중앙대학교
School of Computer Science and Engineering

Advanced Image Processing

Project: Traffic Sign Recognition

Final Report

Team 5

2022-12

Contents

1. Problems and Proposal

- **Intelligent Transportation System (ITS)** has emerged as an important part of a smart city.
- **A traffic sign recognition (TSR)** module has been one of the most vital modules assisting vehicles and devices in ITS.
- TSR requires:

  Accurate and automatic operation(1)

  Preserving privacy among vehicles, devices(2)
- Central ML/AI methods only satisfy the requirement(1).

Reduce traffic congestion

Reduce environmental pollution

Advantages

Figure 1. Intelligent Transportation System (ITS)

ITS Vehicles: Self-driving cars
ITS Devices: CCTVs, intelligent traffic lights

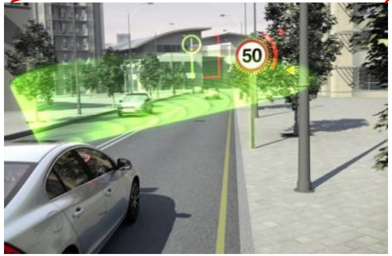


Figure 2. Traffic Sign Recognition (TSR)

1. Problems and Proposal

- Moreover, central ML methods has **drawbacks**:

  - Low preserving privacy

  - The requirement of high performance of computation machines for training process

  - The requirement of massive data storage

  - The requirement of wide bandwidth for data transportation

Because of requirements of the traffic sign recognition in the intelligent transportation system, drawbacks of central ML methods, we provide 3 contributions in the traffic sign recognition topic.

1. Problems and Proposal

Our contributions:

1. The federated learning framework is applied into the TSR module in ITS for maintaining the preserving privacy, guaranteeing an efficient and accurate learning process, and saving resources (bandwidth, storage, energy)
2. We utilize a truly lightweight traffic sign recognition method, LeNet-5[1], which can be compatible with most vehicles and devices with medium or low computational ability in ITS
3. We conduct experiments with the large number of slaves in the federated learning assisted traffic sign recognition.

2. Proposed Traffic Sign Recognition System

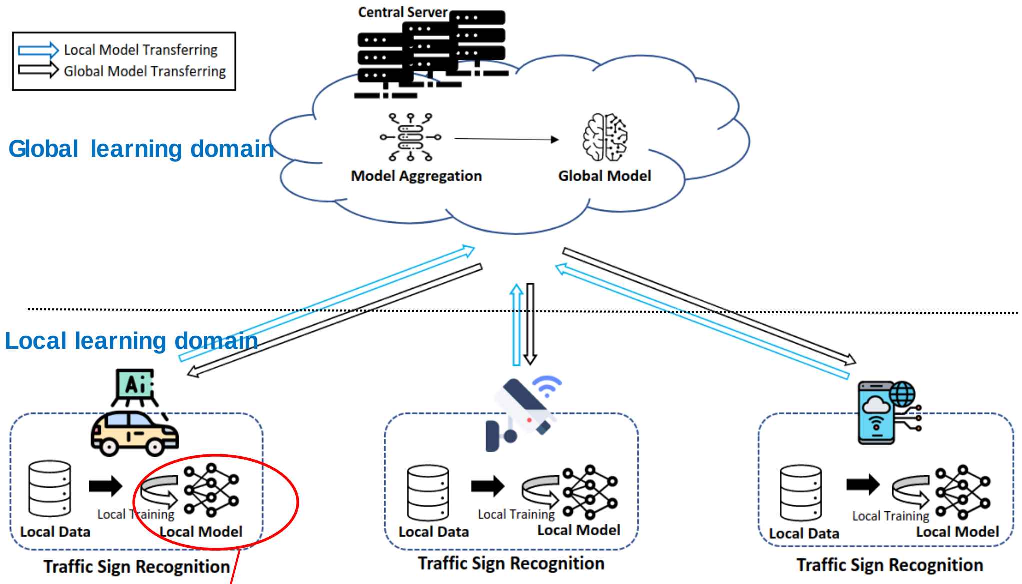


Figure 3. Proposed Traffic Sign Recognition System

LeNet-5 is used as the recognition learning model

- In this proposal for the TSR, we use the **federated learning** framework as the traffic sign recognition learning framework, LeNet-5 as the learning model for the TSR.
- **2 main domains** in this proposal:

Local learning domain:

- Vehicles, devices perform the local training of TSR on their own local data.
- Transfer local models to the global learning domain.
- Receive updated global model from the global learning domain.

