Federated Learning with Autoencoders for Image Classification in IoT Environments

Authors: André Gonçalves
Bruno Olivieri
Markus Endler

LAC Laboratory for Advanced Collaboration
DI Department of Informatics
PUC-Rio Pontifical Catholic University of Rio de Janeiro

Laboratory for Advanced Collaboration
DEPARTAMENTO DE INFORMÁTICA PUC-RIO
PUC RIO

Current Challenges in IoT

Privacy & Resource Constraints

- Traditional centralized approaches face multiple challenges:
  - Data privacy concerns when transmitting sensitive information
  - High communication costs for continuous data transmission
  - Battery drain from constant data uploads
  - Limited bandwidth in IoT networks

Proposed Solution

Federated Learning + IoT

Use a non-supervisioned approach for image classification.

- Supervised learning limitations:
  - Expensive and time-consuming labeling process
  - Often impractical in real-world IoT deployments
  - Need for continuous data updates

Proposed Solution

Key Benefits

Privacy preservation through local processing

- Reduced communication overhead
- No requirement for labeled data
- Scalable architecture

Proposed Solution

System Overview

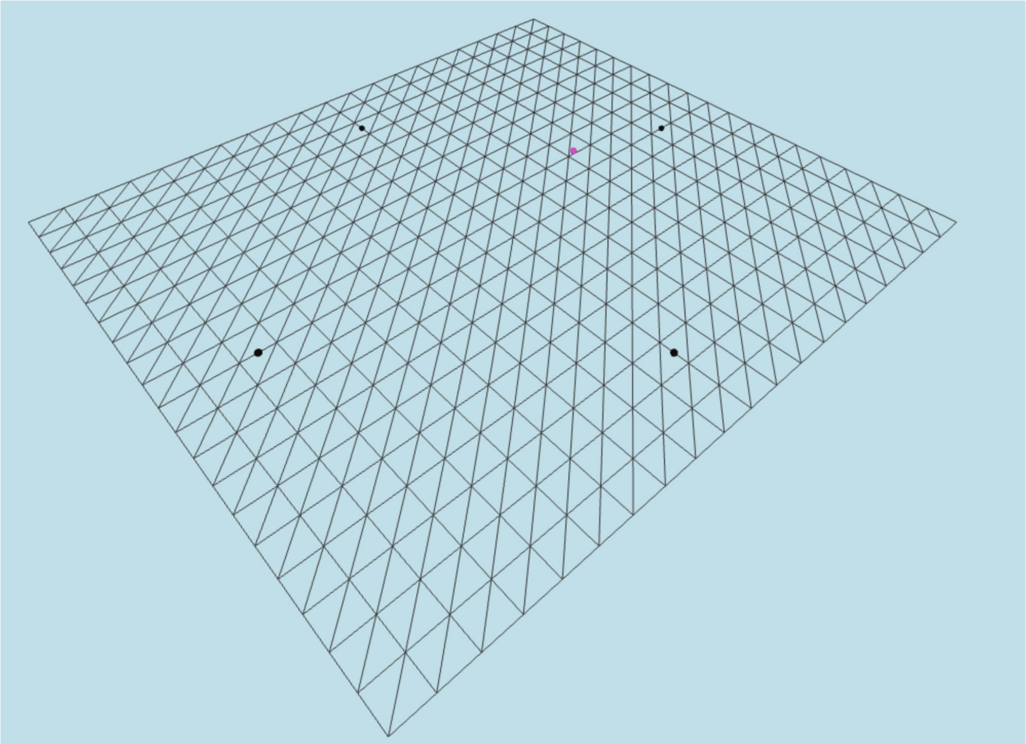
Integration of three concepts:

1. **IoT Sensors:** Collect raw data and train autoencoder models locally.
2. **UAVs:** Collect trained models, aggregate them into a global model, and redistribute the updated global model.
3. **Autoencoder:** Encoder compresses data, decoder reconstructs data, classification head performs classification.

Implementation Specifics

Experimental Setup

- Environment Configuration:
  - GrADyS-SIM NG simulator
  - Grid size: 200×200 units
  - 4 sensors at fixed coordinates
  - UAV communication range: 30 units



Technical Architecture

Data Distribution

- Dataset: CIFAR-10
  - Equally divided among 4 sensors
  - Each sensor processes unique data subset

Protocol Implementation

- Communication Protocol
  - Model Update Request from UAV
  - Local Model Updates from Sensors
  - Global Model Distribution by UAV
  - Quantization and compression before transmission

