Decentralized DNN Architectures

A. Kaimakamidis, N. Tzavidas, D. Papaioannou, Prof. I. Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 7.6

Artificial Intelligence & Information Analysis Lab

Decentralized DNN Architectures

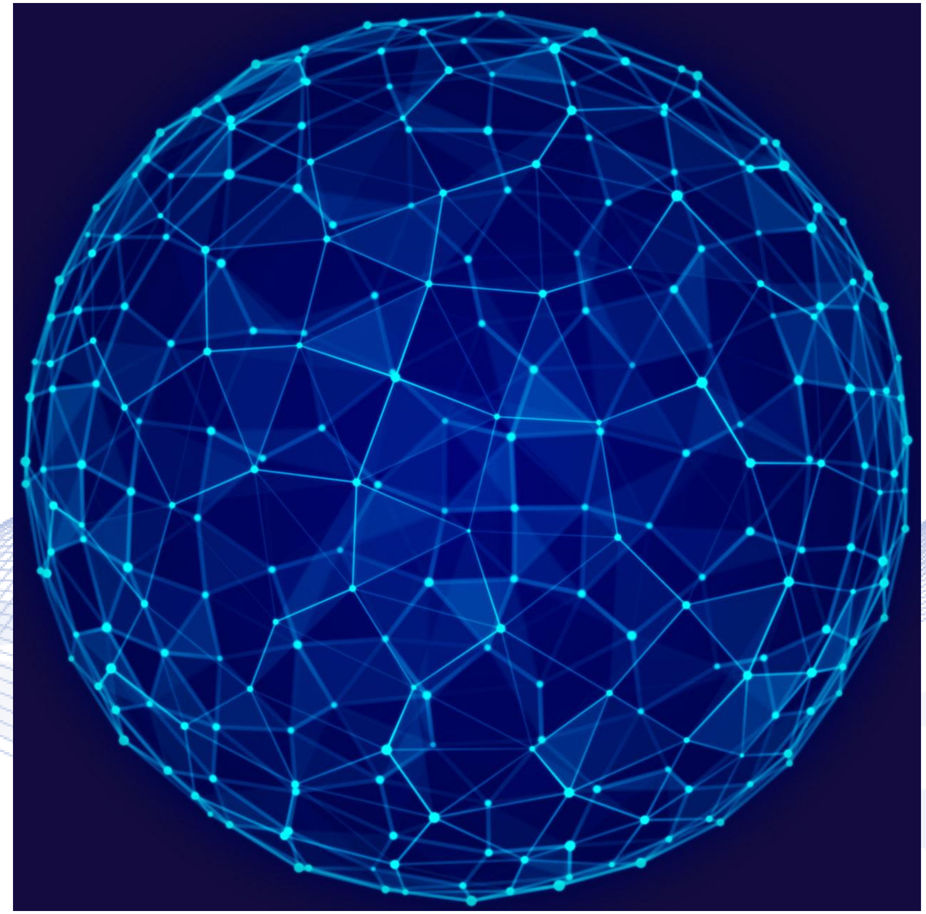
- **Decentralized DNN Architectures**
- Learning-by-Education Node Community (LENC) Framework
- LENC Framework Applications
- LENC Framework Experiments
- LENC Architecture Implementation

Decentralized DNN Architectures

Definition

Decentralized Deep Neural Network architectures distribute computation and decision-making across multiple nodes or devices, offering advantages in:

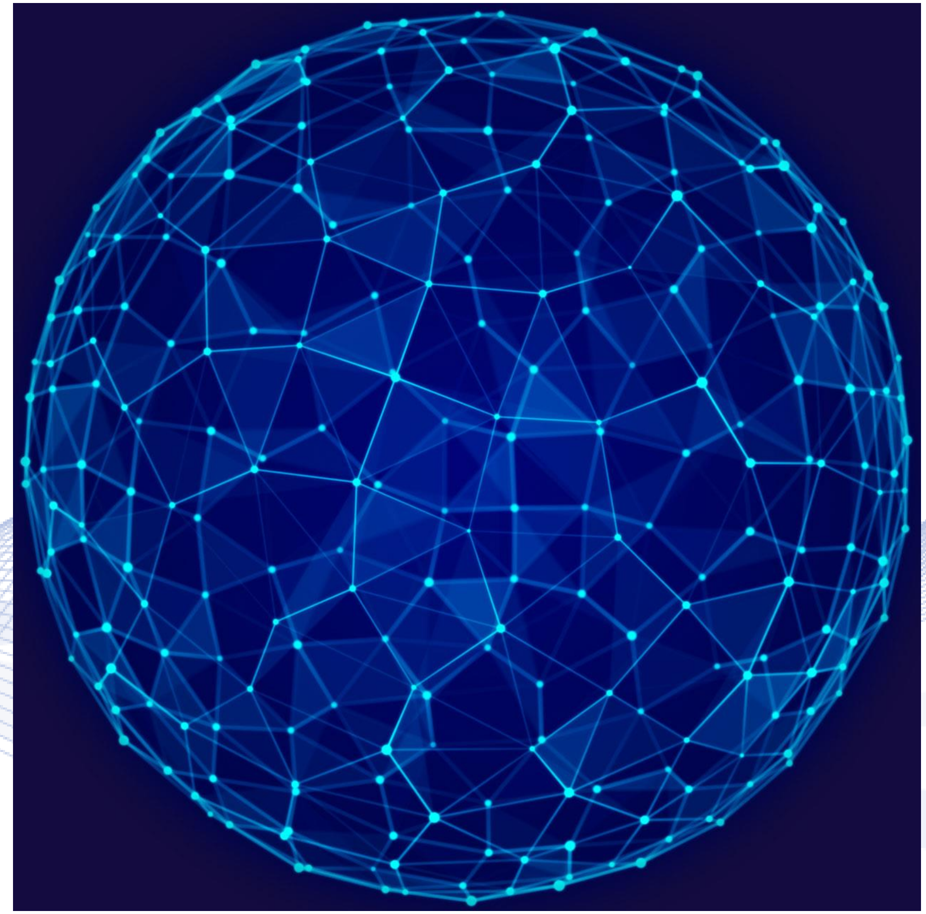
- ***scalability,***
- ***privacy, and***
- ***robustness***.



Decentralized DNN Architectures

Decentralized DNN advantages

- ***Distribution***: Data and computations are spread across multiple nodes or devices.
- ***Collaboration***: Nodes can cooperate for DNN model training or inference.
- ***Privacy Preservation***: Data remain local, thus enhancing privacy and security.
- ***Fault Tolerance***: Resilience to individual node failures or attacks.

