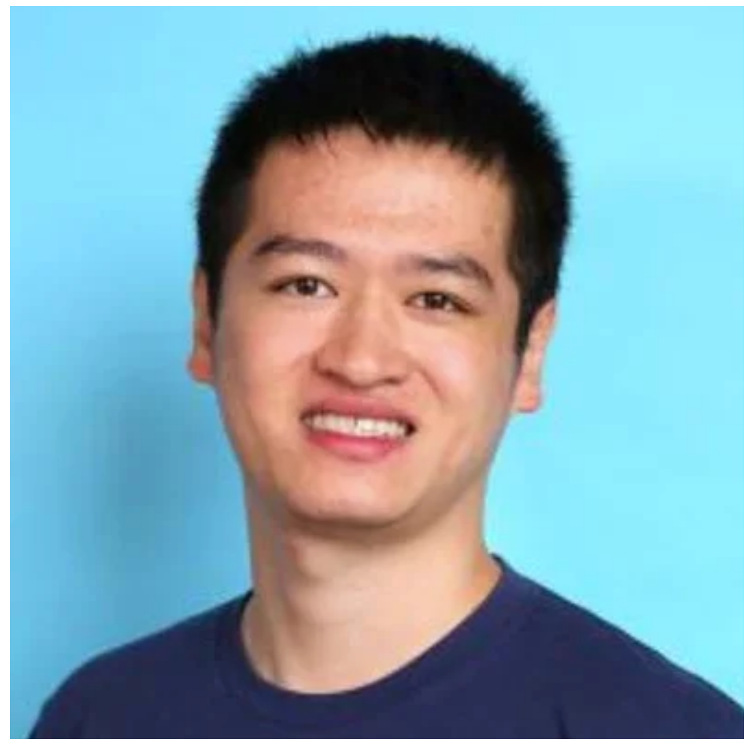
Berkeley University of California
University of Michigan
Caltech
UC Davis

ECCV

Pose-Aware Self-Supervised Learning with Viewpoint Trajectory Regularization

Jiayun Wang



On job market!
peterw@caltech.edu

Yubei Chen

Stella X. Yu


NM2255 HD

- Car heading towards the camera
- At danger!

Video credit: The BEST Car Donuts. Youtube


NM2255 HD

- Car heading towards the camera
- At danger!

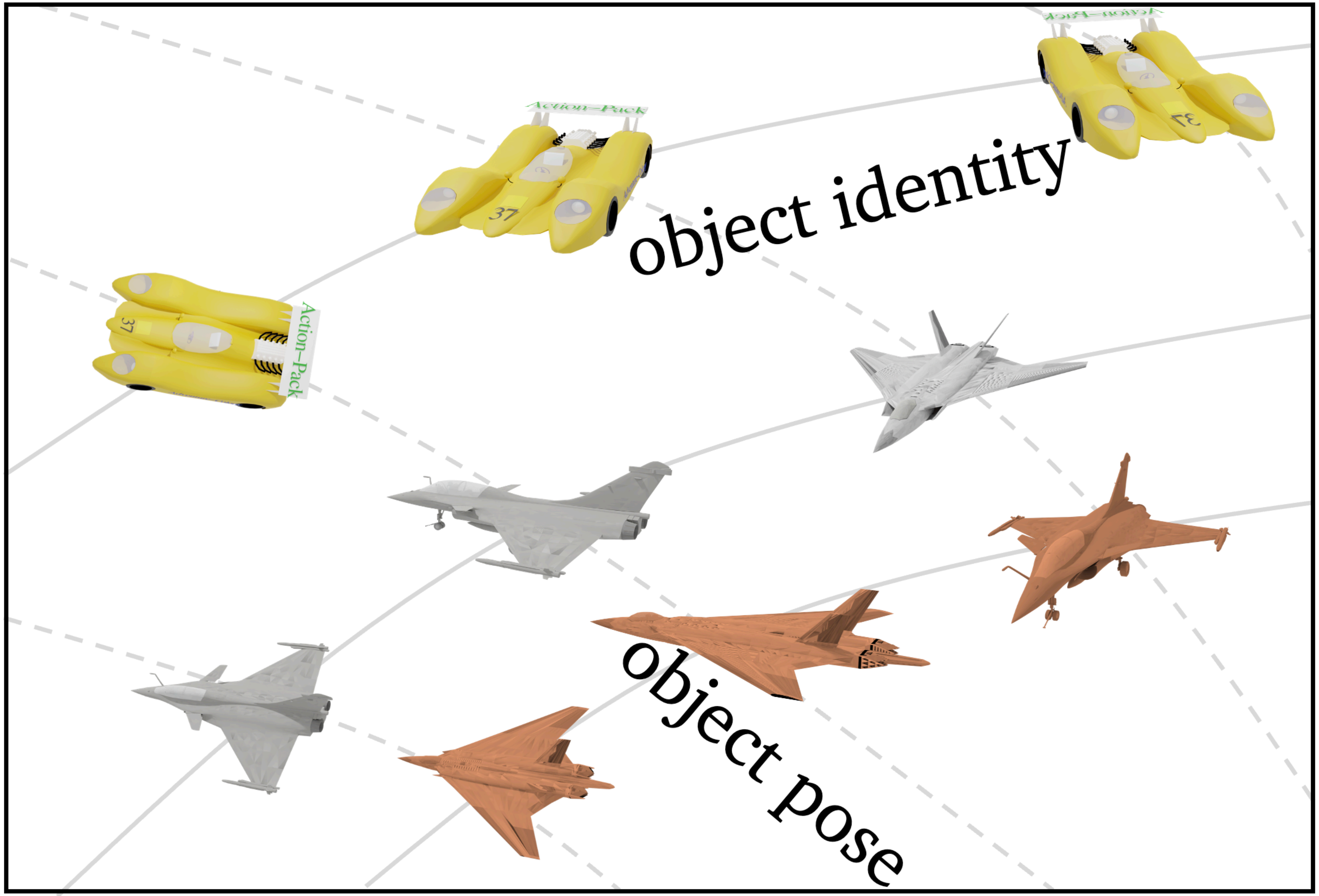
- Car heading away from the camera
- No danger

Recognition needs to understand both aspects:

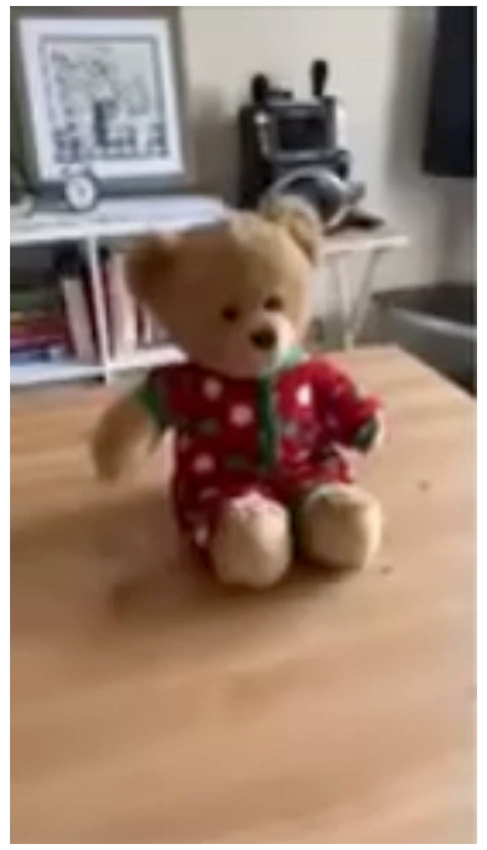
- *What* is the object
- *How* is it presented

Video credit: The BEST Car Donuts. Youtube

Can we learn **disentangled** semantic-pose representation?



Settings for disentangled representation learning?



Scenario: a robot moves around in the environment

A natural **data acquisition** scheme:

- No labels → Self-supervised learning (SSL)
- Adjacent images of the same object from a smooth viewpoint trajectory

