TEXAS
The University of Texas at Austin

UT Swarm Lab

# Exploiting **Distribution Constraints** for **Scalable** and **Efficient** Image **Retrieval**

**Mohammad Omama**, Po-han Li, Sandeep Chinchali

ICLR 2025

# Two Key Problems with Image Retrieval

- Dataset Specific Models
  - SOTA image retrieval methods train large models for each dataset.
  - This is not **scalable**.

- Huge Embedding Sizes
  - SOTA image retrieval methods use large embeddings.
  - Retrieval speed is directly proportional to embedding size.
  - This is not **efficient**.

# How are these problems addressed (existing approaches)?

- Dataset Specific Models
    - Use off the shelf foundation models like DINO/CLIP
    - They lack in performance!

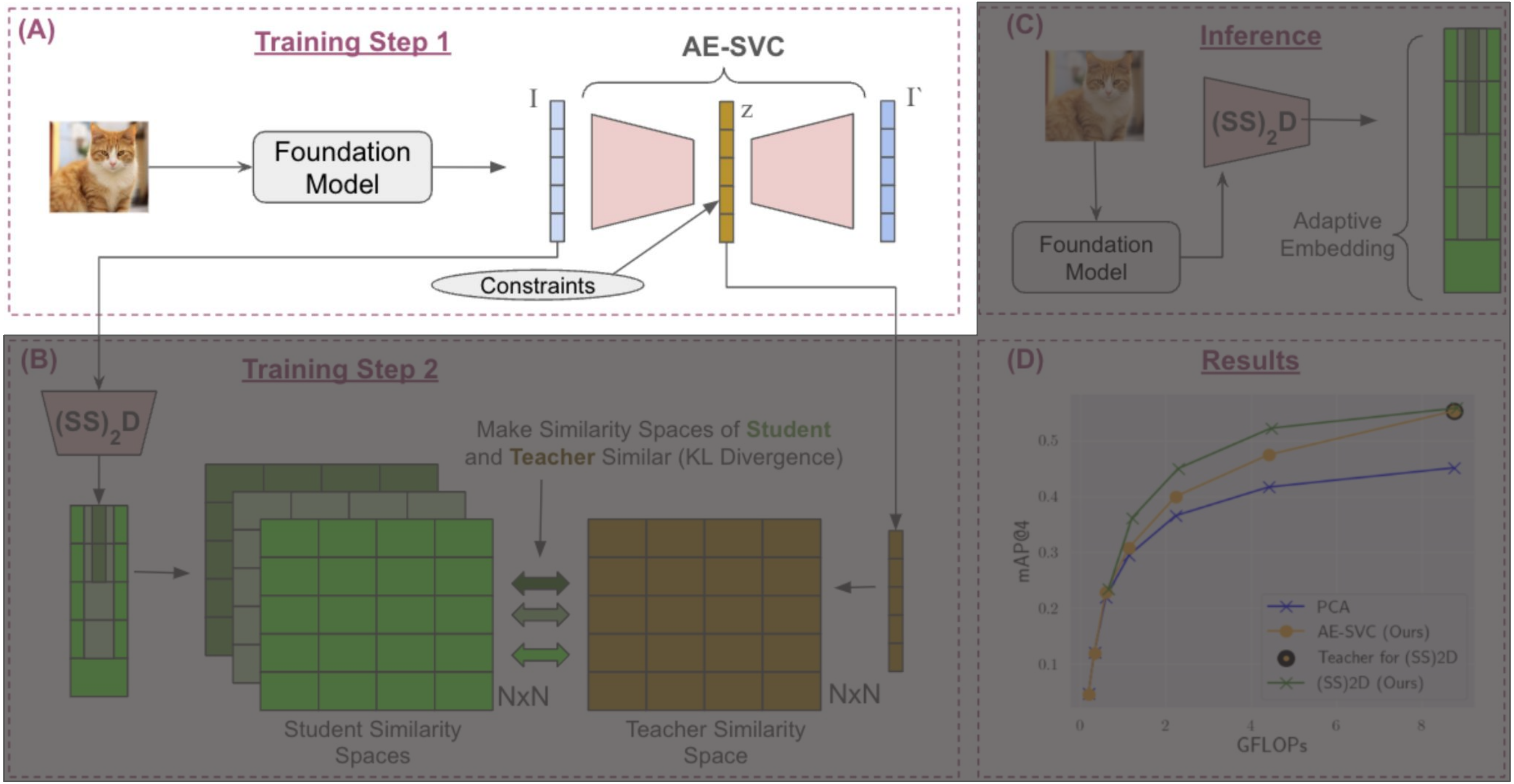
- Huge Embedding Sizes
    - Dimensionality reduction with PCA/Auto-encoders
    - They are not tuned for retrieval!
    - Further Auto-encoders require training separate model for each dimension size.