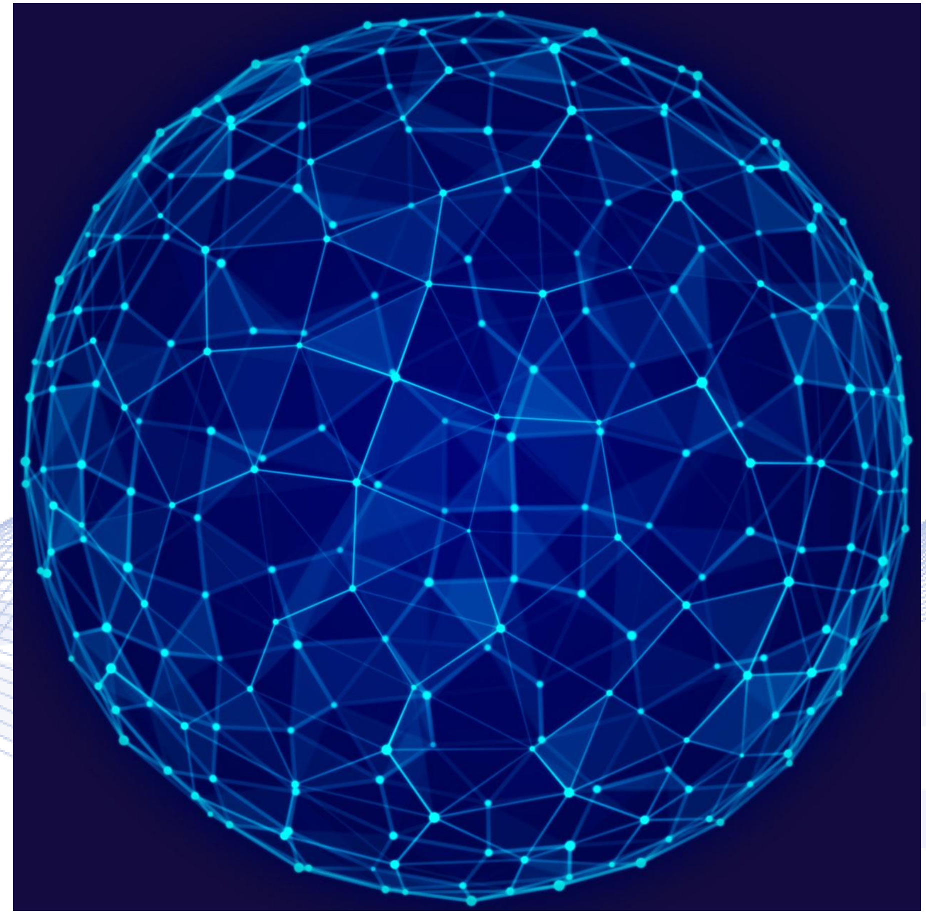


1. Types:
   1. Federated Learning: Training a global

Decentralized DNN Architectures

Decentralized DNN computation

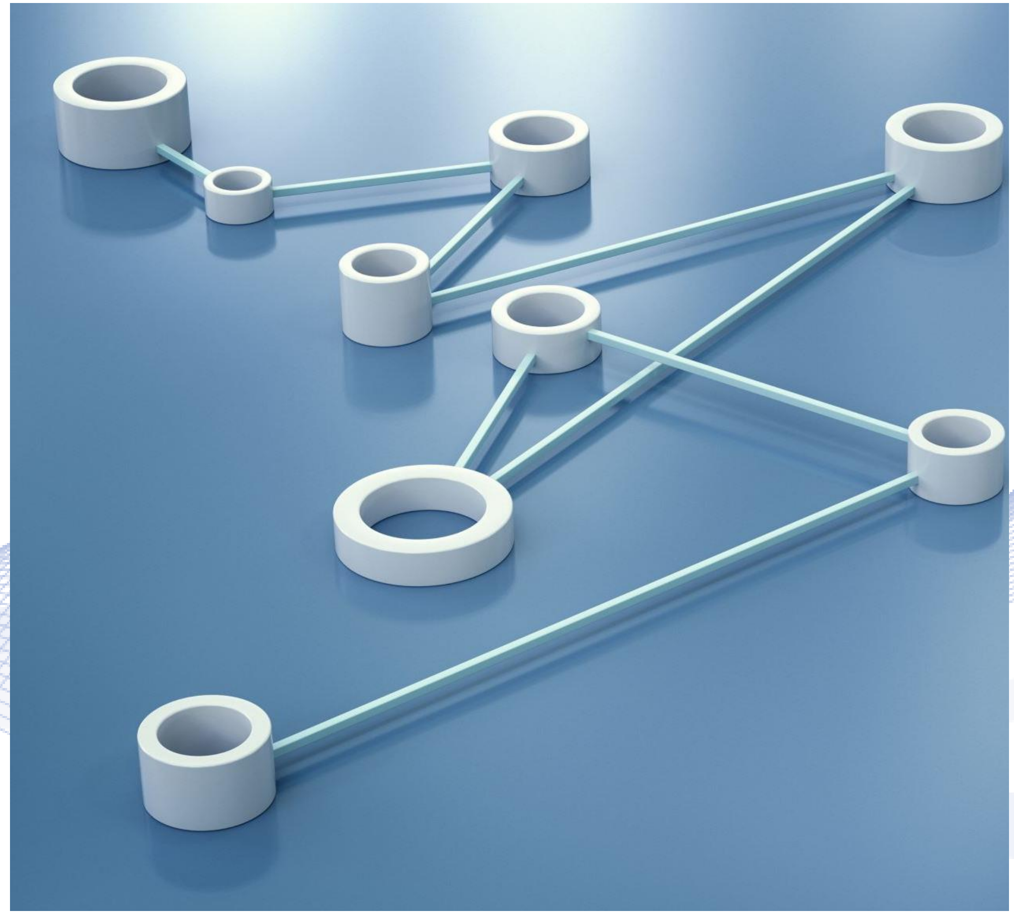
- ***Peer-to-Peer Networks***: Collaborative learning (training) among peers, without a central server.
- ***Cloud DNN Computing***: Running DNN training and/or inference on cloud nodes.
- ***Edge Computing***: Running DNN inference or lightweight training directly on edge devices.



Decentralized DNN Architectures

Peer-to-peer DNN computing.

- ***Decentralization***: Reduced dependency on central servers, enhancing scalability and robustness.
- ***Resource Efficiency***: Idle computational resource utilization across peers.
- ***Resilience*** to node failures or attacks.
- ***Community-driven Innovation*** through collaborative research and knowledge exchange.

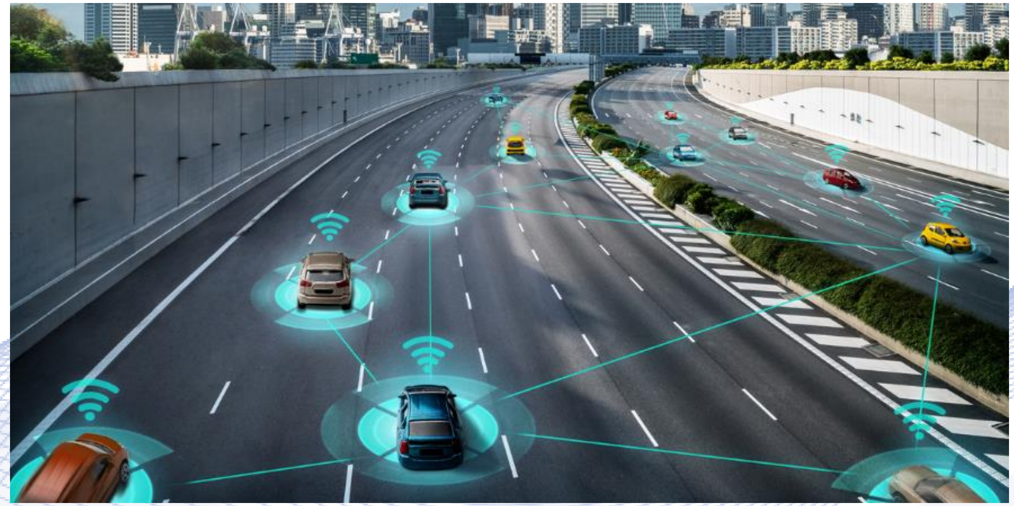


Artificial Intelligence & Information Analysis Lab

Decentralized DNN Architectures

Edge DNN Computing

- ***Low Latency***: Decision-making without reliance on distant servers.
- ***Bandwidth Efficiency***: No transfer of large data volumes to central servers.
- ***Privacy Preservation***: Sensitive data can be processed locally, enhancing privacy.
- ***Offline Capability***: DNN operation in disconnected or low-connectivity environments.



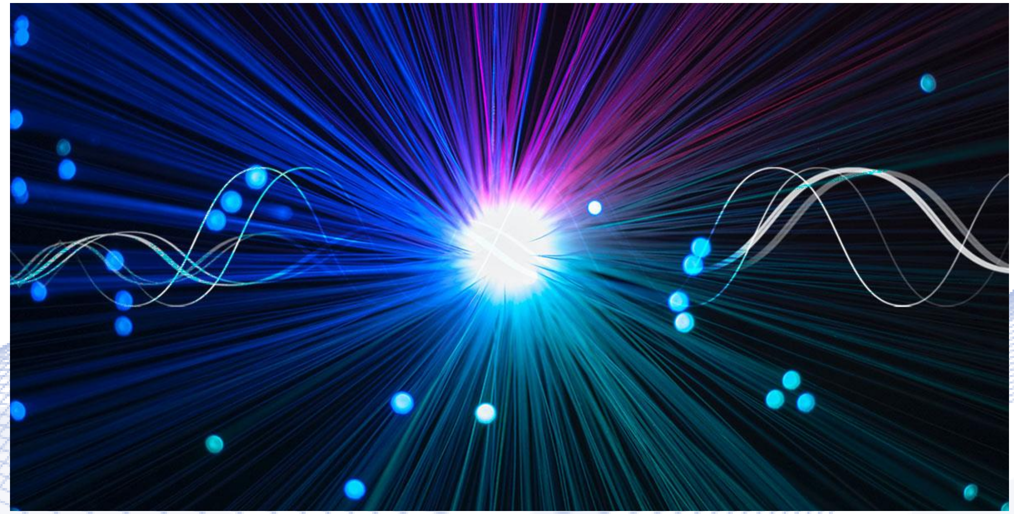
Decentralized DNN Architectures

- Decentralized DNN Architectures
- **Learning-by-Education Node Community (LENC) Framework**
- LENC Framework Applications
- LENC Framework Experiments
- LENC Architecture Implementation

LENC Framework

In ***Knowledge Distillation***, a compact DNN model (***student model***), learns from a larger, more complex DNN model (***teacher model***), by mimicking its outputs or internal representations.

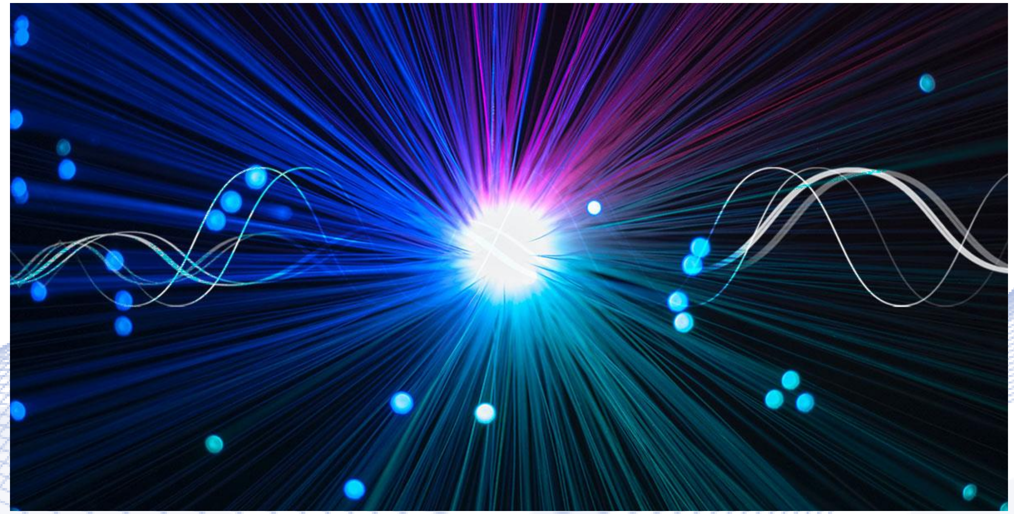
- ***Teacher-Student DNN architectures***.



LENC Framework

Knowledge Distillation process.

- ***Training***: The Student DNN model is trained using a combination of the original training data and the Teacher DNN model predictions or intermediate representations.
- ***Objective Function***: The KD objective is to minimize the discrepancy between the student DNN predictions/ representations from the teacher DNN ones.

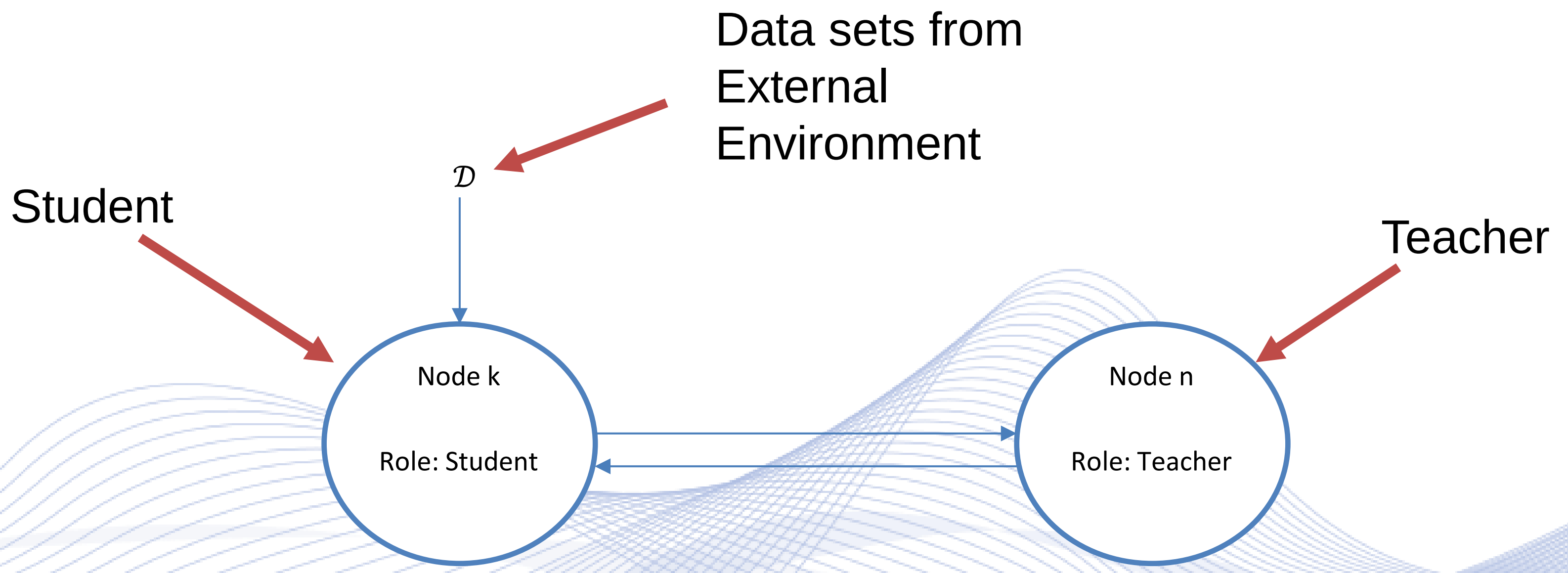


LENC Framework

Teacher-Student Learning for Humans: The student asks for tutoring on unknown data coming from her/his external environment.

