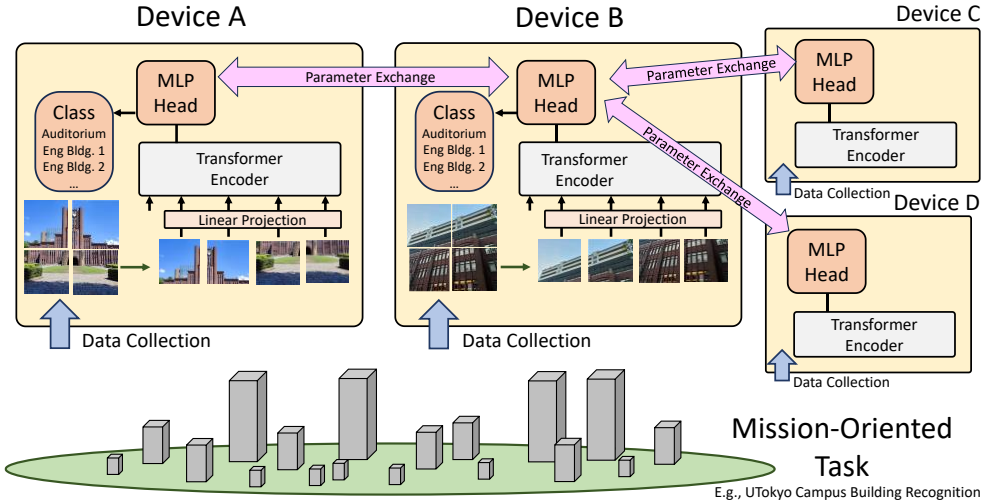
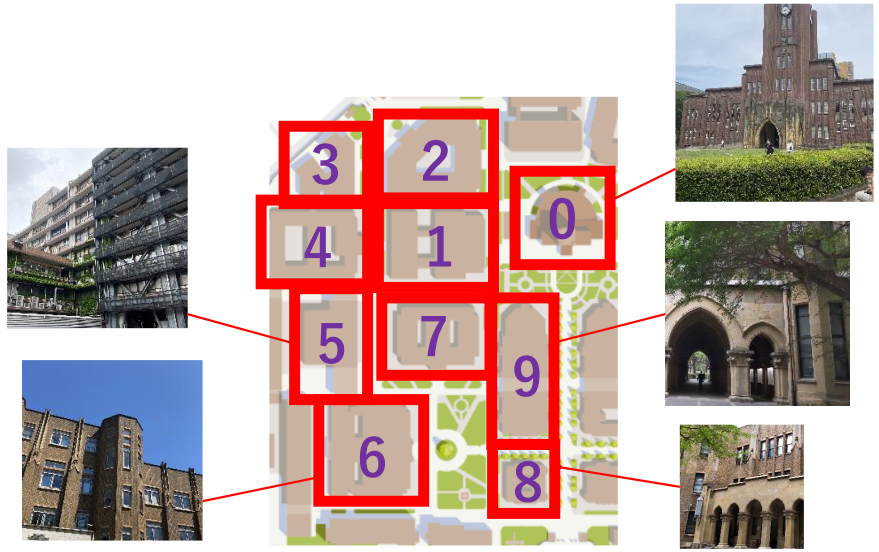
Tuning Vision Transformer with Device-to-Device Communication for Targeted Image Recognition





Hideya Ochiai, Atsuya Muramatsu, Yudai Ueda,
Ryuhei Yamaguchi, Kazuhiro Katoh, Hiroshi Esaki

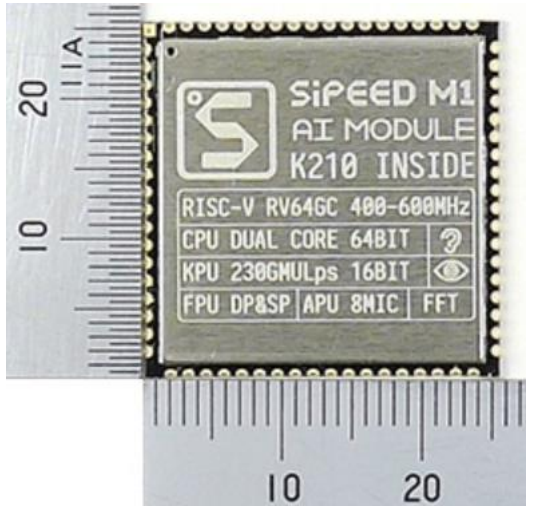
The University of Tokyo, Japan

Table of Contents

Table of Contents

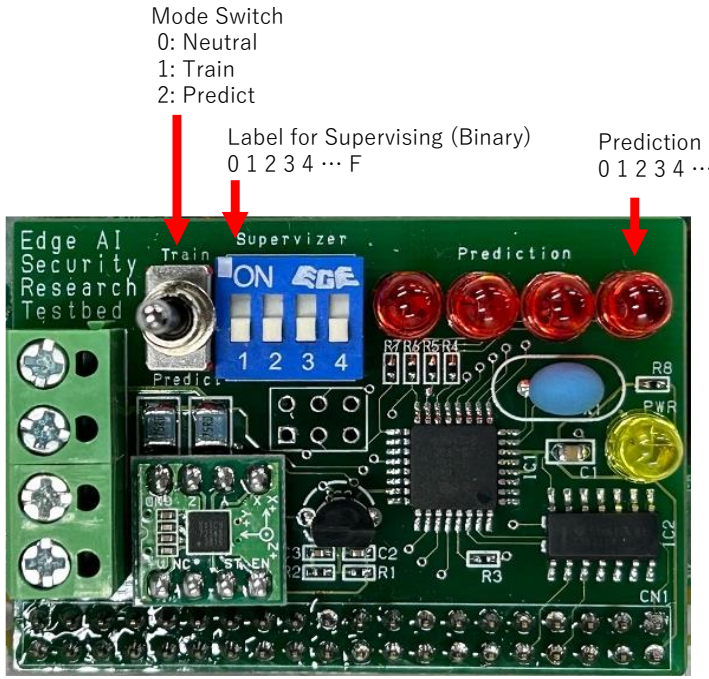
Toward On-Device Training

- Machine Learning has evolved with cloud computing.
- Nowadays :
  - AI Chips are available for 10 USD.
  - Machine Learning shifts onto IoT Edges.



AI Chip (10USD)

