Bridging OOD Generalization and Detection: A Graph-Theoretic View

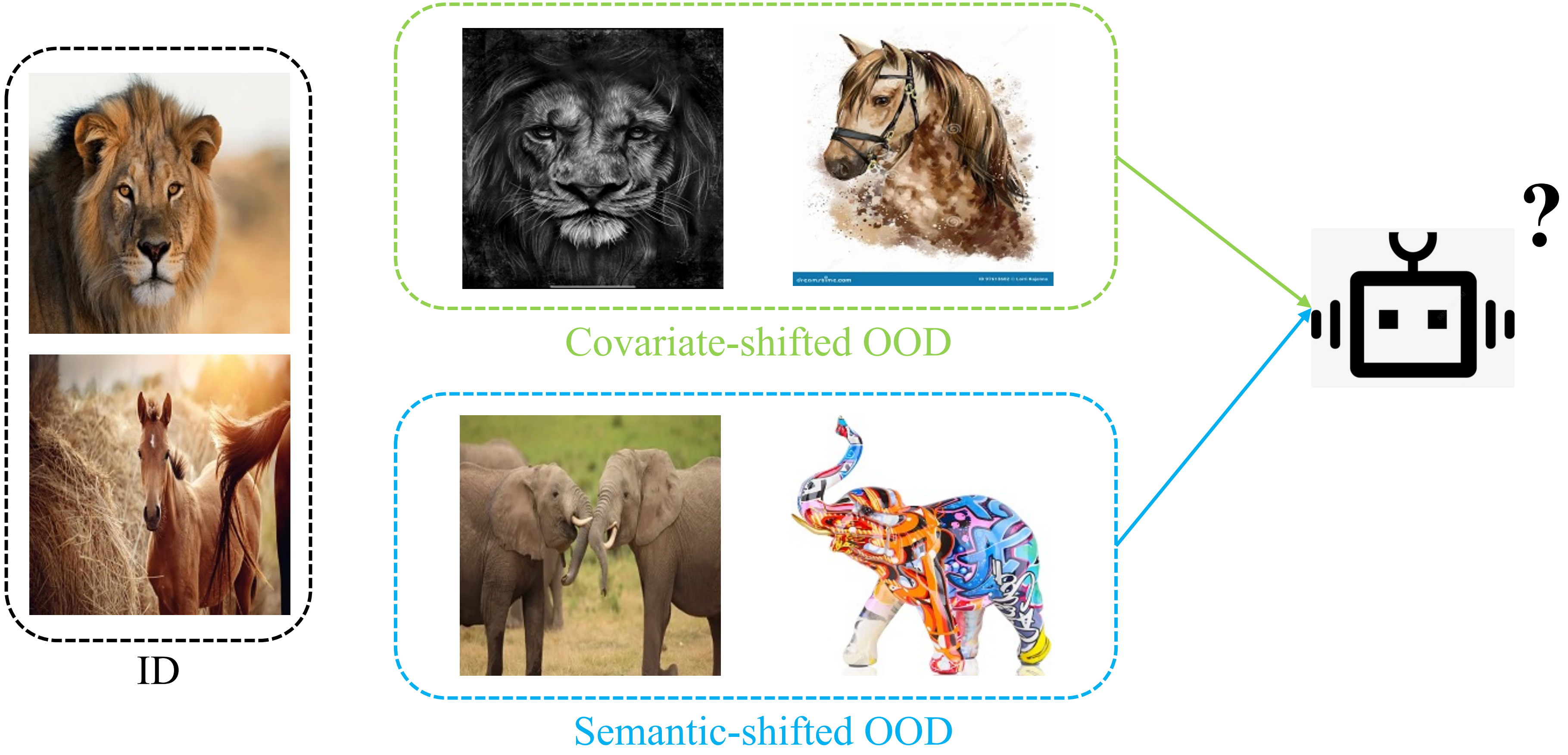
Han Wang[1], Yixuan Li[2]

[1] University of Illinois Urbana-Champaign, [2] University of Wisconsin, Madison