LENC Framework

The ***LENC framework*** is a network of N interacting LENC nodes.



Teacher-Student Learning for the LENC framework nodes: The Student LENC nodes asks for tutoring from a Teacher LENC node on unknown data.

LENC Framework

- A ***LENC class*** can have one Teacher and multiple Student nodes.
- LENC can support multiple Teachers and Students.
- Students can choose their Teacher that knows best their task.
- Teachers may learn as well.
- Teacher/student roles may reverse for certain tasks.
- A ***classification task*** is defined on a group of semantic classes.
- Regression or clustering taks can be defined as well.

LENC Framework

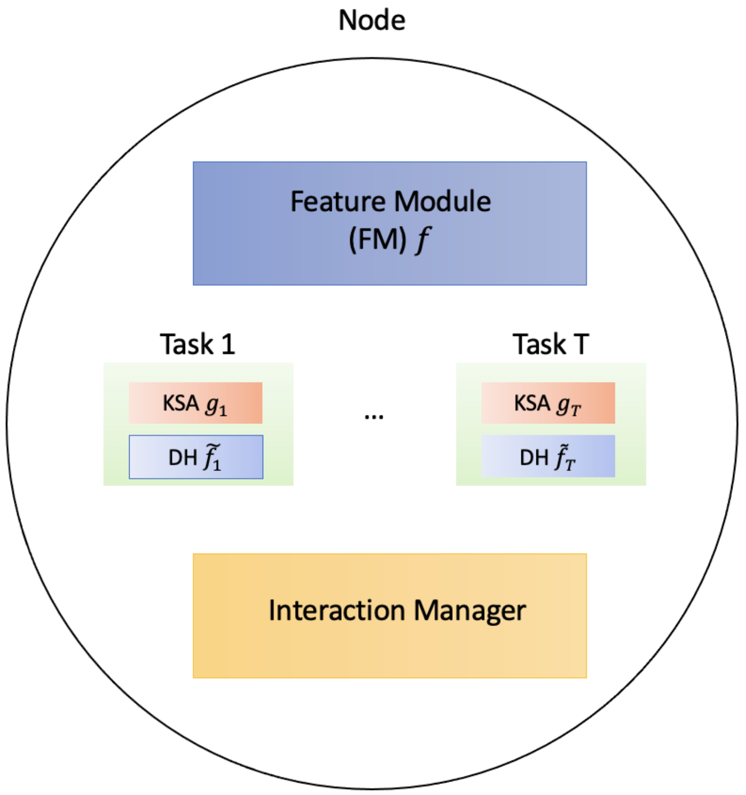
- Students can cooperate with each other during learning.
- Teachers can pull together their knowledge.
- LENC nodes can have a ***cooperating*** or ***competing*** behavior.
- Some LENC nodes may be ***malicious***.

LENC Framework

LENC node structure.

Each LENC node can be trained on various DNN tasks (data classes $1, \cdots, T$).

- ***Feature Module*** (FM) f.
- ***Decision Heads*** (DH) $\tilde{f}_{\mathrm{i}}$, $i = 1, \cdots, T$ (one per task).
- ***Knowledge Self-Assessment*** (KSA) ***Modules*** g_{i}, $i = 1, \cdots, T$.
- ***Interaction Manager*** (IM) interacts with other LENC Ims and receives external environment data sets $\mathcal{D}$.



LENC Node architecture.

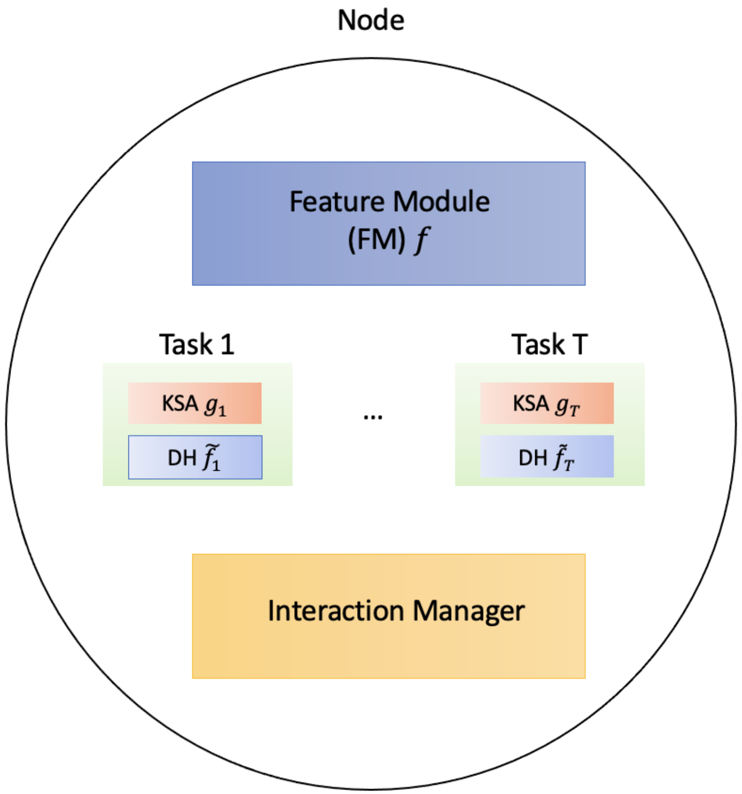
LENC Framework

Feature Module

- ***Feature Module*** (FM) DNN is shared among tasks:

$$\mathbf{f} = f(\mathbf{x}; \mathbf{w}_S).$$

- Its structure is described by $\mathcal{S}_S$.
- It is parametrized by $\mathbf{w}_S$.



LENC Node architecture.

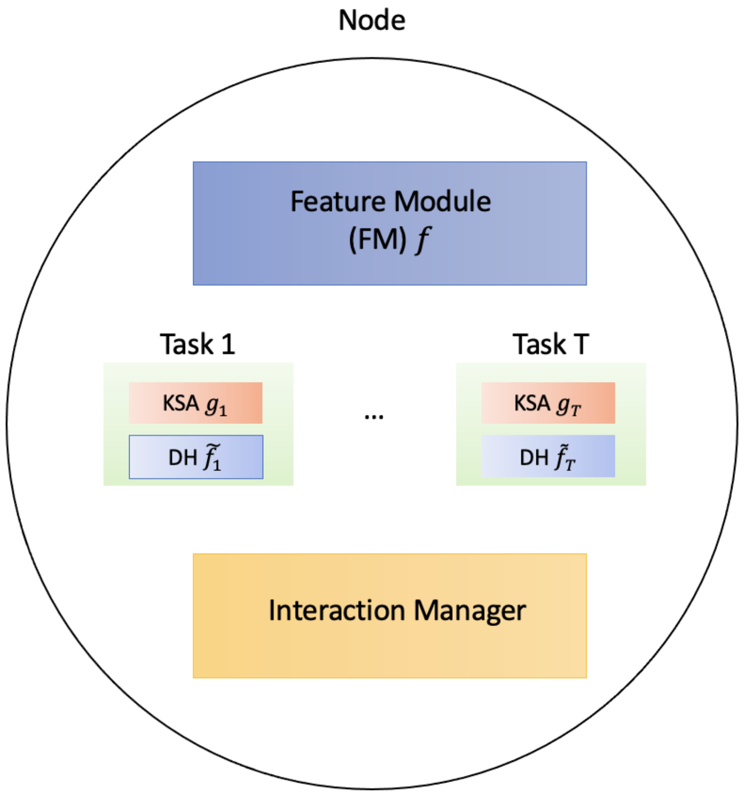
LENC Framework

Knowledge Self-Assessment Module

- It decides whether input data $\mathbf{x}$ of an input dataset $\mathcal{D}$ belongs to the same probability distribution of the data used for LENC node training for each task.
- It comprises an ***Out-of-Distribution*** (OOD) detector:

$$g_i(\mathbf{x}): \mathcal{D} \longrightarrow \{0,1\}, \quad i = 1, \cdots, T.$$

- It classifies new data samples $\mathbf{x} \in \mathcal{D}$ as in- or out-of-distribution for each task.



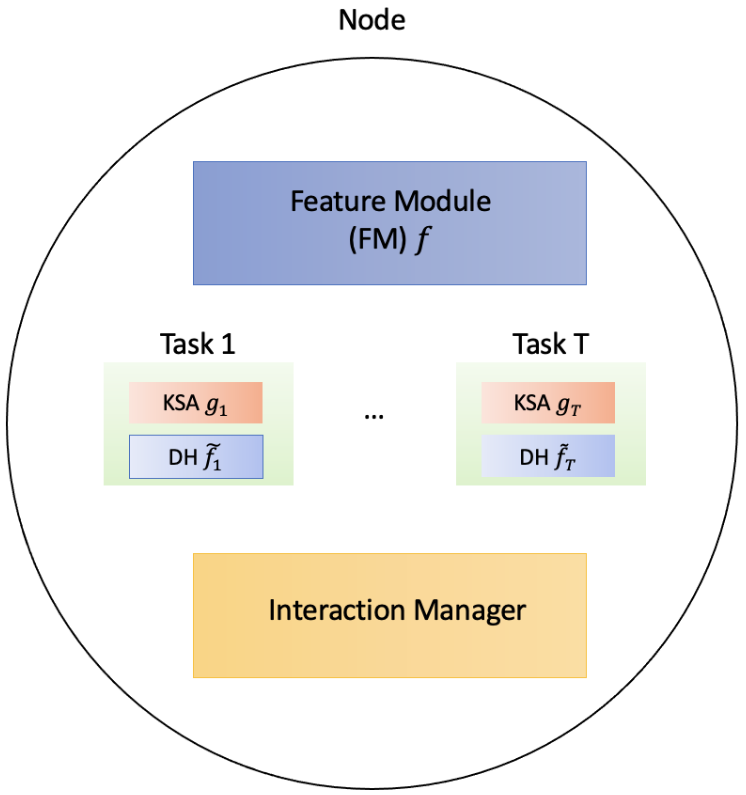
LENC Node architecture.

LENC Framework

- The KSA module is used to automatically detect the Decision Head j out of $\tilde{f}_i, i = 1, \cdots, T$ that will be used for LENC node decision making.
- The decision minimizes:

$$argmin_j(g_1, \cdots, g_T).$$

- Decision Head $\tilde{f}_j$ has been trained on sample data that are similar to current input $\mathbf{x}$.



LENC Node architecture.