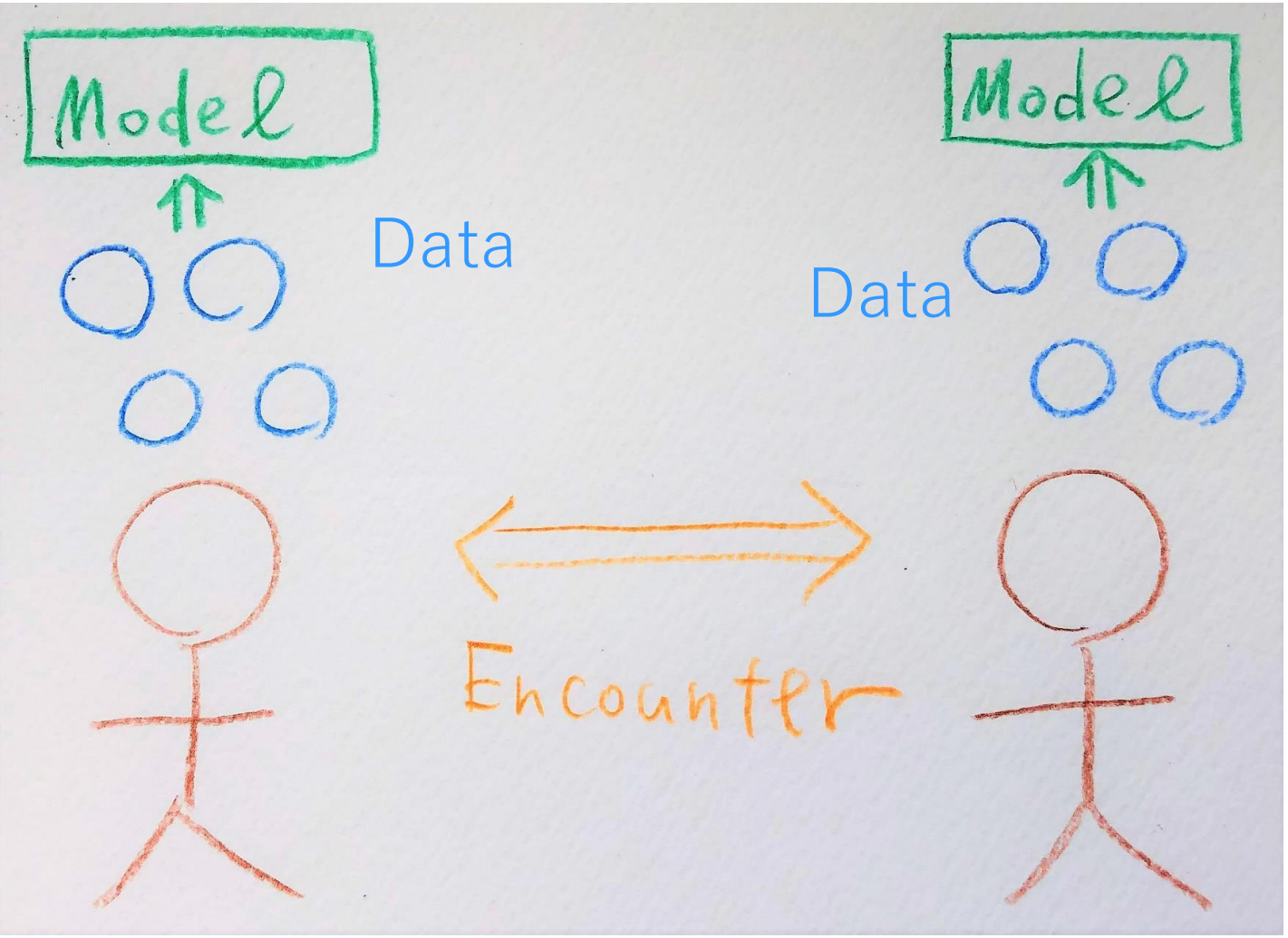
- Issues:
  - Learning on an Edge may lead to model overfit.
- Approach of this issue:
  - Collaborative Model Tuning with Device-to-Device Communication.
- Focus of this research:
  - Image recognition with Vision Transformer
  - Building Recognition Task of the University of Tokyo



Edge AI Testbed

Table of Contents

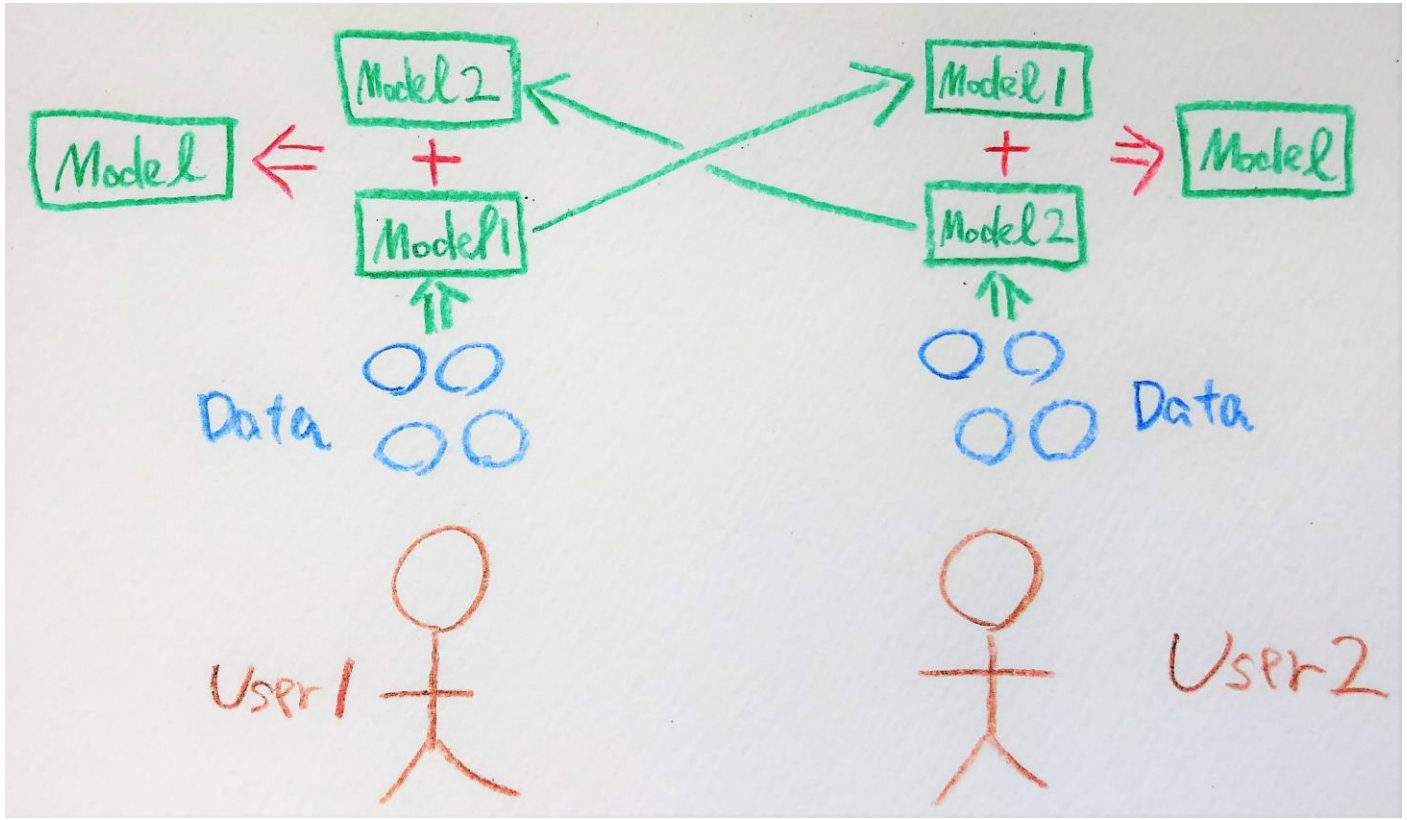
Previous Studies (1/5): Wireless Ad Hoc Federated Learning (WAFL) [1] with Device-to-Device Communication



1. Each node individually trains its ML model using its local data.
2. Each node encounters the other.
3. They can communicate with local wireless communication media such as Wi-Fi Ad Hoc mode or Bluetooth

[1] Ochiai, Hideya, et al. "Wireless ad hoc federated learning: A fully distributed cooperative machine learning." *arXiv preprint arXiv:2205.11779* (2022).