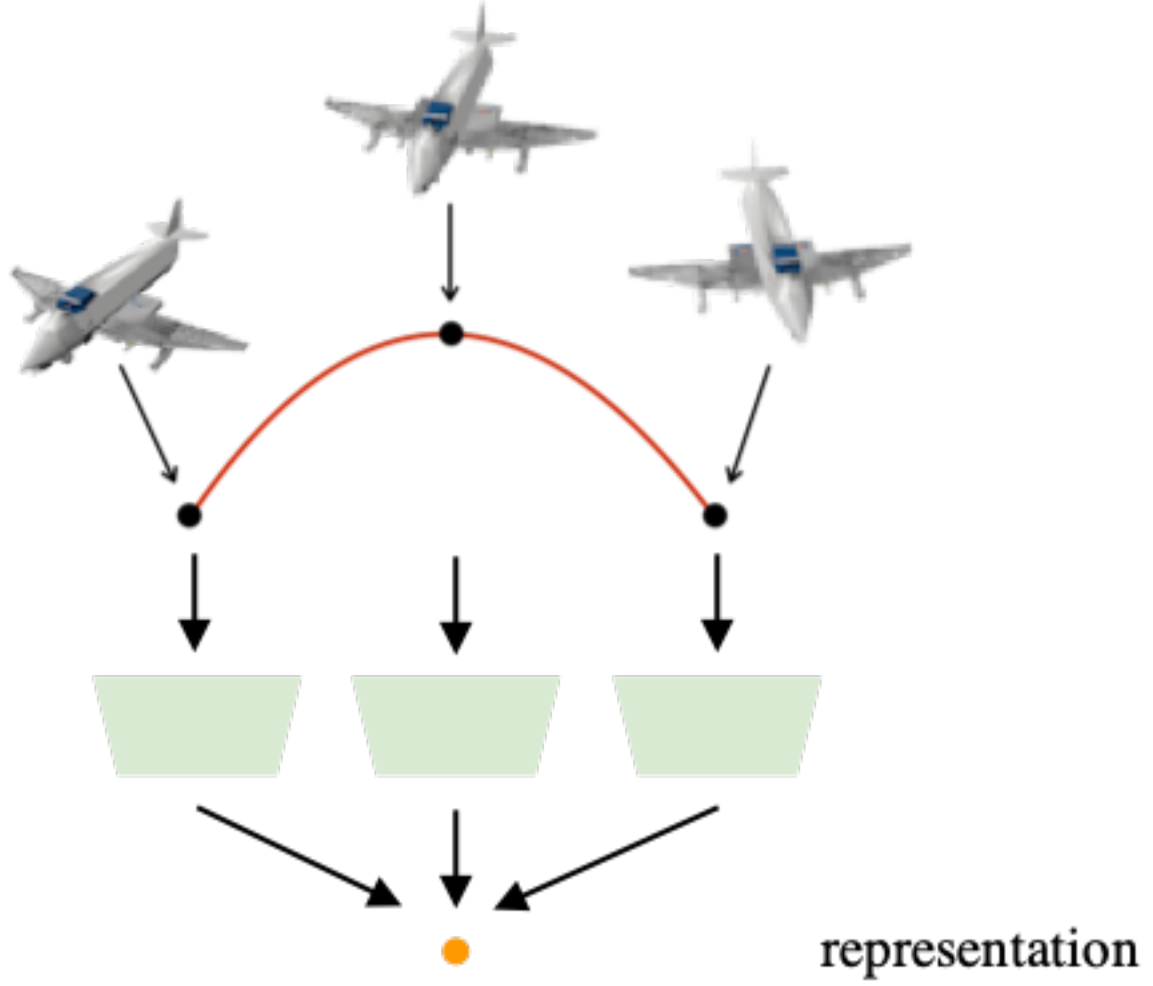
Common objects in 3d. ICCV 2021

Existing SSL vs Ours

Existing SSL

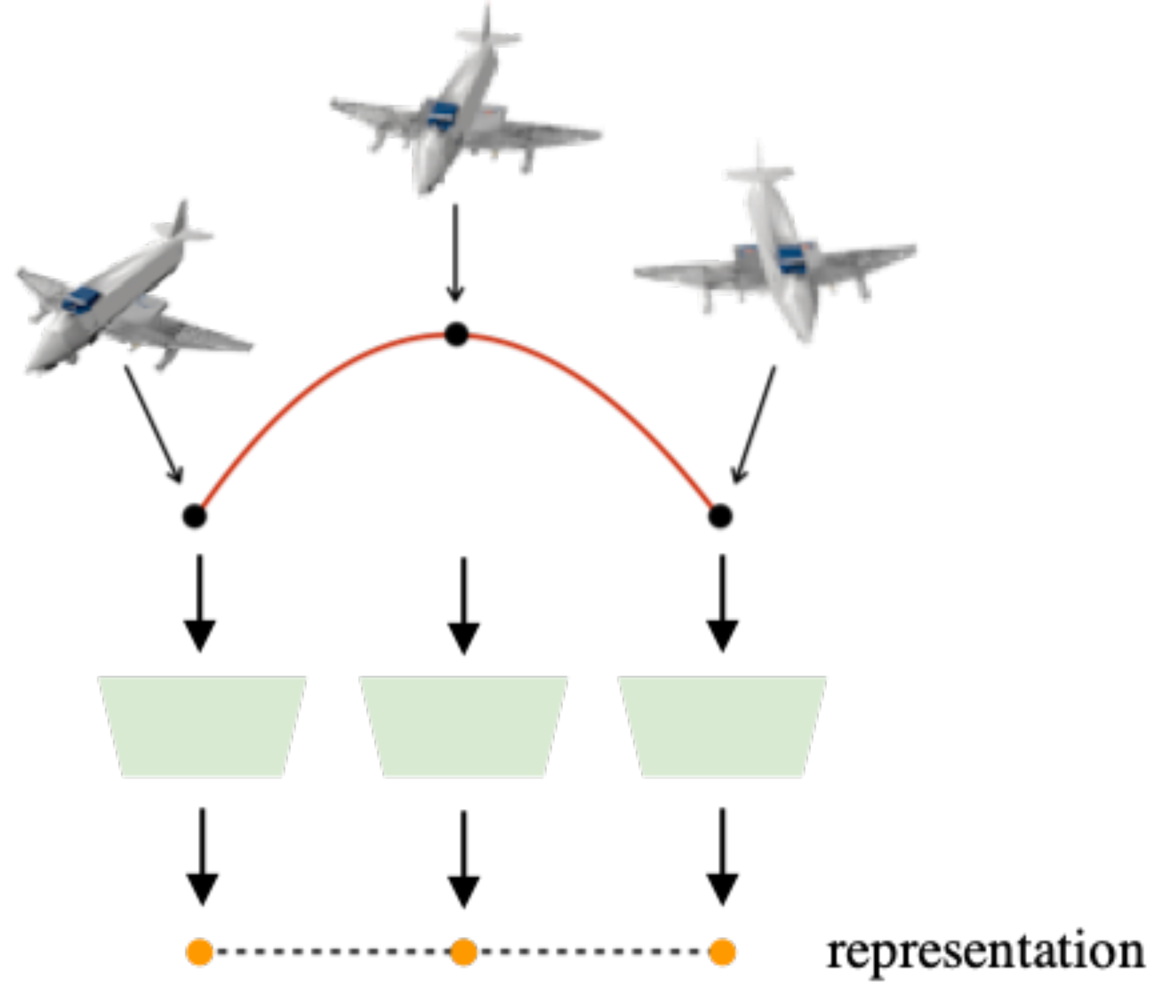
VICReg, SimCLR, SimSiam, MoCo,…

- Invariant representation
- Object identity only



Ours

- Equivariant representation
- Object identity + pose



Existing SSL vs Ours

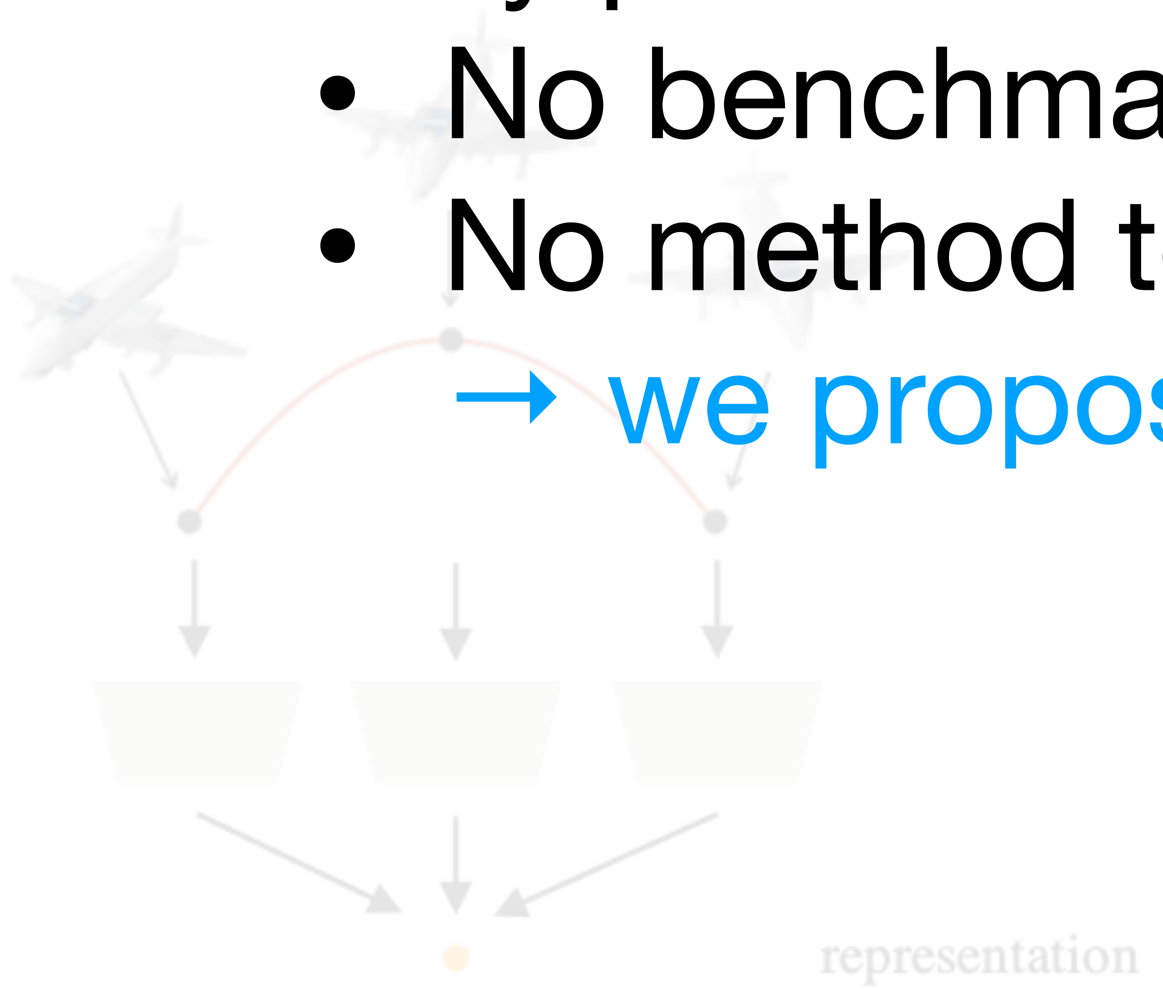
Existing SSL

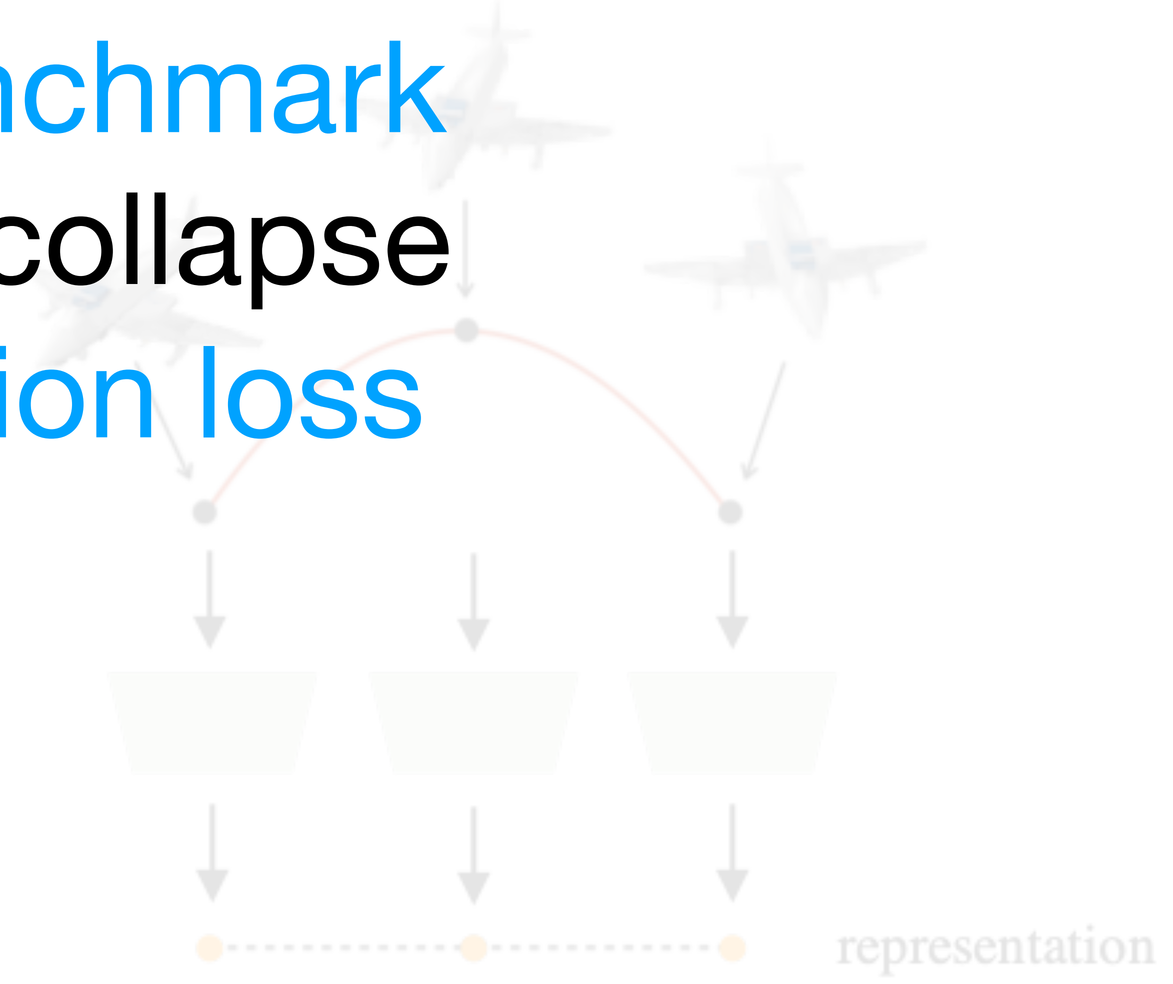
VICReg, SimCLR, SimSiam, MoCo,...

