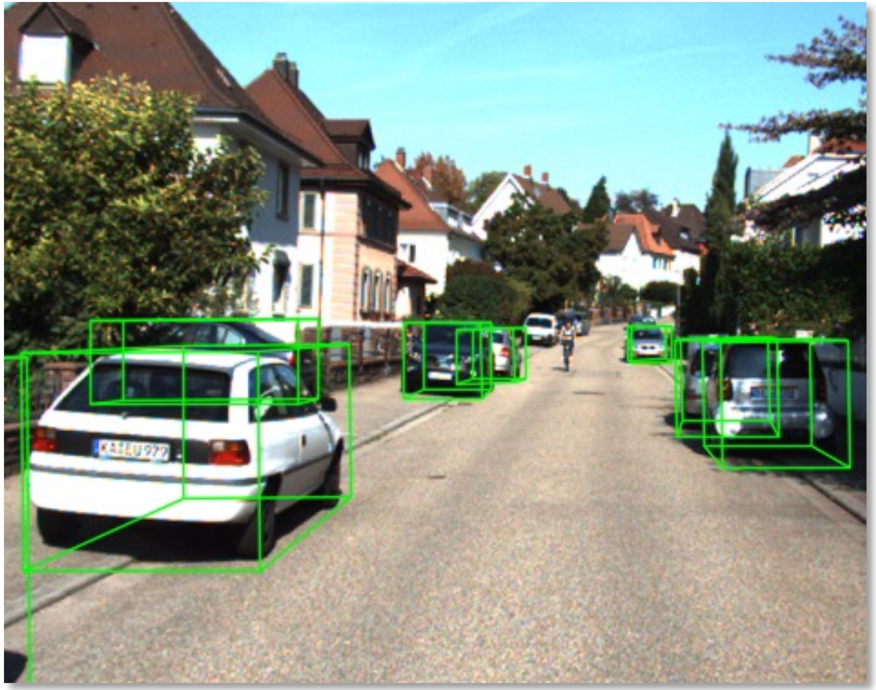
3D Object Detection: critical for autonomous driving, detect both classes and depth



Background

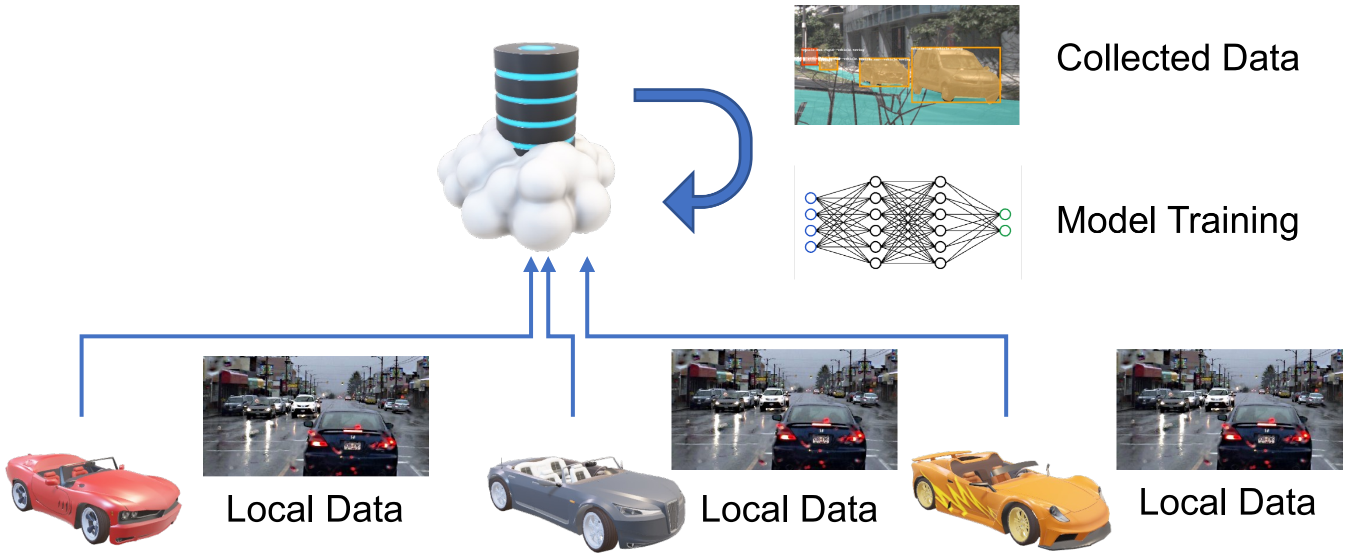
Monocular 3D Object Detection: use camera to establish 3D representation of surroundings