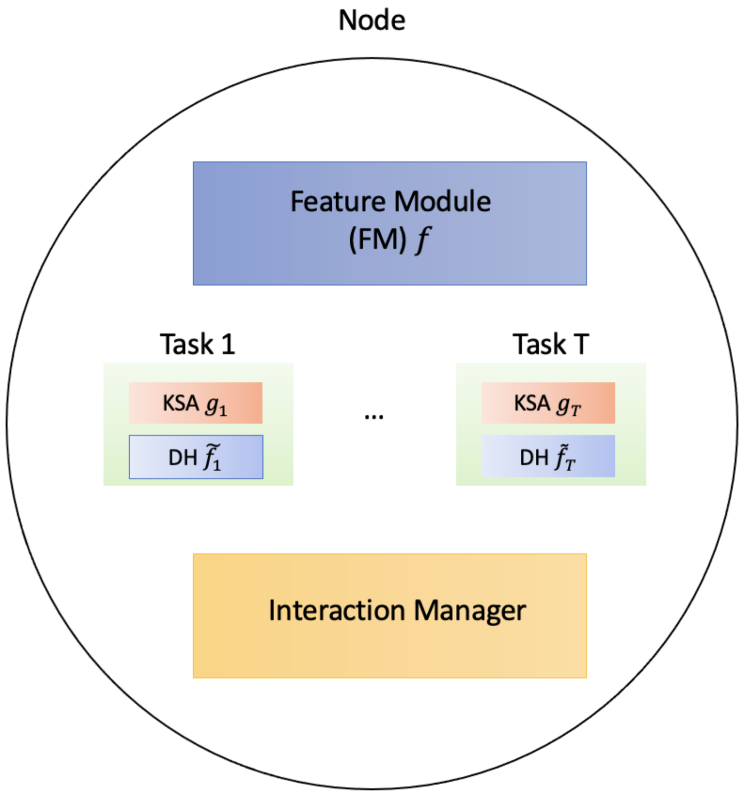
LENC Framework

Decision Heads

- There are T Decision Heads $\tilde{f}_i, i = 1, \cdots, T$ (one per task).
- $\mathcal{S}_i, \mathbf{w}_i, i = 1, \cdots, T$: DH structure description and parameter vector.
- LENC Node Decision is made by concatenating FM and DH inference:

$$\mathbf{f} = f(\mathbf{x}; \mathbf{w}_s), \tilde{y}_j = \tilde{f}_j(\mathbf{f}; \mathbf{w}_j),$$
$$j = argmin(g_1, \cdots, g_T).$$

$\mathbf{x}$: input vector.

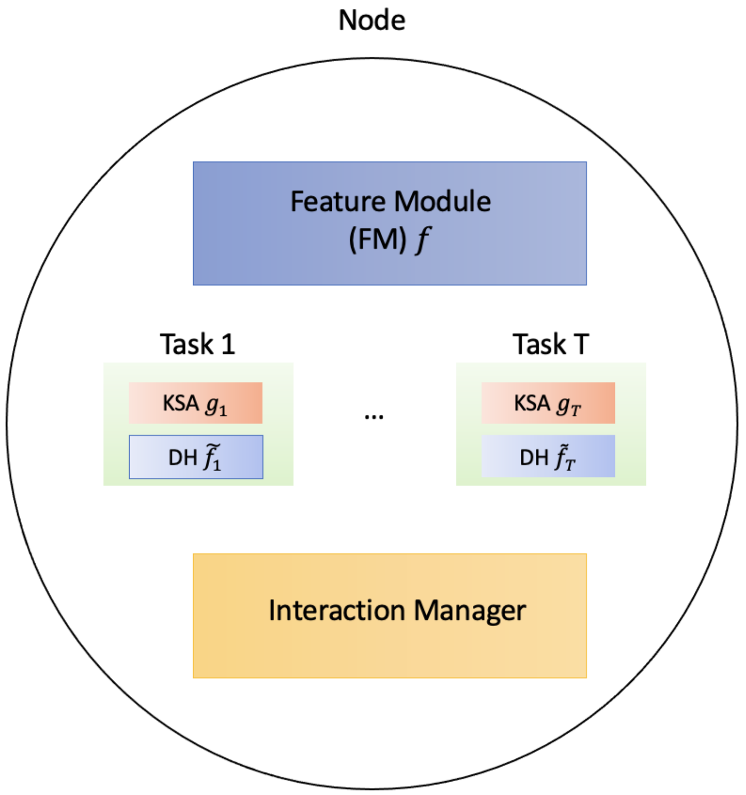


LENC Node architecture.

LENC Framework

Interaction Manager handles:

- Inter-node communications.
- Communications between the nodes and the external environment.
- Communication of LENC nodes components, such as data, activations, weights and structure.



LENC Node architecture.

Artificial Intelligence & Information Analysis Lab

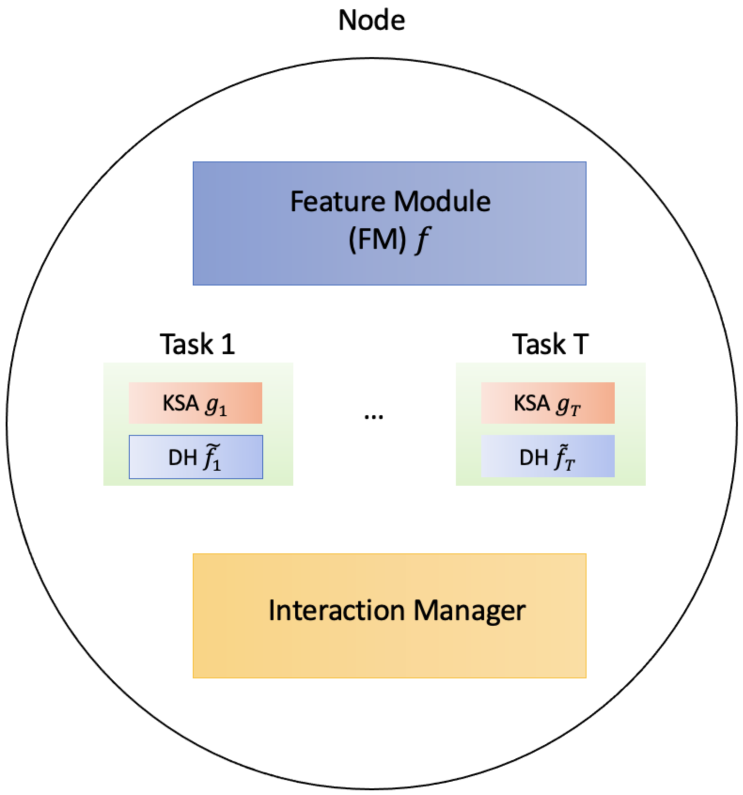
LENC Framework

Key Interaction Manager Functions for LENC node k:

- It receives data sets $\mathcal{D}$ from the environment.
- It transmits data sets $\mathcal{D}'$ to other nodes and receives their responses:

$$\mathcal{D}' = \{q_i, \ \ i = 1, \cdots, N, i \neq k\}.$$

- $q_k = 0$, if the node is not aware of the task or,
- q_k is a scalar number measuring its knowledge on the task.

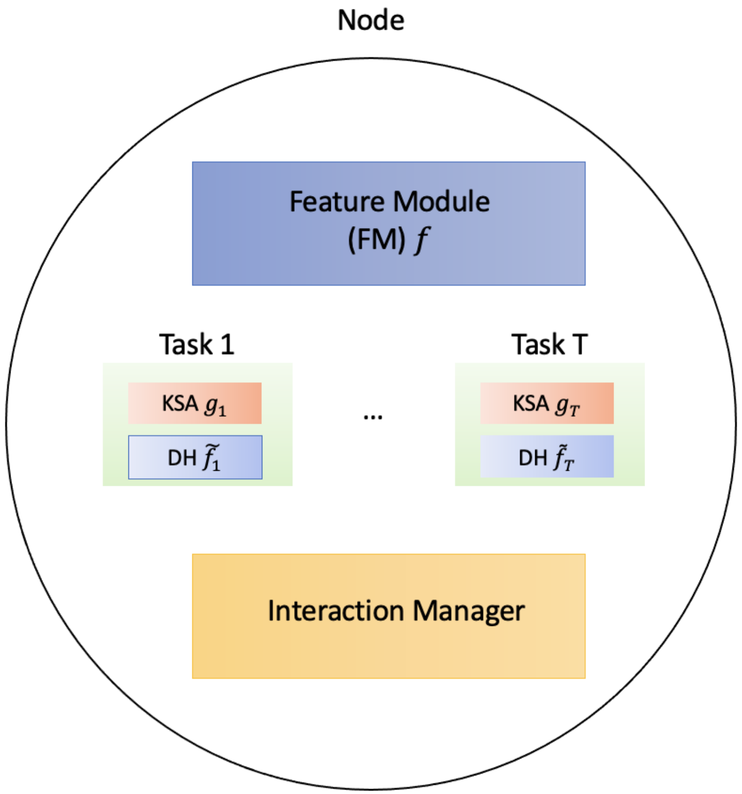


LENC Node architecture.

LENC Framework

Policies to compute q_k for teacher selection when the k^{th} node is aware about the task:

- ***Accuracy***: q_k can be the optionally stored average classification accuracy a_j^n.
- ***ODD score***: q_k can be a function of an ODD score g_j internally computed by the j^{th} KSA module of the k^{th} node given $\mathcal{D}'$.
- ***Disagreement***: q_k can be a scalar measure of the disagreement between the current Student LENC node and the k^{th} LENC node.



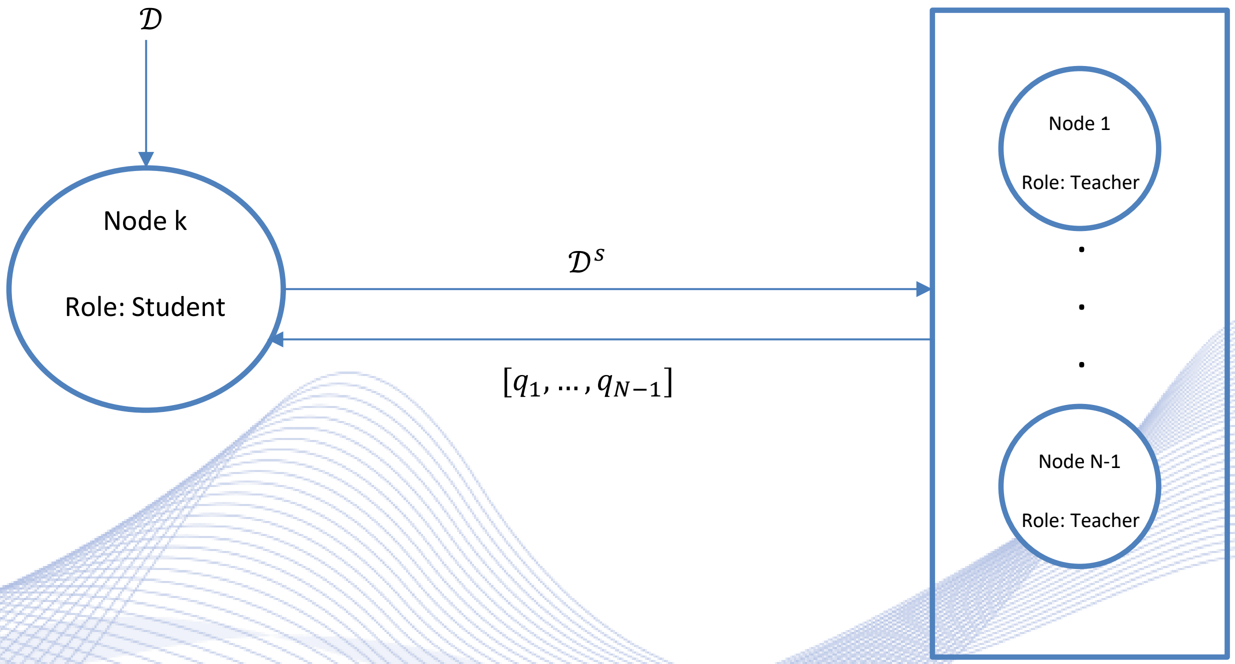
LENC Node architecture.