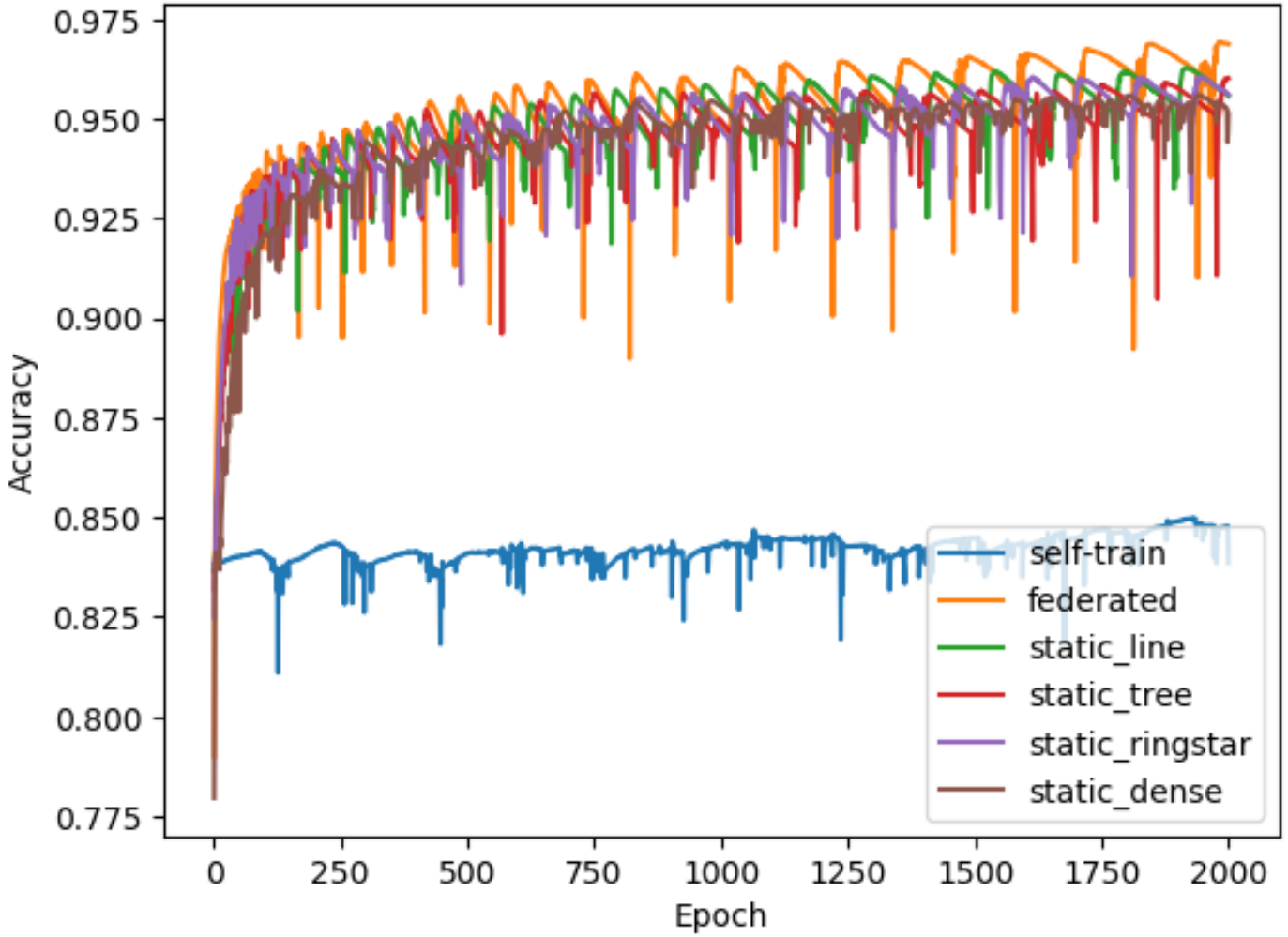
Previous Studies (2/5): Wireless Ad Hoc Federated Learning (WAFL) [1] with Device-to-Device Communication



1. Each node individually trains its ML model using its local data.
2. Each node encounters the other.
3. They can communicate with local wireless communication media such as Wi-Fi Ad Hoc mode or Bluetooth
4. They exchange and aggregate the models to develop a new model.
5. This enables collaborative training.

[1] Ochiai, Hideya, et al. "Wireless ad hoc federated learning: A fully distributed cooperative machine learning." *arXiv preprint arXiv:2205.11779* (2022).

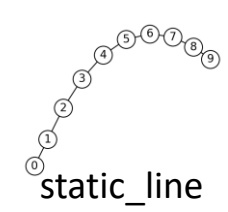
Previous Studies (3/5): Wireless Ad Hoc Federated Learning (WAFL) [1] with Device-to-Device Communication

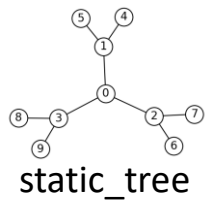


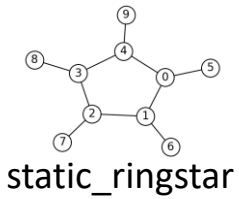
WAFL has achieved almost the same performance as conventional Federated Learning.

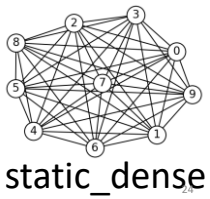
10.0-11.5% Improvement

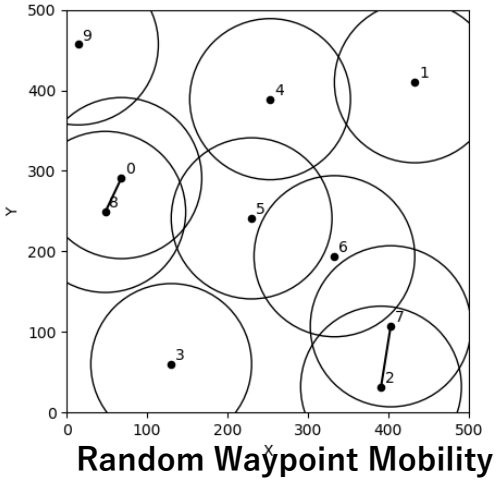
Accuracy of Self-Train







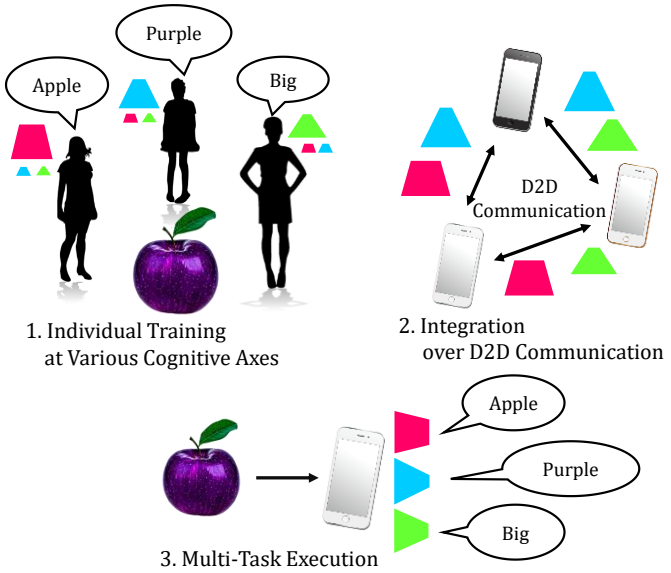




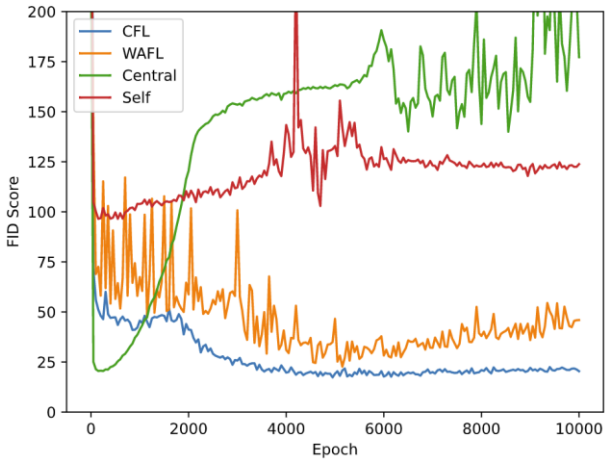
Ochiai, Hideya, et al. "Wireless ad hoc federated learning: A fully distributed cooperative machine learning." *arXiv preprint arXiv:2205.11779* (2022).

