CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

Towards Realistic Evaluation of OOD-Generalization in DML

CVPR Tutorial: **Deep Visual Similarity and Metric Learning**

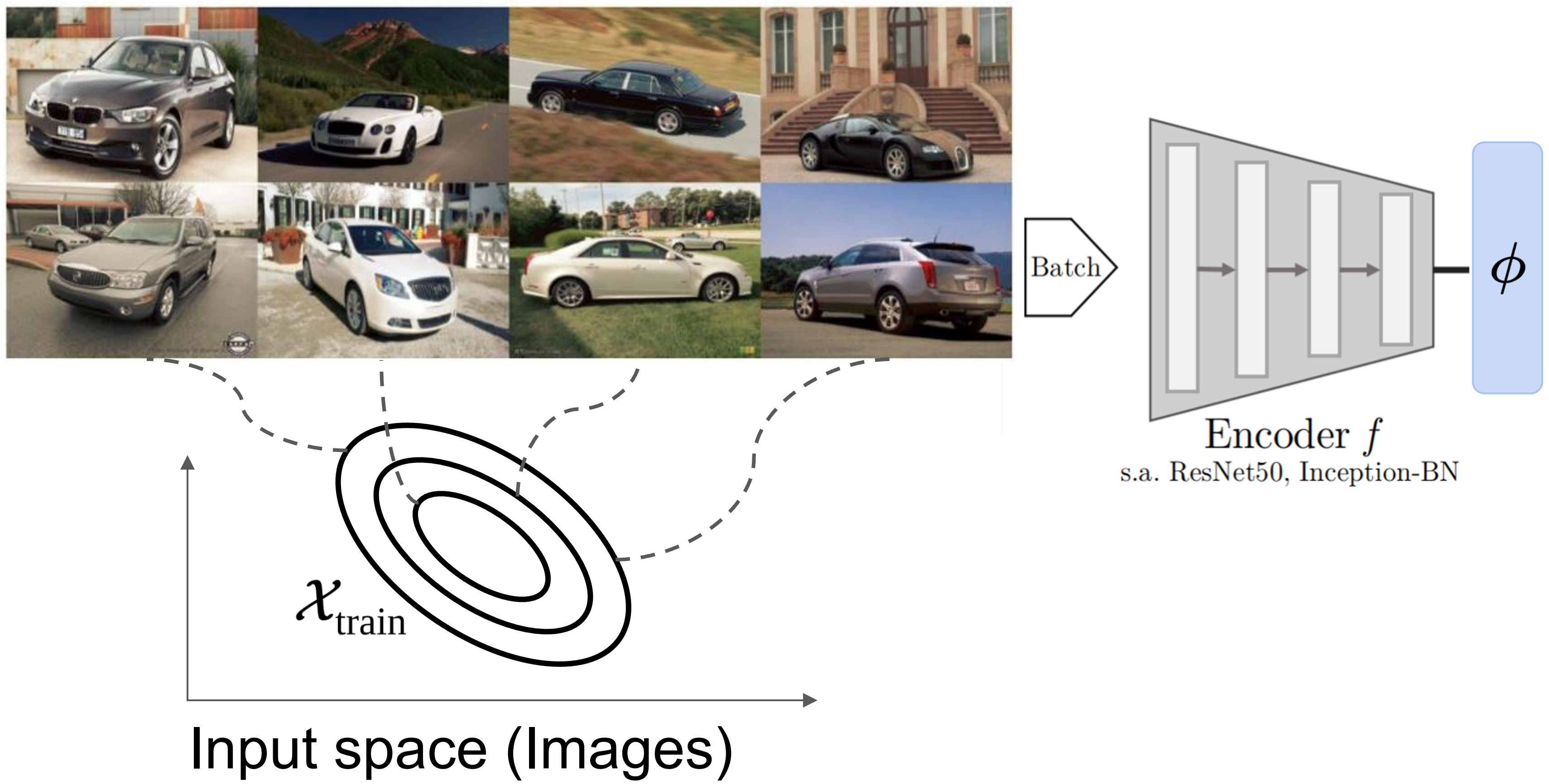
Timo Milbich

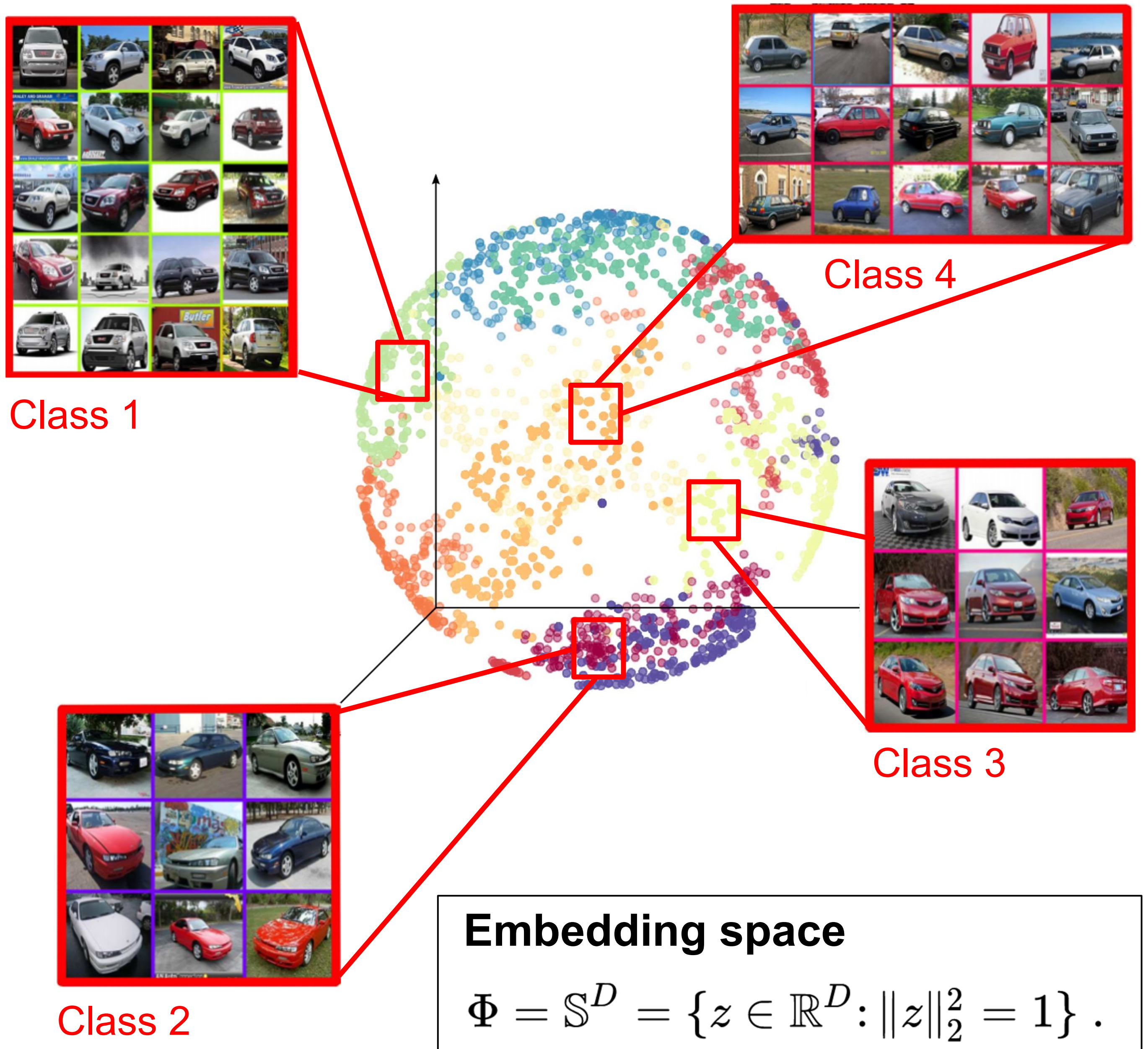
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Technical University of Munich TUM

Visual Similarity Learning

Learn **representation** $\phi(x)$ which reflects **semantic similarity** $d(\phi(x_i), \phi(x_j))$ within training distribution $\mathcal{X}_{\text{train}}$.



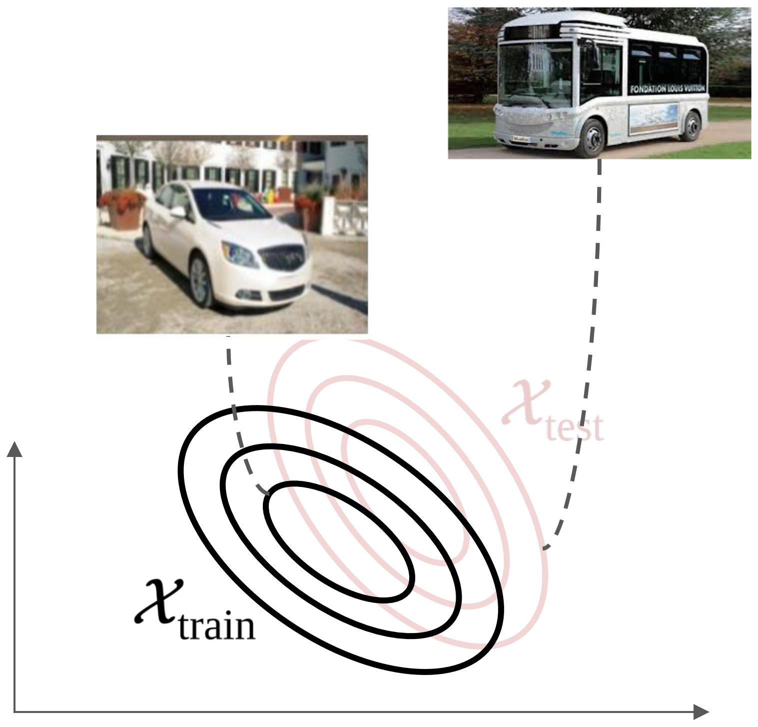


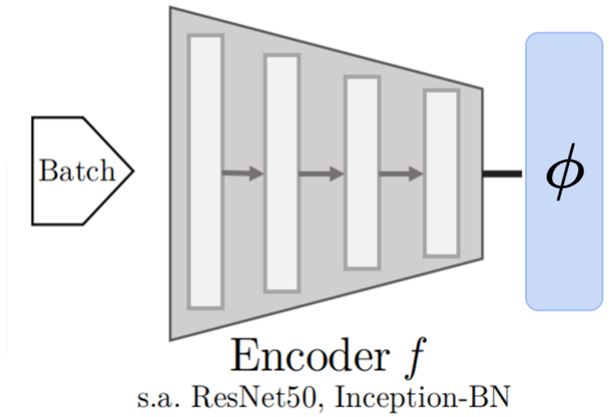
Embedding space

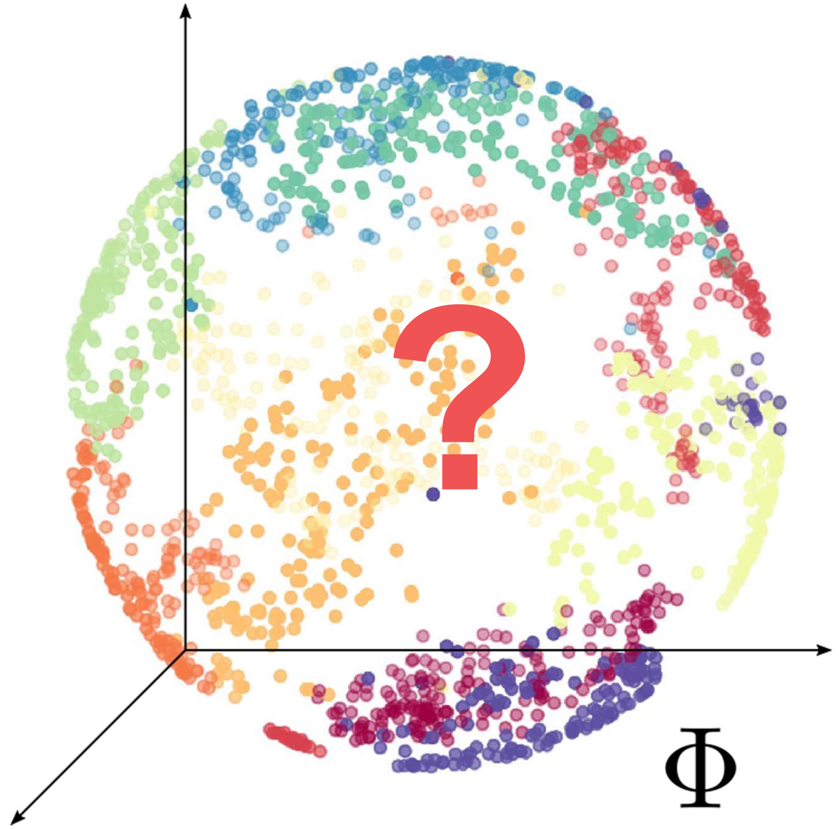
$$\Phi = \mathbb{S}^D = \{z \in \mathbb{R}^D \colon \|z\|_2^2 = 1\} \, .$$

Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual **class changes**?

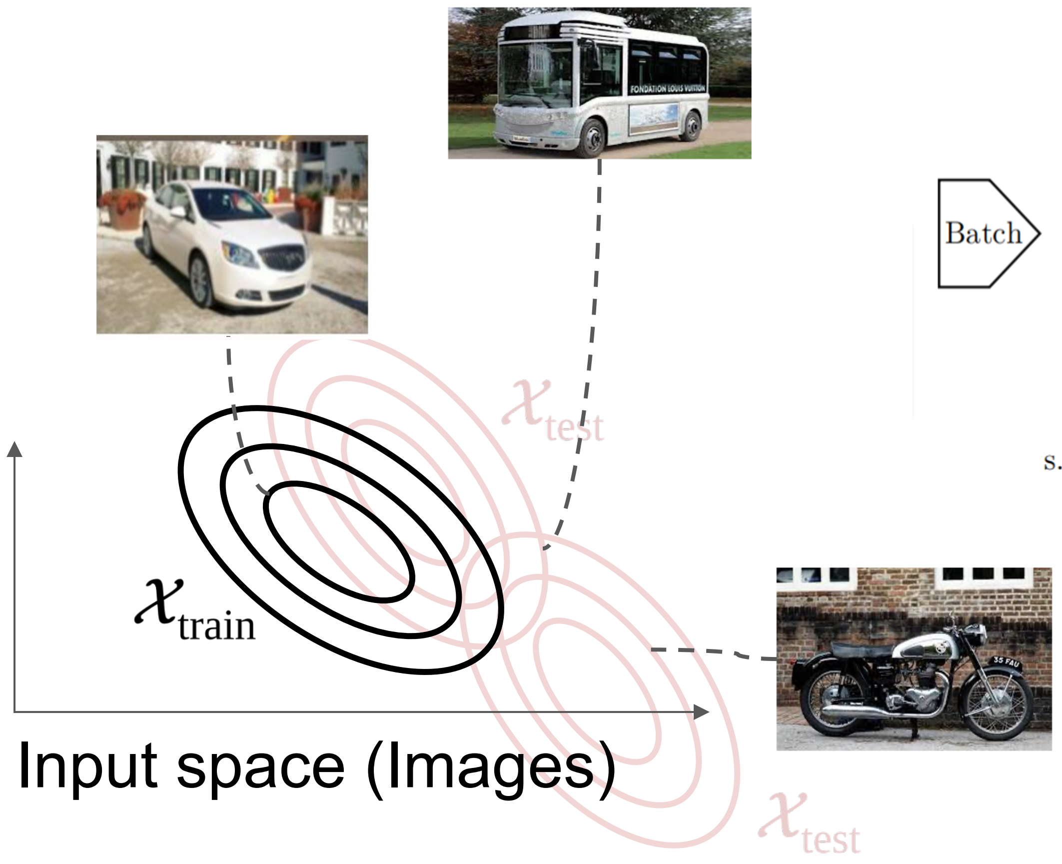


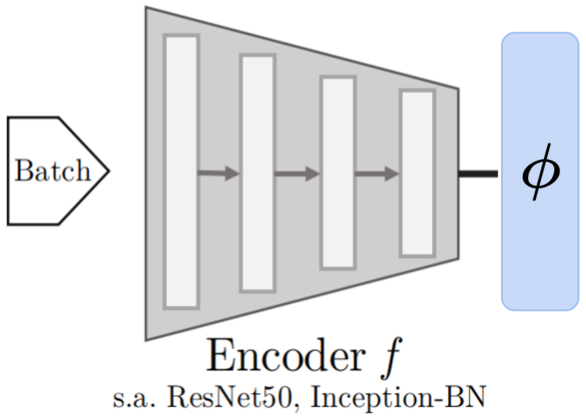


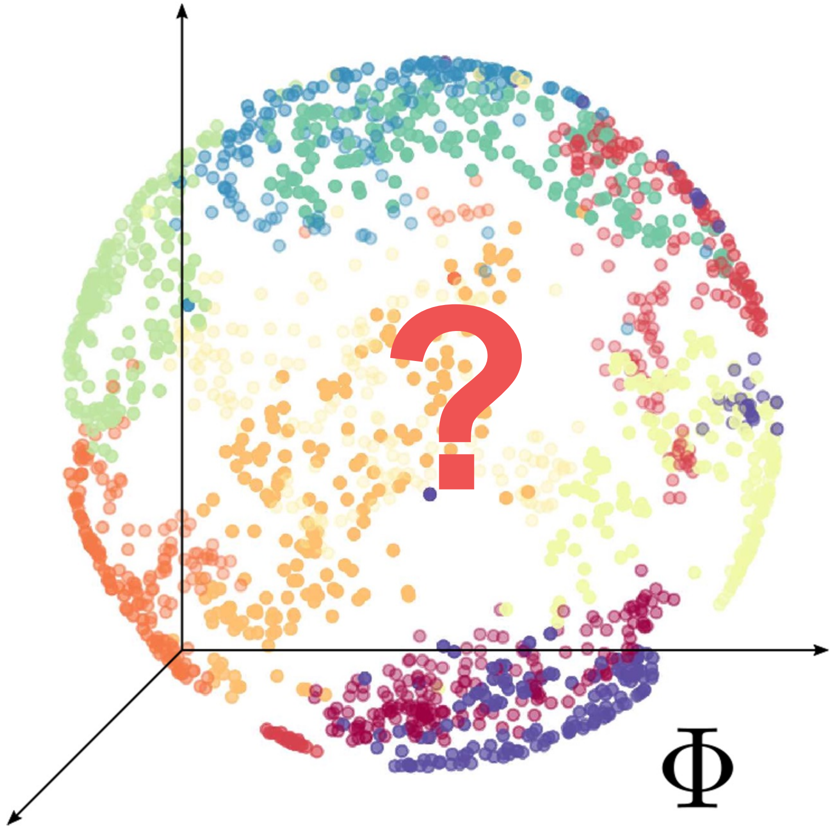


Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual **class changes**?

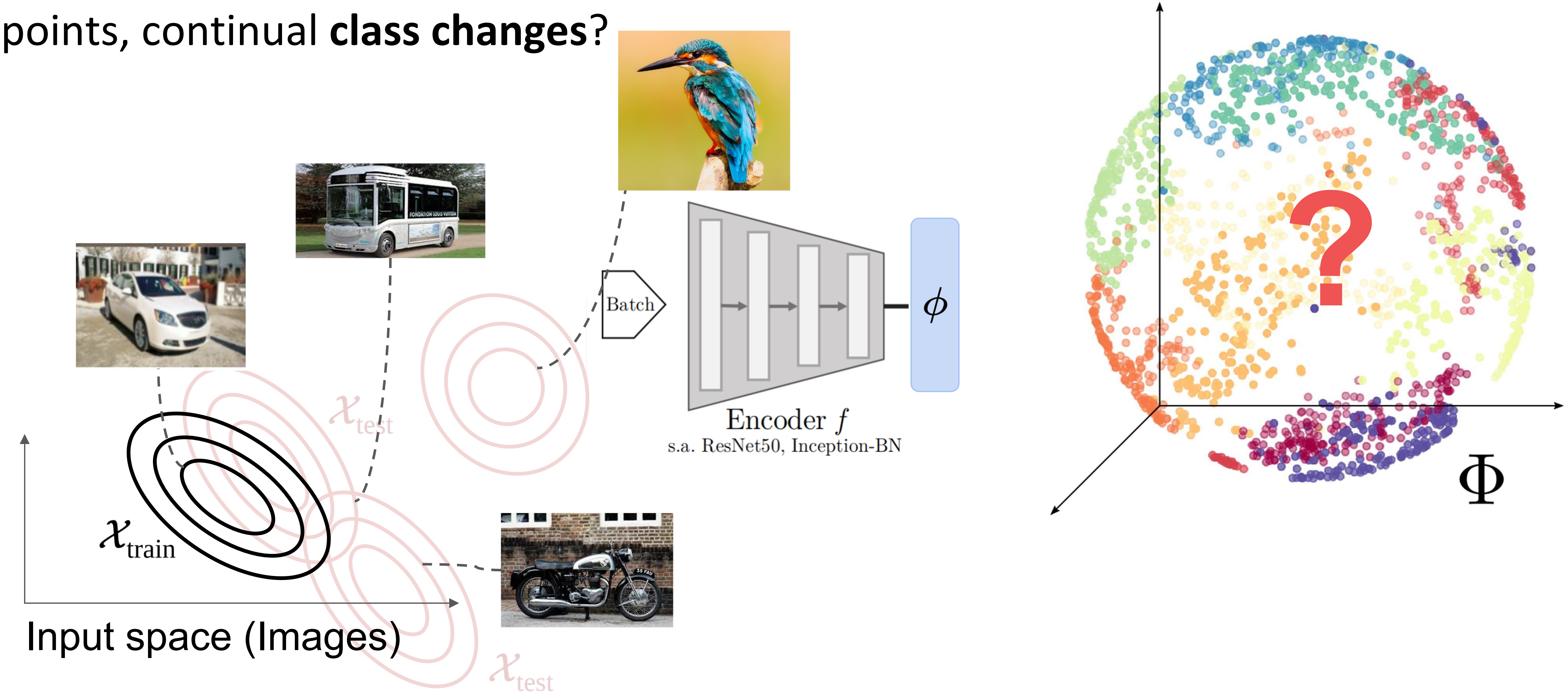






Visual Similarity Learning

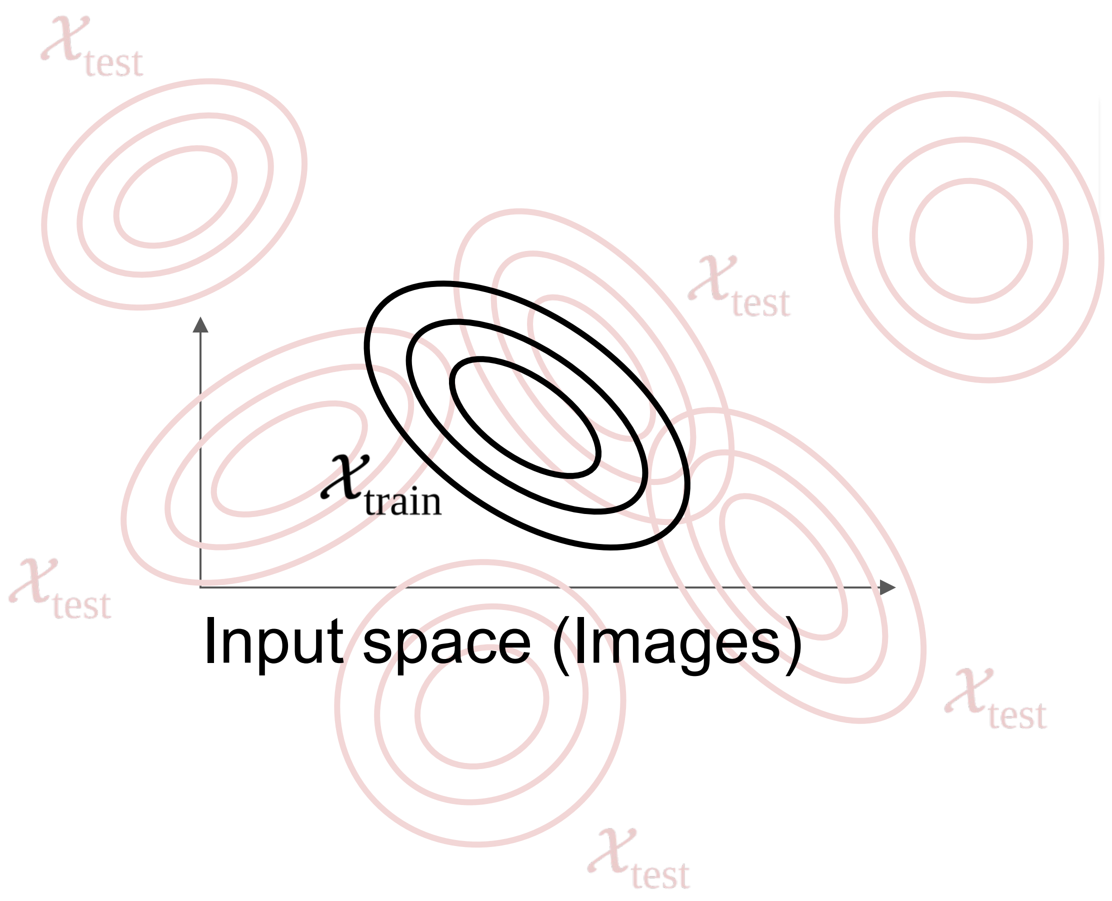
Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual **class changes**?

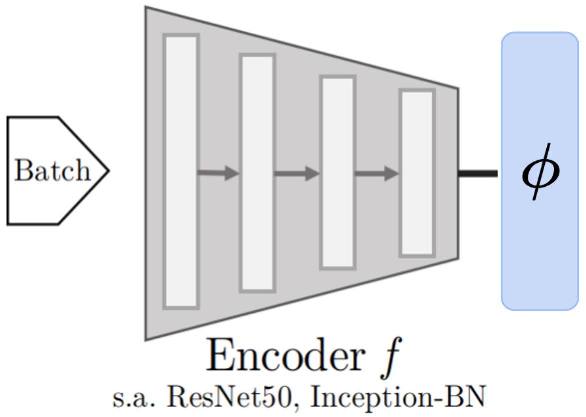


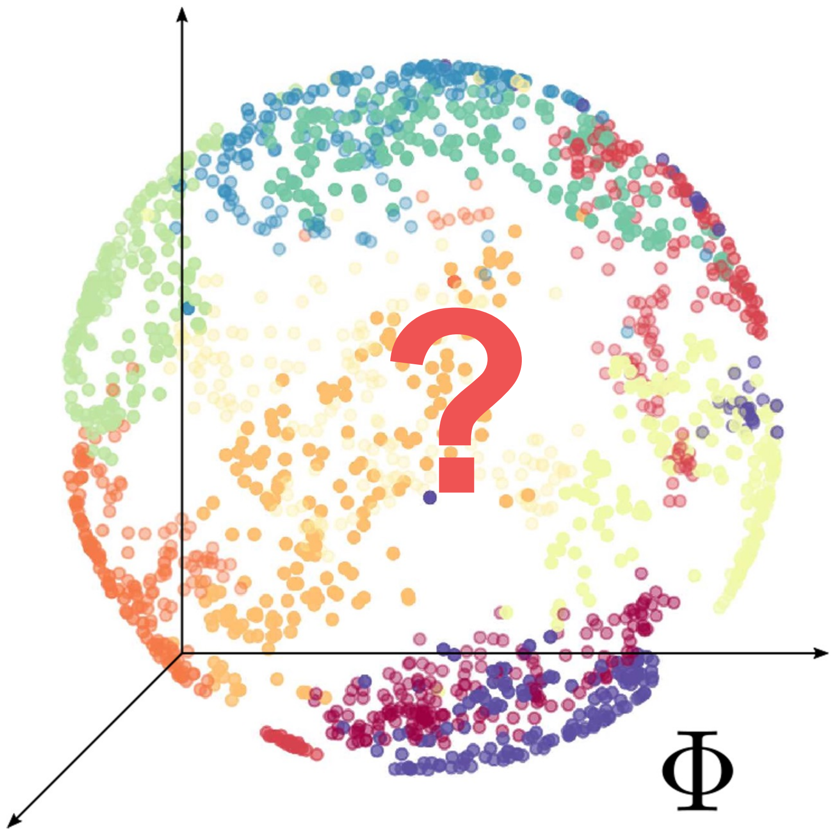
Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual **class changes**?

Evaluation needs to consider broad range of test distributions and difficulty.



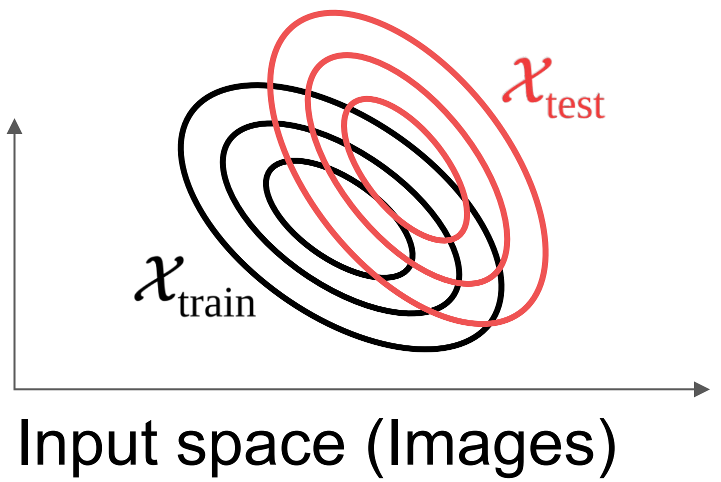


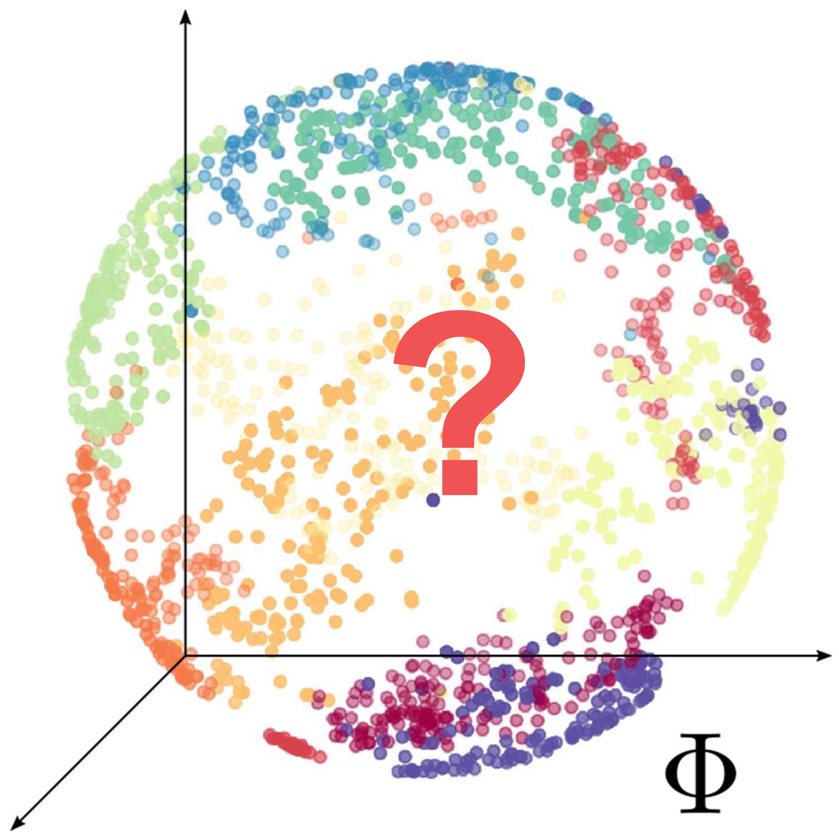


Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual class **changes**?

Evaluation needs to consider **broad range of test distributions** and **difficulty.**





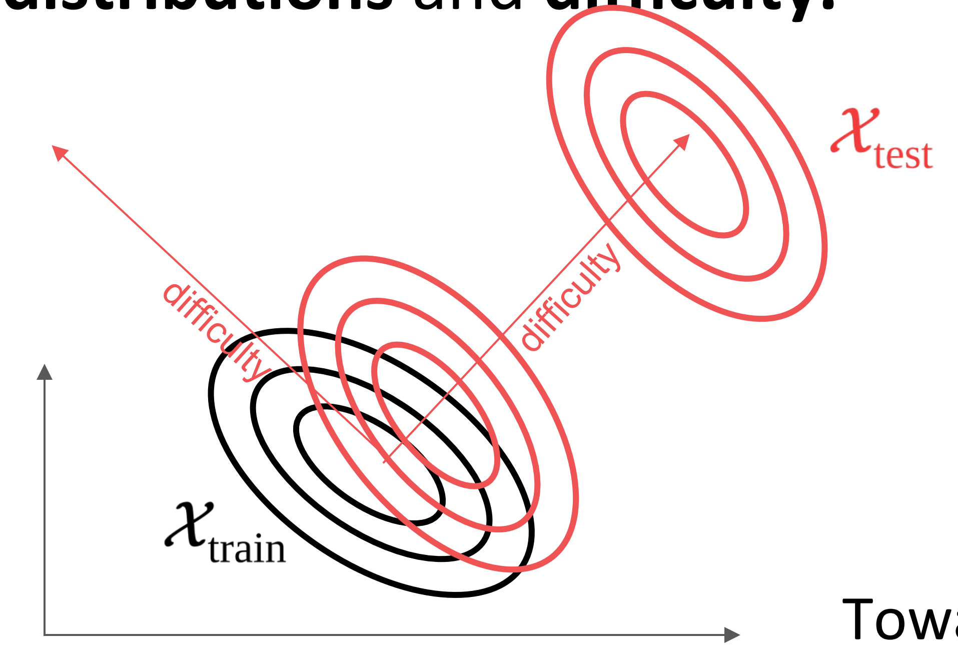
In practice:

- Only consider a **single** data split ($\mathcal{X}_{train}$, $\mathcal{X}_{test}$) for evaluation
- **random**, **fixed** problem difficulty
- **Hyperparameter overfitting**

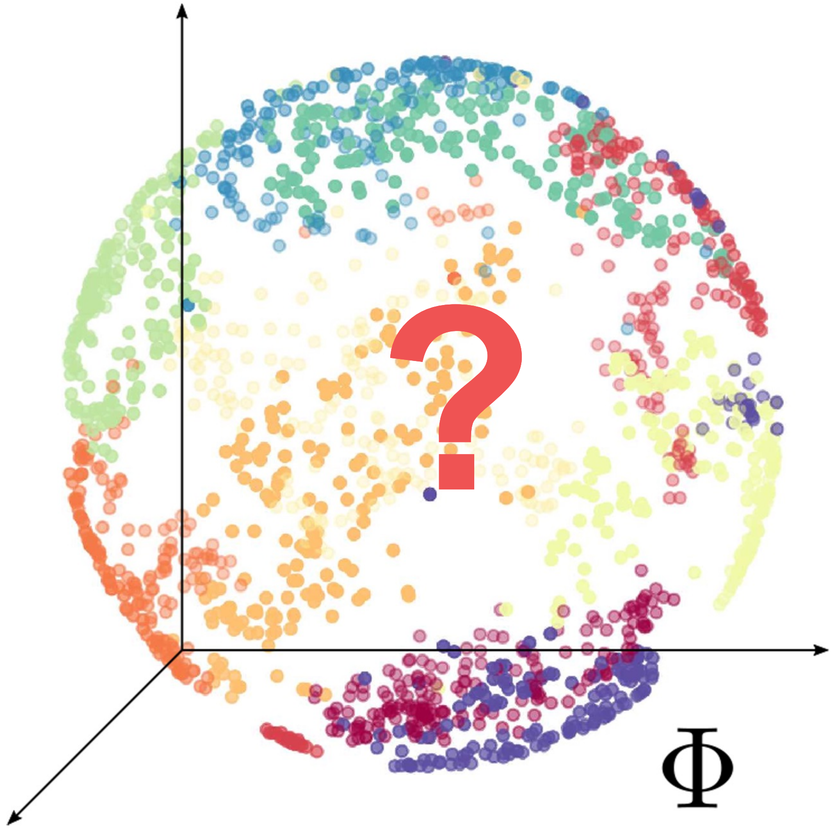
Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual class **changes**?

Evaluation needs to consider **broad range of test distributions** and **difficulty.**



Input space (Images)



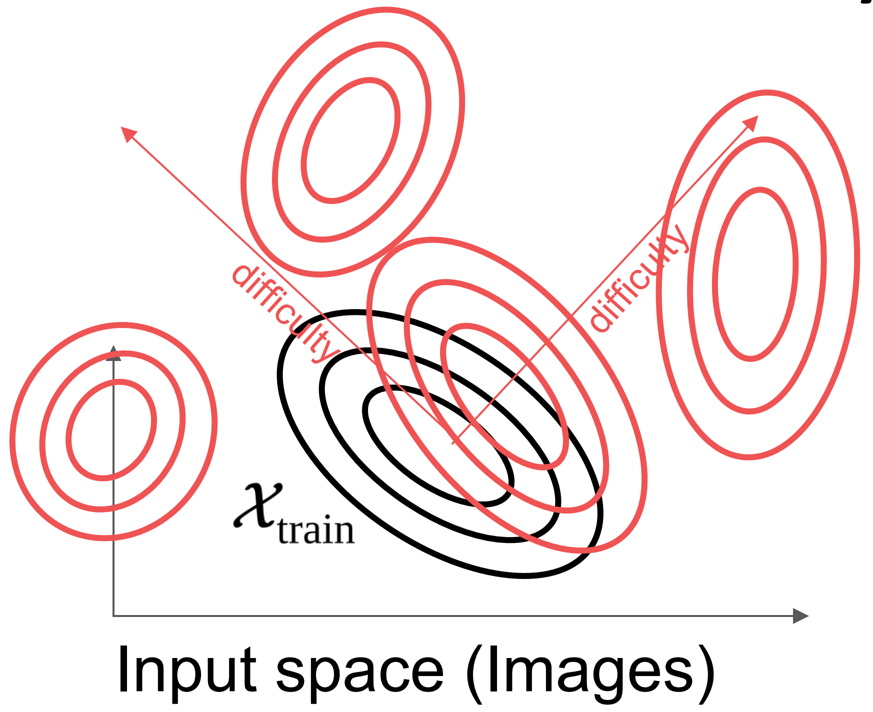
Towards **realistic evaluation** protocols for OOD Generalization:

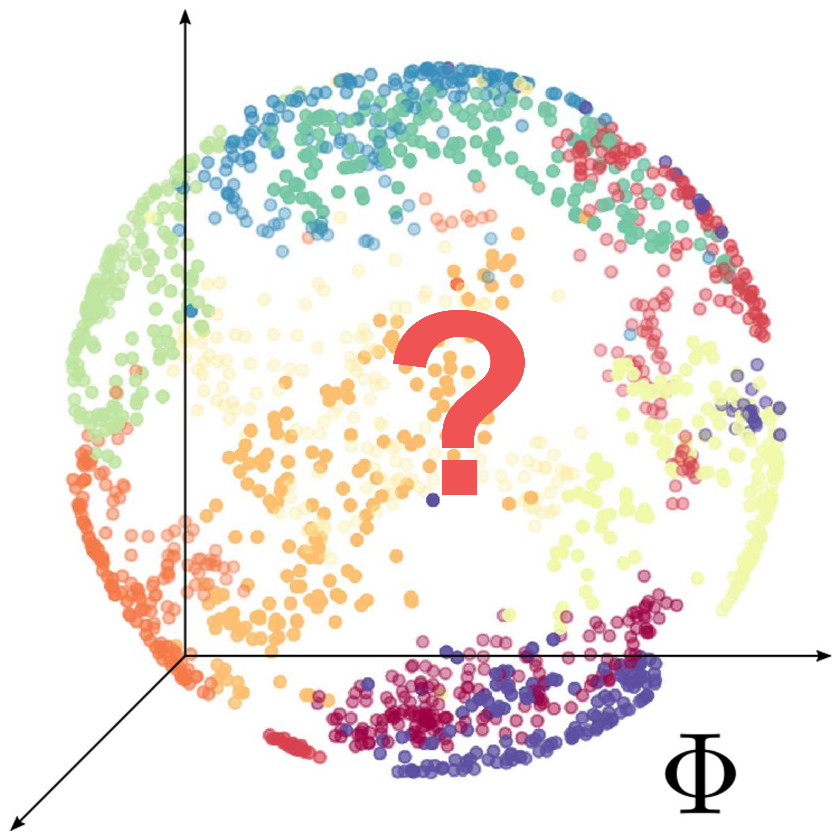
- **measure, change** and **control difficulty** of learning problems.

Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual class **changes**?

Evaluation needs to consider **broad range of test distributions** and **difficulty.**





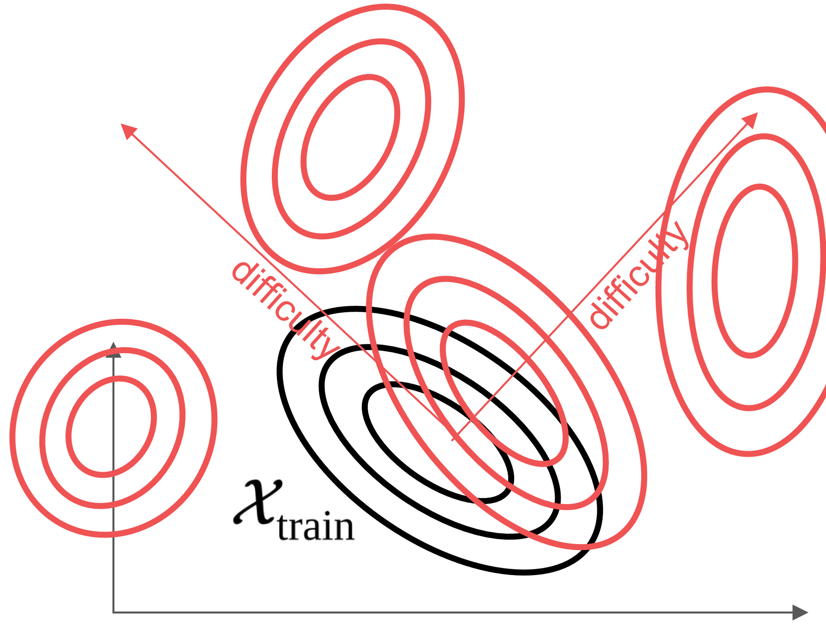
Towards **realistic evaluation protocols** for OOD Generalization:

- **measure, change** and **control difficulty** of learning problems.
- consider **multiple learning problems** of **different difficulties** (i.e. data splits).

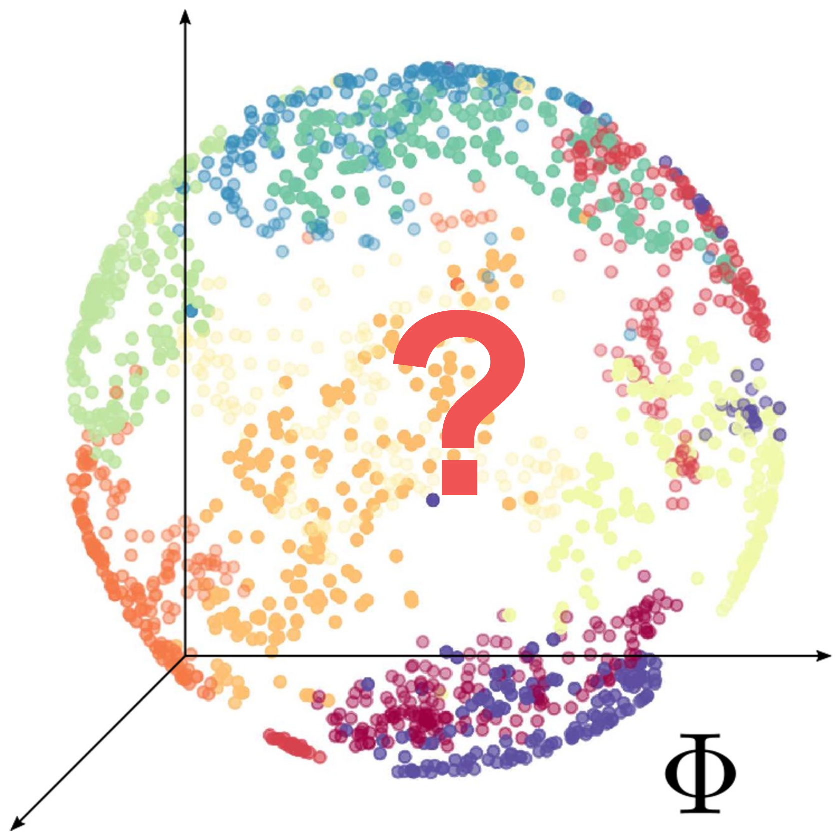
Visual Similarity Learning

Out-Of-Distribution-Generalization:
How well does Φ capture **unseen** classes, **unknown** surroundings, viewpoints, continual class **changes**?

Evaluation needs to consider **broad range of test distributions** and **difficulty.**



Input space
(Cars196, Cub200, SOP,...)



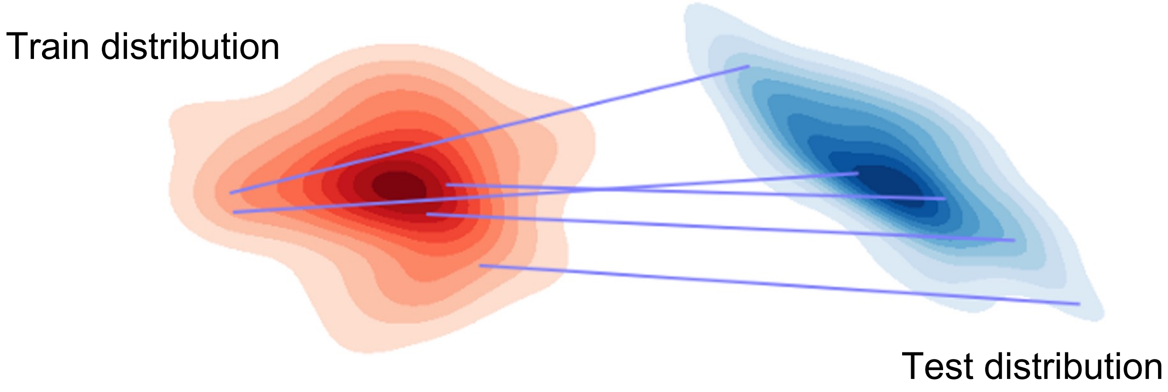
Towards **realistic evaluation protocols** for OOD Generalization**:**

- **measure, change** and **control difficulty** of learning problems.
- consider **multiple learning problems** of **different difficulties** (i.e. data splits).
- built **individually for datasets** for better comparison between data splits.

ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

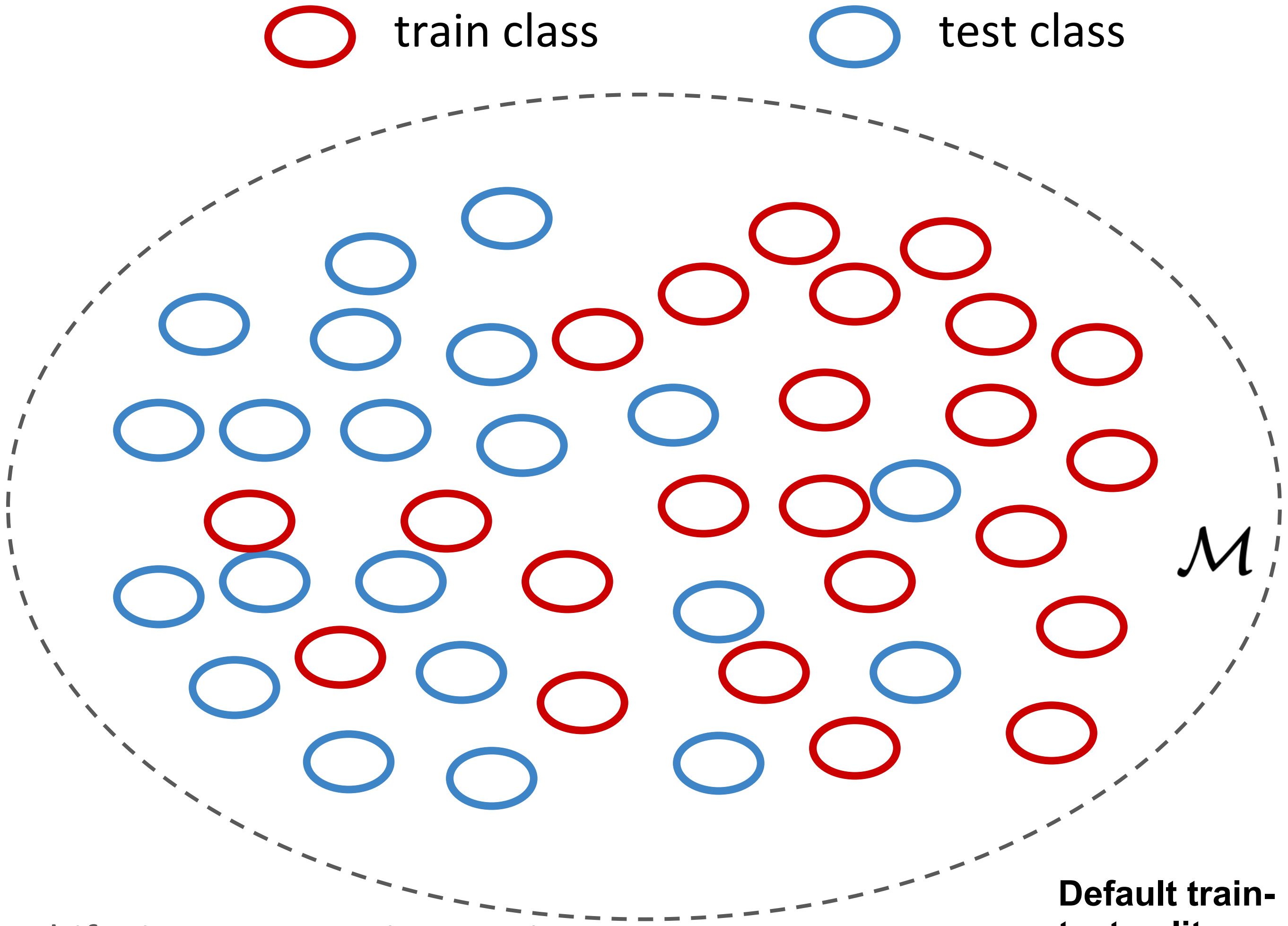


Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.

Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

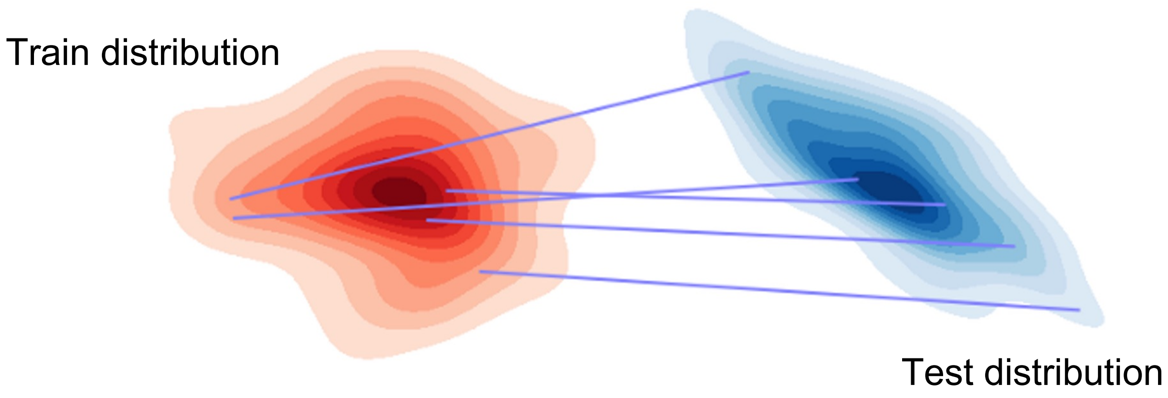
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

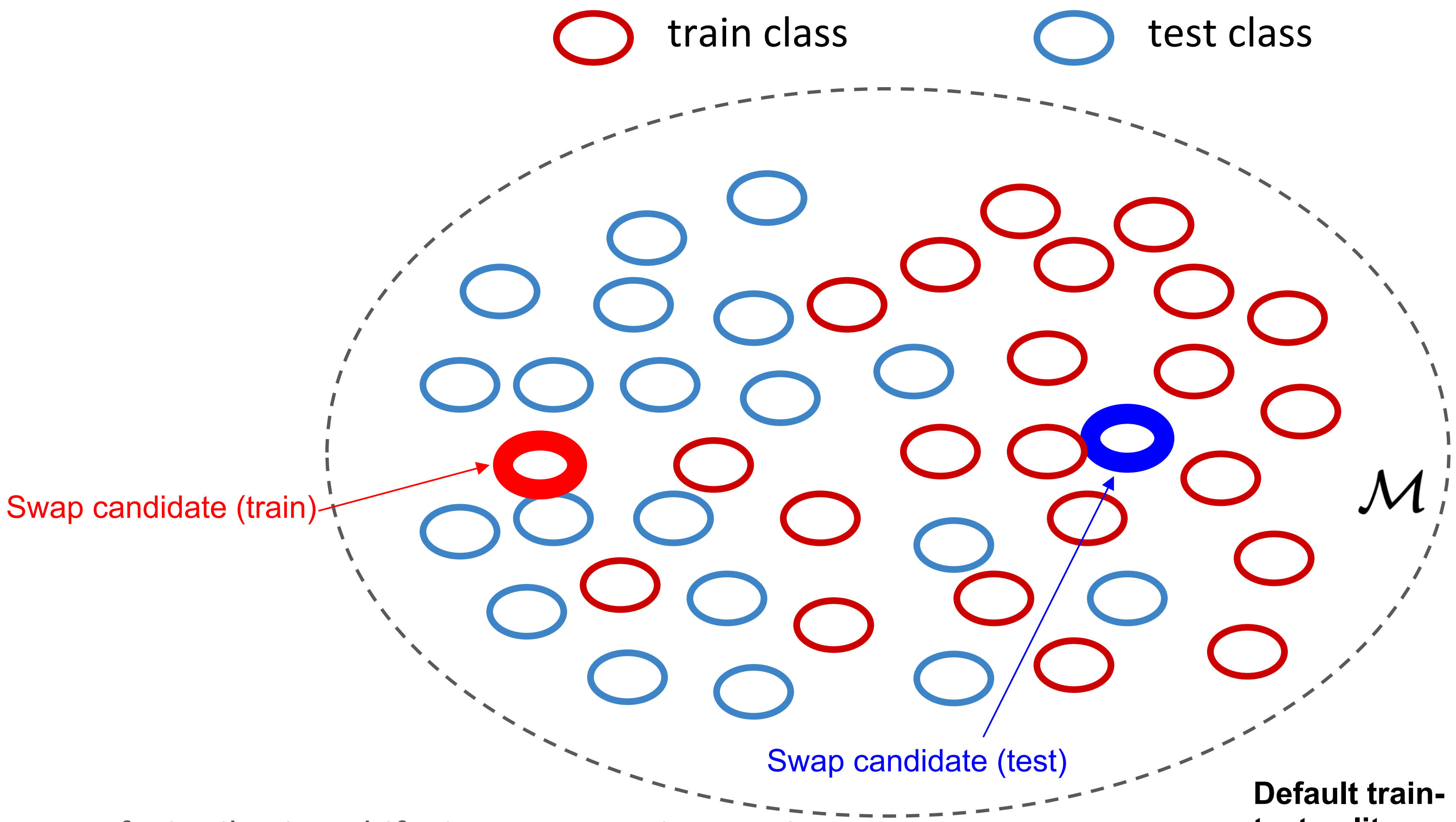
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



Default train-test split

[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

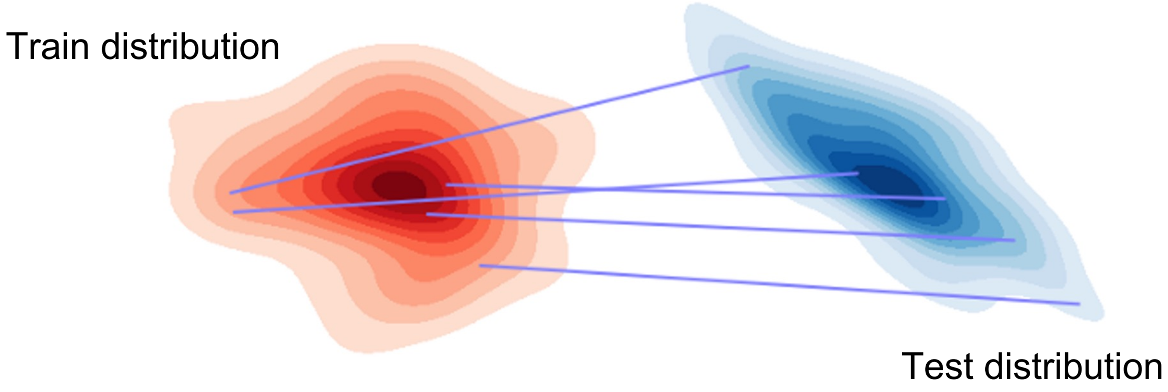
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

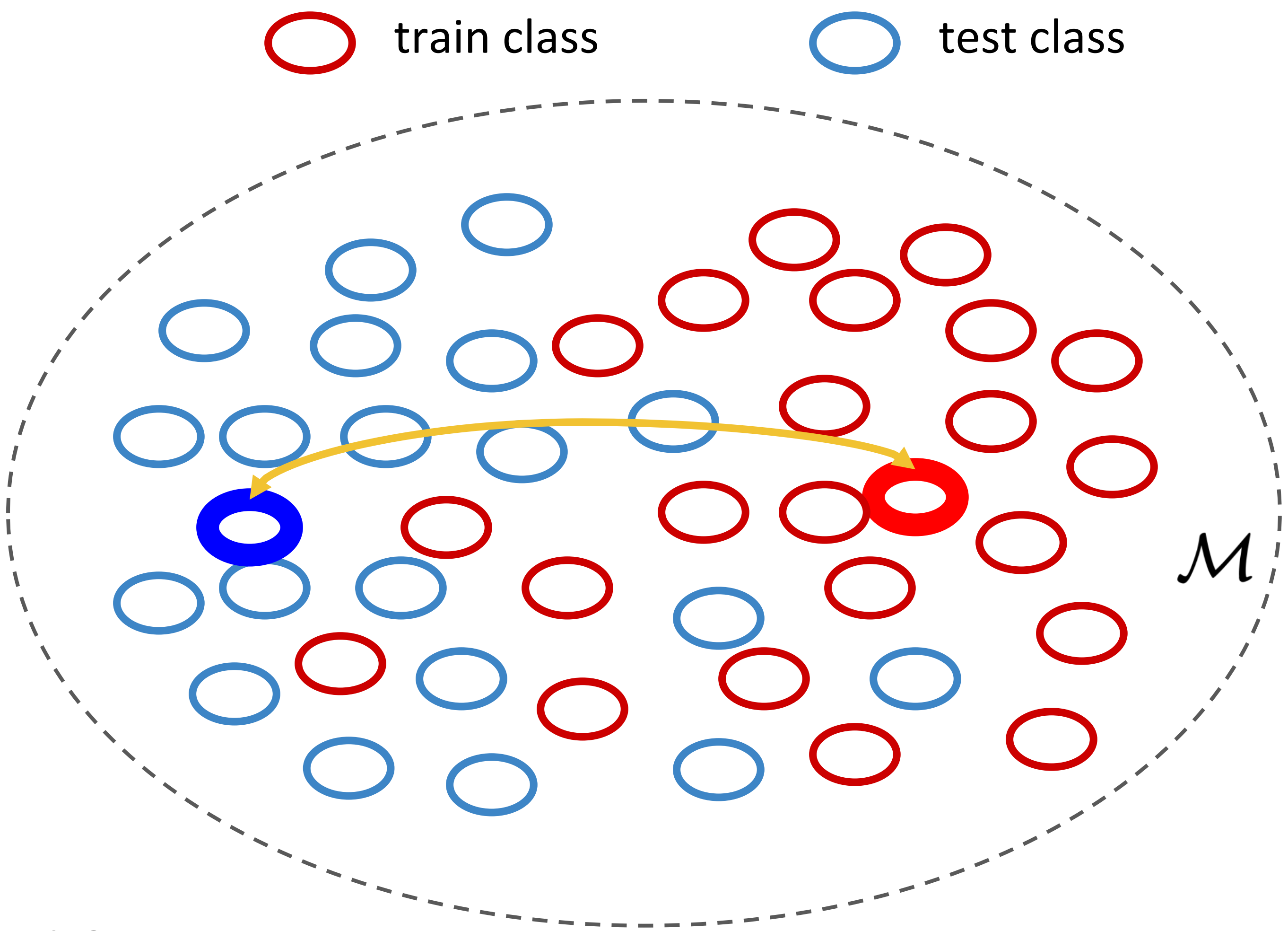
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

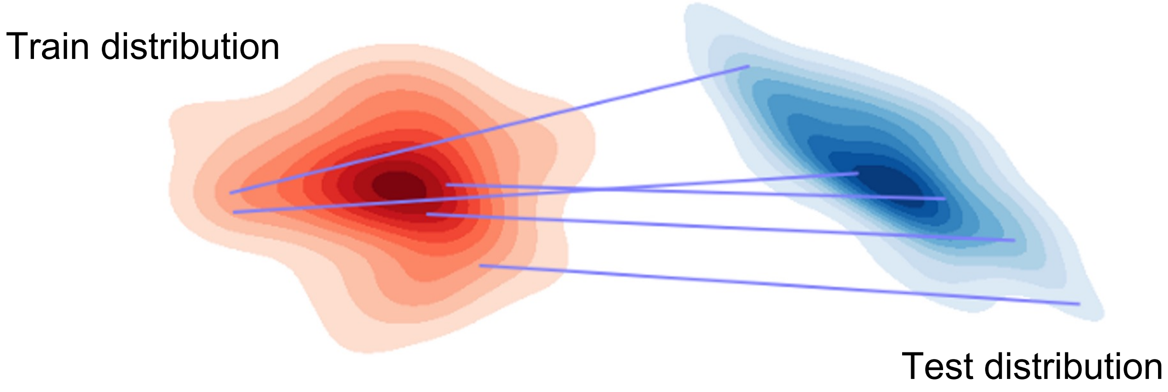
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

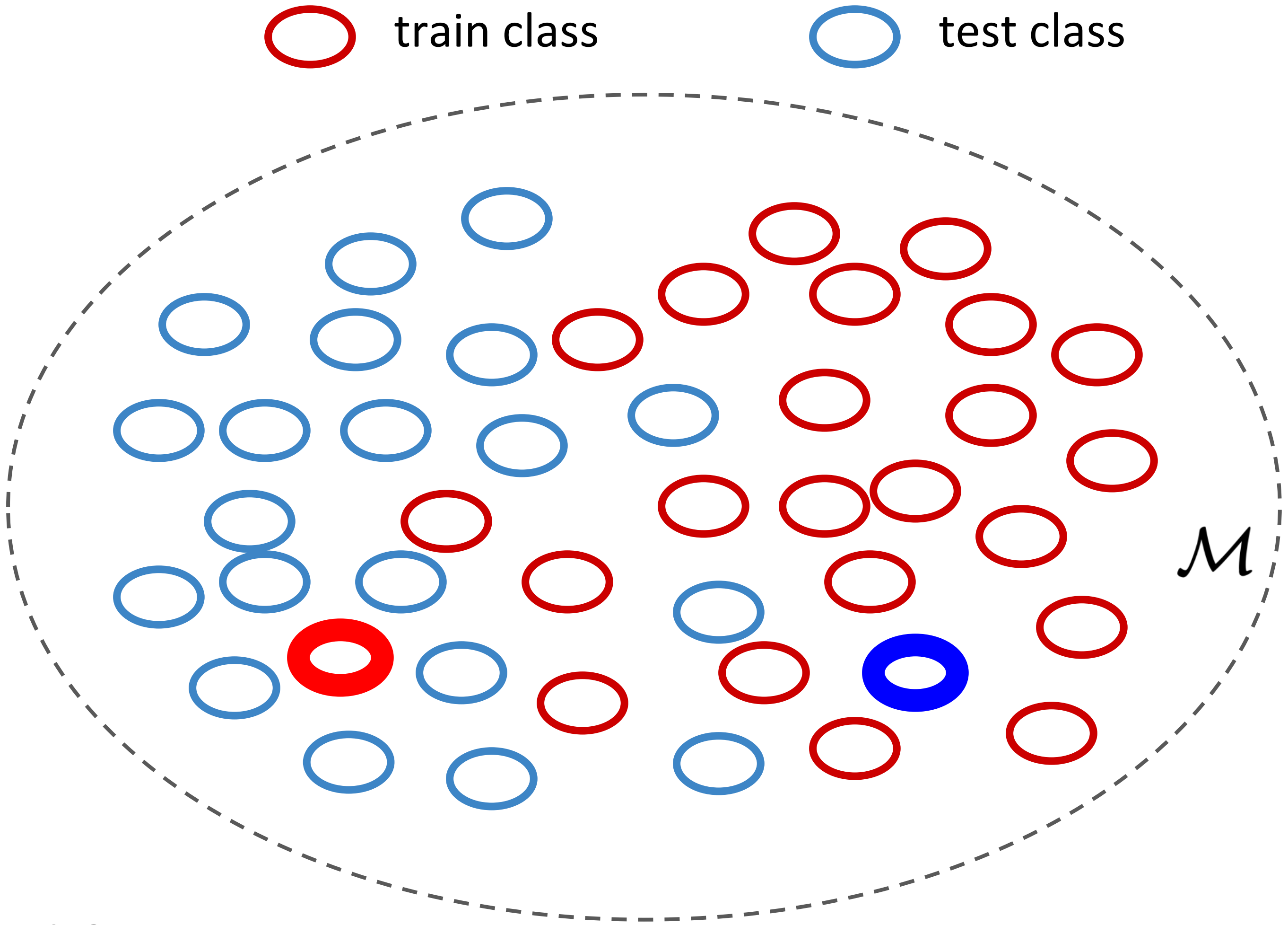
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

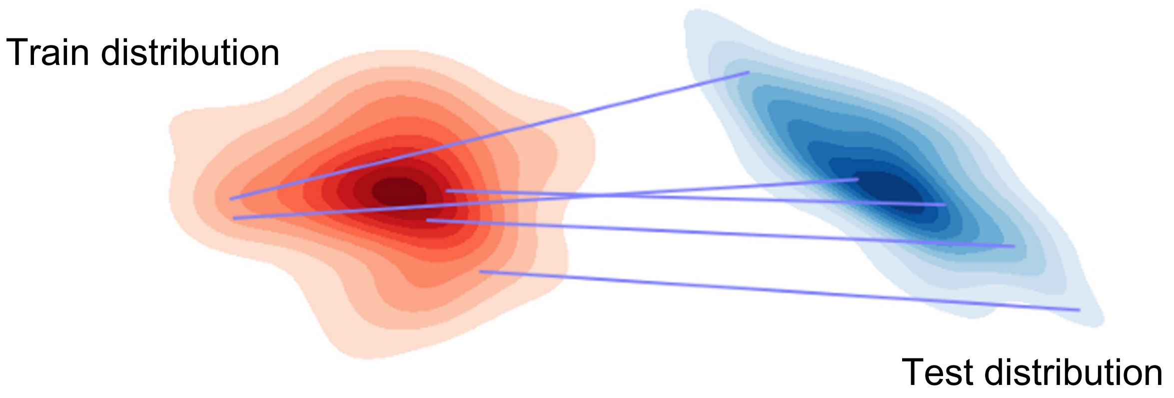
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

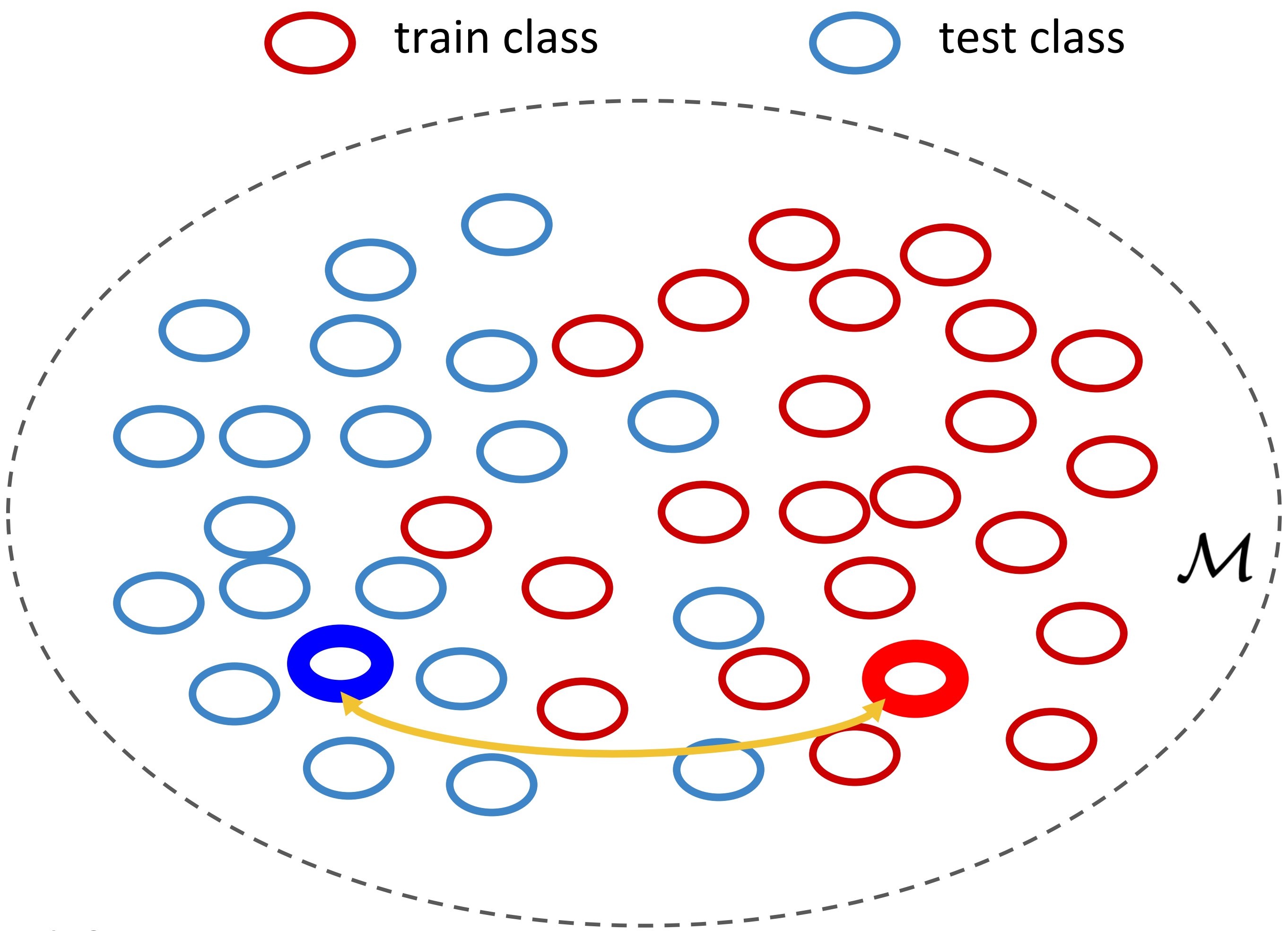
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

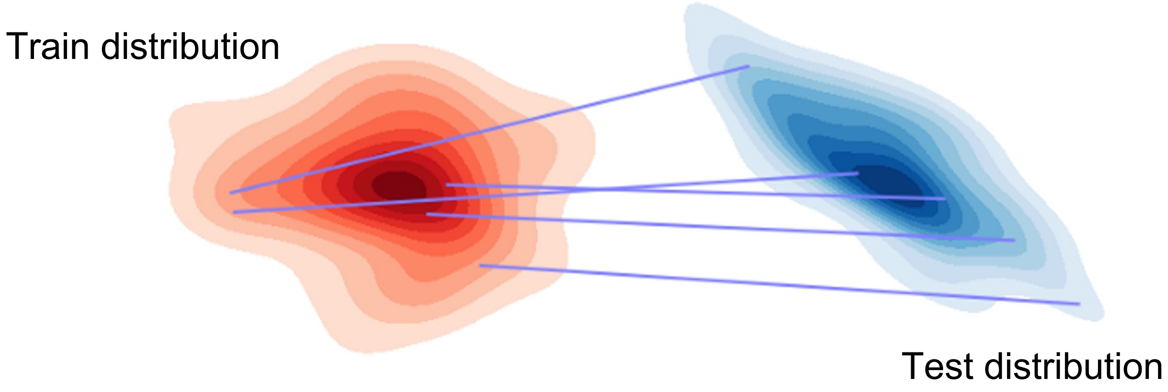
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

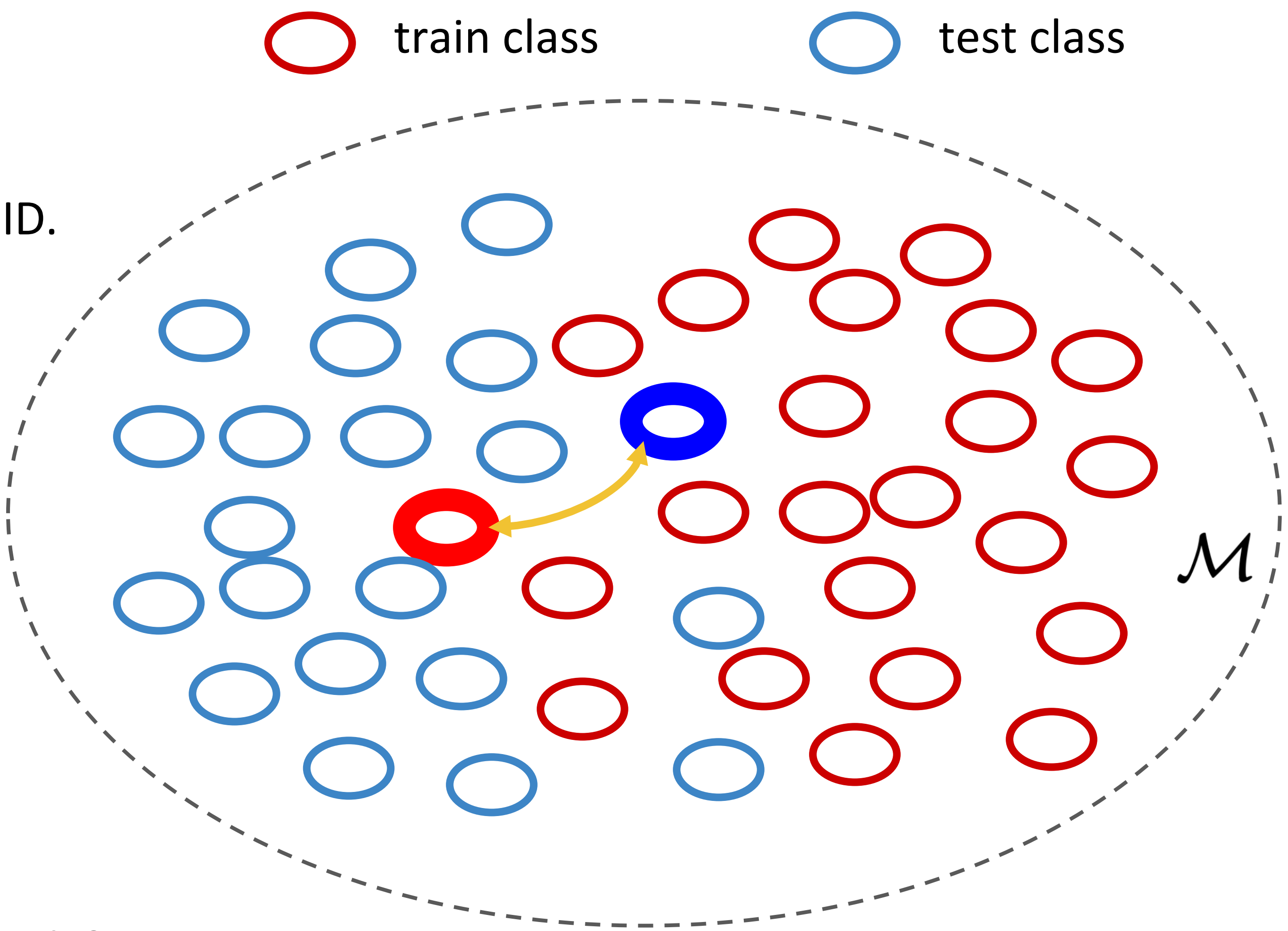
(1) $\mathcal{X}_{\text{train}}$ default training data, $\mathcal{X}_{\text{test}}$ default test data

(2) Identify classes $C_{\text{train}} \subset \mathcal{X}_{\text{train}}$ and $C_{\text{test}} \subset \mathcal{X}_{\text{test}}$ whose exchange will increase FID.

$$C^*_{\text{train}} = \underset{C_{\text{train}} \in \mathcal{X}_{\text{train}}}{\arg\max} \|\mu_{C_{\text{train}}} - \mu_{\mathcal{X}_{\text{train}}}\|_2 - \|\mu_{C_{\text{train}}} - \mu_{\mathcal{X}_{\text{test}}}\|_2$$

$$C^*_{\text{test}} = \underset{C_{\text{test}} \in \mathcal{X}_{\text{test}}}{\arg\max} \|\mu_{C_{\text{test}}} - \mu_{\mathcal{X}_{\text{test}}}\|_2 - \|\mu_{C_{\text{test}}} - \mu_{\mathcal{X}_{\text{train}}}\|_2$$

(1) Swap classes C_{train} and C_{test} between $\mathcal{X}_{\text{train}}$ and $\mathcal{X}_{\text{test}}$.



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

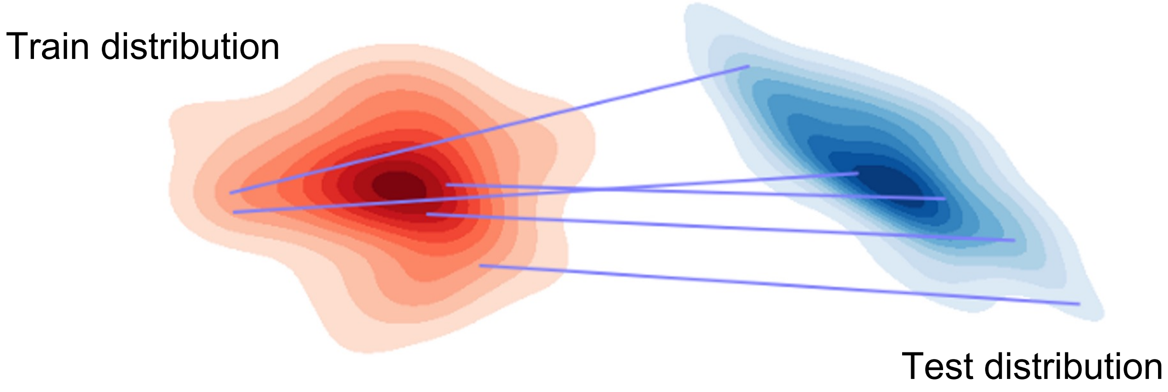
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

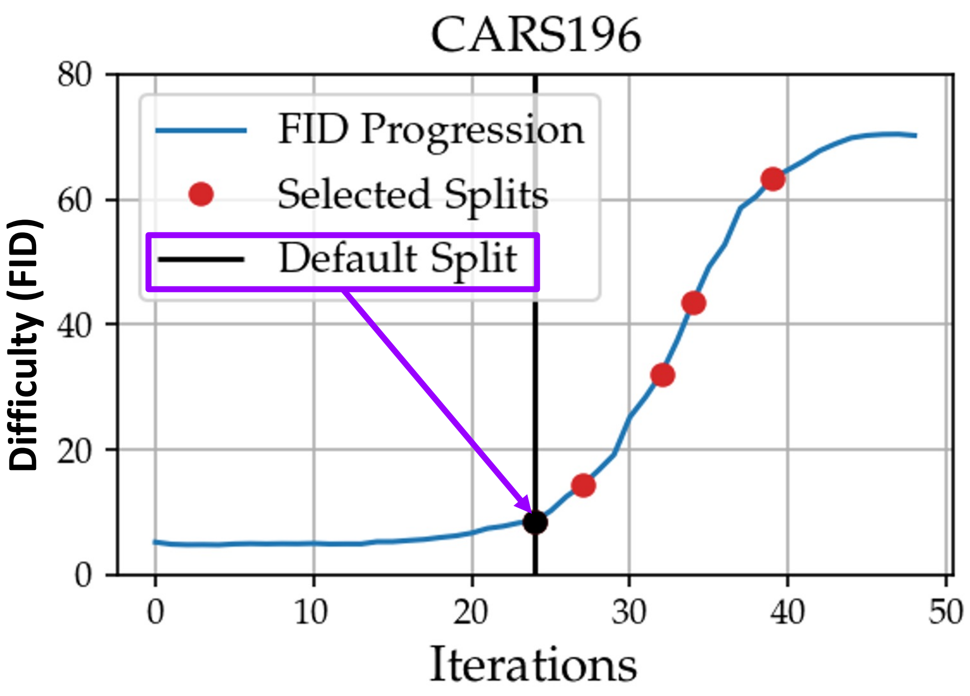
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

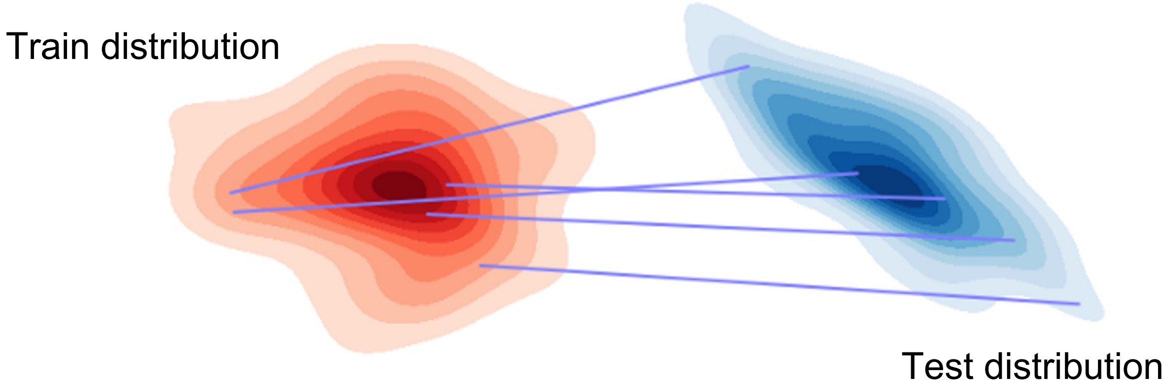
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

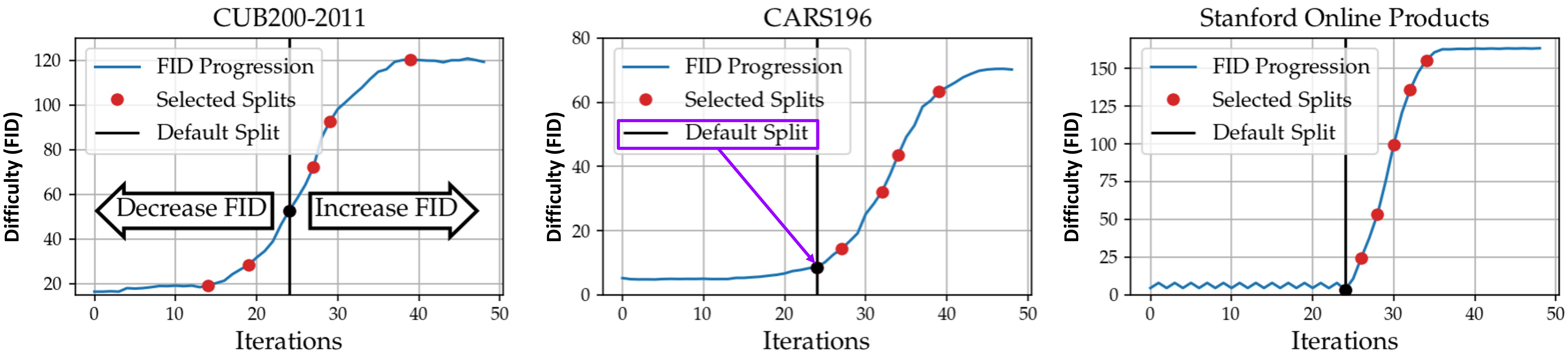
$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

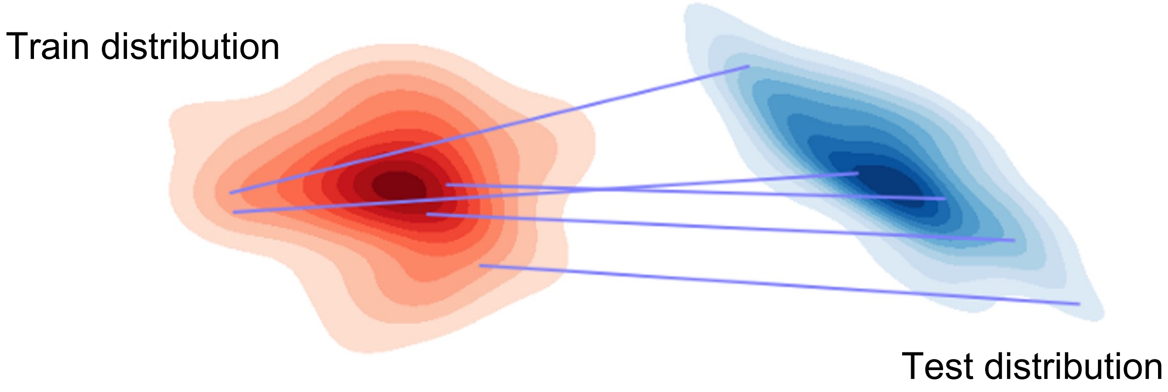
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

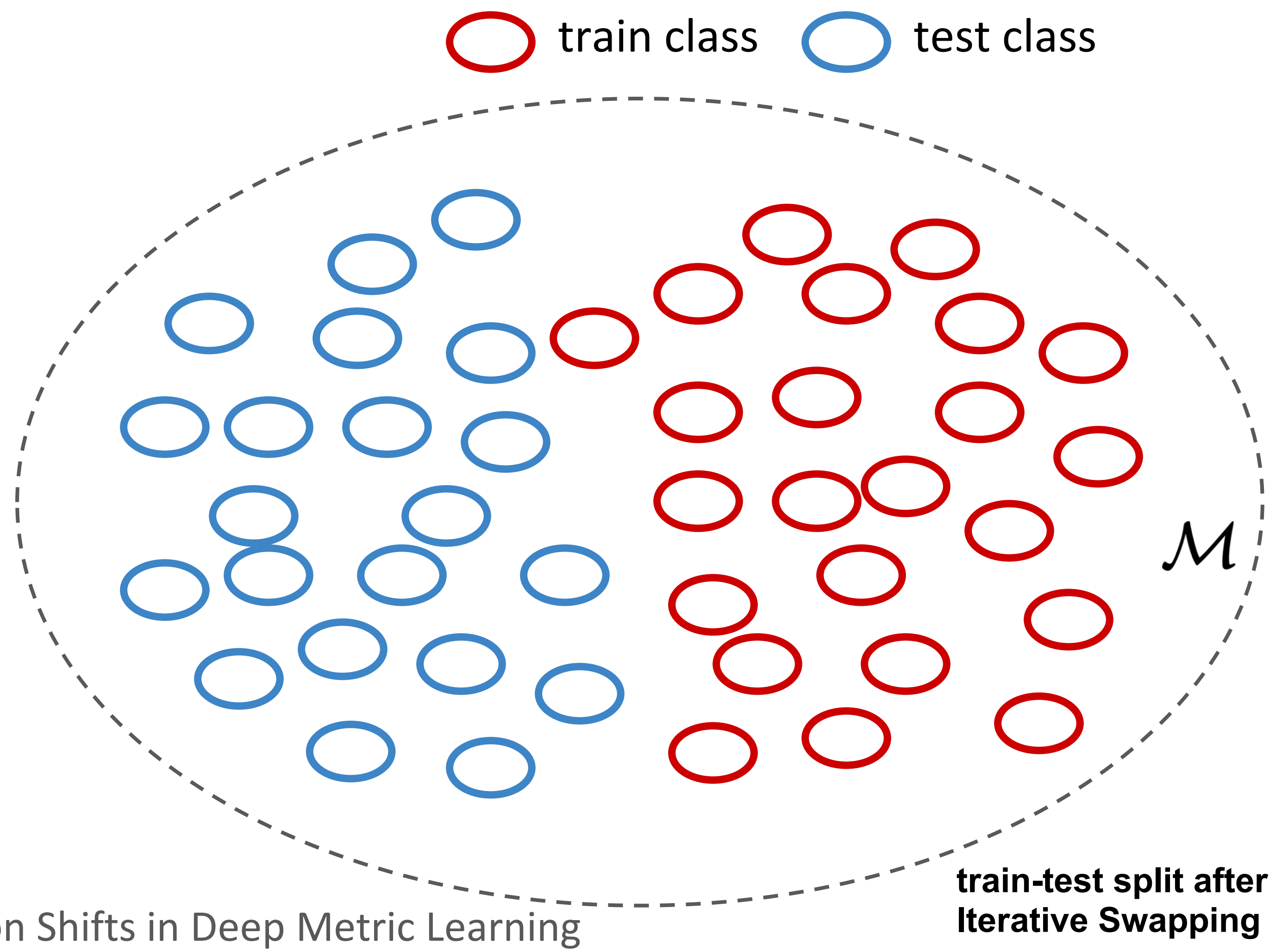
Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

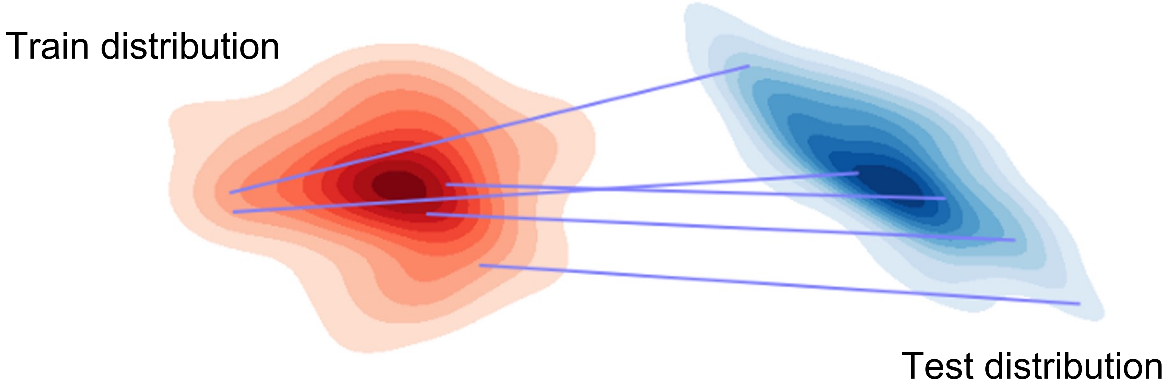
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

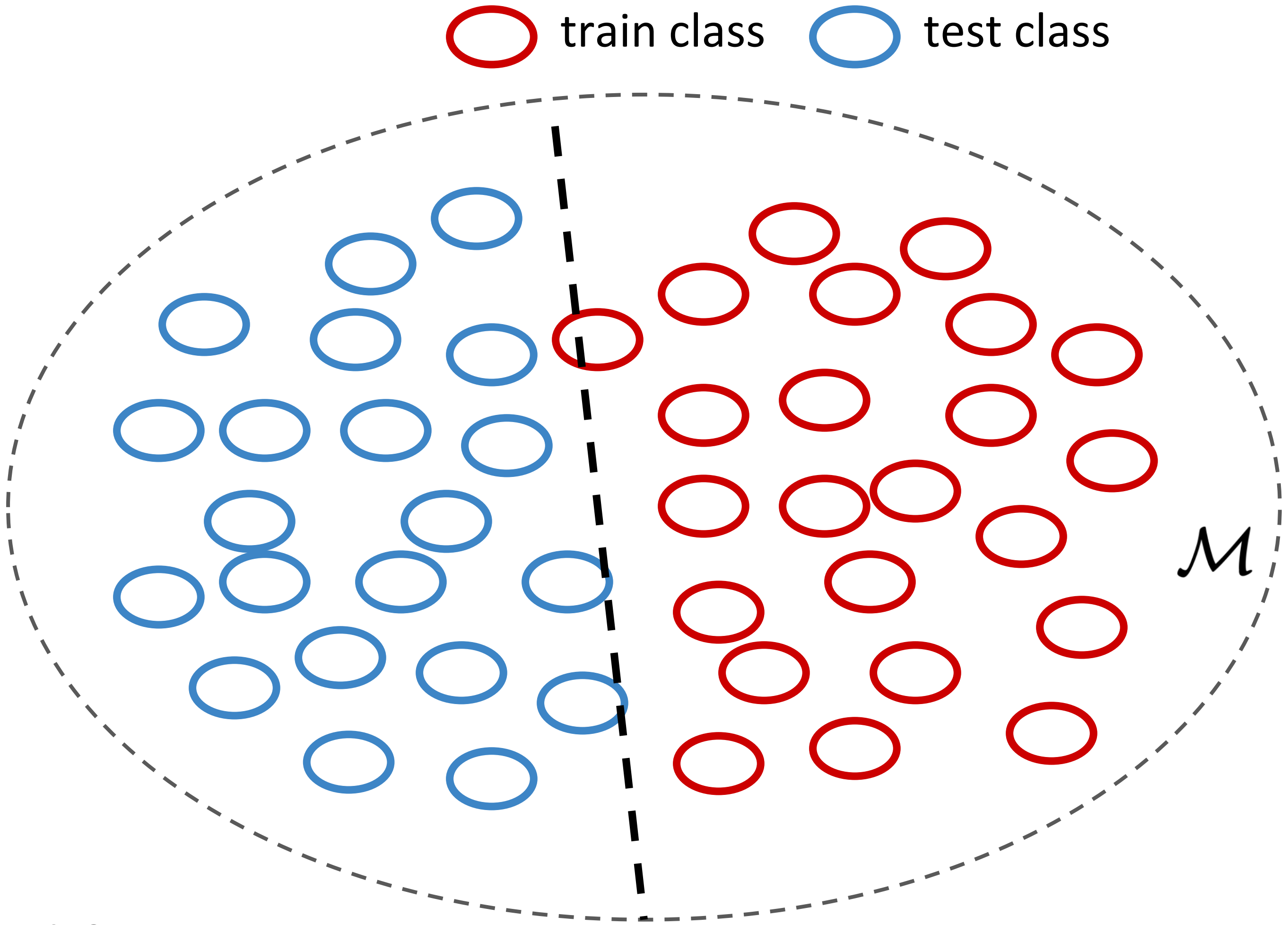
Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

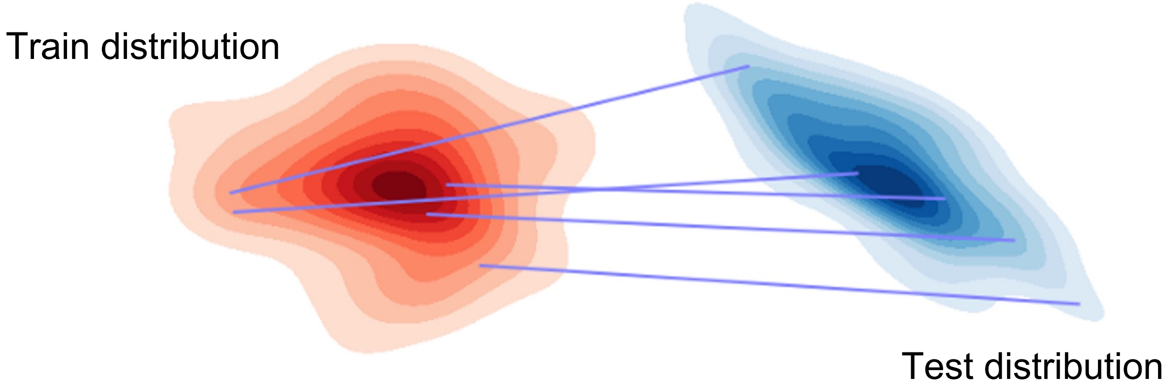
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

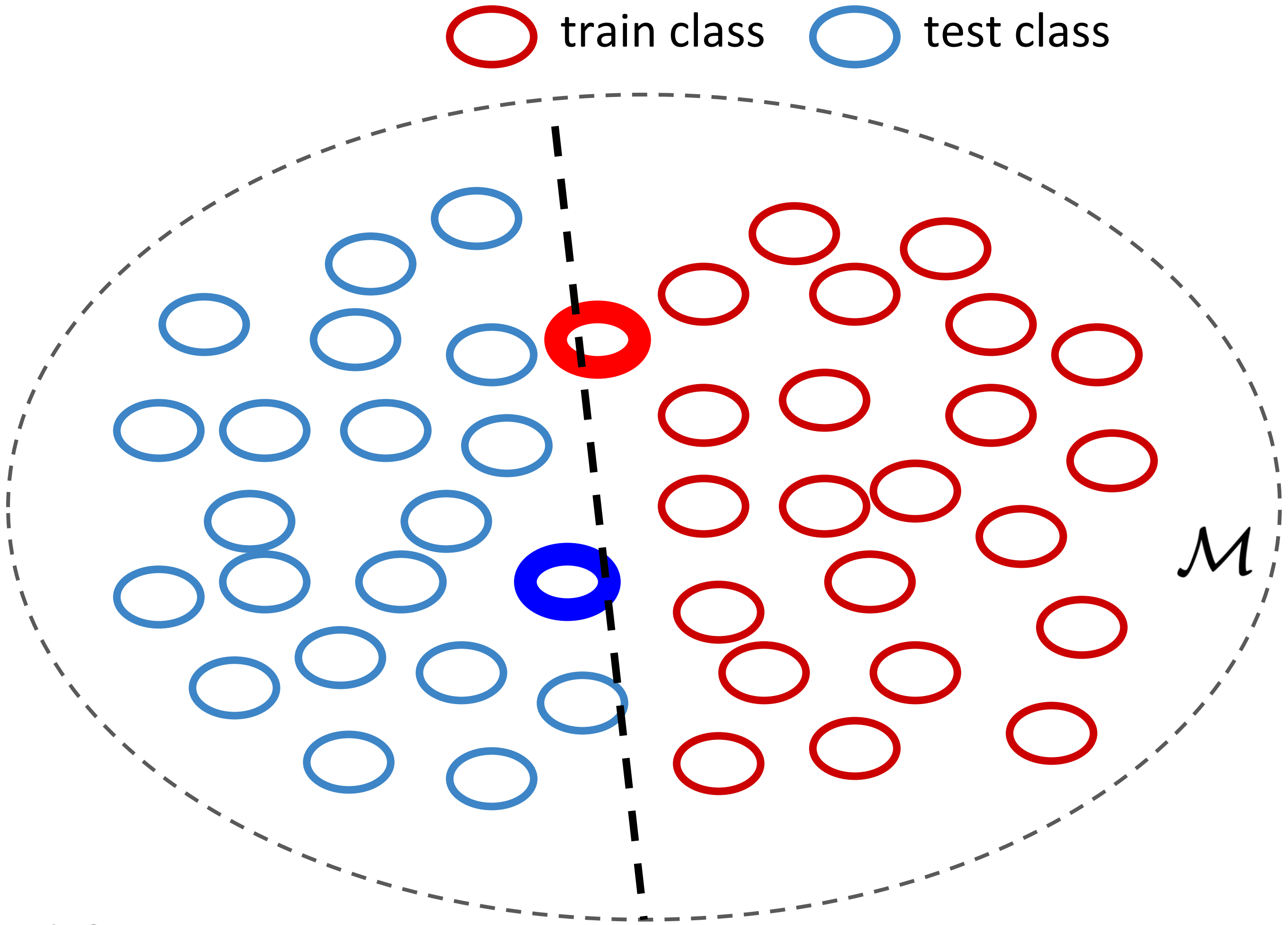
Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

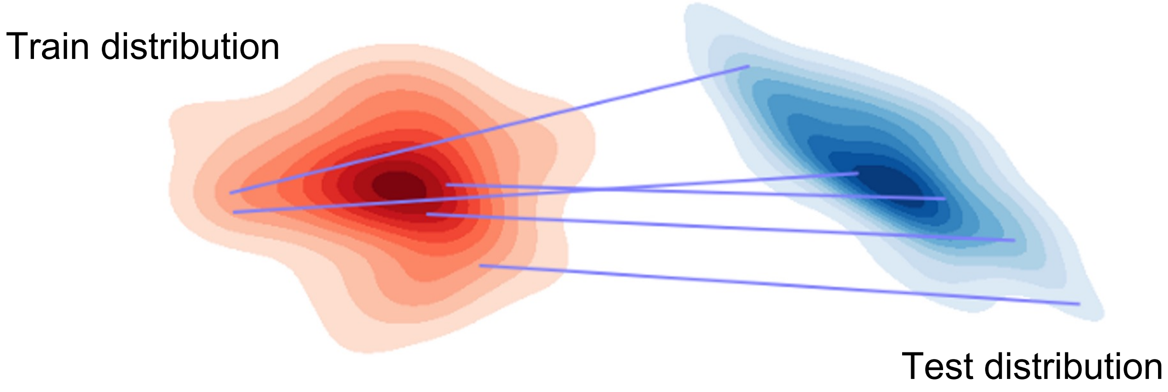
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

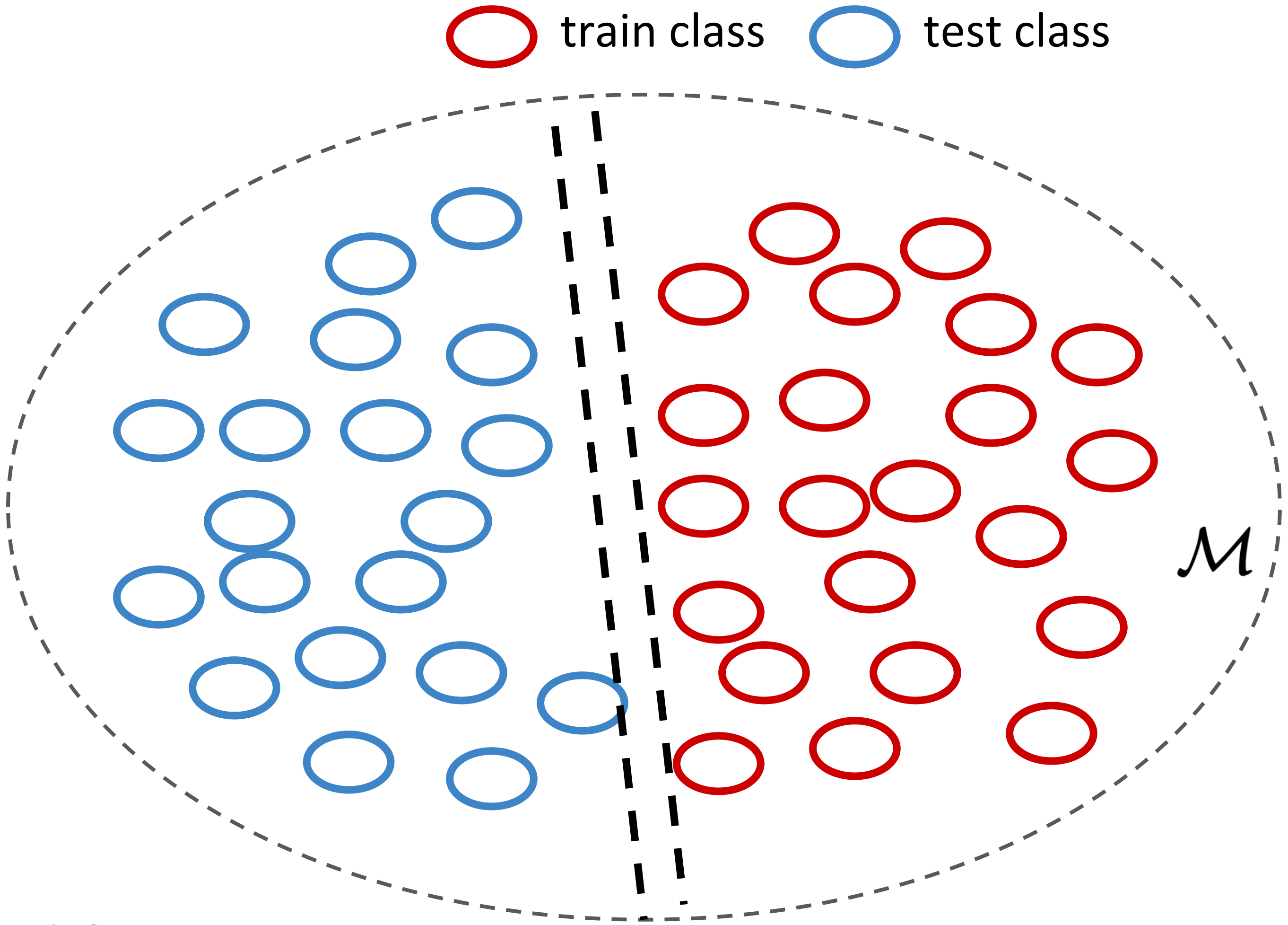
Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

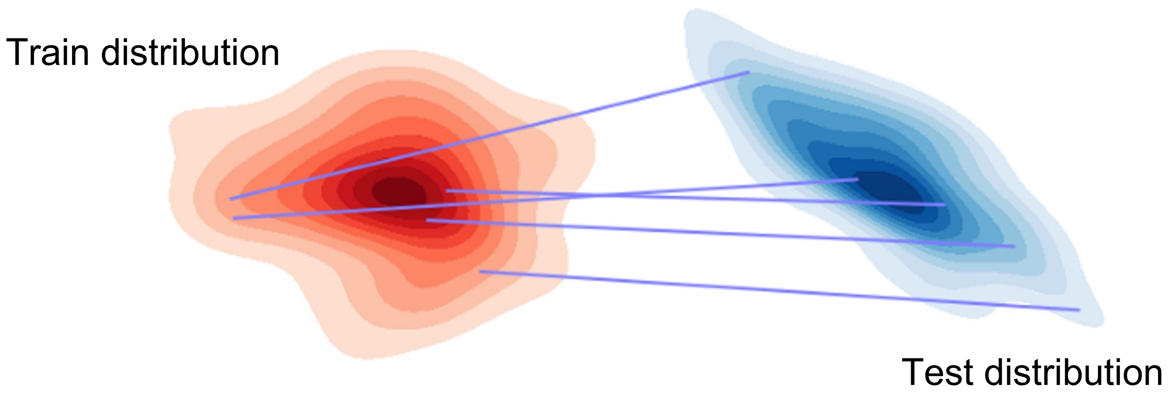
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

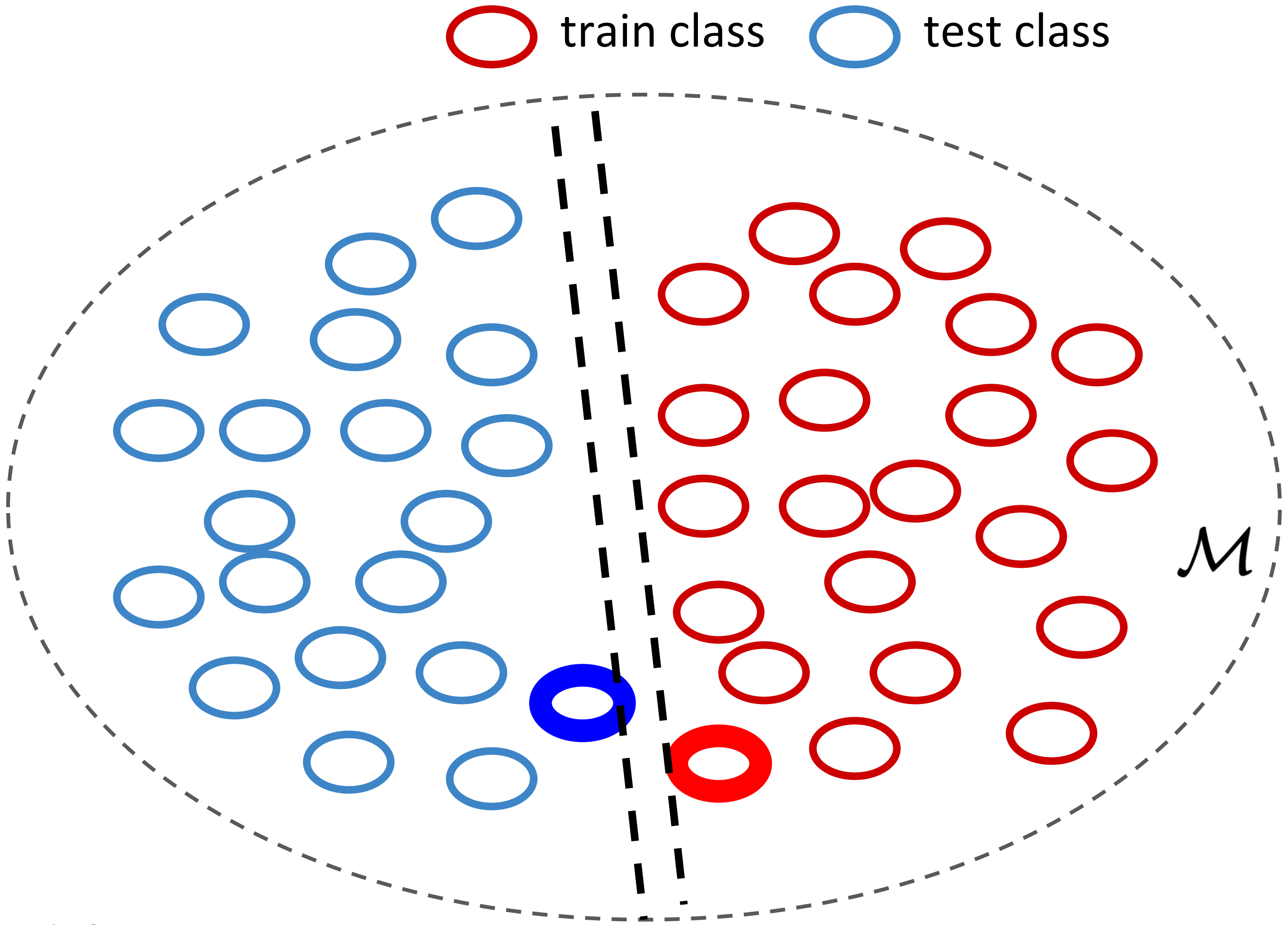
Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal

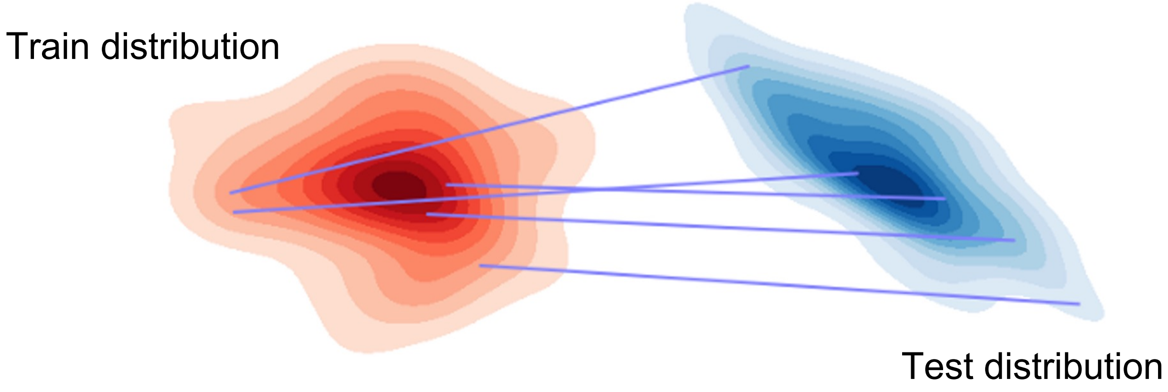


[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

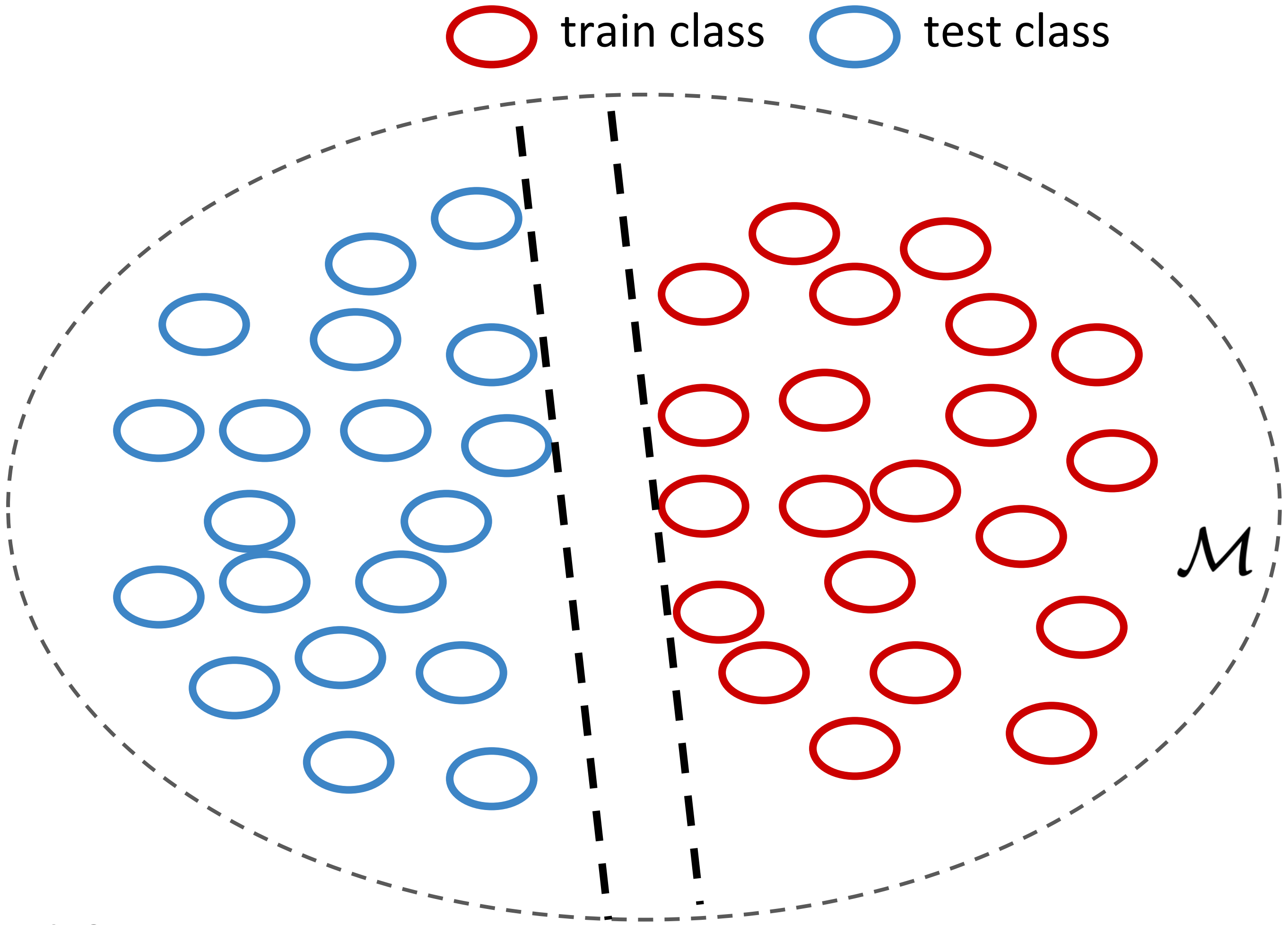


Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.

Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

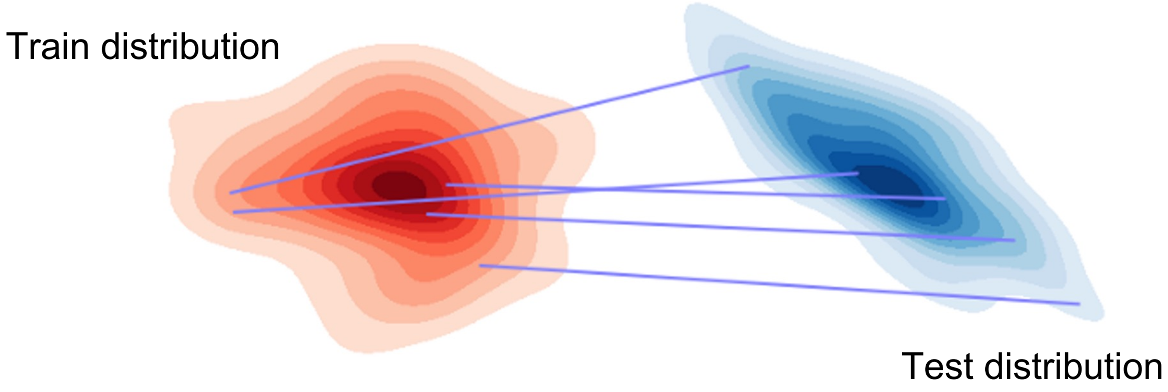
ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.



Iterative Class Swapping

Iterative Class Removal

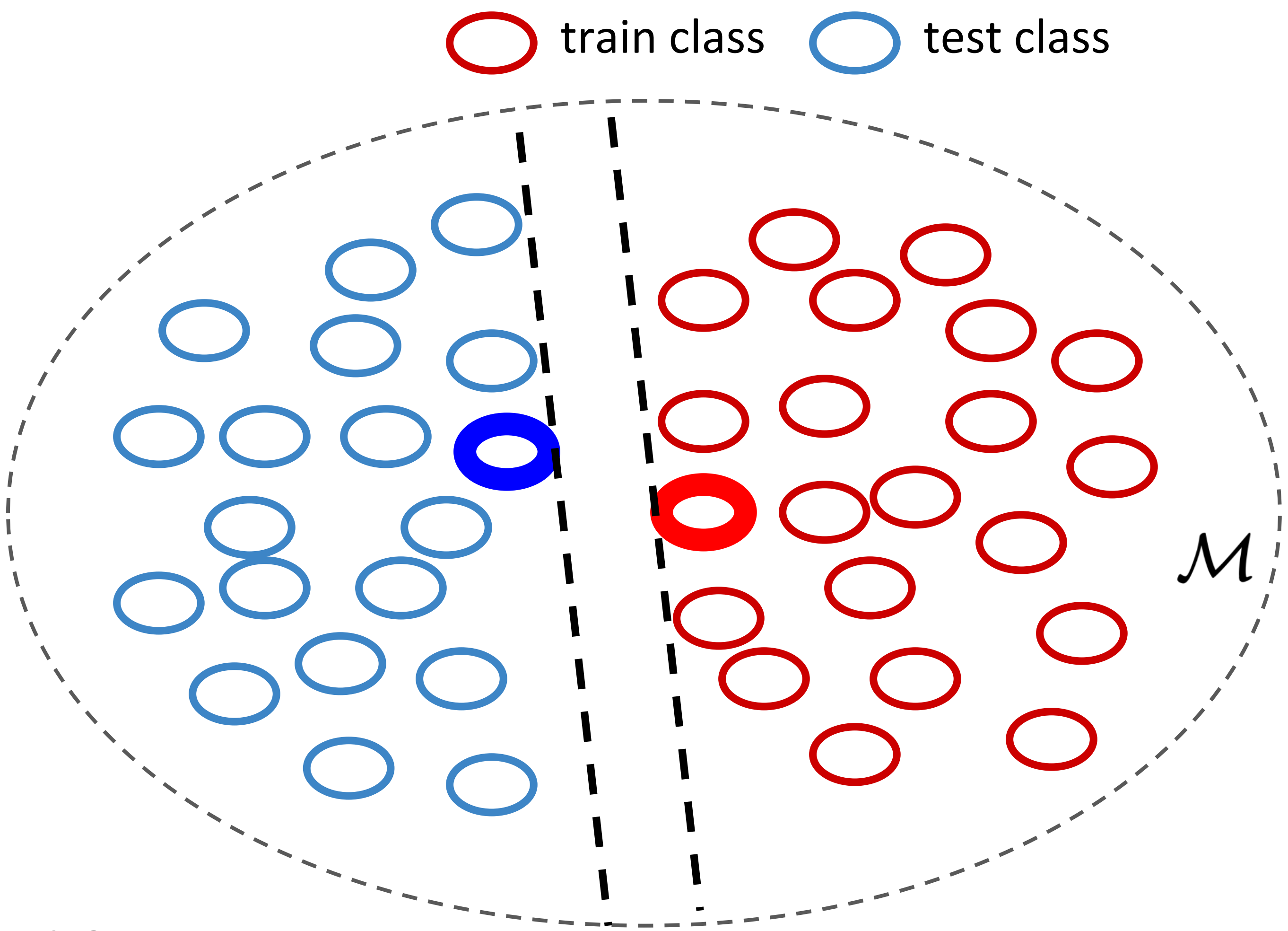
(1) $\mathcal{X}_{\text{train}}$ training data after swapping, $\mathcal{X}_{\text{test}}$ test data after swapping.

(2) Identify class $C_{\text{train}} \subset \mathcal{X}_{\text{train}}$ closest to test distribution $\mathcal{X}_{\text{test}}$ and vice versa.

$$C^*_{\text{train}} = \underset{C_{\text{train}} \in \mathcal{X}_{\text{train}}}{\arg\max} \|\mu_{C_{\text{train}}} - \mu_{\mathcal{X}_{\text{test}}}\|_2$$

$$C^*_{\text{test}} = \underset{C_{\text{test}} \in \mathcal{X}_{\text{test}}}{\arg\max} \|\mu_{C_{\text{test}}} - \mu_{\mathcal{X}_{\text{train}}}\|_2$$

(1) Remove classes C_{train} and C_{test} from $\mathcal{X}_{\text{train}}$ and $\mathcal{X}_{\text{test}}$.

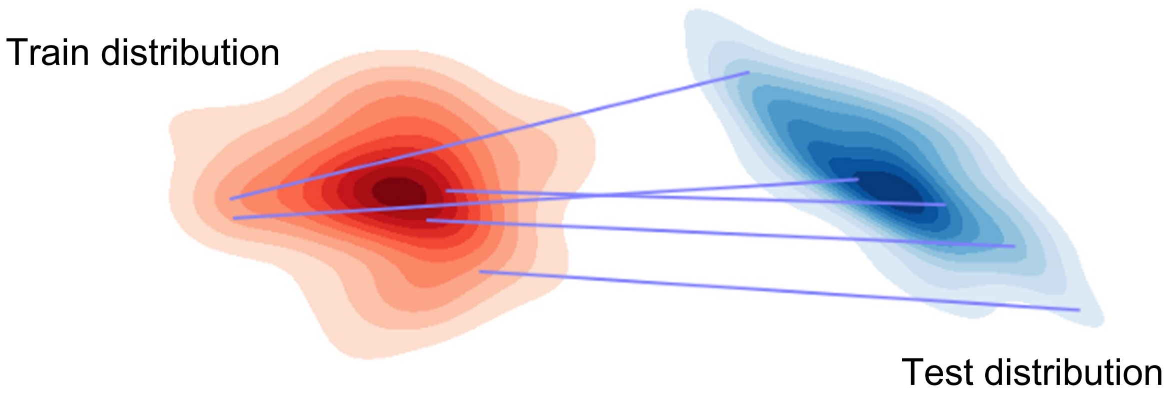


[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

ooDML: Towards Evaluating OOD Generalization

Frechet Inception Distance (FID) to measure distance between data distributions $\mathcal{X}_1$ and $\mathcal{X}_2$.

$$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathcal{X}_1} + \Sigma_{\mathcal{X}_2} - 2(\Sigma_{\mathcal{X}_1}\Sigma_{\mathcal{X}_2})^{\frac{1}{2}})$$

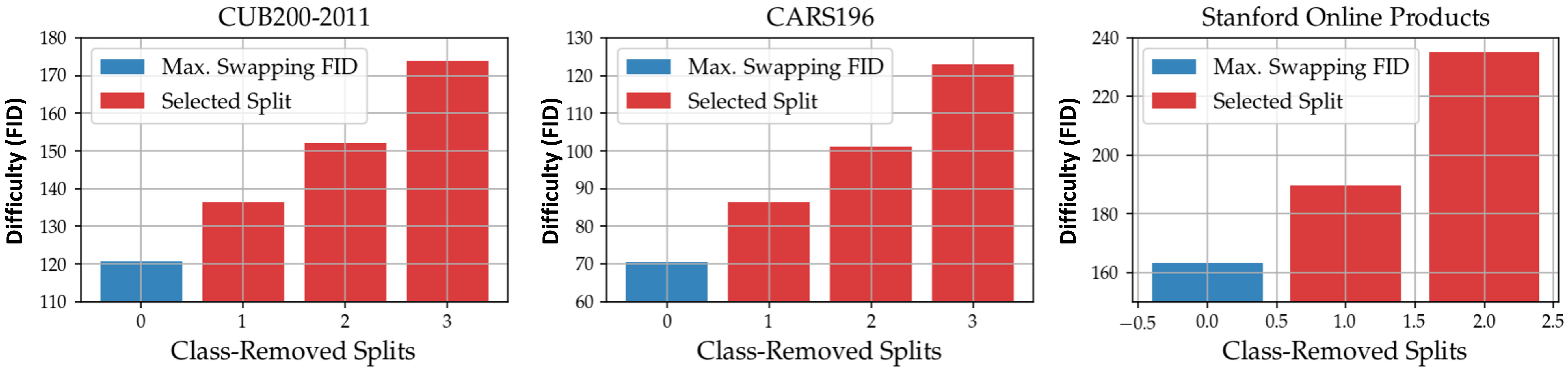


Assumptions:

- Use pretrained ImageNet representation (InceptionV3) as metric space $\mathcal{M}$.
- Data distributions are approximately Gaussian in $\mathcal{M}$.

Iterative Class Swapping

Iterative Class Removal



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

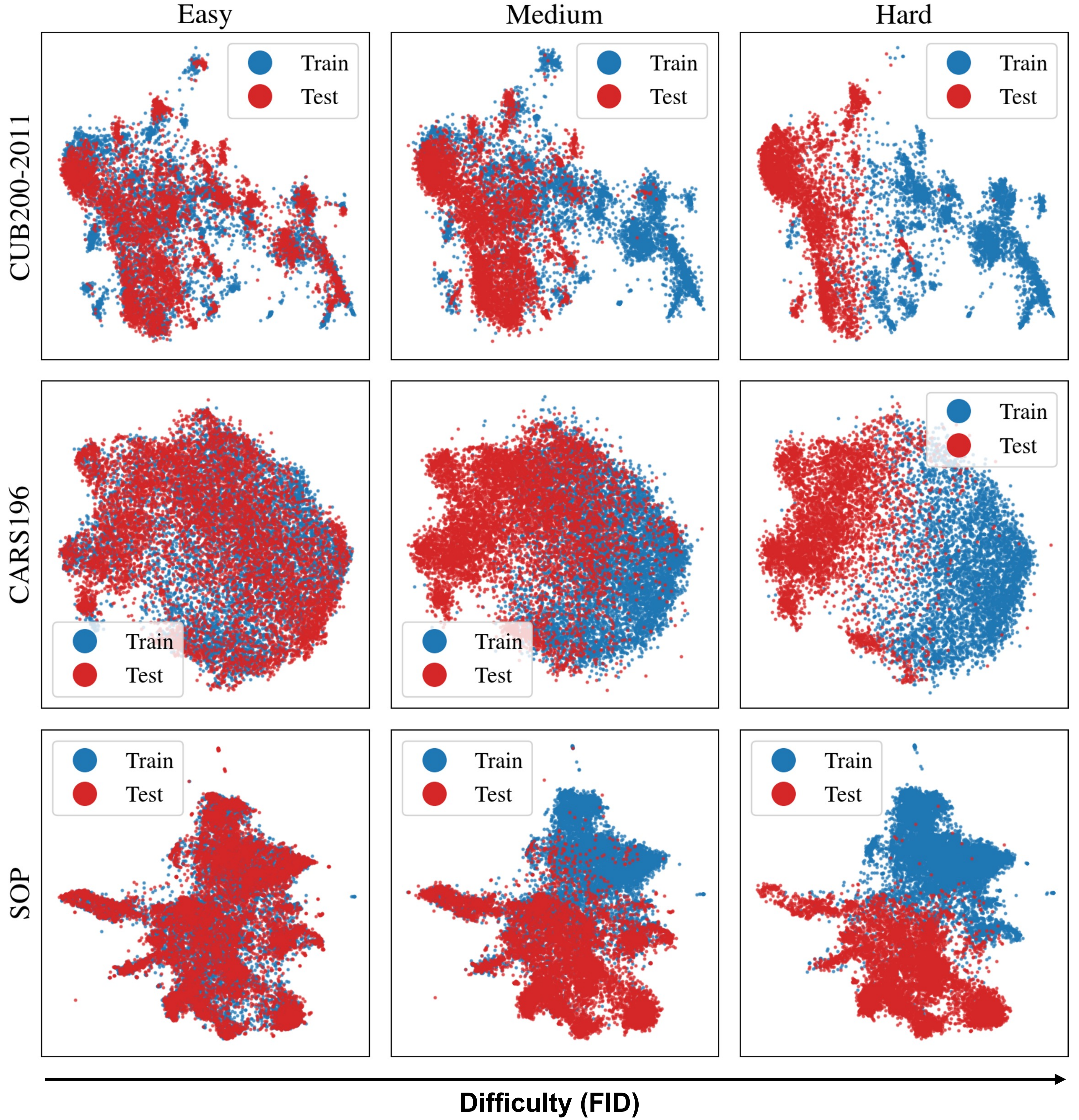
ooDML: Towards Evaluating OOD Generalization

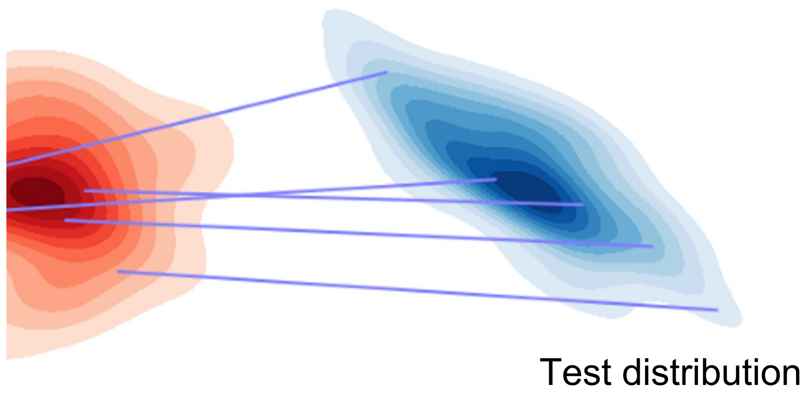
Frechet Inception Distance (FID) t
$\mathcal{X}_1$ and $\mathcal{X}_2$.

$d(\mathcal{X}_1, \mathcal{X}_2) \triangleq \|\mu_{\mathcal{X}_1} - \mu_{\mathcal{X}_2}\|_2^2 + \mathrm{Tr}$

Assumptions:

- Use pretrained ImageNet r
- Data distributions are appr





Iterative Class Swapping

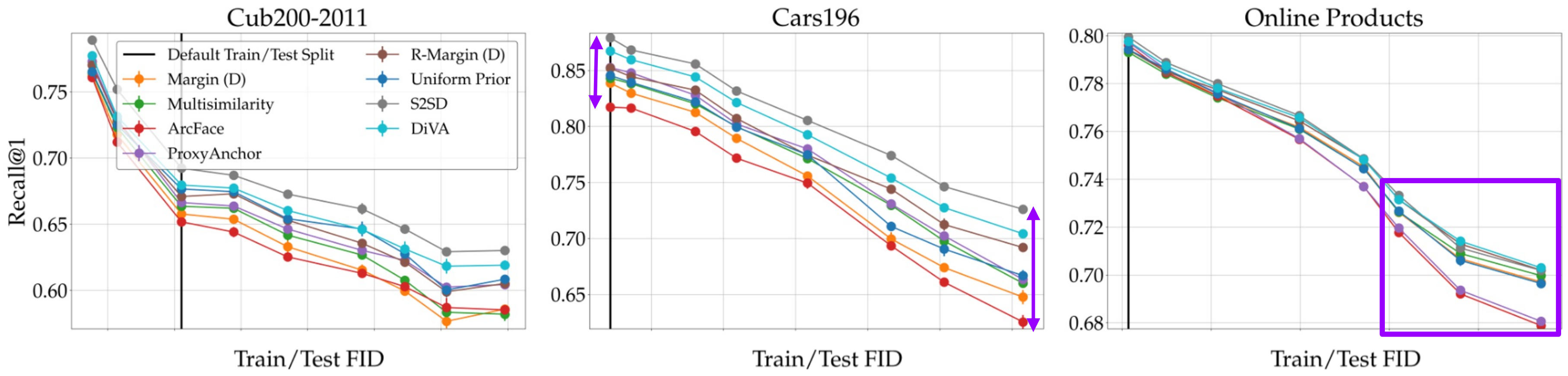
Iterative Class Removal

Procedures is also applicable to other fields and domains!

Assessing Generalization using ooDML benchmark

Benchmark learning concepts:

- performance monotonically decreases
- S2SD most robust too OOD shifts
- proxy-learning seems to lacks behind on SOP

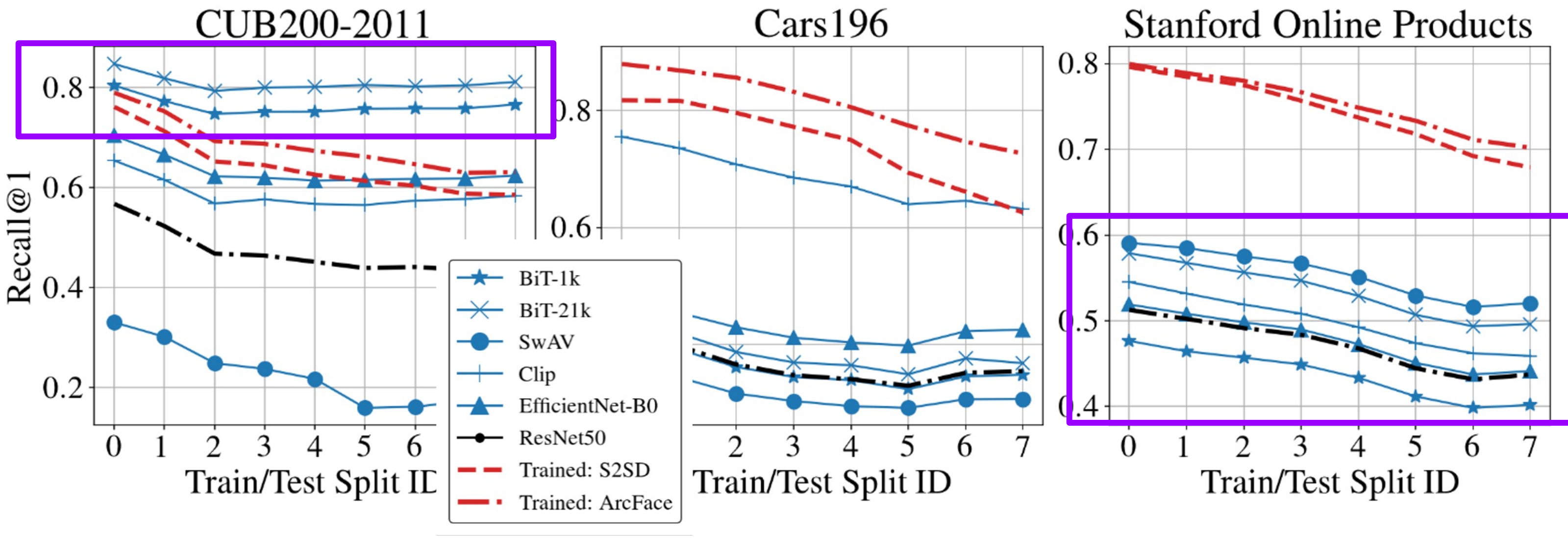


[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

Assessing Generalization using ooDML benchmark

Generic representations for DML:

- CLIP, Vision Transformer (huge pretraining+architecture, text labels) performance surprisingly strong on some datasets
- only explicit adaptation to training data closer to test distributions provides reliable generalization

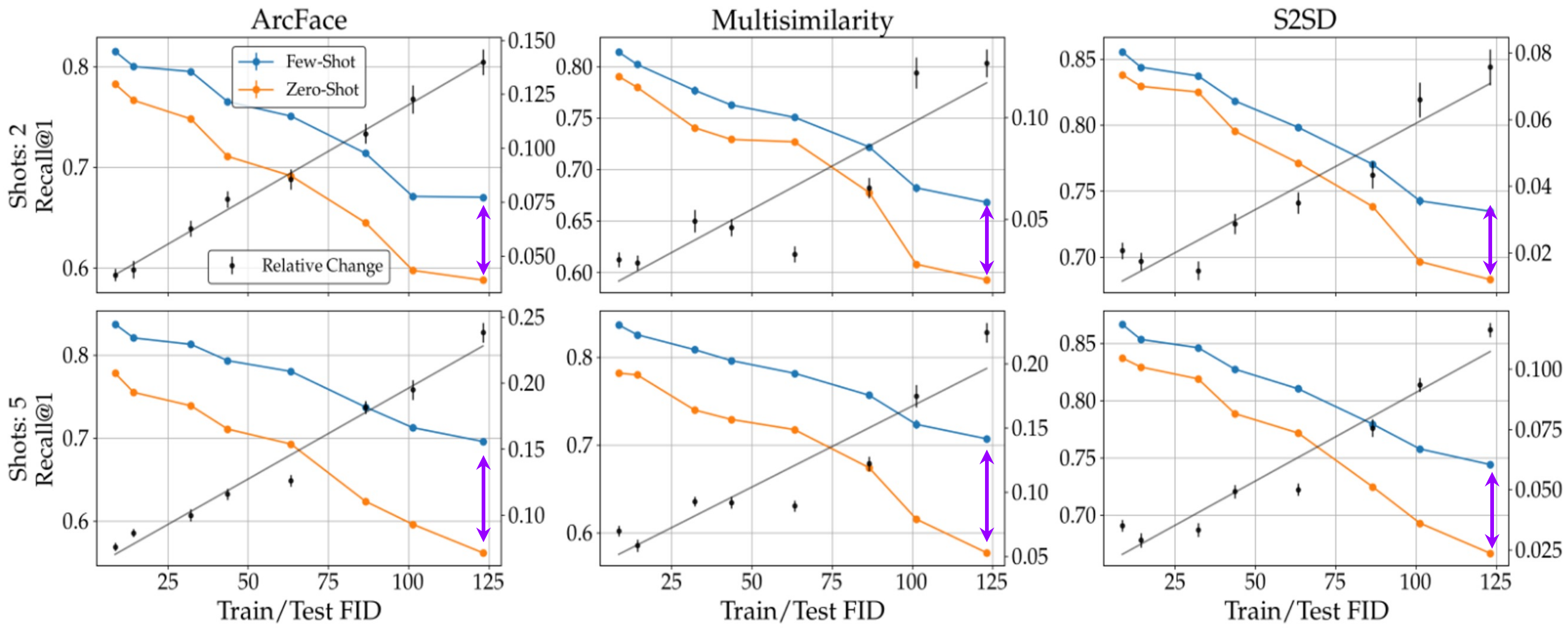


[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

Assessing Generalization using ooDML benchmark

Introducing Few-shot Learning to DML:

- even few examples of unseen test distribution yield consistent improvement of OOD generalization.
- gains across all evaluated DML methods



[1] Milbich, Roth et al.; NeuRIPS 2021; Characterizing Generalization under Out-Of-Distribution Shifts in Deep Metric Learning

Questions?