Artificial Intelligence & Information Analysis Lab

LENC Framework

LENC Teacher selection.

- The External Environment sends an input data set $\mathcal{D}$ to LENC student node k.
- Its KSA Module checks if the data distribution is known.
- If not, the data stream is sent to other nodes.
- The nodes respond with q_i, $i = 1, \cdots, N, i \neq k$.
- ***The student selects one (best) or more teachers***, based on the scalar metric q_i measuring their performance on $\mathcal{D}$.



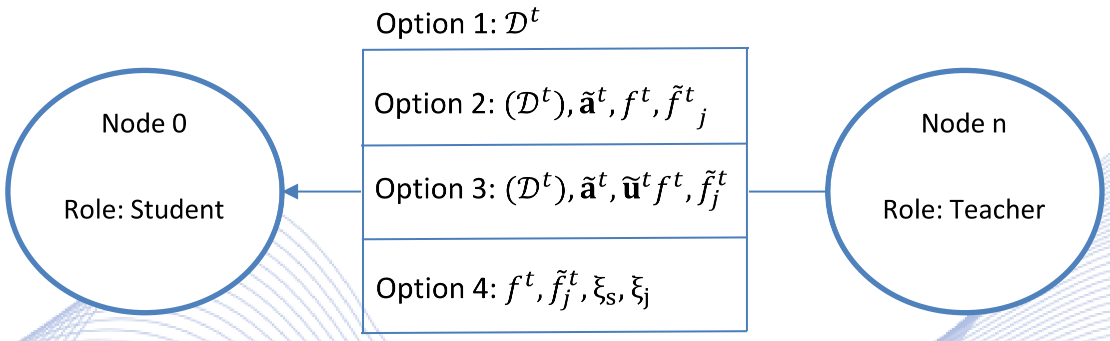
LENC Teacher selection.

LENC Framework

LENC Student node training (option 1):

Training Data Transmission.

- The Teacher LENC node sends its related training data set $\mathcal{D}^t$ to the Student LENC node.
- The Student LENC uses these training data to learn the new task.



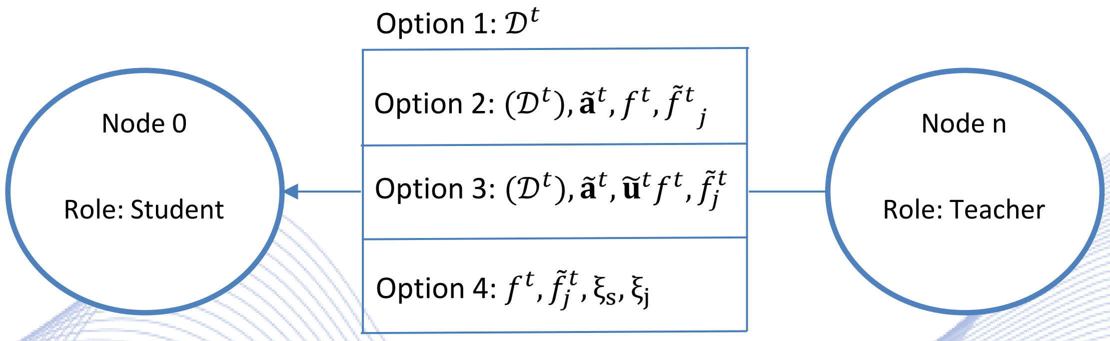
The LENC framework transfer learning options

LENC Framework

LENC Student node training (option 2):

Soft-Output Activation Transmission.

- The Teacher LENC Node sends its training data set $\mathcal{D}^t$, its soft-output activations $\tilde{\mathbf{a}}^t$ and its FM structure $\mathcal{S}_s$ (for and DH structure $\mathcal{S}_j$ for the task j.
- The Student LENC node uses KD to for training using Teacher LENC node guidance.



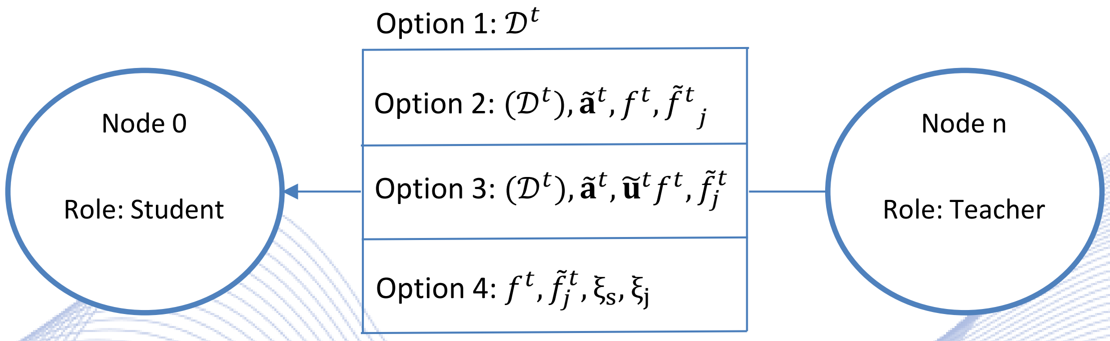
The LENC framework transfer learning options

LENC Framework

LENC Student node training (option 3):

Feature Activation Transmission.

- LENC Teacher node sends its training data set $\mathcal{D}^t$, its soft-output activations $\tilde{\mathbf{a}}^t$, its feature activations $\tilde{\mathbf{u}}^t$ and its structure $\mathcal{S}_s$ and $\mathcal{S}_j$ for the task j.
- Student LENC node uses KD to for training using the teacher's guidance.



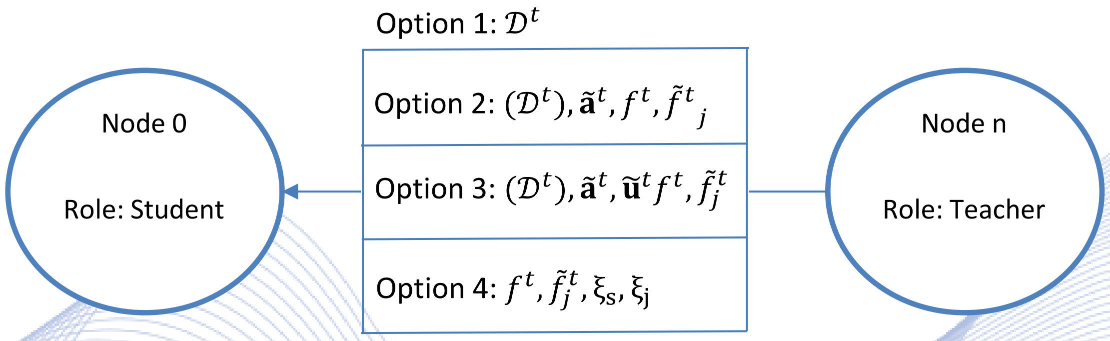
The LENC framework transfer learning options

LENC Framework

- ***LENC Student node training*** (option 4):

DNN weights transmission.

- Teacher LENC node sends its FM and DH structures $\mathcal{S}_s, \mathcal{S}_j$ and its FM and DH weights $\mathbf{w}_s, \mathbf{w}_j$ for the task j.
- The Student LENC node just copies of the Teacher model f^t and $\tilde{f}_j^t$.

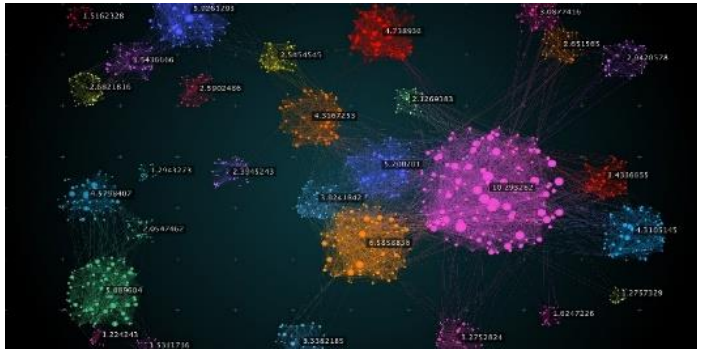


The LENC framework transfer learning options

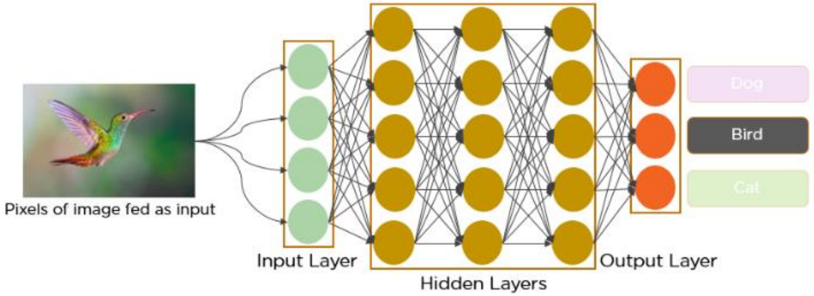
Decentralized DNN Architectures

- Decentralized DNN Architectures
- Learning-by-Education Node Community (LENC) Framework
- **LENC Framework Applications**
- LENC Framework Experiments
- LENC Architecture Implementation

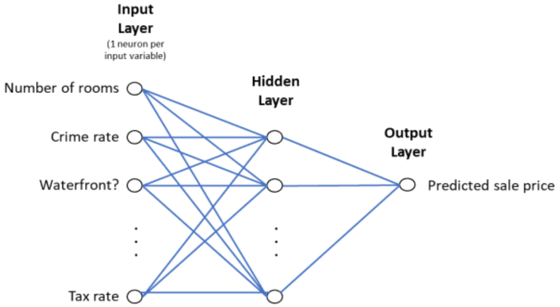
LENC Framework Applications



Clustering.

