HVCLIP: High-dimensional Vector in CLIP for Unsupervised Domain Adaptation

Noranart Vesdapunt **Kah Kuen Fu** **Yue Wu**

Xu Zhang **Pradeep Natarajan**

amazon

Problem Setting

Clipart

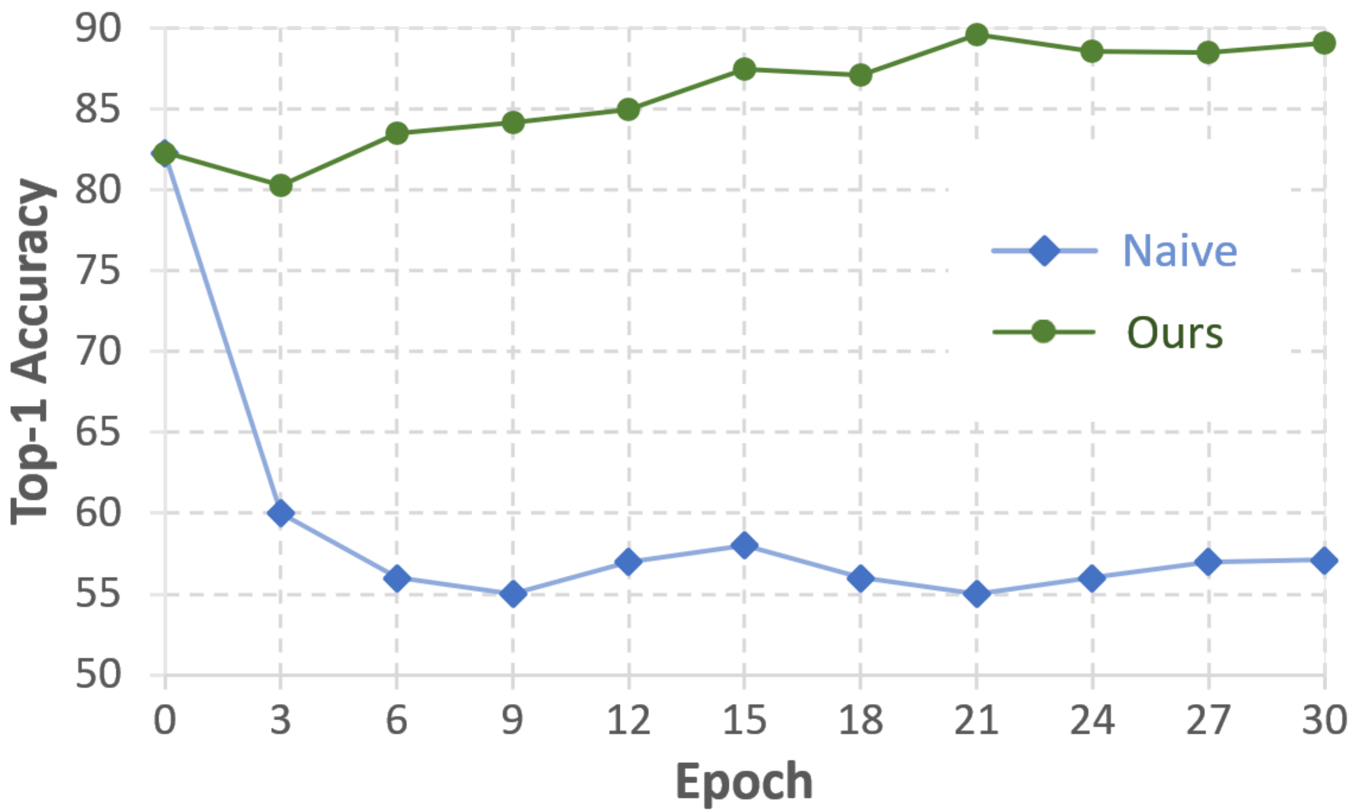
Painting

Real-world

Drawing

- Unsupervised Domain Adaptation (UDA) aims to bridge the gap between source and target domains.
- Training set is source (label) and target (unlabeled) images, and test set is target (label).

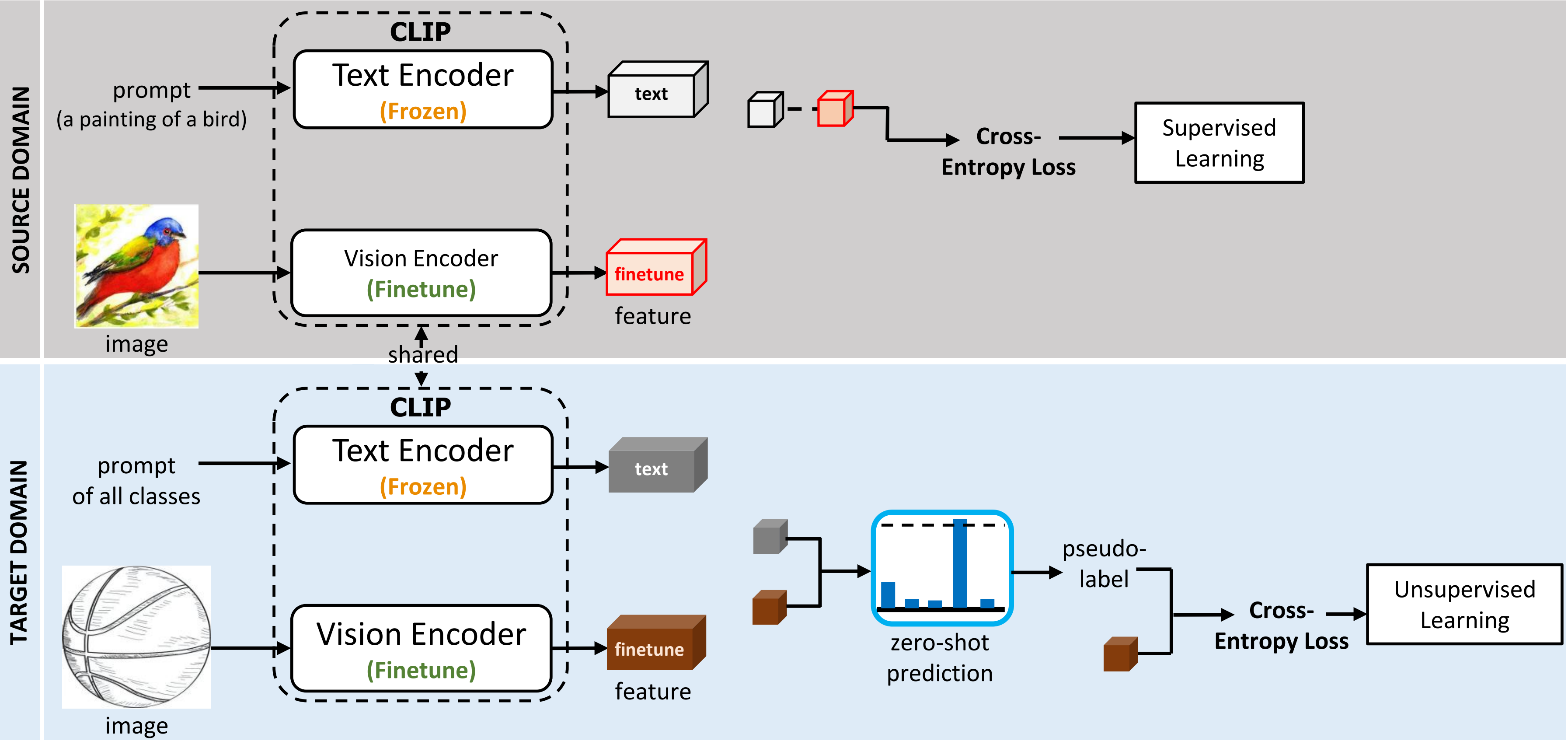
Motivation



Fine-tune CLIP on VISDA-2017

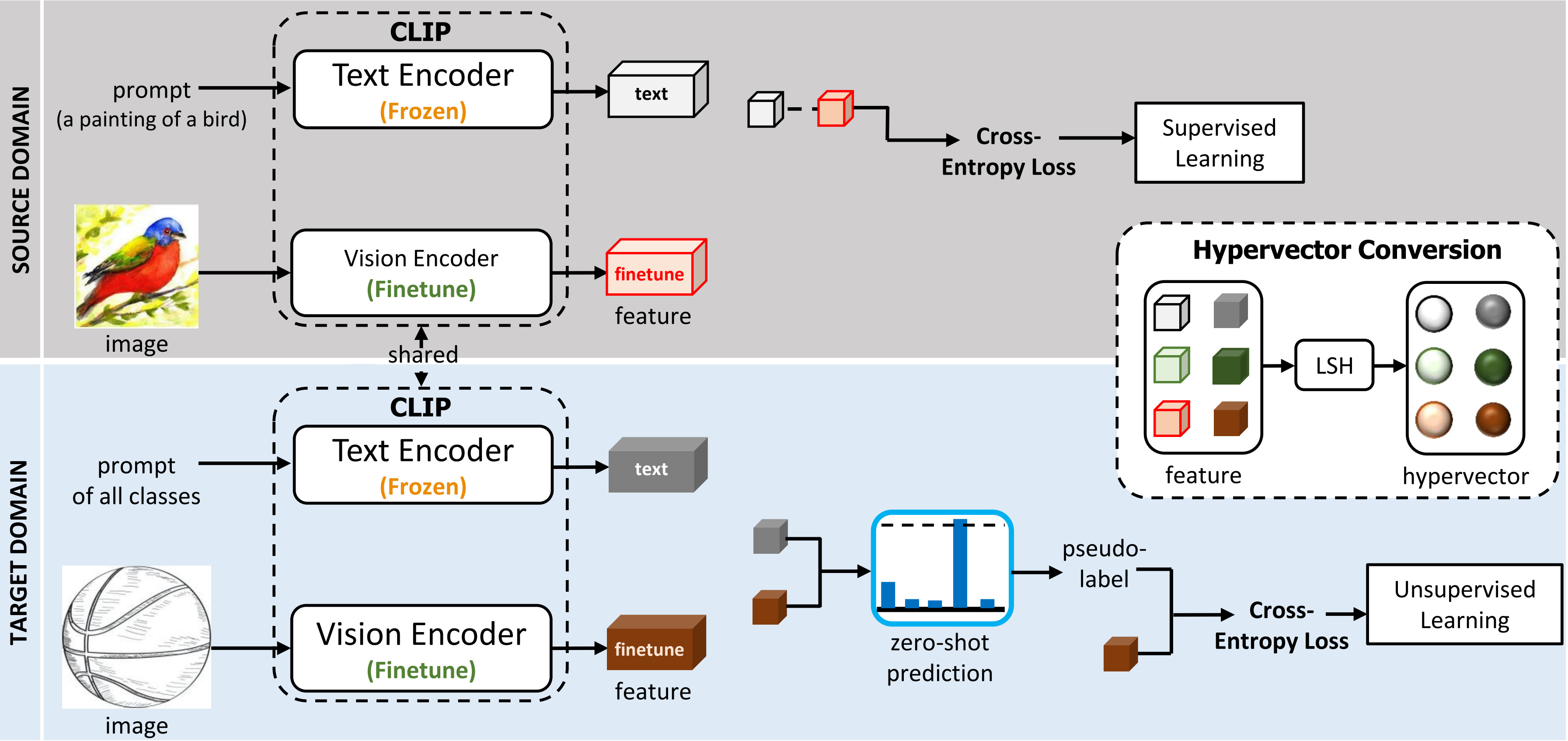
- CLIP achieves competitive accuracy out of the box.
- But fine-tuning CLIP quickly overrides pre-trained knowledge (catastrophic forgetting).
- We proposed to mitigate catastrophic forgetting by hypervector.

Method



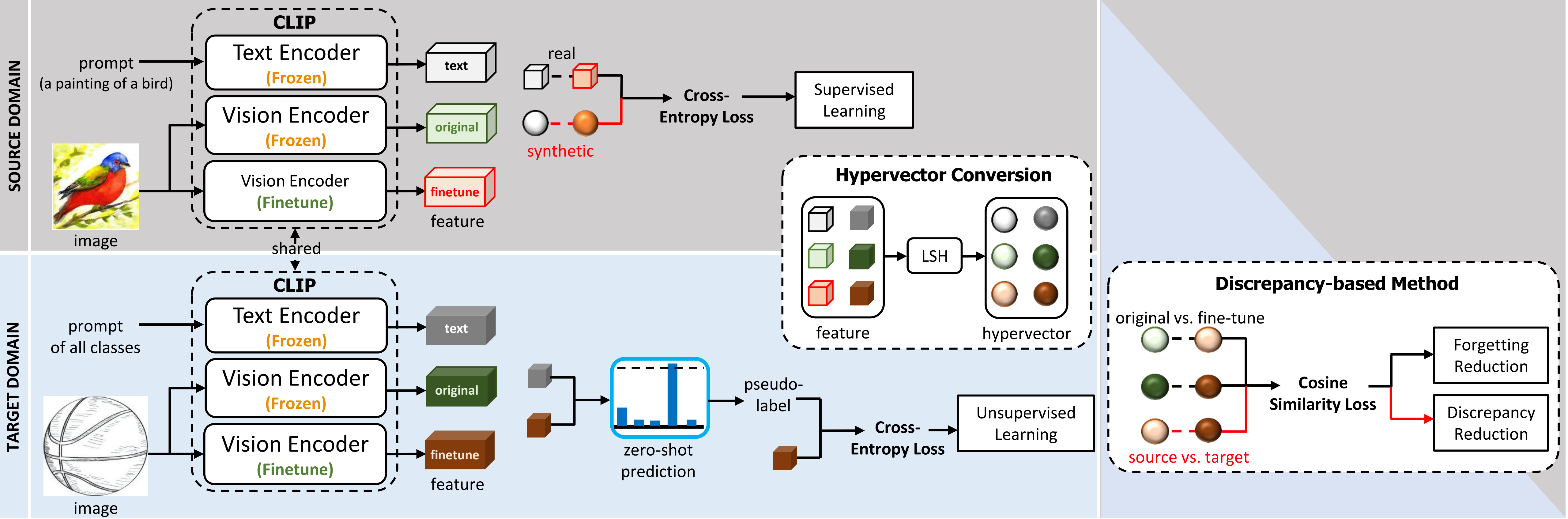
The baseline of CLIP in UDA is to fine-tune with cross entropy loss in source domain, and use zero-shot prediction to generate pseudo-label for target domain

