ICML
International Conference
On Machine Learning

# A Geometry-Based View of Mahalanobis OOD Detection

Denis Janiak, Jakub Binkowski, Tomasz Kajdanowicz

Wroclaw University of Science and Technology

**TL;DR:** the *same* Mahalanobis detector swings wildly across backbones — we show *representation geometry* explains it, and tune it with one ID-only knob.

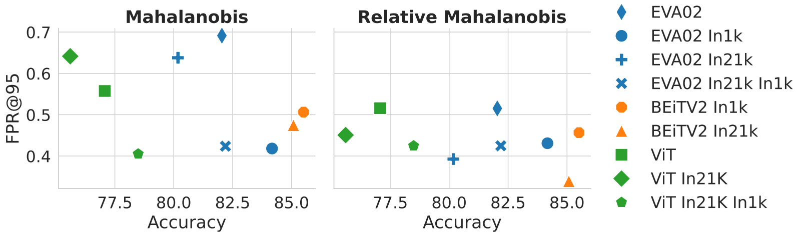
Wrocław Tech

Department of Artificial Intelligence

- Mahalanobis = *simplest* post-hoc OOD detector:

$$S(z) = (z - \mu)^\top \Sigma^{-1} (z - \mu).$$

- Still competitive on modern vision backbones.
- But: same detector, very different results across pretraining / fine-tuning.

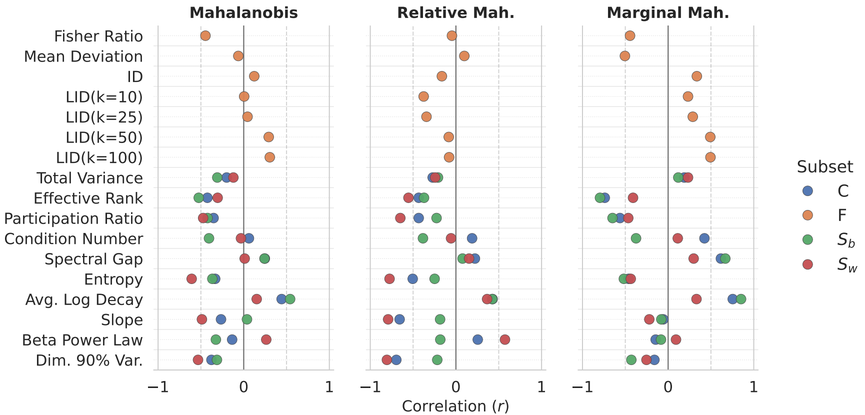



FPR@95 vs. Accuracy on NINCO.

**Research Question**

*Which properties of an ID feature space decide whether Mahalanobis succeeds or fails?*

# Different detectors read different geometry cues




**F**: feature metrics; **C**, $S_b$, $S_w$: covariance / scatter spectra.

**RMD** → within-class compactness ($S_w$)

**MMD** → global manifold ($C$, $S_b$)

**MD** → in between (cluster & global)

# ID geometry predicts when Mahalanobis will fail

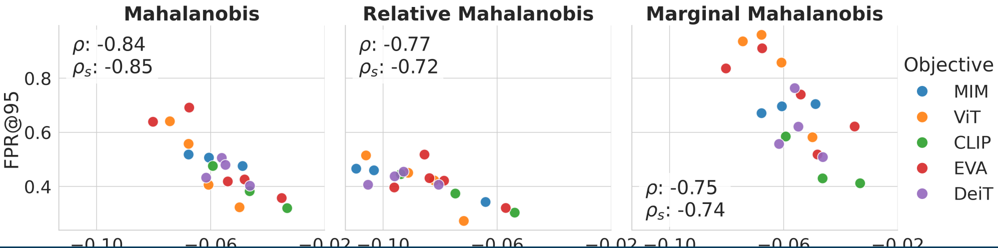
Two ID-only scalars:

- $m_k$ — local intrinsic dimensionality (manifold richness, $k$NN-MLE).
- $|s|$ — within-class spectral slope (class compactness, decay of $S_w$).

## Compensatory trade-off

$$P(\beta) = m_k(\beta)\,|s(\beta)|$$

Predicts Mahalanobis OOD across models & variants — $\rho \approx -0.85$.

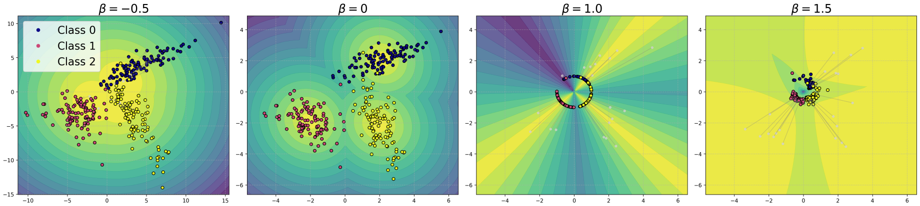


# We tune geometry without changing feature directions

**Direction-preserving transformation (radial scaling)**

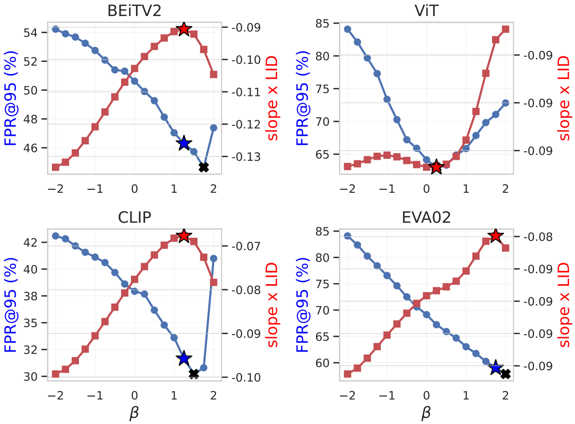
$$\phi_\beta(z) = \frac{z}{\|z\|^\beta}$$

- Detector form *unchanged* (still quadratic).
- $\beta = 0$: original $\beta = 1$: unit sphere (MD++).
- $\beta$ moves the rep along the $m_k \times |s|$ curve — **tunes geometry, not the detector**.




Larger $\beta$ contracts norms $\rightarrow$ tighter clusters & localized decision regions; smaller $\beta$ broadens them.

- **ID-only rule:** sweep $\beta$, evaluate $P(\beta) = m_k(\beta)\,|s(\beta)|$, take its interior turning point $\hat{\beta}$. No OOD data.
- Beats fixed $\beta = 0$ & $\beta = 1$; one $\hat{\beta}$ per model; approaches *oracle*-$\beta$ FPR.




Larger $\beta$ contracts norms $\rightarrow$ tighter clusters & localized decision regions; smaller $\beta$ broadens them.

# ID-only $\beta$ selection beats fixed normalization

- **ID-only rule:** sweep $\beta$, evaluate $P(\beta) = m_k(\beta)\,|s(\beta)|$, take its interior turning point $\hat{\beta}$. No OOD data.
- Beats fixed $\beta=0$ & $\beta=1$; one $\hat{\beta}$ per model; approaches *oracle*-$\beta$ FPR.
- *Why:* score = *size* (how far the residual reaches, $\sim m_k$) $\times$ *stretch* (how whitening reweights it, $\sim |s|$).

Avg. FPR@95 over 12 backbones $\times$ 5 OOD sets

| Detector | Avg. FPR@95 |
|---|---|
| KNN | 40.1 |
| ViM | 38.2 |
| MD ($\beta=0$) | 37.2 |
| MD++ ($\beta=1$) | 36.0 |
| **RS-MD** (proxy $\hat{\beta}$, ours) | **35.3** |
| RMD ($\beta=0$) | 37.2 |
| RMD++ ($\beta=1$) | 36.1 |
| **RS-RMD** (proxy $\hat{\beta}$, ours) | **35.8** |

# ID-only $\beta$ selection beats fixed normalization

- **ID-only rule:** sweep $\beta$, evaluate $P(\beta) = m_k(\beta)\,|s(\beta)|$, take its interior turning point $\hat{\beta}$. No OOD data.
- Beats fixed $\beta = 0$ & $\beta = 1$; one $\hat{\beta}$ per model; approaches *oracle*-$\beta$ FPR.
- *Why:* score = *size* (how far the residual reaches, $\sim m_k$) $\times$ *stretch* (how whitening reweights it, $\sim |s|$).
- Largest gains on **near-OOD** (the practically hard regime).

Near vs. far split: proxy $\hat{\beta}$ vs. fixed $\beta = 1$

| Det. | Selector | Far | Near | All |
|---|---|---|---|---|
| MD | $\beta = 1$ | 20.6 | 59.0 | 36.0 |
| MD | **Proxy** | **20.2** | **57.8** | **35.3** |
| RMD | $\beta = 1$ | 21.7 | 57.8 | 36.1 |
| RMD | **Proxy** | **21.4** | **57.5** | **35.8** |

Near = NINCO, SSB-Hard;

Far = iNaturalist, OpenImages-O, Textures.

# Normalization is geometry tuning, not preprocessing

## Takeaways

- Mahalanobis OOD is **representation-geometry** dependent.
- $m_k|s|$ is a simple **ID-only diagnostic**.
- $\beta$-scaling tunes geometry and **beats fixed normalization**.

## Limitations

- ImageNet ID + vision backbones.
- Proxy is not the oracle on every dataset.
- One global $\beta$ per model.

---

**Thank you for watching!**

`denis.janiak@pwr.edu.pl` · Paper →

# Selected references I


- Lee et al. *A simple unified framework for detecting OOD samples*. NeurIPS 2018.
- Ren et al. *A simple fix to Mahalanobis distance for OOD detection* (RMD). 2021.
- Müller et al. *Mahalanobis++: L2-normalized features for OOD*. 2025.
- Ma et al. *Characterizing adversarial subspaces using LID*. ICLR 2018.
- Bitterwolf et al. *In or out? (NINCO)*. ICML 2023.
- Zhang et al. *OpenOOD*. NeurIPS 2022.