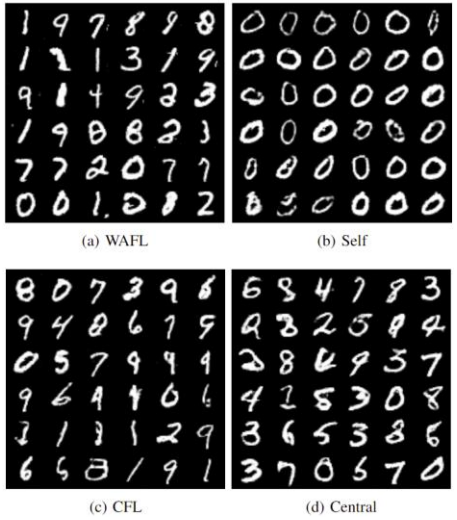
Previous Studies (4/5): Extensions and Variations of WAFL (1/2)

WAFL- Multi-Task Learning[1]

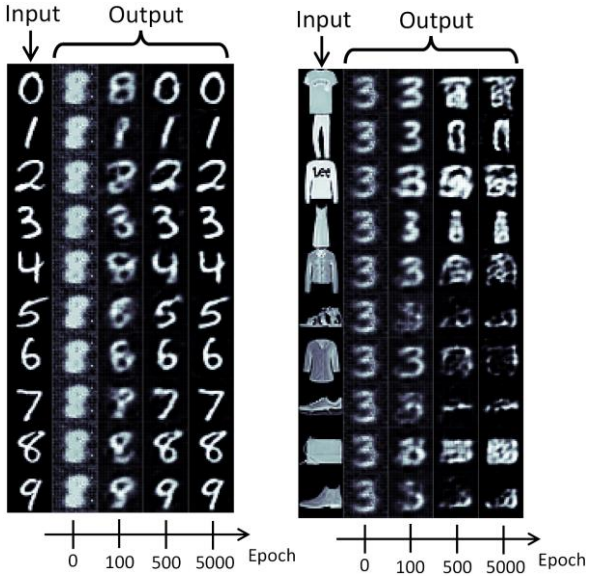


WAFL-GAN[2]





WAFL-Autoencoder[3]



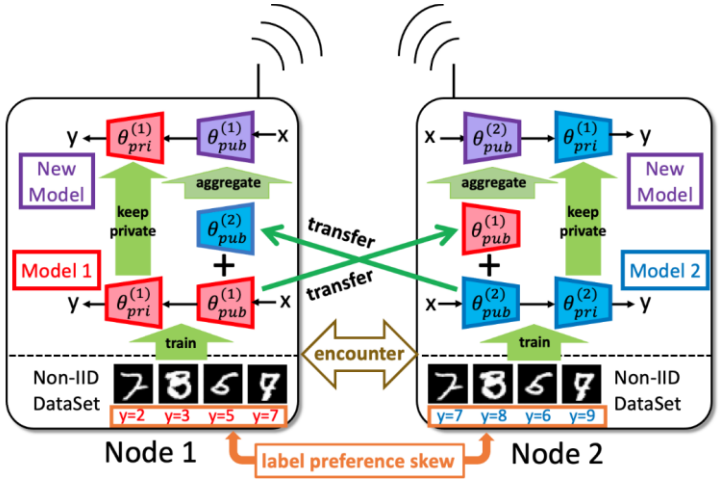
[1] Ryusei Higuchi, Hiroshi Esaki, Hideya Ochiai, "Collaborative Multi-Task Learning across Internet Edges with Device-to-Device Communications“, IEEE Cybermatics Congress, 2023 (under review).

[2] Eisuke Tomiyama, Hiroshi Esaki, Hideya Ochiai, "WAFL-GAN: Wireless Ad Hoc Federated Learning for Distributed Generative Adversarial Networks", IEEE International Conference on Knowledge and Smart Technology, 2023.

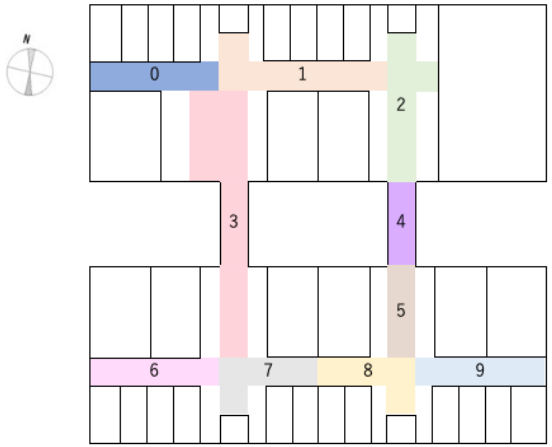
[3] Hideya Ochiai, Riku Nishihata, Eisuke Tomiyama, Yuwei Sun, and Hiroshi Esaki, "Detection of Global Anomalies on Distributed IoT Edges with Device-to-Device Communication", ACM MobiHoc AIoT workshop, 2023 (to be presented).

Previous Studies (5/5): Extensions and Variations of WAFL (2/2)

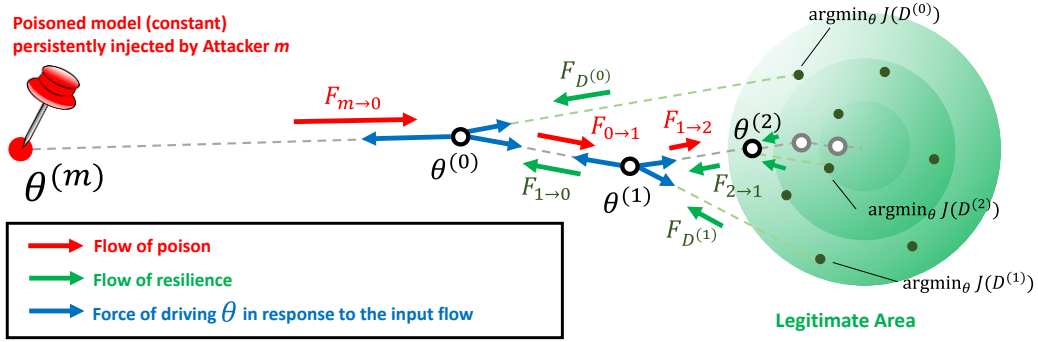
WAFL-Personalization[4]



WAFL-Localization[5]



WAFL-Security[6]



[4] Ryusei Higuchi, Hiroshi Esaki, and Hideya Ochiai, "Personalized Wireless Ad Hoc Federated Learning for Label Preference Skew", IEEE World Forum on Internet of Things, 2023 (accepted).

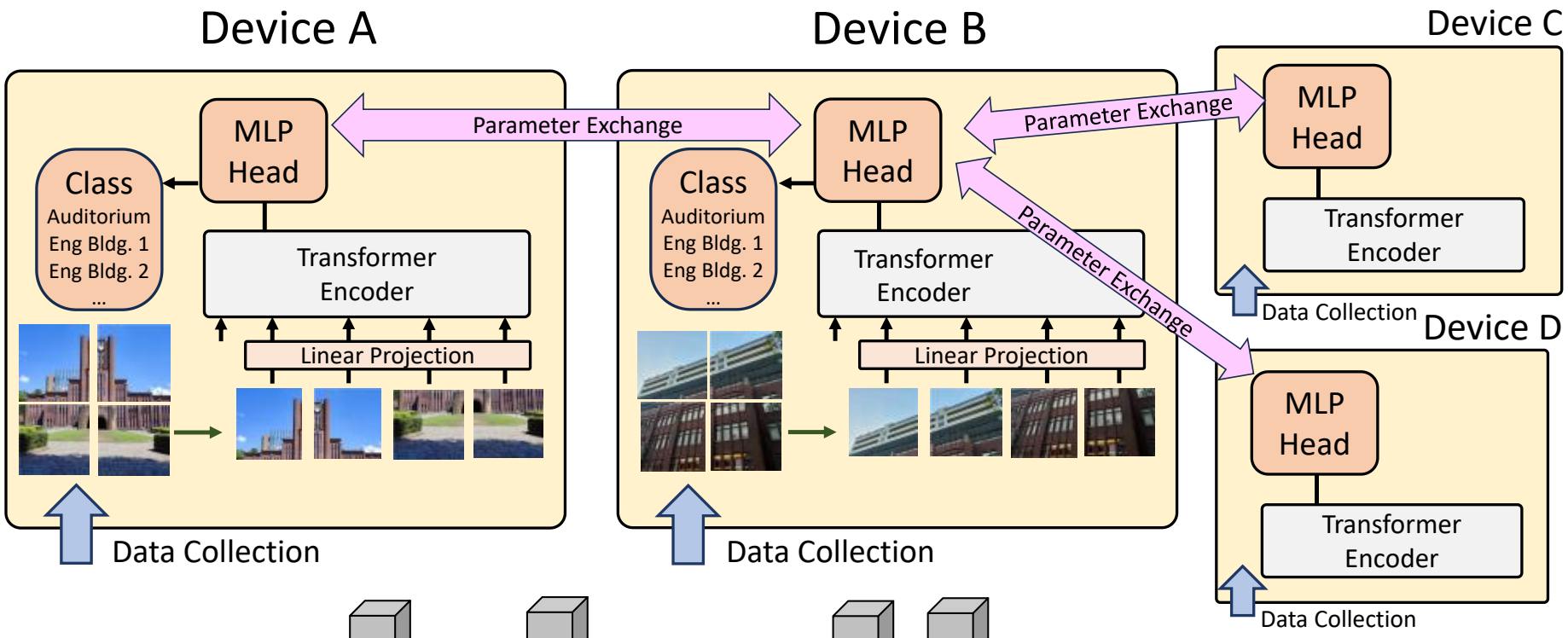
[5] Yusuke Sugizaki, Hideya Ochiai, Muhammad Asad, Manabu Tsukada, and Hiroshi Esaki, "Wireless Ad-Hoc Federated Learning for Cooperative Map Creation and Localization Models", IEEE World Forum on Internet of Things, 2023 (accepted).