11/12/24

Harmonizing OOD generalization and detection

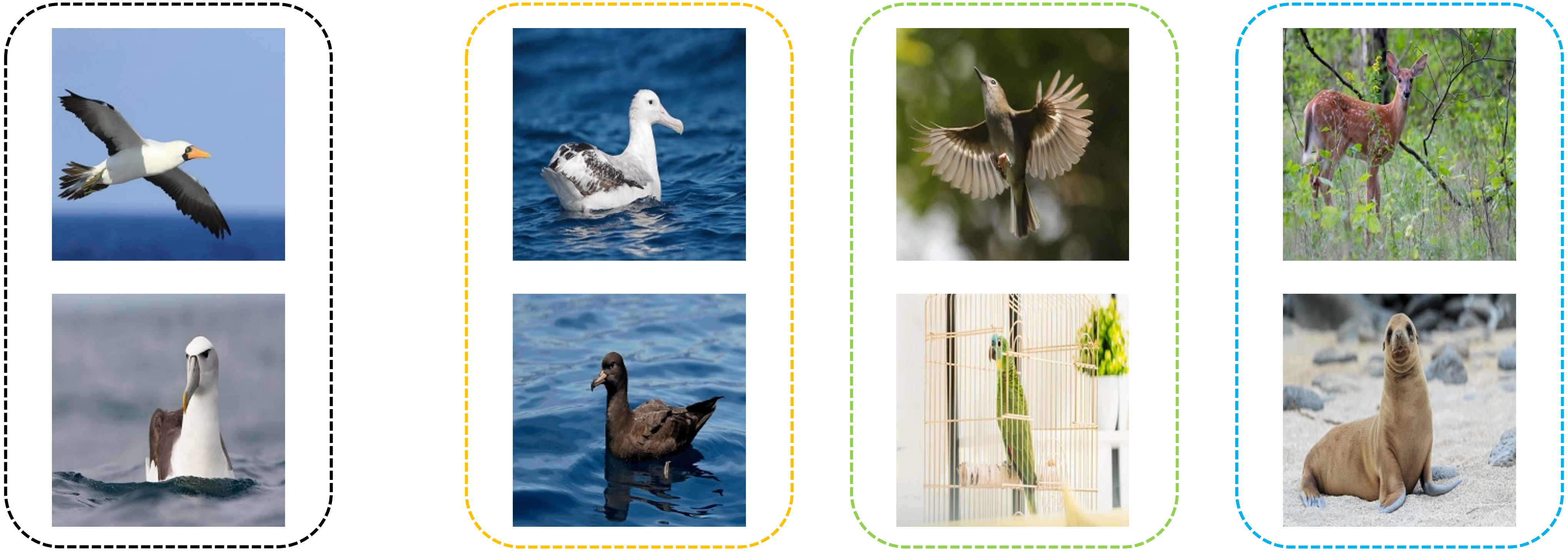
- OOD generalization: generalize to **covariate-shift OOD** data
- OOD detection: reject unknown **semantic-shift OOD** data



A Real-World Scenario

Labeled ID data
(known class)

Unlabeled wild data



ID

Covariate-shifted OOD

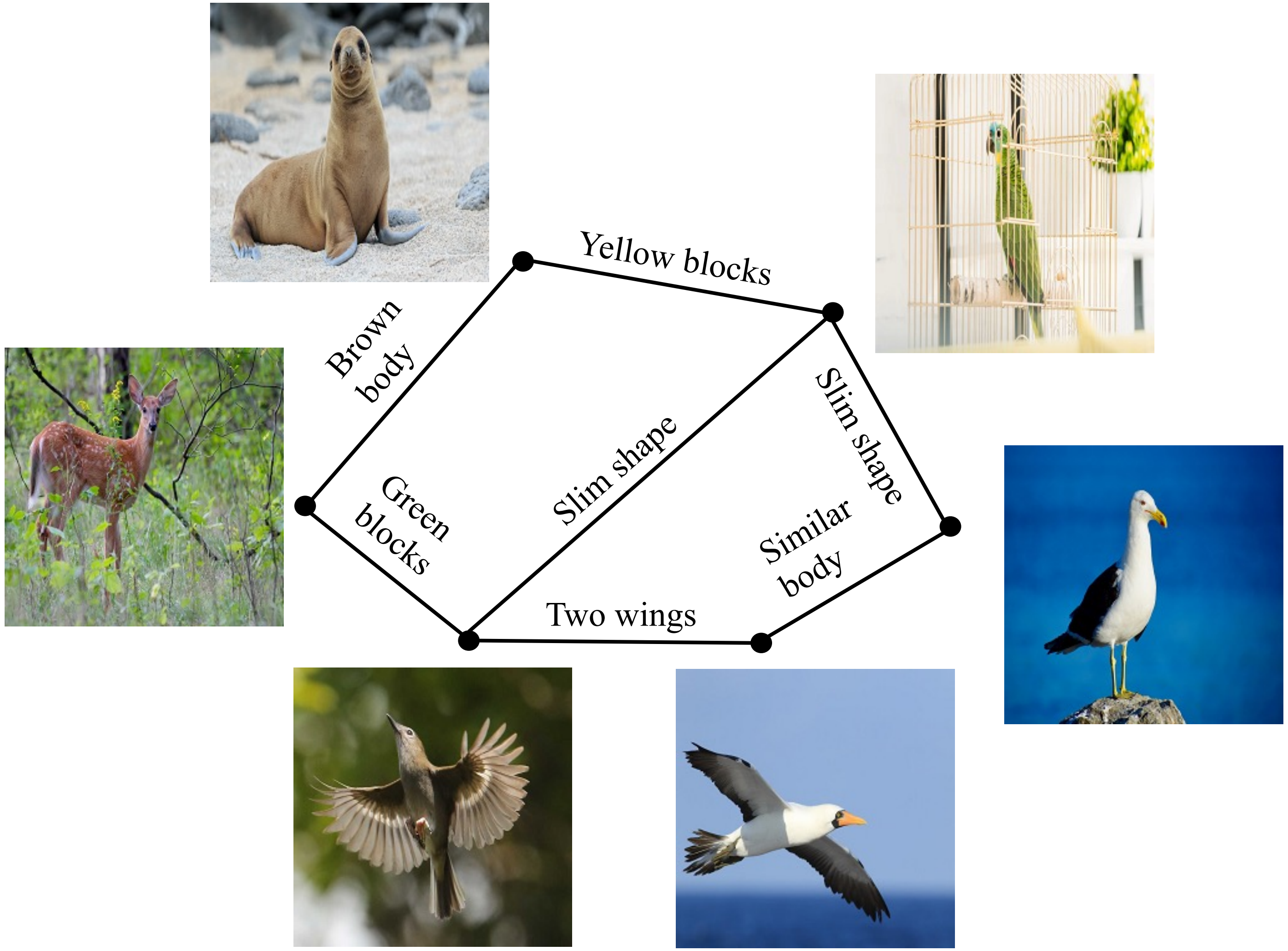
Semantic-shifted OOD

[1] Bai, Haoyue, et al. "Feed two birds with one scone: Exploiting wild data for both out-of-distribution generalization and detection." *International Conference on Machine Learning*. PMLR, 2023.

We can perform analysis on a semantic graph!

Node: Image

Edge: Semantic connection between two images



Augmentation Graph

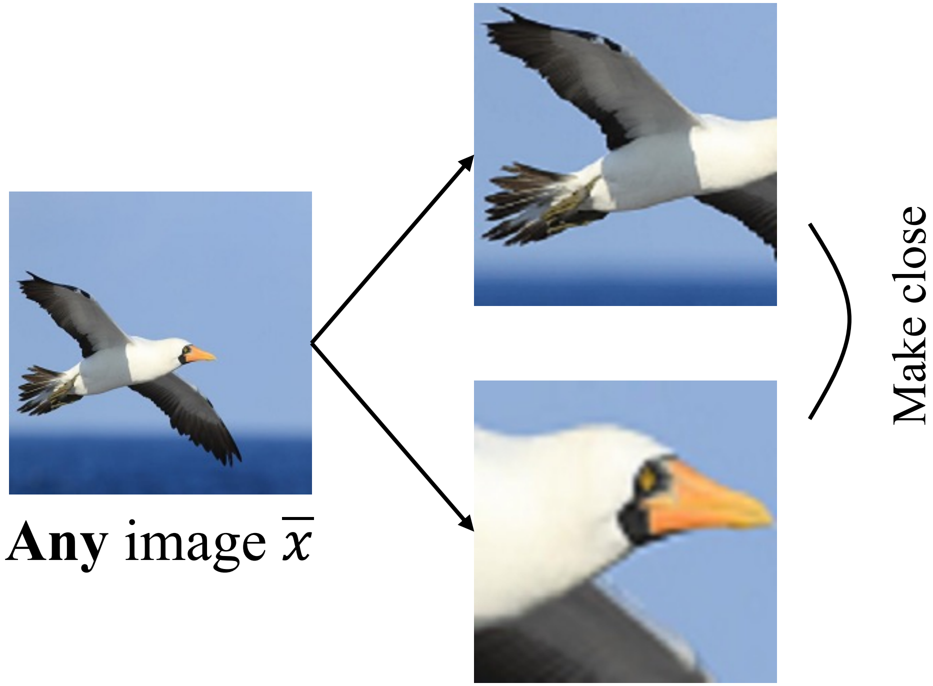
Node: Augmented Images

...

Edge Weight (semantic connections): Probability of two images being considered as positive pairs

Two Cases of Positive Pairs

Unsupervised Case $w_{xx'}^{(u)}$

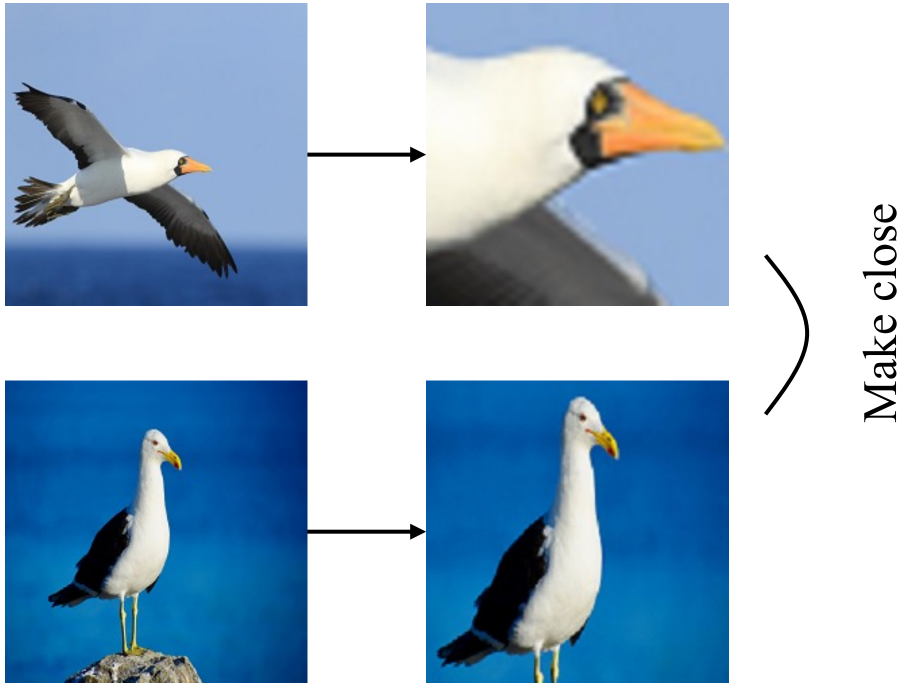


Any image $\bar{x}$

Augmented image x, x'

$$w_{xx'}^{(u)} \triangleq E_{\bar{x}\sim P}\, \mathcal{T}(x|\bar{x})\mathcal{T}(x'|\bar{x})$$

Supervised Case $w_{xx'}^{(l)}$



Any image pairs $\bar{x}, \bar{x}'$ in the **same class** $\mathcal{Y}_l$

Augmented image x, x'

$$w_{xx'}^{(l)} \triangleq \sum_{i\in\mathcal{Y}_l} E_{\bar{x}_l\sim P_{l_i}} E_{\bar{x}_l'\sim P_{l_i}}\, \mathcal{T}(x|\bar{x}_l)\mathcal{T}(x'|\bar{x}_l')$$

Spectral Contrastive Learning with Wild Data

Edge weights: $w_{xx'} = \eta_u w_{xx'}^{(u)} + \eta_l w_{xx'}^{(l)}$

Adjacency Matrix: $A = \eta_u A^{(u)} + \eta_l A^{(l)},\ \ where\ entry\ A_{xx'} = w_{xx}'$

Normalized Adjacency Matrix: $\tilde{A} \triangleq D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, where D is a diagonal matrix with $D_{xx} = w_x$

$$\min_{F \in R^{N \times k}} \boxed{\mathcal{L}_{mf}(F, A) = \left\|\tilde{A} - FF^\top\right\|_F^2}$$

$$F^\top = \sqrt{w_x} f(x)$$

$$\mathcal{L}(f) \triangleq \underbrace{-2\eta_l \mathcal{L}_1(f) - 2\eta_u \mathcal{L}_2(f)}_{\text{Positive pairs}} + \underbrace{\eta_l^2 \mathcal{L}_3(f) + 2\eta_l \eta_u \mathcal{L}_4(f) + \eta_u^2 \mathcal{L}_5(f)}_{\text{Negative pairs}}$$