- Rich attributes, beyond geometry
- Affordable for all vehicles



Background

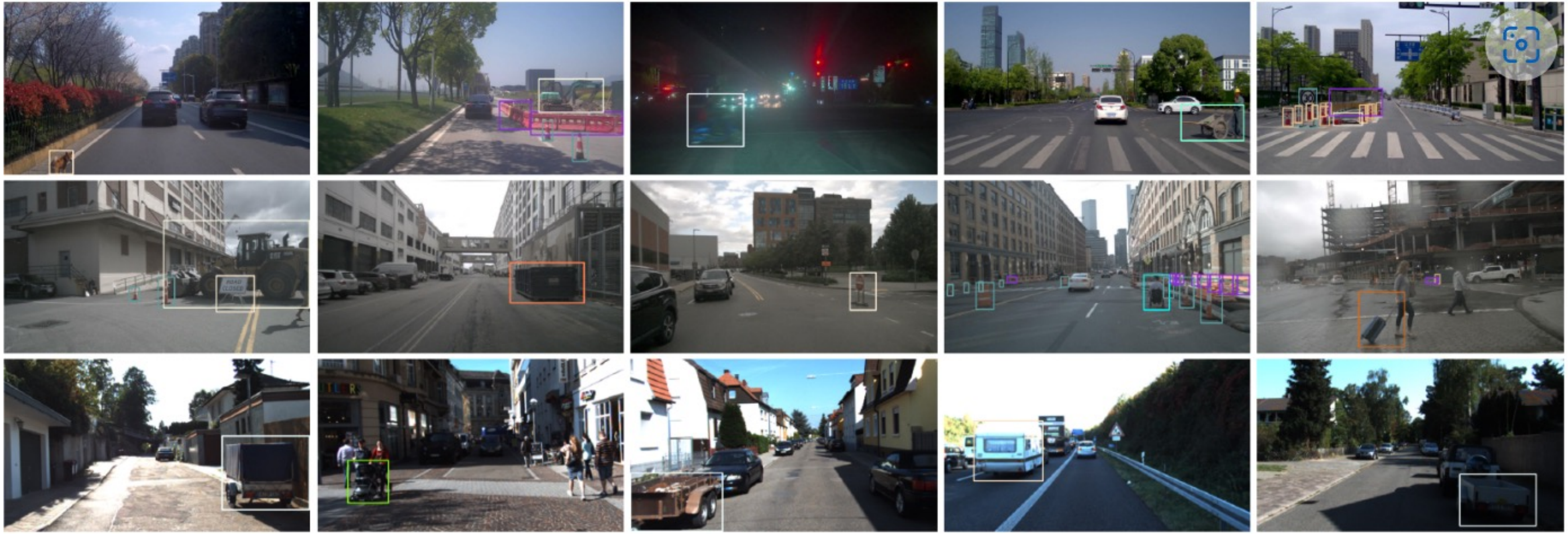
Existing Training Method: offline centralized training with large scale dataset