Global learning domain:

- Collects local models.
- Aggregates all local models into the global model.
- Broadcasts the global model to the local learning domain.

2. Proposed Traffic Sign Recognition System

2.1. Federated Learning in TSR

- Our federated learning framework follows key steps of the general federated learning process shown in the Fig. 4.

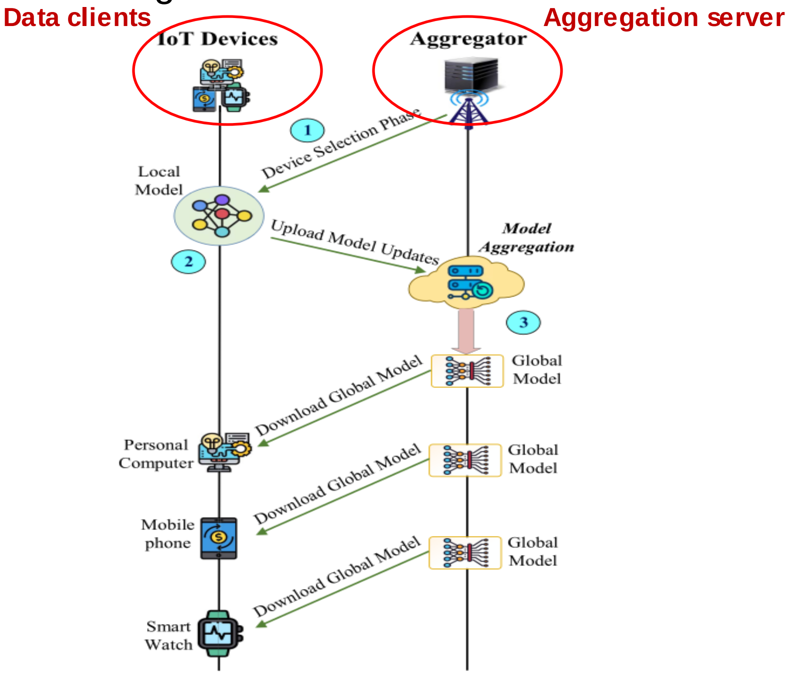


Figure 4. General federated learning process

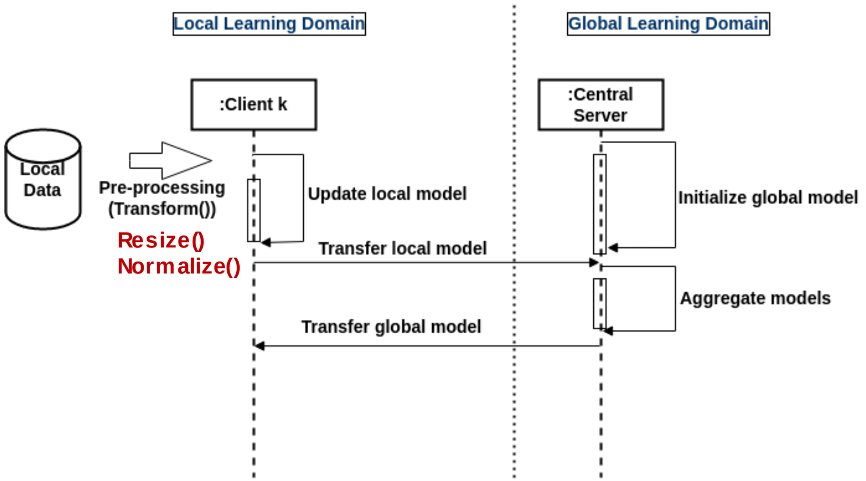


Figure 5. Training process of federated learning-assisted traffic sign recognition

2. Proposed Traffic Sign Recognition System

2.1. Federated Learning in TSR

- We adopt a well-known federated learning method, Federated Averaging [2].

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
  initialize w_0
  for each round $t = 1, 2, \dots$ **do**
    $m \leftarrow \max(C \cdot K, 1)$
    $S_t \leftarrow$ (random set of m clients)
    for each client $k \in S_t$ **in parallel do**
      $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w)**:** // *Run on client* k
  $\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
  for each local epoch i from 1 to E **do**
    for batch $b \in \mathcal{B}$ **do**
      $w \leftarrow w - \eta \nabla \ell(w; b)$
  return w to server

2. Proposed Traffic Sign Recognition System

2.2. LeNet-5 in TSR

- We utilize LeNet-5[1] as the recognition learning model. This is a lightweight pre-trained model, which can be suitable for medium, low computational capacity of data clients in the local learning domain.