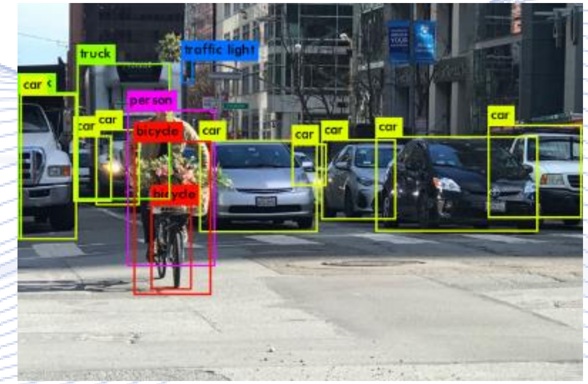


Classification.



Regression.

Image segmentation.



Object detection.

LENC Framework Applications

Federated Learning: Training a global DNN model across decentralized nodes, while keeping data on-device.

- ***Privacy Preservation***: Data remain on local devices, ensuring privacy.
- ***Communication efficiency***: No large data volume transfer to a central server is needed.
- ***Scalability***: Large-scale diverse data sources can be accomodated.
- ***Adaptability***: Non-identically distributed data can be supported.
- ***Distributed*** rather than decentralized DNN FL computing.

LENC Framework Applications

Federated Learning

- One LENC node is the master node (***aggregator***).
- All LENC nodes have the same structure $\mathcal{S}$ and are trained using their local data.
- The master node uses training option 4 to receive the weights of all other nodes with the same structure within the community.
- The master node aggregates the weights of all participating nodes.
- The process is repeated until convergence.

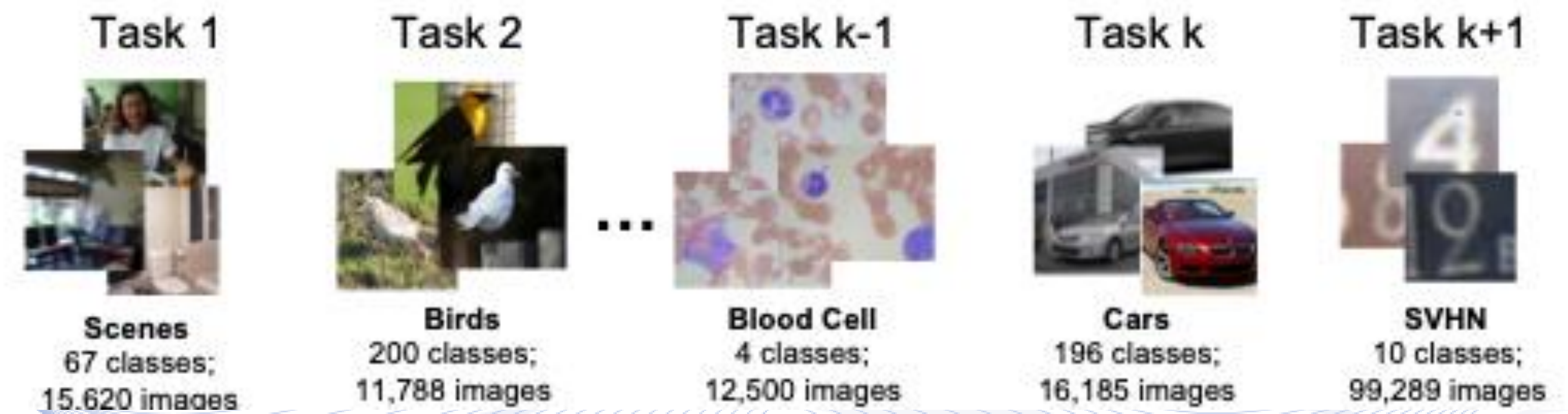
LENC Framework Applications

Peer-to-Peer Learning

- LENC node training options 1-4 constitute forms of Peer-to-Peer Learning.
- Nodes act exclusively to enhance their knowledge.
- No need for a central server.
- Retaining knowledge within the node community.

LENC Framework Applications

Continual Learning



Learn different tasks (with different semantic classes) sequentially, without forgetting.

Artificial Intelligence & Information Analysis Lab

LENC Framework Applications

Edge Computing – Decentralized Inference

- Raw data is processed locally on LENC nodes.
- Nodes use real-time inference on their data.
- Lightweight training of Decision Modules is done directly on nodes.
- A master node (server) can be defined to aggregate inference results.
- Inference can local without centralized decision-making.

LENC Framework Applications

DNN performance 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 DNN model behavior and also consider possible privacy constraints.
- Private weights, architecture, training dataset, etc.

Decentralized DNN Architectures

- Decentralized DNN Architectures
- Learning-by-Education Node Community (LENC) Framework
- LENC Framework Applications
- **LENC Framework Experiments**
- LENC Architecture Implementation

CKD Experiment

Collaborative Knowledge Distillation (CKD) ***Experiment***.

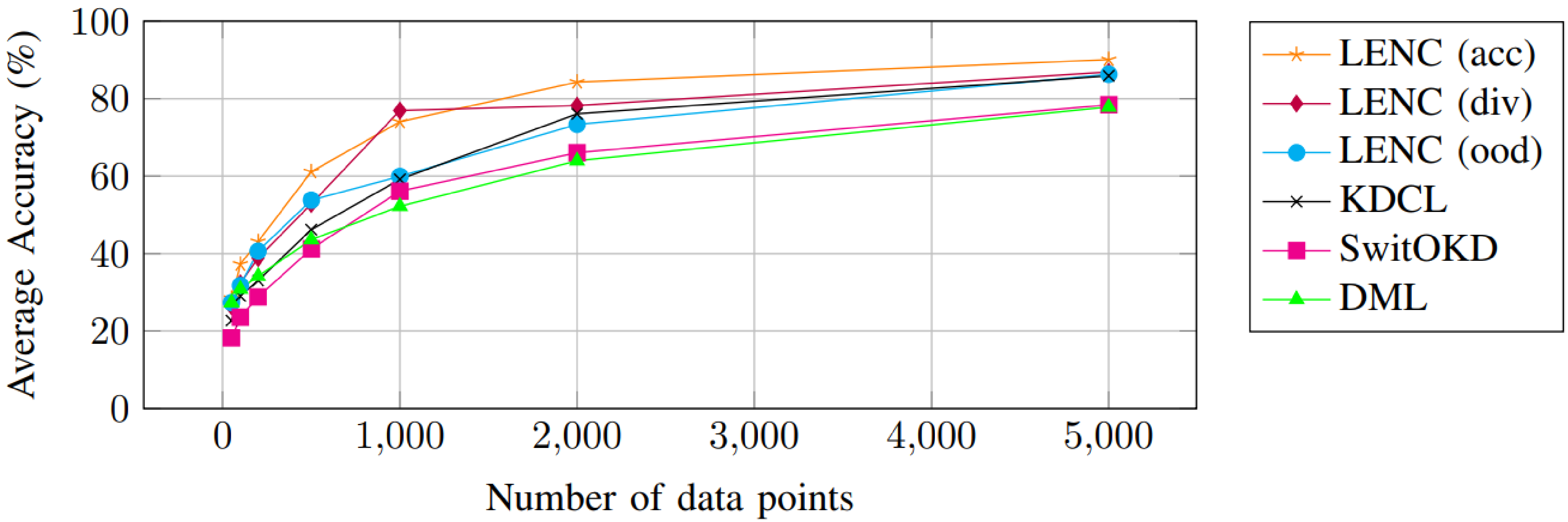
- Four LENC nodes are initialized.
- One of them (Teacher LENC node) is pretrained on a ***classification*** dataset (CIFAR10, SVHN, MNIST, FashionMNIST).
- Each node (including the Teacher) takes the LENC Student role exactly once every ***education cycle***.
- All Student LENC nodes use the LENC framework option 2 (Knowledge Distillation).
- After 5 education cycles, we observe the results.

CKD Experiment

Average test classification accuracy (%) of the 3 student LENC nodes and of competing CKD methods, for incoming data sets $\mathcal{D}^t$ having 1000 or 5000 samples (from C10 or C100 CIFAR).

Dataset	Students	Stream Size	DML	KDCL	SwitOKD	LENC (proposed)
C10	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**
C100	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**

CKD Experiment



Average student LENC node classification accuracy (%) for varying D^s sizes in the CIFAR-10 dataset for 3 alternative LENC teacher selection policies against that of competing methods.

CKD Experiment

Experimental conclusions

- LENC framework outperforms existing CKD methods, when digesting un-labelled incoming data samples, under the assumption that the sole expert indeed knows data similar to the incoming ones.

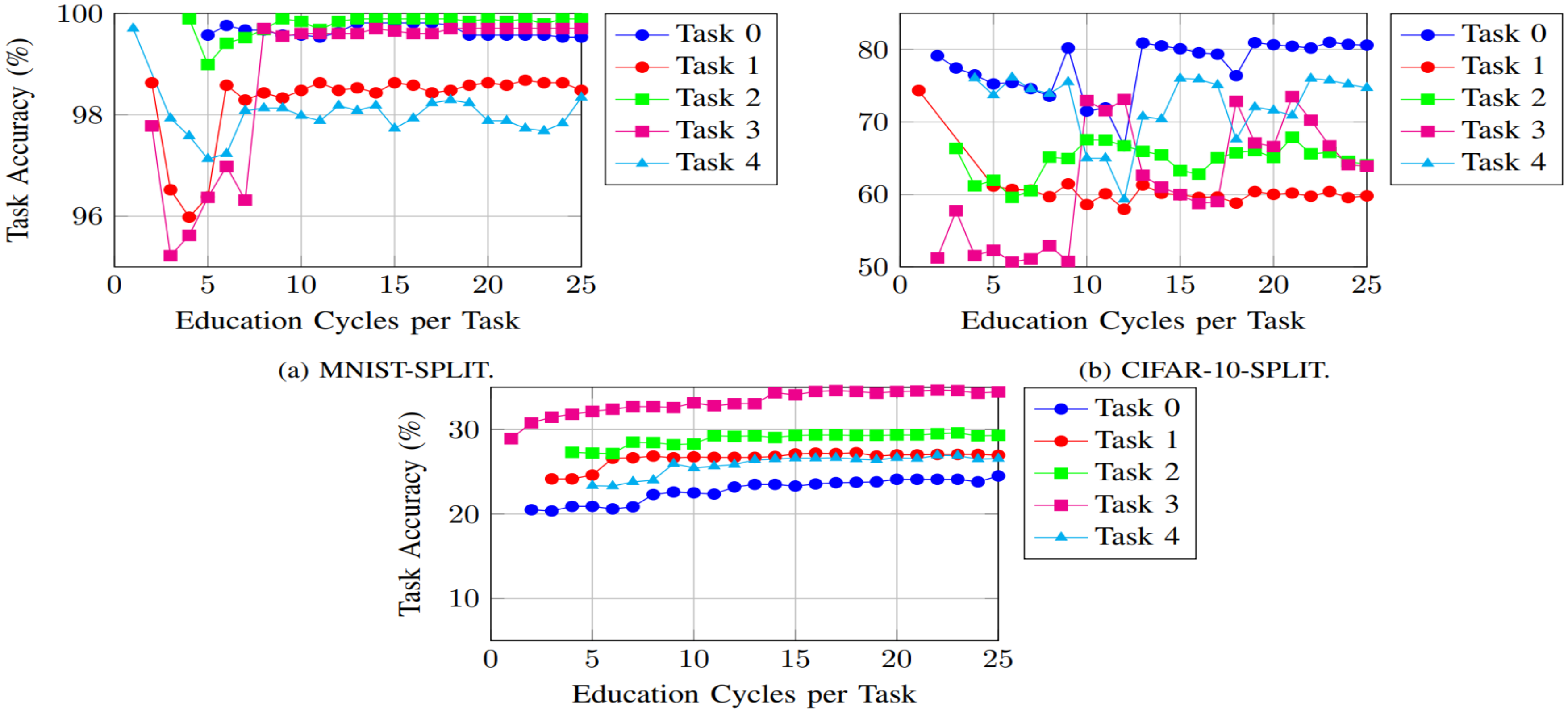
- LENC proves to be the most tolerant to small batch sizes, thus showcasing its usability in an important real world use-case: when a node faces unknown current input data and needs to acquire relevant knowledge as soon as possible, in order to respond immediately.