Background

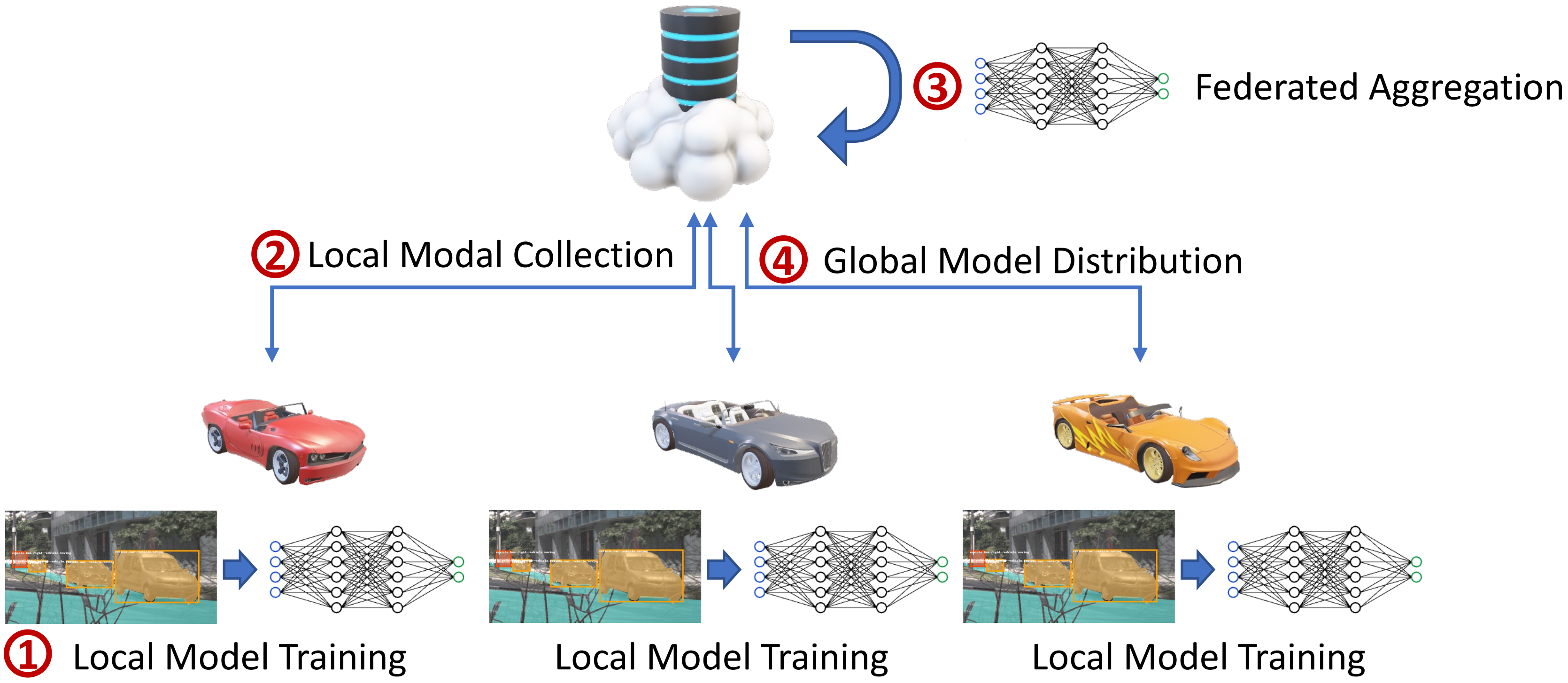
Existing Training Method: offline centralized training with large scale dataset

Challenge: How to collect a large amount of data across different areas and evolving scenarios



