FORLA: Federated Object-Centric Representation Learning with Slot Attention

Guiqiu Liao, Matjaz Jogan, Eric Eaton, Daniel A. Hashimoto

NeurIPS 2025

Penn
University of Pennsylvania

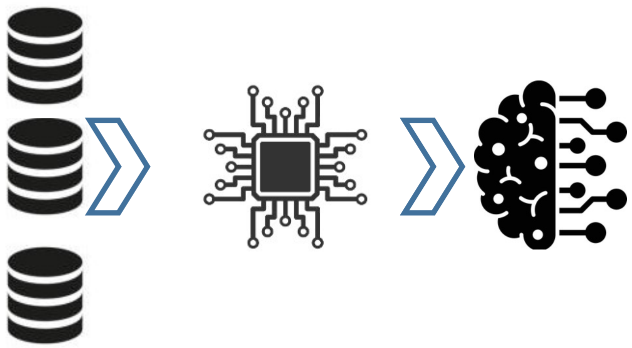
Penn Medicine

Penn Engineering
GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

PCASO Lab

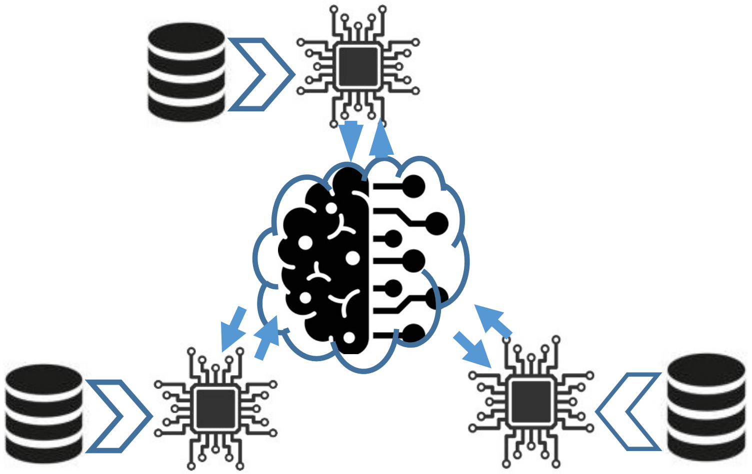
Motivation & Background

Centralized Learning



VS

Federated learning



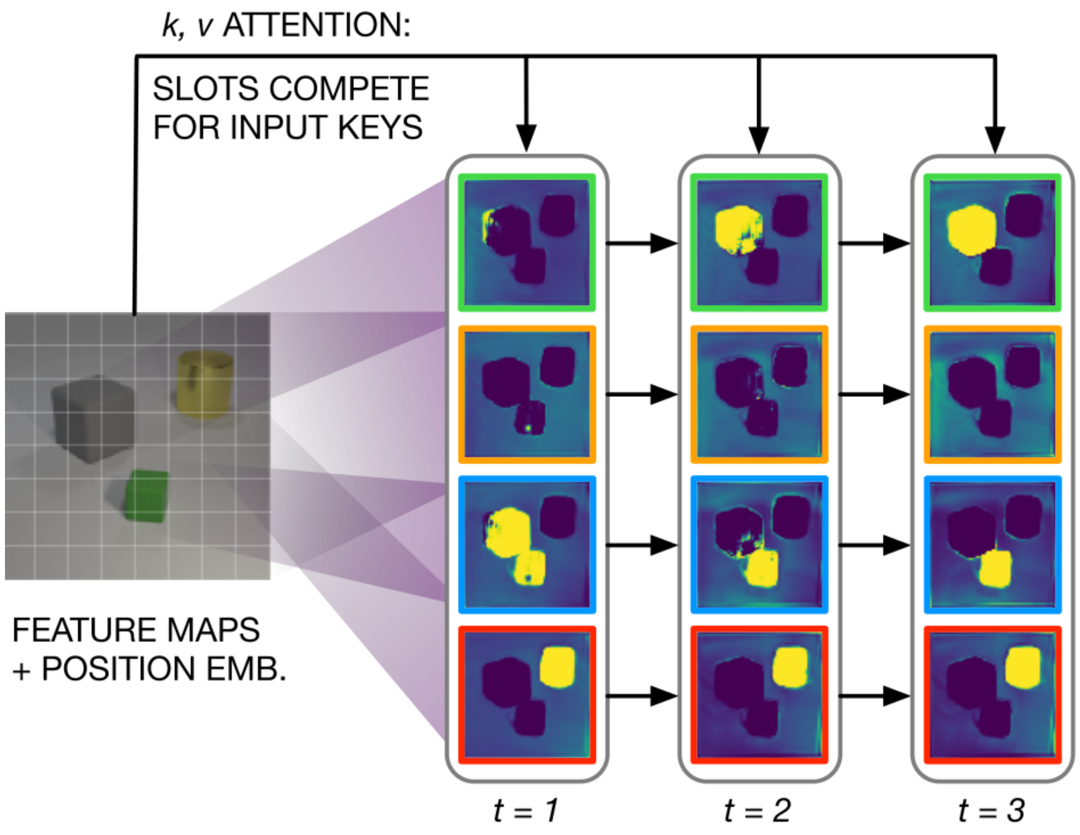
- ❑ Can we learn disentangled universal representations (foundation models) from distributed, non-IID datasets?