Method



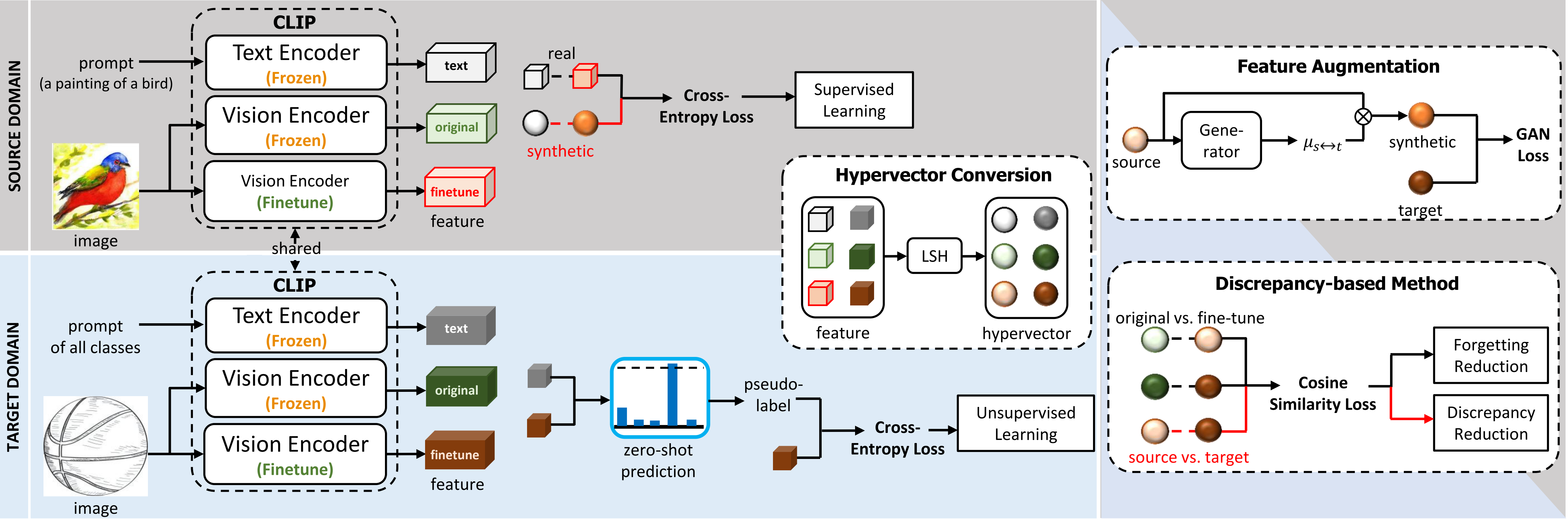
We convert features into high-dimensional vector space (hypervector). Hypervector is robust to catastrophic forgetting due to the large feature dimension (e.g., 2 millions).

Method



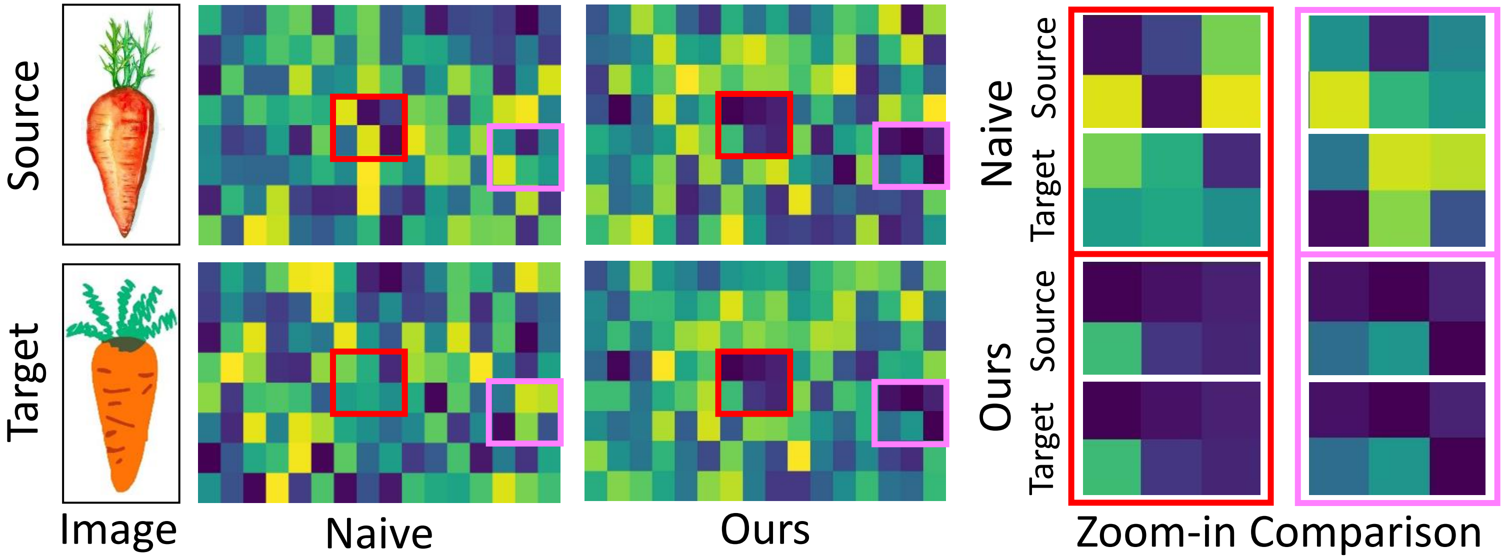
We propose Discrepancy-based Method in hypervector space to reduce the gap between source vs. target domain, and fine-tune vs. original CLIP features.

Method



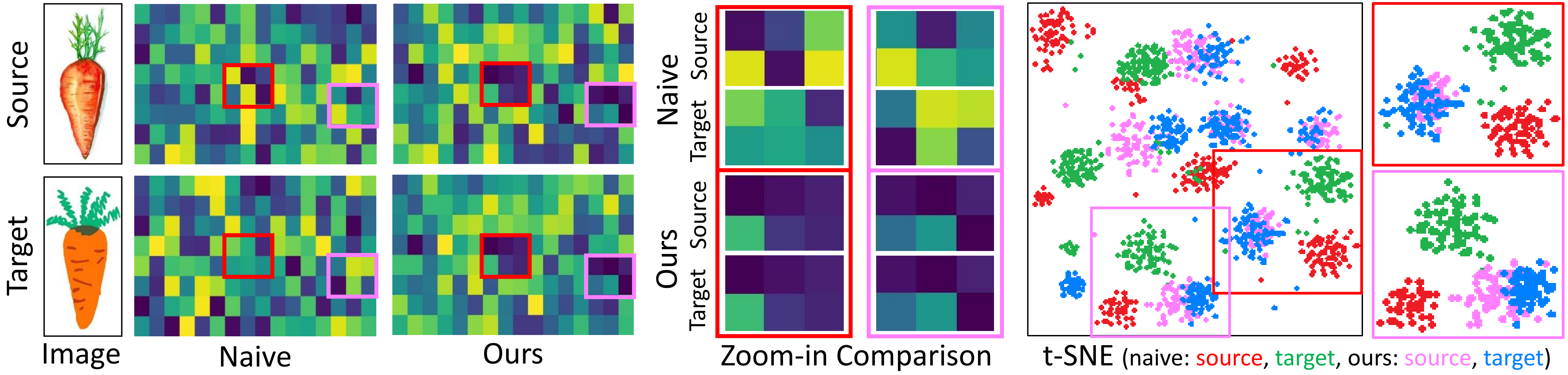
We propose Feature Augmentation in hypervector space by synthesizing target domain features from source domain features.

Visualization



Our features between source vs. target domain are more similar than the naïve fine-tuning

Visualization



Our t-SNE visualization also shows that our features between source vs. target domain are more similar than the naïve fine-tuning.

Results

Dataset	State-of-the-Art		Ours
DomainNet	LLaVO	64.7%	66.9%
VISDA-2017	VFR	91.7%	92.5%
Office-31	PMTrans	95.0%	95.5%
Office-Home	LLaVO	91.6%	92.0%

We achieved the new state-of-the-art on 4 popular UDA datasets.