- Invariant representation
- Object identity only

Ours

- Equivariant representation
- Object identity + pose



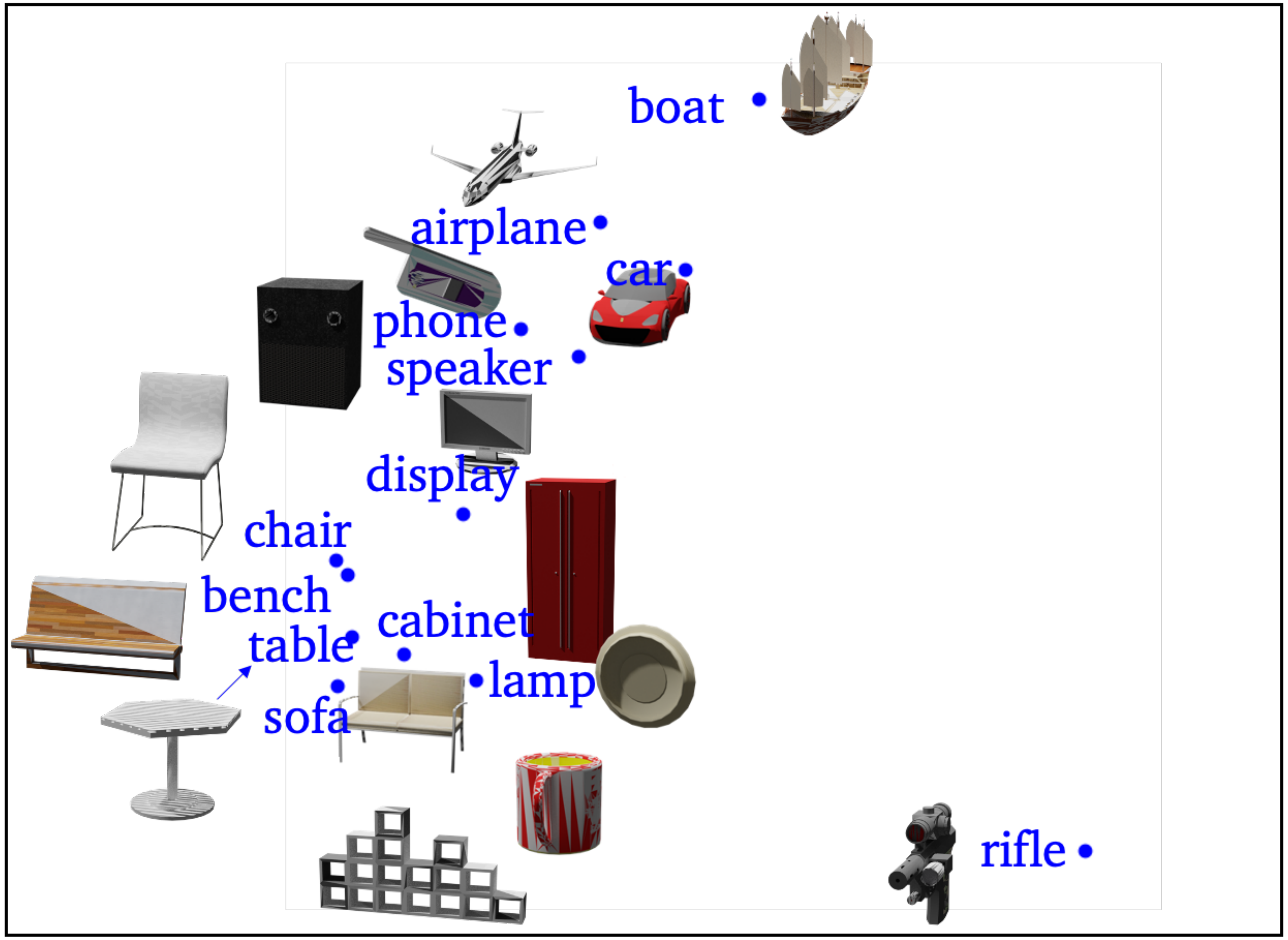


Why pose-aware SSL is hard?

- No benchmark → we propose a benchmark
- No method to avoid representation collapse
  → we propose trajectory regularization loss