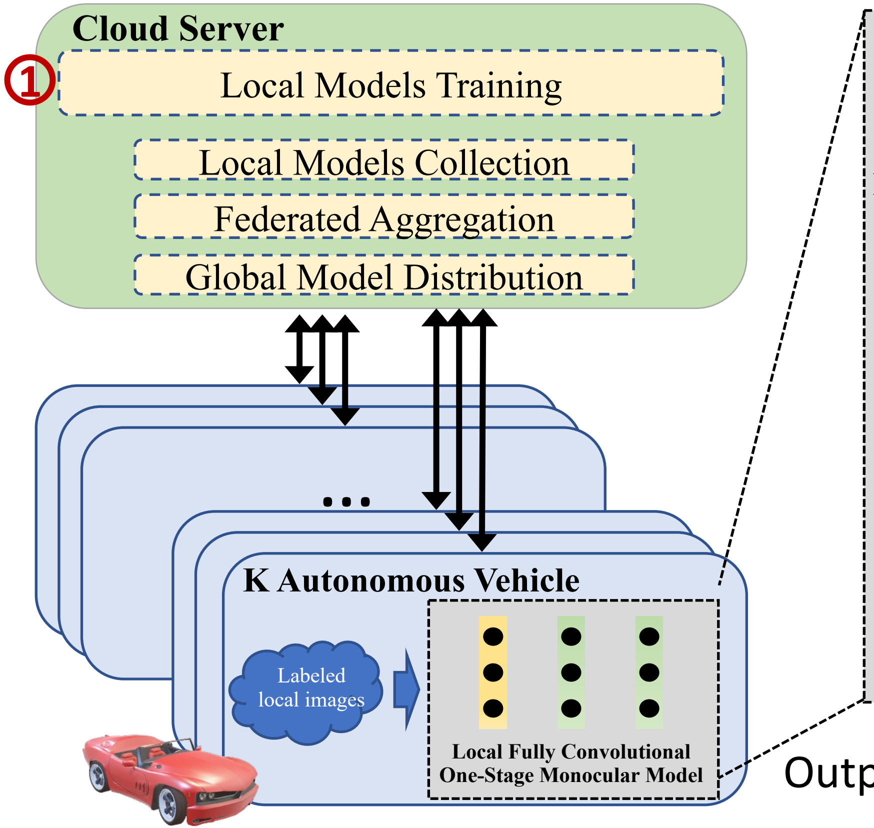
Our Method: Architecture

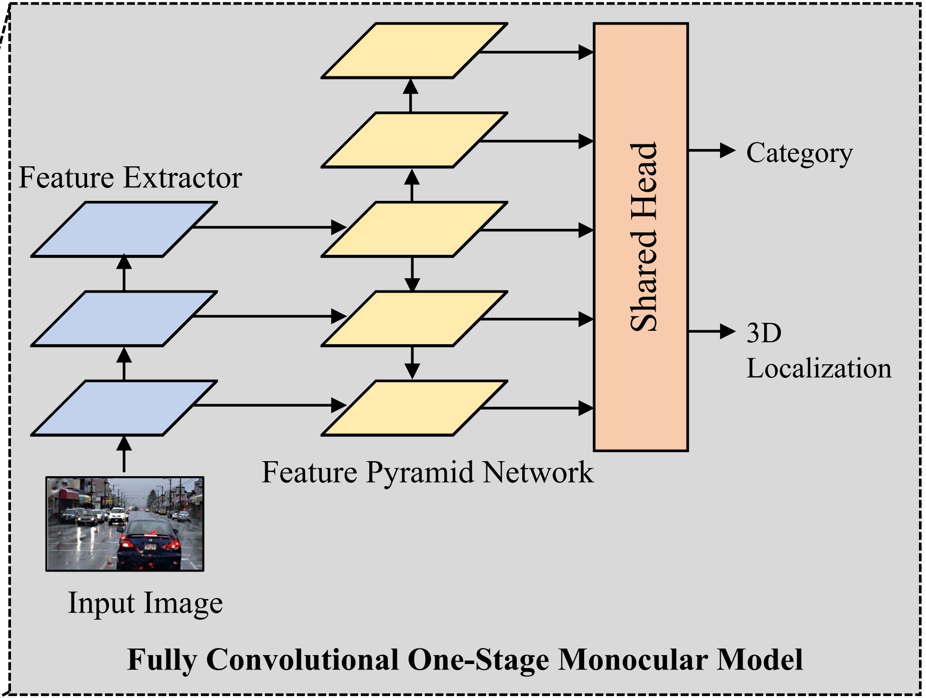
Our proposed method: leverage federated learning for distributed and continuous learning