- ➢ Self-supervised object-centric learning offers the right inductive bias for learning semantically meaningful representations.

- ➢ Federated learning across curated datasets integrates local, object-level knowledge across clients.

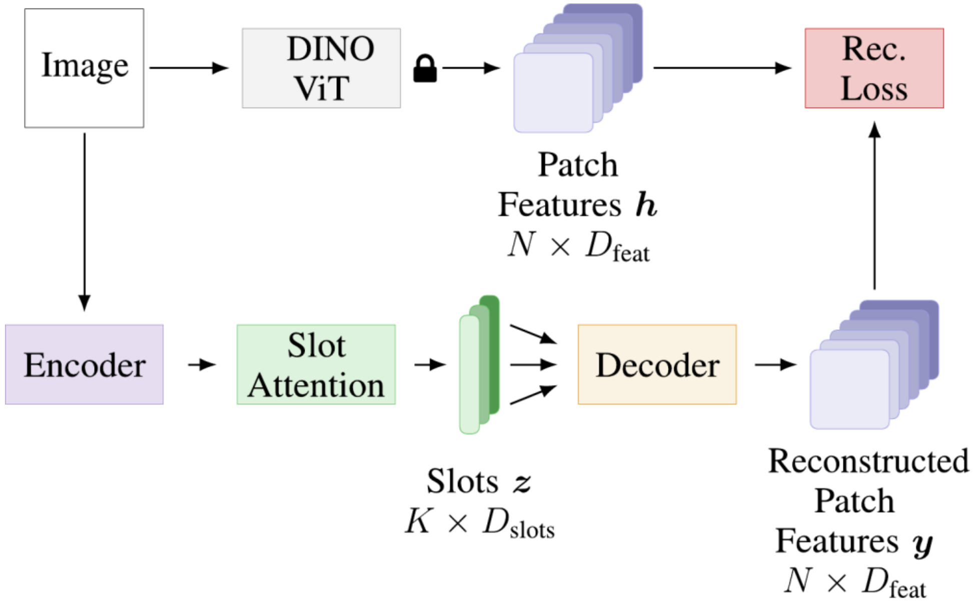
Motivation & Background

- Slot Attention model groups input pixels into slots via iterative attention
- Learning follows a reconstruction objective without labels

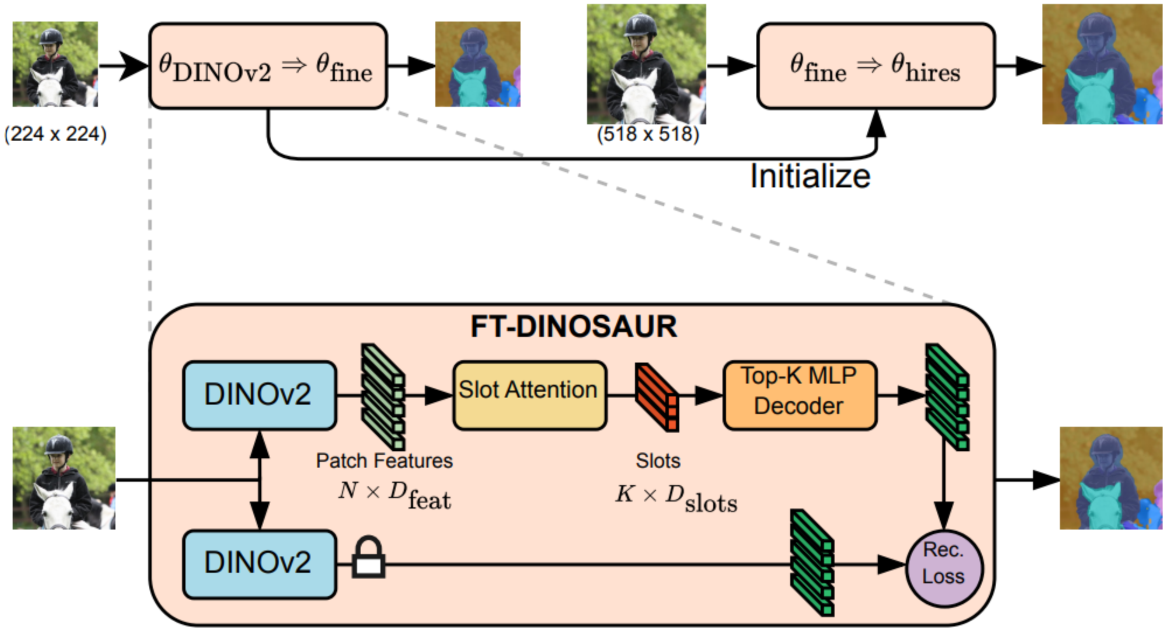


Vanilla slot attention module. Locatello, Francesco, et al., "Object-centric learning with slot attention." NeurIPS 2020.

Motivation & Background



DINOSAUR. Maximilian Seizer, et al., ICLR 2023.