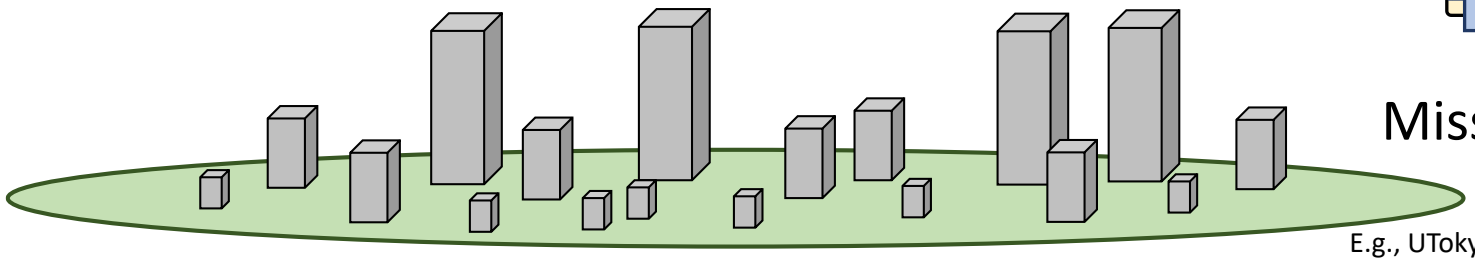
[6] Naoya Tezuka, Hideya Ochiai, Yuwei Sun, Hiroshi Esaki, "Resilience of Wireless Ad Hoc Federated Learning against Model Poisoning Attacks", IEEE International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA), 2022.

For more complex photos, we propose Vision-Transformer-based WAFL (WAFL-ViT).

Table of Contents

Wireless Ad Hoc Federated Learning for Mission-Oriented Image Recognition





Mission-Oriented Task

E.g., UTokyo Campus Building Recognition

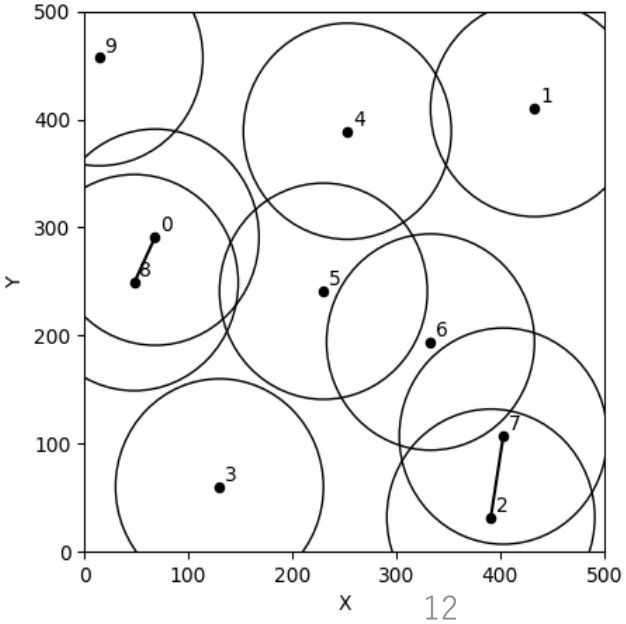
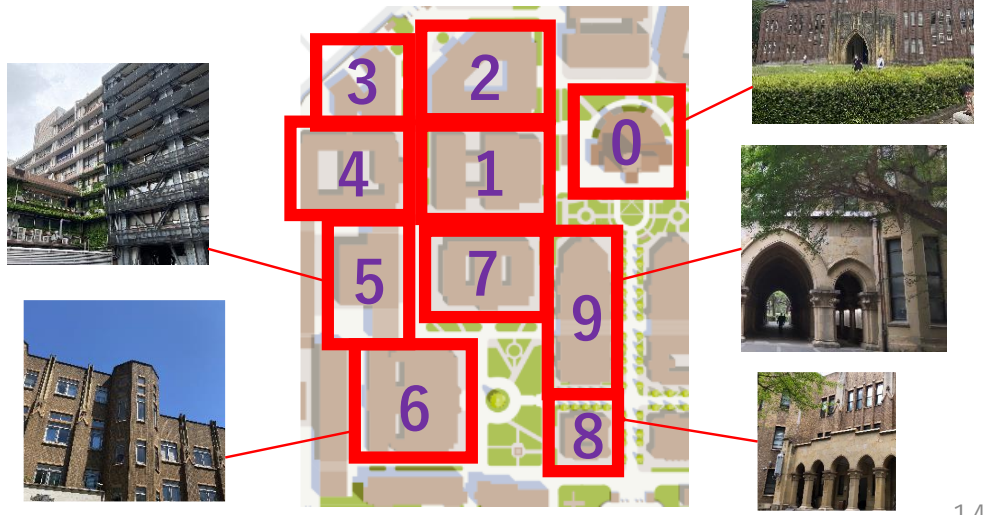


Table of Contents

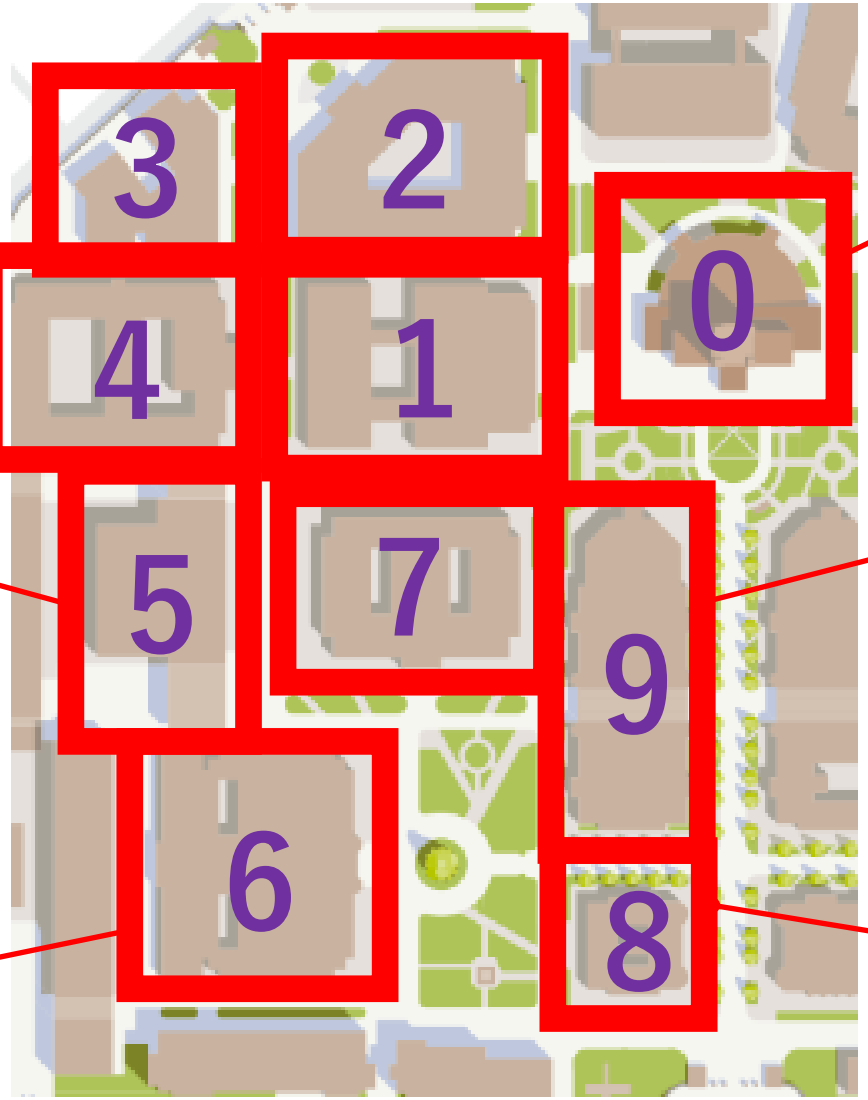
UTokyo Building Recognition Dataset for Distributed Machine Learning Study

- For a Mission-Oriented Image Recognition:
  - We focus on the task of building recognition of the UTokyo Campus.
  - This is a local problem, but such local problems are everywhere.



UTokyo Campus Map (Hongo Campus)

Definition of Building Numbers in the Target Area



0: Yasuda Auditorium

1: Engineering Bldg. 2

2: Engineering Bldg. 3

3: Engineering Bldg. 13

4: Engineering Bldg. 4

5: Engineering Bldg. 8

6: Engineering Bldg. 1

7: Engineering Bldg. 6

8: Reppinkan

9: Law & Letters Bldg. 1

Characteristics of the UTBR Images

- Target buildings were taken:
  - from the front, rear, and sides,
  - sometimes closely, looking up,
  - or from afar, or with a telescopic mode,
  - containing trees, clouds, and the sun.

- These characteristics are not available in MNIST or CIFAR-10.

Examples of Label 0 (Yasuda Auditorium) Photos -- taken from Front, Rear, and Side

Taken from Front

Taken from Rear

Taken from Side

All of these photos are Label 0: Yasuda Auditorium.

Examples of Label 0 (Yasuda Auditorium) Photos -- taken from Afar, or Closely

Taken from Afar

Taken Closely

All of these photos are Label 0: Yasuda Auditorium.

Examples of Label 0 (Yasuda Auditorium) Photos -- taken with the Sun and Trees

Taken with the Sun

Taken with Trees

All of these photos are Label 0: Yasuda Auditorium.

Characteristics of the UTBR Images

All of these photos are Label 0: Yasuda Auditorium.

Taken from Front

Taken from Rear

Taken from Side

Taken with the Sun

Taken from Afar

Taken Closely

Taken with Trees

So many variations (in X) for a single label (Y)

Distribution of the Training Data

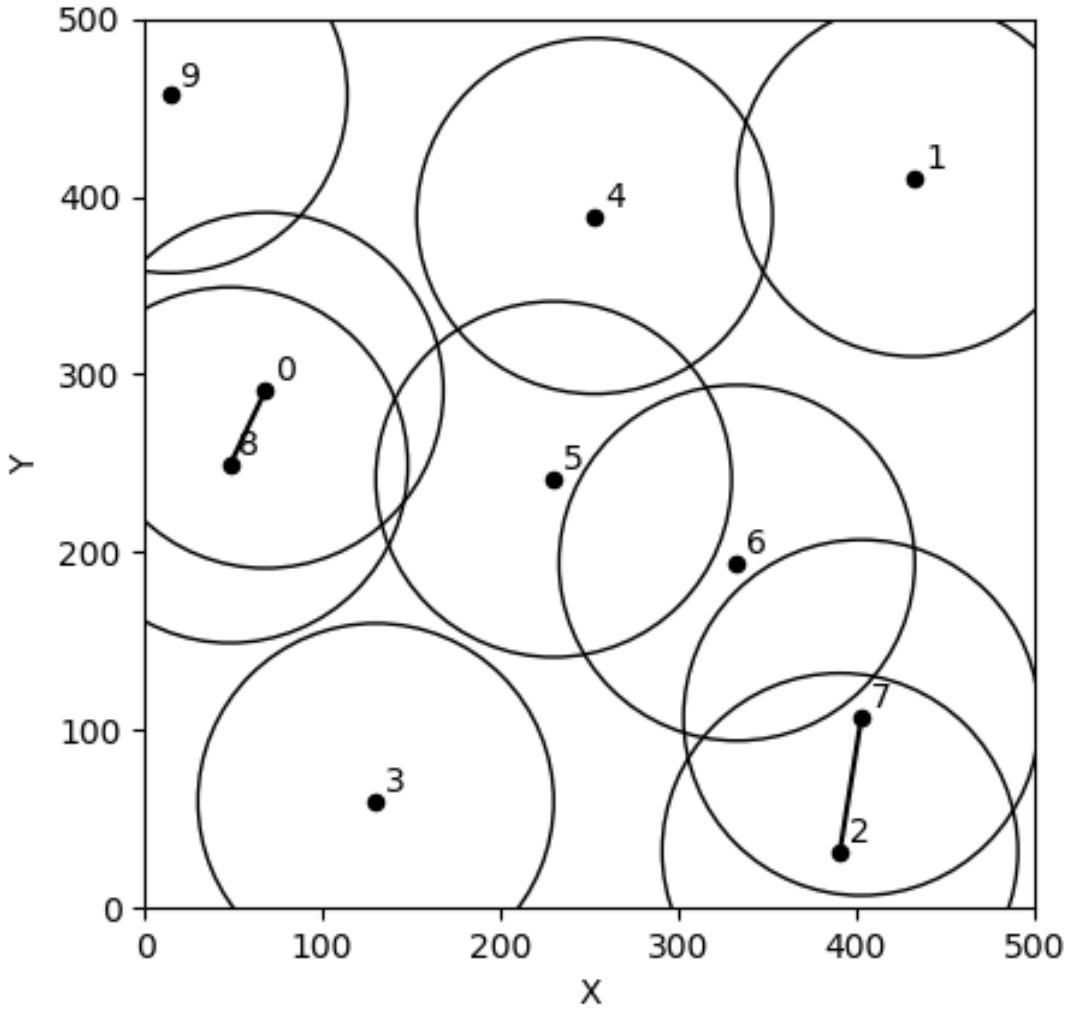
Device	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	Sum
0	11	16	10	10	12	13	12	17	14	20	135
1	11	16	10	10	12	13	12	17	14	20	135
2	11	16	10	10	11	14	12	17	14	20	135
3	11	16	10	9	12	14	12	17	14	20	135
4	11	16	10	9	12	14	12	17	13	21	135
5	11	15	11	9	12	13	13	16	14	21	135
6	10	16	10	10	12	13	12	17	14	21	135
7	10	16	10	10	12	13	12	17	14	21	135
8	10	16	10	10	12	13	12	17	14	21	135
9	10	16	10	10	12	13	12	17	14	20	134

Each device has approximately 10-20 data samples per label,
the dataset of each device may not cover all possible shooting conditions

Table of Contents

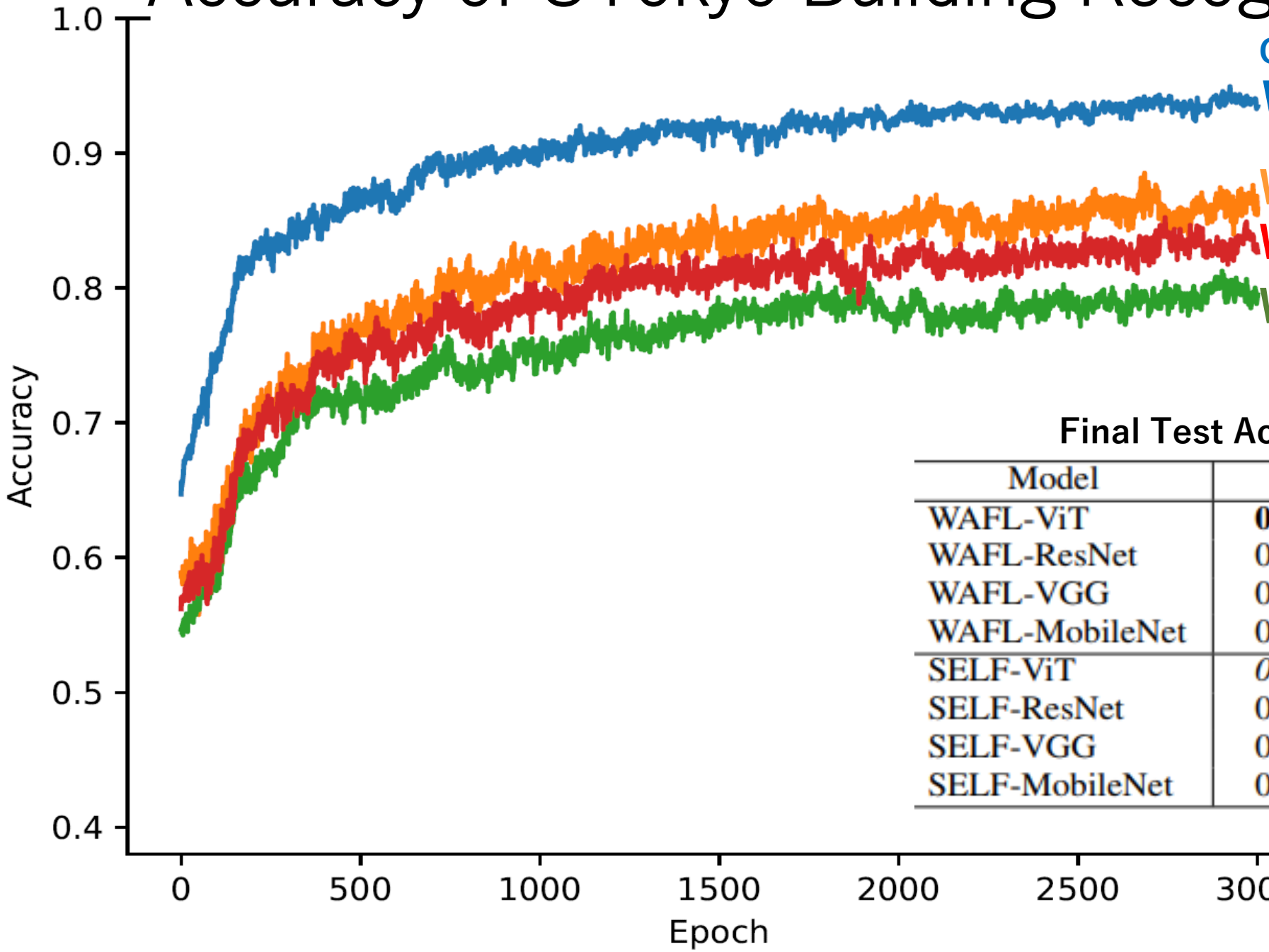
Evaluation: Experiment Settings

- ML model (with pre-trained parameters)
  - Vision Transformer-B/16
  - ResNet-152
  - MobileNet-V2
  - VGG-19-BN
- Dataset
  - UTokyo Building Recognition (UTBR)
- Mobility Pattern
  - Random Waypoint Mobility (RWP)
- Simulation
  - We carried out the experiment by simulation on a single computer.



Random Waypoint Mobility

Accuracy of UTokyo Building Recognition



Our Proposal

WAFL-ViT: 0.937

WAFL-ResNet: 0.859

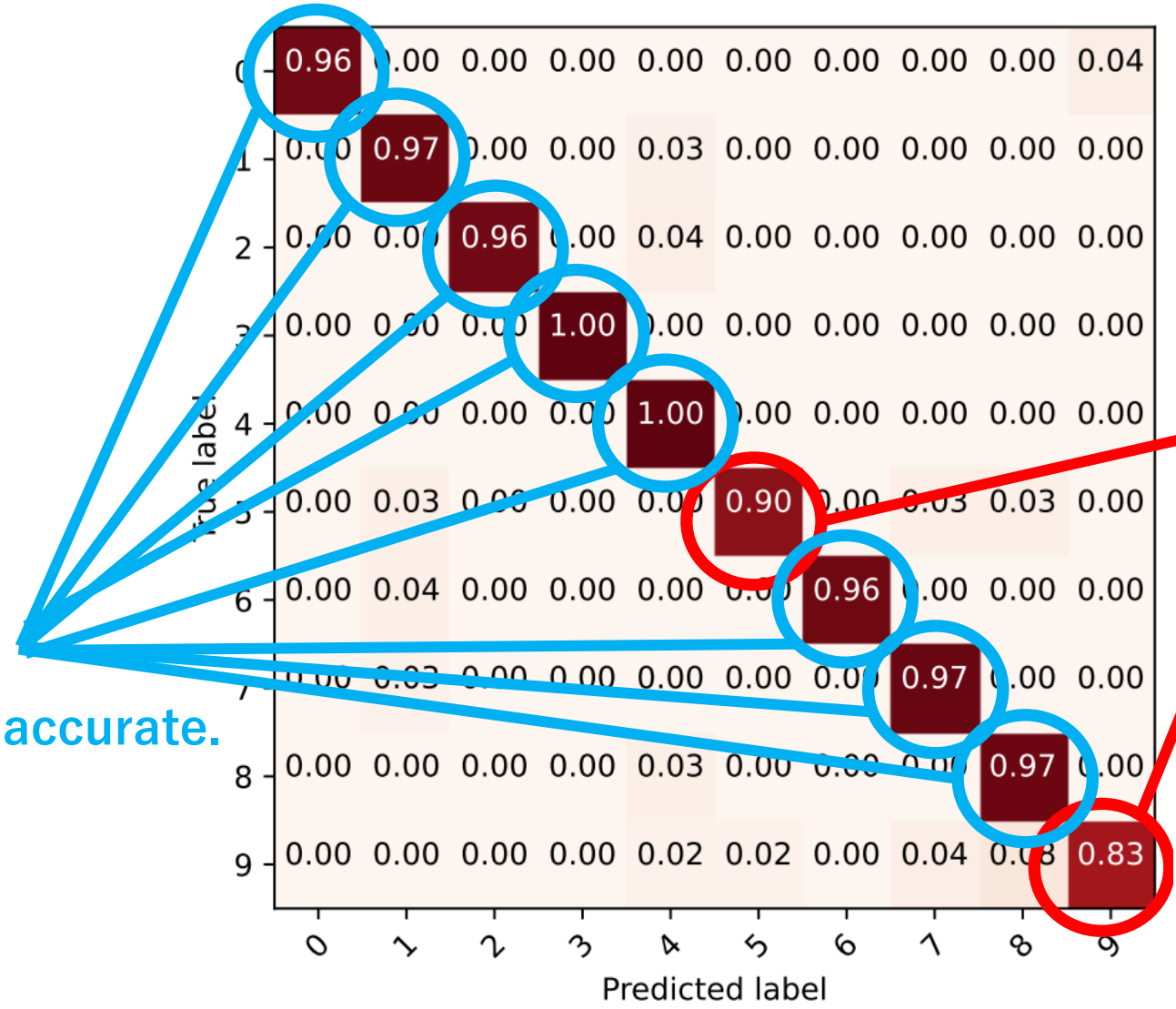
WAFL-MobileNet: 0.829

WAFL-VGG: 0.795

Final Test Accuracy

Model	Accuracy	
WAFL-ViT	**0.937±0.009**	Collaborative Training
WAFL-ResNet	0.859±0.023	
WAFL-VGG	0.795±0.020	
WAFL-MobileNet	0.829±0.021	
SELF-ViT	*0.700±0.028*	Individual Training
SELF-ResNet	0.635±0.021	
SELF-VGG	0.574±0.022	
SELF-MobileNet	0.603±0.029	

Confusion Matrix @ Device 2



Majority were more than 96% accurate.

Some photos of 5 and 9 were misclassified.

The footprints of the models to be exchanged over device-to-device communications (lower better)

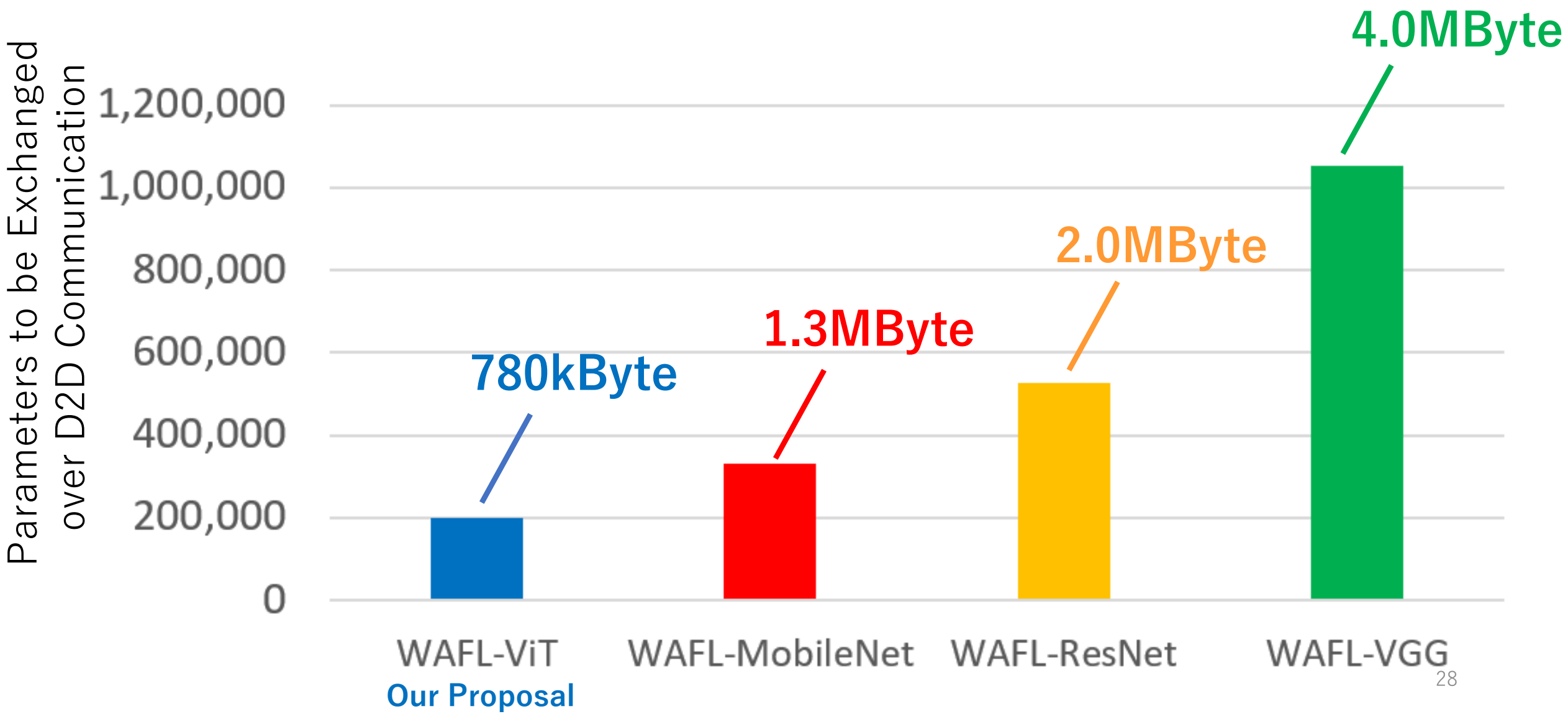


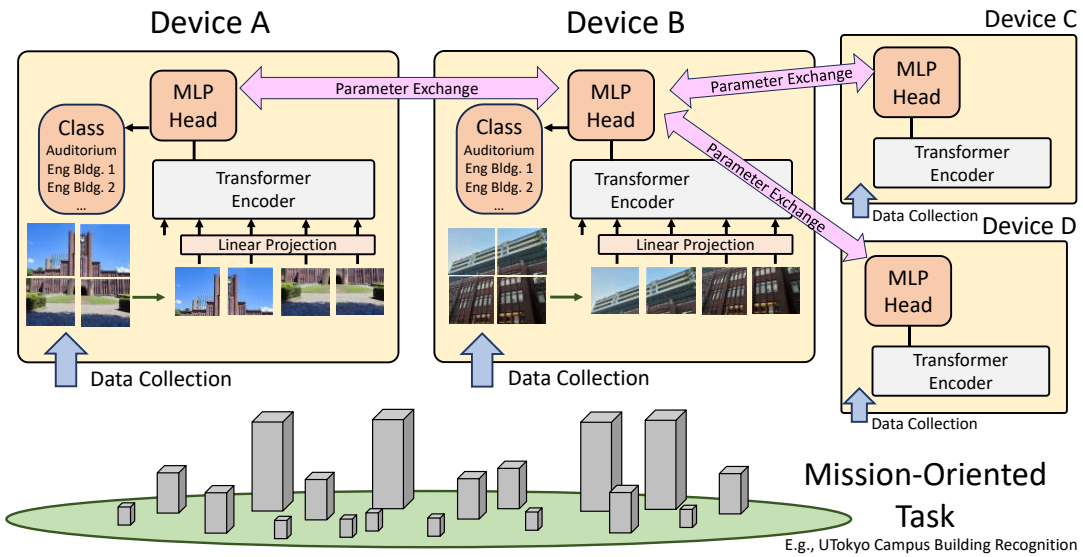
Table of Contents

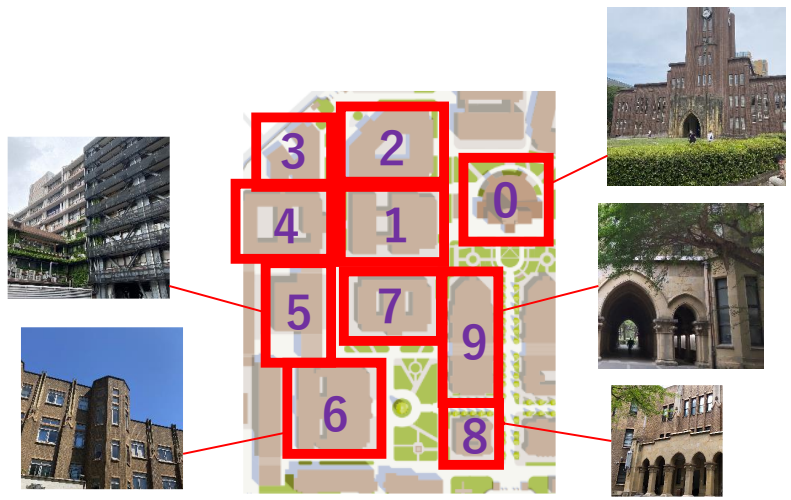
Conclusion

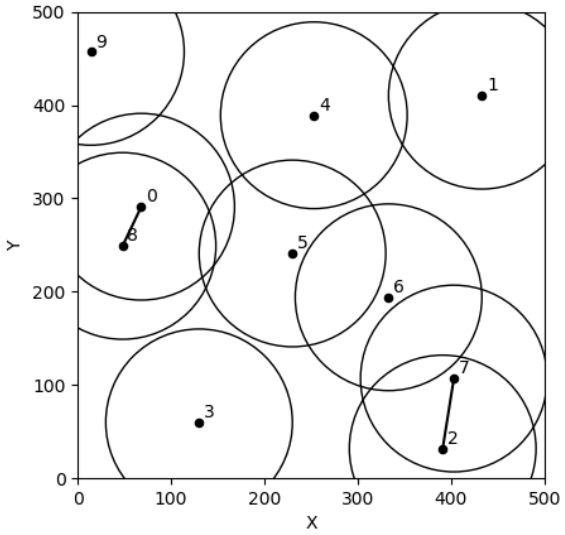
- WAFL-Vision Transformer (WAFL-ViT) is practically useful for image recognition in collaborative learning scenario with D2D communication.

- WAFL-ViT outperformed other image recognition models such as WAFL-ResNet, MobileNet, and VGG, in terms of
  - Prediction Accuracy
  - Computation Load
  - Communication Load

Thank you.







Final Test Accuracy

Model	Accuracy
WAFL-ViT	**0.937±0.009**
WAFL-ResNet	0.859±0.023
WAFL-VGG	0.795±0.020
WAFL-MobileNet	0.829±0.021
SELF-ViT	*0.700±0.028*
SELF-ResNet	0.635±0.021
SELF-VGG	0.574±0.022
SELF-MobileNet	0.603±0.029