Our Method: Local 3D Monocular Object Detection

The proposed system architecture

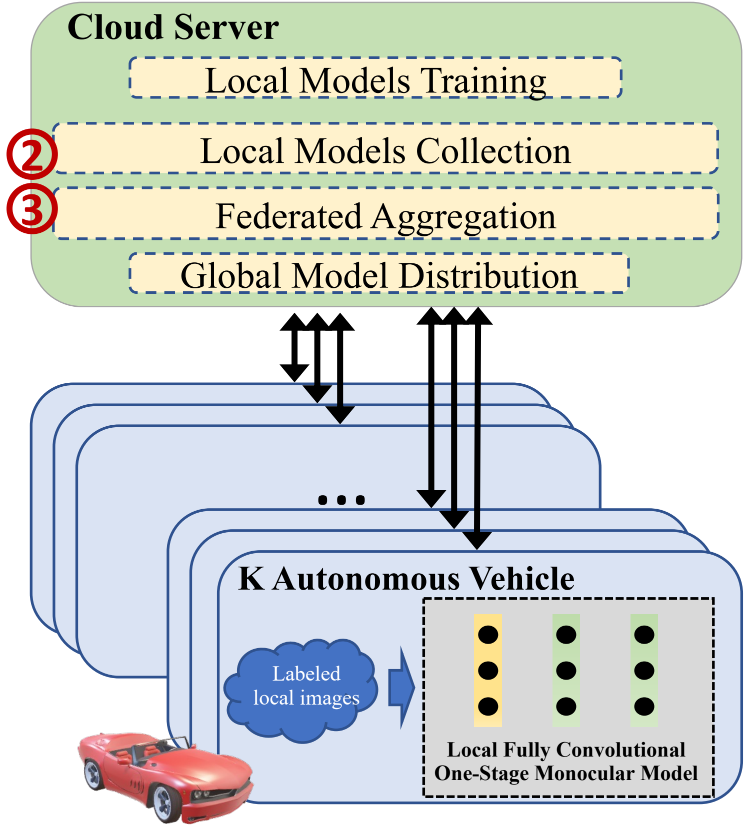




Output a local model weight ω_t^i at each training epoch

Our Method: Model Collection and Federated Average

The proposed system architecture

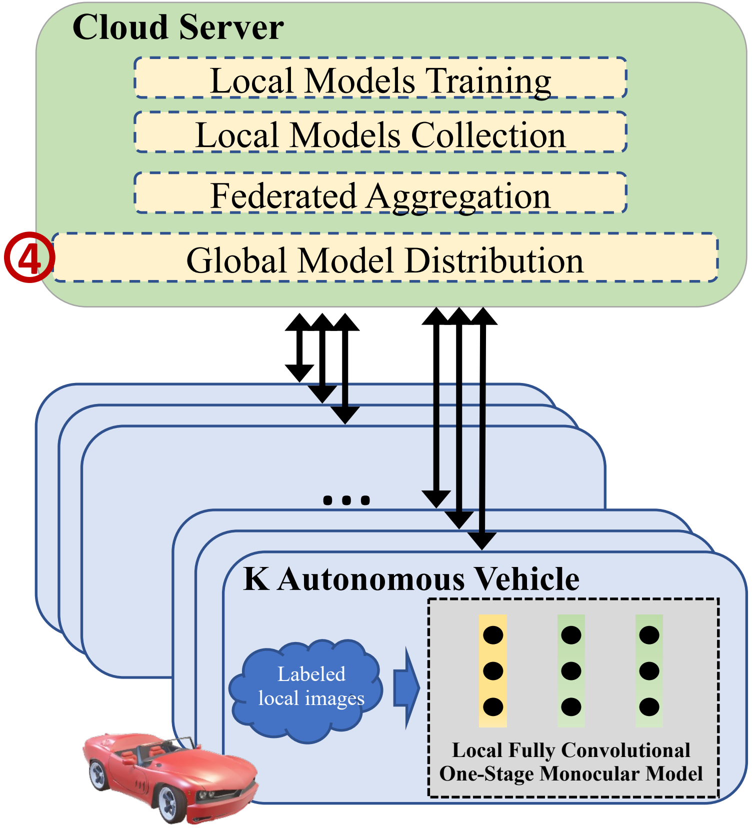


1. Receiving $\omega_t^1 \dots \omega_t^k$ from K vehicles
2. Updating global model parameters:

$$W_{t+1} \leftarrow \sum_{k=1}^{k} \frac{n_k}{n} \omega_t^k$$

Our Method: Global Model Distribution

The proposed system architecture



1. Receiving W_{t+1} from the cloud server
2. Updating local model parameters:

$$\omega_t^i \leftarrow W_{t+1}$$

Evaluation

Evaluation Setup

Real-word Dataset: nuScenes

1,200,000 camera images captured from **1000 scenes**, **23 annotated object classes** including vehicles, pedestrians, etc.

Simulated Autonomous Vehicles

We split the dataset into **10 parts** to simulate that we have **10 autonomous vehicles** participating in the federated learning process

Training Process

We train the model for **12 epochs** at a batch size of **4**. At the end of each epoch, the model parameters from each vehicle we be averaged

Evaluation

Evaluation Metrics

NuScenes Detection Score (NDS):

The NDS metric is specific to the nuScenes dataset, and measures the overall performance of the model in terms of both precision and recall.

Evaluation

Evaluation Metrics

<u>Mean Average Precision (mAP)</u>:

Measures the average precision of the model across all classes and all thresholds.

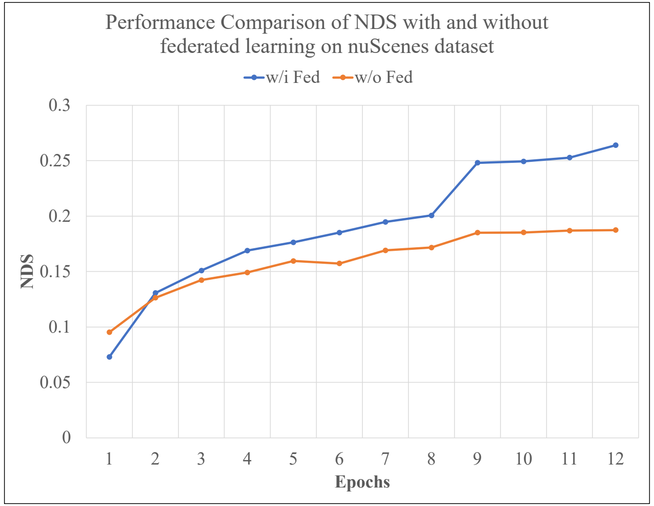
Calculated by averaging the precision of the model at different recall levels

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \qquad \text{where}$$

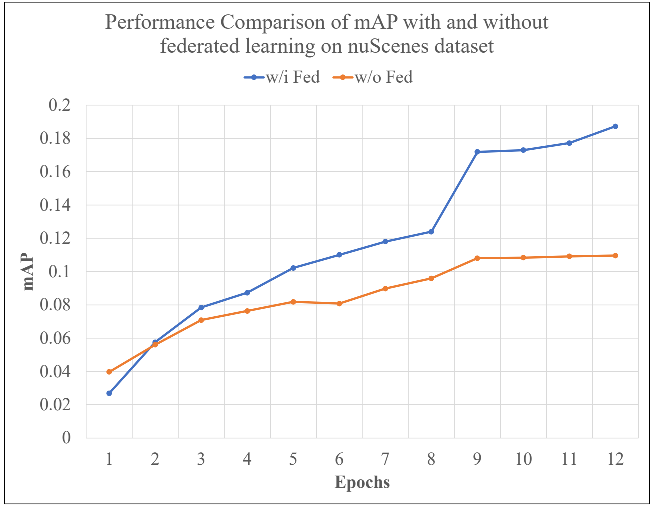
n = the number of classes
AP_k = the average precision of class k

Evaluation Results

NDS



mAP



Method	Backbone	Data Ratio	NDS	mAP
Baseline	ResNet101	10%	0.187	0.110
Ours	ResNet101	10%	**0.264(+41.18%)**	**0.187(+70%)**

Summary

Integrate federated learning with 3D monocular object detection

- Enable continuous learning with large scale dataset from evolving driving scenarios
- Improve detection accuracy and robustness

Thank you