IMPERIAL

Representation Learning Through Collaboration

Stefanos A Frilingos[1, 2]

Introduction

A novel method for a network of agents to learn a shared task while seeing the world through a unique lens.

Core Technologies

Self-Supervised Learning (SSL): learn meaningful features from unlabeled data. By comparing augmented views of an image, the model builds a **robust visual grammar** without manual labeling.

VICReg avoids representation collapse by regularizing the **Variance**, **Invariance** and **Covariance** of the learned features.

Decentralized Optimization: a network of agents collaboratively train a model **without a central server**. Agents communicate directly with peers in a graph, sharing knowledge while keeping data private.

Problem: Learning Different Domains
Imagine autonomous farm cameras, each with a systematic **domain shift** in its **grapes**. When each camera trains a model on its unique data, it learns an incompatible "visual language," leading to **misaligned** feature spaces. Naively averaging these models results in poor performance.

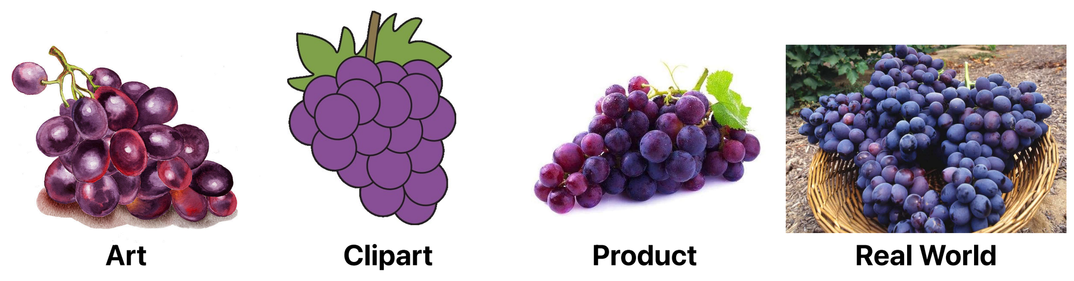

Figure 1: The 4 Unique Image Domains: Art, Clipart, Product, RealWorld

How can specialist agents with different data domains collaboratively train a robust, shared model that understands all their unique perspectives?

Our Approach

Use a small public set of images to enable specialist agents to collaborate on a **shared classifier** while maintaining their own **personalized backbones**.

Goal: Align the agents' latent spaces regardless of domain; same classes should map to a similar representation.

Algorithm Logic

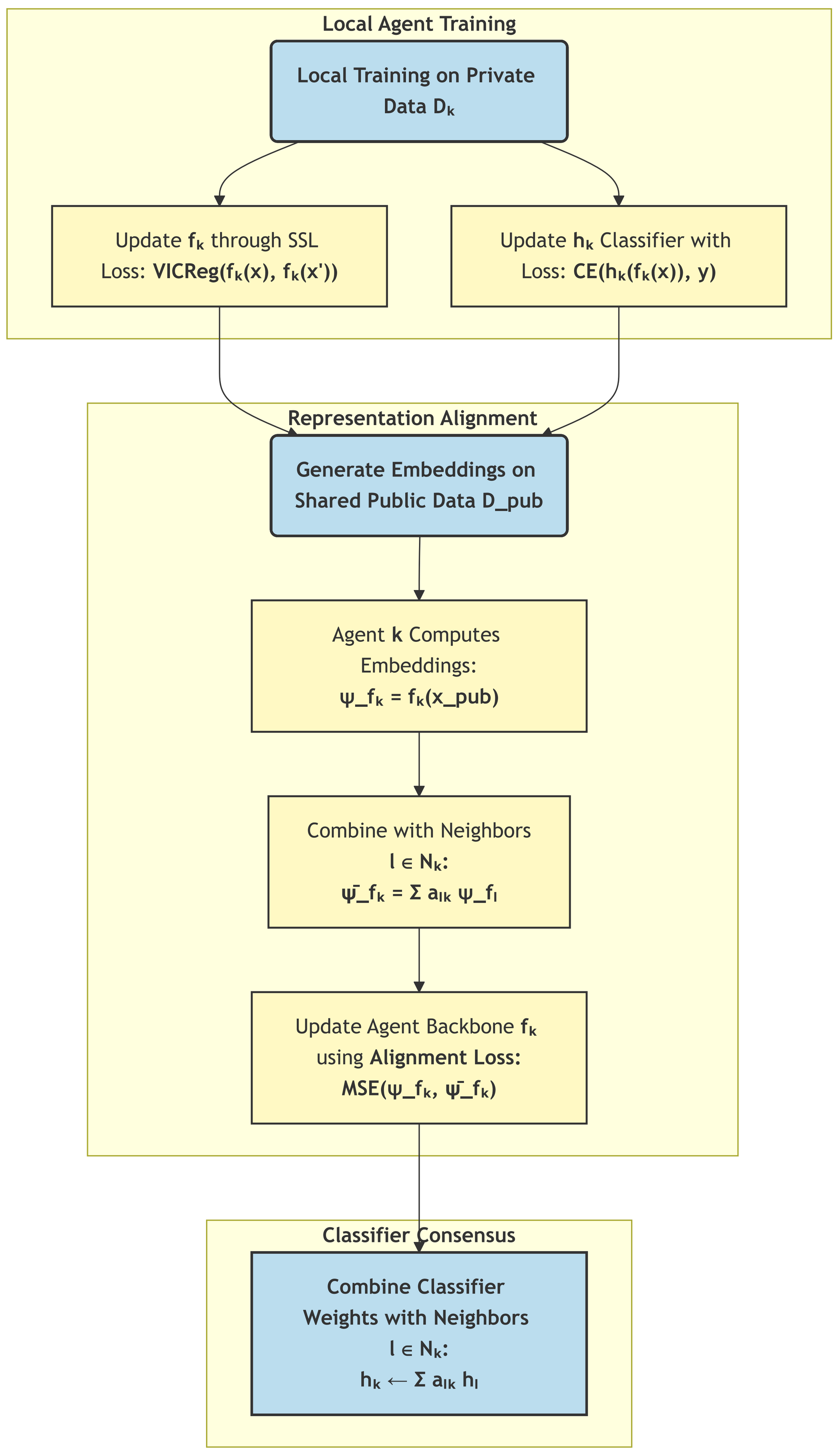

Figure 2: Flowchart of our novel peer-to-peer training protocol per communication round

Results

- Our method **outperforms centralized**, **federated**, and P2P **disconnected** baselines in multi-domain evaluation
- Removing either alignment or classifier sharing significantly degrades accuracy, highlighting the **necessity of dual-communication**
- The disconnected model achieves high k-NN accuracy (51.5%), indicating strong local feature learning, but fails at scale

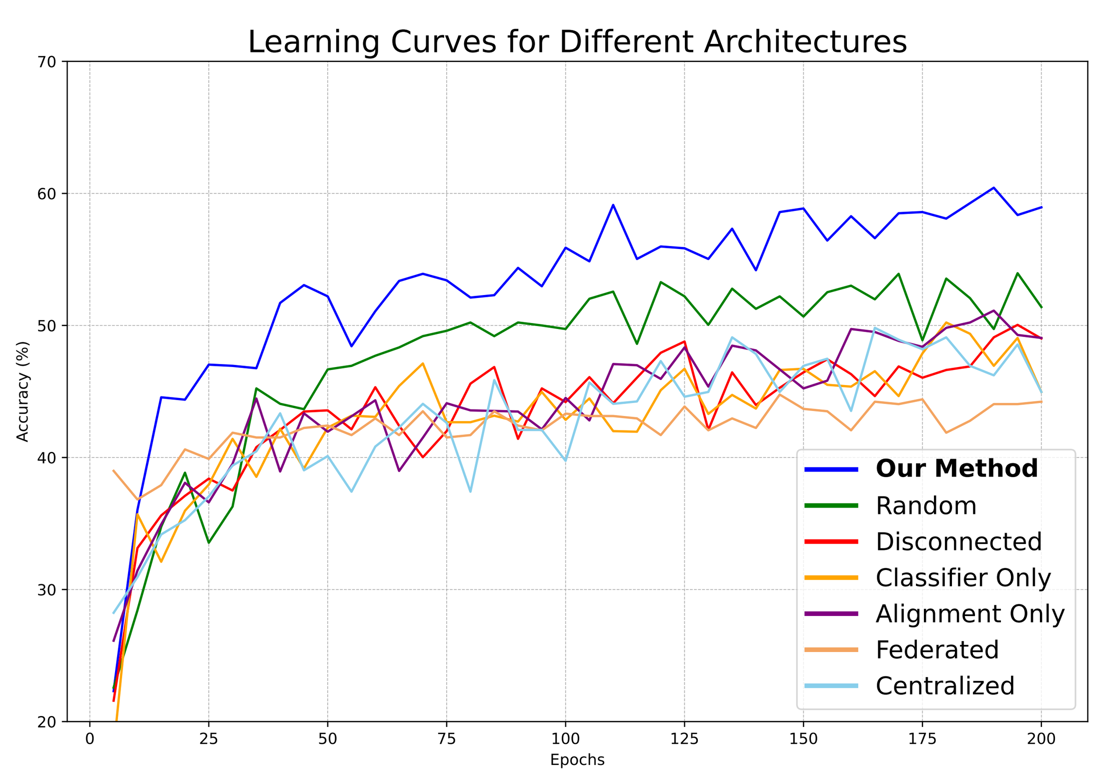

Figure 3: Learning curves on OfficeHome show our decentralized approach (P2P Fully Connected) achieves significantly higher final linear accuracy.

- A **random** topology trails the fully connected, but **still surpasses** centralized and federated baselines

Final Test Accuracy

Table 1: Final Top-1 Accuracy (%) on OfficeHome using a Linear NN Classifier and k-NN Voting Evaluation.

Architecture	Linear	k-NN
Centralized	48.9%	44.2%
Federated (FedAvg)	47.3%	45.7%
Ours (Connected)	**58.9%**	**51.9%**
P2P Disconnected	37.3%	51.5%

Analysis & Conclusion

Representation Alignment: Angle analysis shows our method keeps agents' **latent features aligned**, while disconnected agents diverge.

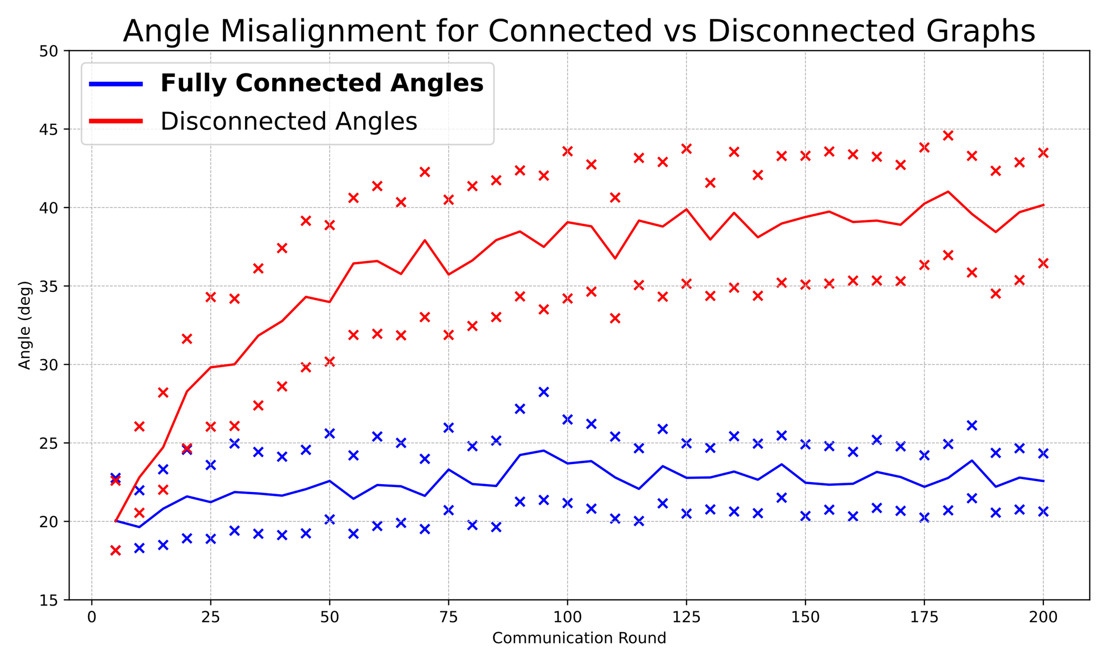

Figure 4: The median angle between agent representations trends downwards in our connected approach, proving successful alignment, while diverging in the disconnected case. The stars show the 25% and 75% percentile of angle disparity.

Feature Quality: t-SNE shows VICReg creates semantically meaningful, separable **clusters**.

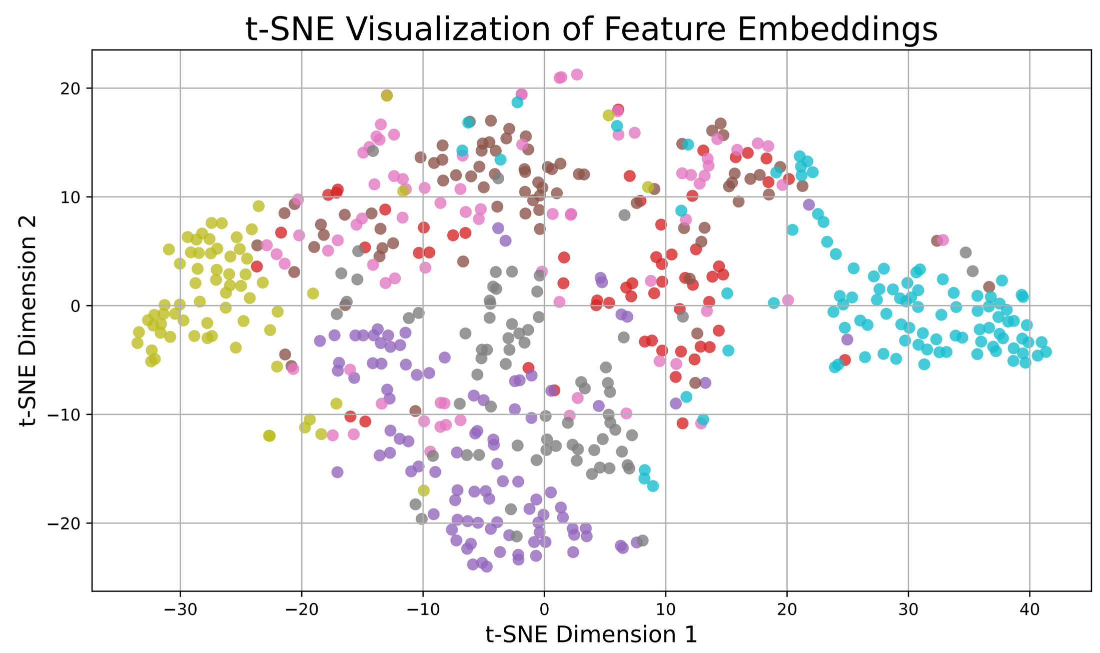

Figure 5: The final representation space of a connected agent shows semantically meaningful clusters, even with high intra-class variance. T-SNE is a non-linear dimensionality reduction technique for visualizing high-dimensional datasets in a low-dimensional space. Each colour signifies a different class with samples from all four domains

Conclusion

We overcome non-IID domains by aligning agents' representations into a shared feature space and with a shared classifier outperform baselines by over 10%.

Affiliations
[1] EEE, Imperial College London
[2] Communications & Signal Processing Group