➡️ We need a solution for **scalable** and **efficient** image retrieval!

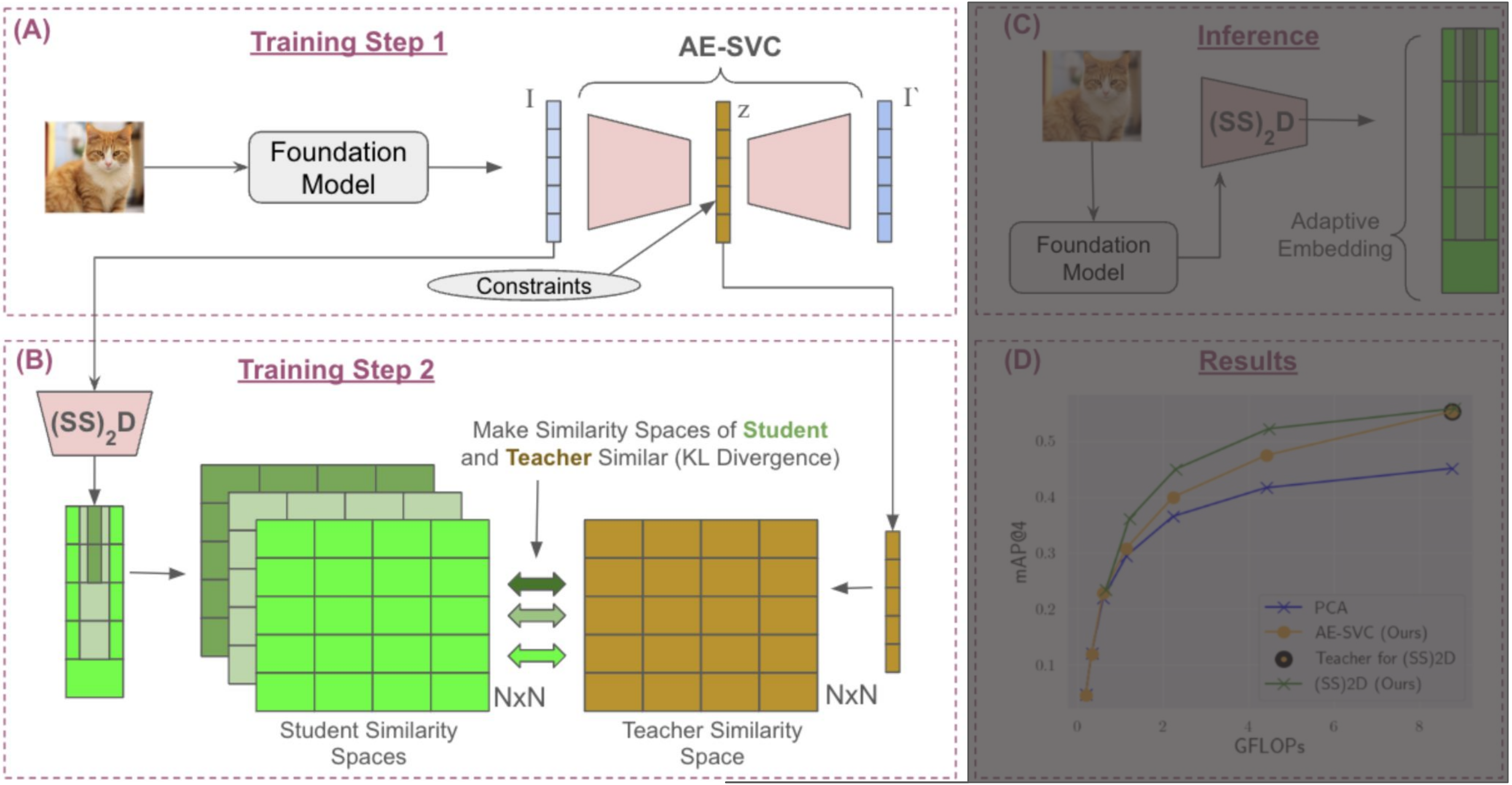
# Our Key Insights

- On the scalability side:
  - Foundation models capture necessary subtleties for effective retrieval. 😊
  - The underlying distribution of their embedding space can negatively impact cosine similarity searches. 😓
  - We need to improve the underlying distribution. 💡
- On the efficiency side:
  - Dimensionality reduction techniques like PCA/Auto-encoders either focus on information maximization or reconstruction quality. 😓
  - We need dimensionality reduction tuned to cosine searches. 💡