Positive pairs

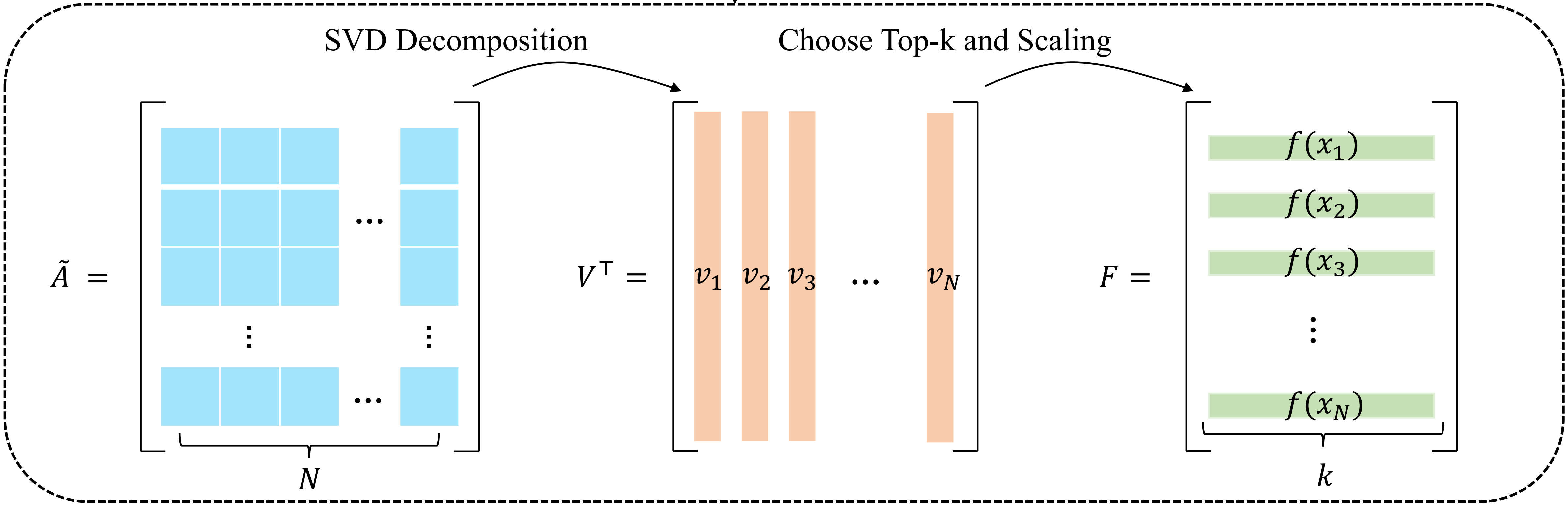
Negative pairs

The closed-form embedding is known!

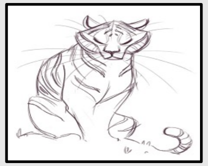
We can analyze the feature space with **spectral analysis** of the adjacency matrix!

$$\boldsymbol{L_{mf}(F, A) = \left\| \widetilde{A} - FF^{\top} \right\|_F^2}$$

Optimal Solution (low-rank approximation)



A Toy Example

Samples	Distribution	Labeled	Class label	Domain label
	ID	✓	Angel	Sketch
	ID	✓	Tiger	Sketch
	Covariate OOD	✗	Angel	Painting
	Covariate OOD	✗	Tiger	Painting
	Semantic OOD	✗	Panda	Cartoon

Evaluation Protocols

OOD generalization:

- Linear probing error: the number of misclassification samples in the covariate-shifted domain.

$$\mathcal{E}(f) \triangleq E_{\bar{x}\sim P_{\text{out}}^{covariate}}[y(\bar{x}) \neq h(\bar{x}; f, M)]$$

OOD detection:

- Separability: the extent of separation between ID and semantic OOD data.

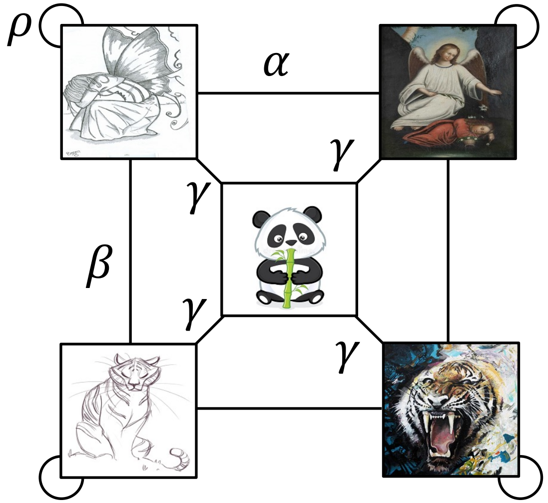
$$S(f) \triangleq E_{\bar{x}_i\sim P_{in},\, \bar{x}_j\sim P_{out}^{semantic}} \left\| f(\bar{x}_i) - f(\bar{x}_j) \right\|_2^2$$

Derive the embedding space by eigen-decomposition

Augmentation Transformation Probability:

$$\mathcal{T}(x|\bar{x}) = \begin{cases} \rho \ if \ y(\bar{x}) = y(x), d(\bar{x}) = d(x); \\ \alpha \ if \ y(\bar{x}) = y(x), d(\bar{x}) \neq d(x); \\ \beta \ if \ y(\bar{x}) \neq y(x), d(\bar{x}) = d(x); \\ \gamma \ if \ y(\bar{x}) \neq y(x), d(\bar{x}) \neq d(x). \end{cases}$$

$$\mathcal{T} = \begin{bmatrix} \rho & \beta & \alpha & \gamma & \gamma \\ \beta & \rho & \gamma & \alpha & \gamma \\ \alpha & \gamma & \rho & \beta & \gamma \\ \gamma & \alpha & \beta & \rho & \gamma \\ \gamma & \gamma & \gamma & \gamma & \rho \end{bmatrix}$$