CL Experiment

Continual Learning (CL) Experiment.

- Six LENC nodes are initialized for each dataset.
- Five of them (Teacher LENC nodes) are trained on a task of the ***classification*** datasets (SPLIT-MNIST, SPLIT-CIFAR-10 and SPLIT-CIFAR-100).
- For example for the SPLIT-MNIST dataset: Node 1 knows classes {0,1}, Node 2 knows classes {2,3} etc.
- The Student LENC node encounters all tasks for a single education cycle and picks the correct teacher for each task.
- The Student LENC nodes use the LENC framework option 2 (Knowledge Distillation) and a ***specialized CL loss*** to learn new tasks without forgetting.
- After 5 education cycles, we observe the results.

CL Experiment



Student classification accuracy per task for a) SPLIT-MNIST, b) SPLIT-CIFAR-10 and c) SPLIT-CIFAR-100.

Artificial Intelligence & Information Analysis Lab

CL Experiment

Experimental conclusions.

- The LENC framework can achieve continual learning and adaptation with only a few randomly sampled batches.

Federated Learning Experiment

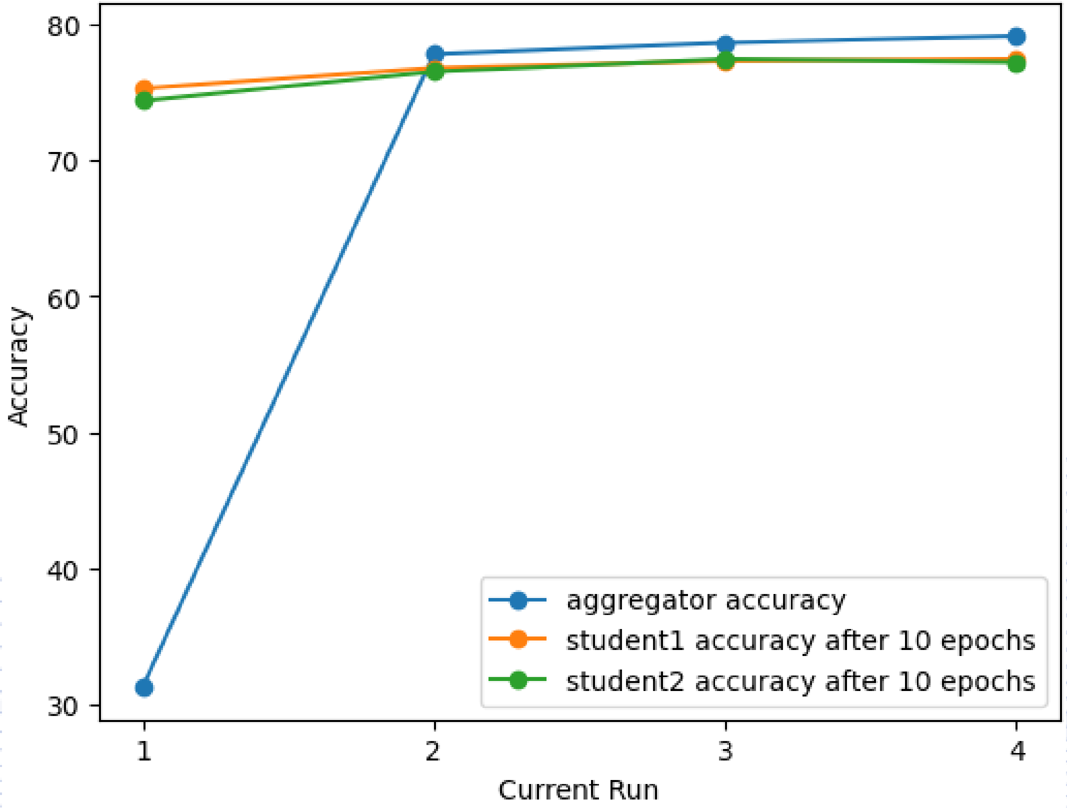
- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Learning
- Learning-by-Education Node Community (LENC) Framework
- Collaborative Knowledge Distillation (CKD) Experiment
- **Federated Learning Experiment**
- Decentralized DNN (D-DNN) Consensus Inference Experiment
- LENC Framework Applications
  - Deep Learning Tasks Supported by LENC Framework
  - Teacher-Classroom Classification
  - Federated Learning
  - Peer-to-Peer Learning
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
- LENC Architecture Implementation

Federated Learning Experiment

Federated Learning Experiment Setup

- 1 master LENC node (aggregator), 2 student LENC nodes with unique datasets.
- Students are trained locally for 50 epochs and send output to Aggregator.
- Aggregator calculates the mean DNN model weights (global) and sends them to student nodes.
- 4 FL rounds in total.

Federated Learning Experiment



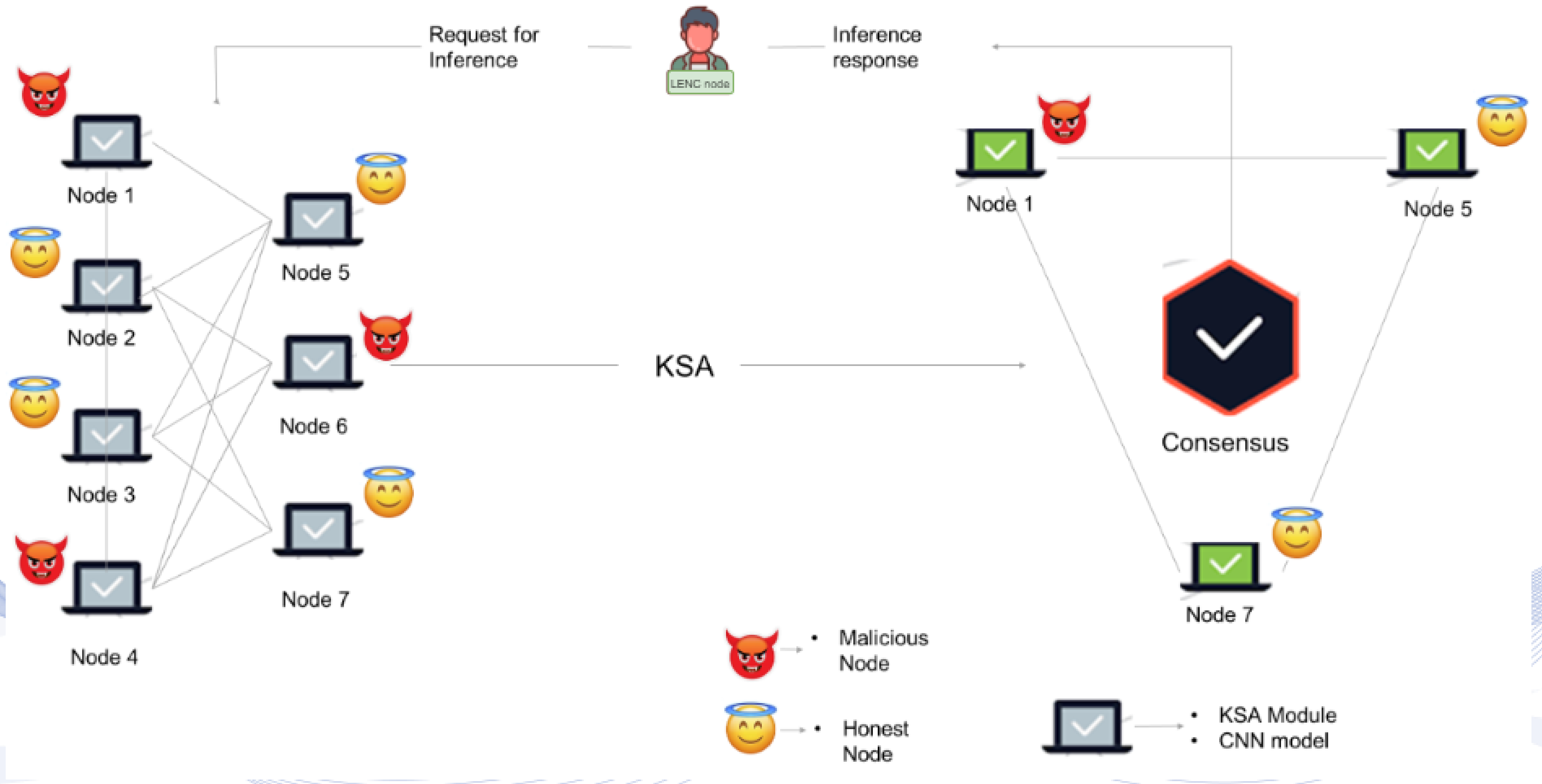
Accuracy report (%) of the aggregator and the two students after each federated run on Cifar 10 test dataset.

D-DNN Inference

Decentralized DNN (D-DNN) Consensus Inference

- Any LENC (master) node can make a request to the LENC community to perform ***Decentralized DNN*** (D-DNN) ***inference*** by providing its own dataset.
- Through the KSA module, LENC nodes that are familiar with the master node data distribution are selected to carry out the inference.
- ***Proof of Quality Inference*** (POQI) Consensus Protocol, is used to establish consensus between the selected teachers regarding their DNN Inference outputs.
- ***Security and integrity*** of the inference results reported to the client are ensured by detecting and excluding malicious DNN nodes.

D-DNN Inference



A master LENC node finds related LENC nodes for POQI consensus.

D-DNN Inference

Dataset	Faulty Nodes	Method	Accuracy (%)						
			N1	N2	N3	N4	N5	N6	N7
Cifar-10	0	Weighted Average	95.12	95.12	95.12	95.12	95.12	95.12	95.12
		Majority Voting	95.05	95.05	95.05	95.05	95.05	95.05	95.05
		PoQI	**95.27**	**95.27**	**95.27**	**95.27**	**95.27**	**95.27**	**95.27**
Cifar-10	1	Weighted Average	16.40	15.87	15.92	16.24	15.35	16.11	-
		Majority Voting	94.63	94.86	94.76	**95.02**	94.72	94.56	-
		PoQI	**94.99**	**94.99**	**94.99**	94.99	**94.99**	**94.99**	-
SVHN	1	Weighted Average	15.27	15.41	15.37	15.33	-	15.13	15.52
		Majority Voting	93.21	93.36	93.17	93.12	-	93.04	**93.77**
		PoQI	**93.42**	**93.42**	**93.42**	**93.42**	-	**93.42**	93.42
SVHN	3	Weighted Average	-	11.14	11.40	-	11.16	11.36	-
		Majority Voting	-	92.56	93.12	-	92.94	91.82	-
		PoQI	-	**93.18**	**93.18**	-	**93.18**	**93.18**	-

LENC node N1-M7 classification accuracy (%) comparison in the presence of 1-3 malicious nodes.