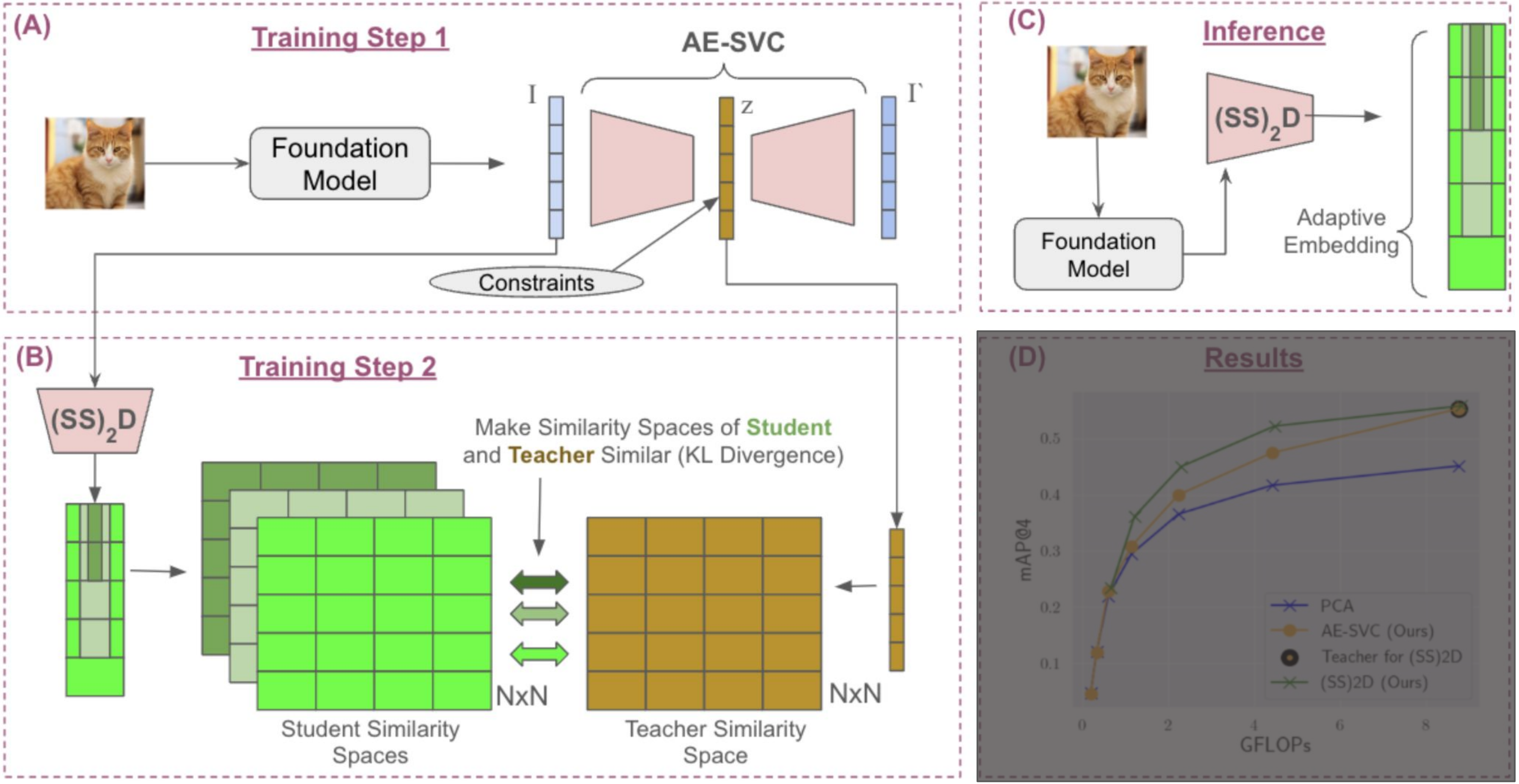
# Our Two-step Solution



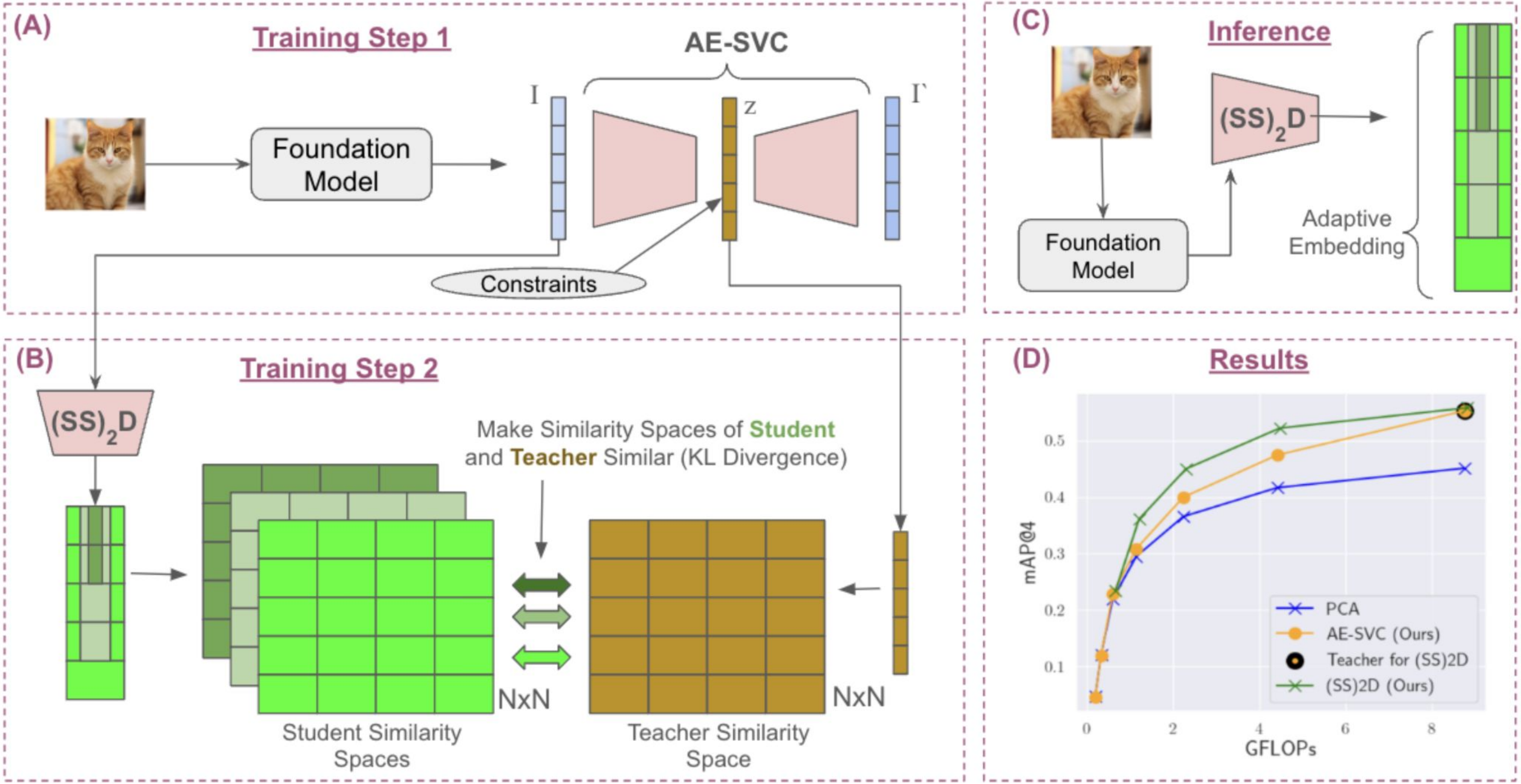
# Our Two-step Solution



# Our Two-step Solution

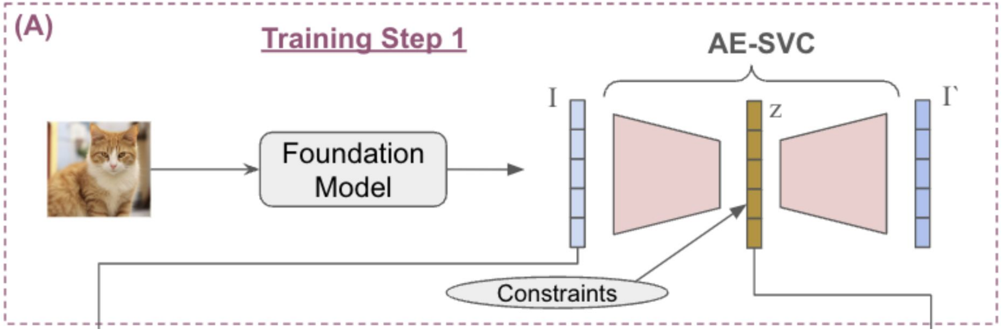


# Our Two-step Solution



# Autoencoders with Strong Variance Constraints (AE-SVC)

# AE-SVC | The Losses Explained




$$\mathcal{L}_{\text{rec}} = \frac{1}{n}\sum_{i=1}^{n} \|I_i - I_i'\|_2^2,$$

$$\mathcal{L}_{\text{cov}} = \left\| \frac{1}{n}(Z-\mu)^\top (Z-\mu) - \mathbb{I} \right\|_F^2,$$

$$\mathcal{L}_{\text{var}} = \frac{1}{d}\sum_{i=1}^{d} \left(\text{Var}(z^i) - 1\right)^2,$$

$$\mathcal{L}_{\text{mean}} = \frac{1}{d}\sum_{i=1}^{d} (\mu^i)^2,$$

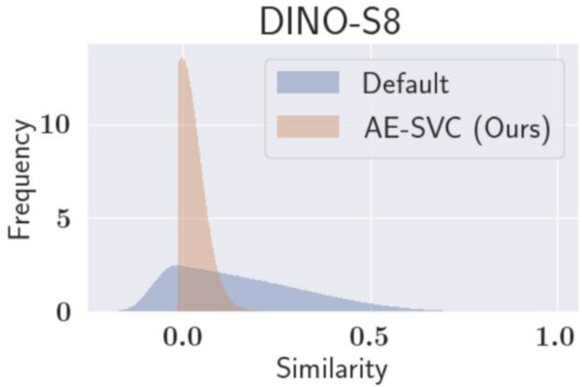
# AE-SVC | Insight Into the Constraints

- Why do these constraints help?