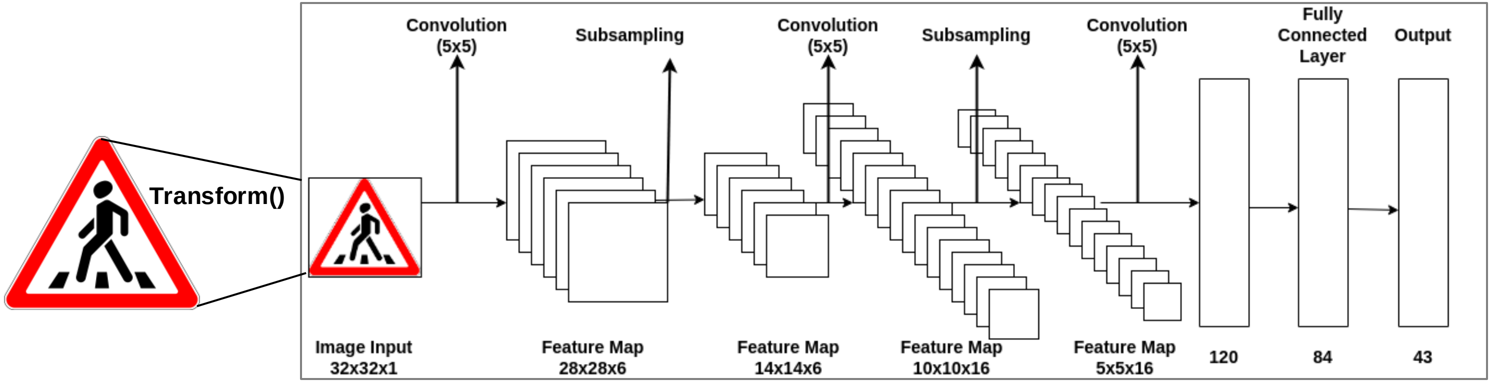


Figure 6. LeNet-5 architecture for traffic sign recognition

2. Proposed Traffic Sign Recognition System

2.2. LeNet-5 in TSR

- The detail of the LeNet-5 architecture:

Table 1. LeNet-5 architecture

Layer	#Filters/ Neurons	Filter size	Stride	Size of feature map	Activation function
Input	-	-	-	32x32x1	
Conv 1	6	5*5	1	28x28x6	ReLU
Avg.pooling 1		2*2	2	14x14x6	
Conv 2	16	5*5	1	10x10x16	ReLU
Avg.pooling 2		2*2	2	5x5x16	
Conv 3	120	5*5	1	120	ReLU
Fully Connected 1	-	-	-	84	ReLU
Fully Connected 2	-	-	-	43	Softmax

2. Proposed Traffic Sign Recognition System

2.3. Transform() in Torchvision

Transform()

Composes of

Resize()

This transform will resize the input image to the given size by using a non-adaptive interpolation method, **bilinear interpolation**.
The interpolation works by using known data to estimate values at unknown points.

Normalize()

This transform will normalize the input in each channel of images.

Output[channel] = (input[channel] - mean[channel])/ std[channel]

2. Proposed Traffic Recognition System

2.3. Transform() in Torchvision

Interpolation

- Image interpolation tries to gain a best approximation of a pixel's intensity based on the values at surrounding pixels.

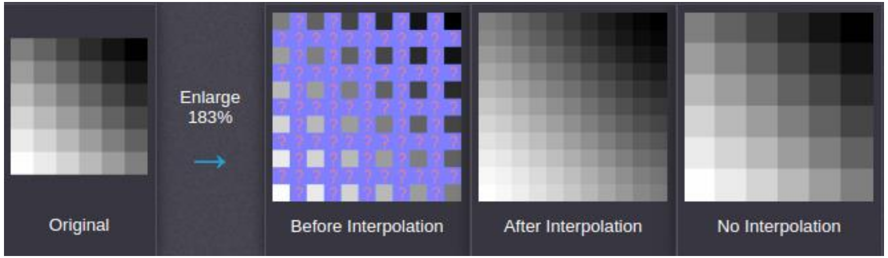


Figure 7. An example of resizing an image by using interpolation

- Interpolation algorithms can be grouped into two categories:

Adaptive algorithms: They change depending on what they are interpolating (sharp edges, smooth textures, .etc)

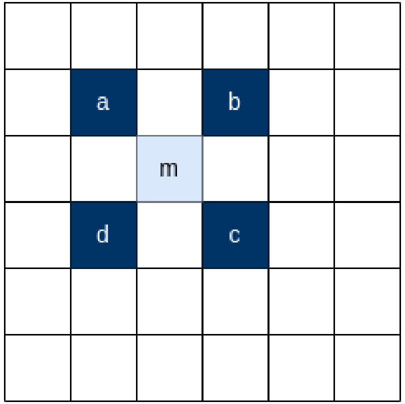
Non-adaptive algorithms: They treat all pixels equally. **Bilinear interpolation** is a non-adaptive algorithm.

2. Proposed Traffic Recognition System

2.3. Transform() in Torchvision

Bilinear Interpolation

- Considers the closest 2x2 neighbourhood of known pixel values surrounding the unknown pixel.
- It takes a weighted average of 4 pixel values as its final interpolated value.



$m = (a + b + c + d)/4$

In this example, distances between the interpolated pixel and the known pixels are equal, so their weights are the same.
When distances are different, weights are different.

Figure 8. An example of bilinear interpolation

3. Experiments

3.1. Experimental Setup

Dataset: German Traffic Sign Recognition Benchmark (GTSRB) [3,4]

· It was created from about 10 hours of video recorded while driving in Germany.

· It consists of about 40.000 colorful photos of traffic signs.

· Their sizes vary from 15x15 to 250x250 pixels.

Tools:

· Coding language: Python

· Libraries/ modules: TensorFlow, NumPy, Matplotlib, Torchvision, Models

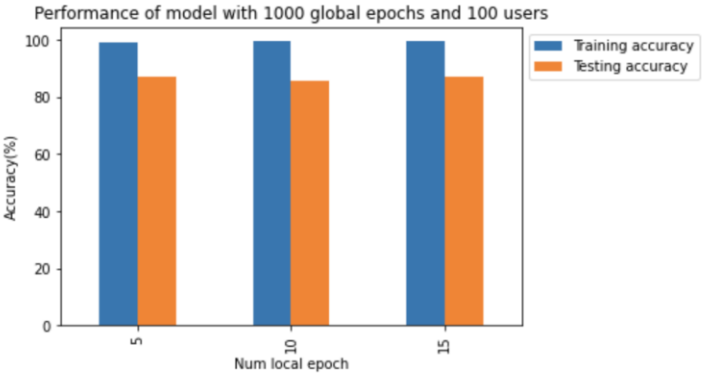
Experimental parameters:

Parameter	Value
K: The number of clients	100, 200
B: Local minibatch size	32
E: The number of local epochs	5, 10, 15
T: The number of epochs	100, 500, 1000
η:Learning rate	0.001

3. Experiments

3.2. Experimental Results

There is no big gap between training accuracy and testing accuracy, when the number of epochs is large enough!



