Decentralized DNN Architectures

A. Kaimakamidis, Prof. I. Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Artificial Intelligence &
Information Analysis Lab

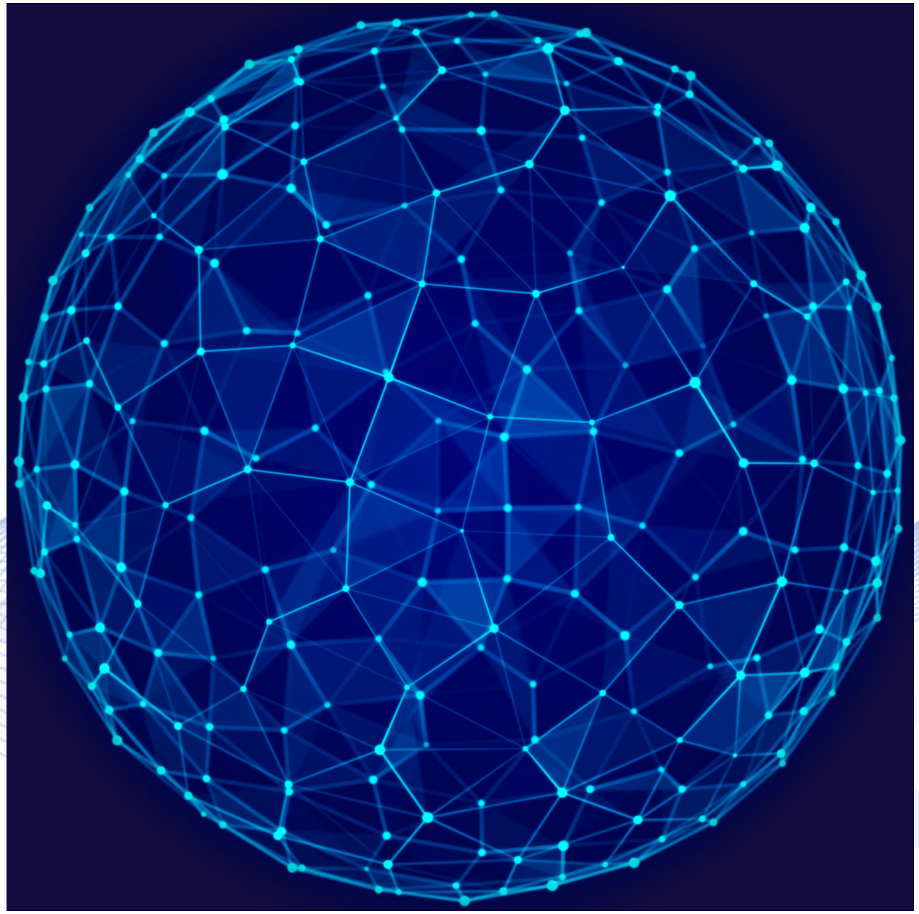
Decentralized DNN Architectures

- **Decentralized DNN Architectures**
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Decentralized DNN Architectures

Definition:

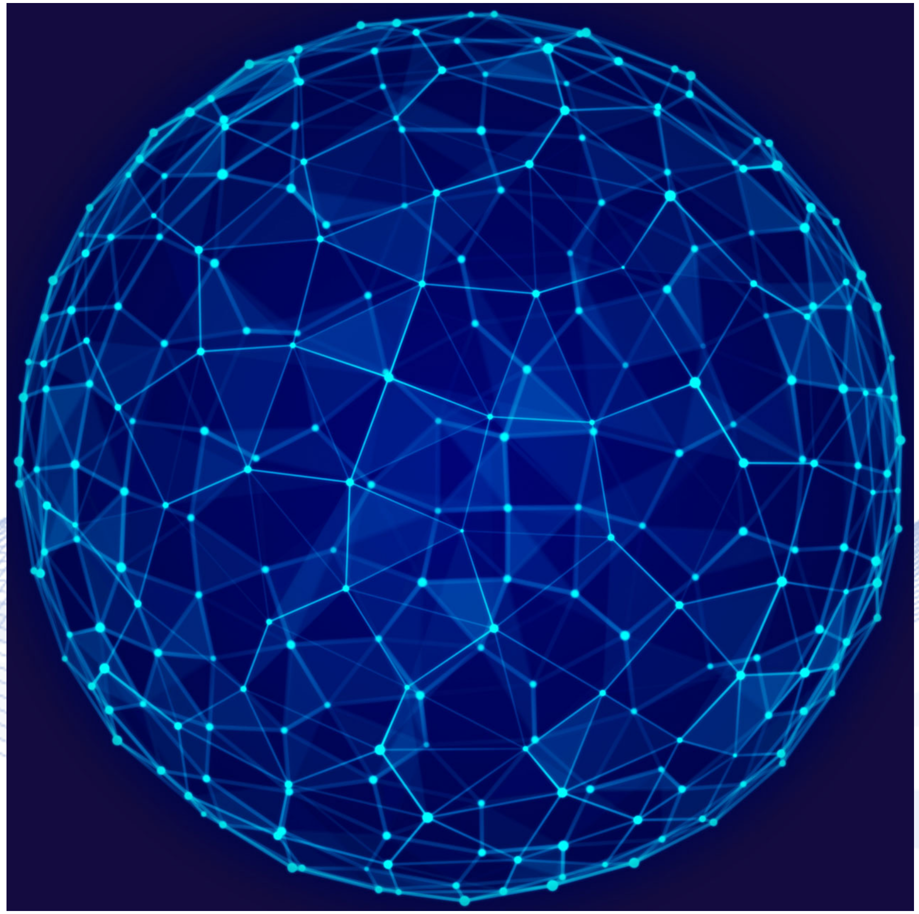
Decentralized Deep Neural Network (DNN) architectures distribute computation and decision-making across multiple nodes or devices, offering advantages in scalability, privacy, and robustness.



Decentralized DNN Architectures

Characteristics:

- Distribution: Computation and data are spread across multiple nodes or devices.
- Collaboration: Nodes cooperate to train or execute models.
- Privacy Preservation: Data remains localized, enhancing privacy and security.
- Fault Tolerance: Resilience to individual node failures or attacks.

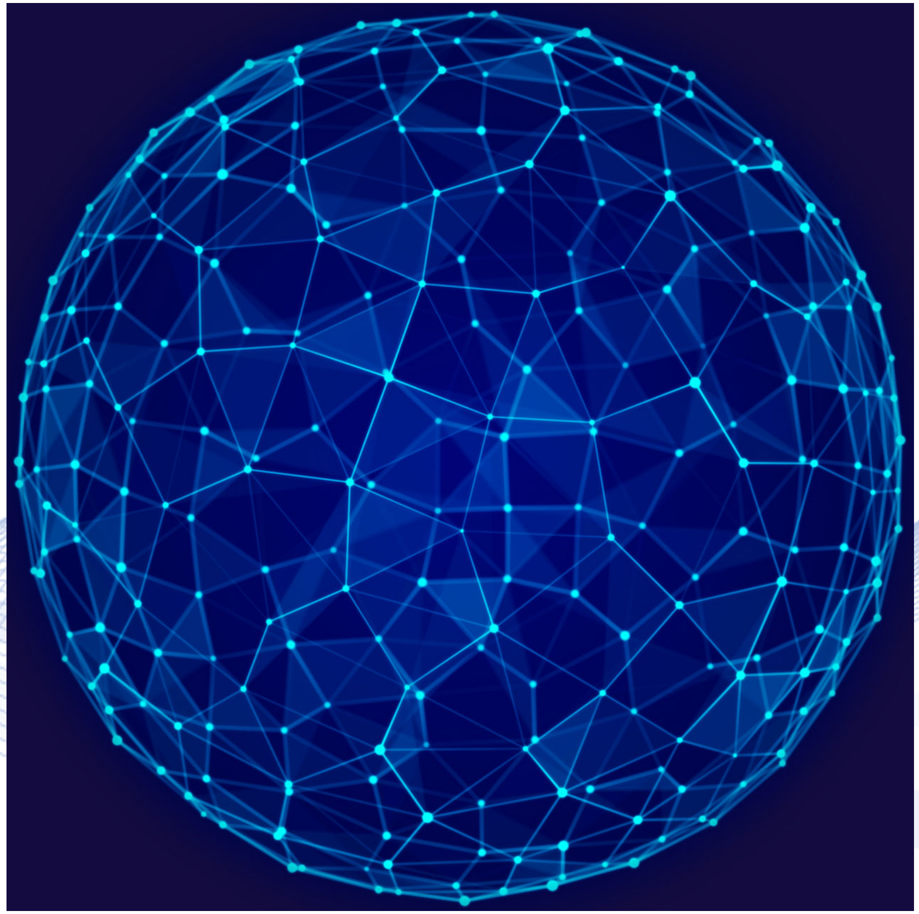


1. Types:
   1. Federated Learning: Training a global

Decentralized DNN Architectures

Types:

1. Federated Learning: Training a global model across decentralized devices while keeping data on-device.
2. Edge Computing: Running inference or lightweight training directly on edge devices.
3. Peer-to-Peer Networks: Collaborative learning among peers without a central server.



Federated Learning

- Decentralized DNN Architectures
  - **Federated Learning**
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Federated Learning

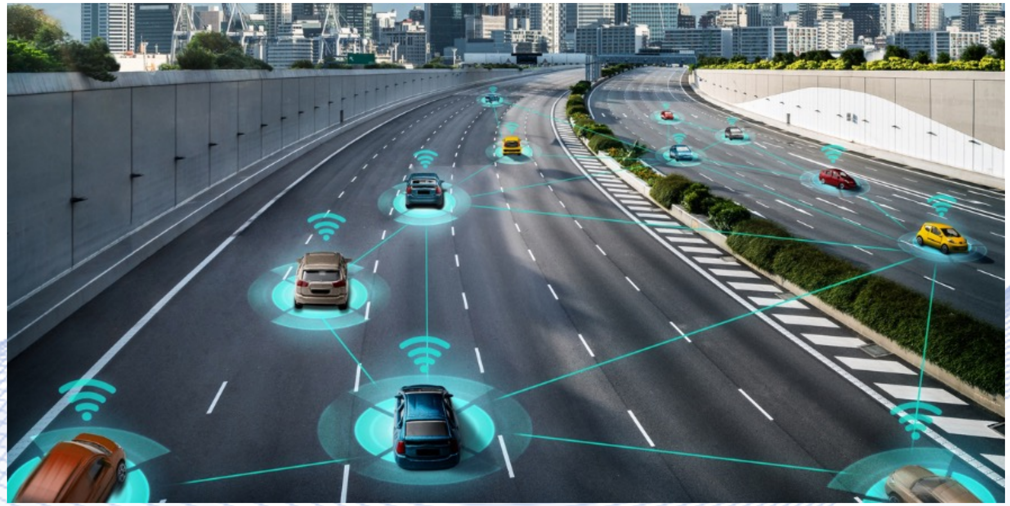
- Privacy Preservation: Data remains on local devices, ensuring privacy.
- Efficiency: Reduces the need to transfer large volumes of data to a central server.
- Scalability: Suitable for large-scale distributed systems with diverse data sources.
- Adaptability: Can accommodate non-IID (non-identically distributed) data across devices.

Edge Computing

- Decentralized DNN Architectures
  - Federated Learning
  - **Edge Computing**
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Edge Computing

- Low Latency: Enables real-time decision-making without reliance on distant servers.
- Bandwidth Efficiency: Reduces the need to transfer large volumes of data to central servers.
- Privacy Preservation: Sensitive data can be processed locally, enhancing privacy.
- Offline Capability: Allows for operation in disconnected or low-connectivity environments.

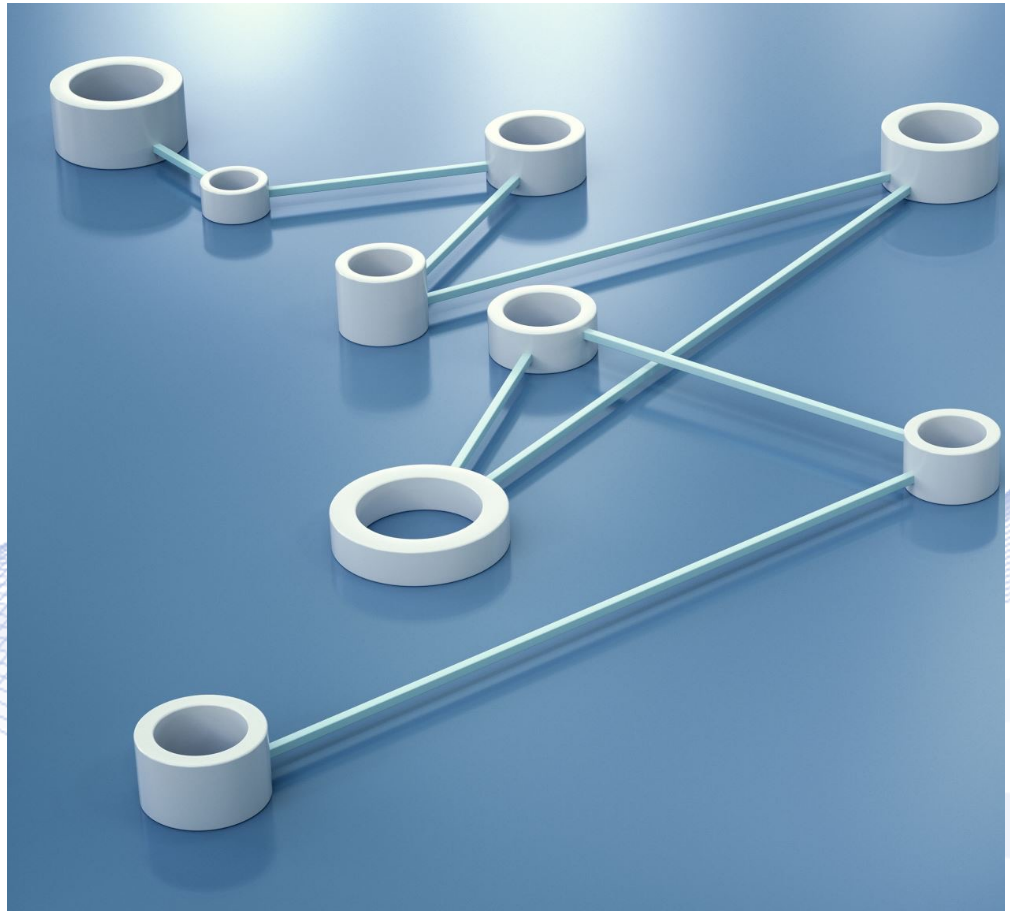


Peer-to-Peer Networks

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - **Peer-to-Peer Networks**
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Peer-to-Peer Networks

- Decentralization: Reduces dependency on central servers, enhancing scalability and robustness.
- Resource Efficiency: Utilizes idle computational resources across peers.
- Resilience: Resilient to node failures or attacks due to distributed nature.
- Community-driven Innovation: Facilitates collaborative research and knowledge exchange.



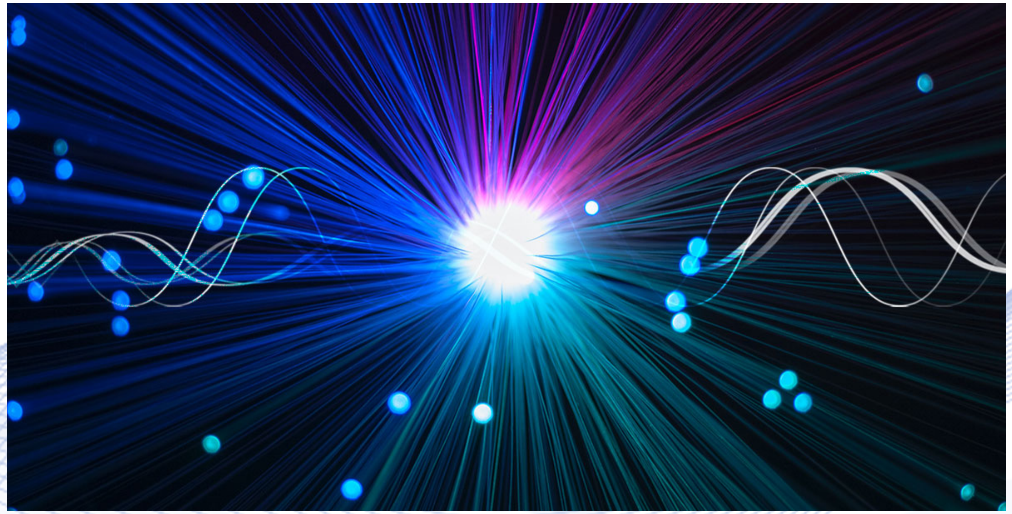
Knowledge Distillation

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- **Knowledge Distillation**
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Knowledge Distillation

Definition:

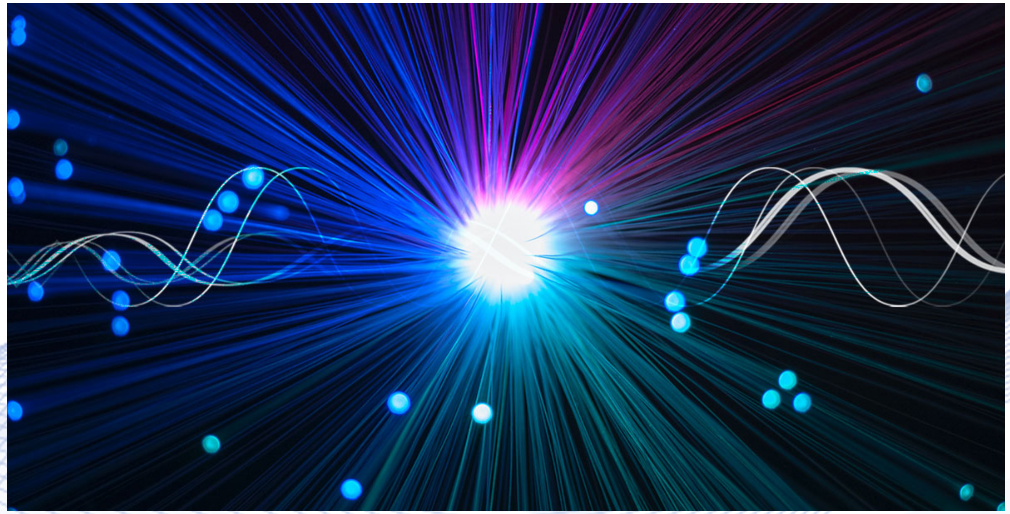
Knowledge Distillation is a technique in machine learning where a compact model, known as the student model, learns from a larger, more complex model, referred to as the teacher model, by mimicking its outputs or internal representations.



Knowledge Distillation

Process:

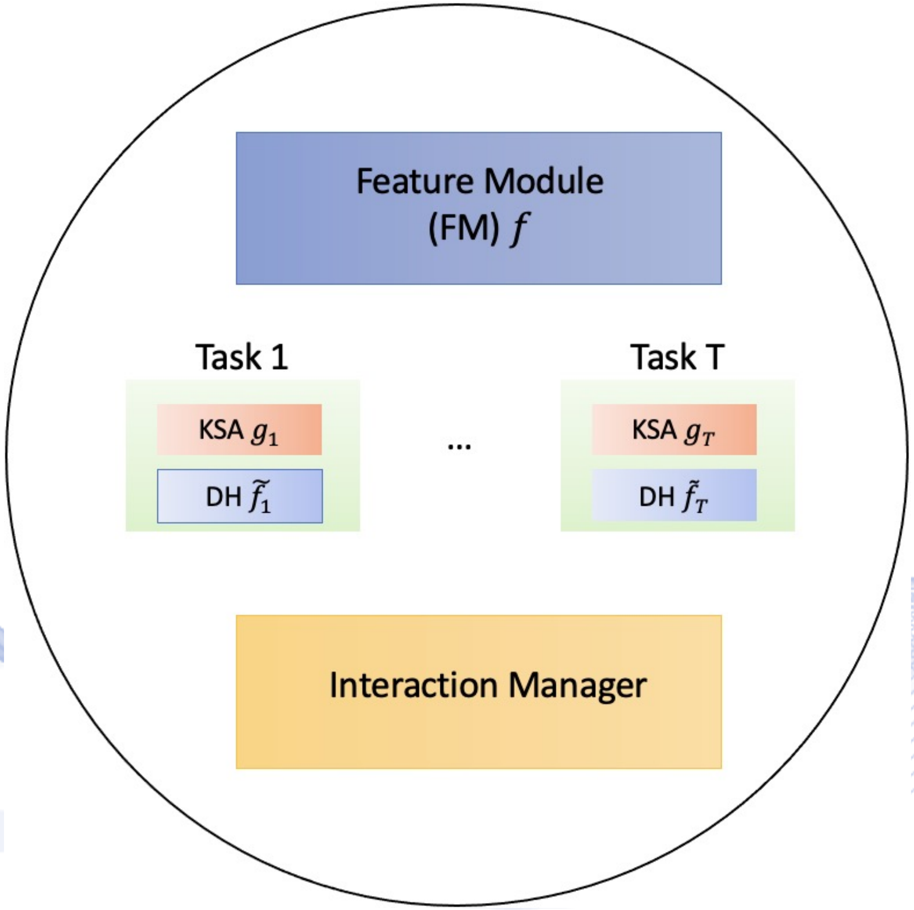
1. Teacher-Student Setup.
2. Training: The student model is trained using a combination of the original training data and the teacher model's predictions or intermediate representations.
3. Objective Function: The objective is to minimize the discrepancy between the student's predictions and the teacher's outputs or representations.



Learning-by-Education Node Community (LENC) Framework

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- **Learning-by-Education Node Community (LENC) Framework**
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

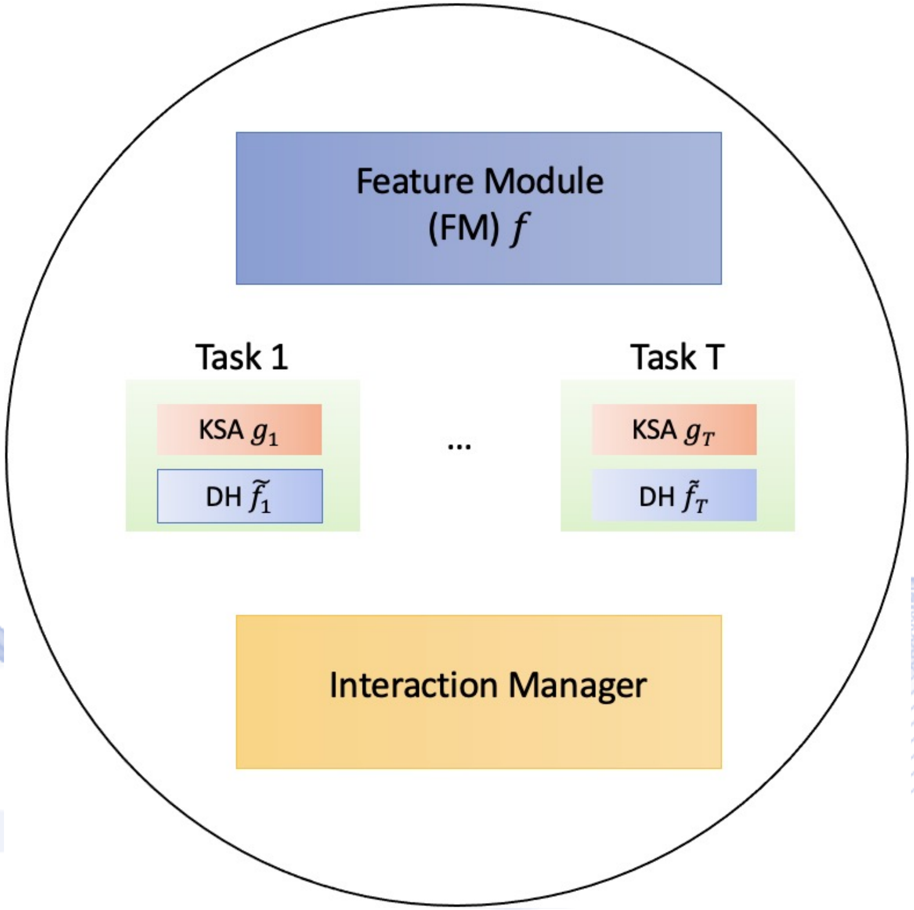
Learning-by-Education Node Community (LENC) Framework



- External environment data streams $\mathcal{D}_s$.
- DNN nodes
  - Feature Module (FM) f.
  - Decision Heads (DH) $\tilde{f}_i$, $i = 1, \cdots, T$.
  - Knowledge Self-Assessment (KSA) Modules g_i , $i = 1, \cdots, T$.
  - Interaction Manager.

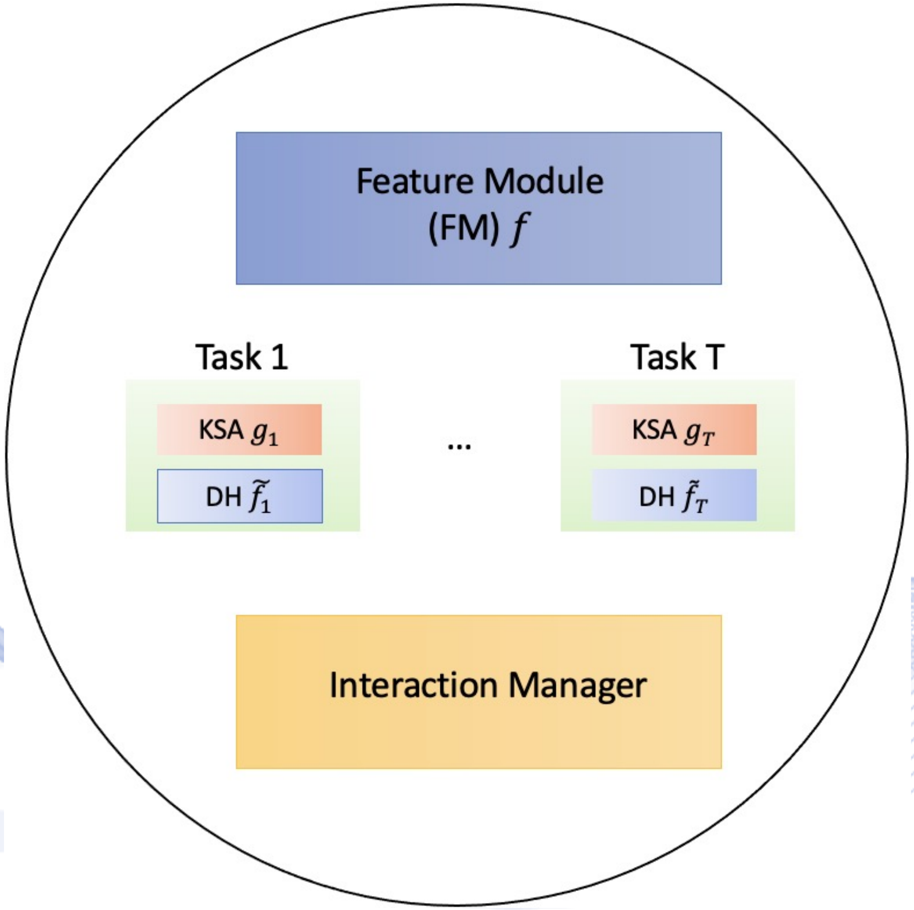
Artificial Intelligence & Information Analysis Lab

Learning-by-Education Node Community (LENC) Framework



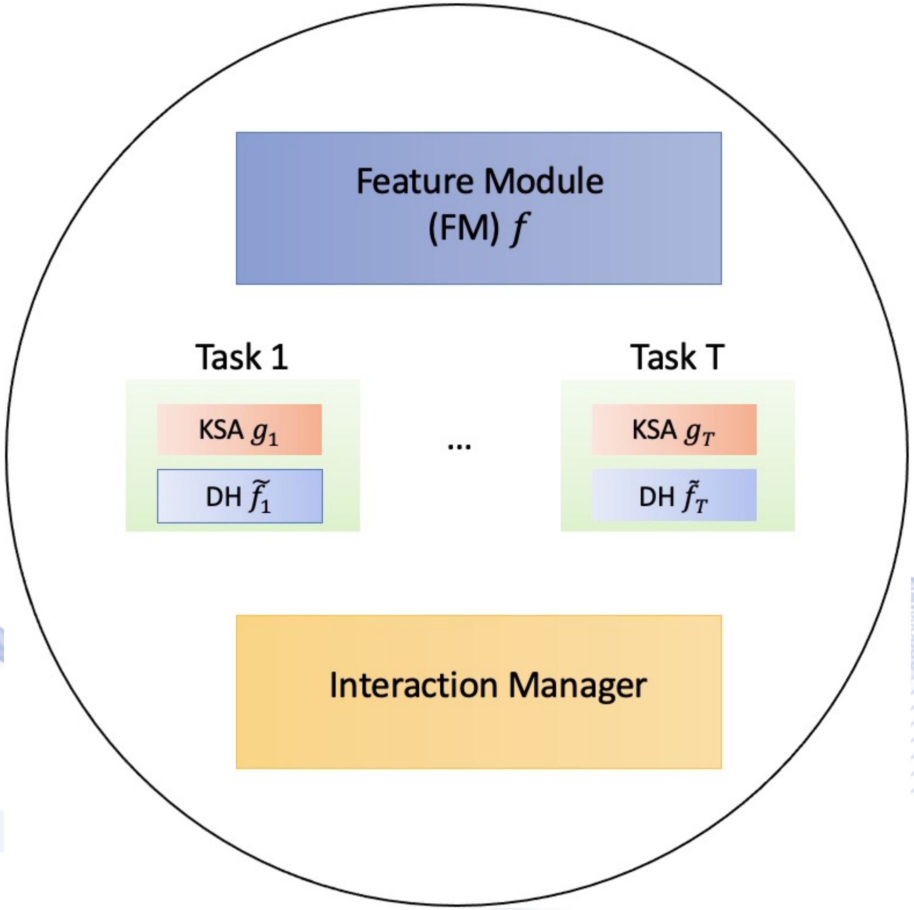
- Knowledge Self-Assessment Modules
  - The KSA Modules consist of an Out-of-Distribution (OOD) detector $g_i(\boldsymbol{x})$: $\mathcal{X}_i \longrightarrow \{0,1\}, i = 1, \cdots, T$.
  - This module classifies new data samples $\boldsymbol{x} \in \mathcal{X}_i,\ i = 1, \cdots, T$ as in or out of distribution.

Learning-by-Education Node Community (LENC) Framework



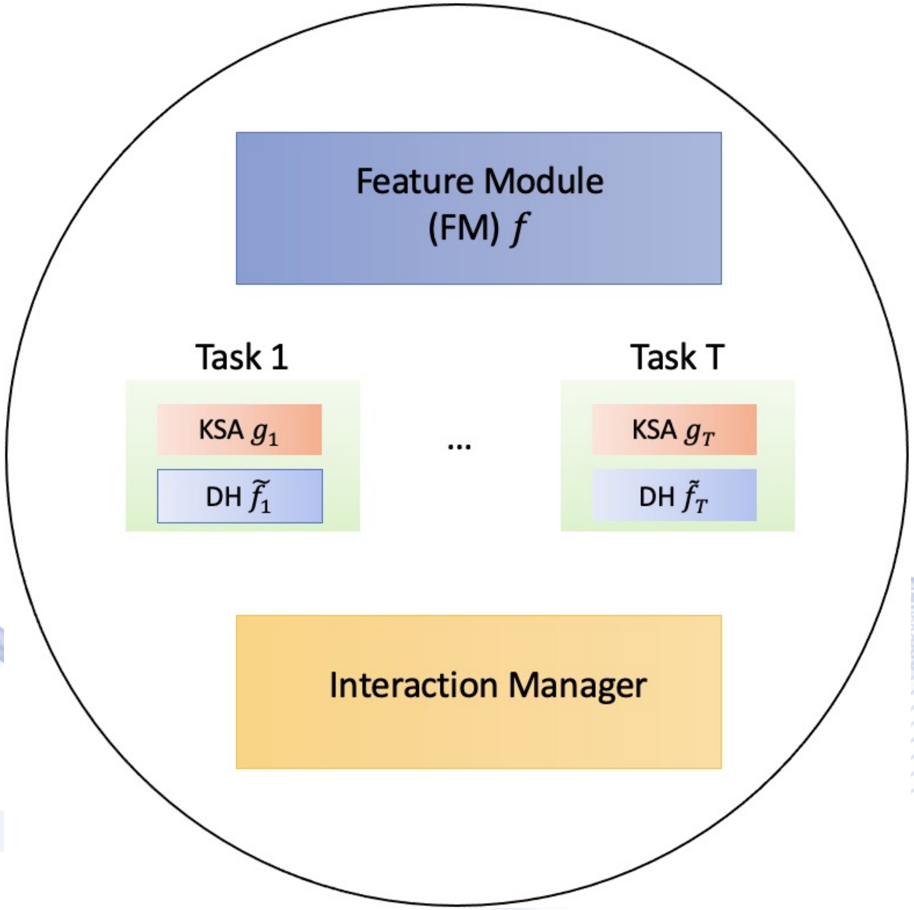
- Knowledge Self-Assessment Modules
  - The KSA module is used to automatically detect which DH $\tilde{f}_i, i = 1, \cdots, T$ will be used for decision making.
  - We define $j = argmax(g_1, \cdots, g_T)$, where j is the index of the task trained on sample data that were like the ones found in $\mathcal{X}_i$.

Learning-by-Education Node Community (LENC) Framework



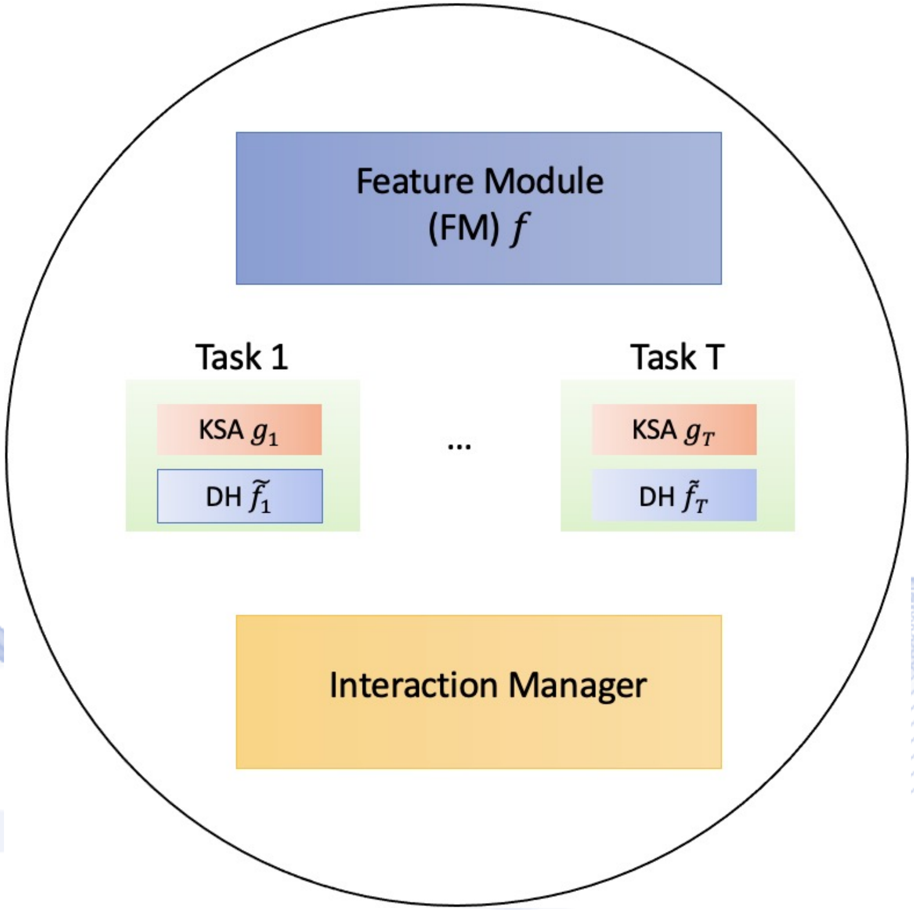
- Feature Module
  - Shared DNN f among tasks, parametrized by w_s.
  - Decision Heads $\tilde{f}_i$, $i = 1, \cdots, T$, parametrized by w_i.
  - Decision (Inference): $\tilde{y}_j = \tilde{f}_j\big(f(\boldsymbol{x}; w_s); w_j\big)$, for an input vector $\boldsymbol{x}$, where $j = argmax(g_1, \cdots, g_T)$.

Learning-by-Education Node Community (LENC) Framework



- Interaction Manager
  - Handles the communications among the nodes.
  - Handles the communications among the nodes and the external environment.

Learning-by-Education Node Community (LENC) Framework



- Interaction Manager
- Three Key Functions:
  - Receives data streams $\mathcal{D}^s$ from the environment.
  - Transmits the data streams $\mathcal{D}^s$ to other nodes and receives their responses$\{q_j, j = 1, \cdots, N, i \neq j\}$, where N is the number of nodes and i is the current node.
  - Sends and receives node components, such as data, activations, weights and structure.

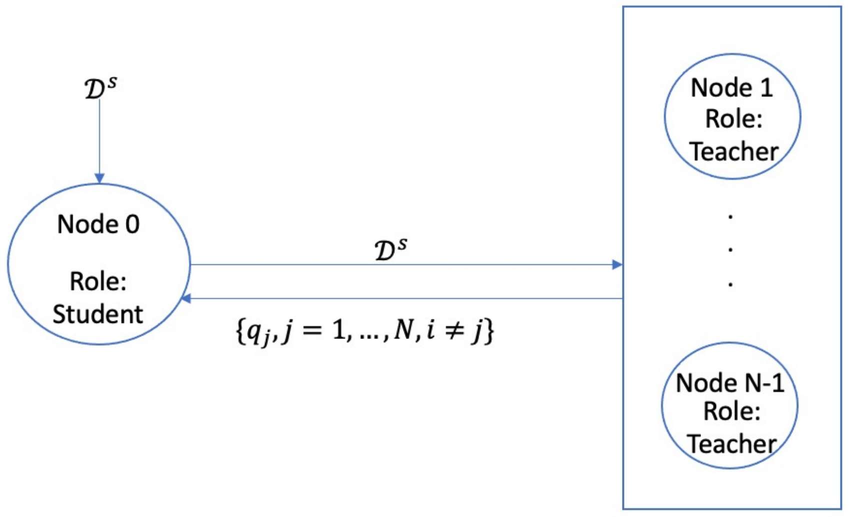
Artificial Intelligence & Information Analysis Lab

Learning-by-Education Node Community (LENC) Framework

Interaction Manager

- Possible ways to compute q_n for each LENC node:
- a) Average Accuracy
  Stored average classification accuracy over past tasks.
- b) OOD Score
  Function of out-of-distribution score from the KSA module, using $\mathcal{D}^s$.
- c) Prediction Disagreement (Churn)
  Accuracy of student predictions on $\mathcal{D}^s$ using the teacher node outputs as pseudo-ground-truth.

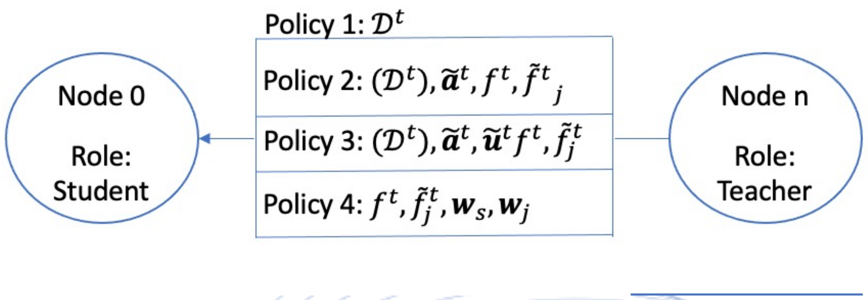
Learning-by-Education Node Community (LENC) Framework



- External Environment sends data stream $\mathcal{D}_s$.
- Node's KSA Module checks if the distribution is known.
- If not the data stream is sent to other nodes.
- The nodes respond with $\{q_j, j = 1, \cdots, N, i \neq j\}$.
- The student node selects a teacher node.

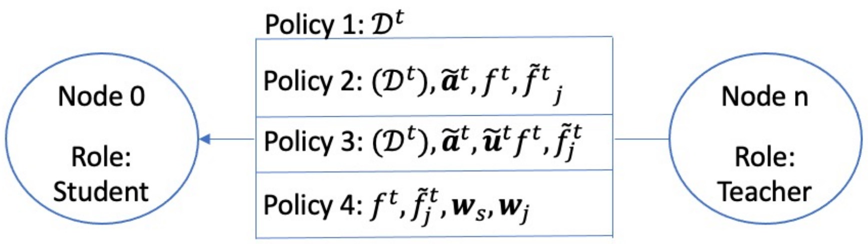
Artificial Intelligence & Information Analysis Lab

Learning-by-Education Node Community (LENC) Framework



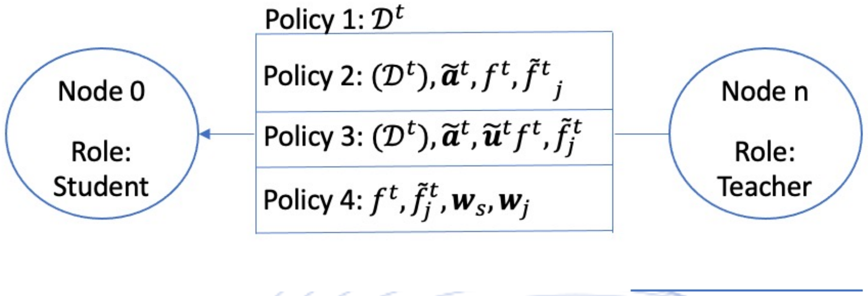
- Option 1: Data Transmission
  - The teacher node sends its training data $\mathcal{D}^t$.
  - The student node uses the training data to learn the task.

Learning-by-Education Node Community (LENC) Framework



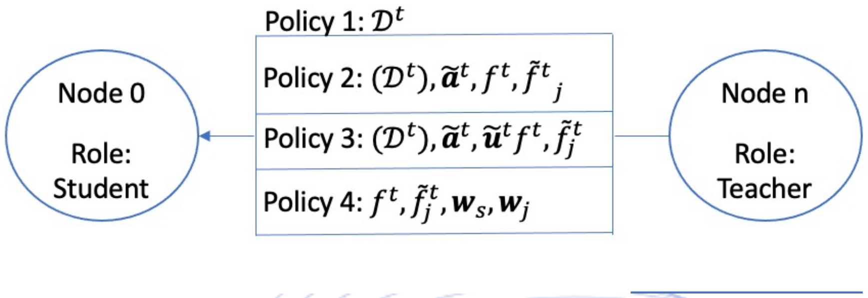
- Option 2: Soft-Output Activation Transmission
  - The teacher node sends its training data $\mathcal{D}^t$, its soft-output activations $\tilde{\boldsymbol{a}}^t$ and its structure f^t and $\tilde{f}_j^t$ for the task j.
  - The student node uses KD to for training using the teacher's guidance.

Learning-by-Education Node Community (LENC) Framework



- Option 3: Feature Activation Transmission
  - The teacher node sends its training data $\mathcal{D}^t$, its soft-output activations $\tilde{\boldsymbol{a}}^t$, its feature activations $\tilde{\boldsymbol{u}}^t$ and its structure f^t and $\tilde{f}_j^t$ for the task j.
  - The student node uses KD to for training using the teacher's guidance.

Learning-by-Education Node Community (LENC) Framework



- Option 4: Weights Transmission
  - The teacher node its structure f^t and $\tilde{f}_j^t$ and its weights ξ_s and ξ_j for the task j.
  - The student node is now a copy of the teacher node's model.

Learning-by-Education Node Community (LENC) Framework

LENC selects the appropriate knowledge transfer policy based on user-defined environmental conditions.

Key Questions:

1. Are there privacy limitations on the model, dataset, or parameters?
2. Are there network traffic limitations?
3. Is there a latency requirement for instant transfer?

Learning-by-Education Node Community (LENC) Framework

Policy Selection Logic:

- Policy 2 (Default)
  - Use when strong privacy restrictions apply
  - Only the first input option ($D^s \rightarrow$ soft activations)
  - Works with any architecture or dataset
- Policy 3
  - Use if the teacher and student share architecture
  - More effective guidance
  - Second input (D_j^t) allowed only if no privacy or traffic limitations

Learning-by-Education Node Community (LENC) Framework

Policy Selection Logic:

- Policy 4
  Use if all apply:
  ▪ Latency-sensitive
  ▪ No privacy limits
  ▪ Student is untrained

- Training-free option

- Policy 1
  Use if all apply:
  ▪ No privacy or traffic limits
  ▪ Teacher’s architecture can be shared
  ▪ Student > Teacher in model complexity

Federated Learning

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - **Federated Learning**
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Federated Learning

- Define a node as *a master node.*
- All nodes with the same structure within the community train themselves using their local data.
- The master node uses Option 4 to receive the weights of all nodes with the same structure within the community.
- The master node aggregates the weights of all participating nodes.
- The process is repeated until convergence.

Peer-to-Peer Networks

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - **Peer-to-Peer Networks**
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

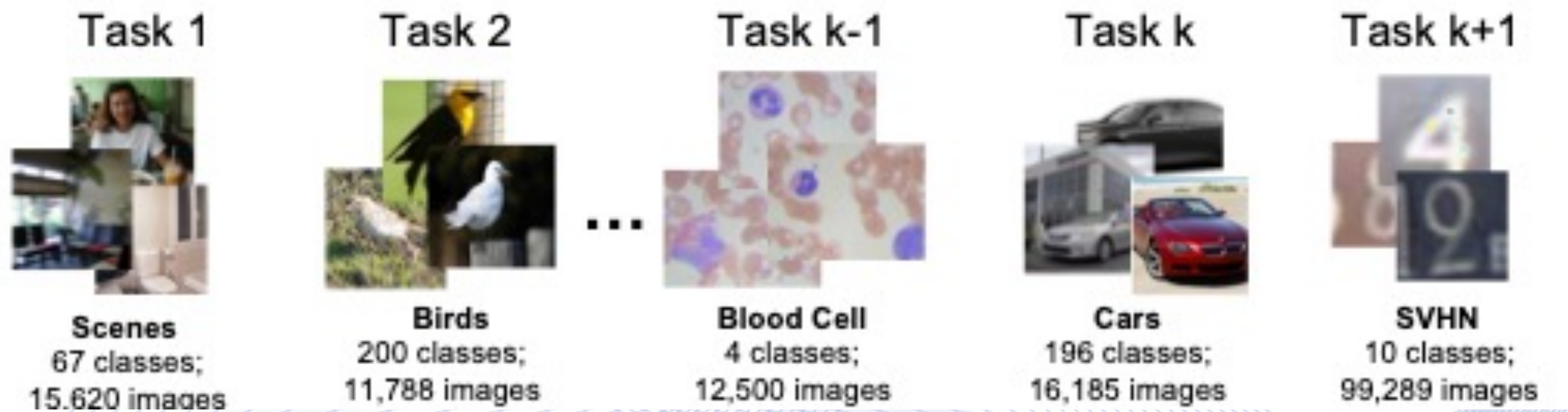
Peer-to-Peer Networks

- Options 1-4 constitute forms of Peer-to-Peer Network interactions.
- Nodes act exclusively to enhance their knowledge.
- No need for a central server.
- Retaining knowledge within the node community.

Continual Learning

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - **Continual Learning**
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Continual Learning



Edge Computing – Decentralized Inference

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - **Edge Computing – Decentralized Inference**
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

Edge Computing – Decentralized Inference

- Raw data is processed locally on nodes.
- Nodes use real-time inference on their data.
- Lightweight training of Feature Modules directly on nodes.
- A master node (server) can be defined to aggregate inference results.
- Generating responses or actions locally without centralized decision-making.

Reproducibility - Privacy

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - **Reproducibility – Privacy**
  - Deep Learning Tasks Supported by LENC Framework
- Experimental Evaluation

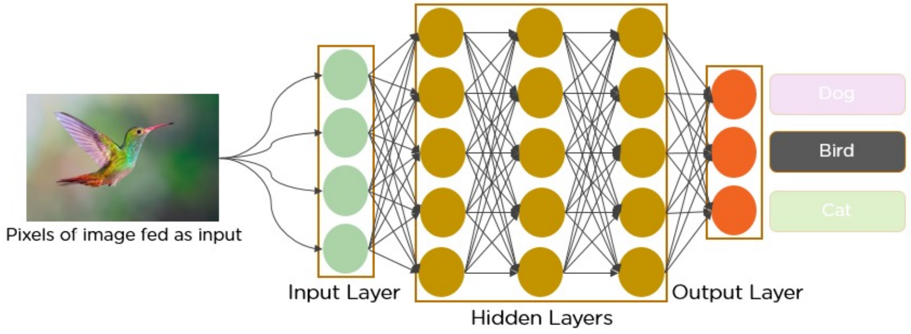
Reproducibility - Privacy

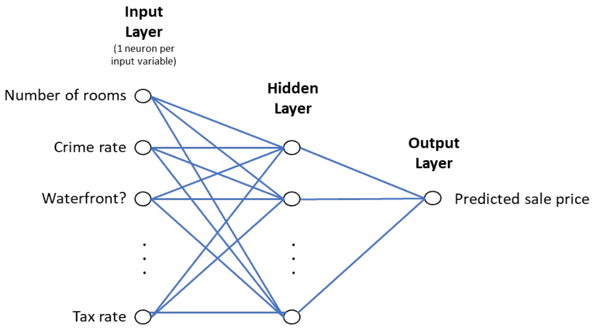
- DNN node 1 is the model of a published paper.
- DNN node 2 wants to replicate the model and the experiments.
- Using variations of Options 1-4 DNN node 2 can replicate the initial model and also consider possible privacy constraints.
- Private weights, architecture, training dataset, etc.

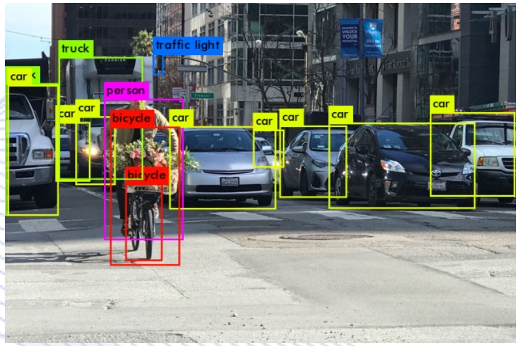
Deep Learning Tasks using the LENC Framework

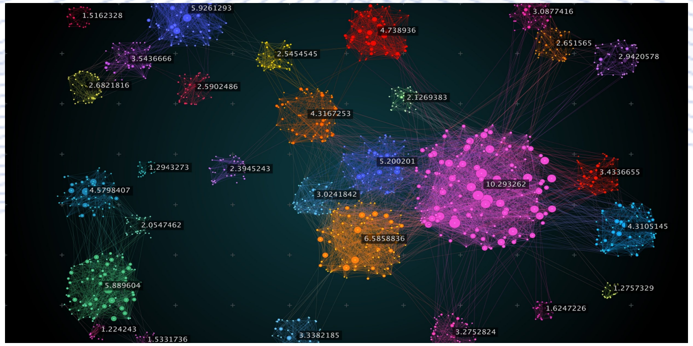
- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - **Deep Learning Tasks Supported by LENC Framework**
- Experimental Evaluation

Deep Learning Tasks using the LENC Framework









Experimental Evaluation

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Networks
- Knowledge Distillation
- Learning-by-Education Node Community (LENC) Framework
  - Federated Learning
  - Peer-to-Peer Networks
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
  - Deep Learning Tasks Supported by LENC Framework
- **Experimental Evaluation**

Experimental Evaluation

Datasets: CIFAR-10 & CIFAR-100.

Architectures: ResNet-18 (teacher), WRN-16-4, VGG11, and additional ResNet-18s (students).

Key Details:

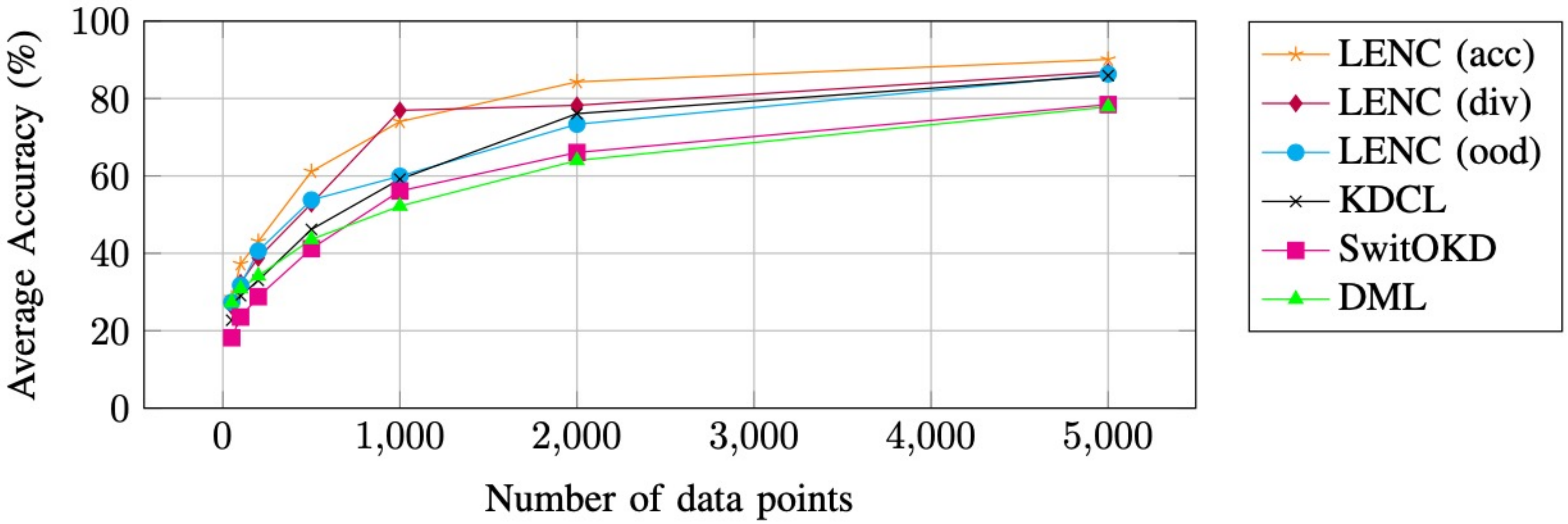
- Pretrained ResNet-18 used as the sole teacher
- Competing CKD methods adapted to use teacher responses (not ground-truth).
- Two stream sizes: 1,000 & 5,000 data points from the teacher's training set.
- 10 sequential data streams $\mathcal{D}_s$, each triggering a knowledge cycle

Experimental Evaluation

Dataset	Students	Stream Size	DML	KDCL	SwitOKD	LENC (proposed)
CIFAR-10	ResNet-18 & ResNet-18	1000	52.20±0.52	62.23±0.15	56.15±0.73	**76.93±0.71**
	WRN-16-4 & VGG11		51.17±0.71	62.09 ± 0.21	57.85±0.80	**70.16±0.82**
	ResNet-18 & ResNet-18	5000	77.85±0.31	85.76±0.07	79.08±0.70	**86.31± 0.32**
	WRN-16-4 & VGG11		75.56±0.82	84.47 ± 0.08	78.79±0.68	**87.12±0.24**
CIFAR-100	ResNet-18 & ResNet-18	1000	9.77±0.25	25.16±0.12	13.71±0.57	**34.96±0.47**
	WRN-16-4 & VGG11		6.12±0.38	27.59±0.19	14.72±0.61	**29.75±0.49**
	ResNet-18 & ResNet-18	5000	31.53±0.31	58.70±0.09	35.31±0.29	**65.02±0.13**
	WRN-16-4 & VGG11		8.30±0.16	56.94±0.12	37.27±0.45	**58.18±0.17**

Comparisons of LENC with competing CKD methods, for incoming data streams Ds of sizes 1000 and 5000. The average test accuracy (%) of the student nodes is reported.

Experimental Evaluation



Average student LENC node classification accuracy (%) for varying Ds sizes in the CIFAR-10 dataset. The 3 alternative LENC teacher selection policies are compared against competing methods.

Experimental Evaluation

- Comparisons of the LENC knowledge transfer policies, for incoming data streams $\mathcal{D}_s$ of sizes 100, 500, 1000, 5000, and 60000 (full dataset).
- Policies 2-3 are independently evaluated with both unlabeled (using $\mathcal{D}_s$) and labeled (using $\mathcal{D}_j^t$) input options.
- The average test classification accuracy (%) of the student LENC nodes is reported.

Dataset	Stream Size	Policy 1	Policy 2	Policy 3
$\mathcal{D}_j^t$	60000	91.97	**93.72**	93.59
$\mathcal{D}^s$	60000	-	91.86	**92.07**
	100	-	**37.75**	37.11
	500	-	61.13	**62.48**
	1000	-	74.04	**74.29**
	5000	-	**90.15**	90.05

Student LENC node classification accuracy (%) for varying Ds sizes in the CIFAR-10 dataset. The 3 alternative knowledge transfer policies are examined.

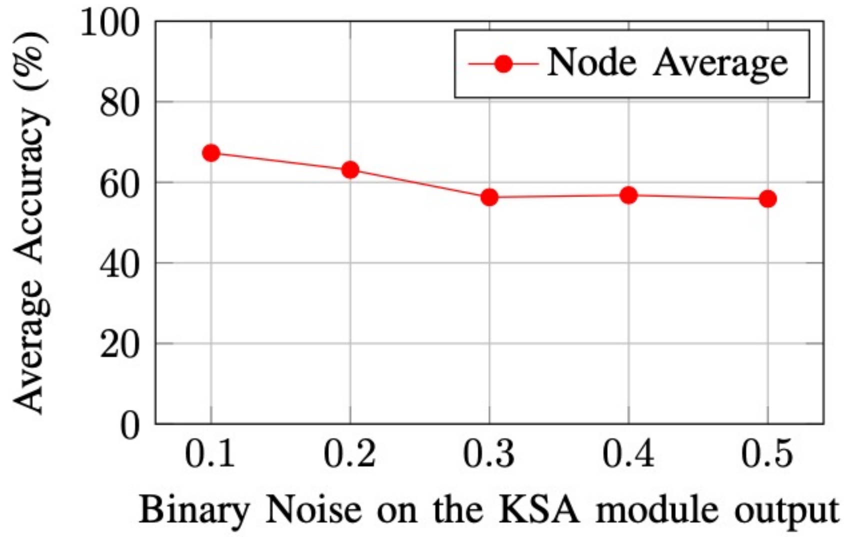
Experimental Evaluation

Experiment Setup:

- Repeated the CKD experiment. Used two untrained ResNet-18 students.
- Data stream: 1,000 CIFAR-10 samples.
- Simulated KSA failure by injecting binary noise into KSA outputs.

Key Observation:

- LENC remained robust despite KSA corruption.
- Only a slight drop in average accuracy was observed.



KSA module robustness analysis by adding binary noise to the KSA modules' output.

Artificial Intelligence & Information Analysis Lab

Bibliography

[1] I. Pitas, “Artificial Intelligence Science and Society Part A: Introduction to AI Science and Information Technology“, Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156460?ref_=pe_3052080_397514860

[2] I. Pitas, “Artificial Intelligence Science and Society Part B: AI Science, Mind and Humans“, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156479?ref_=pe_3052080_397514860

[3] I. Pitas, “Artificial Intelligence Science and Society Part C: AI Science and Society”, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156487?ref_=pe_3052080_397514860

[4] I. Pitas, “Artificial Intelligence Science and Society Part D: AI Science and the Environment“, Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156495?ref_=pe_3052080_397514860

Bibliography

[KAI2024] Kaimakamidis, A., Mademlis, I., & Pitas, I. (2024). Collaborative Knowledge Distillation via a Learning-by-Education Node Community. arXiv preprint arXiv:2410.00074.

[ZHA2021] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216, 106775.

[MAS2020] Masinde, N., & Graffi, K. (2020). Peer-to-peer-based social networks: A comprehensive survey. SN Computer Science, 1(5), 299.

[BEL2021] Bellavista, P., Foschini, L., & Mora, A. (2021). Decentralised learning in federated deployment environments: A system-level survey. ACM Computing Surveys (CSUR), 54(1), 1-38.

Bibliography

[OUY2021] Ouyang, S., Dong, D., Xu, Y., & Xiao, L. (2021). Communication optimization strategies for distributed deep neural network training: A survey. Journal of Parallel and Distributed Computing, 149, 52-65.

[REN2023] Ren, W. Q., Qu, Y. B., Dong, C., Jing, Y. Q., Sun, H., Wu, Q. H., & Guo, S. (2023). A survey on collaborative DNN inference for edge intelligence. Machine Intelligence Research, 20(3), 370-395.

[HIN2015] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr