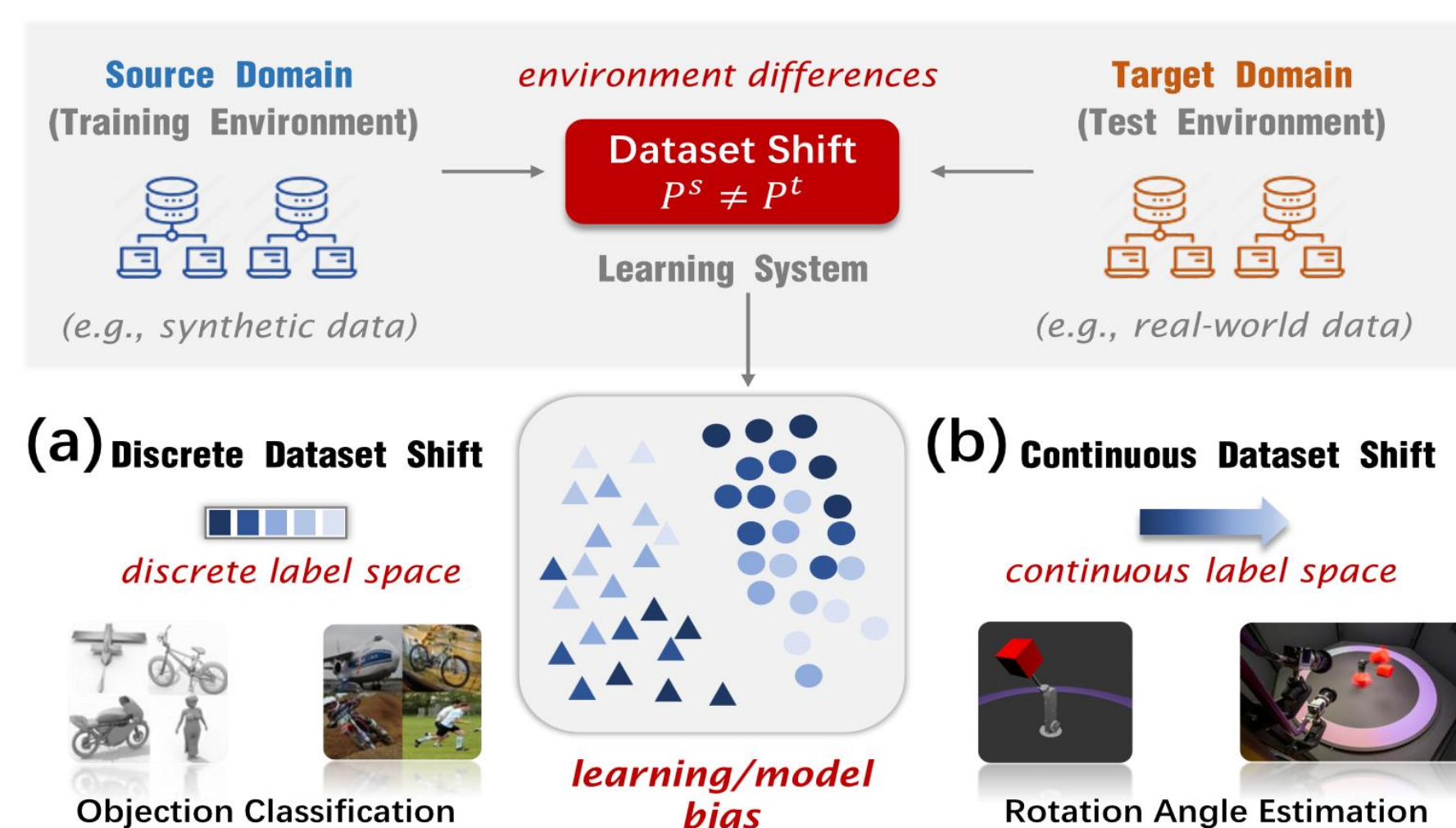


TRLD:

Correct the data distribution shift without erasing the continuous label geometry.

THEORY GAP → PRINCIPLE → KERNEL PROXY → EVIDENCE

1 PROBLEM

Continuous labels break the
usual DA recipeThe target is not only $p_Z^s = p_Z^t$.
Aligning labeling rule across domains!

2 THEORY GAP

Alignment alone is insufficient

$$|\varepsilon_{ps} - \varepsilon_{pt}| \leq D(P_{Z|Y}^s, P_{Z|Y}^t) + c(Y)$$

DISCRIMINABILITY

Preserve an identifiable,
label-dominated
geometry

$$\max I(Z; Y)$$

TRANSFERABILITY

Remove domain
information
conditioned on labels

$$Z \perp E | Y$$



Key principles:

• Conditional shift correction

Align $p_{Z|Y}$ across source and target,
rather than matching only p_Z .

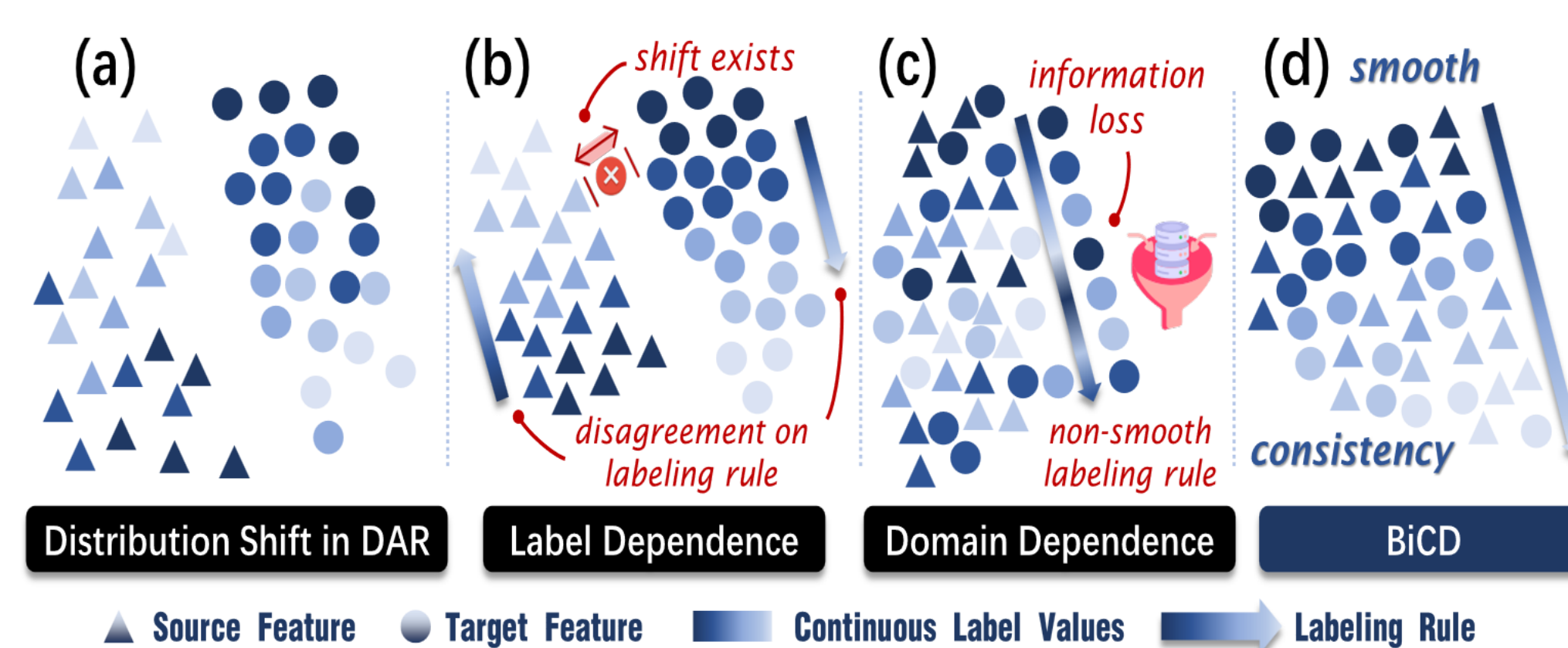
• Continuous label geometry

Nearby labels remain smooth while
distinct label-conditioned supports
stay identifiable.

• A computable objective

Avoid explicit estimation of high-
dimensional conditional distributions.BiCD unifies these requirements into
statistical dependence learning.