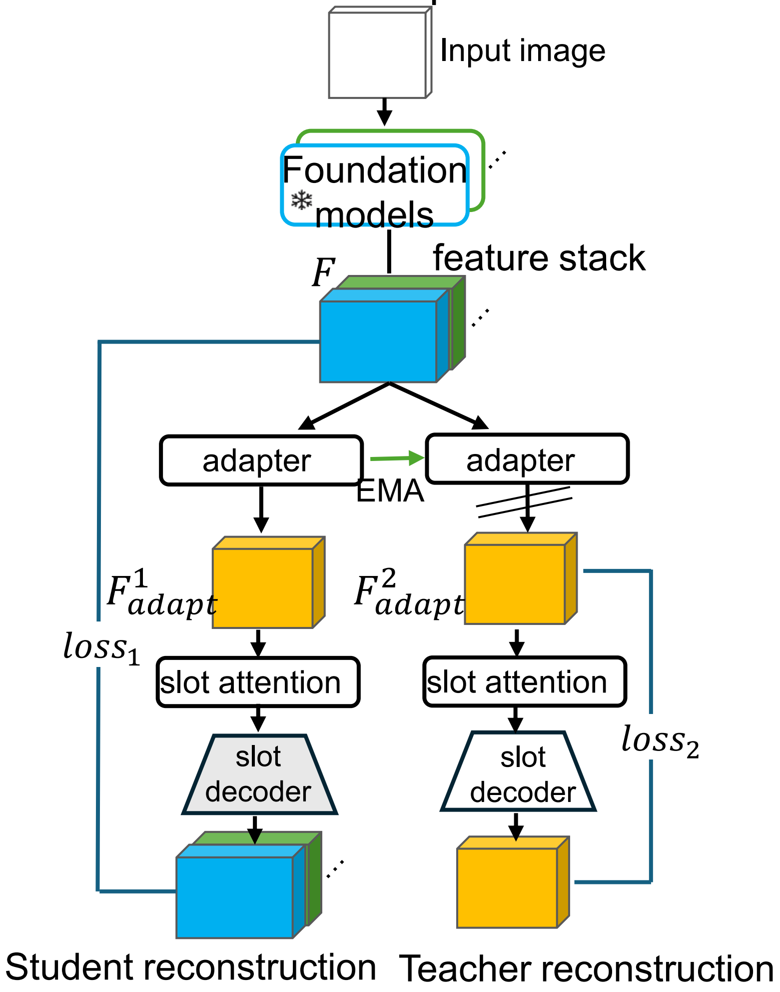
FT-DINOSAUR. Aniket Didolkar, et al, ICLR 2025.

- Slot Attention in real-world images with a foundation model
- Slot Attention can be used to fine-tune DINO-based foundation models

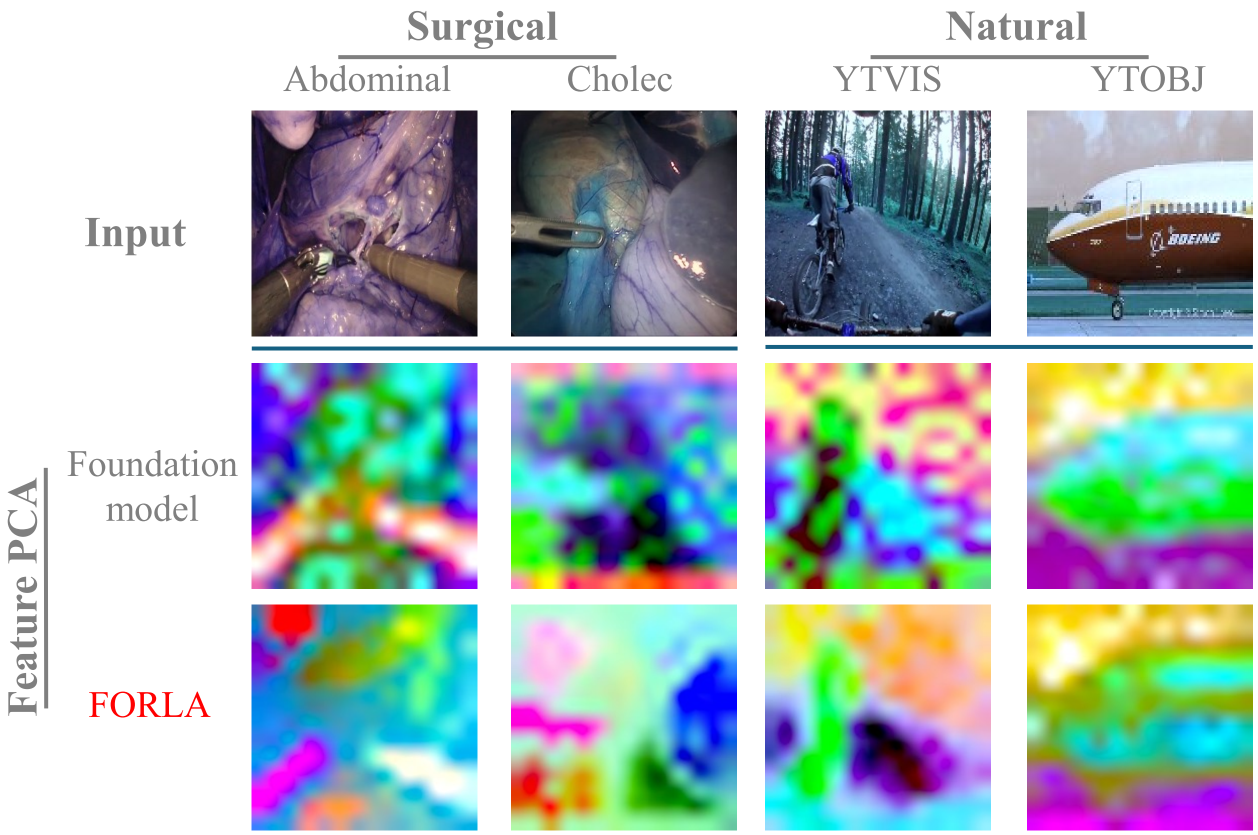
➢ Our key novelty: Bridging Federated Learning and Object-centric Representation Learning