Decentralized DNN Architectures

- Decentralized DNN Architectures
- Learning-by-Education Node Community (LENC) Framework
- LENC Framework Applications
- LENC Framework Experiments
- **LENC Architecture Implementation**

LENC Architecture Implementation

- Decentralized DNN Architectures
  - Federated Learning
  - Edge Computing
  - Peer-to-Peer Learning
- Learning-by-Education Node Community (LENC) Framework
- Collaborative Knowledge Distillation (CKD) Experiment
- Federated Learning Experiment
- Decentralized DNN (D-DNN) Consensus Inference
- LENC Framework Applications
  - Deep Learning Tasks Supported by LENC Framework
  - Teacher-Classroom Classification
  - Federated Learning
  - Peer-to-Peer Learning
  - Continual Learning
  - Edge Computing – Decentralized Inference
  - Reproducibility – Privacy
- **LENC Architecture Implementation**

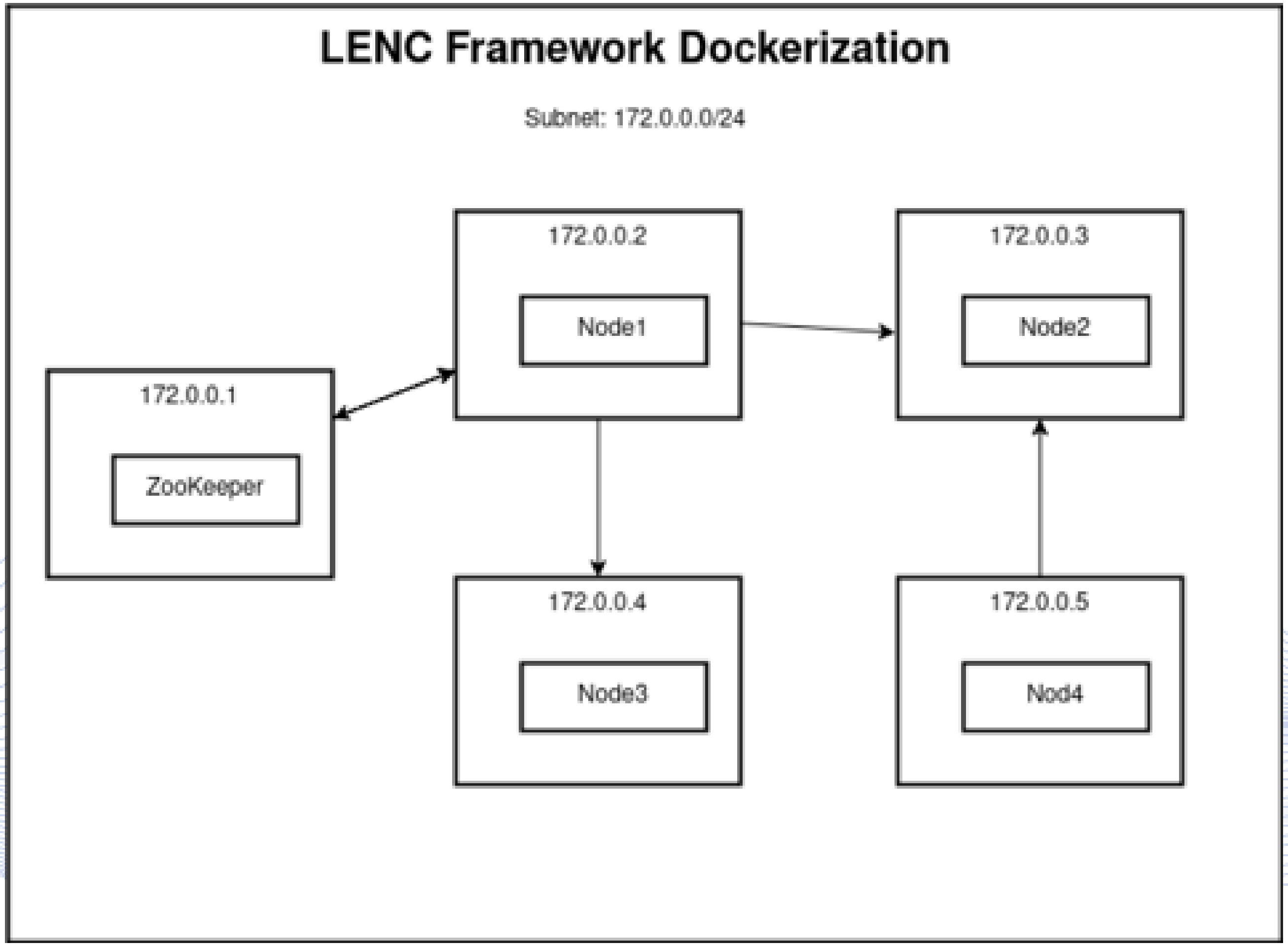
LENC Architecture Implementation

- We implement the LENC framework on multiple devices by using ***Docker*** and ***Zookeeper***.
- Now each LENC node is physically located in a single terminal.
- The LENC nodes use the Zookeeper for “online” ***node discovery*** and receive the IP addresses of all “online” LENC nodes.

LENC Architecture Implementation

- Each LENC node occupies one ***Docker container***.
- Zookeeper-like instance for service discovery and coordination.
- Environment simulation using Docker network capabilities:
  - The network has a predefined IP mask (172.0.0.0/24).
  - Each container has its own virtual IP address (172.0.0.1-.256).
- All communications (including file sharing) use ***sockets***:
  - Each node acts both as a server and client.
  - All listening on port 60.000.
  - Zookeeper webservice listens on port 8080.

LENC Architecture Implementation



LENC Docker network outline.

Artificial Intelligence & Information Analysis Lab

LENC Architecture Implementation

Dockerized LENC training procedure

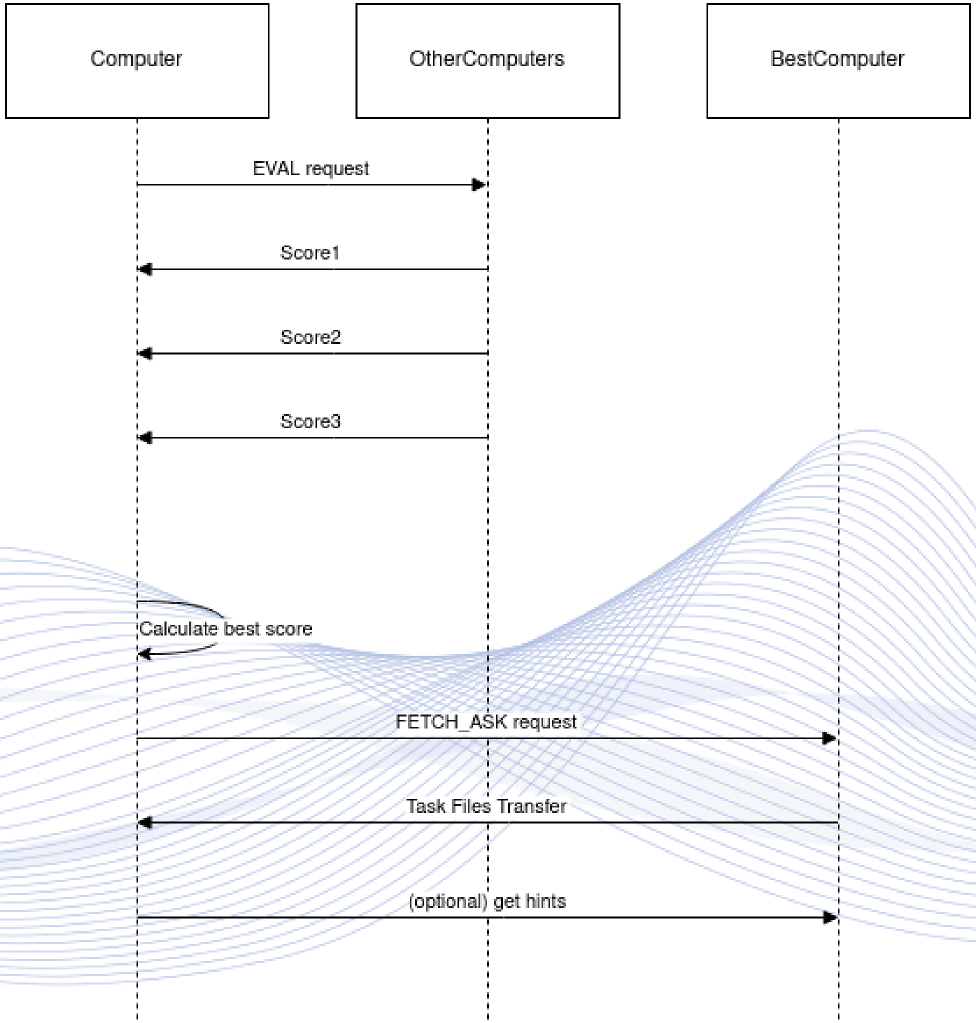
- Teacher LENC nodes download datasets and teach their underlying DNNs by creating new tasks.
- A Student LENC node with a novel dataset can search for the most suitable Teacher by sending "EVAL" requests, along with the dataset to all available LENC nodes.
- They test the dataset to their networks and return the resulting accuracy score to the Student LENC node.
- The Student LENC node picks the best teacher, e.g., the one with the max recognition score.

LENC Architecture Implementation

Dockerized LENC training procedure

- The Student LENC node requests the files needed according to the option configuration (FETCH TASK request) from the picked potential Teacher LENC node.
- Files are ‘.bin’ for datasets and ’.pth’ for teacher soft-output activation, weights and structure.
- The Student LENC node uses the received files to train a new Decision Head, but without forgetting the previous tasks learned.

LENC Architecture Implementation



TEMA training process flow.

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*.

[KAI2023] Kaimakamidis, A., & Pitas, I. (2023). Facilitating Experimental Reproducibility in Neural Network Research with a Unified Framework. In Proceedings of the IEEE/ACM 10th International Conference on Big Data Computing, Applications and Technologies (BDCAT ‘23) (Article 14, 1–5). Association for Computing Machinery.

[PAP2024] D. Papaioannou, V. Mygdalis, I. Pitas. (2024). Proof of Quality Inference (PoQI): An AI Consensus Protocol for Decentralized DNN Inference Frameworks. In Proceedings of the IEEE/ISCC 4th International Workshop on Distributed Intelligent Systems.

Bibliography

[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.

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