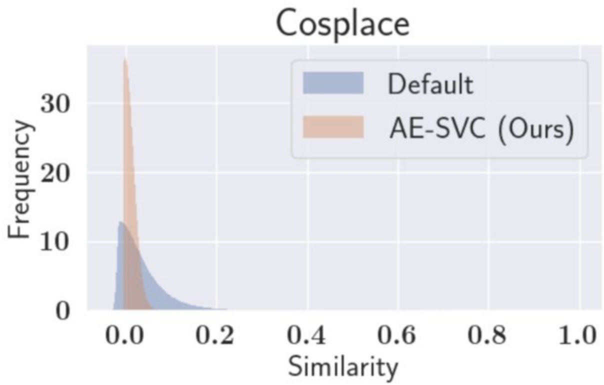
Constraints on
**Embedding** Distributions

Minimum Variance in
**CosSim** Distributions

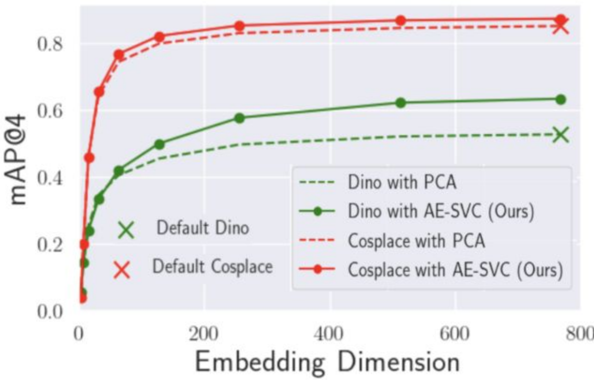
# AE-SVC | Insight Into the Constraints




(a)



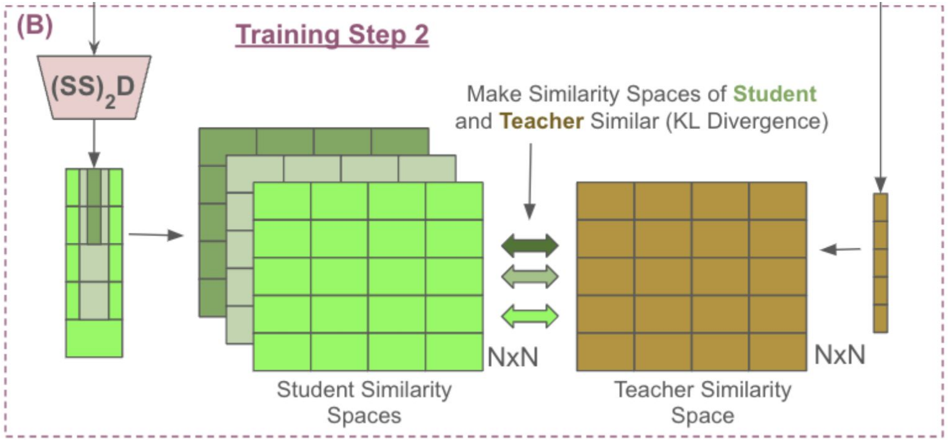

(b)




(c)

# Single Shot Similarity Space Distillation
# $(SS)_2D$

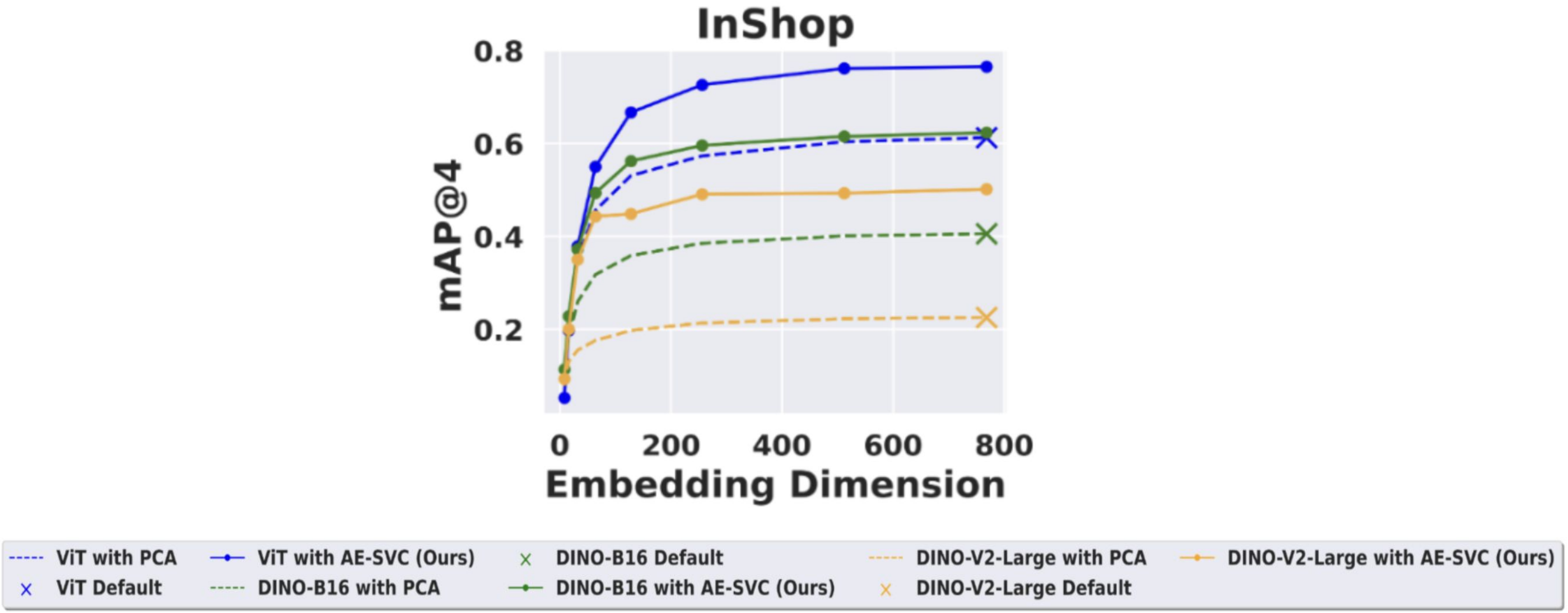
# $(SS)_2D$ | Losses Explained



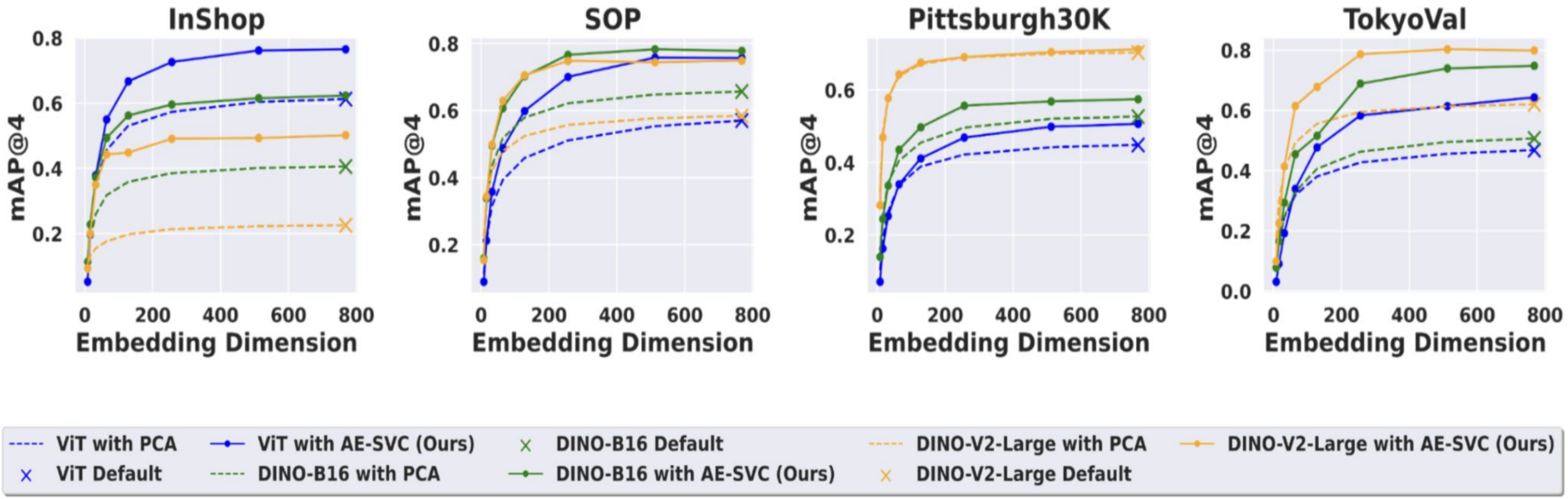

$$l^m = \sum_i \mathrm{D_{KL}}(\tilde{\mathrm{C}}_i^m \| \mathrm{C}_i).$$

$$L_{(SS)_2D} = \sum_m l^m = \sum_m \sum_i \mathrm{D_{KL}}(\tilde{\mathrm{C}}_i^m \| \mathrm{C}_i).$$

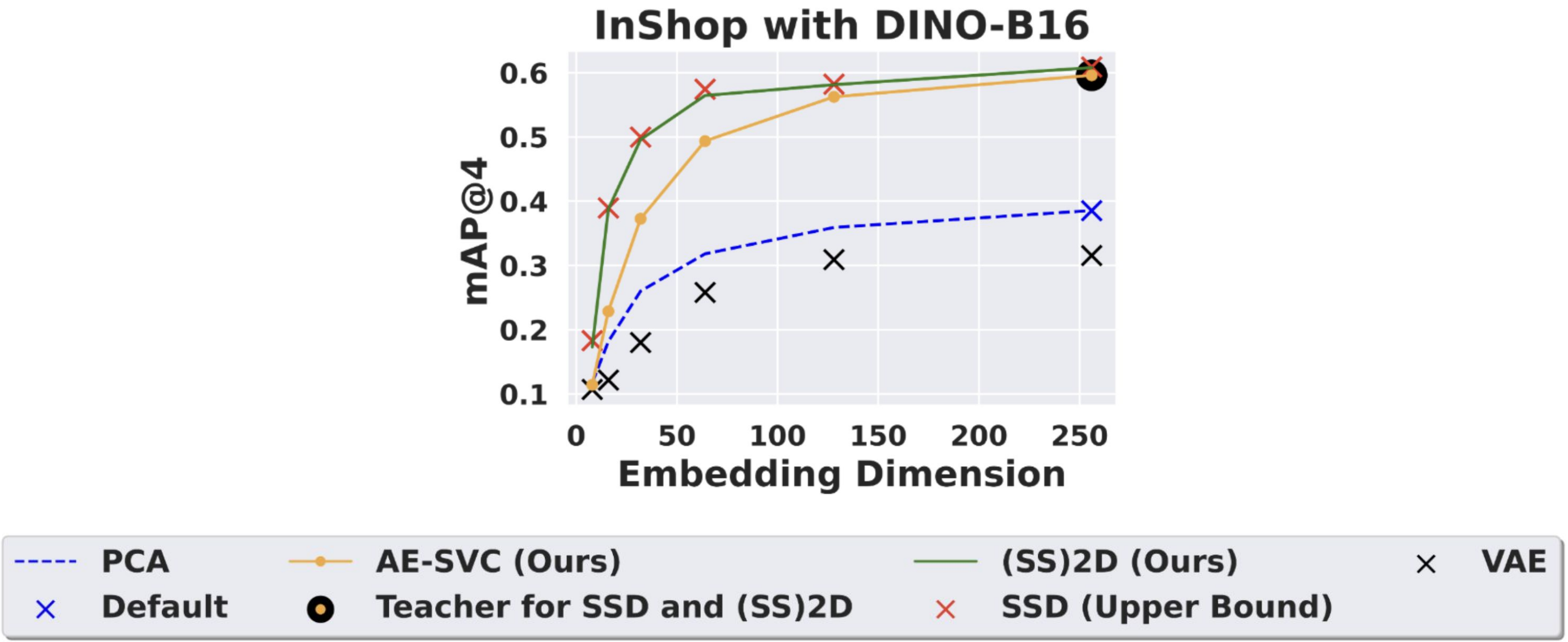
# AE-SVC | Results



# AE-SVC | Results



# (SS)2D | Results



# (SS)2D | Results