Method

Architecture per client

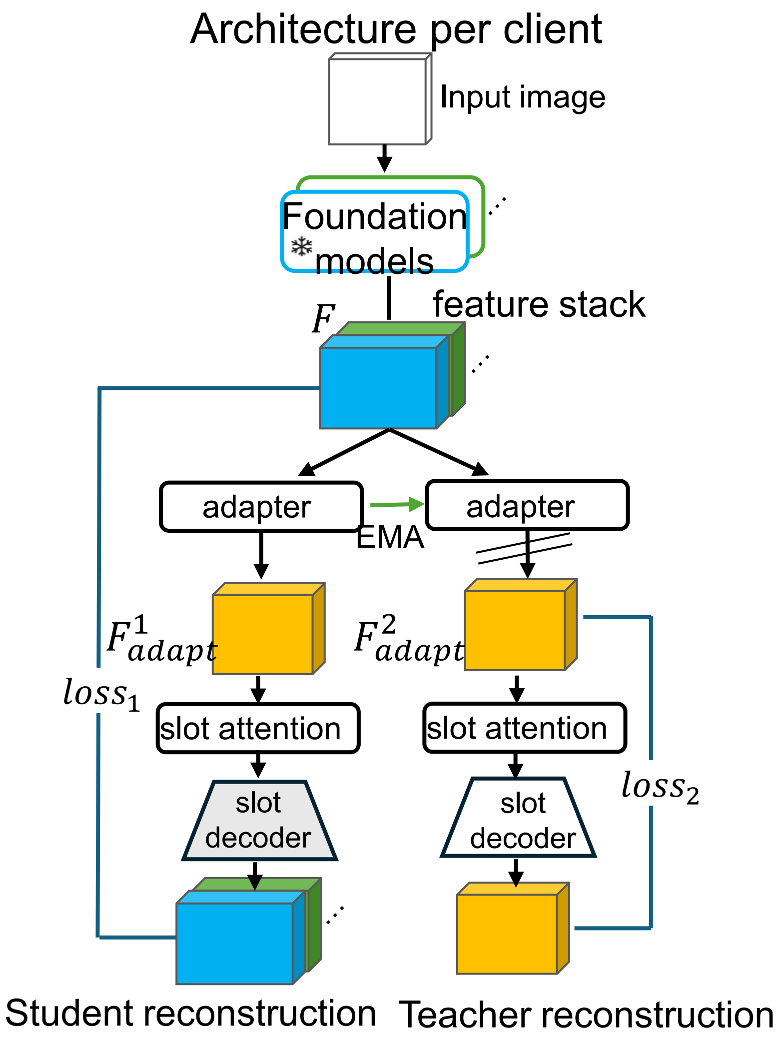


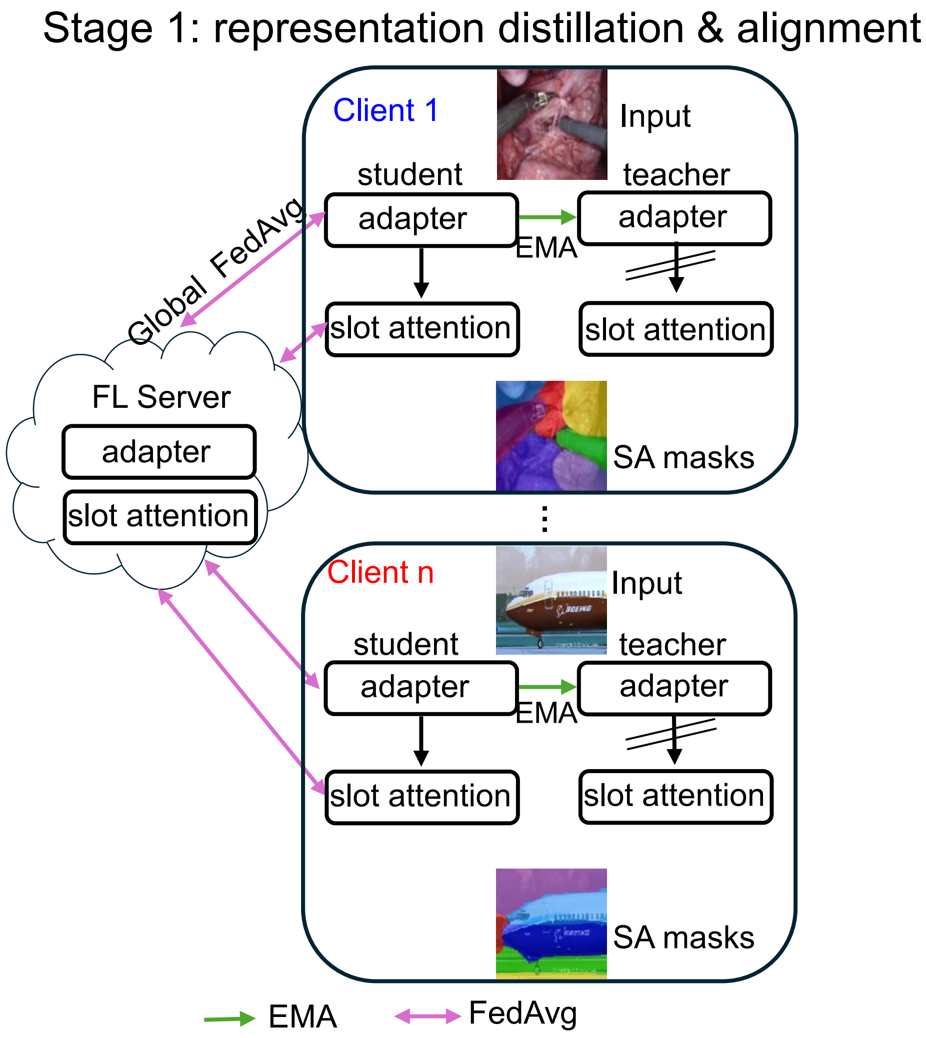
Feature adaptation results



- ➢ Observation: Slot attention can adapt the feature to better decompose images into different objects
- ▪ Inspired design: Another branch learns to reconstruct the adapted feature

