WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

NEURAL INFORMATION PROCESSING SYSTEMS

Dream the Impossible: Outlier Imagination with Diffusion Models

NeurIPS 2023

Xuefeng Du
UW-Madison

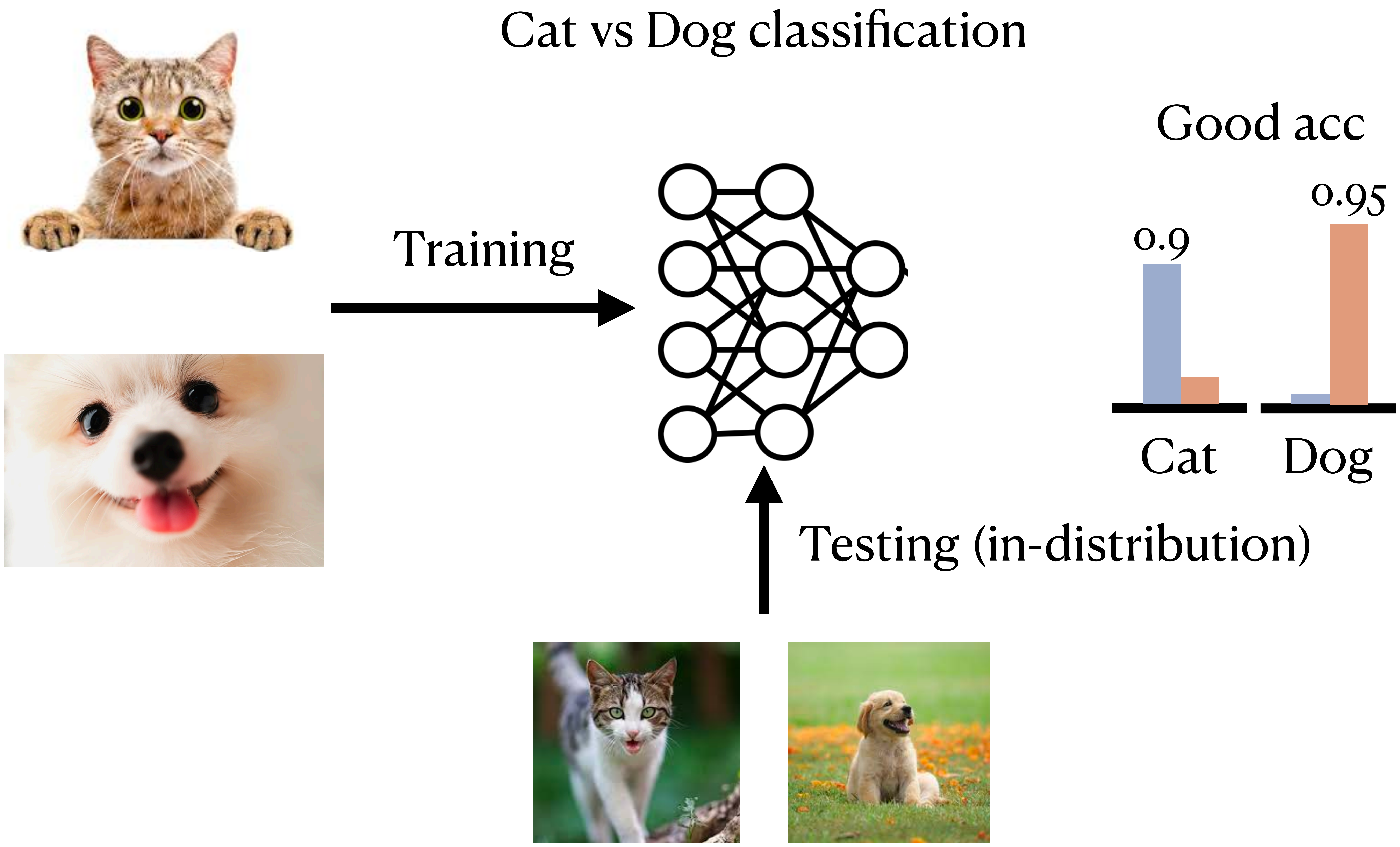
Yiyou Sun
UW-Madison

Jerry Zhu
UW-Madison

Sharon Yixuan Li
UW-Madison

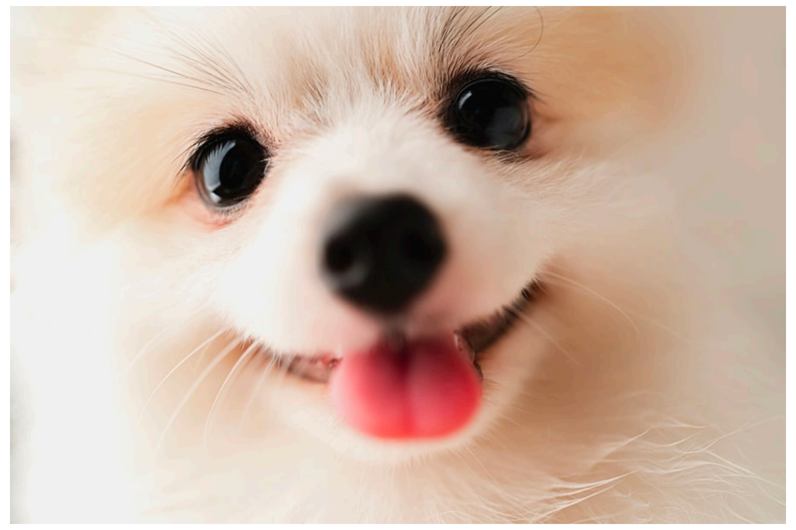
Background

Closed-world ML on In-Distribution (ID) Data

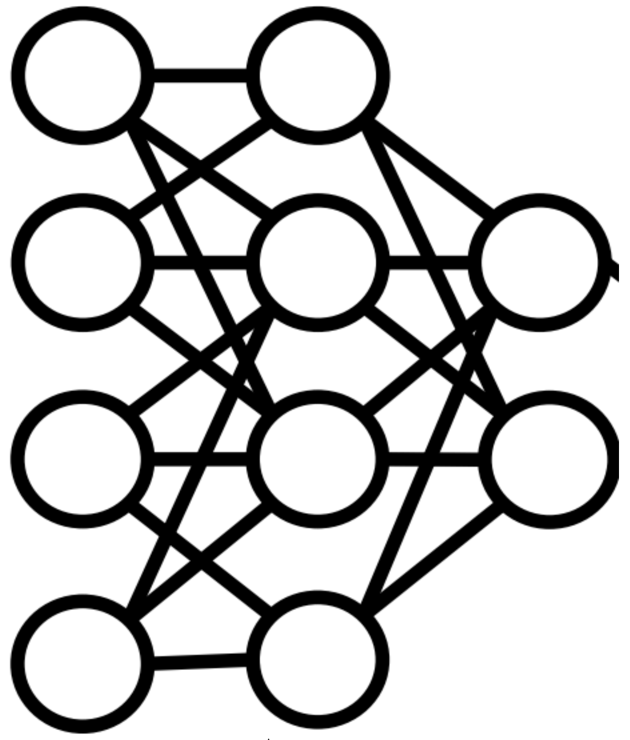


ML Meets Out-of-distribution (OOD) Data

Cat vs Dog classification



Training



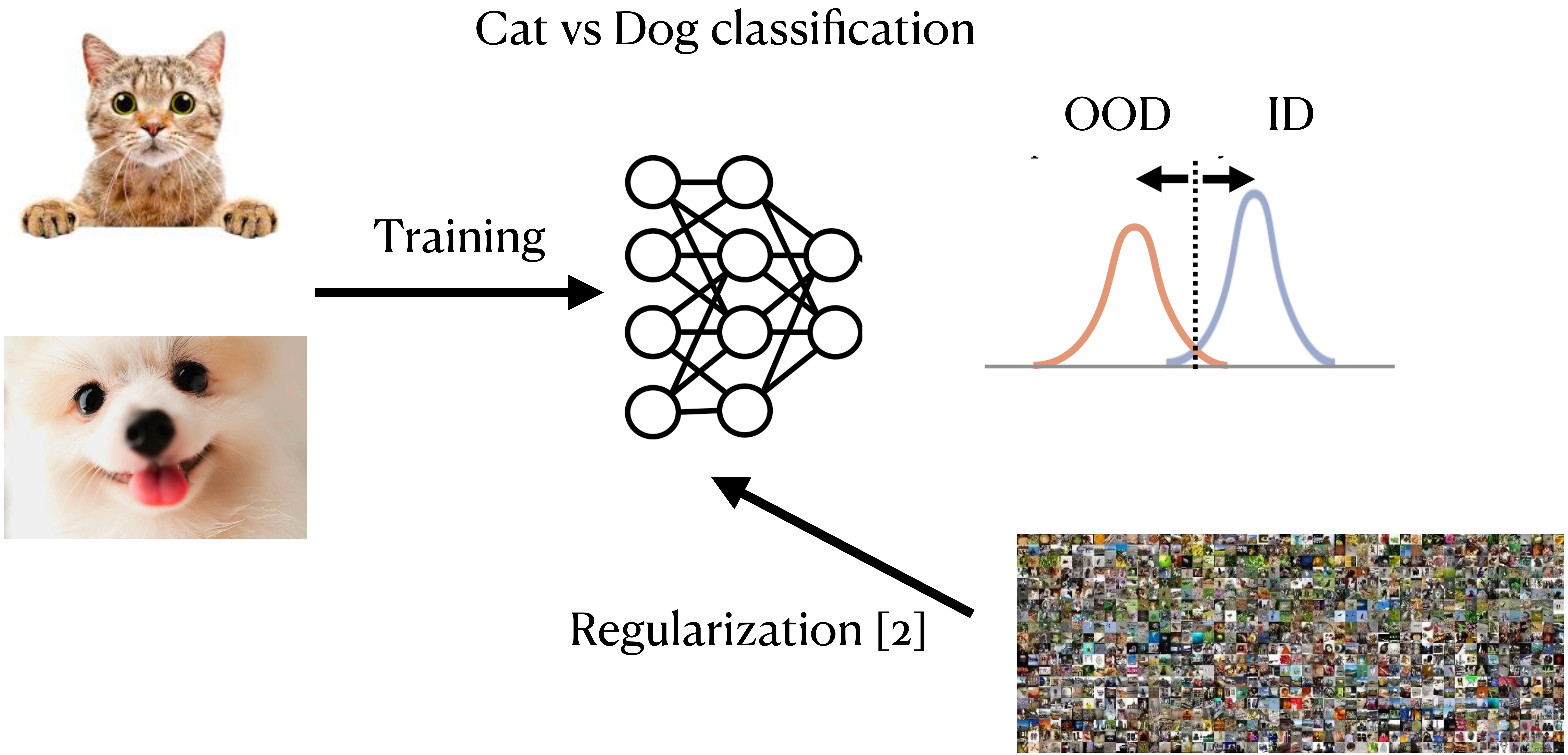
Overconfident
but **wrong** predictions [1]!

Testing (out-of-distribution)

How to solve this problem?

[1] Du et.al., Vos: Learning what you don't know by virtual outlier synthesis, ICLR 2022

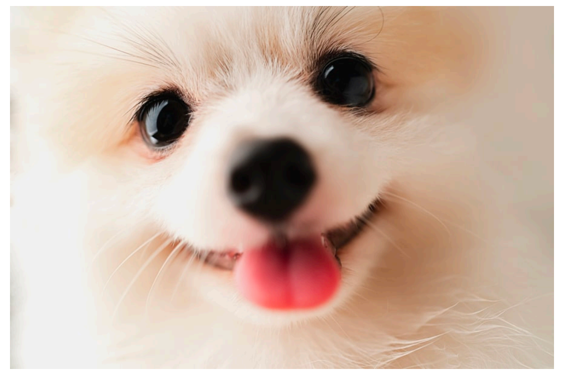
OOD Detection with Real Outliers



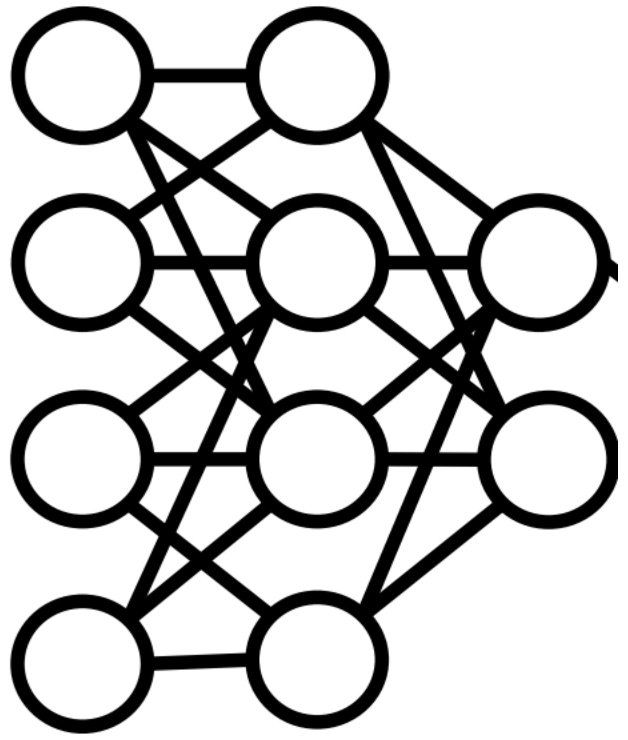
[2] Hendrycks et.al., Deep Anomaly Detection with Outlier Exposure, ICLR 2019

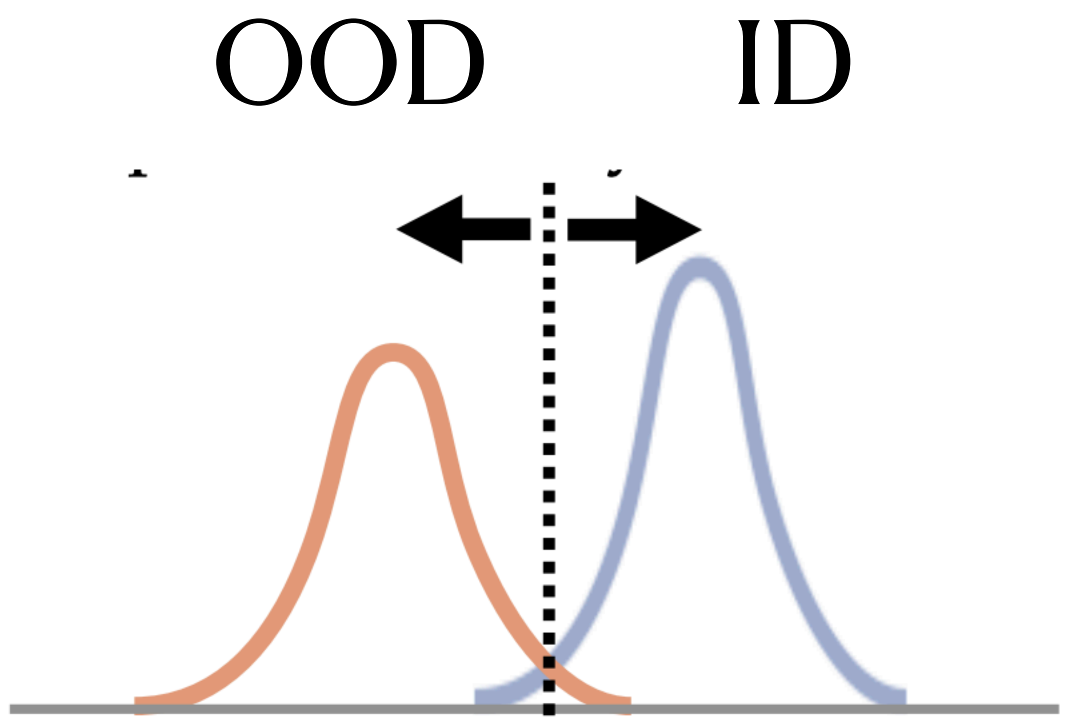
OOD Detection with Real Outliers

Cat vs Dog classification



Training





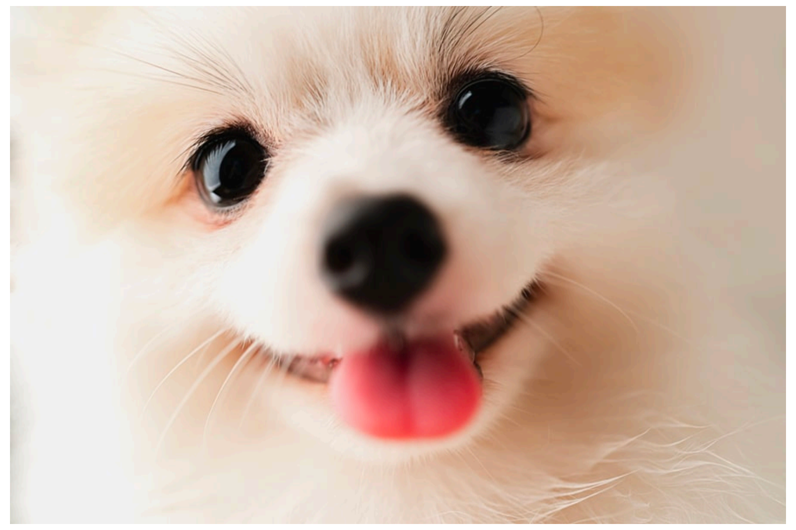
Costly to prepare,
Not easy to adapt!

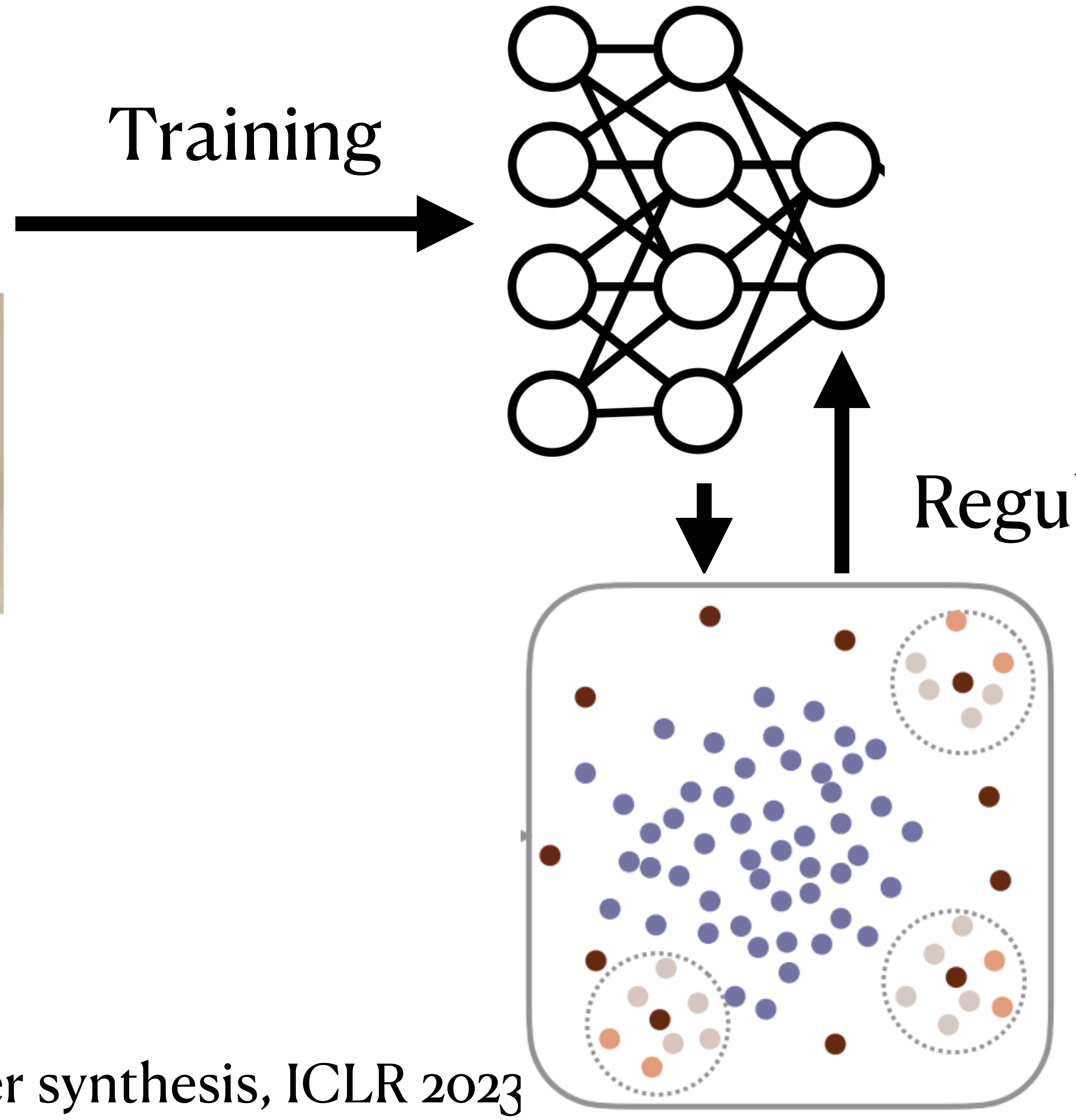
Regularization [2]

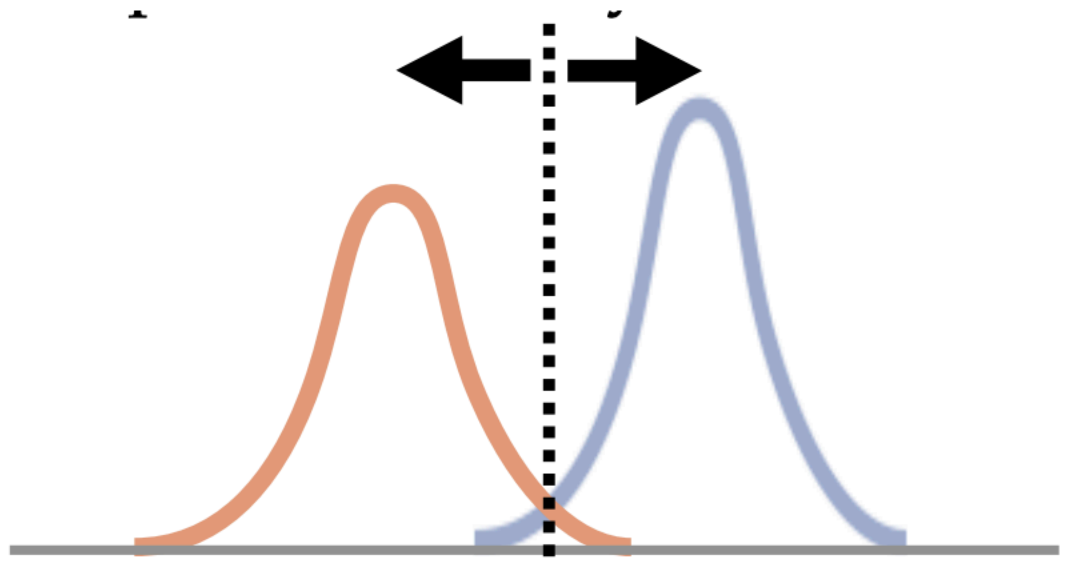
[2] Hendrycks et.al., Deep Anomaly Detection with Outlier Exposure, ICLR 2019

OOD Detection with Virtual Outliers

Cat vs Dog classification





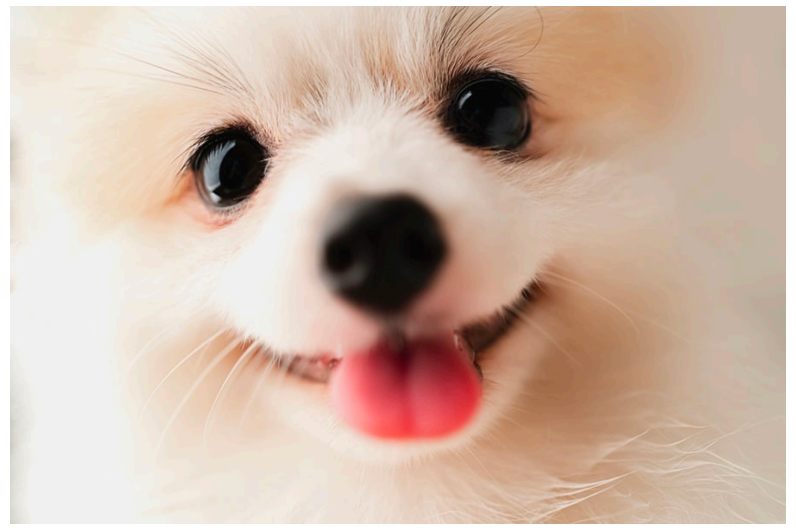


Synthesizing the outlier
in the feature space!

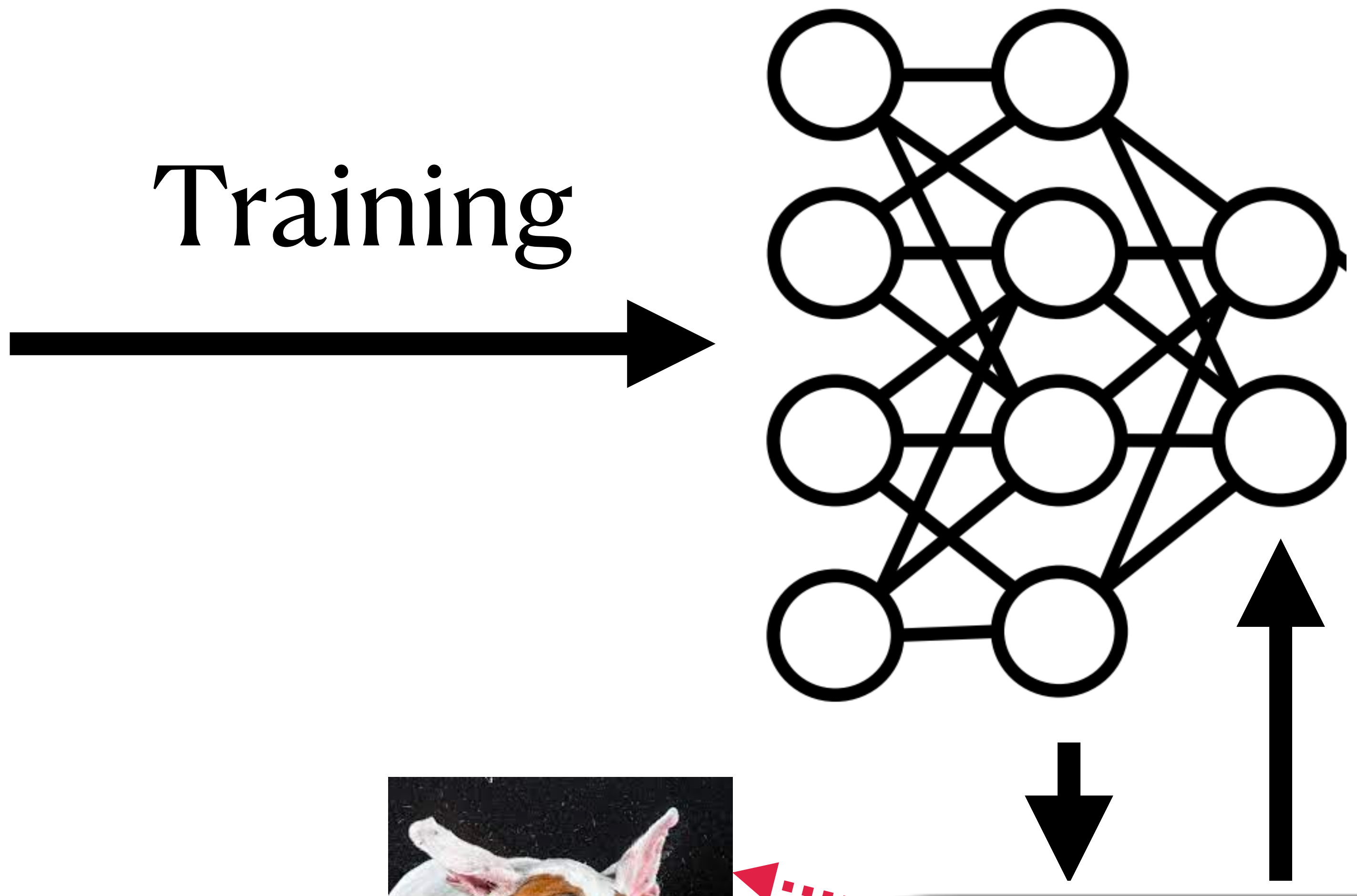
[3] Tao et.al.,Non-parametric outlier synthesis, ICLR 2023

OOD Detection with Virtual Outliers

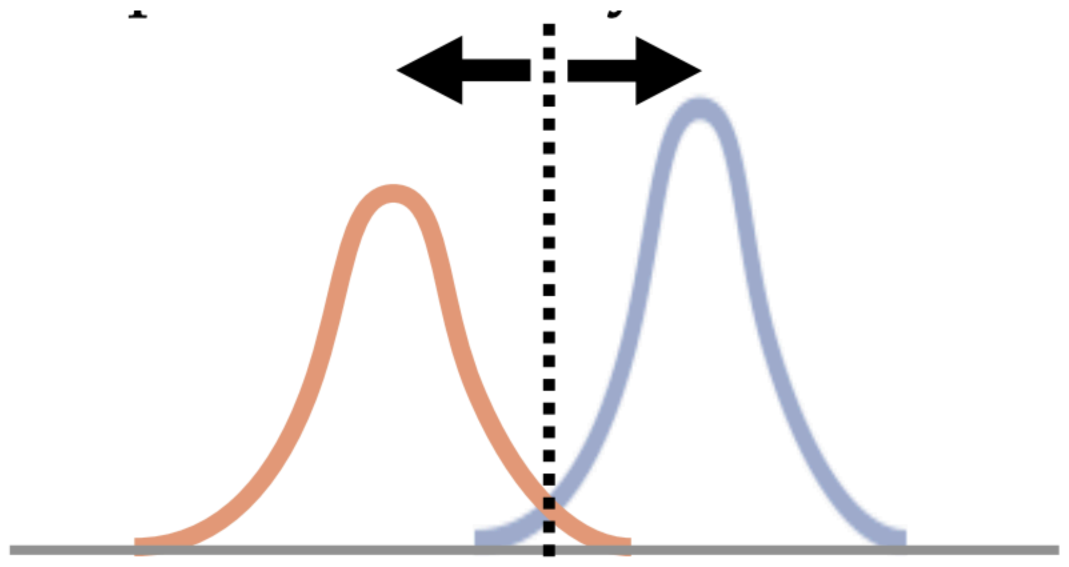
Cat vs Dog classification



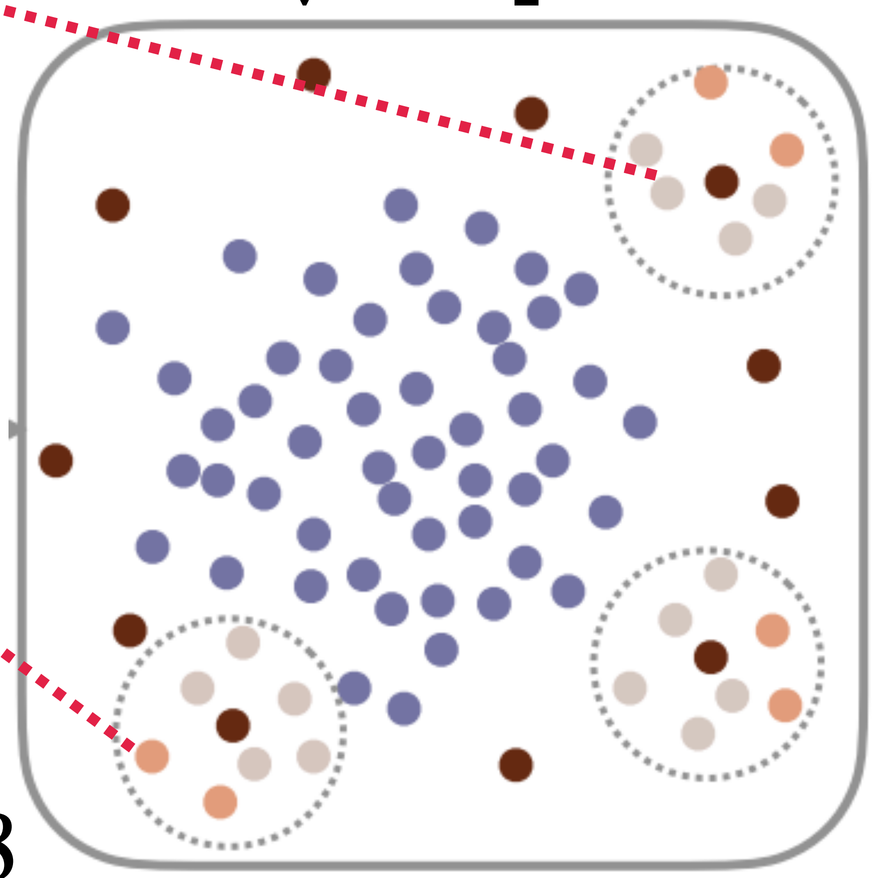
Training

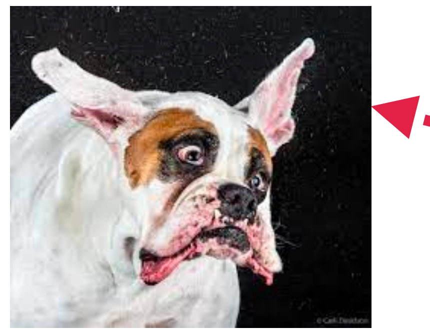


OOD ID



Regularization [1,3]





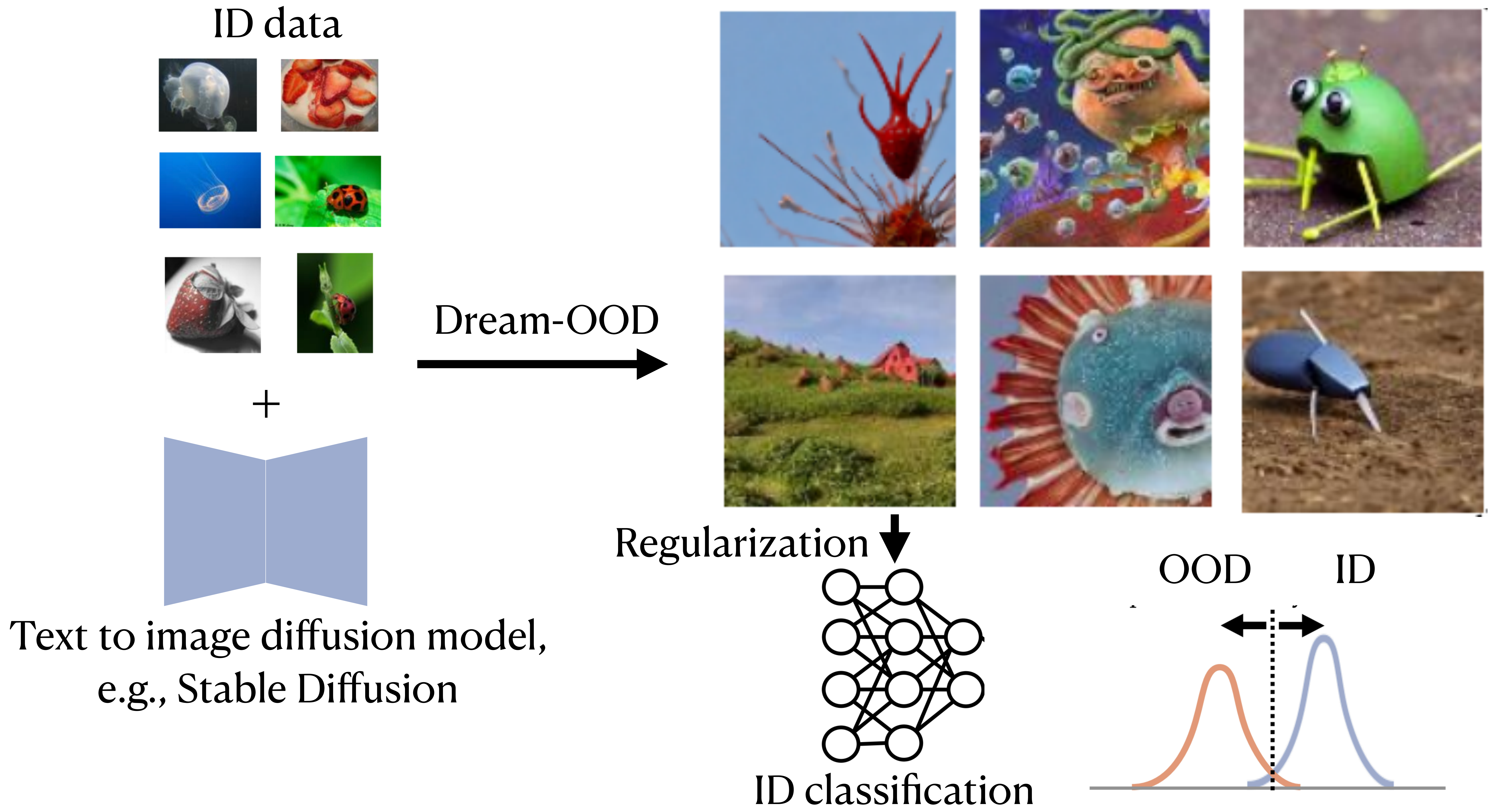
……

Cheap, flexible and adaptive,
But how to interpret them?

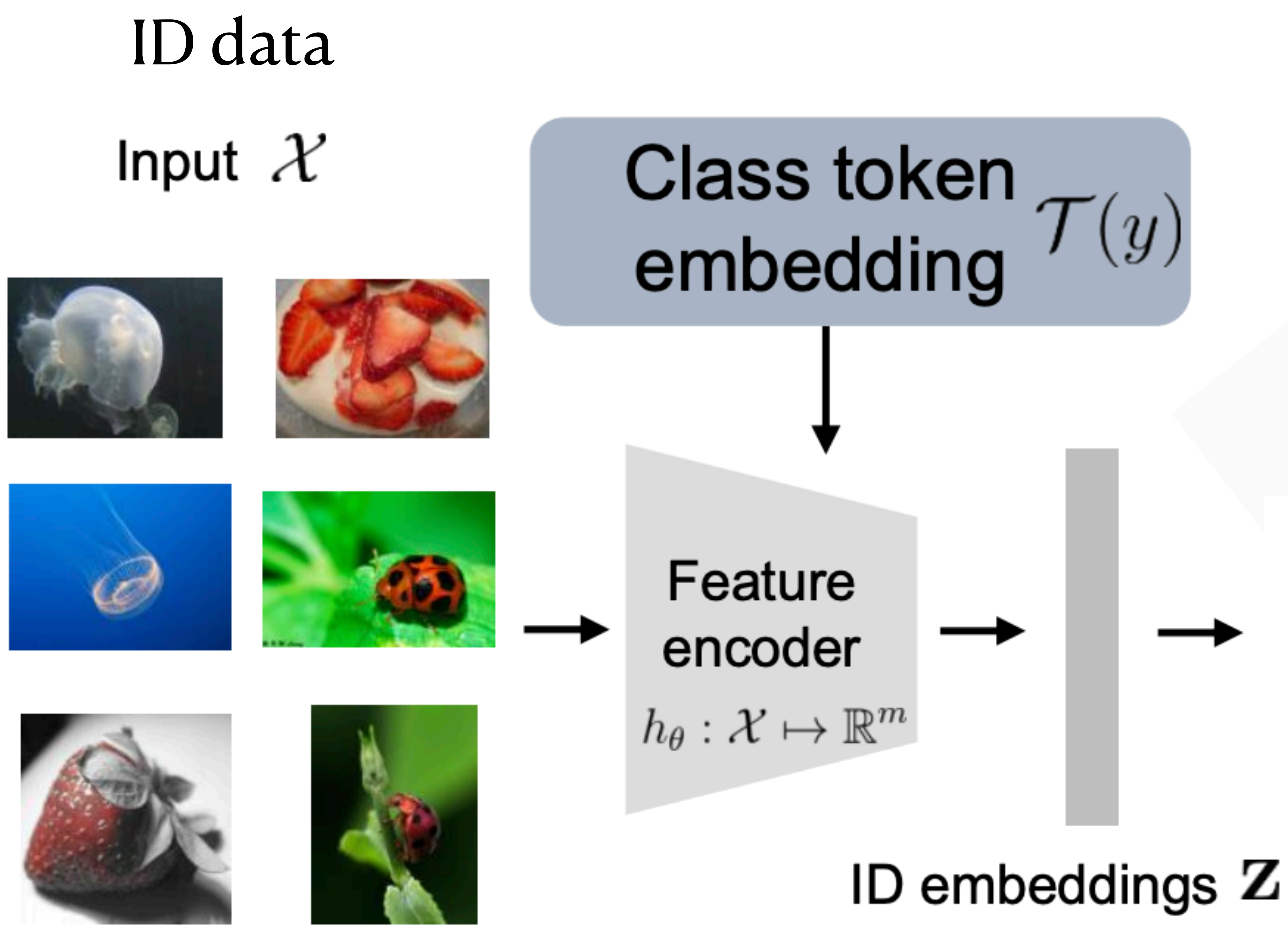
[3] Tao et.al.,Non-parametric outlier synthesis, ICLR 2023

Dream-OOD

Dream-OOD: Outlier Imagination with Diffusion Models

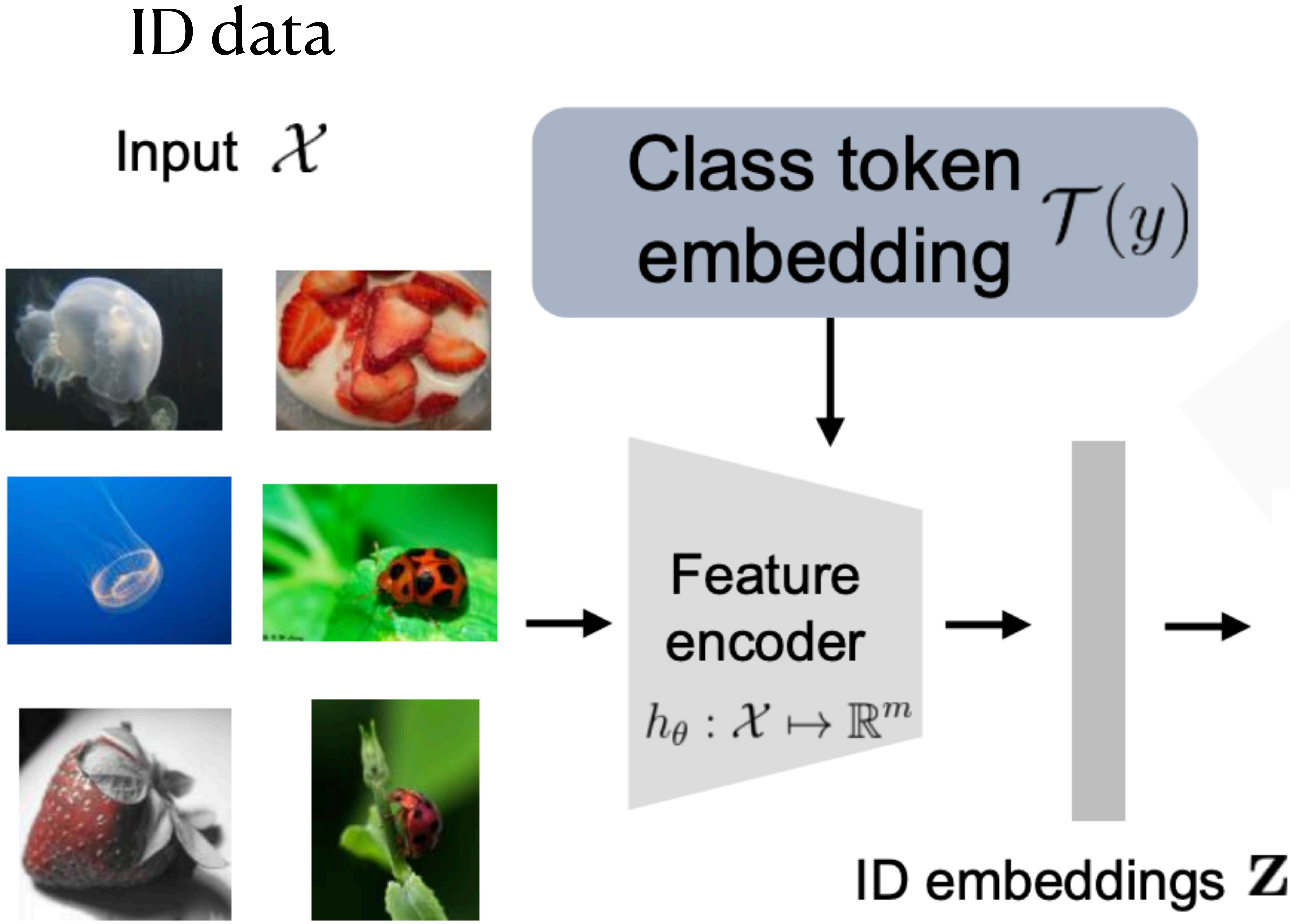


Dream-OOD: Learning the Text-Conditioned Latent Space



$$\mathcal{L} = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}}[-\log \frac{\exp\left(\mathcal{T}(y)^\top \mathbf{z}/t\right)}{\sum_{j=1}^{C}\exp\left(\mathcal{T}(y_j)^\top \mathbf{z}/t\right)}]$$

Dream-OOD: Learning the Text-Conditioned Latent Space



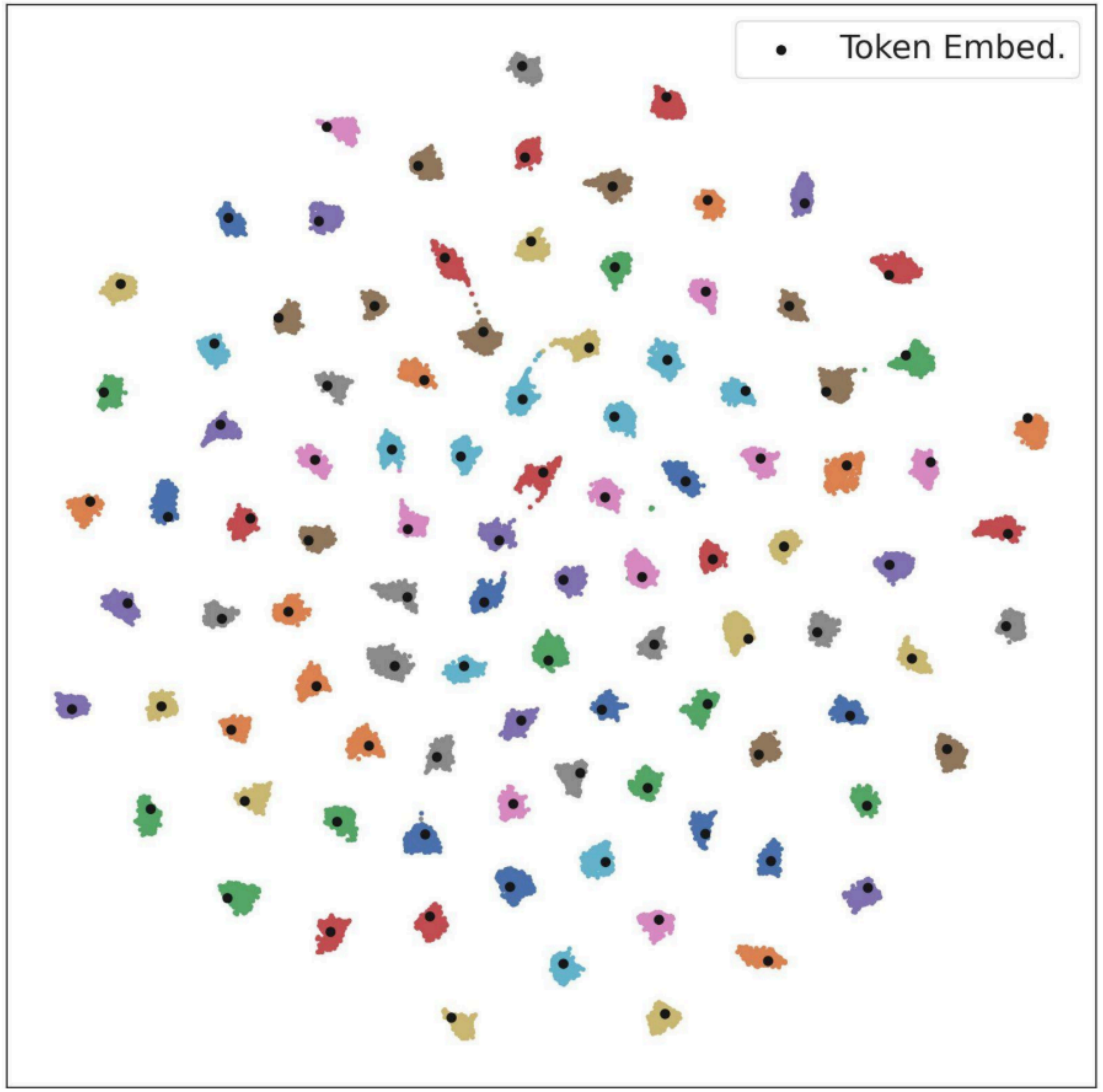
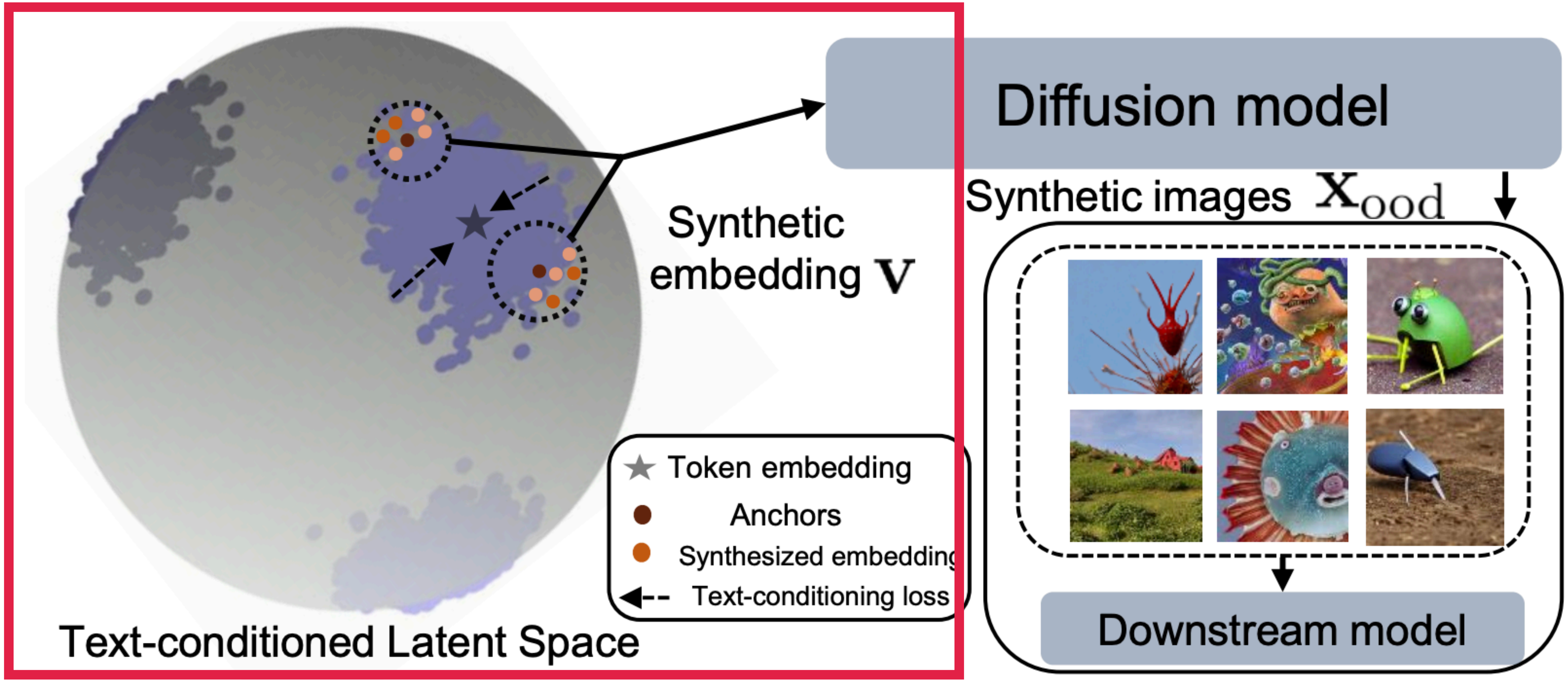


Figure 3: **TSNE visualization of learned feature embeddings using $\mathcal{L}$.** Black dots indicate token embeddings, one for each class.

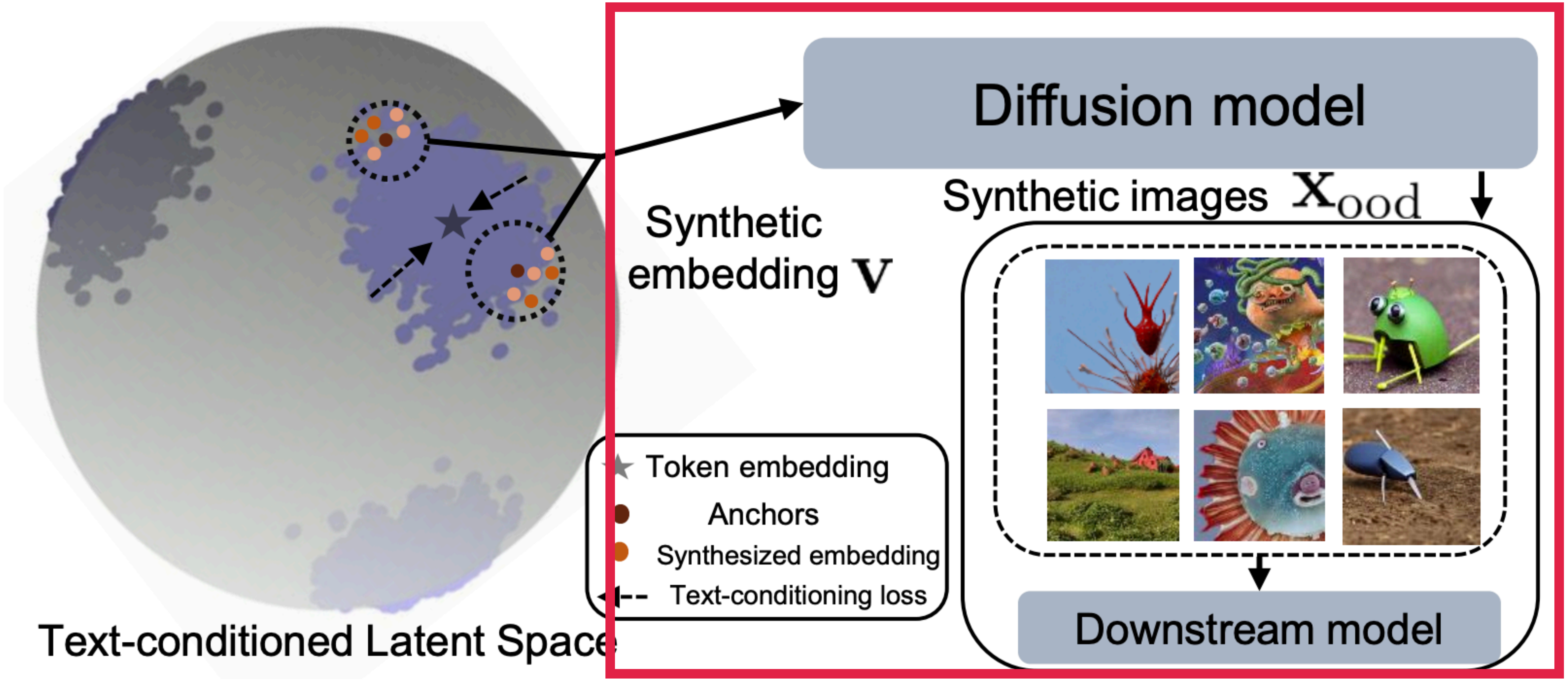
Dream-OOD: Outlier Imagination via Text-Conditioned Latent

1. *Sample OOD in the latent space*: draw new embeddings $\mathbf{v}$ that are in the low-likelihood region of the text-conditioned latent space.



Dream-OOD: Outlier Imagination via Text-Conditioned Latent

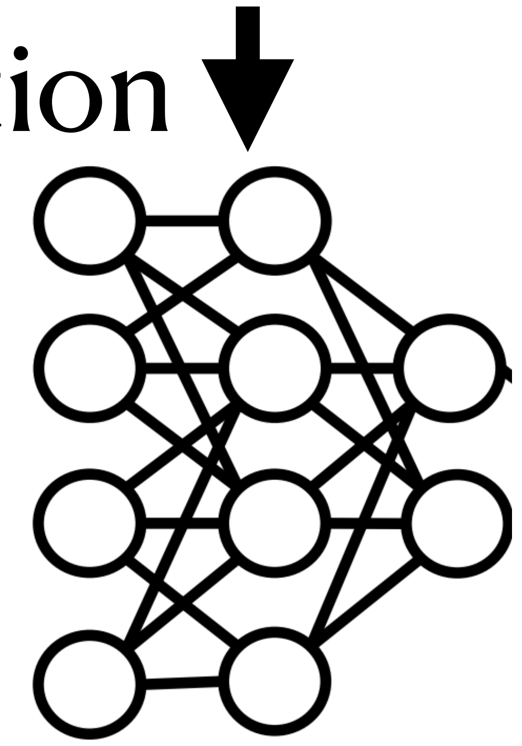
2. *Image generation:* decode $\mathbf{v}$ into a pixel-space OOD image via diffusion model.

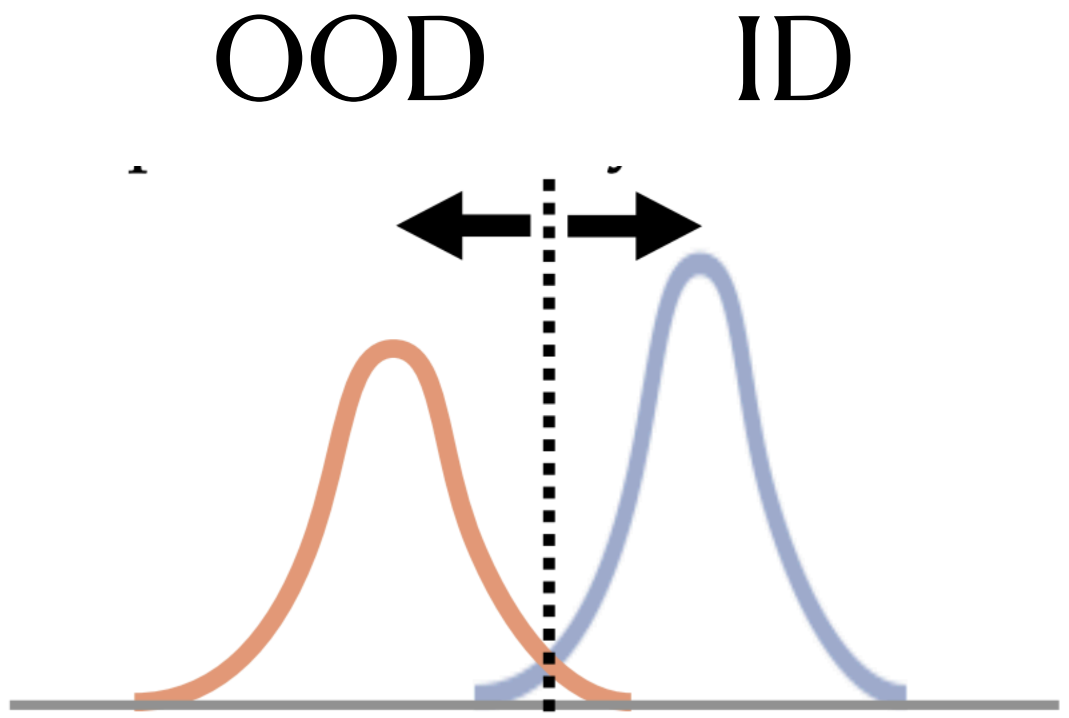


$$\mathbf{x}_{\text{ood}} \sim P(\mathbf{x}|\mathbf{v})$$

Dream-OOD: Learning with Imagined Outlier Images

Regularization





Level-set estimation loss [1]

$$\mathcal{L}_{\text{ood}} = \mathbb{E}_{\mathbf{x}_{\text{ood}}}\left[-\log \frac{1}{1+\exp^{\phi(E(f_\theta(\mathbf{x}_{\text{ood}})))}}\right] + \mathbb{E}_{\mathbf{x}\sim\mathbb{P}_{\text{in}}}\left[-\log \frac{\exp^{\phi(E(f_\theta(\mathbf{x})))}}{1+\exp^{\phi(E(f_\theta(\mathbf{x})))}}\right]$$

Experiments

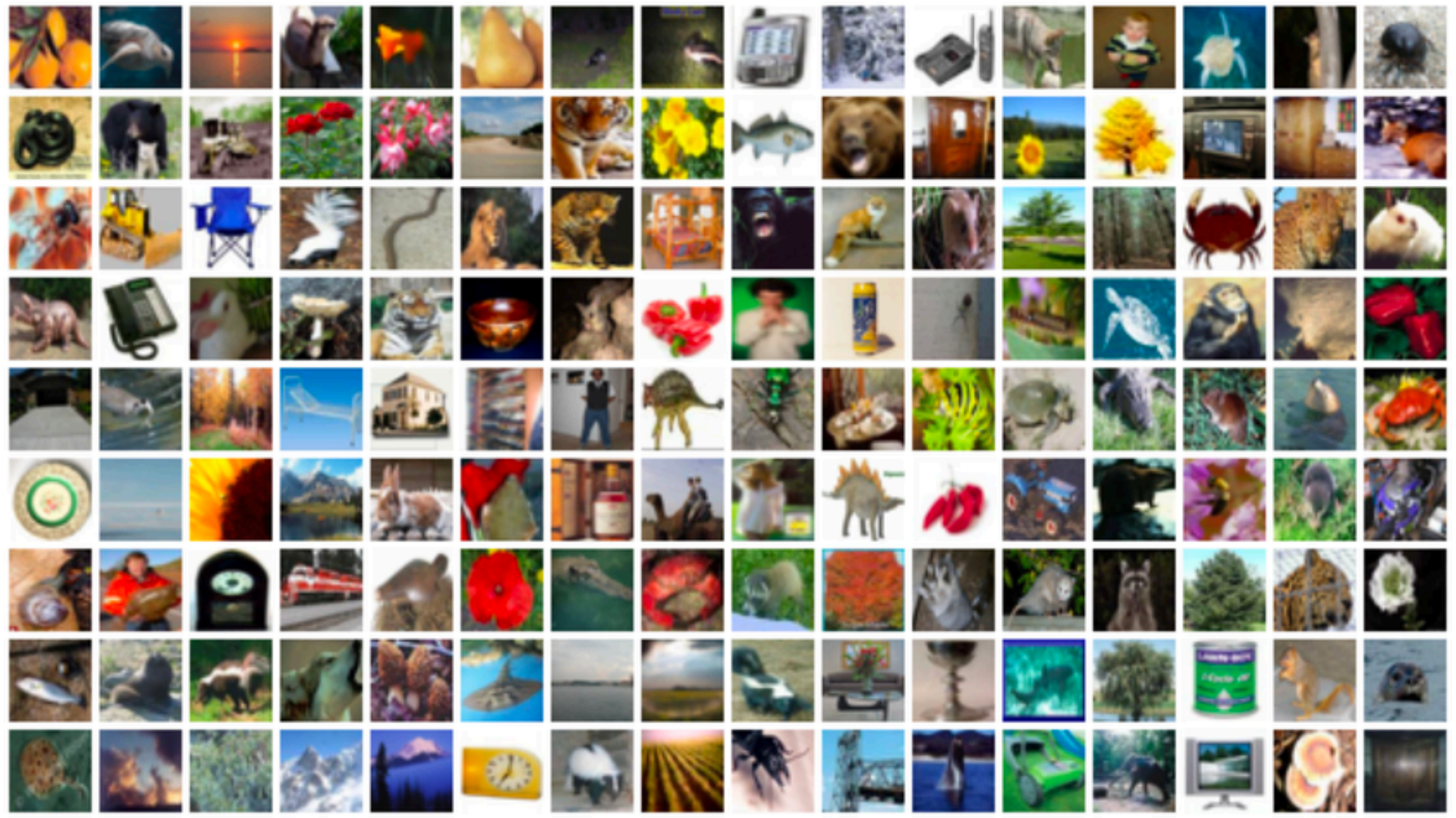
Dataset

In-distribution

OOD

ImageNet

Cifar-100



,etc.

Dream-OOD can Significantly Improve OOD Detection

Methods	iNaturalist		Places		Sun	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUR
MSP [33]	31.80	94.98	47.10	90.84	47.60	90.
ODIN [52]	24.40	95.92	50.30	90.20	44.90	91.
Mahalanobis [51]	91.60	75.16	96.70	60.87	97.40	62.
Energy [56]	32.50	94.82	50.80	90.76	47.60	91.
GODIN [40]	39.90	93.94	59.70	89.20	58.70	90.
KNN [96]	28.67	95.57	65.83	88.72	58.08	90.
ViM [103]	75.50	87.18	88.30	81.25	88.70	81.
ReAct [94]	22.40	96.05	45.10	92.28	37.90	93.
DICE [95]	37.30	92.51	53.80	87.75	45.60	89.
Synthesis-based methods						
GAN [50]	83.10	71.35	83.20	69.85	84.40	67.
VOS [18]	43.00	93.77	47.60	91.77	39.40	93.
NPOS [98]	53.84	86.52	59.66	83.50	53.54	87.
Dream-OOD (Ours)	**24.10**$\pm$0.2	**96.10**$\pm$0.1	**39.87**$\pm$0.1	**93.11**$\pm$0.3	**36.88**$\pm$0.4	**93.3**

(Table header spanning all dataset columns: OOD Dataset)

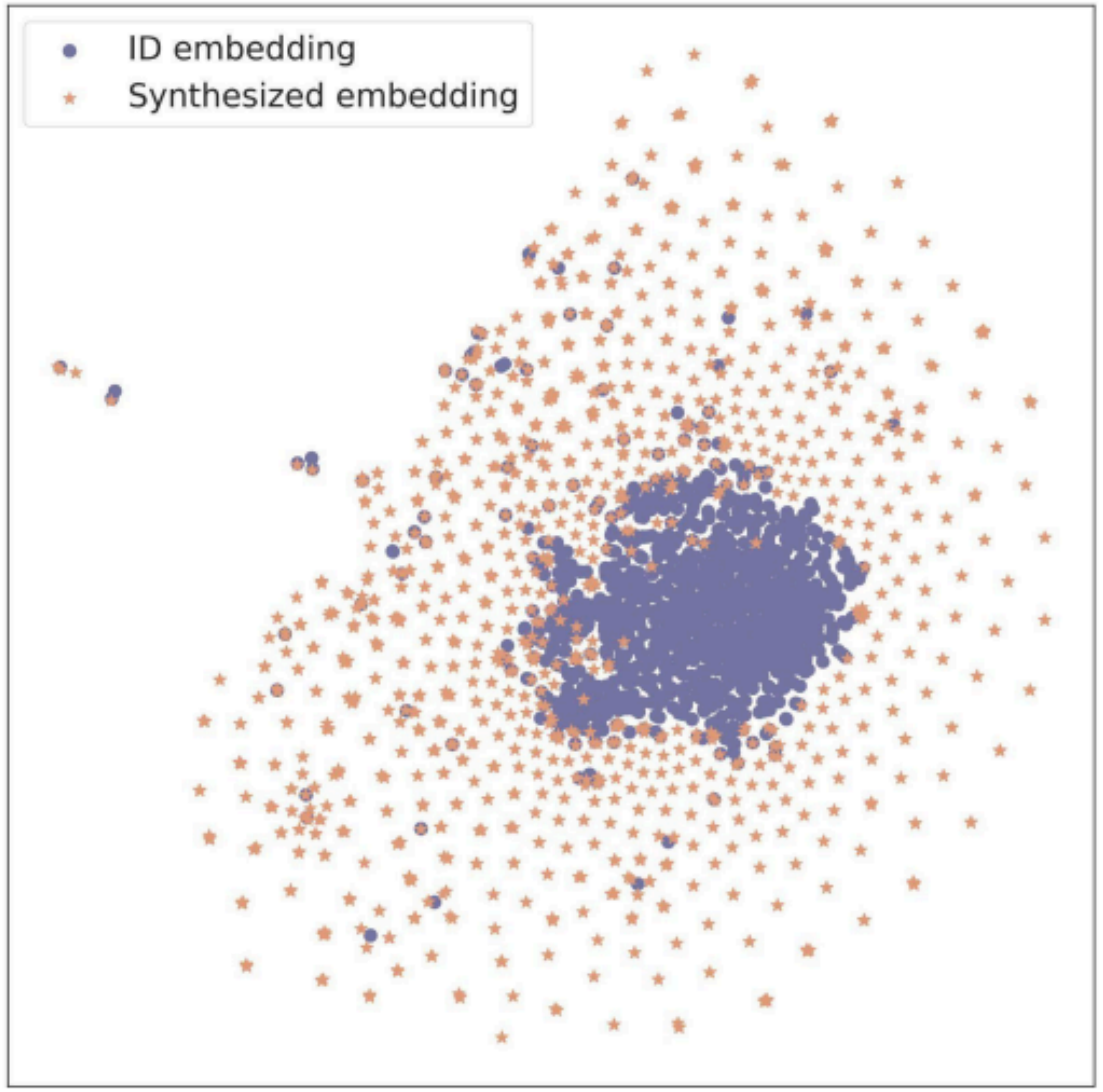


Figure 4: TSNE visualization of ID embeddings (purple) and the sampled outlier embeddings (orange), for the class "hen" in ImageNet.

Synthesized outlier embeddings (in orange) reside in the boundary of ID features!

Please check the paper for more results, including the improved model generalization.

Summary

- Machine learning models can make overconfident predictions on OOD data.
- Existing works are either costly in preparation or lacks interpretability.
- Dream-OOD mitigates the problem via diffusion models by
  ① Learning a text-conditioned latent space.
  ② Sampling outlier embeddings in the latent space.
  ③ Decoding the embeddings into outlier images with diffusion models.

GitHub
Paper: https://arxiv.org/pdf/2309.13415
Code: https://github.com/deeplearning-wisc/dream-ood