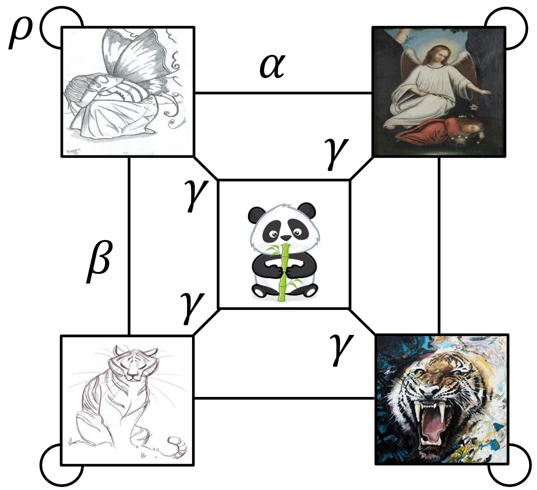


Linear Probing Error & Separability

Evaluation I: linear probing error

- ID classifier of Misclassification in covariate OOD data

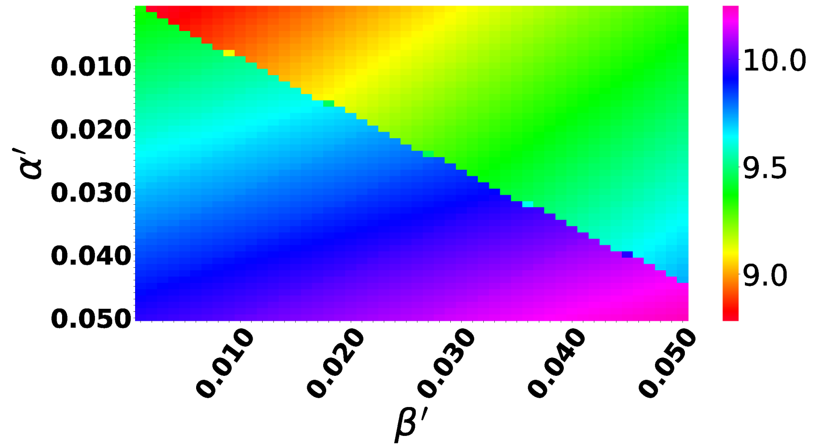
$$\varepsilon(f_1) = \begin{cases} 0, & if\ \frac{9}{8}\alpha > \beta \\ 2, & if\ \frac{9}{8}\alpha < \beta \end{cases}$$



Evaluation II: separability

- Euclidean distance between ID and semantic OOD data

$$\mathcal{S}(f_1) = \begin{cases} (7 + 12\beta' + 12\alpha')\left(\frac{1-2\beta'}{3}\left(1-\beta'-\frac{3}{4}\alpha'\right)^2 + 1\right), if\ \frac{9}{8}\alpha > \beta \\ (7 + 12\beta' + 12\alpha')\left(\frac{2-3\beta'}{8}\left(1-\beta'-\frac{3}{4}\alpha'\right)^2 + 1\right), if\ \frac{9}{8}\alpha < \beta \end{cases}$$



Impact of Semantic OOD Data

Evaluation I: Linear Probing Error

- ID classifier of Misclassification in covariate OOD data

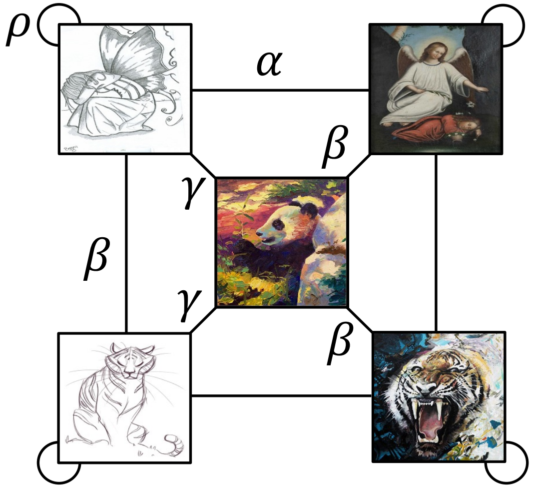
$$\varepsilon(f_2) = 0, if\ \alpha > 0, \beta > 0$$

Better OOD generalization performance!

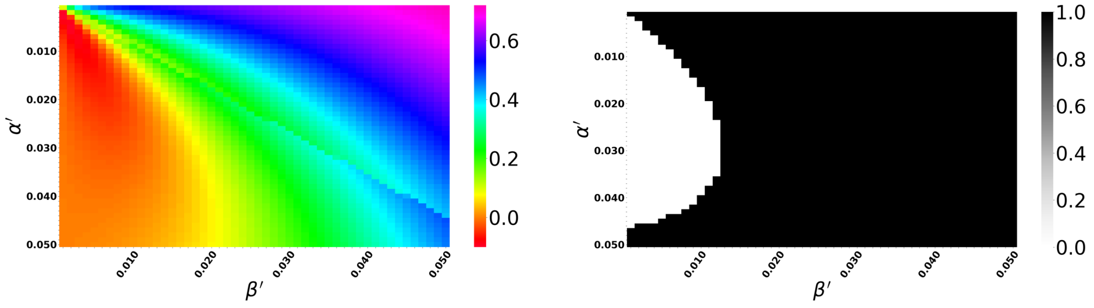
Evaluation II: Separability

- Euclidean distance between ID and semantic OOD data

$$\mathcal{S}(f_1) - \mathcal{S}\ (f_2) = \begin{cases} > 0, if\ \alpha', \beta' \in black\ areas \\ < 0, if\ \alpha', \beta' \in white\ areas \end{cases}$$



$$\mathcal{T} = \begin{bmatrix} \rho & \beta & \alpha & \gamma & \gamma \\ \beta & \rho & \gamma & \alpha & \gamma \\ \alpha & \gamma & \rho & \beta & \beta \\ \gamma & \alpha & \beta & \rho & \beta \\ \gamma & \gamma & \beta & \beta & \rho \end{bmatrix}$$



Heatmap of $\boldsymbol{S(f_1) - S(f_2)}$ $\quad$ $\mathbf{1}(\boldsymbol{S(f_1) - S(f_2)})$

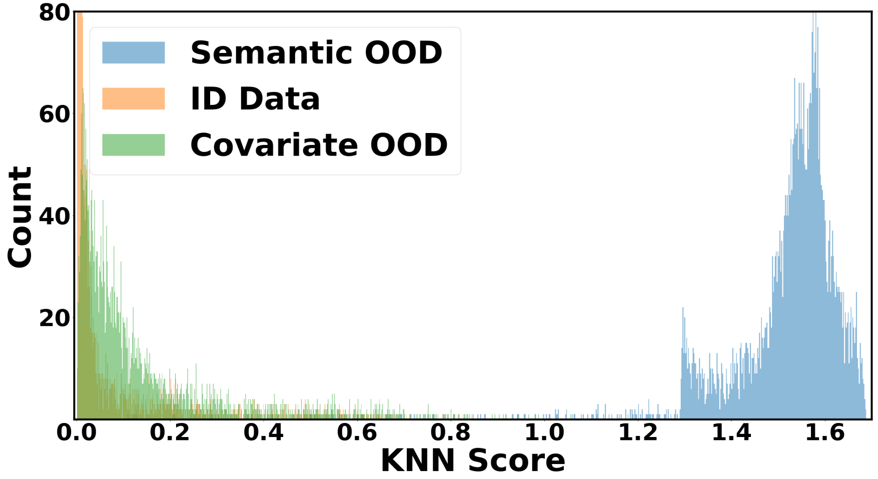
Empirical Experiments

Dataset setup:

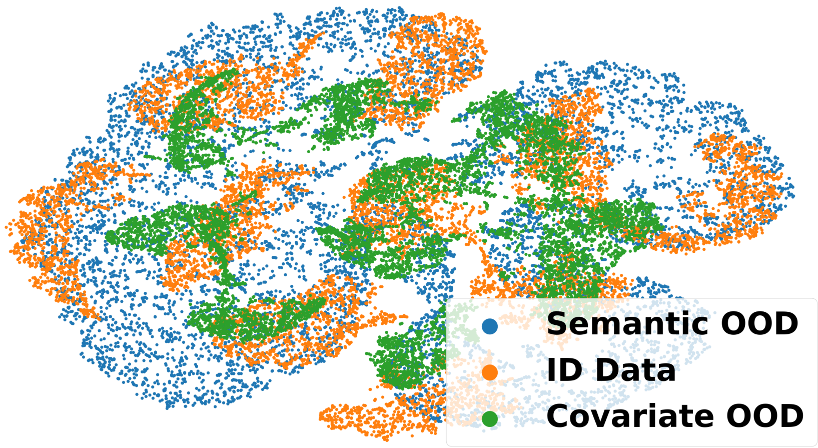
- ID: CIFAR-10. Covariate OOD: CIFAR-10-C. Semantic OOD: SVHN, LSUN-C, Textures, etc.

Method	SVHN $\mathbb{P}_{\text{out}}^{\text{semantic}}$, CIFAR-10-C $\mathbb{P}_{\text{out}}^{\text{covariate}}$				LSUN-C $\mathbb{P}_{\text{out}}^{\text{semantic}}$, CIFAR-10-C $\mathbb{P}_{\text{out}}^{\text{covariate}}$				Textures $\mathbb{P}_{\text{out}}^{\text{semantic}}$, CIFAR-10-C $\mathbb{P}_{\text{out}}^{\text{covariate}}$			
	OOD Acc.↑	ID Acc.↑	FPR↓	AUROC↑	OOD Acc.↑	ID Acc.↑	FPR↓	AUROC↑	OOD Acc.↑	ID Acc.↑	FPR↓	AUROC↑
OOD detection												
MSP	75.05	94.84	48.49	91.89	75.05	94.84	30.80	95.65	75.05	94.84	59.28	88.50
ODIN	75.05	94.84	33.35	91.96	75.05	94.84	15.52	97.04	75.05	94.84	49.12	84.97
Energy	75.05	94.84	35.59	90.96	75.05	94.84	8.26	98.35	75.05	94.84	52.79	85.22
Mahalanobis	75.05	94.84	12.89	97.62	75.05	94.84	39.22	94.15	75.05	94.84	15.00	97.33
ViM	75.05	94.84	21.95	95.48	75.05	94.84	5.90	98.82	75.05	94.84	29.35	93.70
KNN	75.05	94.84	28.92	95.71	75.05	94.84	28.08	95.33	75.05	94.84	39.50	92.73
ASH	75.05	94.84	40.76	90.16	75.05	94.84	2.39	99.35	75.05	94.84	53.37	85.63
OOD generalization												
ERM	75.05	94.84	35.59	90.96	75.05	94.84	8.26	98.35	75.05	94.84	52.79	85.22
IRM	77.92	90.85	63.65	90.70	77.92	90.85	36.67	94.22	77.92	90.85	59.42	87.81
GroupDRO	77.27	**94.97**	23.78	94.93	77.27	94.97	6.90	98.51	77.27	**94.97**	62.08	84.60
Mixup	79.17	93.30	97.33	18.78	79.17	93.30	52.10	76.66	79.17	93.30	58.24	75.70
VREx	76.90	91.35	55.92	91.22	76.90	91.35	51.50	91.56	76.90	91.35	65.45	85.46
EQRM	75.71	92.93	51.86	90.92	75.71	92.93	21.53	96.49	75.71	92.93	57.18	89.11
SharpDRO	79.03	94.91	21.24	96.14	79.03	94.91	5.67	98.71	79.03	94.91	42.94	89.99
Learning w. $\mathbb{P}_{wild}$												
OE	37.61	94.68	0.84	99.80	41.37	93.99	3.07	99.26	44.71	92.84	29.36	93.93
Energy (w. outlier)	20.74	90.22	0.86	99.81	32.55	92.97	2.33	**99.93**	49.34	94.68	16.42	96.46
Woods	52.76	94.86	2.11	99.52	76.90	**95.02**	1.80	99.56	83.14	94.49	39.10	90.45
Scone	84.69	94.65	10.86	97.84	84.58	93.73	10.23	98.02	**85.56**	93.97	37.15	90.91
SLW (Ours)	**86.62**$_{\pm 0.3}$	93.10$_{\pm 0.1}$	**0.13**$_{\pm 0.0}$	**99.98**$_{\pm 0.0}$	**85.88**$_{\pm 0.2}$	92.61$_{\pm 0.1}$	**1.76**$_{\pm 0.8}$	99.75$_{\pm 0.1}$	81.40$_{\pm 0.7}$	92.50$_{\pm 0.1}$	**12.05**$_{\pm 0.8}$	**98.25**$_{\pm 0.2}$

Further Analysis



(a) OOD detection score distribution



(b) T-SNE visualization of embeddings

[1] Sun, Yiyou, et al. "Out-of-distribution detection with deep nearest neighbors." *International Conference on Machine Learning*. PMLR, 2022.

Summary

- Propose a novel **graph-theoretic framework** for understanding both OOD generalization and detection

- Provide theoretic insight by analyzing **closed-form solutions for OOD generalization and detection error**

- Demonstrate **strong OOD generalization and detection capabilities** and provide empirical evidence of its robustness and **alignment with our theoretical analysis**