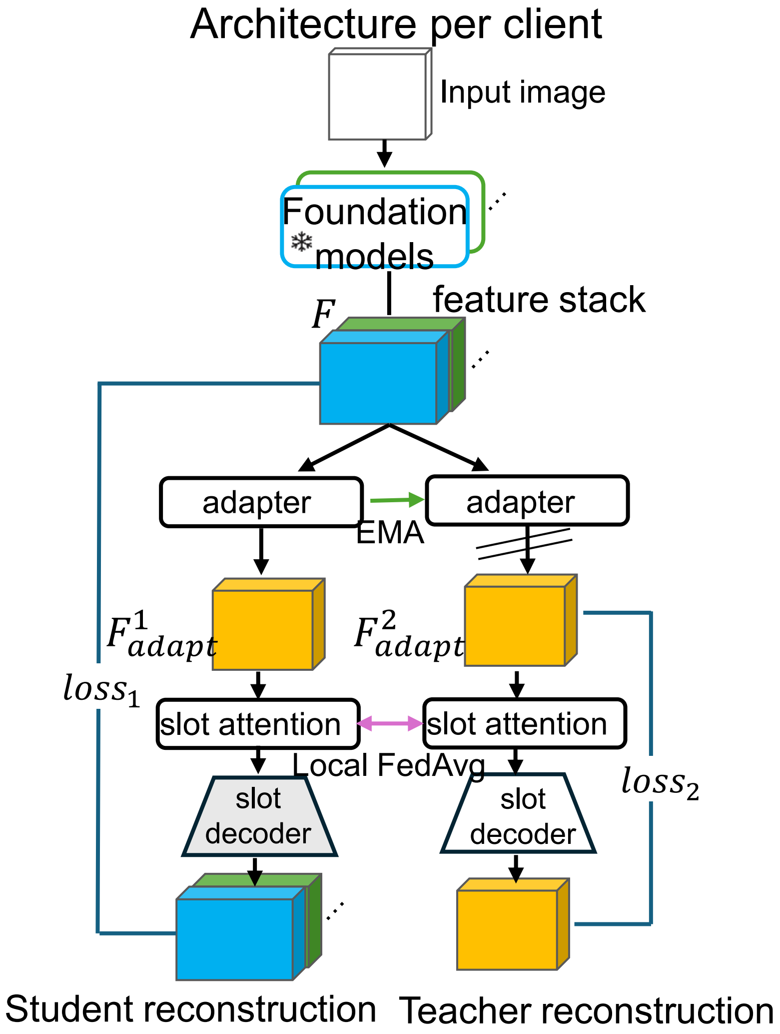
Method

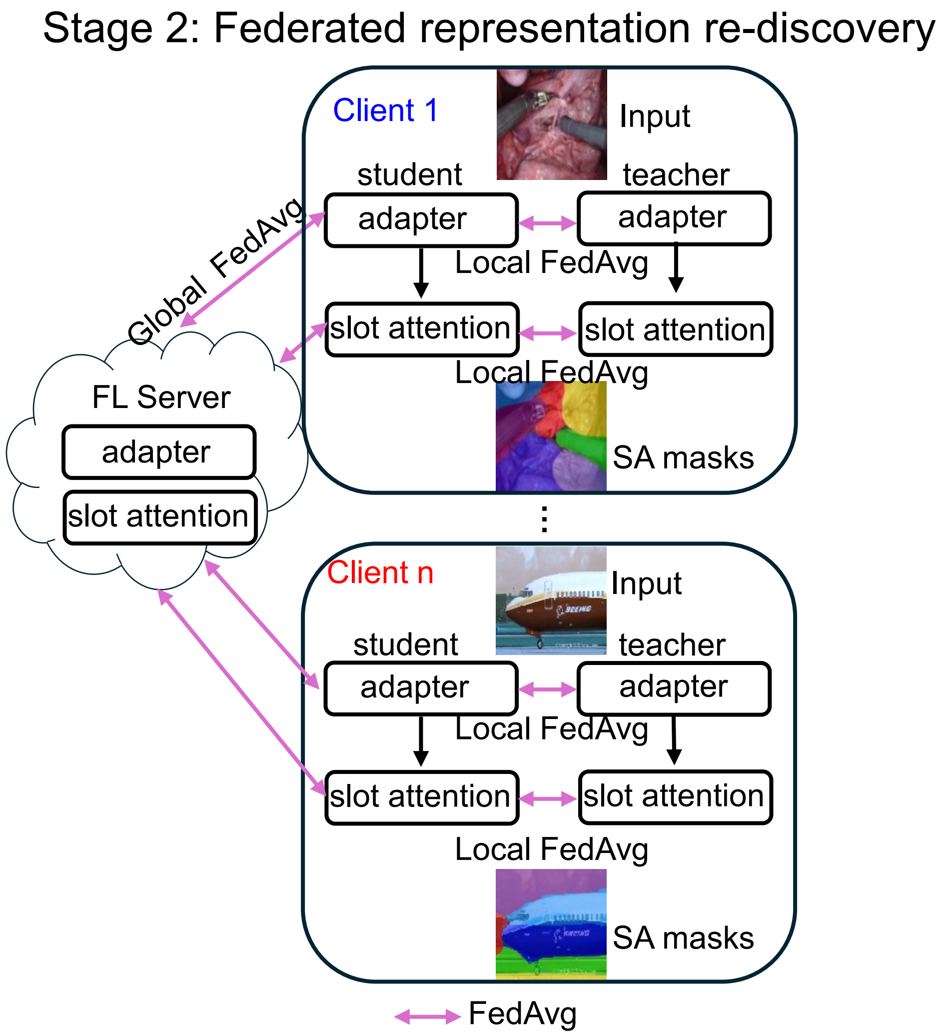




- A two-stage framework for federated object-centric representation learning