3 PRINCIPLES

Two dependences define the
ideal representations(b) MAXIMIZE
conditional label
dependence(c) MINIMIZE
conditional domain
dependence

$$\min_Z \mathcal{L}_{\text{BiCD}} = -I_{Ml}(Z; Y|E) + I_{Ml}(Z; E|Y)$$

4 LEARNING MODEL

A trainable kernel-based proxy

A. Exact identity

$$I(Z; E|Y) - I(Z; Y|E) \equiv I(Z; E) - I(Z; Y)$$

B. NOCCO kernel proxy

$$\hat{J}_{\text{BiCD}}^n = \text{Tr}(\mathbf{R}_E \mathbf{R}_Z) - \text{Tr}(\mathbf{R}_Y \mathbf{R}_Z)$$

C. Plug-and-play learning

$$\inf_{h,g} \mathcal{L}_{\text{Base}}(h, g) + \lambda_{\text{Dom}} \hat{I}_{\text{NOCCO}}^{(n)}(Z; E) - \lambda_{\text{Lab}} \hat{I}_{\text{NOCCO}}^{(n)}(Z; Y)$$



The optimum admits structural properties:

TRANSFERABILITY

 $\{P_{Z^*|Y=y} | y \in \mathcal{Y}\}$ are
linearly independent

DISCRIMINABILITY

$$\Omega_Y \cap \Omega_{Y'} = \emptyset$$

SEPARABILITY

$$P_{Z^*|Y}^s = P_{Z^*|Y}^t$$

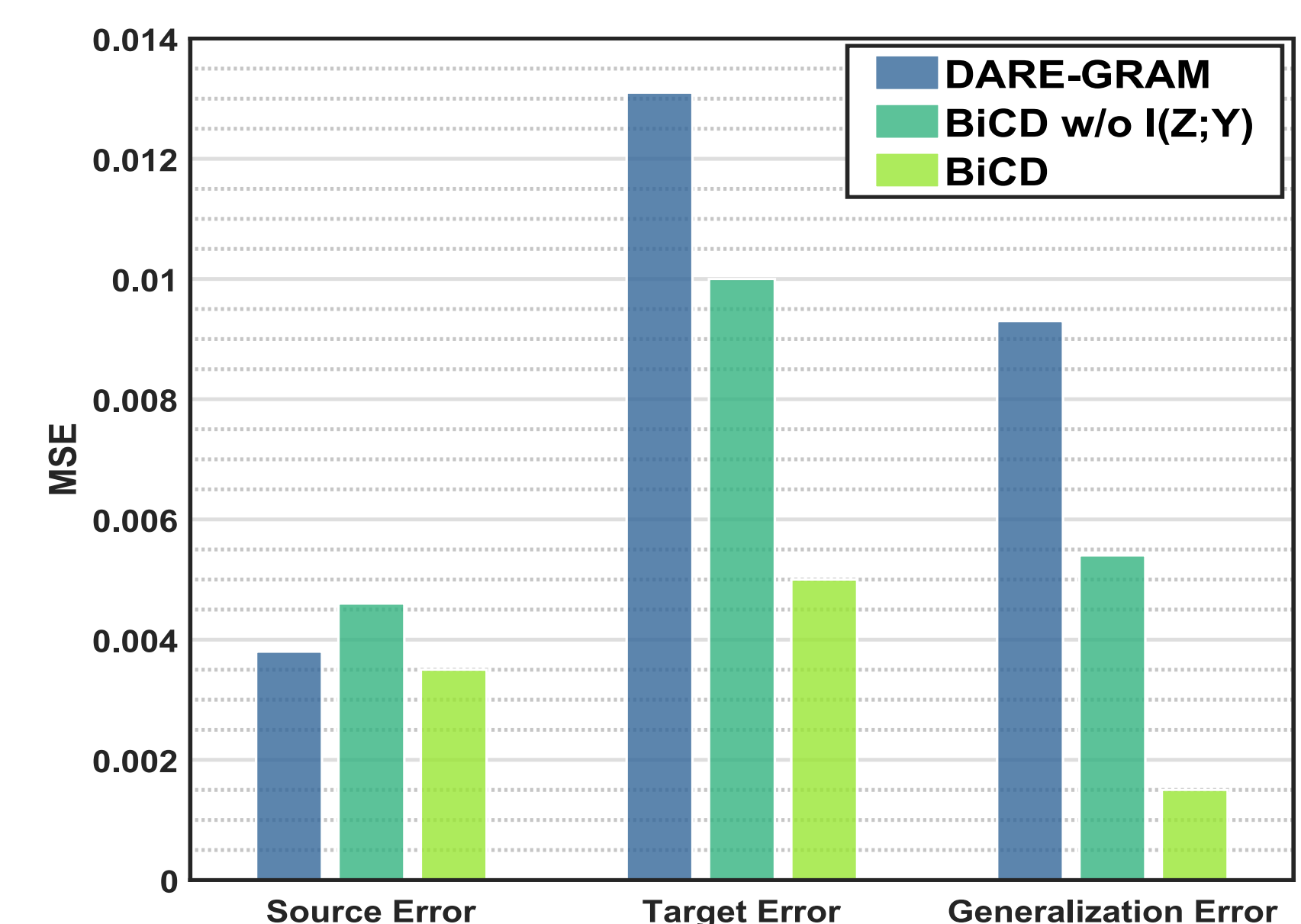
5 EVIDENCE

The representation changes in
the predicted way

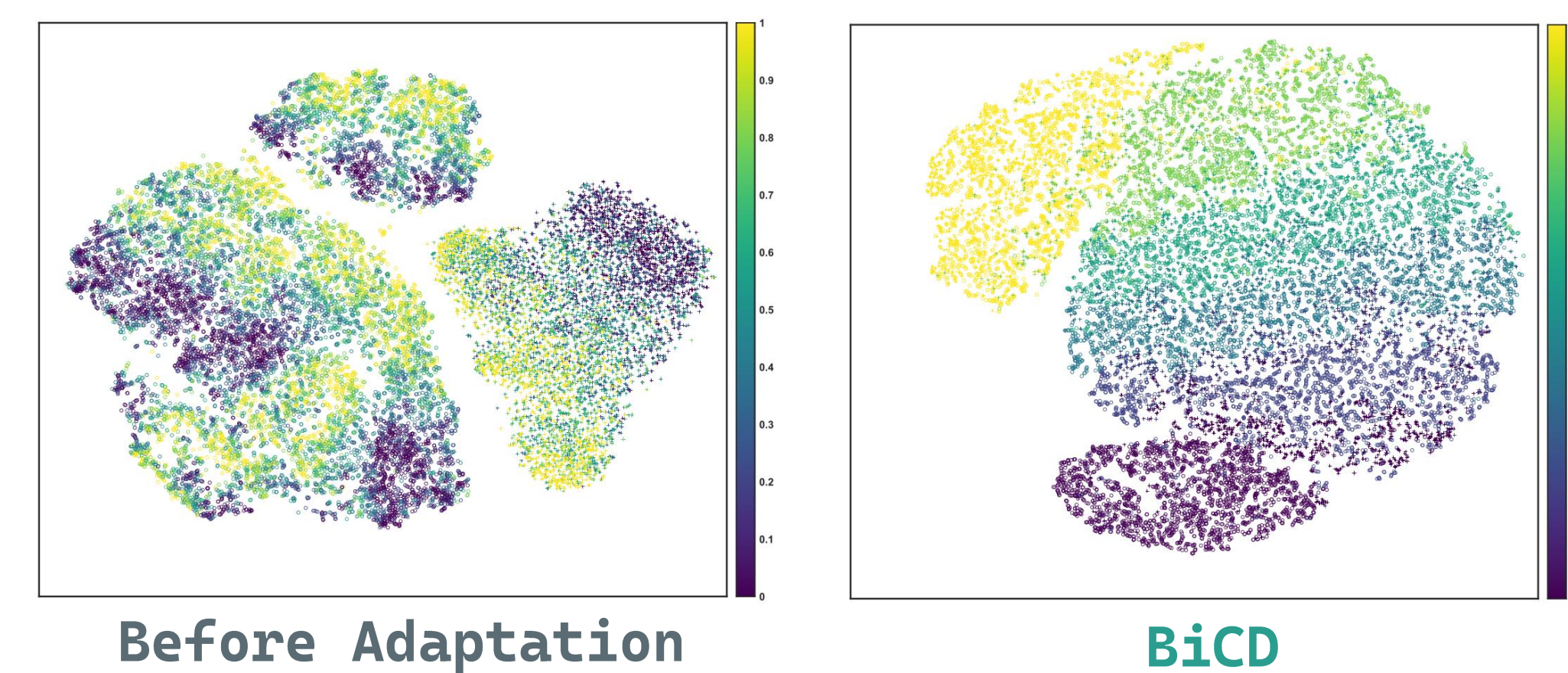
A. Comparison with SOTA methods

MPI3D MAE	BIWI MAE	ABLATION MAE
0.13	0.24	0.202
vs 0.16 DARE/UGA	vs 0.26 DARE-GRAM	vs 0.248 base

B. Generalization risk

BiCD yields the lowest target risk and
the smallest source-target risk gap.

C. Representation Geometry

Domain alignment and a smooth label
manifold are ensured together.

D. Ablation: both dependences matter

Base	0.248
Single Dependence	0.215
Bi-level Dependence	0.202