Meta Learning in Decentralized Neural Networks: Towards More General AI

Yuwei Sun

The University of Tokyo, RIKEN

Decentralized neural networks

Key idea: Performing knowledge sharing and social learning among neural network models can benefit their performance in unseen tasks.

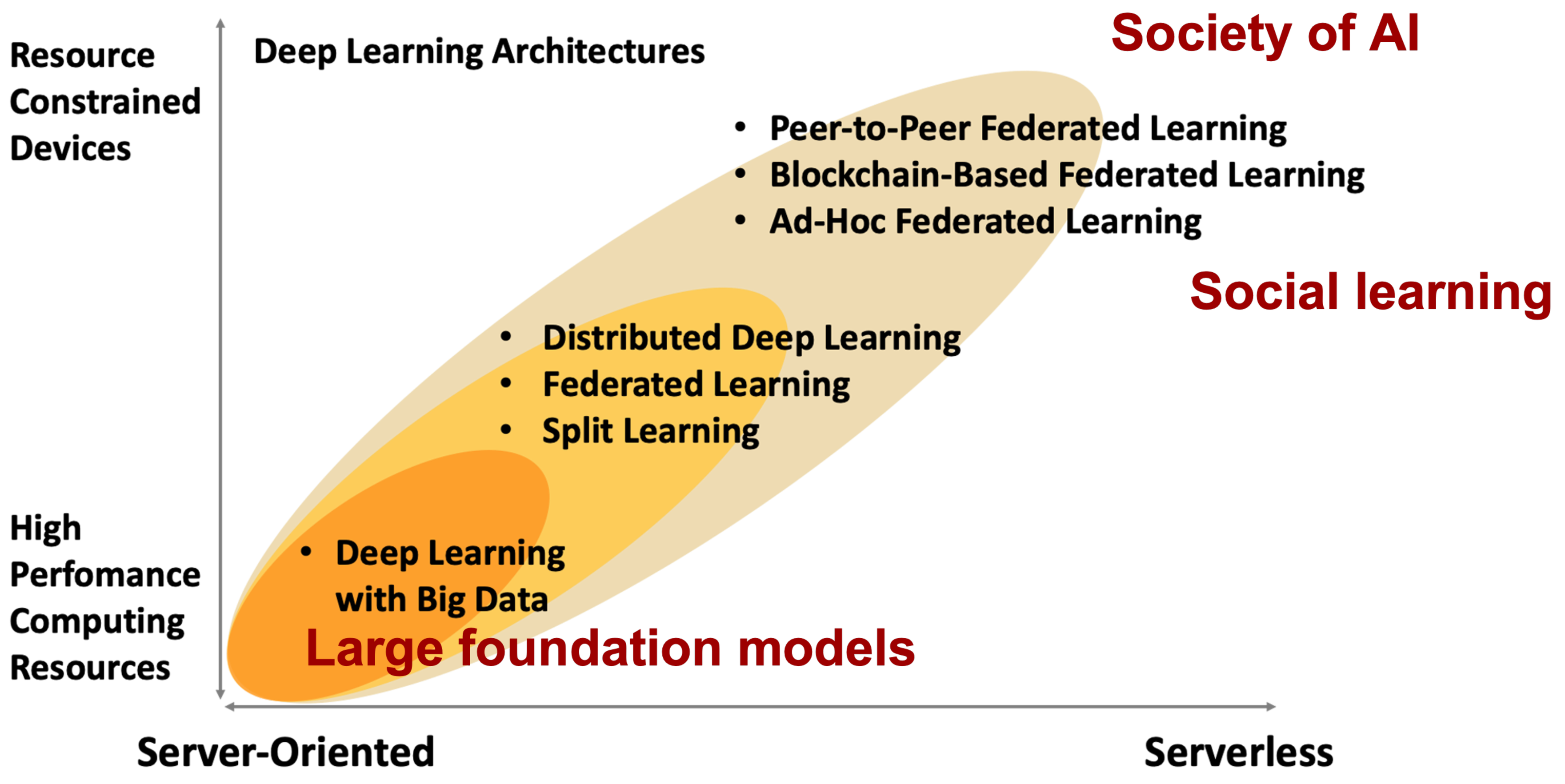

Figure 1. Decentralized neural networks are built upon the multi-access edge computing of different models [1].

- It is intractable to build a predictive model that can represent all possible predictions simultaneously and is generalizable to all types of unseen data. A more effective approach is to leverage knowledge from previous experiences across different neural networks when tackling new tasks.
- The learned knowledge refers to either the model parameters, or the processed representations, such as feature maps, produced by these models.
- Similar to modular neural networks, a complex problem is divided into smaller subproblems that can be solved by specialized sub-networks. The cognitive architecture theory of Global Workspace explains how neural networks cooperate and compete to solve problems via a shared feature space for knowledge sharing.

Learning from replica neural networks

Key idea: Create a network of interconnected neural network models with similar architecture and functionality.

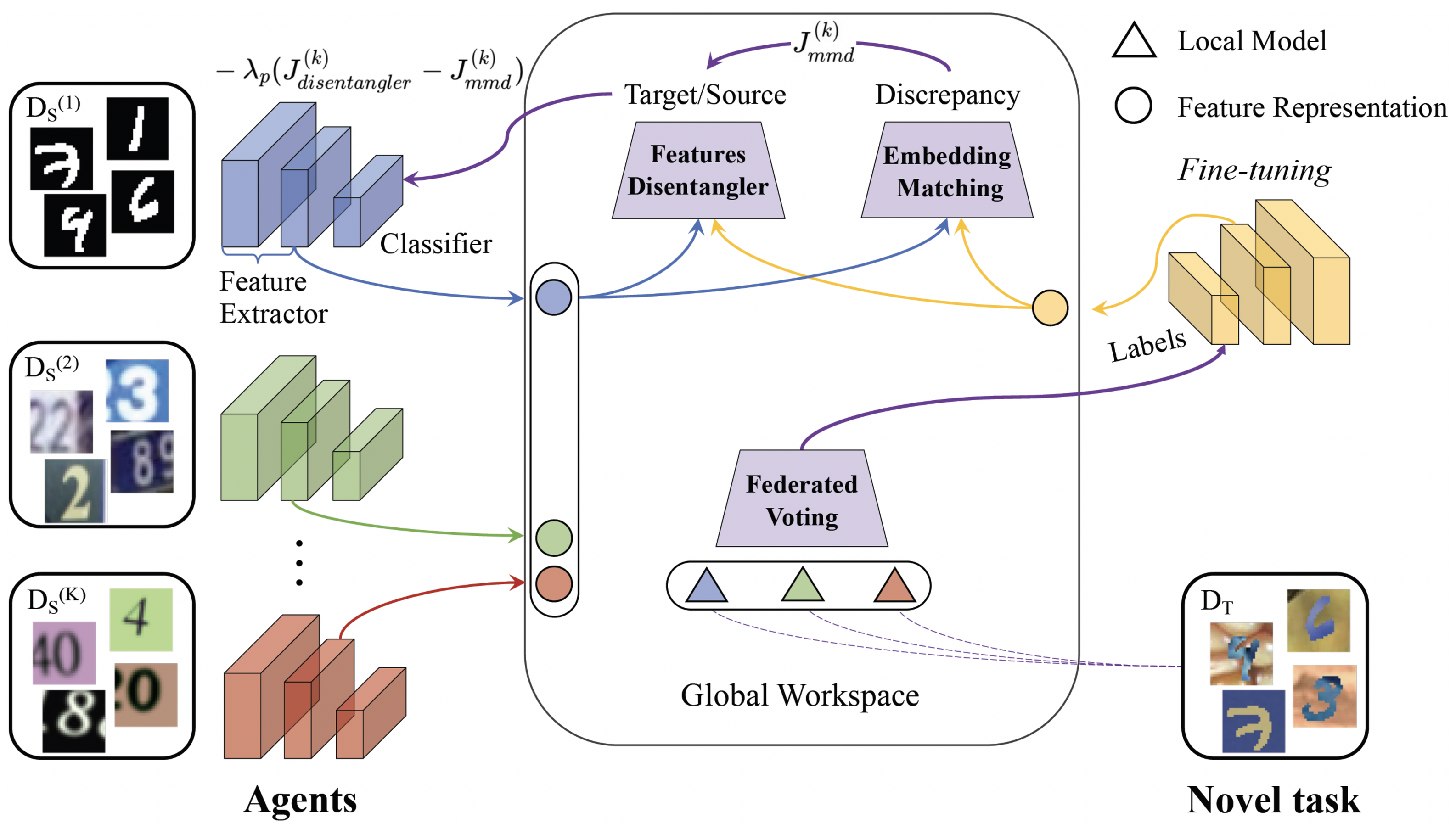

Figure 2. Reusing distilled knowledge and discarding irrelevant information from a trained model on a source domain enables more efficient training in the target domain [2].

- When the data distributions during training and testing are different, the performance of the trained model on the test data will be degraded, referred to as the Distributional Shift problem.
- Multiple agents are trained to solve separate tasks using samples with different styles due to their data collection environments. The goal is to tackle a new task leveraging the reusable information from these agents' local tasks.
- To improve the performance in an unseen target domain, we propose a multi-source domain adaptation method in federated learning, by disentangling domain-invariant stationary features and reusing model parameters of different agents.

Building a hierarchy of neural networks

Key idea: Hierarchical neural networks consist of multiple neural networks in a form of an acyclic graph, each with a specific task. A meta model learns the optimized policy to facilitate learning in hierarchical neural networks.

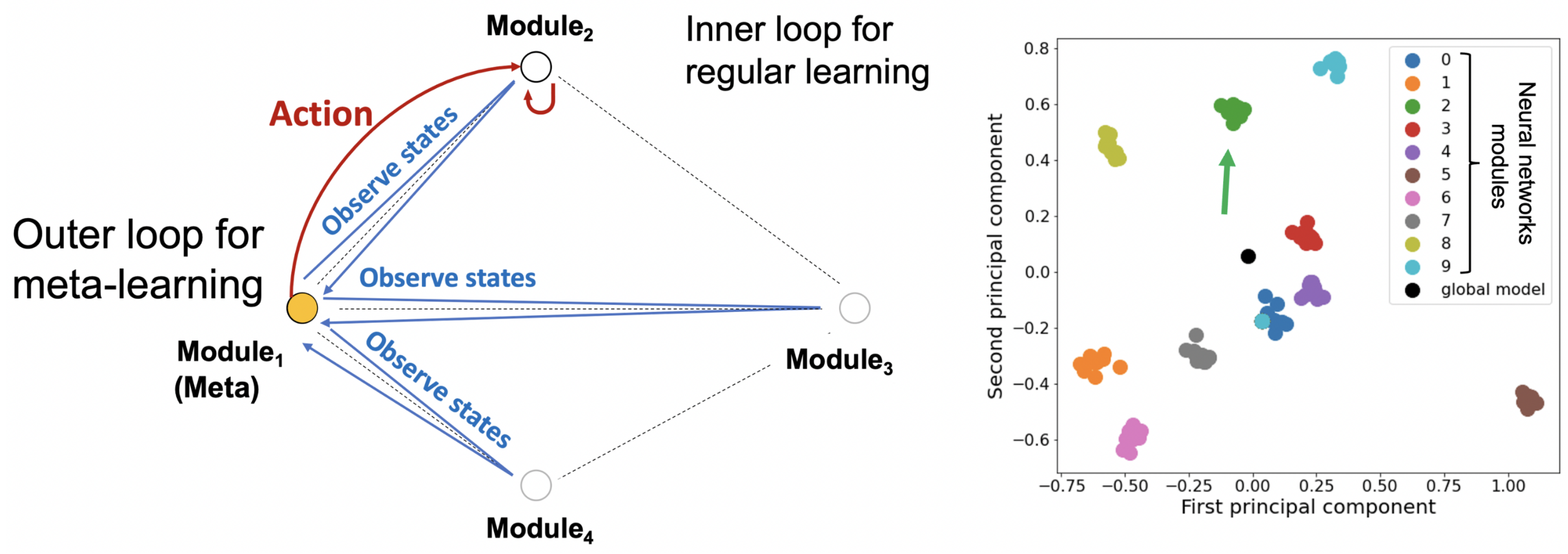

Figure 3. The hierarchical neural networks leverage multiple fast inner learning loops for task-specific knowledge and a slow outer learning loop for a policy to efficiently reuse these task-specific knowledge [3].

- The goal is to learn the optimized learning policy of leveraging different models' knowledge using reinforcement learning.
- The connections among models in hierarchical neural networks are formed on the fly by trying out different alternatives.
- Improved performance and reduced training cost over time.

Leveraging different modality experts

Key idea: Self-supervised learning enables knowledge transfer across modalities without labels, obtaining reusable cross-modal representations for downstream tasks.

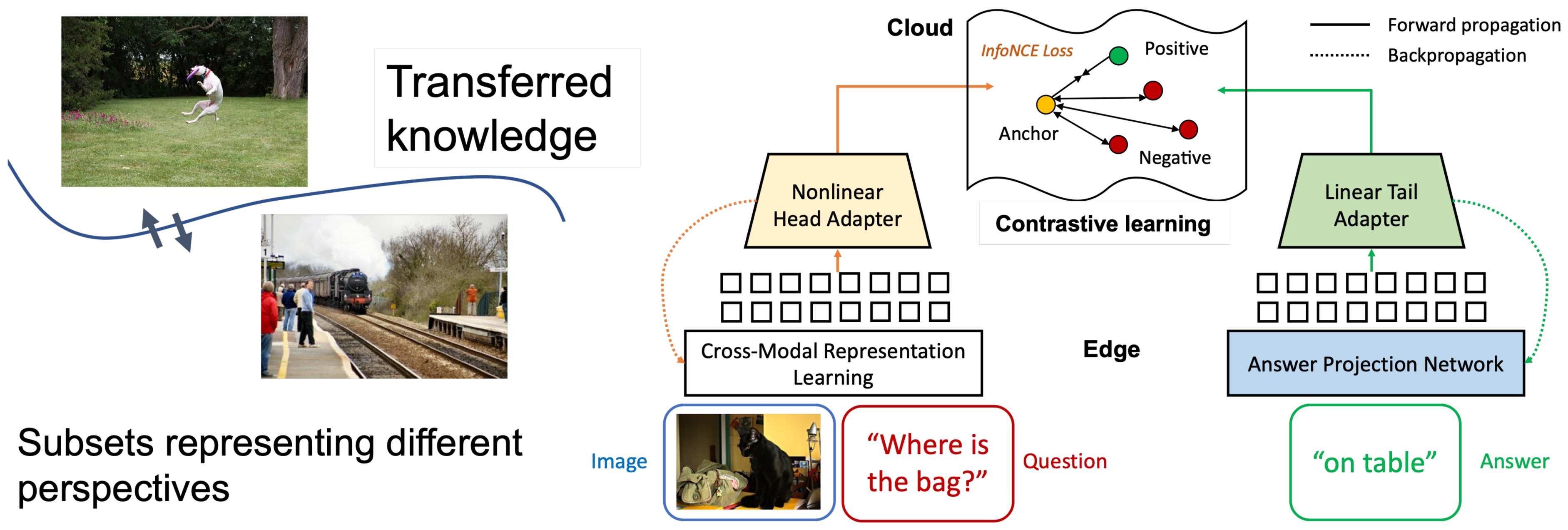

Figure 4. We use contrastive learning to train multiple models on different data distributions for Visual Question Answering, and then combine the learned knowledge to improve performance on various tasks [4].

- In practical applications, it is often the case that multiple models are employed for different users, each with a unique perspective on the data distribution.
- Transferring and combining knowledge from these models effectively addresses future tasks.
- We align representations of different modalities by using contrastive learning to encourage the similarity of relevant model outputs and discourage irrelevant ones.
- Learned edge models are aggregated and broadcast to users for improved model generality.

References

[1] Yuwei Sun, Hideya Ochiai, and Hiroshi Esaki. Decentralized deep learning for mobile edge computing: A survey on communication efficiency and trustworthiness. *IEEE Transactions on Artificial Intelligence*, 2022.

[2] Yuwei Sun, Ng Chong, and Ochiai Hideya. Feature distribution matching for federated domain generalization. *ACML*, 2022.

[3] Yuwei Sun and Hideya Ochiai. Homogeneous learning: Self-attention decentralized deep learning. *IEEE Access*, 2021.

[4] Yuwei Sun and Hideya Ochiai. Unicon: Unidirectional split learning with contrastive loss for visual question answering. *NeurIPS workshop*, 2022.

[5] Yuwei Sun. Meta learning in decentralized neural networks: Towards more general ai. *AAAI*, 2023.

ywsun@g.ecc.u-tokyo.ac.jp