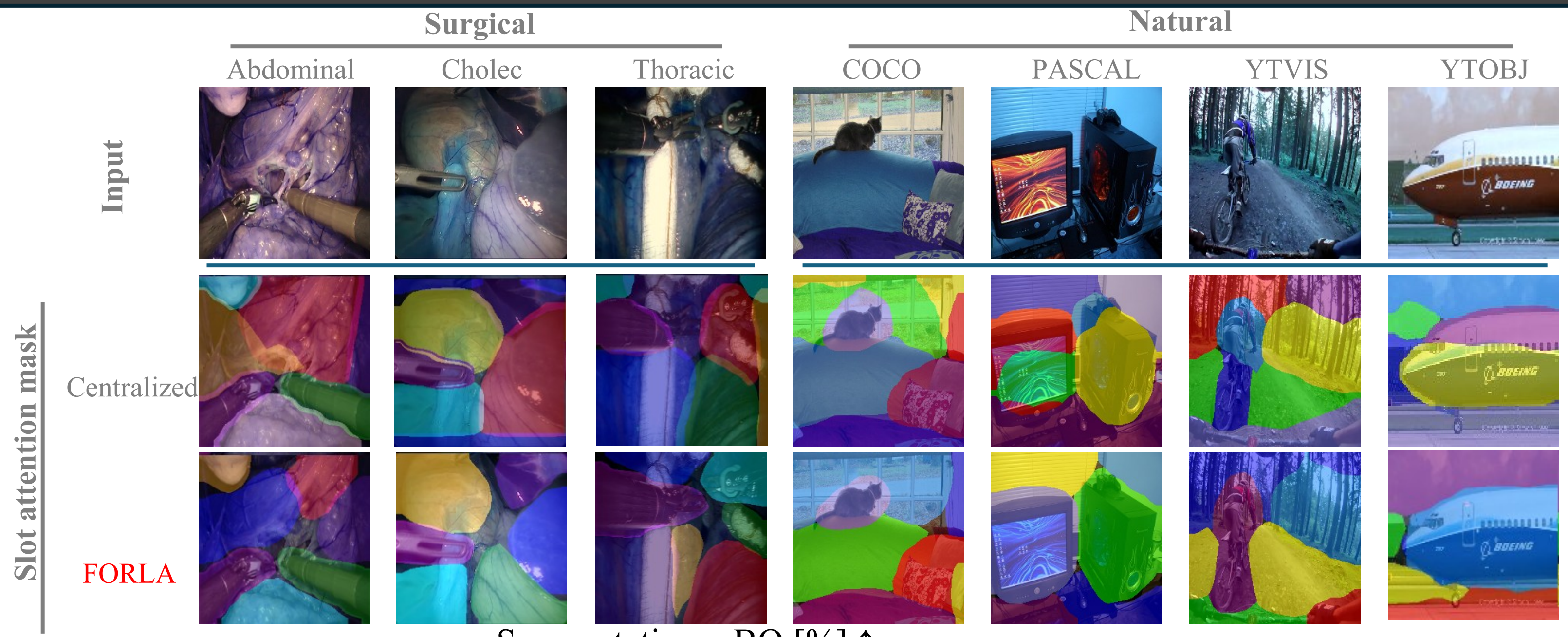
Method





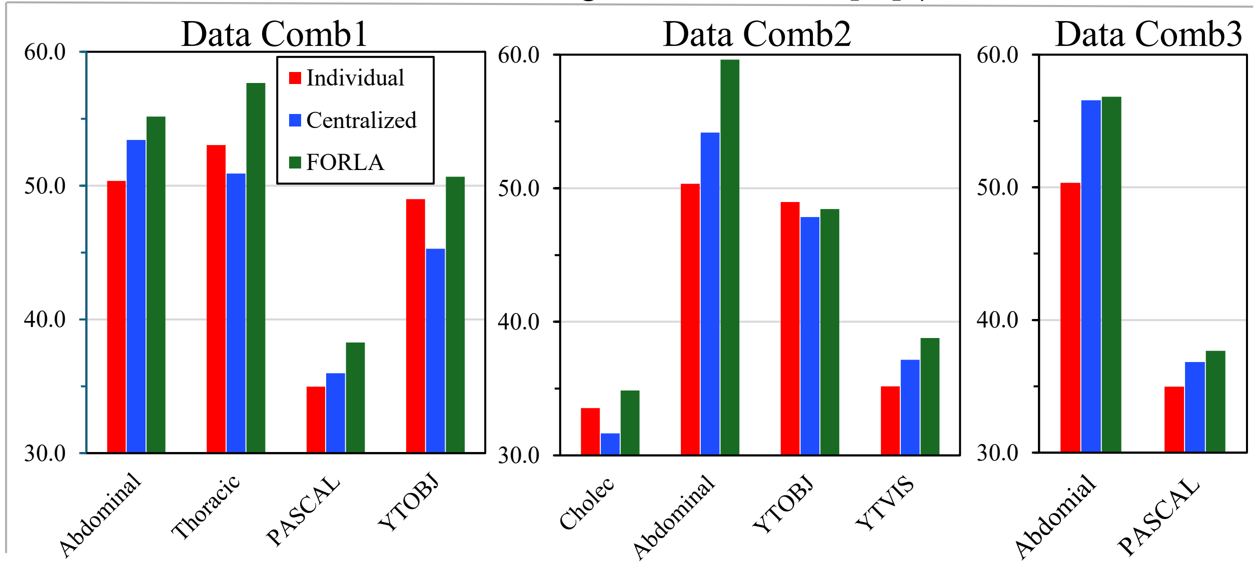
- A two-stage framework for federated object-centric representation learning