Technical Architecture

Network Design

- Traditional centralized approaches face multiple challenges:
  - Three-component architecture:
    1. Encoder Network:
       1. Input: 32×32×3 images
       2. Two convolutional layers with batch normalization
       3. Output: 8×8×64 latent representation
    2. Decoder Network:
       1. Input: 8×8×64 latent space
       2. Two transposed convolutional layers
       3. Output: 32×32×3 reconstructed image
    3. Classification Head:
       1. Processes latent representation
       2. Two fully connected layers
       3. Output: Class probabilities

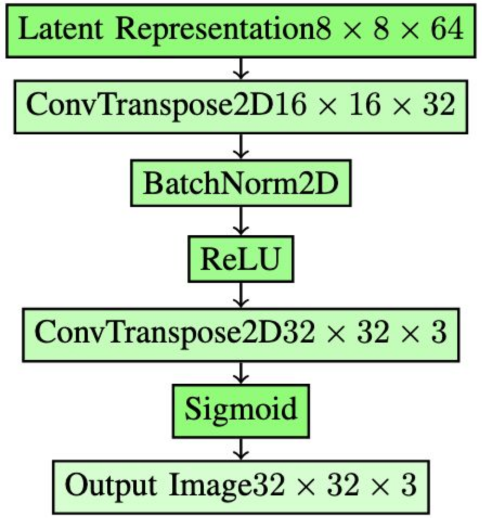


Fig. 3: Decoder Network Layers

Implementation Specifics

Optimization Methods

- Model Size Reduction:
  - Quantization: 74.4% size reduction
    - Autoencoder: 2.197MB → 0.562MB
    - Supervised model: 2.415MB → 0.619MB
  - Gzip compression for transmission

Parameter	**Autoencoder**	**Supervised Model**
Training Approach	Unsupervised (Autoencoder)	Supervised (Direct Classification)
Number of Training Cycles	80	80
Duration per Run (seconds)	15,000	15,000
Learning Rate	0.001	0.001
Batch Size	32	32
Evaluation Metrics	MSE, ARI, Accuracy, Clustering Accuracy, Confusion Matrix	Loss, Accuracy, ARI, Clustering Accuracy, Confusion Matrix

