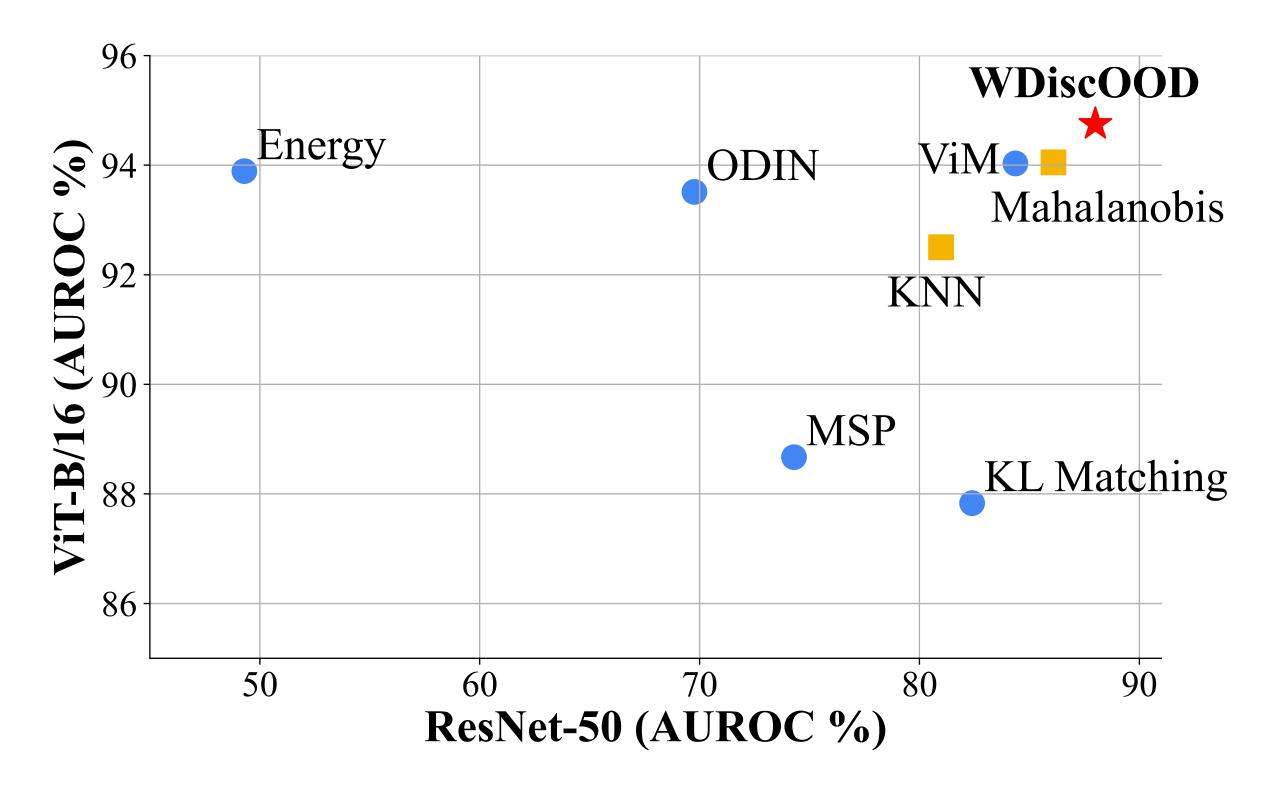
Robotics & Georgia Tech Intelligent Machines

Background & Goal

- Deep vision models prone to generate incorrect predictions when given unfamiliar data (Out-of-distribution, OOD) vrelative to the training data (In-distribution, ID)
- We study the OOD detection problem, where the goal is to develop a mechanism to distinguish between ID and OOD data
- We aim to jointly reason about classagnostic and class-specific information in the feature space

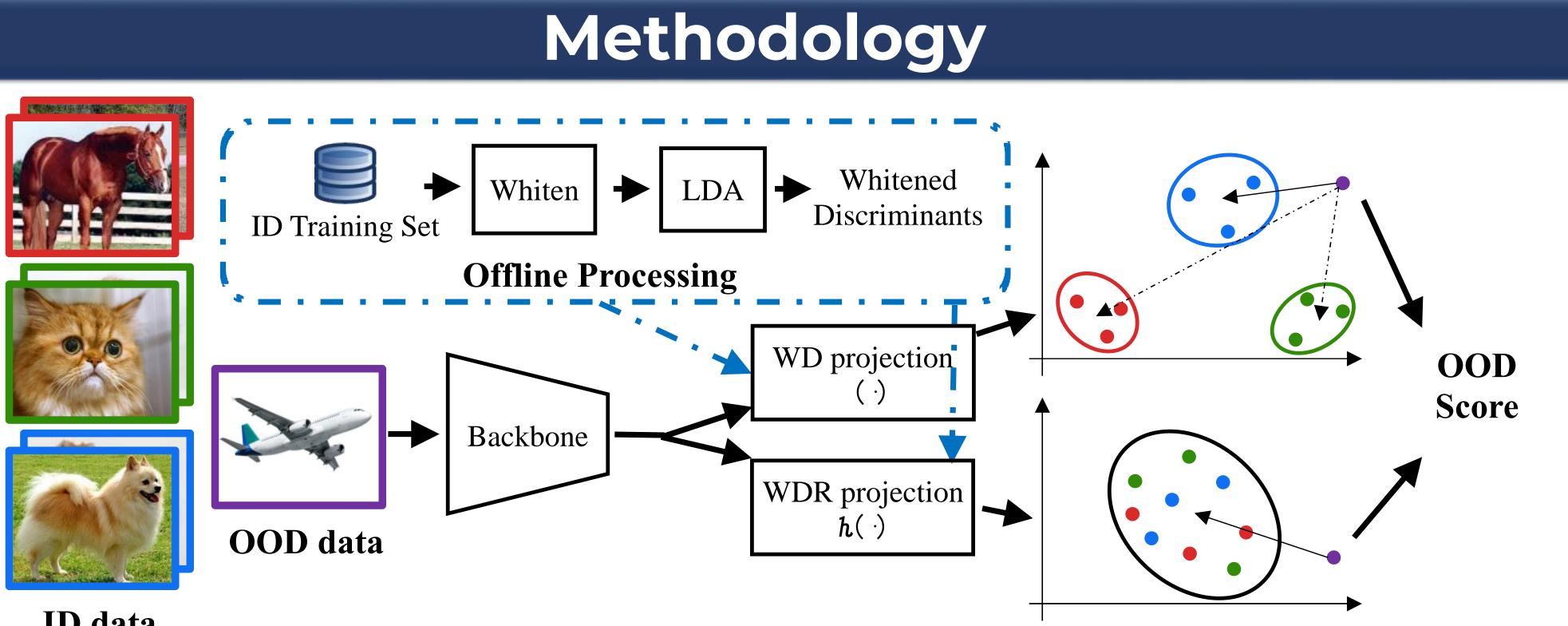
Contributions

- A new OOD scoring function based on Whitened Linear Discriminant Analysis (WLDA) in the feature space.
- A new insight on the efficacy of the Whitened Discriminative Residual (WDR) Subspace on OOD detection.
- New state-of-the-art results achieved on the large-scale ImagetNet OOD detection benchmark, under various settings including various visual classifiers (CNN & ViT) and contrastive visual encoders (SupCon & CLIP)



WDiscOOD: Out-of-Distribution Detection via Whitened Linear Discriminant Analysis

Yiye Chen, Yunzhi Lin, Ruinian Xu, Patricio A. Vela Institute for Robotics and Intelligent Machine, Georgia Tech



ID data

1. Data whitening

	 $c^{-1/2}$	-
${\mathcal X}$	$S_{Z,W}$	Z

- χ
- Whitened feature Original feature
- Covariance matrix for z
- 3. OOD score

9	$v(x) = W^T x$
$g(\cdot)$	Whitened Discri
$h(\cdot)$	Whitened Discri
W	Stack of top disc
Q	Eigenvalues of <i>V</i>

 $s_g(x) = -\min_{C} ||g(x) - \mu_{C}^{WD}||_2; \ s_h(x) = -||h(x) - \mu^{WD}||_2$

Results – Classifiers (ResNet-50 & ViT)

											0			
	Text			UN		aces		ıralist	0	Net-O	-	Img-O		rage
Method	FPR95↓A	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95	AUROC↑
Classifier-dependent methods														
MSP [21]	72.98	74.92	70.98	78.75	<u>73.43</u>	76.65	60.90	84.40	95.65	53.13	69.73	81.17	73.94	74.84
Energy [33]	95.74	48.60	97.93	50.12	97.77	48.90	98.12	50.86	92.80	48.23	95.41	52.33	96.30	49.84
ODIN [32]	75.94	69.33	75.51	74.05	77.54	71.28	68.60	79.88	94.95	51.19	73.98	76.15	77.75	70.31
MaxLogit [20]	75.92	69.33	75.51	74.05	77.55	71.28	68.57	79.88	94.95	51.19	73.97	76.15	77.74	70.31
KLMatch [20]	57.57	86.09	70.36	<u>82.91</u>	74.04	<u>80.65</u>	46.83	90.81	89.75	68.86	58.21	88.31	66.13	82.94
ReAct [40]	98.05	34.51	99.66	23.68	99.80	22.86	100.00	23.13	99.40	37.31	99.86	23.86	99.46	27.56
ViM [44]	<u>25.18</u>	<u>92.63</u>	<u>69.22</u>	81.39	74.90	76.40	<u>30.02</u>	<u>93.38</u>	<u>76.15</u>	<u>77.08</u>	<u>46.70</u>	88.60	53.70	<u>84.91</u>
					Fea	ture spa	ice met	hods						
Maha [<mark>31</mark>]	31.17	91.62	66.29	<u>84.31</u>	70.27	<u>81.45</u>	25.64	<u>95.38</u>	81.45	75.65	44.36	91.41	53.20	86.64
KNN [41]	23.26	93.11	88.59	74.01	89.00	71.07	74.60	85.83	71.05	81.15	70.29	84.01	69.47	81.53
WDiscOOD	<u>29.20</u>	<u>91.90</u>	56.83	86.74	64.40	83.13	22.39	95.59	81.60	75.52	<u>44.67</u>	<u>90.51</u>	49.85	87.23
	Text	ures	SU	UN	Pla	aces	iNatu	ıralist	Imgl	Net-O	Open	[mg-O	Ave	rage
Method					_	aces AUROC↑			0		-	0		0
Method				AUROC↑	FPR95↓		FPR95↓	AUROC↑	FPR95↓		-	0		0
Method MSP [21]	FPR95↓A		FPR95↓	AUROC↑	FPR95↓ Classif	AUROC↑	FPR95↓. endent i	AUROC↑ method	FPR95↓	AUROC↑	-	AUROC↑		AUROC↑
	FPR95↓A	AUROC↑ 85.42	FPR95↓. 53.22	AUROC↑	FPR95↓ Classif 57.75	AUROC↑ i er-depe	FPR95↓. endent 1 13.66	AUROC↑ method	FPR95↓. s 51.75	AUROC↑	FPR95↓/	AUROC↑ 92.48	FPR95↓4	AUROC↑ 88.89
MSP [21]	FPR95↓A 52.43 36.13	AUROC↑ 85.42	FPR95↓. 53.22	AUROC↑ 86.93 <u>93.28</u>	FPR95↓ Classifi 57.75 42.80	AUROC↑ i er-depe 85.72	FPR95↓. endent 1 13.66	AUROC↑ method 97.00	FPR95↓ s 51.75 <u>30.30</u>	AUROC↑ 85.81	FPR95↓/ 31.99	AUROC↑ 92.48 96.87	FPR95↓4	AUROC↑ 88.89 94.11
MSP [21] Energy [33]	FPR95↓A 52.43 36.13 38.57	AUROC↑ 85.42 91.25 90.86	FPR95↓ 53.22 <u>34.44</u> 37.45	AUROC↑ 86.93 <u>93.28</u> 92.81	FPR95↓ Classifi 57.75 42.80 44.68	AUROC↑ ier-depe 85.72 90.98	FPR95↓. endent 1 13.66 5.60 6.03	AUROC↑ method 97.00 98.94	FPR95↓ s 51.75 <u>30.30</u> 33.50	AUROC↑ 85.81 93.36	FPR95↓/ 31.99 16.06	AUROC↑ 92.48 96.87 96.54	FPR95↓4 43.47 27.56	AUROC↑ 88.89 94.11 93.73
MSP [21] Energy [33] ODIN [32]	FPR95↓A 52.43 36.13 38.57 38.56	AUROC↑ 85.42 91.25 90.86 90.86	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81	FPR95↓ Classifi 57.75 42.80 44.68 44.68	AUROC↑ ier-depe 85.72 90.98 <u>90.66</u>	FPR95↓. endent 1 13.66 5.60 6.03 6.03	AUROC↑ method 97.00 98.94 98.81	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50	AUROC↑ 85.81 93.36 92.69	FPR95↓4 31.99 16.06 17.83	AUROC↑ 92.48 96.87 96.54 96.54	FPR95↓4 43.47 27.56 29.68	AUROC↑ 88.89 94.11 93.73 93.73
MSP [21] Energy [33] ODIN [32] MaxLogit [20]	FPR95↓A 52.43 36.13 38.57 38.56	AUROC↑ 85.42 91.25 90.86 90.86 85.12	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45 56.04	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81	FPR95↓ Classifi 57.75 42.80 44.68 44.68 61.08	AUROC↑ ier-depe 85.72 90.98 <u>90.66</u> 90.66	FPR95↓. endent 1 13.66 5.60 6.03 6.03	AUROC↑ method 97.00 98.94 98.81 98.81 96.32	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50 49.90	AUROC↑ 85.81 93.36 92.69 92.69	FPR95↓4 31.99 16.06 17.83 17.83	AUROC↑ 92.48 96.87 96.54 96.54 91.93	FPR95↓4 43.47 27.56 29.68 29.68	AUROC↑ 88.89 94.11 93.73 93.73 88.05
MSP [21] Energy [33] ODIN [32] MaxLogit [20] KLMatch [20]	FPR95↓A 52.43 36.13 38.57 38.56 51.22 36.35	AUROC↑ 85.42 91.25 90.86 90.86 85.12 91.17	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45 56.04 34.55	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81 85.45 93.22	FPR95↓ Classifi 57.75 42.80 44.68 44.68 61.08 43.32	AUROC↑ ier-depe 85.72 90.98 <u>90.66</u> <u>90.66</u> 83.86	FPR95↓. endent 1 13.66 5.60 6.03 13.68 5.61	AUROC↑ method: 97.00 98.94 98.81 98.81 96.32 98.94	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50 49.90 <u>30.30</u>	AUROC↑ 85.81 93.36 92.69 92.69 85.62 93.40	FPR95↓4 31.99 16.06 17.83 17.83 31.38 <u>16.01</u>	AUROC↑ 92.48 96.87 96.54 96.54 91.93	FPR95↓4 43.47 27.56 29.68 43.88 27.69	AUROC↑ 88.89 94.11 93.73 93.73 88.05 94.07
MSP [21] Energy [33] ODIN [32] MaxLogit [20] KLMatch [20] ReAct [40]	FPR95↓A 52.43 36.13 38.57 38.56 51.22 36.35	AUROC↑ 85.42 91.25 90.86 90.86 85.12 91.17	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45 56.04 34.55	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81 85.45 93.22	FPR95↓ Classifi 57.75 42.80 44.68 44.68 61.08 <u>43.32</u> 44.23	AUROC↑ ier-depe 85.72 90.98 <u>90.66</u> <u>90.66</u> 83.86 <u>90.83</u>	FPR95↓ 2 endent 1 13.66 5.60 6.03 6.03 13.68 5.61 <u>1.40</u>	AUROC↑ nethod 97.00 98.94 98.81 98.81 96.32 98.94 <u>99.68</u>	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50 49.90 <u>30.30</u>	AUROC↑ 85.81 93.36 92.69 92.69 85.62 93.40	FPR95↓4 31.99 16.06 17.83 17.83 31.38 <u>16.01</u>	AUROC↑ 92.48 96.87 96.54 96.54 91.93 96.88	FPR95↓4 43.47 27.56 29.68 43.88 27.69	AUROC↑ 88.89 94.11 93.73 93.73 88.05 94.07
MSP [21] Energy [33] ODIN [32] MaxLogit [20] KLMatch [20] ReAct [40]	FPR95↓A 52.43 36.13 38.57 38.56 51.22 36.35	AUROC↑ 85.42 91.25 90.86 90.86 85.12 91.17 <u>91.38</u>	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45 56.04 34.55 32.47	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81 85.45 93.22	FPR95↓ Classifi 57.75 42.80 44.68 44.68 61.08 <u>43.32</u> 44.23 Fea	AUROC↑ ier-depe 85.72 90.98 90.66 90.66 83.86 90.83 89.86	FPR95↓ endent 1 13.66 5.60 6.03 13.68 5.61 <u>1.40</u> ace met	AUROC↑ methods 97.00 98.94 98.81 98.81 96.32 98.94 <u>99.68</u> hods	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50 <u>49.90</u> <u>30.30</u> <u>31.80</u>	AUROC↑ 85.81 93.36 92.69 92.69 85.62 93.40	FPR95↓4 31.99 16.06 17.83 17.83 31.38 <u>16.01</u> 16.61	AUROC↑ 92.48 96.87 96.54 96.54 91.93 96.88 <u>97.10</u>	FPR95↓4 43.47 27.56 29.68 43.88 27.69 <u>27.53</u>	AUROC↑ 88.89 94.11 93.73 93.73 88.05 94.07 <u>94.25</u>
MSP [21] Energy [33] ODIN [32] MaxLogit [20] KLMatch [20] ReAct [40] ViM [44]	FPR95↓A 52.43 36.13 38.57 38.56 51.22 36.35 38.67	AUROC↑ 85.42 91.25 90.86 90.86 85.12 91.17 <u>91.38</u> <u>91.67</u>	FPR95↓ 53.22 <u>34.44</u> 37.45 37.45 56.04 34.55 32.47 35.37	AUROC↑ 86.93 <u>93.28</u> 92.81 92.81 85.45 93.22 93.41	FPR95↓ Classifi 57.75 42.80 44.68 44.68 61.08 <u>43.32</u> 44.23 Fea 46.08	AUROC↑ ier-depe 85.72 90.98 90.66 90.66 83.86 90.83 89.86 ture spa	FPR95↓. endent 1 13.66 5.60 6.03 13.68 5.61 <u>1.40</u> ice met <u>0.96</u>	AUROC↑ methods 97.00 98.94 98.81 98.81 96.32 98.94 <u>99.68</u> hods <u>99.78</u>	FPR95↓ s 51.75 <u>30.30</u> 33.50 33.50 49.90 <u>30.30</u> <u>31.80</u> 30.45	AUROC↑ 85.81 93.36 92.69 92.69 85.62 93.40 <u>94.05</u>	FPR95↓4 31.99 16.06 17.83 17.83 31.38 <u>16.01</u> 16.61 13.85	AUROC↑ 92.48 96.87 96.54 96.54 91.93 96.88 <u>97.10</u> 97.50	FPR95↓A 43.47 27.56 29.68 43.88 27.69 <u>27.53</u>	AUROC↑ 88.89 94.11 93.73 93.73 88.05 94.07 94.25

WDiscOOD achieves superior results compared to a large set of baselines for ImageNet classifiers with various backbones including ResNet-50 and Vision Transformer (ViT)

2. Discriminative & Residual Decomposition

 $h(x) = (I - QQ^T)x$

riminative (WD) space projection riminative Residual (WDR) space projection criminants in x space

$$\left| \frac{WDR}{2} \right|_{2}; \quad s(x) = s_{g}(x) + \alpha s_{h}(x)$$

Results – Contrastive Models (SupCon & CLIP)

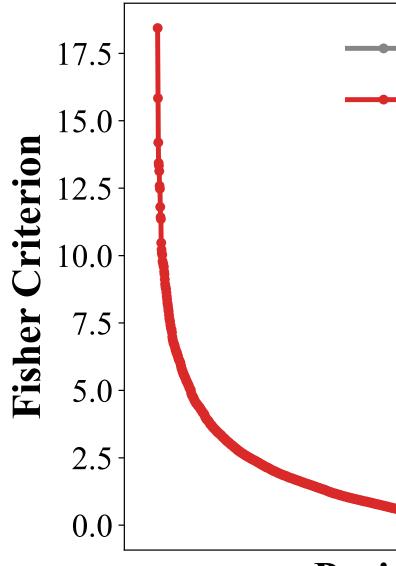
Method

Mahalan KNN **WDisc**

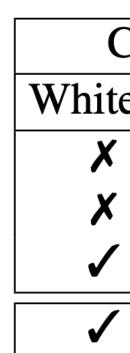
- model on the ImageNet dataset.

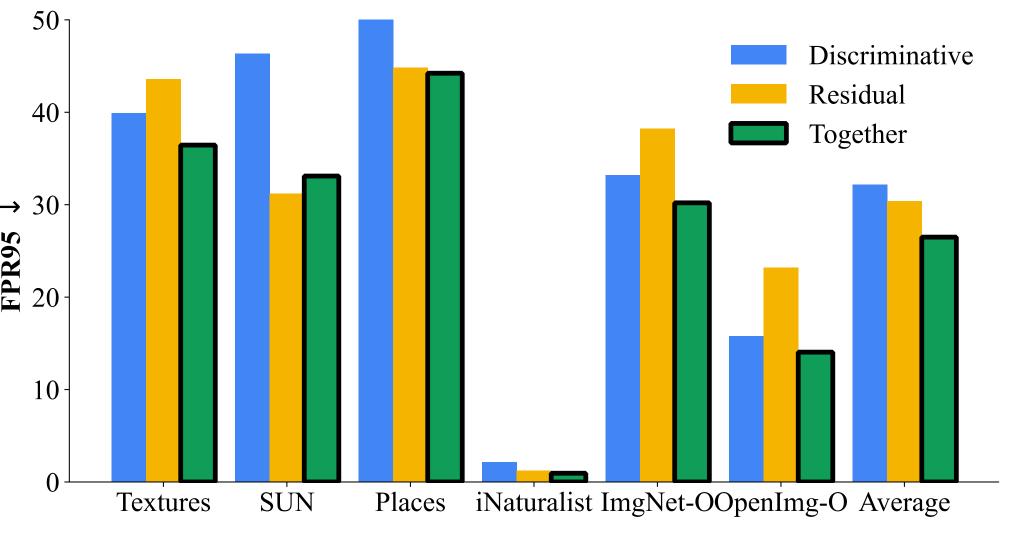
Method Understanding & Ablation Study

The separation of class-agnostic and class specific information



The importance of data whitening for OOD detection





Evaluate under FPR95, which is the lower the better



a	SupC	on [27]	CLIP [36]			
d	FPR95↓	AUROC↑	FPR95↓	AUROC↑		
nobis	46.95	89.78	78.00	75.31		
	42.51	90.35	82.59	67.22		
OOD	40.10	90.89	77.57	75.74		

• WDiscOOD is applicable for contrastive models as it is a feature space method that does not rely on any task head.

• It outperforms other feature space methods for SupCon and CLIP

- Discriminant Orthogonals **Discriminants**
- Fisher Criterion values are lower for discriminant orthogonals than discriminants
- This suggests that the feature projections in WD space are maximally separated into classes, and are closely clustered in WDR space.

Projection Direction

Cont	fig	Resl	Net-50	ViT			
ten [†]	Dist	FPR95↓	AUROC ↑	FPR95↓	AUROC ↑		
,	Maha	53.65	86.20	29.81	93.47		
,	Eucl	74.56	81.17	32.21	93.52		
•	Maha	49.85	87.23	26.60	94.40		
•	Eucl	49.86	87.23	26.49	94.41		

Feature whitening greatly improves the performance of feature-distance-based OOD detection, regardless of the distance type.

► WDR and WD spaces are complementary

- The integrated score $s(\cdot)$ performs better than individual components $s_q(\cdot)$ and $s_h(\cdot)$.
- Residual space is more critical than the discriminant space.