Results

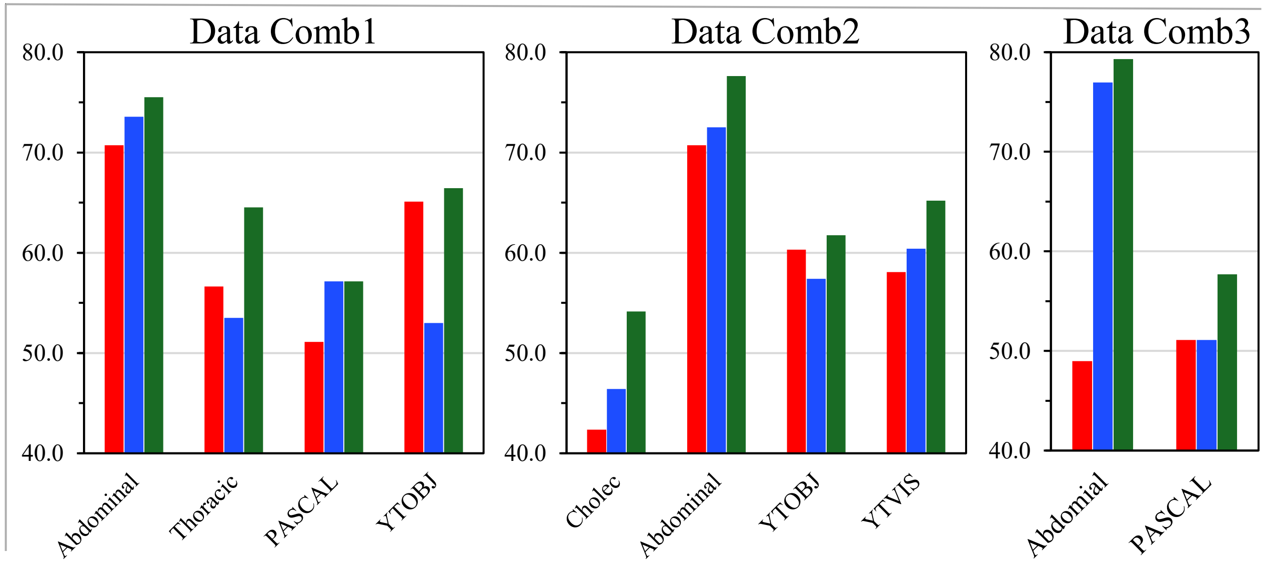


- FORLA outperforms centralized training
- Robust to different domain combinations

Segmentation mBO [%] ↑

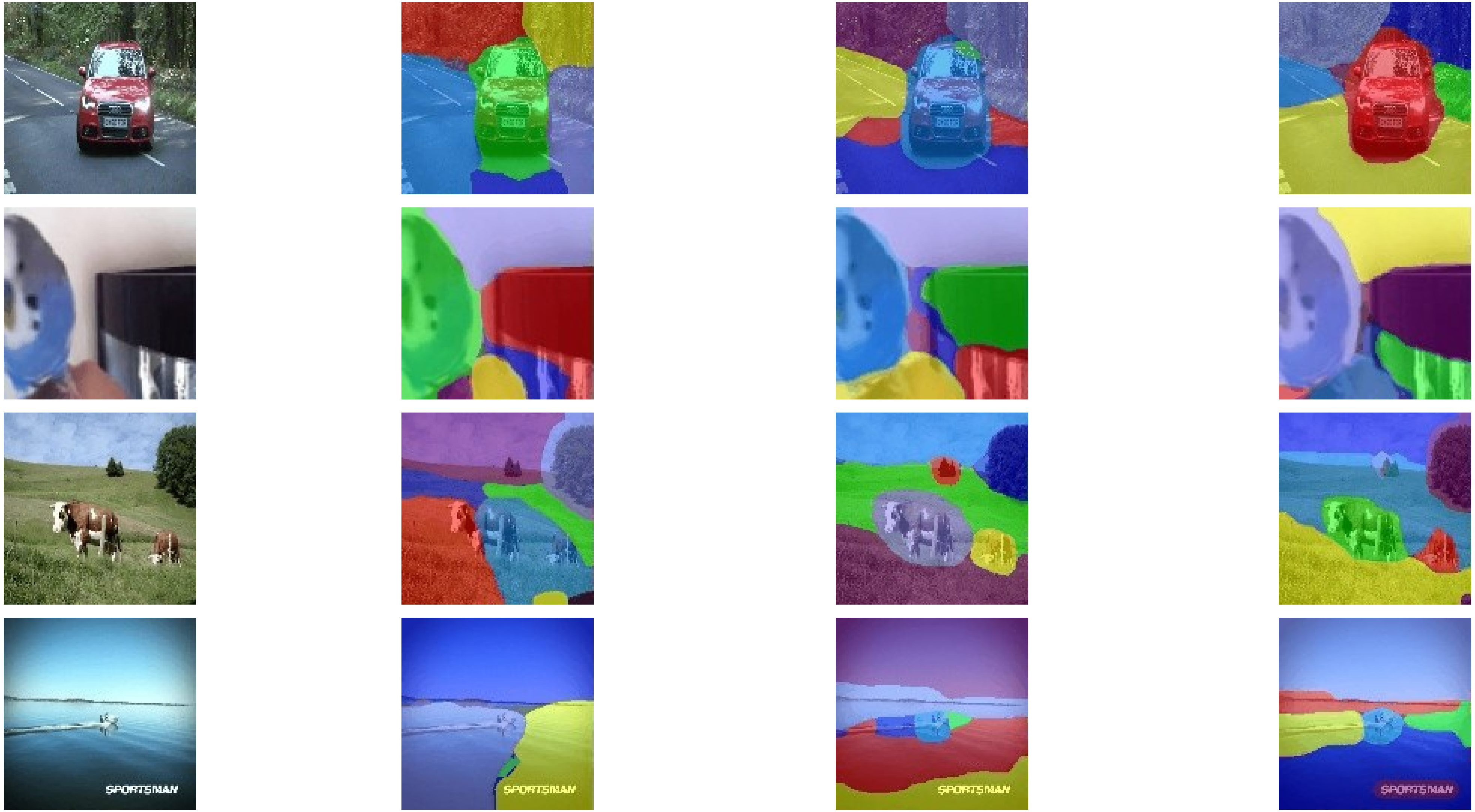


Localization CorLoc [%] ↑



Results

Input | SAM – SA adapted | DINO – SA adapted | FORLA



Conclusion & Future work

- Conclusion

- FORLA is the first to explore federated object-centric learning.
- Mixing all datasets for centralized unsupervised object discovery does not necessarily improve scalability.
- FL can also be a powerful algorithm for knowledge distillation (KD) through teacher–student collaboration.

- Future work

- Incorporate dynamic slot numbers
- Extend FORLA to video-based Slot Attention models

Check our project page: https://forla-research.github.io/