Results Analysis

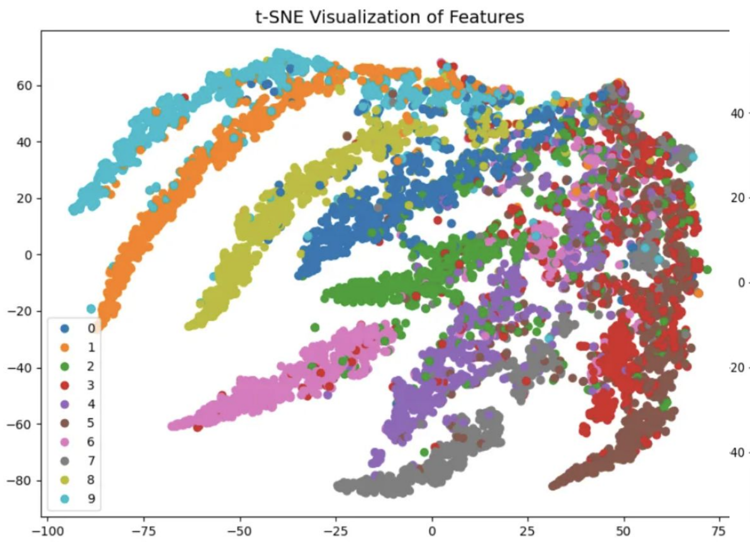
Clustering Accuracy

Autoencoder Model

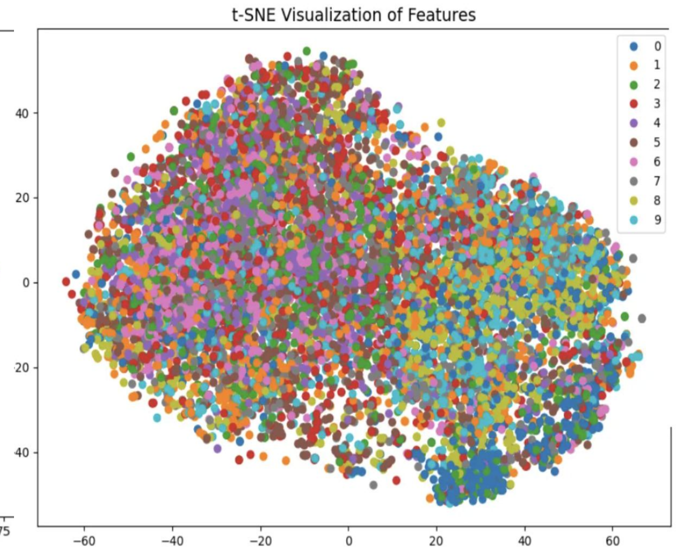
- Clustering accuracy: 19.75%

Supervised Model

- Clustering accuracy: 27.42%



Supervised



Autoencoder

Results Analysis

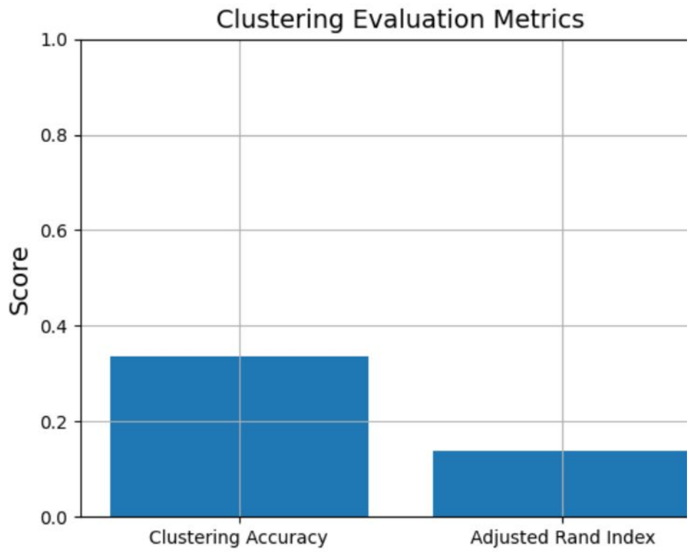
Overall Accuracy

Autoencoder Model

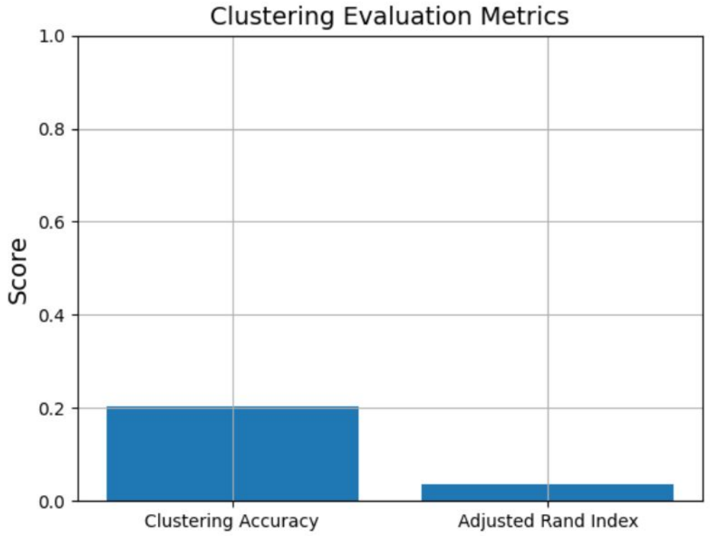
- Classification accuracy: 74.97%
- Mean reconstruction loss: 0.2618

Supervised Model

- Classification accuracy: 82.4%



Supervised



Autoencoder

Strengths

- Handles unlabeled data effectively, making it suitable for scenarios where labeling is costly or impractical.
- Reduces communication overhead by transmitting compressed representations instead of raw data
- Preserves data privacy by keeping raw data on devices and sharing only model updates
- Efficiently extracts meaningful features from data, even with limited labeled data, enabling effective unsupervised learning.

Limitations

- Generally lower classification accuracy compared to supervised models, especially when abundant labeled data is available for training the supervised model.

- Clustering accuracy may be limited, suggesting that extracted features might not be sufficiently discriminative for optimal clustering performance.

- The primary focus on reconstruction might lead to a trade-off with classification performance, requiring careful consideration in applications where classification is the primary goal.

Conclusions

- Autoencoders can effectively extract meaningful features from image data in an unsupervised manner.
- Autoencoder-based approach significantly reduces communication overhead compared to traditional supervised learning.
- The proposed system enhances data privacy by keeping raw image data on local devices.
- While the supervised learning model achieved higher classification accuracy (82.4%), the autoencoder-based approach offers a viable alternative when labeled data is scarce or unavailable.
- The relatively low clustering accuracy of both models suggests that the extracted features might not be optimally discriminative for clustering tasks.

Future Directions

- Explore hybrid models combining autoencoders and supervised learning.
- Advanced clustering algorithms for improved class separation.
- Optimize data transmission protocols (quantization, compression).
- Develop robust training for non-IID data distributions.
- Ensure scalability and energy efficiency for larger IoT networks.

Acknowledgments & Contact

- Research supported by AFOSR grant FA9550-23-1-0136

- Contact: {agoncalves,bolivieri,endler}@inf.puc-rio.br
- Departamento de Informática, PUC-Rio

Questions?