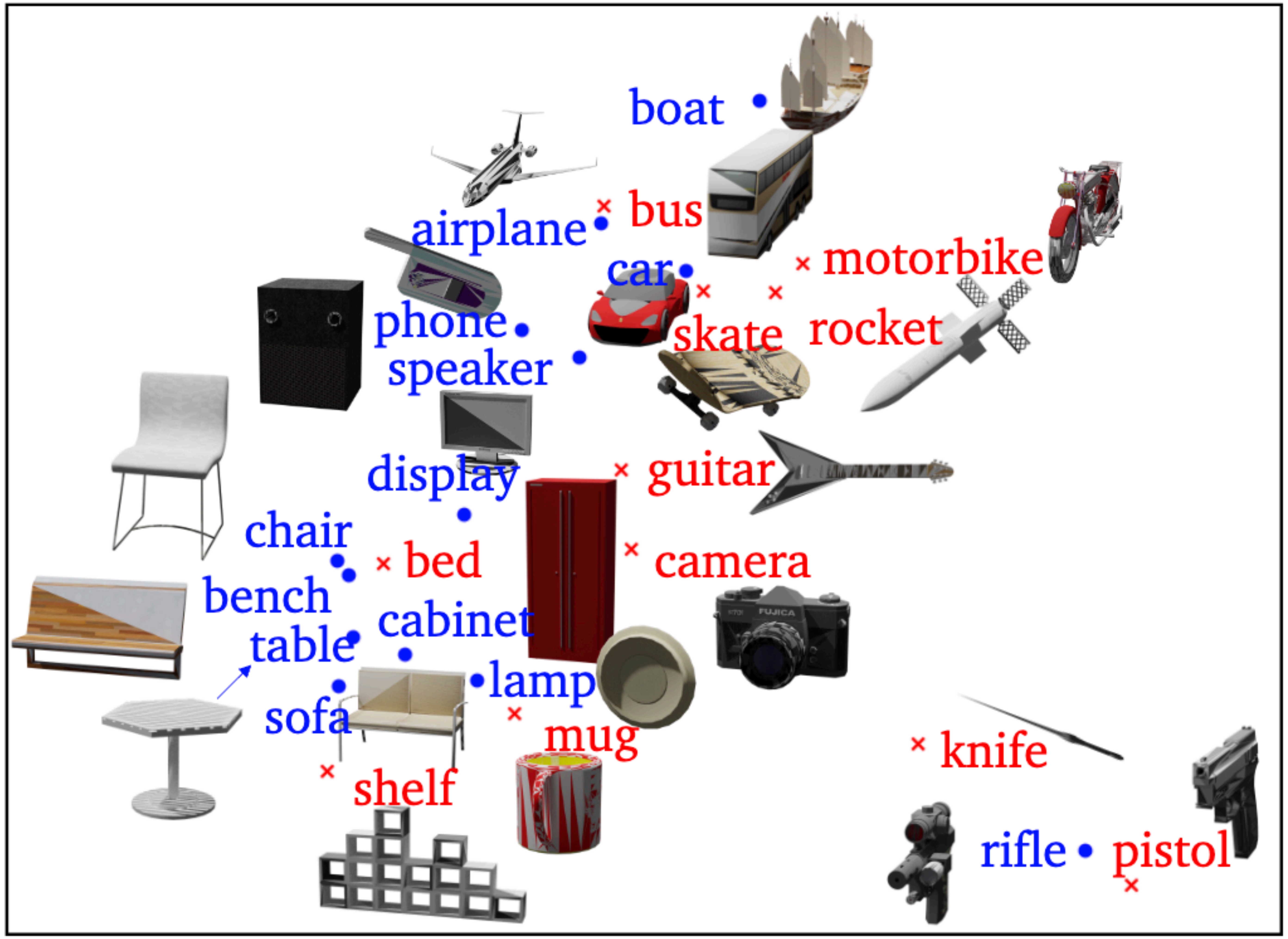
Benchmark: Data

13 in-domain

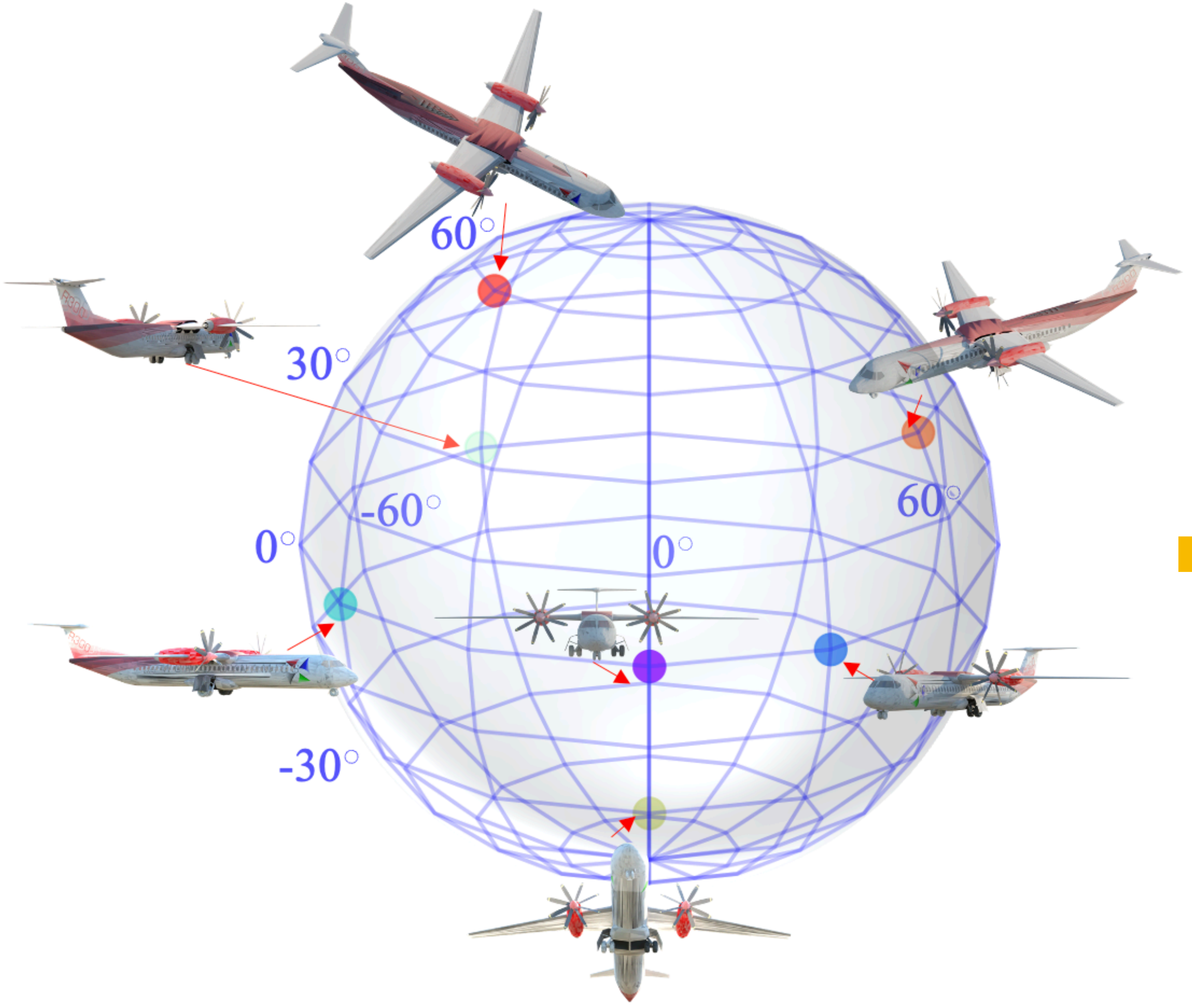


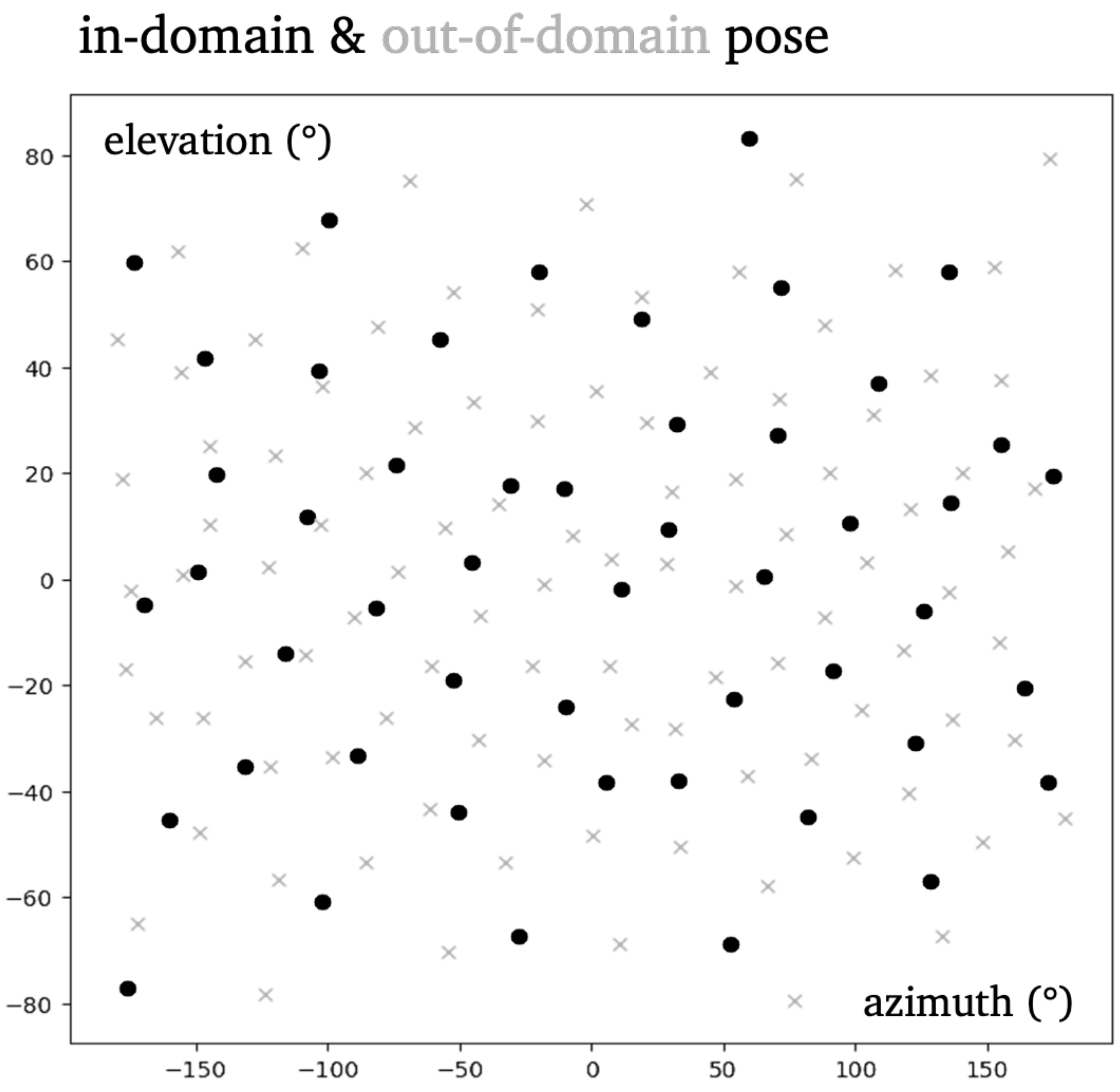
Benchmark: Data

13 in-domain & 20 out-of-domain semantic categories

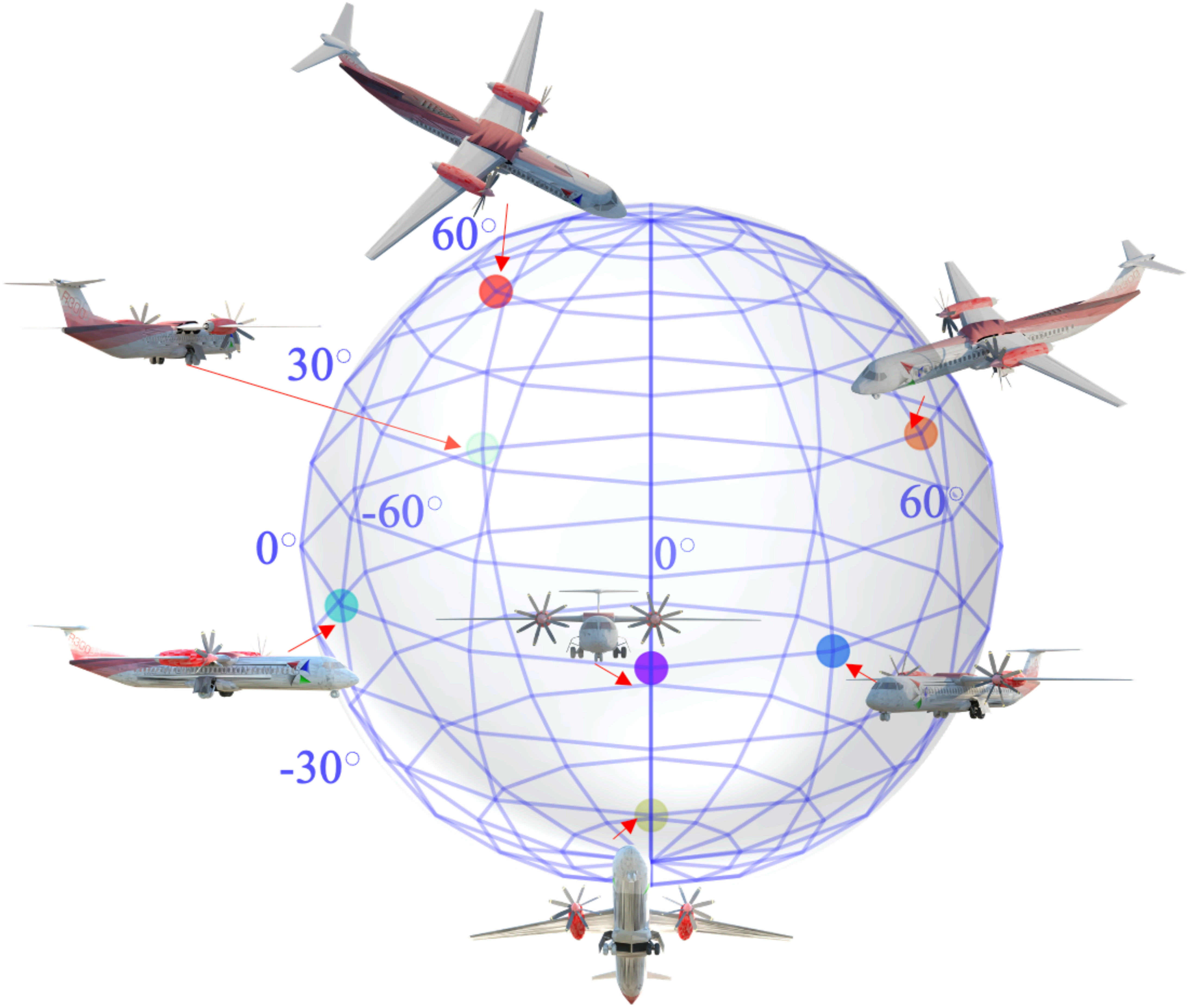


Benchmark: Data



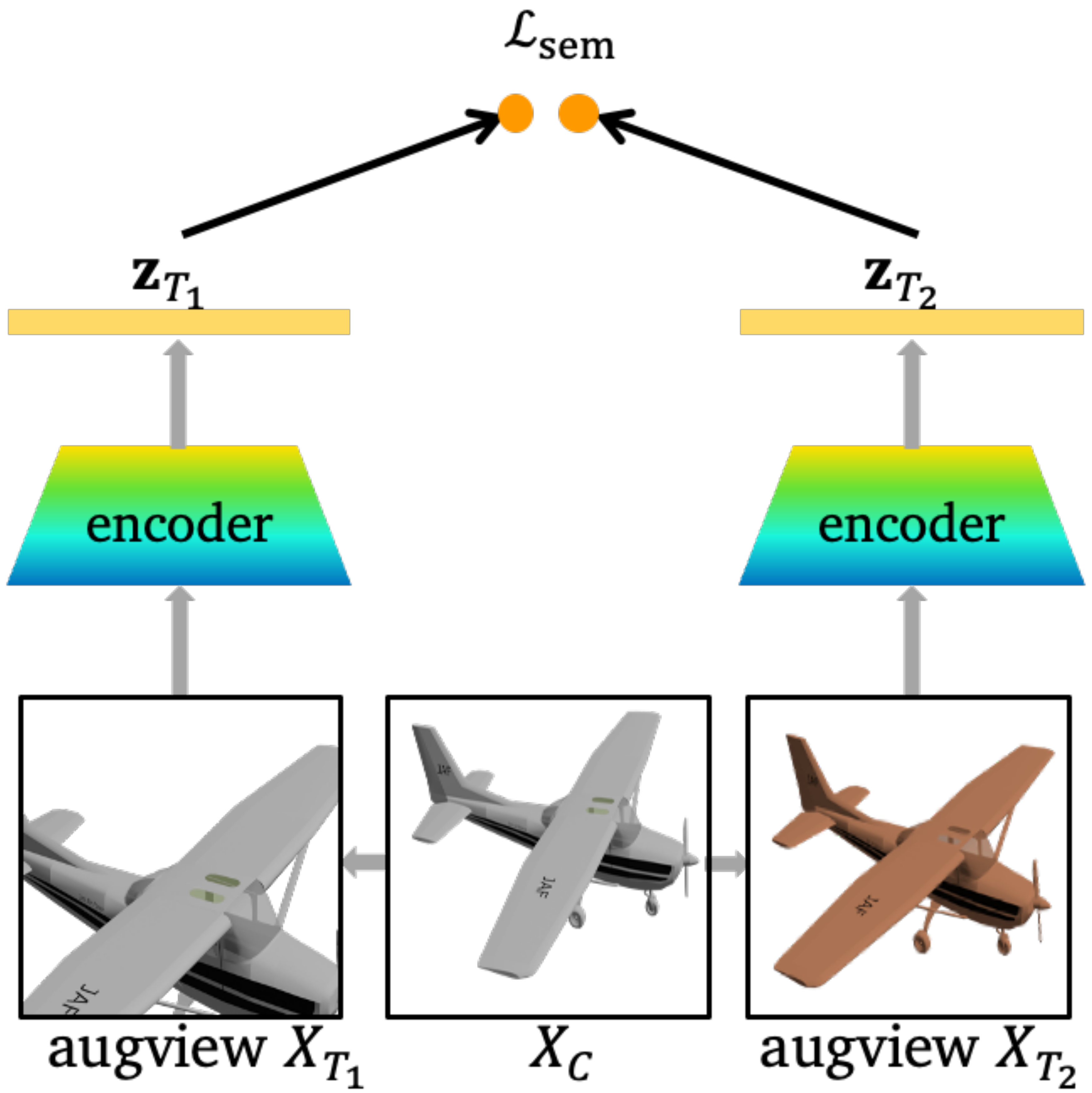


Benchmark: Tasks



- Semantic classification
- Absolute pose → global pose
- Relative pose → generalizability
  - Category-specific pose free
  - Generalize to open categories

Self-Supervised Representation Learning: Invariance



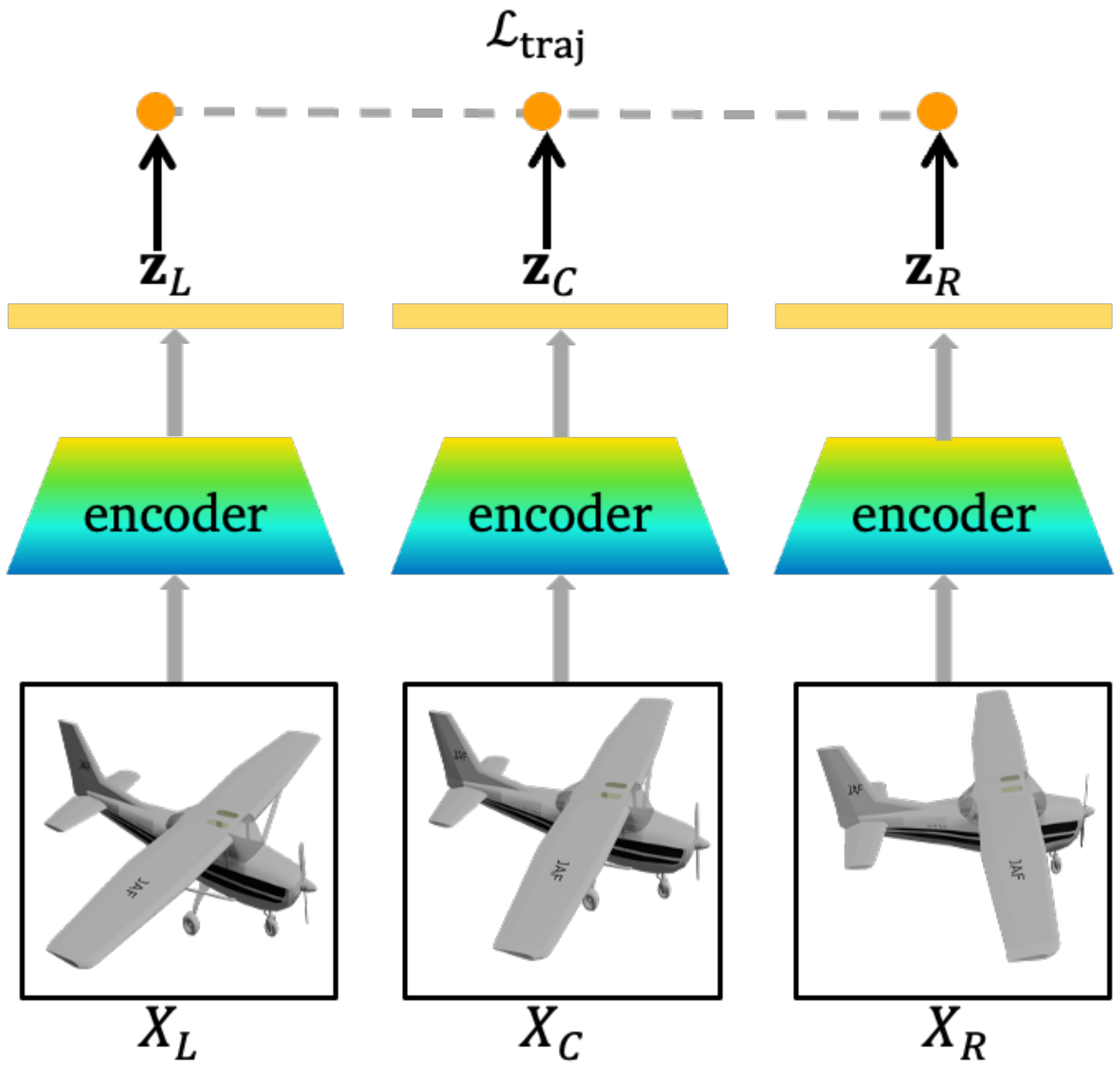
Example: VICReg

Augmentations:

- Random crops
- Color jittering
- Gaussian Blur

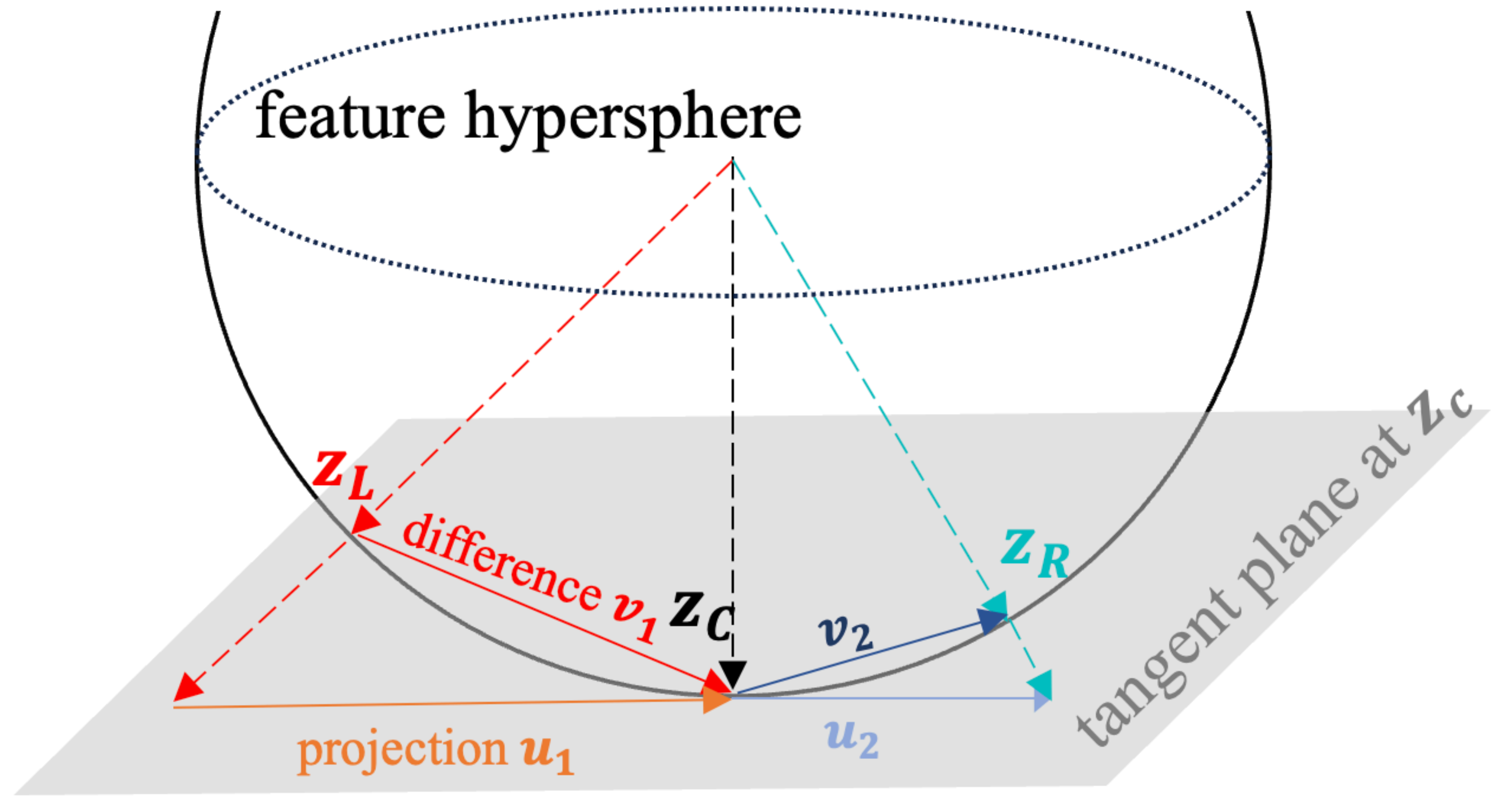
VICReg: Variance-invariance-covariance regularization for self-supervised learning. ICLR 2022

Self-Supervised Representation Learning: Trajectory Regularization

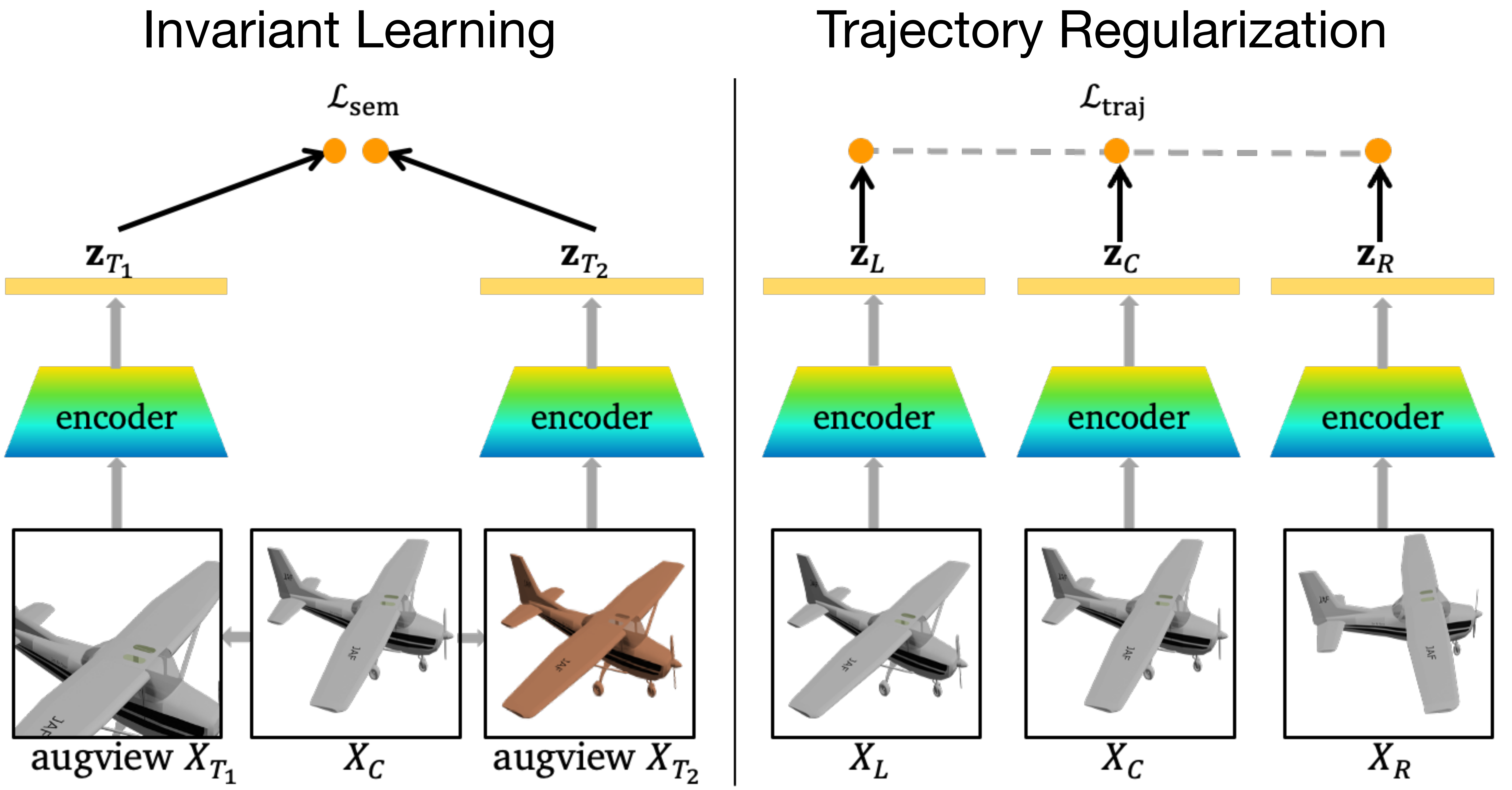


Line up 3 embeddings via trajectory regularization:

$$\mathscr{L}_{\text{traj}}(\mathbf{z}_L, \mathbf{z}_C, \mathbf{z}_R) = -\frac{\mathbf{u}_1 \cdot \mathbf{u}_2}{\|\mathbf{u}_1\| \|\mathbf{u}_2\|}$$

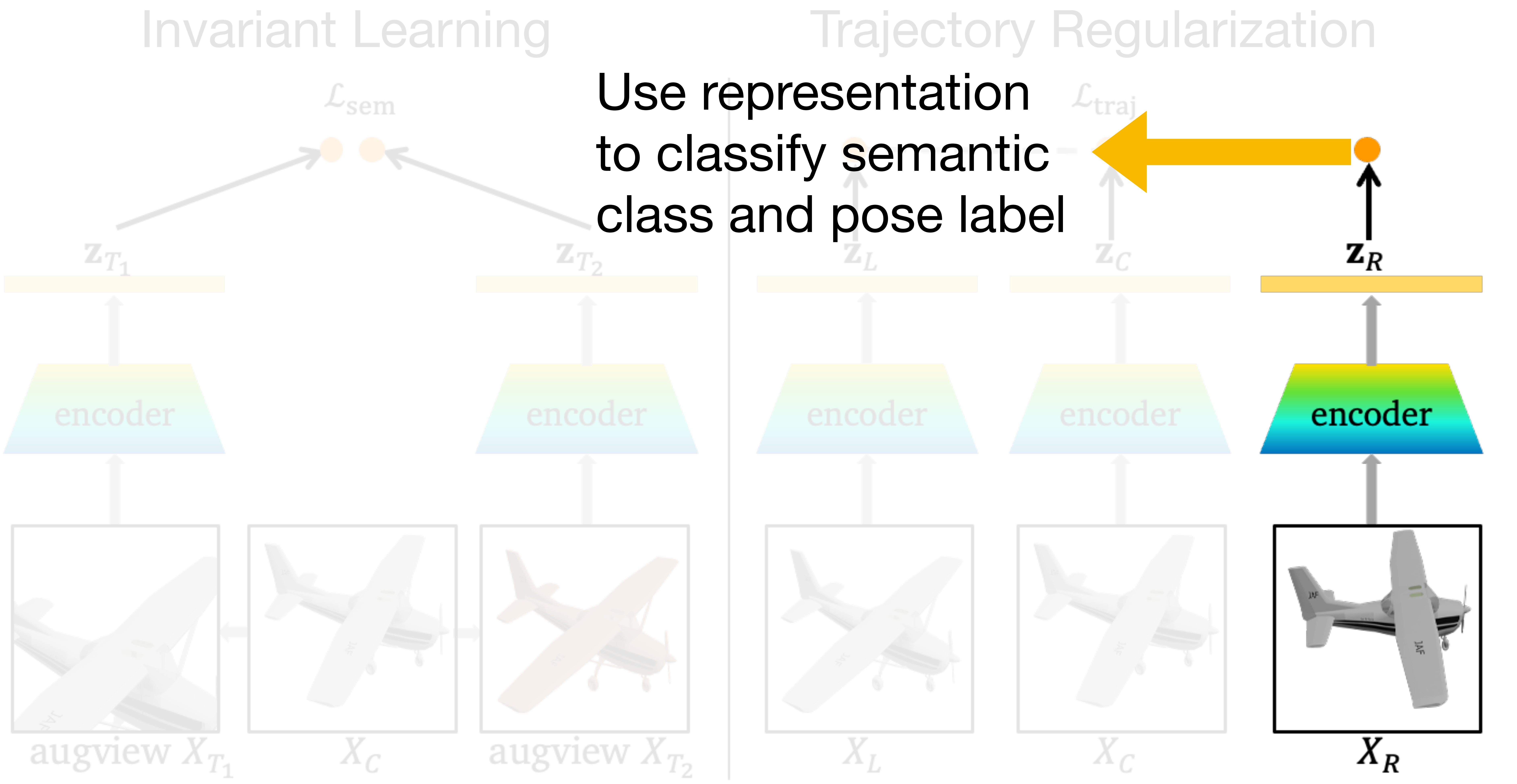


Self-Supervised Representation Learning



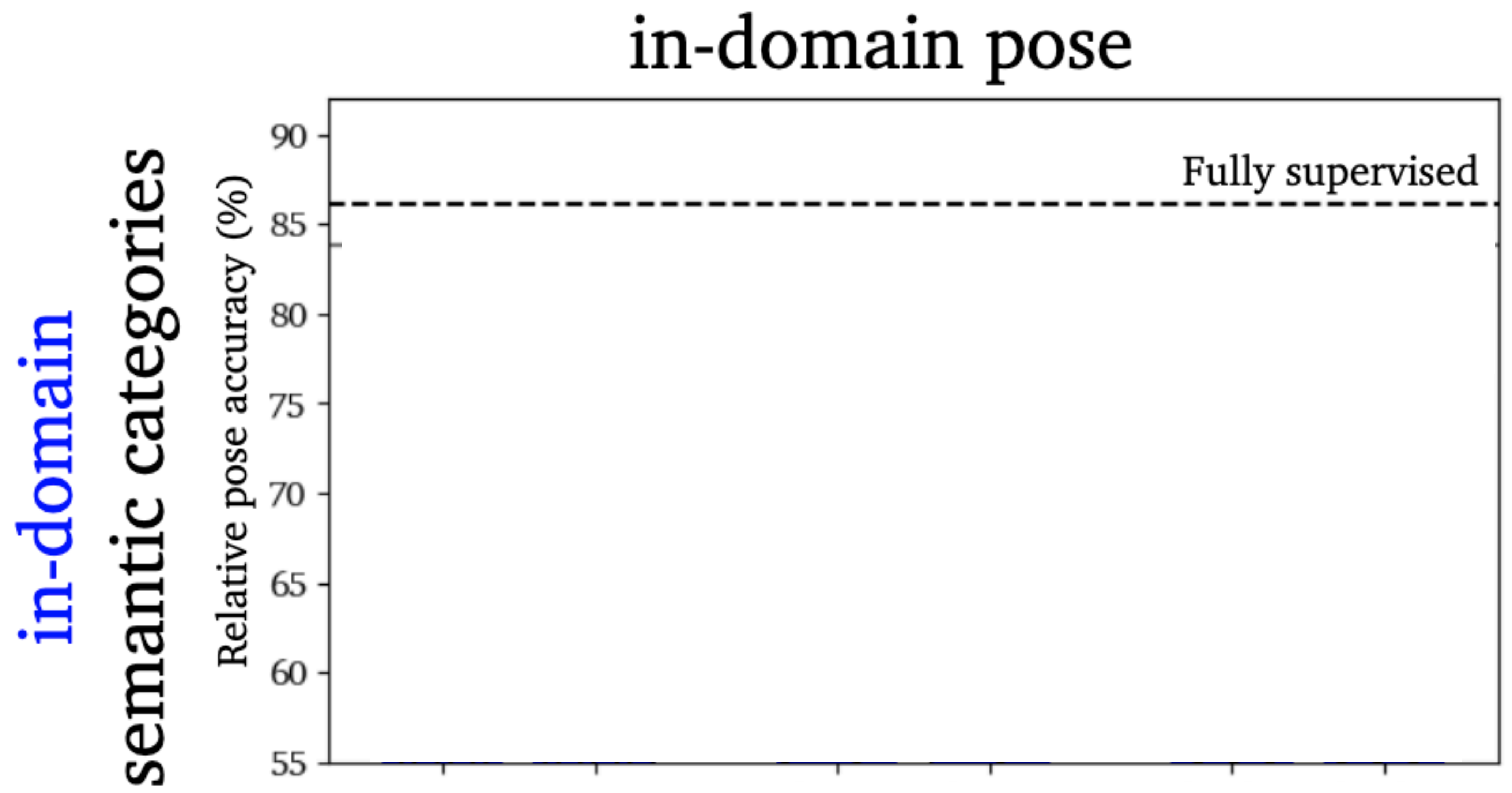
Final loss is a combination: $\mathscr{L} = \mathscr{L}_{\text{sem}}(\mathbf{z}_{T_1}, \mathbf{z}_{T_2}) + \lambda \mathscr{L}_{\text{traj}}(\mathbf{z}_L, \mathbf{z}_C, \mathbf{z}_R)$

After Representation Learning: Probing

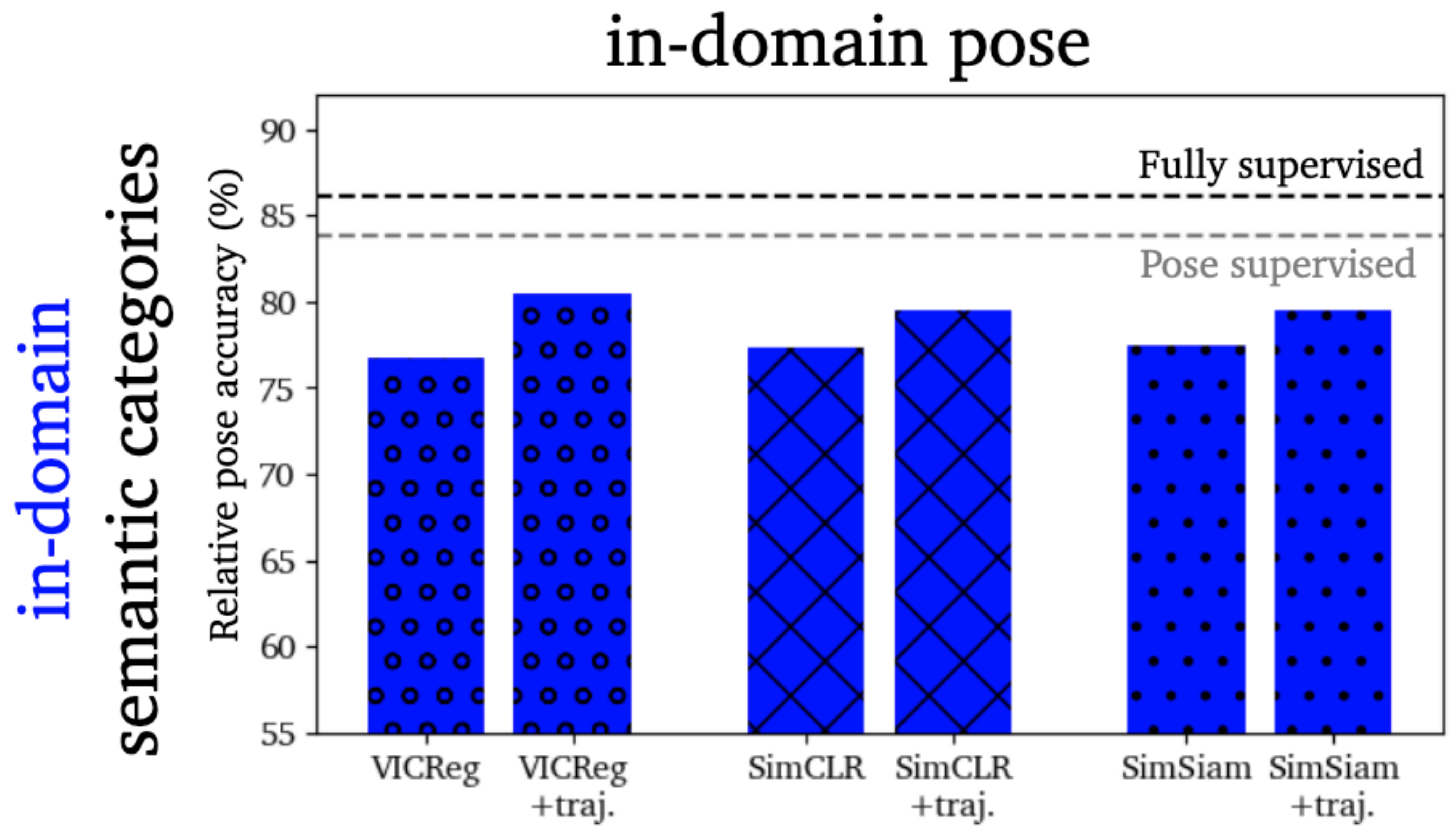


Final loss is a combination: $\mathscr{L} = \mathscr{L}_{\text{sem}}(\mathbf{z}_{T_1}, \mathbf{z}_{T_2}) + \lambda \mathscr{L}_{\text{traj}}(\mathbf{z}_L, \mathbf{z}_C, \mathbf{z}_R)$

Improved In-Domain and Out-of-Domain Pose Accuracy

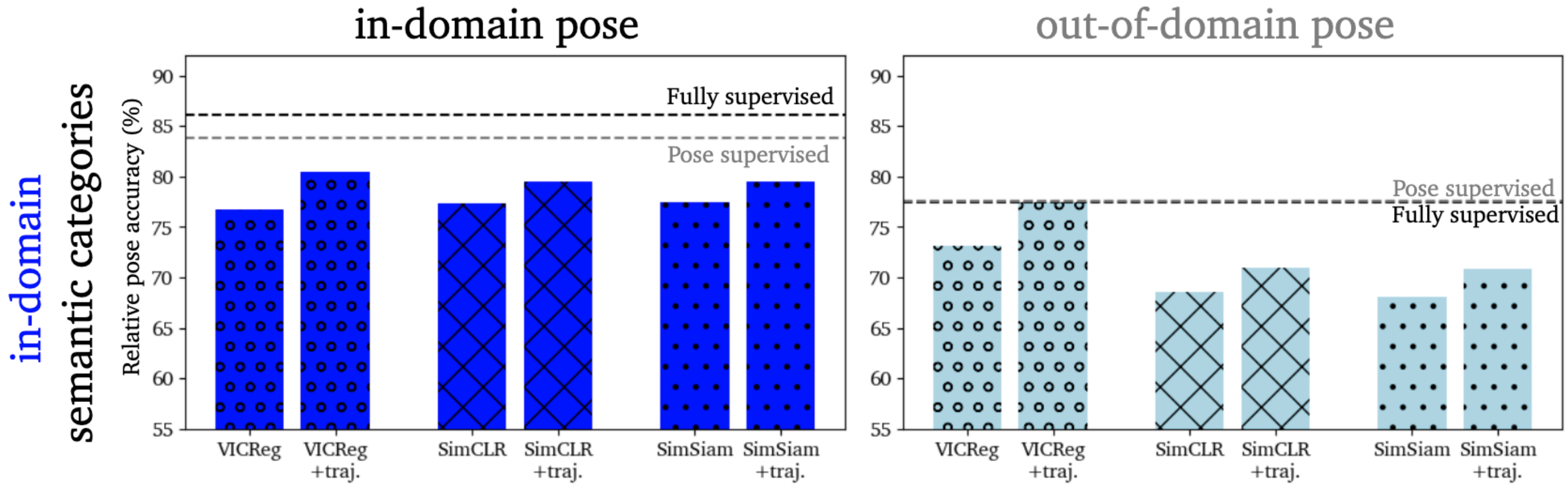


Improved In-Domain and Out-of-Domain Pose Accuracy



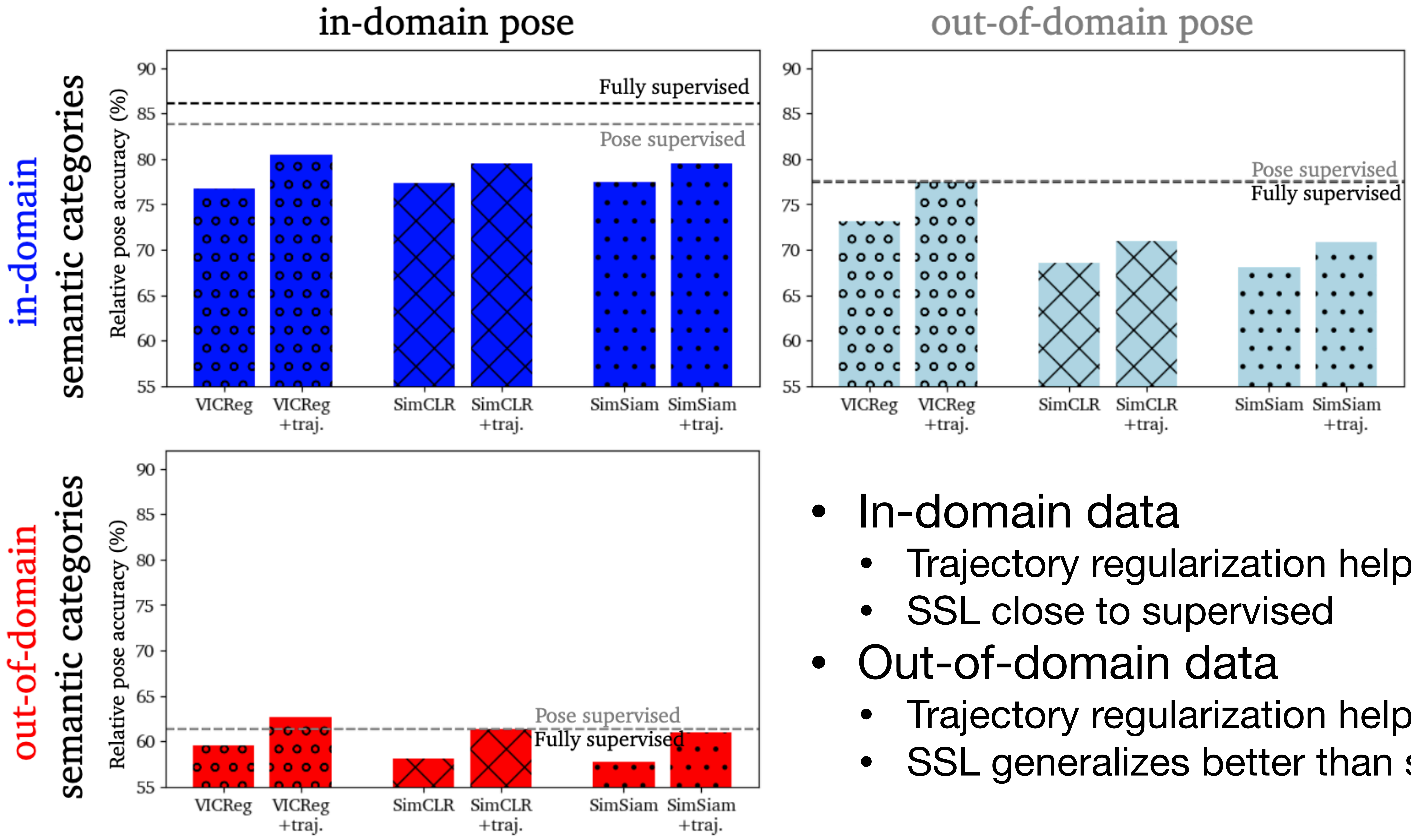
- In-domain data
  - Trajectory regularization helps
  - SSL close to supervised

Improved In-Domain and Out-of-Domain Pose Accuracy



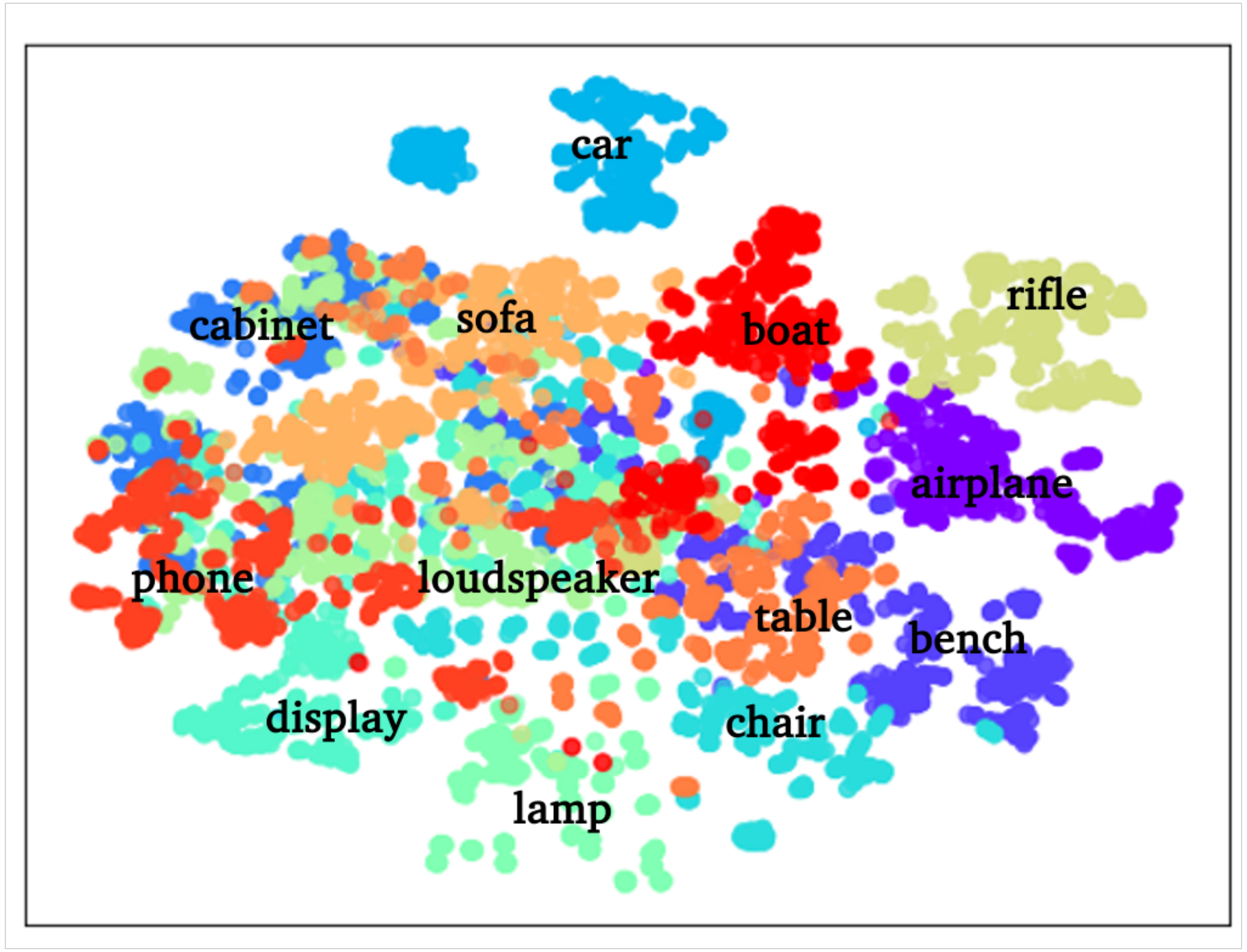
- In-domain data
  - Trajectory regularization helps
  - SSL close to supervised
- Out-of-domain data
  - Trajectory regularization helps
  - SSL generalizes better than supervised

Improved In-Domain and Out-of-Domain Pose Accuracy



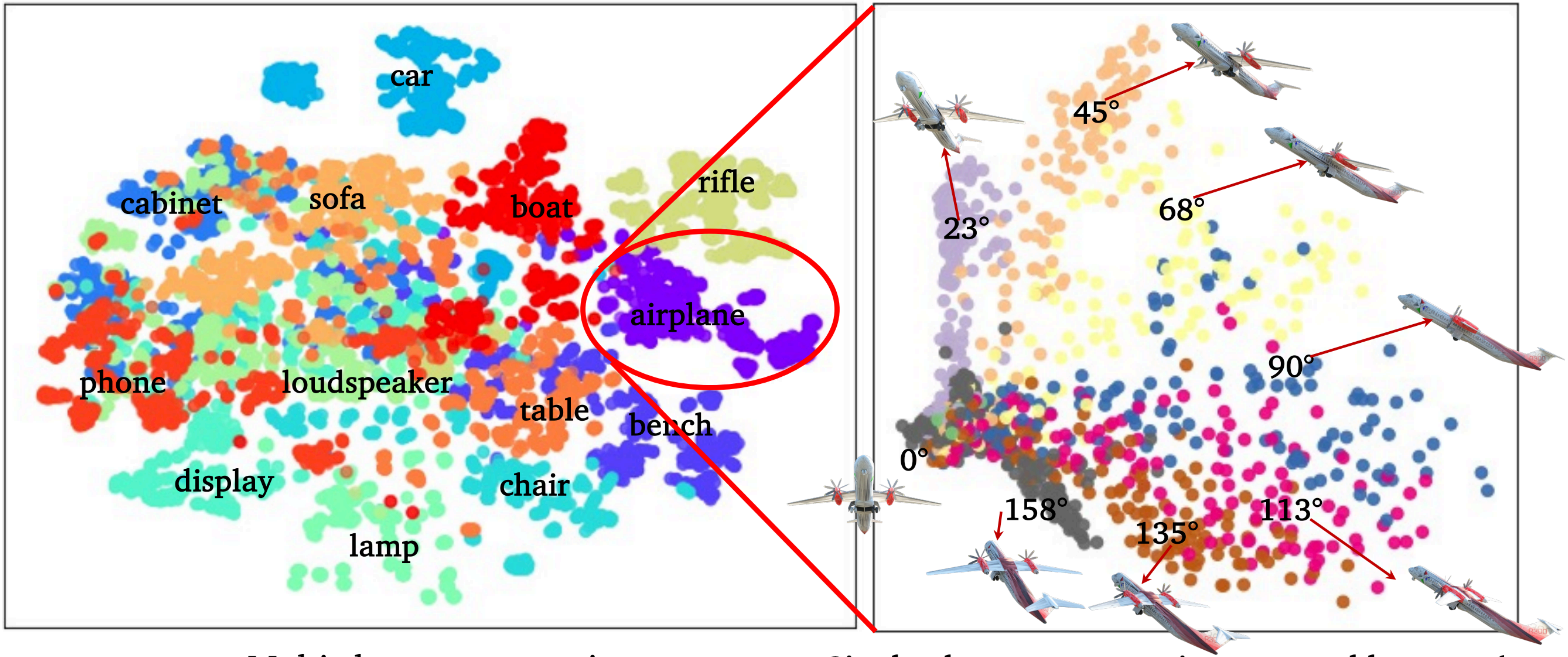
- In-domain data
  - Trajectory regularization helps
  - SSL close to supervised
- Out-of-domain data
  - Trajectory regularization helps
  - SSL generalizes better than supervised

Visualizing Representation



Multi-class representation

Visualizing Representation

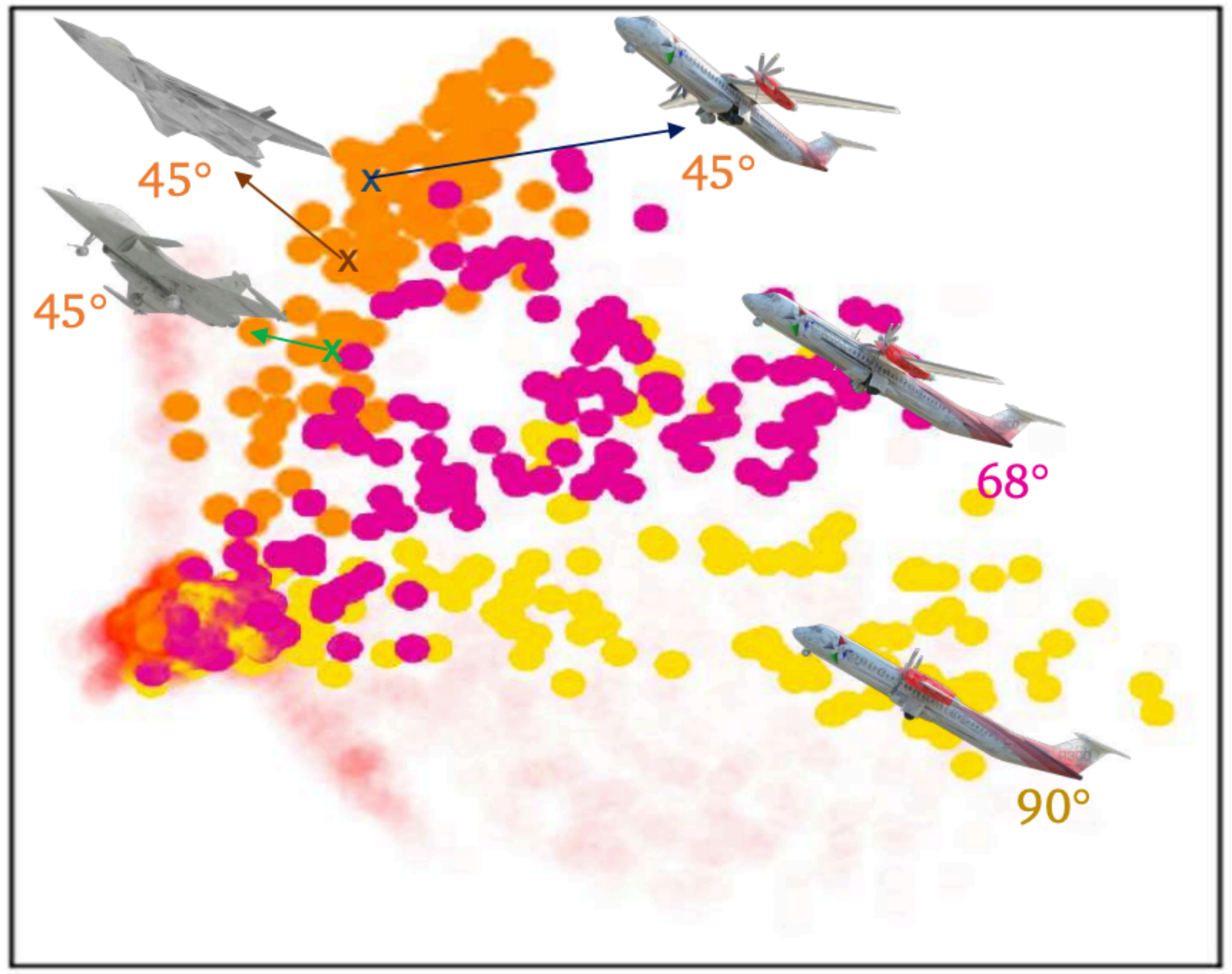


Multi-class representation

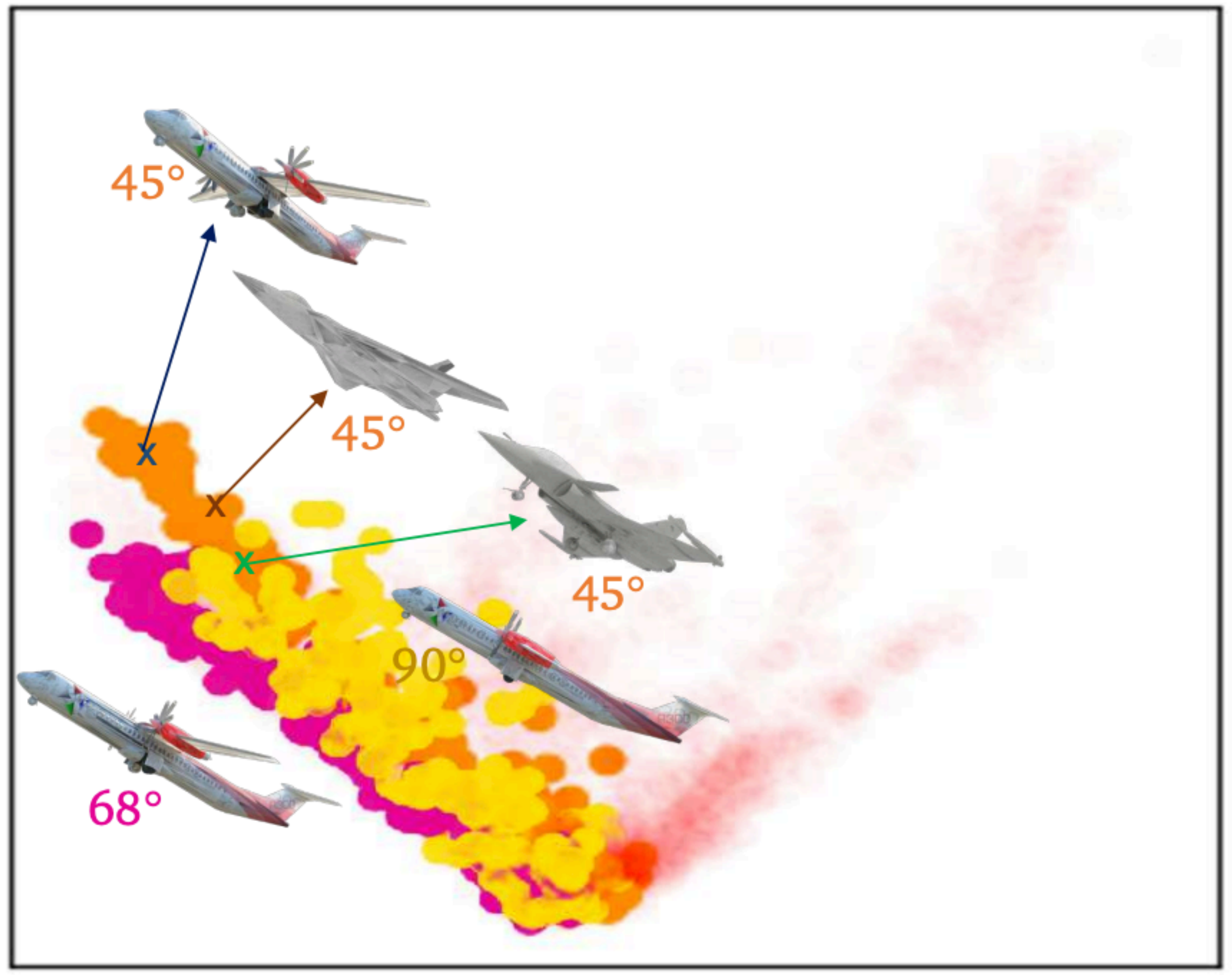
Single-class representation grouped by pose (aero)

Emergent pose-semantic representation without labels!

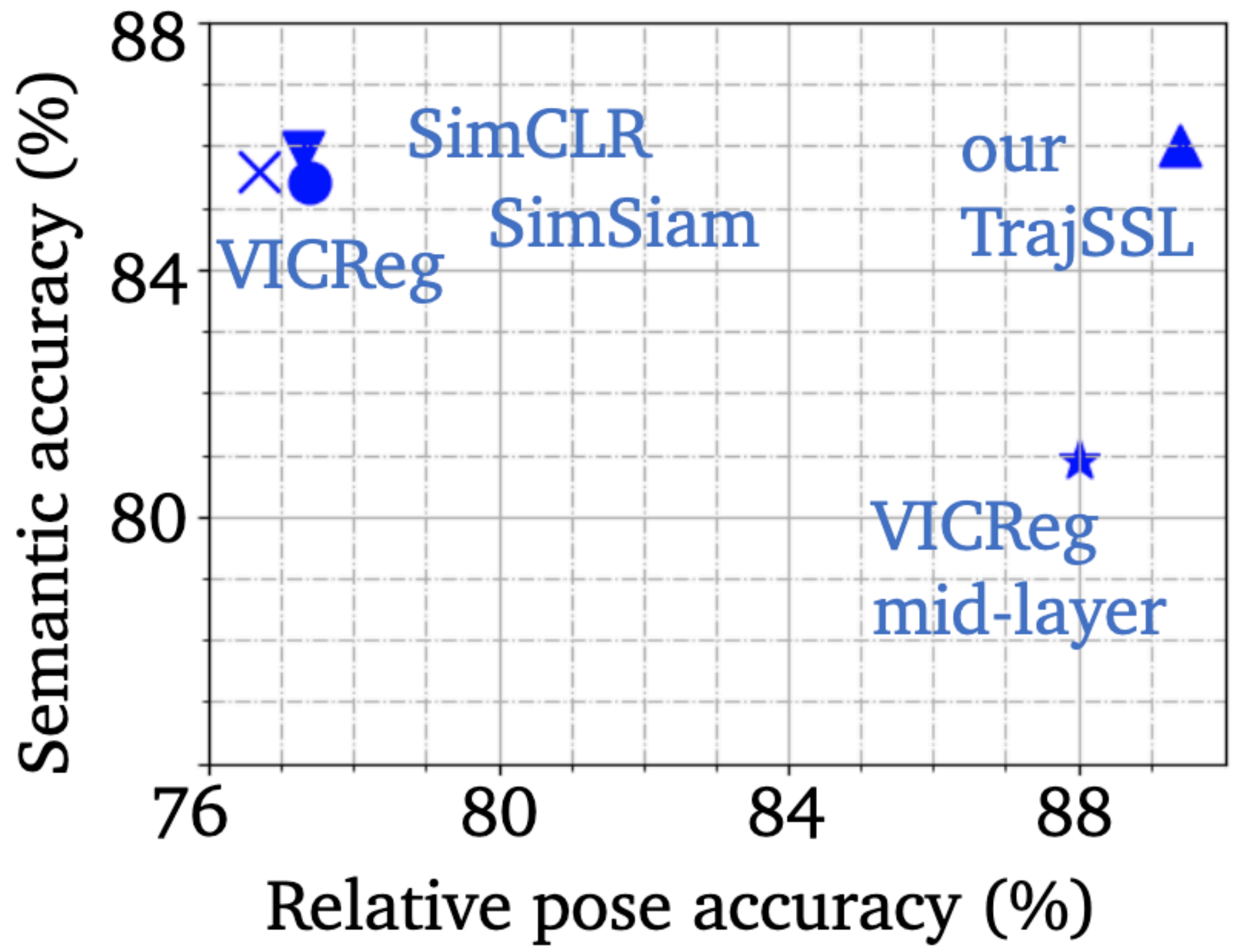
Compared to Baseline SSL



VICReg
+trajectory regularization



VICReg



VICReg: Variance-invariance-covariance regularization for self-supervised learning. ICLR 2022

Thank you!

Please come to our poster: #256

Paper/code/data:

Contact: peterw@caltech.edu