

DMQA Open Seminar

OOD Detection for Image Classification Part5 : NC-based Approach

2026. 09. 11

고려대학교 산업경영공학과

Data Mining & Quality Analytics Lab.

임새린

발표자 소개



❖ 임새린 (Saerin Lim)

- 고려대학교 산업경영공학과 Data Mining & Quality Analytics Lab.
- Ph.D. Student (2021.03 ~ Present)
- 지도 교수: 김성범 교수님

❖ Research Interest

- Self- and Semi-supervised learning, OOD detection

❖ Contact

- E-mail : momo_om@korea.ackr

목차

Contents

1. OOD Detection
2. Neural Collapse
3. Method – NECO, D-KNN
4. Conclusion

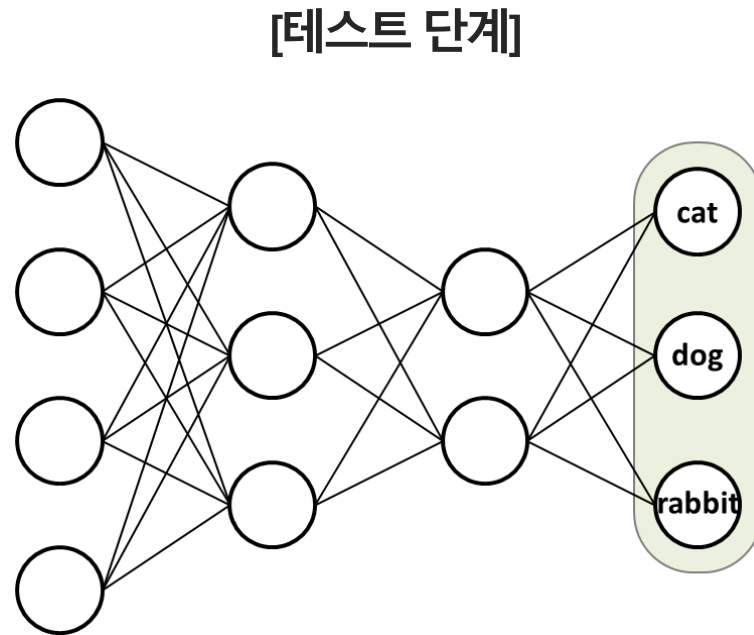
OOD Detection

Problem Definition

- ❖ 일반적으로 분류 모델은 테스트 단계에서 **모델이 학습한 ID 클래스 중 하나의 샘플이 들어올 것을 가정**



테스트 샘플



고양이, 강아지, 토끼를
분류하는 모델



softmax



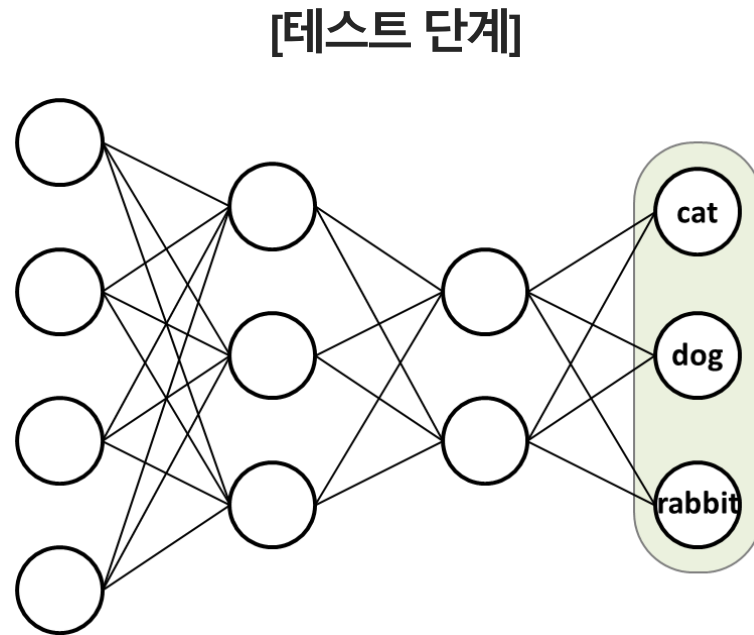
OOD Detection

Problem Definition

- ❖ 일반적으로 분류 모델은 테스트 단계에서 **모델이 학습한 ID 클래스 중 하나의 샘플이** 들어올 것을 가정
- ❖ 하지만 현실에서는 **학습한 클래스와 다른 OOD 샘플이** 들어올 수 있음



OOD sample



고양이, 강아지, 토끼를
분류하는 모델

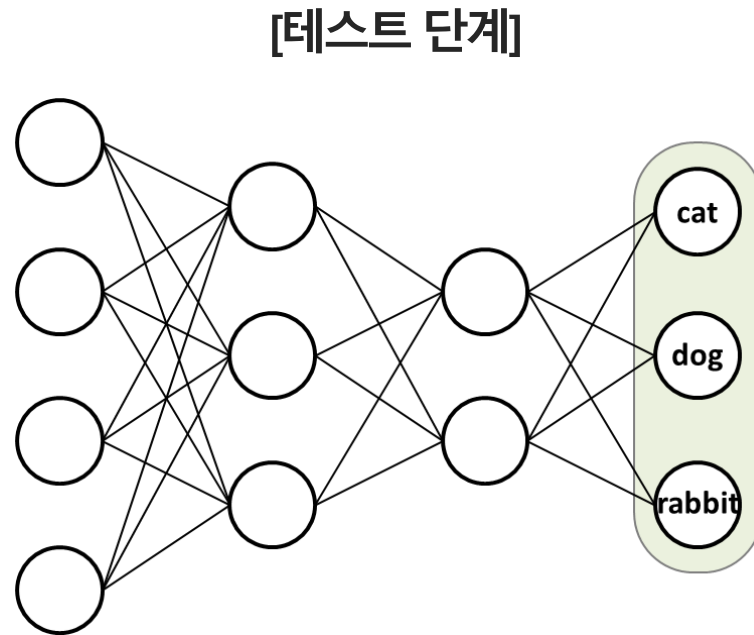
OOD Detection

Problem Definition

- ❖ 하지만 모델은 ID class의 개수에 맞게 학습이 되어, OOD 샘플이 들어오더라도 무조건 **하나의 ID 클래스**로 예측함



OOD sample



토끼로 예측

OOD Detection

Problem Definition

❖ 하지만 모델은 ID class의 개수에 맞게 학습이 되어, OOD 샘플이 들어오더라도 무조건 하나의 ID 클래스로 예측함

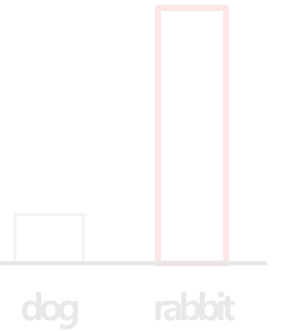
[테스트 단계]

OOD detection의 목적

분류 모델은 무조건 하나의 ID class로 예측하니,
ID와 OOD를 구분할 수 있는 스코어 함수를 찾자!



OOD sample



토끼로 예측

고양이, 강아지, 토끼를
분류하는 모델

OOD Detection

Overconfidence Problem

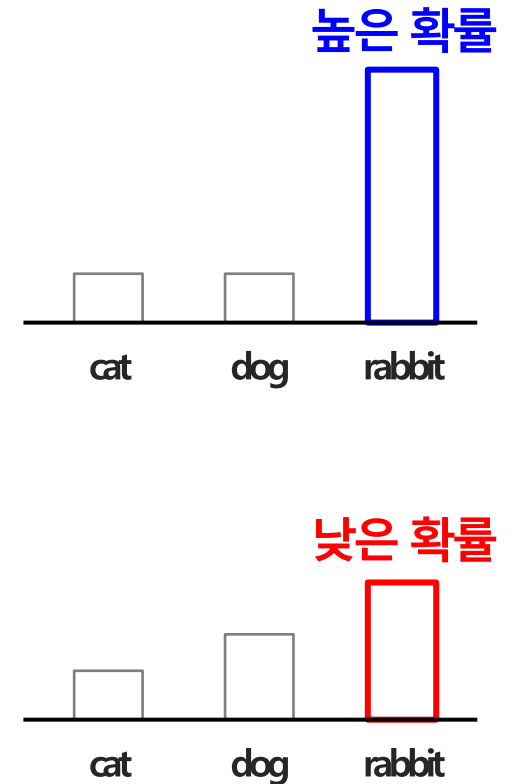
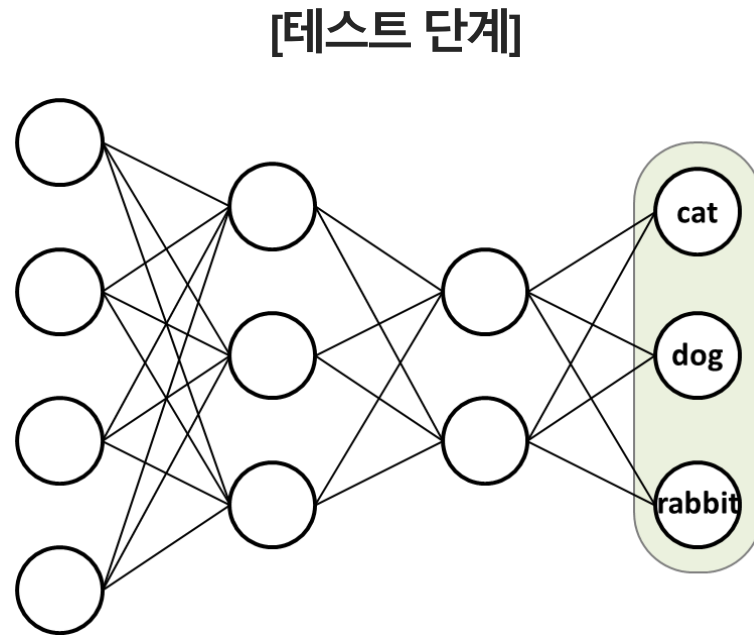
- ❖ OOD 탐지 문제를 처음 다룬 연구에서 **최대 확률값 (Maximum Softmax Prob. MSP)**으로 OOD 탐지가 가능함을 보임



ID sample



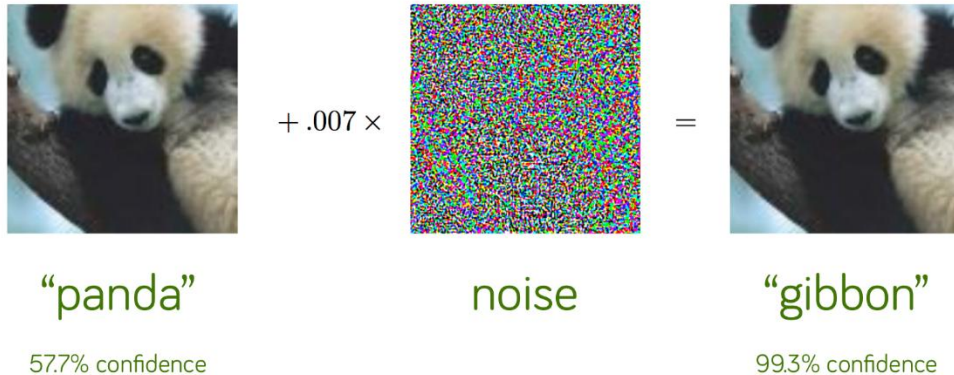
OOD sample



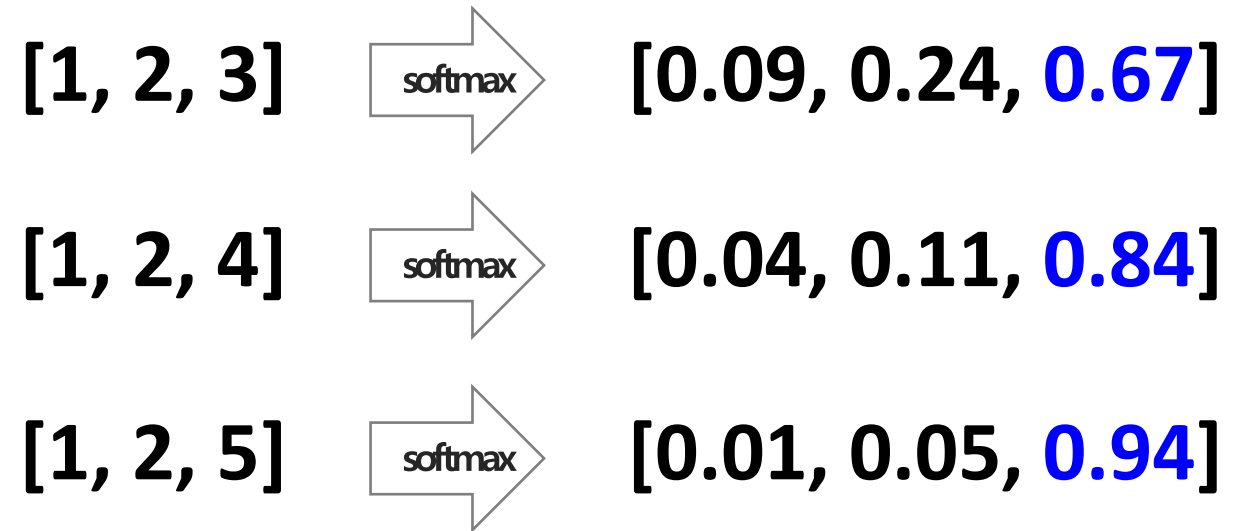
OOD Detection

Overconfidence Problem

- ❖ 하지만 신경망 모델은 작은 변화에도 logit이 크게 바뀌게 되며, softmax 함수는 이 변화를 증폭시키는 역할을 함
- ❖ 즉, OOD 샘플이 들어오더라도 우연히 높은 최대확률값을 가지는 경우가 많음 (Overconfidence Problem)



Noise에 민감한 신경망 모델



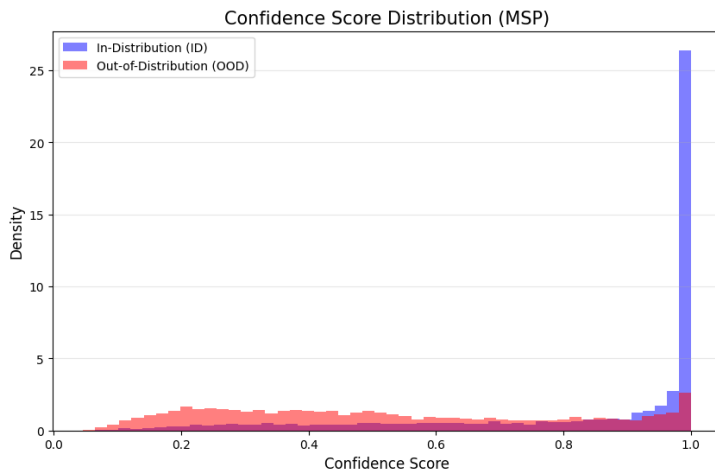
Noise를 증폭시키는 Softmax 함수

OOD Detection

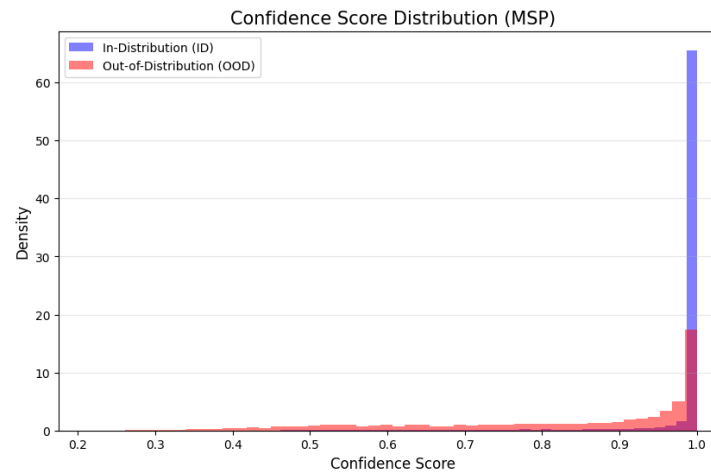
Overconfidence Problem

- ❖ 하지만 신경망 모델은 작은 변화에도 logit이 크게 바뀌게 되며, softmax 함수는 이 변화를 증폭시키는 역할을 함
- ❖ 즉, OOD 샘플이 들어오더라도 우연히 높은 최대확률값을 가지는 경우가 많음 (Overconfidence Problem)

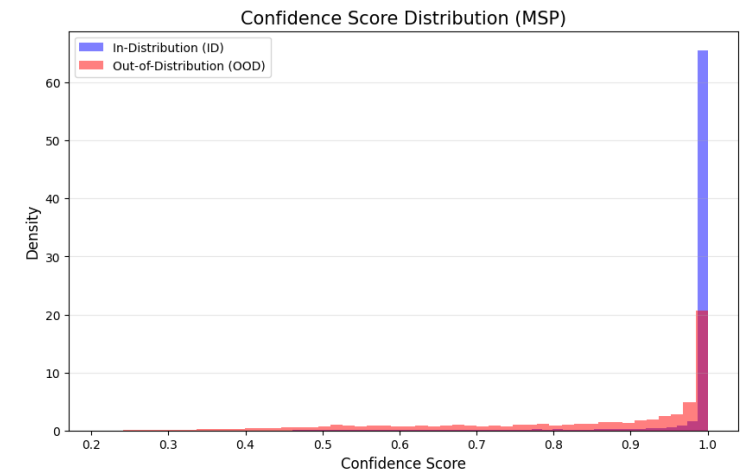
ID Dataset = CIFAR10



OOD = MNIST



OOD = Texture



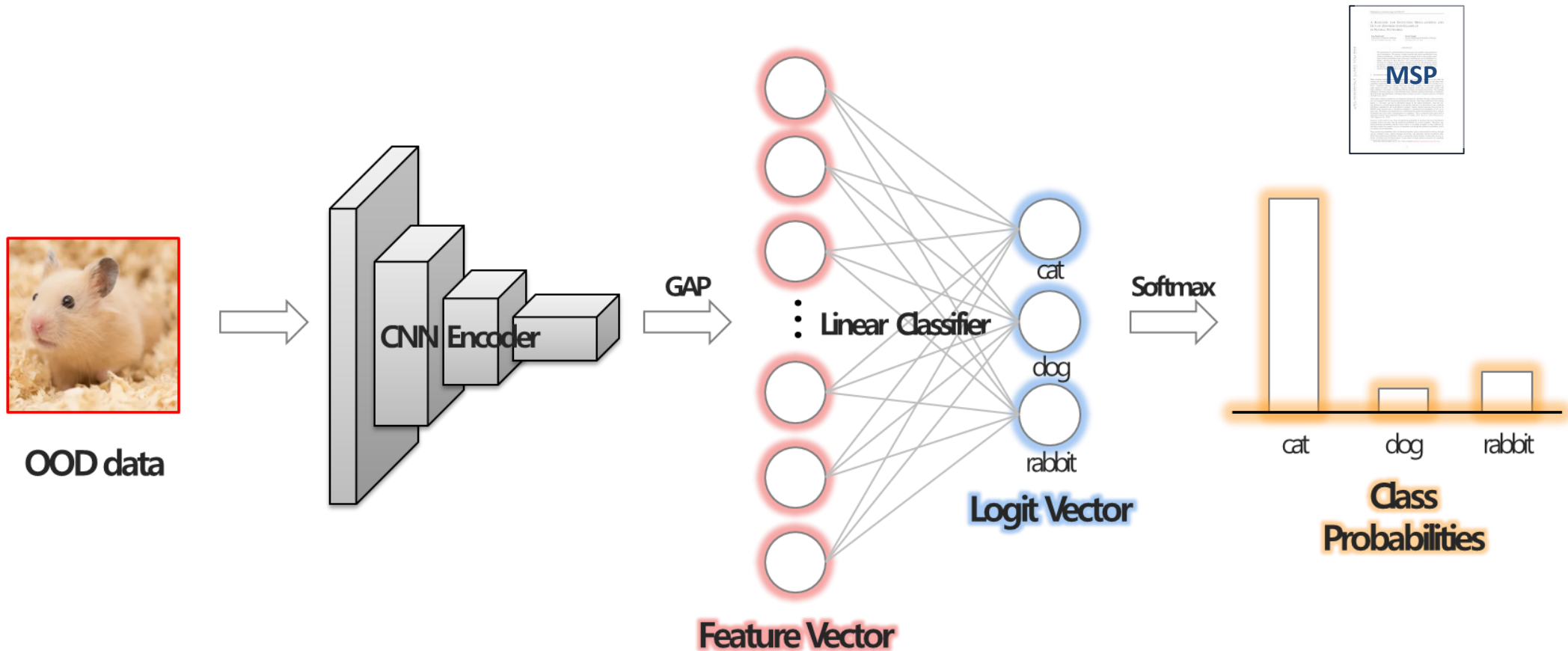
OOD = CIFAR100



OOD Detection

OOD Detection Approach

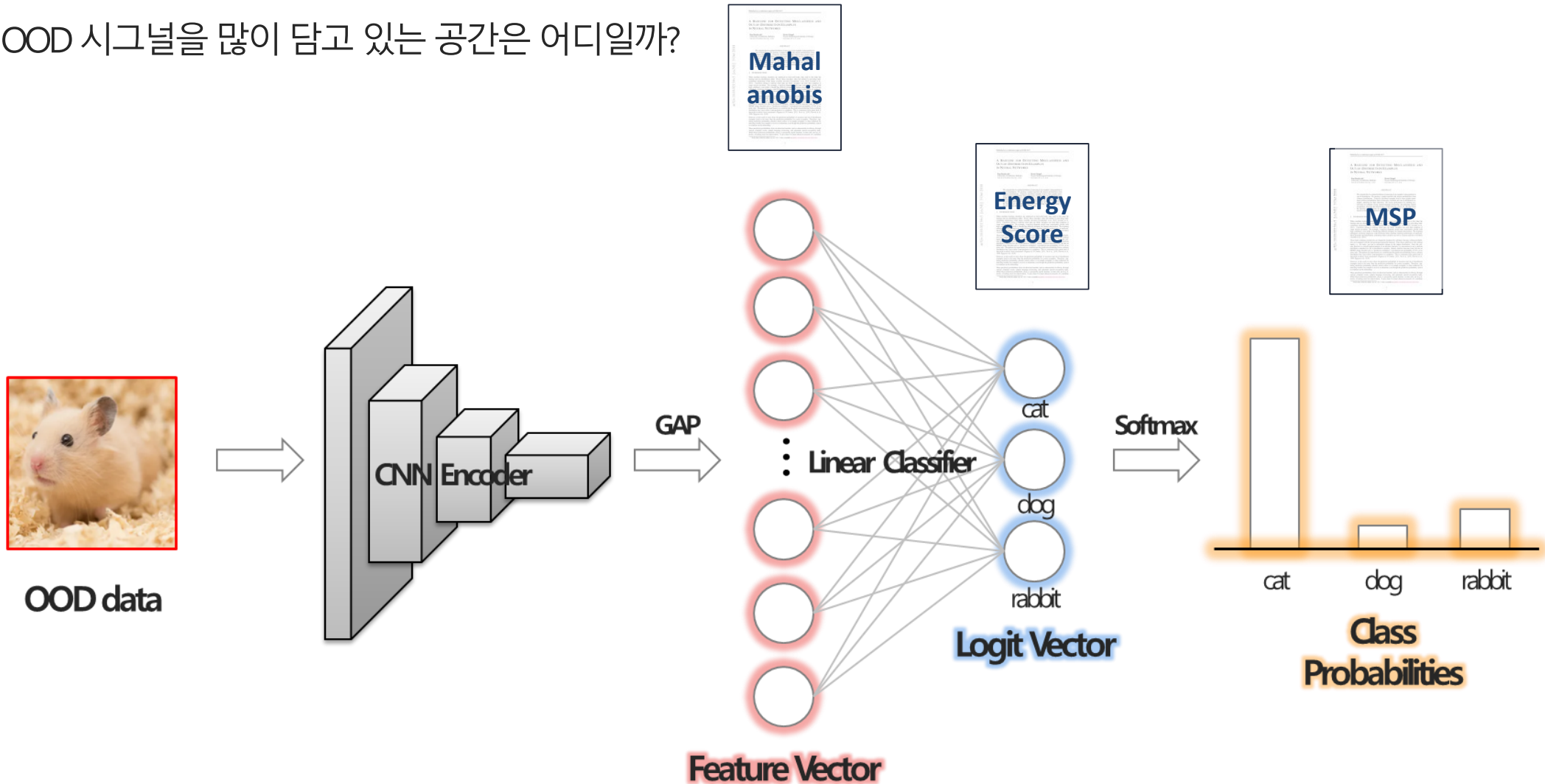
- ❖ OOD 시그널을 많이 담고 있는 공간은 어디일까?



OOD Detection

OOD Detection Approach

❖ OOD 시그널을 많이 담고 있는 공간은 어디일까?



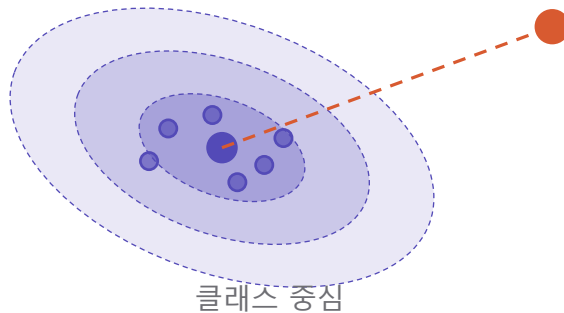
OOD Detection

Distance-based OOD Detection in Feature Space

- ❖ 특징 공간에서 ID 샘플이 이루는 구조를 기준으로, 테스트 샘플이 그 구조에서 얼마나 떨어져 있는지를 활용
- ❖ 구조를 확률 분포로 가정하는 파라메트릭 방식과, 가정 없이 이웃까지의 거리로 추정하는 논파라메트릭 방식으로 나뉨
- ❖ 이러한 방법론들은 ID 샘플이 어떤 구조를 이루는지를 가정하거나 추정할 뿐, 설명하지 못함

Mahalanobis

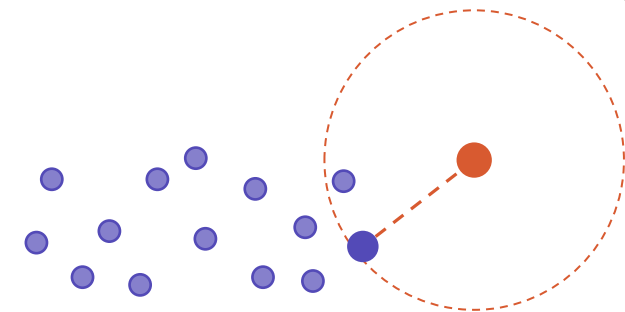
파라메트릭



클래스별 Gaussian 분포를 가정하고, 공분산으로 방향별 스케일을 보정한 거리를 OOD score로 사용

KNN

논파라메트릭



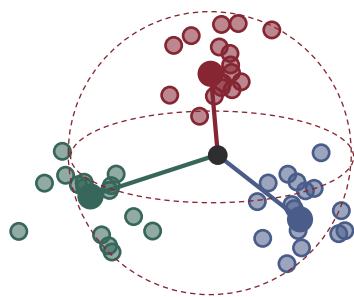
분포 가정 없이, 학습 feature vector에서 k번째 최근접 이웃까지의 거리로 국소 밀도를 직접 추정

Neural Collapse

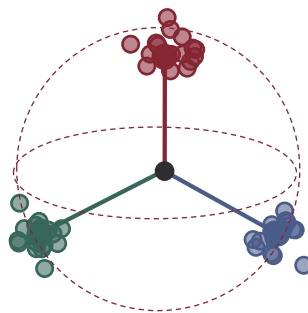
Neural Collapse

- ❖ Neural Collapse (NC)란 분류 모델 학습의 마지막 단계에 도달했을 때, 특징 공간이 **특정 기하학적 구조로 수렴**하는 현상 [1]
- ❖ 같은 클래스의 특징은 클래스 중심 벡터로 수렴하고, 각 클래스 중심 벡터들은 등각·등크기를 만족
- ❖ 즉, D차원의 특징 공간에서 실제로 ID 샘플이 존재하는 공간은 C-1차원에 집중되어 있음
- ❖ 이 Simplex ETF 구조가 특정 조건 하에서 Cross-Entropy Loss를 활용한 분류 모델의 전역 최적해라는 것이 증명 [2]

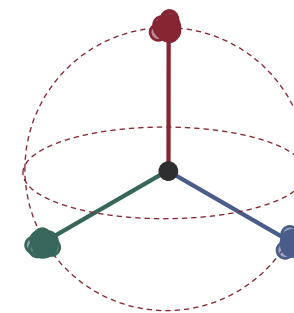
① 학습 초기 단계



② 학습 중반 단계



③ 학습 마지막 단계



Simplex Equiangular Tight Frame (ETF) 구조

[1] Papayan, V., Han, X. Y., & Donoho, D. L. (2020). Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40), 24652-24663.

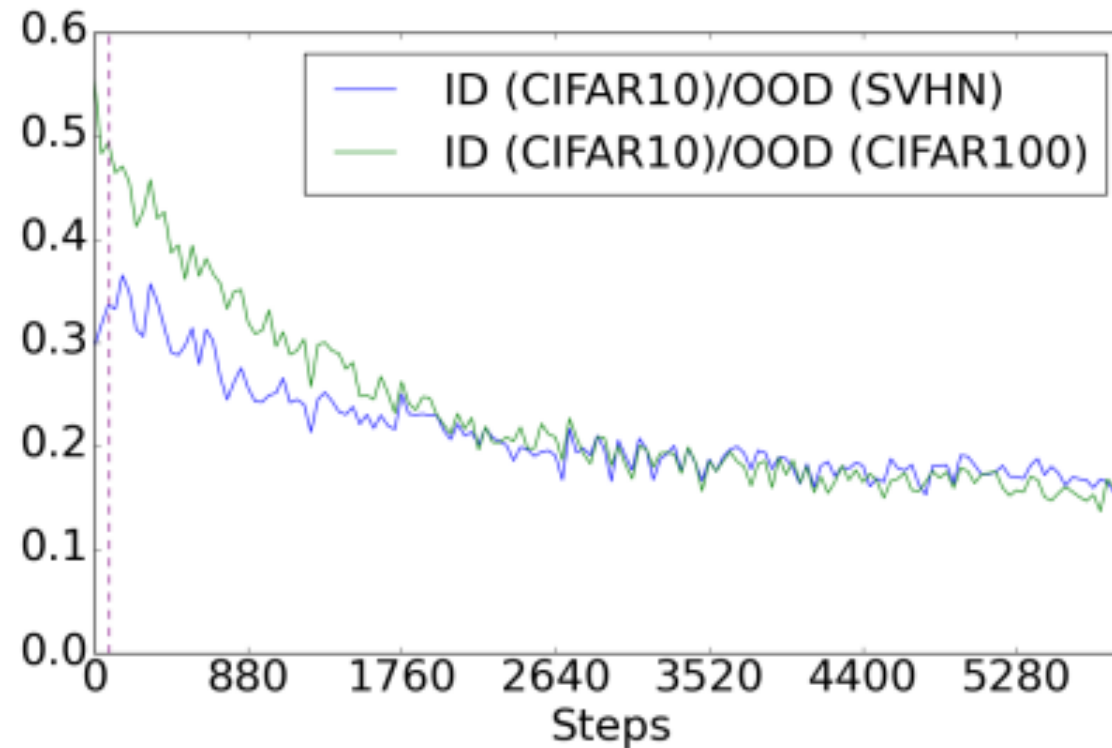
[2] Zhu, Z., Ding, T., Zhou, J., Li, X., You, C., Sulam, J., & Qu, Q. (2021). A geometric analysis of neural collapse with unconstrained features. *Advances in Neural Information Processing Systems*, 34, 29820-29834.

NECO: NEural Collapse Based Out-of-distribution detection (2024, ICLR)

NECO

Motivation

- ❖ Simplex ETF 공간이 ID 샘플의 영역이라면 OOD 샘플은 어디에 있을까?
- ❖ 학습이 진행될수록 OOD 샘플의 중심 벡터가 ID 클래스의 중심 벡터들과 수직을 이루는 방향으로 수렴하는 것을 발견

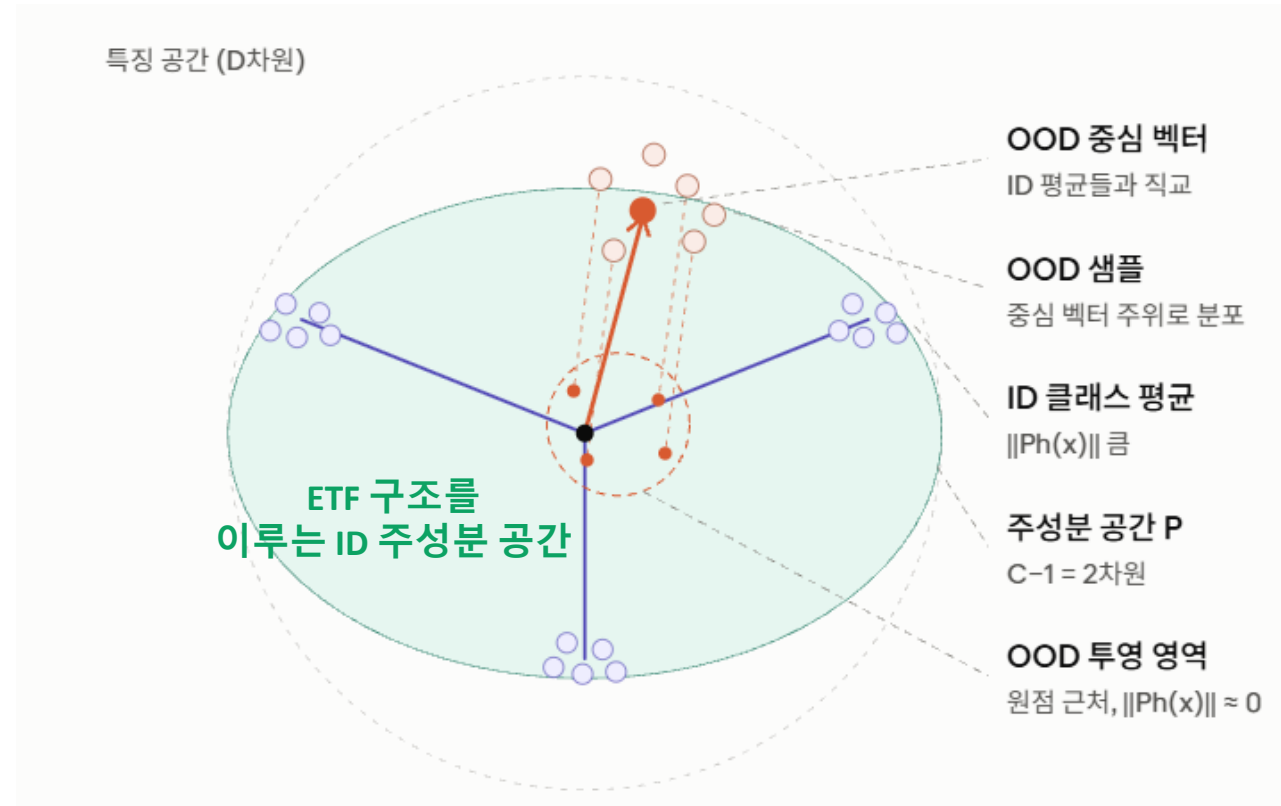


ID 클래스 중심 벡터와 OOD 샘플 중심 벡터의 평균 유사도

NECO

Motivation

- ❖ OOD 중심 벡터가 ID 중심 벡터들과 수직이라면, OOD 중심 벡터를 ID 주성분 공간에 투영했을 때 원점으로 수렴해야 함
- ❖ 개별 OOD 샘플도 이 경향성을 따라서 ID 주성분 공간에 투영한다면 원점 근처에 존재하게 될 것



NECO

Motivation

- ❖ OOD 중심 벡터가 ID 중심 벡터들과 수직이라면, OOD 중심 벡터를 ID 주성분 공간에 투영했을 때 원점으로 수렴해야 함
- ❖ 개별 OOD 샘플도 이 경향성을 따라서 ID 주성분 공간에 투영한다면 원점 근처에 존재하게 될 것

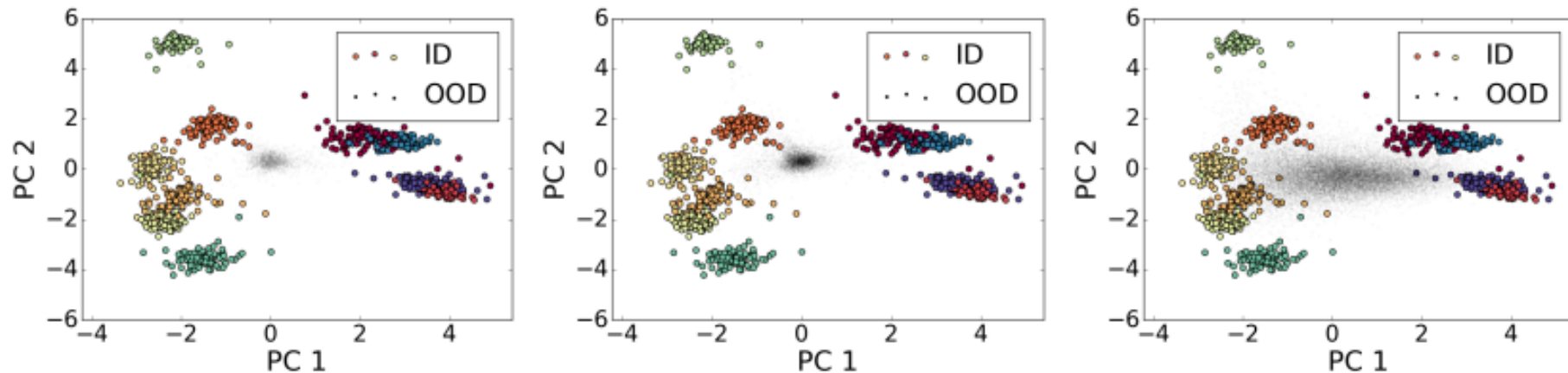
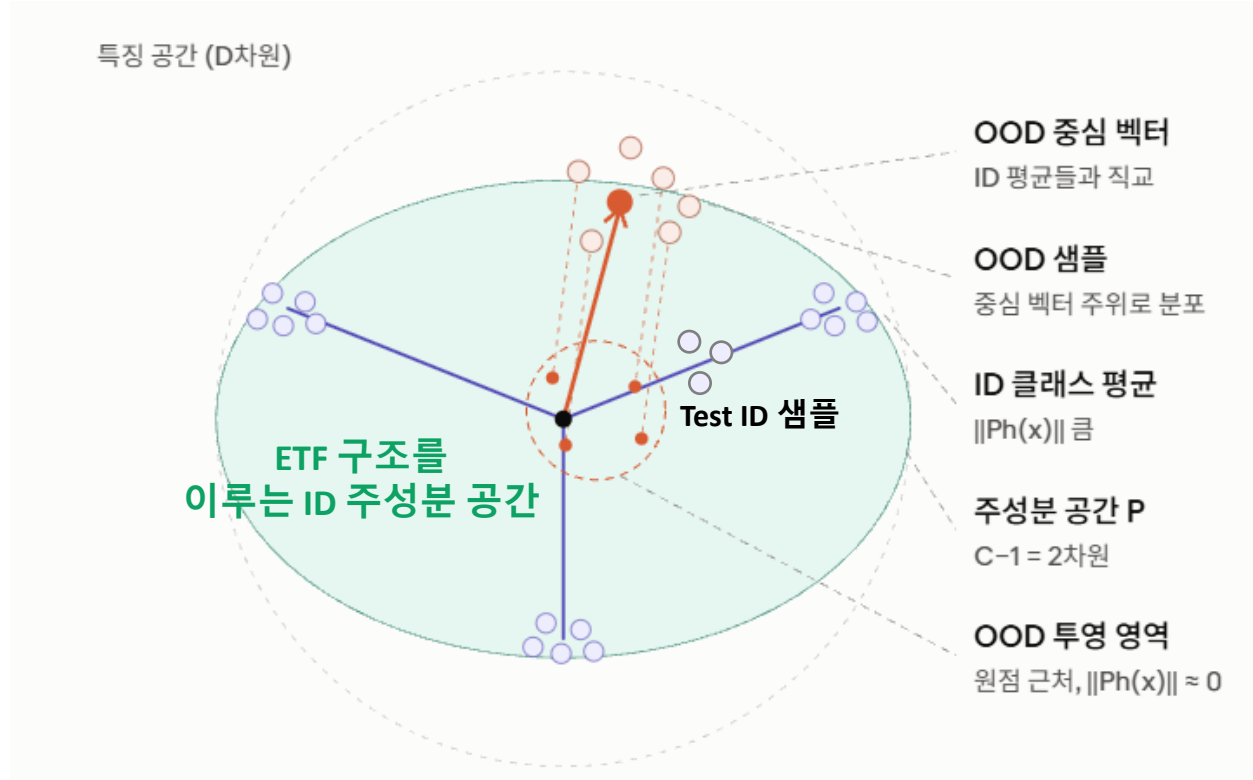


Figure 2: Feature projections on the first 2 principal components of a PCA fitted on CIFAR-10 (ID) using ViT penultimate layer representation. OOD data are ImageNet-O (left), Textures (middle), and SVHN (right). The Figure shows how NC1 (1) property is satisfied by ID data, and that OOD data lie around the origin.

NECO

Methodology

- ❖ 저자들은 이 직관을 기반으로 OOD Score를 정의
- ❖ ID 학습 데이터에 PCA를 적용하여, ID 데이터가 차지하는 공간을 대변하는 주성분 공간 P 를 계산
- ❖ 샘플의 특징 벡터 자체의 크기에 영향을 받지 않도록 원래 크기로 나누어 비율로 사용
- ❖ 클래스 정보를 주입하기 위해서 MaxLogit을 곱하여 최종적인 OOD Score 계산



주성분 공간 P , 특징 벡터 $h(x)$ 가 있을 때,

$$NECO(x) = \frac{\|Ph(x)\|}{\|h(x)\|}$$

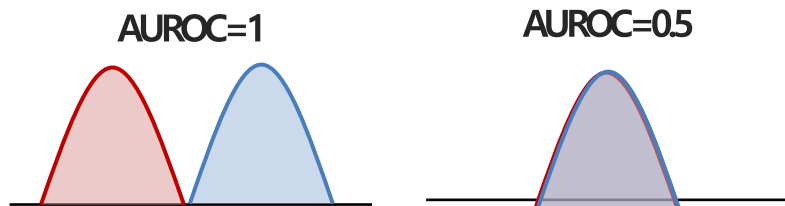
$$S(x) = NECO(x) \times MaxLogit$$

NECO

Experimental Results

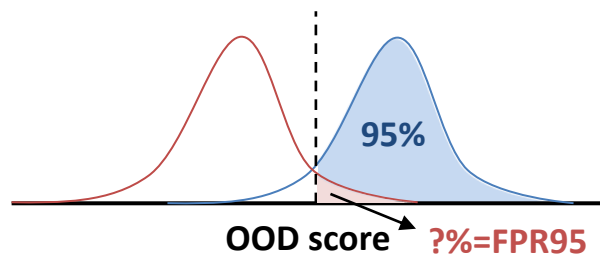
AUROC

- 전반적으로 ID와 OOD를 얼마나 잘 구분하는지 평가
- 1에 가까울수록 좋으며, 0.5는 랜덤하게 추측함을 의미



FPR95

- OOD탐지가 얼마나 실용적인지를 평가
- ID를 95% 통과시킬 수 있는 임계값을 기준으로 OOD를 ID로 잘못 예측한 비율
- 0에 가까울수록 좋은 지표



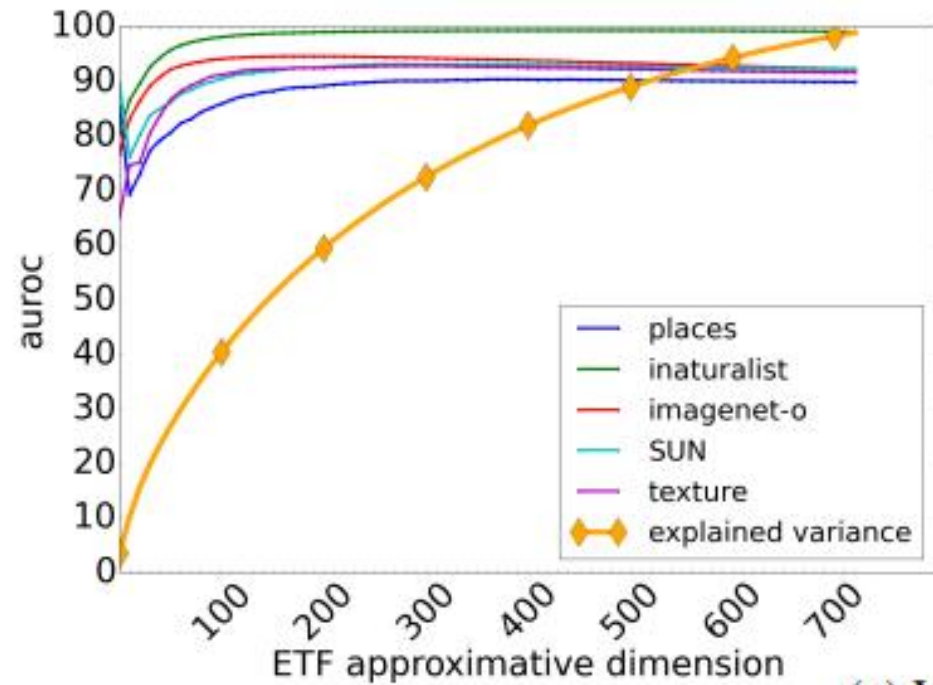
Model	Method	ImageNet-O		Textures		iNaturalist		SUN		Places365		Average	
		AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓
ViT-B/16	Softmax score	85.31	52.65	86.64	49.40	97.21	12.14	86.64	53.75	85.52	56.93	88.26	44.97
	MaxLogit	92.23	36.40	91.69	36.90	98.87	5.72	92.15	39.88	90.05	46.75	92.99	33.13
	Energy	93.03	31.65	92.13	34.15	99.04	4.76	92.65	36.90	90.38	43.72	93.45	30.24
	Energy+ReAct	93.08	31.25	92.08	34.50	99.02	4.91	92.56	36.94	90.19	44.25	93.39	30.37
	ViM	94.01	28.95	91.63	38.22	99.59	2.00	92.73	34.00	89.67	44.12	93.56	29.46
	Residual	92.10	38.35	87.71	52.34	99.36	2.79	89.46	40.91	85.60	53.93	90.85	37.66
	GradNorm	84.70	34.85	86.29	35.76	97.45	6.17	86.67	39.49	83.59	46.94	87.74	32.64
	Mahalanobis	94.00	30.90	91.69	37.93	99.67	1.55	91.38	38.48	88.66	46.09	93.08	30.99
	KL-Matching	82.90	53.40	84.61	52.38	95.07	15.31	83.25	62.10	82.00	65.99	85.57	49.84
	ASH-B	69.28	85.25	66.05	83.29	72.62	80.62	59.36	91.53	55.08	92.68	64.48	86.67
	ASH-P	93.31	29.00	91.55	37.58	98.75	6.25	92.96	33.50	90.07	41.81	93.33	29.63
	ASH-S	92.85	29.45	91.00	39.73	98.50	7.28	92.61	33.12	89.55	41.68	92.90	30.25
	NuSA	92.48	34.30	88.28	51.24	99.30	3.12	89.26	40.08	85.28	54.51	90.92	36.65
	NECO (ours)	94.53	25.20	92.86	32.44	99.34	3.26	93.15	33.98	90.38	42.66	94.05	27.51
SwinV2	Softmax score	61.09	89.60	81.72	60.91	88.59	47.66	81.24	66.85	81.09	68.27	78.75	66.66
	MaxLogit	61.34	87.95	80.36	59.55	86.47	50.51	78.17	68.50	77.43	69.41	76.75	67.18
	Energy	61.32	85.75	77.91	64.44	81.85	63.44	73.80	76.89	73.09	76.68	73.59	73.44
	Energy+ReAct	68.38	83.85	84.56	59.86	90.23	51.97	82.61	69.27	81.41	70.19	81.44	67.03
	ViM	69.06	83.45	81.50	61.18	87.54	54.09	75.24	73.92	73.12	76.46	77.29	69.82
	Residual	66.82	83.80	77.36	65.00	83.23	59.77	71.03	76.41	68.90	78.53	73.47	72.70
	GradNorm	37.39	95.95	33.84	93.31	31.82	95.01	31.97	96.29	33.06	95.73	33.62	95.26
	Mahalanobis	71.87	86.05	84.51	63.35	89.81	57.10	80.28	75.39	78.52	77.10	80.99	71.80
	KL-Matching	58.60	87.50	75.30	71.69	82.93	58.29	73.72	76.53	72.11	78.91	72.53	74.58
	ASH-B	47.07	96.35	38.59	97.50	48.62	97.55	52.11	95.64	52.93	96.18	47.86	96.64
	ASH-P	37.38	97.95	26.51	98.90	20.73	99.28	24.49	99.08	26.12	98.93	27.05	98.83
	ASH-S	40.36	95.10	36.08	94.63	16.15	99.50	22.21	98.53	23.72	98.05	27.70	97.16
	NuSA	56.50	91.95	62.72	83.14	64.01	83.58	55.97	91.28	54.44	92.71	58.73	88.53
	NECO (ours)	65.03	80.55	82.27	54.67	91.89	34.41	82.13	62.26	81.46	64.08	80.56	59.19

NECO

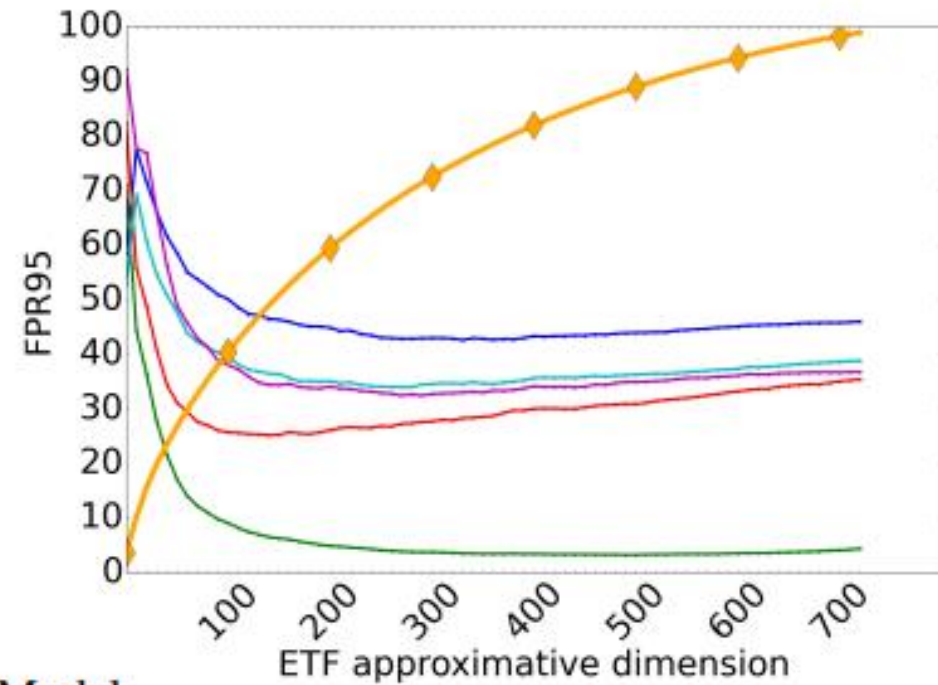
Experimental Results

Model	Method	ImageNet-O		Textures		iNaturalist		SUN		Places365		Average	
		AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓	AUROC↑	FPR↓
ViT	NECO w/o Maxlogit	93.27	30.05	89.02	46.96	99.48	1.47	89.90	38.24	85.89	52.15	91.51	33.77
	NECO	94.53	25.20	92.86	32.44	99.34	3.26	93.15	33.98	90.38	42.66	94.05	27.51
SwinV2	NECO w/o Maxlogit	63.17	87.95	75.52	69.63	90.72	36.37	81.59	57.19	80.52	60.77	78.31	62.38
	NECO	65.03	80.55	82.27	54.67	91.89	34.41	82.13	62.26	81.46	64.04	80.56	59.19
DeiT	NECO w/o Maxlogit	56.21	95.80	60.02	96.68	71.41	84.49	57.29	84.10	55.54	82.74	60.10	88.76
	NECO	62.72	84.10	80.82	59.47	89.54	42.26	77.64	65.55	76.48	68.13	77.44	63.90

클래스 정보인 MaxLogit 유무에 따른 실험 결과



(a) ViT Model



주성분 차원에 대한 민감도 분석

Breaking Semantic Hegemony: Decoupling Principal and Residual Subspaces for Generalized OOD Detection

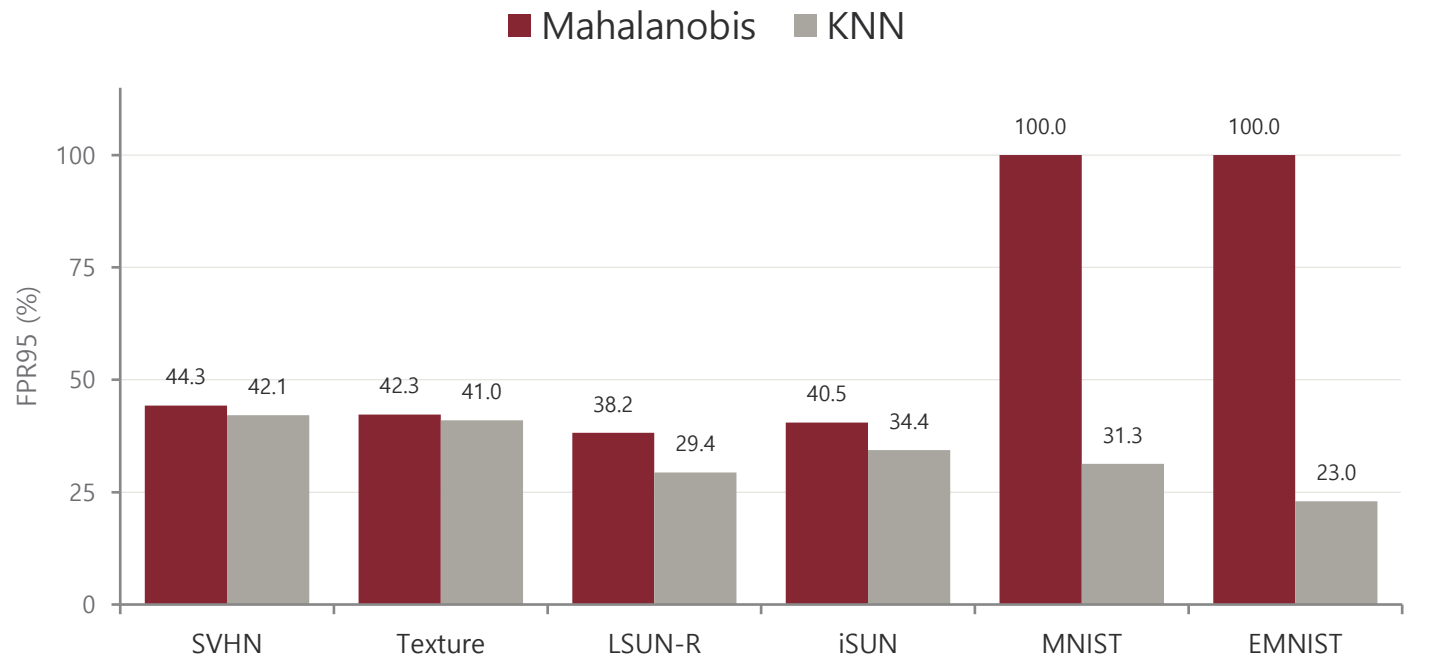
(2026, arXiv)

D-KNN

Motivation

- ❖ 저자들은 기존 Feature Space 거리 기반 OOD 탐지 방법론들이 의미적으로 단순한 OOD 샘플에서 오히려 실패함을 발견
- ❖ 이러한 반직관적 현상을 Simplicity Paradox라 명명하고, 그 원인을 특징 공간의 기하 구조에서 찾고자 함

CIFAR-100 기준 OOD 데이터셋별 FPR95 (낮을수록 좋음)



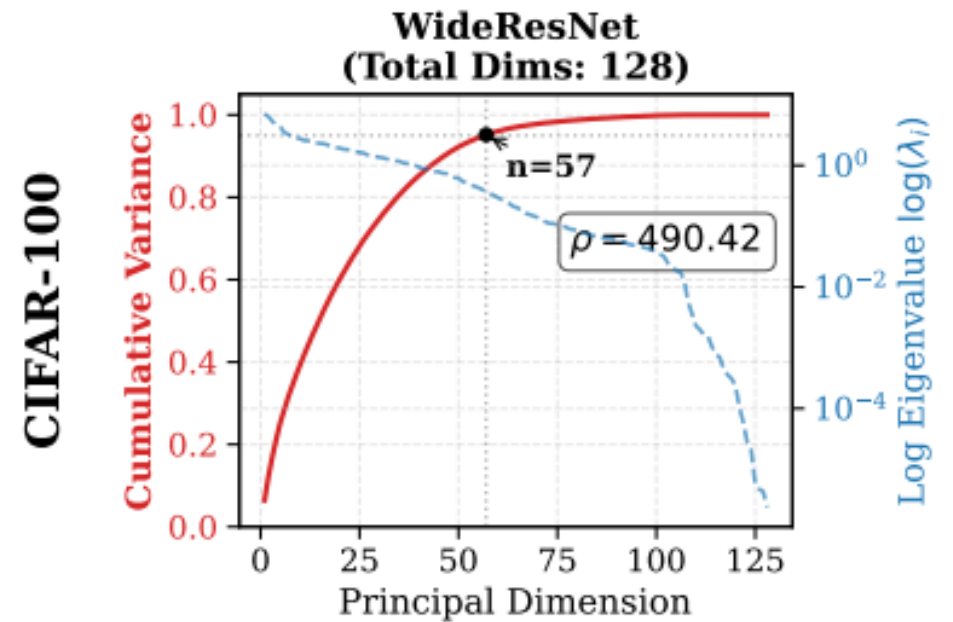
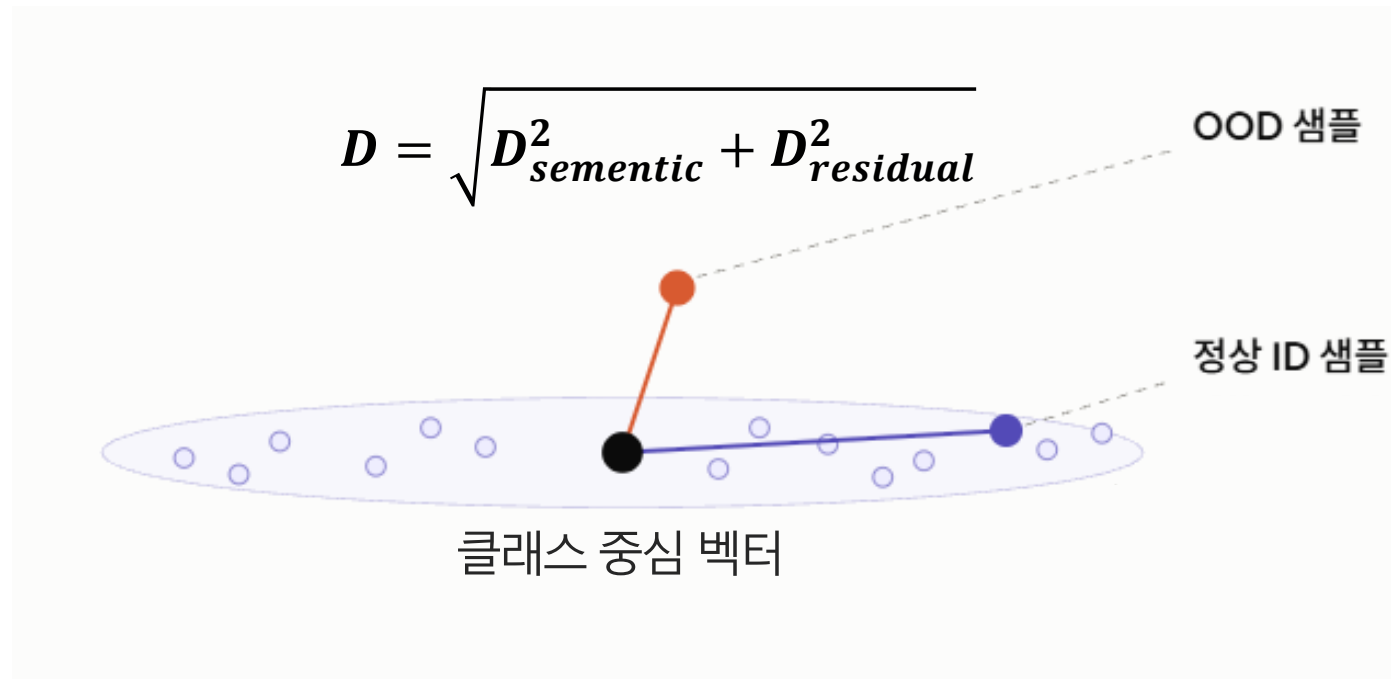
의미적으로 복잡한 OOD

의미적으로 단순한 OOD

D-KNN

Motivation

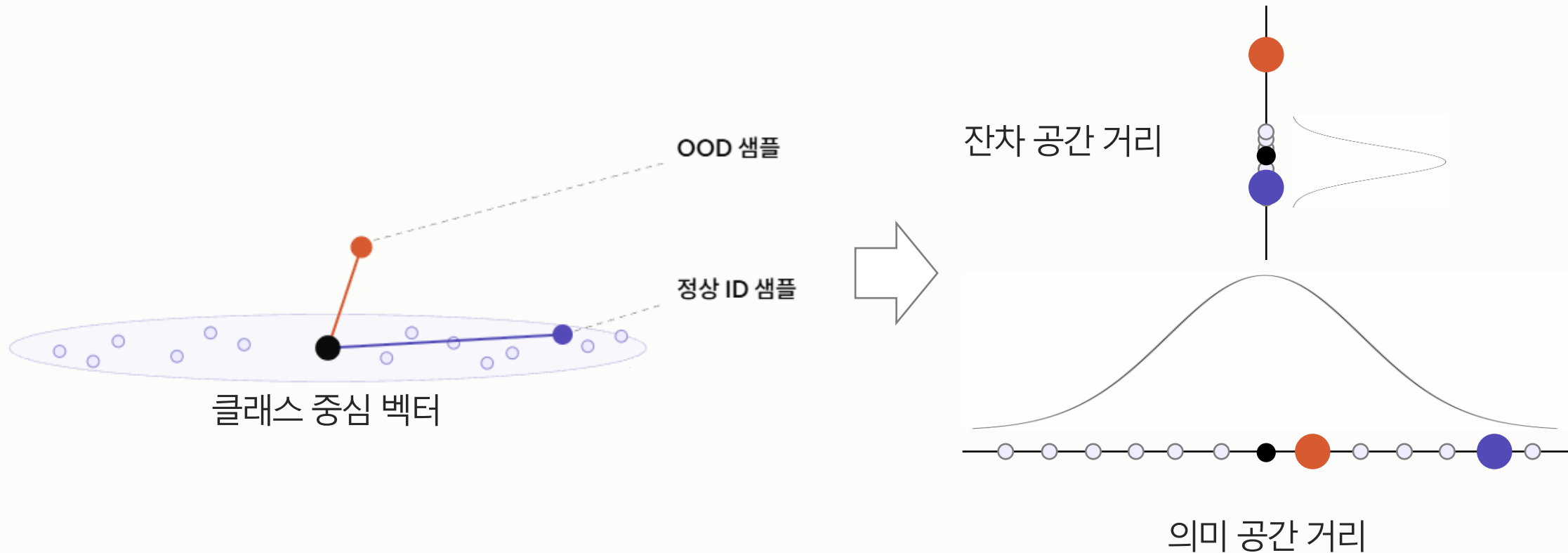
- ❖ 저자들은 NC로 인해 ID 특징의 분산이 특정 축으로 몰리고, 이로 인해 거리 지표가 왜곡된다고 가설을 세움
 - ❖ 실제로 측정한 결과 주성분 방향과 나머지 방향의 분산 비율이 최대 490배로, 가설이 성립함을 확인
- * Mahalanobis 거리는 축별 분산을 고려하지만, NC로 인해 대부분의 축 분산이 0에 가까워짐으로써 공분산 역행렬의 추정 오차가 증폭



D-KNN

Methodology

- ❖ 의미 성분의 지배로 인해 발생하는 편향을 제거하기 위해 의미 공간과 잔차 공간을 분리해서 거리를 계산
- ❖ 그 후, 표준화를 거쳐 동등한 스케일로 만들고 두 거리를 가중 평균하여 OOD Score로 활용



D-KNN

Methodology

- ❖ 의미 성분의 지배로 인해 발생하는 편향을 제거하기 위해 의미 공간과 잔차 공간을 분리해서 거리를 계산
 - ❖ 그 후, 표준화를 거쳐 동등한 스케일로 만들고 두 거리를 가중 평균하여 OOD Score로 활용
- * 실제로는 클래스 중심과의 거리가 아닌 KNN 거리를 활용

① 기하학적 공간 분리

Orthogonal Decomposition

PCA로 구한 주성분 공간(의미 공간)

$$P_{prin} = VV^T$$

주성분 공간과 직교하는 잔차 공간

$$P_{res} = I - P_{prin}$$

L2-정규화된 특징 벡터

$$z = \varphi(h(x)) = h(x) / \|h(x)\|$$

각 공간으로 투영된 특징 벡터

$$z_{prin} = \varphi((z - \mu)P_{prin} + \mu)$$

$$z_{res} = \varphi((z - \mu)P_{res} + \mu)$$

② 공간별 거리 계산

Dual-Space KNN

의미 공간에서 k이웃 거리

$$s_p(z) = \|z_{prin} - n_k^{(p)}\|$$

잔차 공간에서 k이웃 거리

$$s_r(z) = \|z_{res} - n_k^{(r)}\|$$

③ 표준화 및 결합

Dual-Space Calibration

표준화된 k이웃 거리

$$\hat{s}_p(z) = (s_p(z) - \mu_{s_p}) / \sigma_{s_p}$$

$$\hat{s}_r(z) = (s_r(z) - \mu_{s_r}) / \sigma_{s_r}$$

가중 평균한 최종 OOD Score

$$S(x) = \alpha \cdot \hat{s}_p + (1 - \alpha) \cdot \hat{s}_r$$

D-KNN

Experimental Results

ID	Method	OOD Datasets							Average
		EMNIST FPR↓ / AUC↑	MNIST FPR↓ / AUC↑	SVHN FPR↓ / AUC↑	Texture FPR↓ / AUC↑	LSUN-R FPR↓ / AUC↑	LSUN FPR↓ / AUC↑	iSUN FPR↓ / AUC↑	
CIFAR-10	MSP	17.2 / 97.7	10.8 / 98.5	55.1 / 90.0	63.7 / 87.2	48.3 / 92.7	22.9 / 97.0	54.0 / 91.0	38.9 / 93.4
	Energy	1.0 / 99.6	0.1 / 99.8	46.1 / 88.2	59.0 / 84.2	23.6 / 95.3	2.9 / 99.3	30.6 / 93.8	23.3 / 94.3
	MaxLogit	1.3 / 99.5	0.2 / 99.8	45.2 / 88.3	57.9 / 84.4	24.2 / 95.2	3.3 / 99.2	31.0 / 93.7	23.3 / 94.3
	Mahalanobis	42.2 / 94.7	26.2 / 96.1	16.8 / 96.9	20.6 / 96.0	59.0 / 89.9	55.8 / 90.8	58.2 / 89.4	39.8 / 93.4
	ViM	25.4 / 95.2	18.5 / 96.9	15.6 / 96.6	20.1 / 96.5	16.5 / 97.2	14.2 / 97.3	18.2 / 96.2	18.4 / 96.5
	ReAct	1.2 / 99.6	0.2 / 99.8	46.5 / 88.4	57.4 / 86.2	24.4 / 95.0	3.2 / 99.2	31.6 / 93.5	23.5 / 94.5
	DICE	0.02 / 99.9	0.01 / 99.9	47.3 / 86.0	60.1 / 81.1	22.8 / 95.1	0.5 / 99.7	29.3 / 93.7	22.9 / 93.6
	Line	4.7 / 98.6	1.2 / 99.7	73.5 / 72.4	75.4 / 69.8	61.2 / 81.3	11.0 / 96.9	69.3 / 78.1	42.3 / 85.5
	FDBD	6.6 / 99.0	2.4 / 99.5	38.4 / 92.5	48.5 / 89.0	32.8 / 93.4	13.0 / 97.7	38.9 / 91.4	25.8 / 94.6
	KNN	30.9 / 95.3	24.2 / 96.3	31.3 / 94.9	40.6 / 92.3	27.5 / 95.0	26.8 / 95.3	31.0 / 93.9	30.3 / 94.7
	D-KNN (Ours)	2.5 / 99.4	2.1 / 99.5	17.2 / 96.7	28.8 / 94.5	26.2 / 94.9	15.7 / 97.2	30.0 / 93.9	17.5 / 96.6
CIFAR-100	MSP	25.2 / 95.2	2.8 / 99.4	84.8 / 71.4	86.1 / 71.4	80.5 / 78.7	49.0 / 88.6	82.3 / 77.7	58.7 / 83.2
	Energy	1.6 / 99.4	0.01 / 99.8	87.2 / 71.4	85.3 / 72.8	79.9 / 81.6	16.7 / 96.8	83.8 / 80.1	50.6 / 86.0
	MaxLogit	3.4 / 99.2	0.01 / 99.8	86.2 / 71.7	84.9 / 73.1	79.3 / 81.9	21.4 / 96.2	83.1 / 80.4	51.6 / 86.1
	Mahalanobis	100.0 / 48.2	100.0 / 43.4	44.3 / 89.9	42.3 / 90.0	38.2 / 91.7	99.7 / 47.6	40.5 / 91.2	72.1 / 71.7
	ViM	77.3 / 80.5	97.0 / 66.8	69.8 / 75.1	79.1 / 67.9	94.5 / 59.0	91.4 / 65.4	94.0 / 58.8	86.2 / 67.6
	ReAct	1.5 / 99.4	0.01 / 99.8	87.3 / 75.0	84.3 / 76.2	79.8 / 81.3	17.1 / 96.8	83.5 / 79.9	50.5 / 86.9
	DICE	0.02 / 99.7	0.01 / 99.8	90.4 / 68.6	83.0 / 72.4	81.6 / 78.8	4.2 / 98.7	83.6 / 78.2	49.0 / 85.2
	Line	2.6 / 99.2	0.01 / 99.8	96.8 / 62.5	81.3 / 69.0	98.2 / 53.3	38.1 / 92.4	98.9 / 56.3	73.6 / 76.1
	FDBD	12.9 / 98.0	0.4 / 99.8	72.0 / 83.6	72.0 / 83.3	58.0 / 89.8	45.8 / 91.2	63.2 / 88.2	46.3 / 90.6
	KNN	23.0 / 96.3	31.3 / 95.4	42.1 / 90.3	41.0 / 89.7	29.4 / 94.3	56.2 / 82.0	34.4 / 92.3	36.8 / 91.5
	D-KNN (Ours)	1.9 / 99.5	2.3 / 99.3	32.0 / 93.4	35.6 / 91.7	29.9 / 94.3	51.3 / 85.4	33.8 / 92.5	26.7 / 93.7

Table 1: Detailed OOD Detection Results on CIFAR-10 and CIFAR-100 (WRN-28-10). Metric: FPR95 (↓) / AUC (↑). All results are percentages (%).

D-KNN

Experimental Results

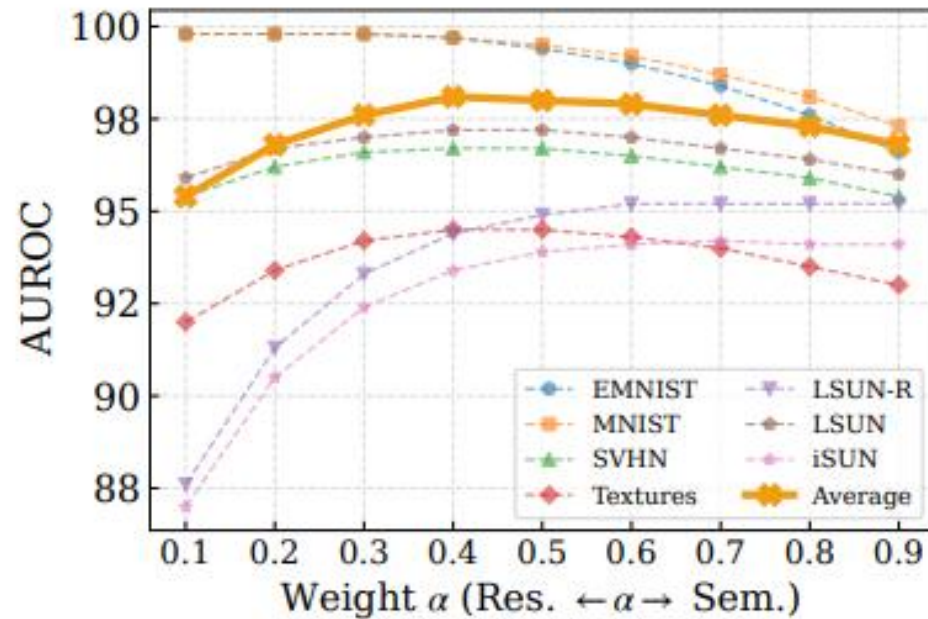


Method	Corruptions Datasets							Average
	Contrast FPR↓ / AUC↑	Frost FPR↓ / AUC↑	Gaussian FPR↓ / AUC↑	Glass Blur FPR↓ / AUC↑	Shot FPR↓ / AUC↑	Speckle FPR↓ / AUC↑	Zoom Blur FPR↓ / AUC↑	
MSP	87.3 / 70.7	86.1 / 69.5	81.8 / 79.7	85.0 / 72.6	82.7 / 78.6	81.7 / 78.3	85.7 / 68.7	84.3 / 74.0
Energy	91.4 / 70.8	85.1 / 72.2	81.9 / 77.9	86.2 / 71.7	79.9 / 78.5	78.6 / 79.2	85.8 / 71.0	84.1 / 74.5
MaxLogit	90.1 / 71.2	84.9 / 72.3	81.7 / 78.6	85.5 / 72.1	80.4 / 78.9	79.0 / 79.6	85.5 / 71.1	83.9 / 74.8
Mahalanobis	59.6 / 84.1	70.4 / 80.5	43.9 / 91.1	66.9 / 82.6	52.5 / 87.7	52.4 / 86.9	75.2 / 76.1	60.1 / 84.1
ViM	74.6 / 69.2	87.4 / 63.5	53.8 / 84.2	89.4 / 60.7	53.9 / 83.2	56.8 / 81.8	92.0 / 59.2	72.6 / 71.7
ReAct	91.4 / 71.9	84.9 / 73.9	82.4 / 78.9	86.3 / 72.1	79.9 / 79.4	78.7 / 80.1	85.8 / 71.9	84.2 / 75.5
Neco	90.0 / 71.5	84.3 / 72.7	79.5 / 79.3	85.0 / 72.6	78.5 / 79.7	76.7 / 80.5	85.3 / 71.5	82.7 / 75.4
KNN	47.0 / 84.9	58.2 / 82.6	31.5 / 93.4	60.0 / 83.5	36.4 / 91.8	32.6 / 92.2	68.5 / 76.6	47.7 / 86.4
D-KNN (Ours)	43.0 / 86.7	58.1 / 83.5	24.0 / 94.9	58.0 / 84.9	30.0 / 93.3	29.4 / 93.2	68.6 / 77.6	44.5 / 87.7

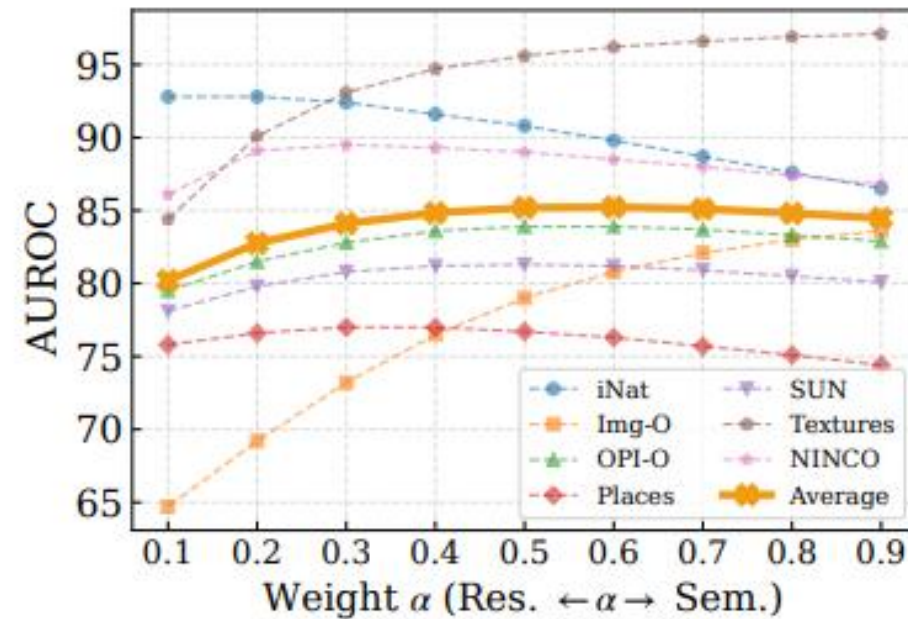
Table 5: **OOD Detection Results on CIFAR-100-C. Metric:** FPR95 (↓) / AUC (↑). All results are percentages (%) and rounded to one decimal place.

D-KNN

Experimental Results



(a) CIFAR-10



(b) ImageNet-1K

α 대한 민감도 분석

Conclusion

Conclusion

- ❖ Neural Collapse는 특징 공간의 기하 구조를 가정하는 대신 설명할 수 있는 이론적 배경을 OOD 탐지에 제공
- ❖ 두 논문은 동일한 NC 현상에서 출발했으나, 이를 활용할 자원으로 볼 것인지 교정할 편향으로 볼 것인지에서 갈라짐

NC를 자원으로 NECO (2024, ICLR)

- ID가 차지한 저차원 공간에 테스트 샘플이 얼마나 정렬되어 있는지를 측정
- OOD가 ID의 ETF 구조에 직교해 간다는 관찰에 기반하여, 주성분 공간 투영
- 원본 벡터와 투영된 벡터의 크기 비율을 스코어로 사용

NC를 편향으로 D-KNN (2026, arXiv)

- 분산 쏠림이 거리 지표를 왜곡한다고 보고 의미 공간과 잔차 공간을 분리
- 각 공간에서 독립적으로 거리를 잰 뒤 표준화하여 결합함으로써, 묻혀 있던 잔차 신호를 되살림

감사합니다