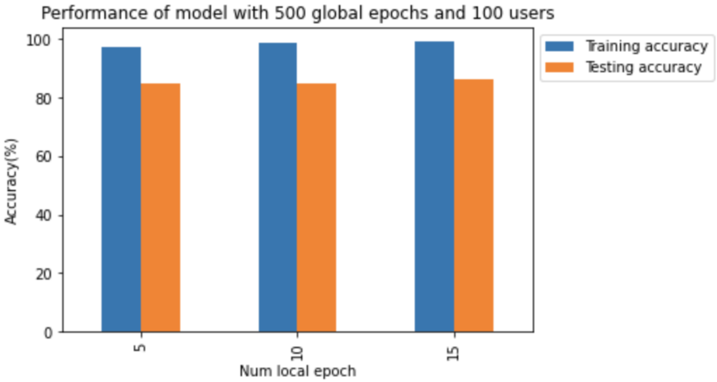


Figure 9. Training and testing accuracy with K = 100, T = 500, 1000

3. Experiments

3.2. Experimental Results

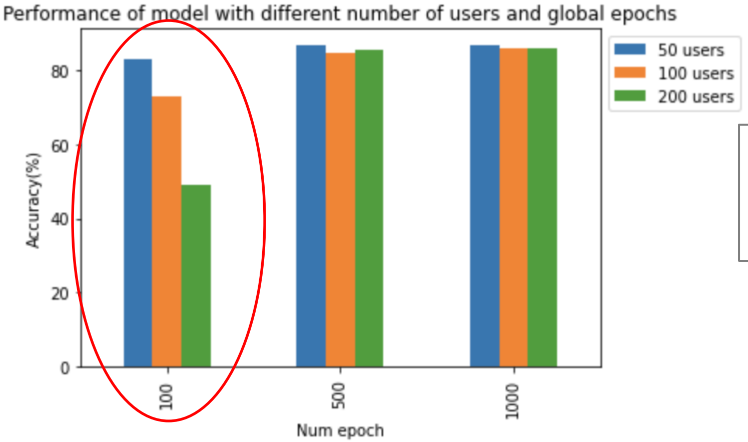


Figure 10. Performance with different number of users and global epochs, K = 50, 100, 200, T = 100, 500, 1000

When the number of global epochs in training is not large enough, the accuracy is affected by the number of users!

3. Experiments

3.2. Experimental Results

The higher the number of local epochs is, the faster the model converges.

The higher the number of users is, the more slowly the model converges.

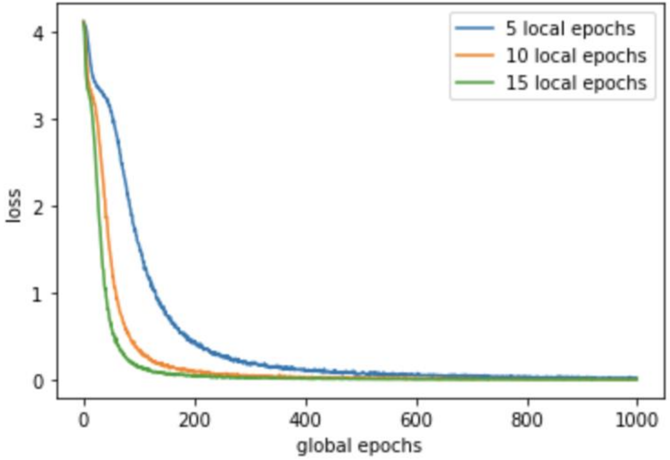


Figure 11. Convergence with T = 1000, E = 5, 10, 15

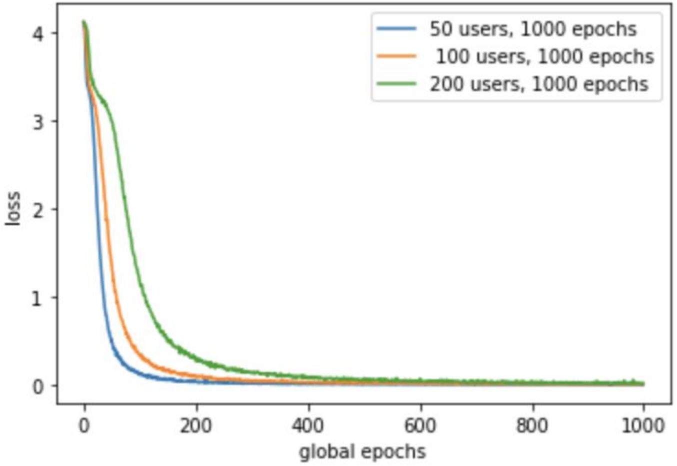


Figure 12. Convergence with T = 1000, K = 50, 100, 200

4. Conclusion

1. Because of the shortage of distributed devices, we simulated the federated learning framework via a provided open-source library. Our experiment could not be highly persuasive. We hope this system can be deployed on real machines in future.
2. We have not provided proof and experimental results explaining two main advantages of the federated learning framework, which are preserving privacy and efficient resource maintaining. We hope we can show it in future.
3. Conventional image processing techniques and modern image processing techniques were not clearly analyzed, compared in our proposal. We hope we can apply them, and dive deeply into them in future.

References

[1] LeCun, Yann et al. "Gradient-based learning applied to document recognition." *Proc. IEEE* 86 (1998): 2278-2324.

[2] McMahan, H. B. et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data." *International Conference on Artificial Intelligence and Statistics* (2016).

[3] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The german traffic sign recognition benchmark: A multi-class classification competition," in *Proc. IJCNN*, 2011, pp. 1453–1460. [Online]. Available: http://benchmark.ini.rub.de/?section=gtsrb

[4] https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign