UNIVERSITY OF CALABRIA
DEPARTMENT OF COMPUTER ENGINEERING, MODELING, ELECTRONICS AND SYSTEMS

CF-OOD: Concentration-Free density estimation for Reliable Out-of-Distribution Detection

Fabrizio Angiulli, Fabio Fassetti, Maria Pia Zupi✉

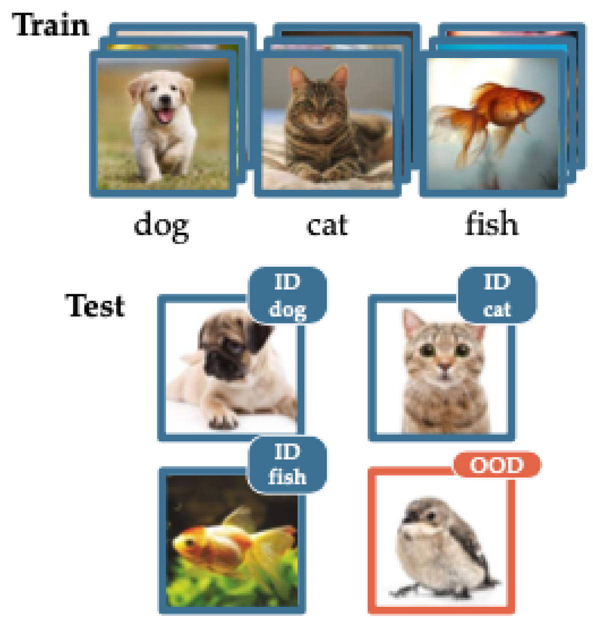
fabrizio.angiulli@unical.it, fabio.fassetti@unical.it, mariapia.zupi@dimes.unical.it

sisap
SISAP 2025
October 1st-3rd
Reykjavik, Iceland

1. Problem: OOD Detection – Seeing the Unseen

What is OOD Detection?

Most deep learning models assume that test data follow the same distribution as training data (in-distribution, ID). In practice, this assumption often fails: unseen or unexpected inputs called *out-of-distribution* (OOD) may lead to incorrect and overconfident predictions.



Why It Matters?

In high-risk applications, misclassifying OOD inputs can have severe consequences:

- **Medical diagnostics**: the model may give overconfident predictions on unseen or unknown conditions, leading to misdiagnosis.;
- **Autonomous driving**: novel road scenarios may cause unsafe actions.

Recognizing OOD inputs enables better safety and robustness.

2. Solution: CFOF-score

Let $\mathcal{D}$ be a reference set of n points in a space $\mathbb{U}$ with a distance function *dist*. Given an instance $x \in \mathbb{U}$, the k-th nearest neighbor of x in $\mathcal{D}$, denoted $nn_k^{\mathcal{D}}(x)$, is the point such that exactly $k-1$ points in $\mathcal{D}$ are closer to x. The set of k nearest neighbors of x is:

$$\mathrm{NN}_k^{\mathcal{D}}(x) = \{nn_i^{\mathcal{D}}(x) : 1 \le i \le k\}.$$

Reverse neighborhood size: how many points have x as one of their k nearest neighbors:

$$\mathrm{N}_k^{\mathcal{D}}(x) = \left|\{y \in \mathcal{D} : x \in \mathrm{NN}_k^{\mathcal{D}}(y)\}\right|.$$

CFOF anomaly score: given $\varrho \in (0,1)$, the CFOF score of x is:

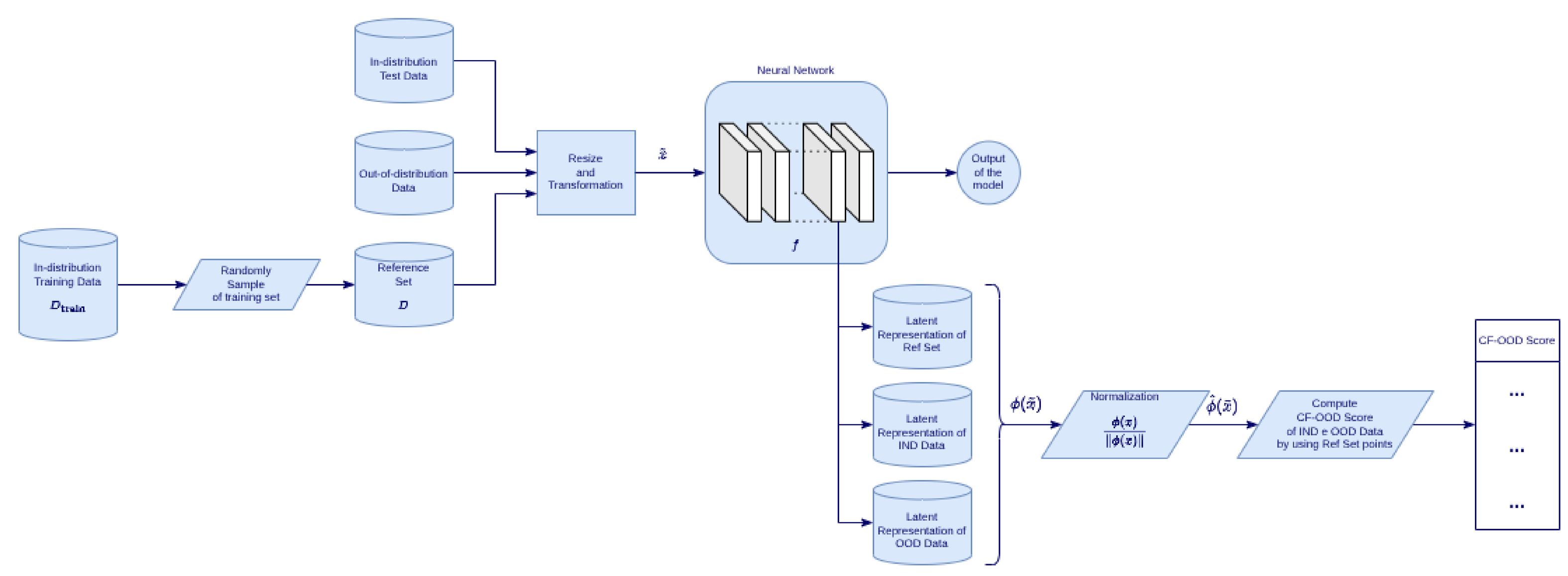
$$\mathrm{CFOF}_{\mathcal{D}}(x) = \min_{1 \le k \le n} \left\{ \frac{k}{n} : \mathrm{N}_k^{\mathcal{D}}(x) \ge n\varrho \right\}.$$

This score measures how many neighbors are needed for x to be "close" to at least a fraction ϱ of the dataset — the smaller the score, the more central x is.

CF-OOD score: applied in latent space via a representation function ϕ:

$$\mathrm{CF\text{-}OOD}_{\mathcal{D}}(x) = \mathrm{CFOF}_{\phi(\mathcal{D})}(\phi(x)).$$

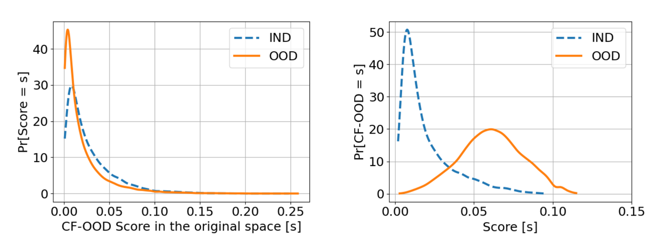
3. Application: CF-OOD Approach



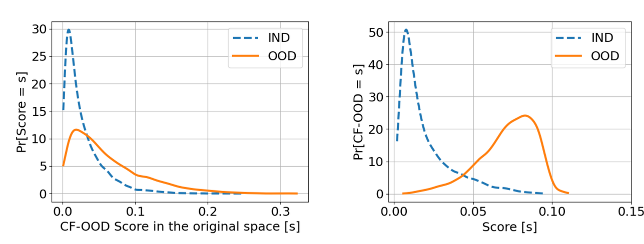
We apply the CF-OOD score to detect OOD samples using latent representations from a pre-trained network. The procedure is:

1. **Model setup:** pre-trained f on $\mathcal{D}_{\text{train}}$.
2. **Reference sampling:** select subset $\mathcal{D} \subset \mathcal{D}_{\text{train}}$ (size α).
3. **Feature extraction:** extract $\phi(\tilde{x})$ from penultimate layer.
4. **Normalization:** compute $\hat{\phi}(\tilde{x}) = \phi(\tilde{x})/\|\phi(\tilde{x})\|$.
5. **Distance computation:** compute Euclidean distances to reference set.
6. **Scoring:** apply CF-OOD via precomputed distance histograms.

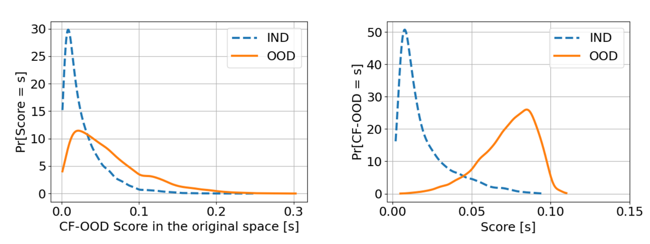
4. Results



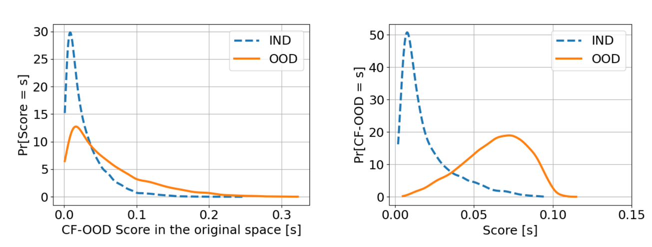
SVHN Dataset



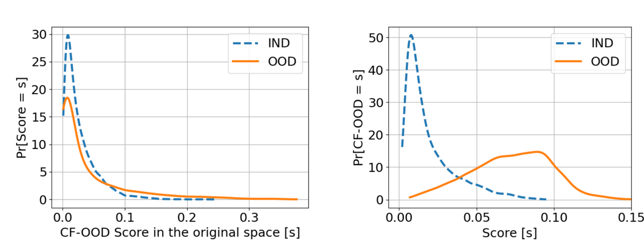
ISUN Dataset



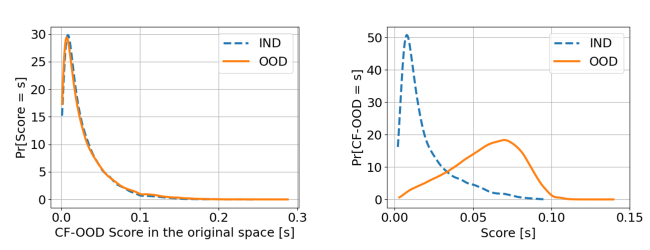
LSUN Dataset



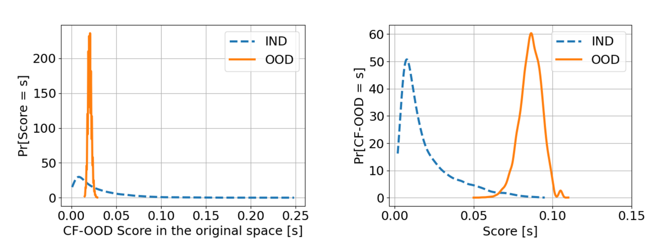
TINY_IMAGENET Dataset



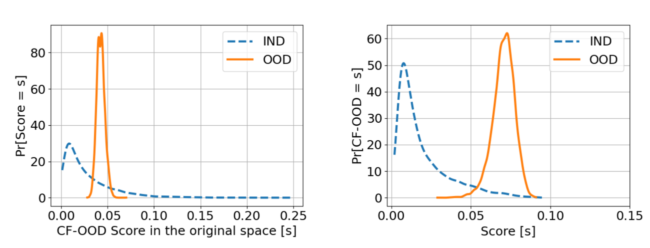
TEXTURES Dataset



PLACES 365 Dataset



UNIFORM Dataset



GAUSSIAN Dataset

- The CFOF score allows effective discrimination in the **latent space**, the distributions become clearly distinguishable. In the **original space**, score distributions largely overlap.
- **Table 1** shows AUROC-based pairwise comparisons. Each cell reports how often the row method outperforms the column. CF-OOD stands out: highest number of wins and rarely underperforms.

	MSP	ODIN	ENERGY	MAHAL	REACT	KNN	CF-OOD
MSP	–	37.5	37.5	57.5	57.5	10.0	5.0
ODIN	62.5	–	50.0	57.5	57.5	20.0	7.5
ENERGY	62.5	47.5	–	60.0	55.0	22.5	12.5
MAHAL	42.5	42.5	40.0	–	57.5	5.0	2.5
REACT	42.5	42.5	42.5	42.5	–	20.0	2.5
KNN+	90.0	80.0	77.5	95.0	80.0	–	15.0
CF-OOD	95.0	92.5	87.5	97.5	97.5	80.0	–

Table 1: Percentage of wins (AUROC)

5. Conclusion

- CF-OOD works across **multiple deep learning models** without retraining or changes to training, making it easy to integrate into existing pipelines.
- Extensive experiments show that **CF-OOD outperforms** state-of-the-art density-based methods (AUROC-based).
- In the latent space, OOD and ID samples become **clearly separable**, enabling robust detection.
- Future work includes broader in-distribution datasets and exploring alternative hidden representations and normalizations.