

Supplementary material for "Outlier detection by ensembling uncertainty with negative objectness"

Anja Delić
 anja.delic@fer.hr
 Matej Grcić
 matej.grcic@fer.hr
 Siniša Šegvić
 sinisa.segvic@fer.hr

University of Zagreb,
 Faculty of Electrical Engineering and
 Computing,
 Zagreb, Croatia

A Correlation analysis of UNO components

Table 1 reports the Pearson correlation coefficient of per-pixel outlier scores s_{Unc} and s_{NO} on Fishyscapes [4] val and RoadAnomaly [34]. We observe that the two UNO components are either mildly correlated or completely uncorrelated. These findings indicate that the performance gains of UNO can be explained by the ensemble learning [30].

Data	FS L&F	FS Static	RoadAnomaly
Outliers	0.27	0.01	0.13
Inliers	0.01	0.15	0.01
All	0.16	0.56	0.41

Table 1: Pearson correlation coefficient of per-pixel scores s_{Unc} and s_{NO} on the three outlier segmentation validation datasets (FS L&F, FS Static and RoadAnomaly).

The benefits of ensembling can also be observed in the feature space. Figure 1 indicates that features from different outlier datasets have different L_2 norms. Outliers that resemble the training negatives typically have a higher norm and small angle to \mathbf{w}_{K+1} , while outliers that are more similar to inliers have lower norm. In the former case, outliers are detected with s_{NO} and with s_{Unc} in the latter case.

B On orthogonality of class vectors

We empirically observe that all class vectors \mathbf{w}_i are mutually orthogonal to each other, $\mathbf{w}_i^T \mathbf{w}_j = 0, \forall i, j \in \{1, 2, \dots, K+1\}$, as shown in Figure 2. The cosine of the angle between any two different class vectors is approximately zero. Such behaviour allows the geometrical interpretation of UNO, as shown in Figure 2 of the main manuscript.

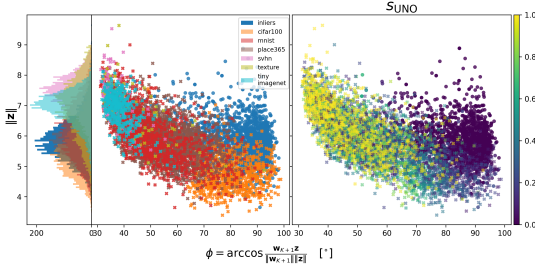


Figure 1: Visualization of the pre-logit space for OpenOOD CIFAR-10. Outlier feature representations either yield an above-average norm with a small angle to \mathbf{w}_{K+1} (e.g. SVHN and Places365) or yield low norm representations (e.g. CIFAR-100 that is similar to inliers).

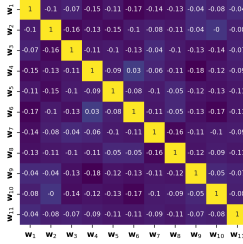


Figure 2: We show that all weight vectors are mutually orthogonal in the form of a heatmap where each element shows the cosine of the angle between two corresponding weight vectors. We use the K+1-way image-wide classifier with the ResNet-18 backbone trained on CIFAR-10 as the example.

C Scene parsing with mask-wide recognition

We extend the Mask2Former [8] architecture with our UNO outlier detector to solve the task of anomaly segmentation. Thus, we describe Mask2Former architecture in detail to make the manuscript self-contained. The Mask2Former architecture consists of three main parts: backbone, pixel decoder, and mask decoder. The backbone extracts features at multiple scales from a given image $\mathbf{x} \in \mathbb{R}^{3 \times H \times W}$. The pixel decoder produces high-resolution per-pixel features $\mathbf{E} \in \mathbb{R}^{E \times H \times W}$ that are fed into the mask decoder. The mask decoder formulates semantic segmentation as a direct set prediction problem by providing two outputs: N mask embeddings \mathbf{q} , and N mask-wide categorical distributions over K+1 classes $P(Y = k | \mathbf{z}_i) = \text{softmax}(\mathbf{W} \cdot \mathbf{z}_i + \mathbf{b})$. The K+1 classes include K inlier classes and one no-object class. Note that \mathbf{z}_i denotes the vector of mask-wide pre-logit activations of the i-th mask. The mask decoder projects pre-logits \mathbf{z}_i into logits by applying the learned matrix \mathbf{W} . The binary masks $\mathbf{m} = \sigma(\text{conv}_{1 \times 1}(\mathbf{E}, \mathbf{q}))$ are obtained by scoring per-pixel features \mathbf{E} with the mask embeddings \mathbf{q} . The sigmoid activation interprets each element of the obtained tensor as a probabilistic assignment of the particular pixel into the corresponding mask. Semantic segmentation can be carried out by classifying each pixel according to a weighted ensemble

of per-mask classifiers $P(y|\mathbf{z})$, where the weights correspond to dense mask assignments \mathbf{m} :

$$\hat{y}[r, c] = \operatorname{argmax}_{k=1, \dots, K} \sum_i^N \mathbf{m}_i[r, c] P(Y = k | \mathbf{z}_i). \quad (1)$$

Default mask-level posterior already includes $K+1$ classes that correspond to K inlier classes and one no-object class. We implement our method by introducing an additional class to learn the negative objectness, which brings us to $K+2$ classes in total. To be consistent with the image-wide setup, in addition to the K inlier classes, we place the outlier class at index $K+1$ and the no-object class at index $K+2$. We expose the segmentation model to negative data by training on mixed-content images [9]. Closed-set recognition can still be carried out by considering only the K inlier logits (1).

We detect anomalies on the mask-level by applying UNO to the mask-wide pre-logits \mathbf{z}_i of the i -th mask. We define the per-pixel outlier score at spatial positions r and c as a sum of mask-level outlier scores weighted with dense probabilistic mask assignments [18]:

$$\mathbf{s}_{\text{UNO}}^{\text{M2F}}[r, c] = \sum_{i=1}^N \mathbf{m}_i[r, c] \cdot \mathbf{s}_{\text{UNO}}(\mathbf{z}_i). \quad (2)$$

D Experimental setup

D.1 Benchmarks and datasets

We evaluate UNO on pixel-level outlier detection benchmarks Fishyscapes [9] and SMIYC [5] and the image-wide OpenOOD [55] benchmark.

Pixel-level benchmarks. Fishyscapes [9] contains datasets with real (FS Lost&Found) and synthetic (FS Static) outliers. SMIYC [5] has two dominant tracks which group anomalies according to size. AnomalyTrack focuses on the detection of large anomalies on the traffic scenes while ObstacleTrack focuses on the detection of small obstacles on the road. Additionally, we validate on the RoadAnomaly [54] dataset which is an early version of the AnomalyTrack dataset.

Image-level benchmarks. OpenOOD [55] proposed a unified benchmark for image-wide OOD detection with large-scale datasets. The test outliers are divided into two groups (Near-OOD and Far-OOD) based on semantical similarity to the inlier classes or observed empirical difficulty. The far-OOD group consists of outliers that are semantically far from the inliers (numerical digits, textural patterns, or scene imagery). The near-OOD group consists of outliers that are semantically similar to the inliers as they all include specific objects. The OpenOOD-CIFAR-10 setup uses the official CIFAR-10 [28] splits as ID train, val and test subsets. The negative dataset corresponds to a subset of Tiny ImageNet (TIN) [80]. The near-OOD datasets include CIFAR-100 [49] and a subset of Tiny ImageNet (TIN) [80] that does not overlap with CIFAR-10 and the negative dataset. The far-OOD group consists of MNIST [12], SVHN [53], Textures [9], and Places365 [52] without images that are related with any of the ID classes. The large-scale OpenOOD ImageNet-200 benchmark considers a subset of 200 classes from ImageNet-1K [11] as the inlier training dataset. The remaining 800 classes are used as the negative dataset. The near-OOD group consists of SSB-hard [43] and NINCO [8] while the far-OOD group includes iNaturalist [47], Textures [9], and OpenImage-O [49].

D.2 Evaluation metrics

We use standard evaluation metrics: area under the precision-recall curve (AP), area under the receiver operating curve (AUROC or AUC), and false positive rate at 95% true positive rate (FPR₉₅). We validate in-distribution performance with accuracy and mIoU. Note that we omit the AUROC metric in the pixel-level experiments since all methods achieve high AUROC within variance.

D.3 Implementation details

Segmentation of road scenes. Our pixel-level experiments build upon a Mask2Former model [8] with an ImageNet-initialized SWIN-L [66] backbone. We pre-train the Mask2Former in closed-set setup for 115K iterations on Cityscapes [10] and Mapillary Vistas [58] with the Cityscapes taxonomy. We use the default hyperparameters [8] and set the batch size to 18. We extend the mask-wide classifier to K+2 classes and fine-tune the model on mixed-content scenes with either real or synthetic negatives for 2K iterations. We use the zero-initialization for the added weights. When training with real negative data, we assemble mixed-content images by pasting three semantically different instances sampled from ADE20K [58] after resizing to the range of [96, 512]. In the case of training with synthetic negatives, we jointly fine tune the K+2-way segmentation model and a flow by adding the loss defined in Equation 6 in the main manuscript to the standard optimization process [8]. We use the DenseFlow-25-6 [14] that we pretrain on Mapillary Vistas for 300 epochs. We generate rectangular patches with spatial dimensions in the range of [48, 512] by sampling the jointly trained flow. We set the loss modulation parameter from Equation 6 to 0.03. We find that training our segmentation model with negatives from scratch, besides from being computationally expensive, leads to overfitting to negatives. Thus, we only finetune the model trained in closed-set manner with both real and synthetic negatives. Additionally, in this case we do not observe feature collapse when utilizing the joint loss (Equation 6 in the main manuscript) since we only jointly train for a small number of iterations. The closed-set training of Mask2Former lasts 48 hours, while the fine-tuning stage takes only 30 minutes on three A6000 GPUs.

Image classification Our image-wide experiments follow the official training setup [58]. When training with real negatives, we train the ResNet-18 [10] backbone with a K+1-way classification layer for 100 epochs from random initialization. We use the SGD optimizer with a momentum of 0.9 and a learning rate of 0.1 with cosine annealing decay schedule. We apply the weight decay of 0.0005. We set the batch size to 128 for CIFAR-10 and 256 for ImageNet-200. Each minibatch contains the equal ratio of all K+1 classes. Specifically, for CIFAR-10 the minibatch contains 117 inlier images and 11 negative images, and 255 inliers and only one negative sample for the ImageNet-200. When training with synthetic negatives we follow the two step procedure as explained in Section 4 of the main manuscript. We use the DenseFlow-25-6 [14] to generate synthetic samples. We pretrain the flow on inlier images, e.g. CIFAR-10 or ImageNet-200, for 300 epoch following the hyperparameters from [14]. In the first step, we jointly train the flow pretrained on inlier images and a randomly initialized K-way classifier with the ResNet-18 backbone according to Equation 6 from the main manuscript. We train for 100 epochs and use the same hyperparameters as described above. In the second step, we freeze the flow and add the K+1-th logit to the classifier and finetune for 20 epochs. We set the learning rate for the backbone to 0.0001 and 0.01 for the final fully connected layer. We set the loss modulation parameter from Equation 6 to 0.03.

E Additional results

We provide a discussion on the choice of negative data, an alternative implementation of negative objectness and the full results on the OpenOOD [69] benchmark.

E.1 Impact of synthetic negatives

Table 2 shows the performance of UNO depending on the source of negative training data. We experiment with random crops from inlier scenes, synthetic negatives generated by a jointly trained normalizing flow, and random instances from the ADE datasets. Training on synthetic negatives is more beneficial than training on inlier even though the flow was jointly trained only for a small number of iterations. For instance, training on inlier crops yields a high FPR₉₅ on Fishyscapes Lost&Found. Contrary, synthetic negatives yield low FPR₉₅ on all three validations sets. Still, there is a performance gap between models that are trained with and without real negative data.

Table 2: Performance of UNO for different training negatives in per-pixel outlier segmentation.

Training negatives	FS L&F		FS Static		RoadAnomaly	
	AP	FPR ₉₅	AP	FPR ₉₅	AP	FPR ₉₅
Inlier crops	67.2	62.5	79.2	0.9	86.5	7.6
Synthetic negatives	74.5	6.9	96.9	0.1	82.4	9.2
ADE20k negatives	81.8	1.3	98.0	0.04	88.5	7.4

E.2 Alternative implementation of the outlier posterior

The outlier posterior $P(y_{\text{NO}}|\mathbf{z})$ can alternatively be modeled with an additional out-of-distribution head [4]. Then, we introduce an additional binary cross-entropy loss term to train the out-of-distribution head that discriminates inliers and outliers. This way the closed set classifier ends up with K classes and is not affected by the negative data. When applied to the mask-wide recognition architecture, the OOD head has 3 outputs 1) inlier, 2) outlier and 3) no-object. Table 3 shows the comparison of the K+2-way classifier proposed in the main paper with a K+1-way classifier and a 3-way OOD head and ablation of UNO components in the pixel-wise outlier detection setup. Our original UNO formulation built atop K+2-way classifier consistently outperforms alternative formulations across all datasets and metrics.

Table 3: Comparison of the K+2-way classifier with the OOD head atop of Mask2Former architecture on Fishyscapes val and RoadAnomaly.

Method	Score	FS L&F		FS Static		RoadAnomaly	
		AP	FPR ₉₅	AP	FPR ₉₅	AP	FPR ₉₅
K+2-way classifier	S _{UNO}	81.8	1.3	98.0	0.0	88.5	7.4
K-way classifier & OOD head	S _{UNO}	81.4	5.3	89.0	0.3	80.6	10.6
K-way classifier & OOD head	S _{Unc}	77.8	2.2	86.4	1.6	76.1	8.4
K-way classifier & OOD head	S _{NO}	79.9	7.3	88.3	0.3	74.5	25.2

Table 4 presents a similar analysis in the image-wide setting using the OpenOOD CIFAR-10 setup. Again, the K+1-way classifier combined with our UNO score outperforms the K-way classifier and a binary OOD head. Still, our UNO score works well even with the alternative formulation.

Table 4: Comparison of the K+1-way classifier with the binary OOD head atop of ResNet-18 on OpenOOD CIFAR-10. We use UNO as the outlier score.

Method	Near-OOD		Far-OOD	
	AUC	FPR ₉₅	AUC	FPR ₉₅
K+1-way classifier	95.00	18.75	97.95	9.30
K-way classifier & OOD head	94.23	19.31	97.76	11.12

E.3 Full results on OpenOOD

Tables 5 and 6 provide the extended results on the OpenOOD [55] benchmark. We compare with post-hoc methods (upper section) and training methods without (middle section) and with the use of real negative data (bottom section). All results are averaged over three runs with variances in subscripts.

References

- [1] Abhijit Bendale and Terrance E. Boult. Towards open set deep networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 1563–1572. IEEE Computer Society, 2016.
- [2] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Dense open-set recognition based on training with noisy negative images. *Image and Vision Computing*, 2022.
- [3] Julian Bitterwolf, Maximilian Müller, and Matthias Hein. In or out? fixing imagenet out-of-distribution detection evaluation. *arXiv preprint arXiv:2306.00826*, 2023.
- [4] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *International Journal of Computer Vision*, 2021.
- [5] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, and Matthias Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. *arXiv preprint arXiv:2104.14812*, 2021.
- [6] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5128–5137, 2021.
- [7] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(11):8065–8081, 2022.
- [8] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022.

Table 5: OOD detection performance on the OpenOOD benchmark, the CIFAR-10 dataset.

Method	Near-OOD			Far-OOD			Acc.
	AUC	FPR ₉₅	AP	AUC	FPR ₉₅	AP	
OpenMax [10]	87.62 _(±0.29)	43.62 _(±2.27)	80.60 _(±0.29)	89.62 _(±0.19)	29.69 _(±1.21)	90.19 _(±0.41)	95.06 _(±0.30)
MSP [10]	88.03 _(±0.25)	48.17 _(±3.92)	85.43 _(±0.36)	90.73 _(±0.43)	31.72 _(±1.84)	93.27 _(±0.14)	95.06 _(±0.30)
TempScale [19]	88.09 _(±0.31)	50.96 _(±4.32)	86.11 _(±0.33)	90.97 _(±0.52)	33.48 _(±2.39)	93.68 _(±0.11)	95.06 _(±0.30)
ODIN [6]	82.87 _(±1.85)	76.19 _(±6.08)	83.03 _(±1.15)	87.96 _(±0.61)	57.62 _(±4.24)	93.14 _(±0.44)	95.06 _(±0.30)
MDS [6]	84.20 _(±2.40)	49.90 _(±3.98)	79.88 _(±3.18)	89.72 _(±1.36)	32.22 _(±3.40)	93.81 _(±0.74)	95.06 _(±0.30)
MDSEns [6]	60.43 _(±0.26)	92.26 _(±0.20)	59.94 _(±0.19)	73.90 _(±0.27)	61.47 _(±0.48)	83.37 _(±0.04)	95.06 _(±0.30)
RMDS [6]	89.80 _(±0.28)	38.89 _(±2.39)	87.52 _(±0.29)	92.20 _(±0.21)	25.35 _(±0.73)	94.21 _(±0.10)	95.06 _(±0.30)
Gram [10]	58.66 _(±4.83)	90.87 _(±1.91)	57.57 _(±5.09)	71.73 _(±3.20)	72.34 _(±6.73)	82.89 _(±3.14)	95.06 _(±0.30)
EBO [6]	87.58 _(±0.46)	61.34 _(±4.63)	87.04 _(±0.27)	91.21 _(±0.92)	41.69 _(±5.32)	94.31 _(±0.09)	95.06 _(±0.30)
OpenGAN [27]	53.71 _(±7.68)	94.48 _(±4.01)	53.35 _(±5.22)	54.61 _(±15.51)	83.52 _(±11.63)	73.34 _(±8.49)	95.06 _(±0.30)
GradNorm [26]	54.90 _(±0.98)	94.72 _(±0.82)	57.95 _(±1.98)	57.55 _(±3.22)	91.90 _(±2.23)	76.75 _(±1.95)	95.06 _(±0.30)
ReAct [6]	87.11 _(±0.61)	63.56 _(±7.33)	86.65 _(±0.19)	90.42 _(±1.41)	44.90 _(±8.37)	93.99 _(±0.45)	95.06 _(±0.30)
MLS [25]	87.52 _(±0.47)	61.32 _(±4.62)	86.88 _(±0.29)	91.10 _(±0.89)	41.68 _(±5.27)	94.21 _(±0.06)	95.06 _(±0.30)
KLM [25]	79.19 _(±0.80)	87.86 _(±6.37)	80.37 _(±0.46)	82.68 _(±0.21)	78.31 _(±4.84)	90.57 _(±0.25)	95.06 _(±0.30)
VIM [6]	88.68 _(±0.28)	44.84 _(±2.31)	86.32 _(±0.39)	93.48 _(±0.24)	25.05 _(±0.52)	96.27 _(±0.24)	95.06 _(±0.30)
KNN [16]	90.64 _(±0.20)	34.01 _(±7.64)	88.50 _(±0.35)	92.96 _(±0.14)	24.27 _(±0.40)	94.93 _(±0.07)	95.06 _(±0.30)
DICE [10]	78.34 _(±0.79)	70.04 _(±0.38)	74.80 _(±2.33)	84.23 _(±1.89)	51.76 _(±4.42)	89.06 _(±1.72)	95.06 _(±0.30)
RankFeat [10]	79.46 _(±2.52)	60.88 _(±4.60)	74.46 _(±3.08)	75.87 _(±5.06)	57.44 _(±7.99)	81.27 _(±3.67)	95.06 _(±0.30)
ASH [16]	75.27 _(±1.04)	86.78 _(±1.82)	77.24 _(±1.26)	78.49 _(±2.58)	79.03 _(±4.22)	88.33 _(±1.39)	95.06 _(±0.30)
SHE [6]	81.54 _(±0.51)	79.65 _(±3.47)	82.04 _(±0.51)	85.32 _(±1.43)	66.48 _(±5.98)	91.26 _(±0.04)	95.06 _(±0.30)
ConfBranch [10]	89.84 _(±0.24)	31.28 _(±0.66)	85.50 _(±0.30)	92.85 _(±0.29)	94.88 _(±0.05)	93.48 _(±0.39)	94.88 _(±0.05)
RotPred [10]	92.68 _(±0.27)	28.14 _(±1.68)	90.47 _(±0.35)	96.62 _(±0.18)	12.23 _(±0.33)	97.54 _(±0.13)	95.35 _(±0.52)
G-ODIN [26]	89.12 _(±0.57)	45.54 _(±2.52)	88.25 _(±0.49)	95.51 _(±0.31)	21.45 _(±1.91)	97.35 _(±0.34)	94.70 _(±0.25)
CSI [16]	89.51 _(±0.19)	33.66 _(±0.64)	86.37 _(±0.25)	92.00 _(±0.30)	26.42 _(±0.29)	93.90 _(±0.33)	91.16 _(±0.14)
ARPL [0]	87.44 _(±0.15)	40.33 _(±0.70)	82.96 _(±0.33)	89.31 _(±0.32)	32.39 _(±0.74)	91.41 _(±0.09)	93.66 _(±0.11)
MOS [16]	71.45 _(±3.09)	78.72 _(±5.86)	72.41 _(±3.05)	76.41 _(±5.93)	62.90 _(±6.62)	85.24 _(±2.92)	94.83 _(±0.37)
VOS [16]	87.70 _(±0.48)	57.03 _(±1.92)	86.57 _(±0.73)	90.83 _(±0.92)	40.43 _(±4.53)	93.95 _(±0.56)	94.31 _(±0.64)
LogitNorm [20]	92.33 _(±0.08)	29.34 _(±0.81)	90.62 _(±0.09)	96.74 _(±0.06)	13.81 _(±0.20)	97.29 _(±0.21)	94.30 _(±0.25)
CIDER [6]	90.71 _(±0.16)	32.11 _(±0.94)	87.97 _(±0.24)	94.71 _(±0.36)	20.72 _(±0.85)	96.19 _(±0.19)	-
NPOS [16]	89.78 _(±0.33)	32.64 _(±0.70)	86.36 _(±0.68)	94.07 _(±0.49)	20.59 _(±0.69)	96.20 _(±0.43)	-
UNO (ours)	91.34 _(±0.33)	31.78 _(±0.82)	87.39 _(±0.35)	92.55 _(±0.25)	20.54 _(±0.87)	93.96 _(±0.25)	95.20 _(±0.25)
MixOE [16]	88.73 _(±0.82)	51.45 _(±7.78)	94.25 _(±0.17)	91.93 _(±0.69)	33.84 _(±4.77)	98.00 _(±0.02)	94.55 _(±0.32)
MCD [26]	91.03 _(±0.12)	30.17 _(±0.06)	87.73 _(±0.36)	91.00 _(±1.10)	32.03 _(±4.21)	94.45 _(±0.85)	94.95 _(±0.04)
UDG [16]	89.91 _(±0.25)	35.34 _(±0.95)	86.89 _(±0.84)	94.06 _(±0.90)	20.35 _(±2.41)	95.23 _(±0.89)	92.36 _(±0.84)
OE [16]	94.82 _(±0.21)	19.84 _(±0.95)	87.39 _(±0.60)	96.00 _(±0.13)	13.13 _(±0.53)	95.03 _(±0.12)	94.63 _(±0.26)
UNO (ours)	94.87 _(±0.07)	9.33 _(±0.50)	94.49 _(±0.13)	97.63 _(±0.72)	9.38 _(±2.65)	99.10 _(±0.25)	94.88 _(±0.19)

Table 6: OOD detection performance on the OpenOOD benchmark, the Imagenet-200 dataset.

Method	Near-OOD			Far-OOD			Acc.
	AUC	FPR ₉₅	AP	AUC	FPR ₉₅	AP	
OpenMax [10]	80.27(±0.10)	63.48(±0.25)	81.42(±0.19)	90.20(±0.17)	33.12(±0.66)	85.13(±0.47)	86.37(±0.08)
MSP [10]	83.34(±0.06)	54.82(±0.35)	85.95(±0.05)	90.13(±0.09)	35.43(±0.38)	88.71(±0.14)	86.37(±0.08)
TempScale [10]	83.69(±0.04)	54.82(±0.23)	86.29(±0.02)	90.82(±0.09)	34.00(±0.37)	89.49(±0.15)	86.37(±0.08)
ODIN [6]	80.27(±0.08)	66.76(±0.26)	85.02(±0.03)	91.71(±0.19)	34.23(±1.05)	91.29(±0.15)	86.37(±0.08)
MDS [6]	61.93(±0.51)	79.11(±0.31)	67.68(±0.42)	74.72(±0.26)	61.66(±0.27)	70.80(±0.61)	86.37(±0.08)
MDSEns [6]	54.32(±0.24)	91.75(±0.10)	64.81(±0.24)	69.27(±0.57)	80.96(±0.38)	69.62(±0.52)	86.37(±0.08)
RMDS [6]	82.57(±0.25)	54.02(±0.58)	83.07(±0.45)	88.06(±0.34)	32.45(±0.79)	82.71(±0.78)	86.37(±0.08)
Gram [6]	67.67(±1.07)	86.40(±1.21)	75.63(±0.78)	71.19(±0.24)	84.36(±0.78)	72.75(±0.25)	86.37(±0.08)
EBO [6]	82.50(±0.05)	60.24(±0.57)	85.48(±0.07)	90.86(±0.21)	34.86(±1.30)	89.85(±0.21)	86.37(±0.08)
OpenGAN [10]	59.79(±3.39)	84.15(±3.85)	66.85(±2.79)	73.15(±4.07)	64.16(±9.33)	66.62(±3.69)	86.37(±0.08)
GradNorm [10]	72.75(±0.48)	82.67(±0.30)	80.19(±0.68)	84.26(±0.87)	66.45(±0.22)	86.54(±0.92)	86.37(±0.08)
ReAct [10]	81.87(±0.98)	62.49(±2.19)	85.38(±0.34)	92.31(±0.56)	28.50(±0.95)	91.31(±0.80)	86.37(±0.08)
MLS [10]	82.90(±0.04)	59.76(±0.59)	85.96(±0.07)	91.11(±0.19)	34.03(±1.21)	90.10(±0.21)	86.37(±0.08)
KLM [10]	80.76(±0.08)	70.26(±0.64)	83.41(±0.23)	88.53(±0.11)	40.90(±1.08)	84.22(±0.47)	86.37(±0.08)
VIM [10]	78.68(±0.24)	59.19(±0.71)	81.61(±0.29)	91.26(±0.19)	27.20(±0.30)	90.01(±0.35)	86.37(±0.08)
KNN [10]	81.57(±0.17)	60.18(±0.52)	85.72(±0.17)	93.16(±0.22)	27.27(±0.75)	93.48(±0.15)	86.37(±0.08)
DICE [10]	81.78(±0.14)	61.88(±0.67)	85.37(±0.13)	90.80(±0.31)	36.51(±1.18)	90.55(±0.29)	86.37(±0.08)
RankFeat [10]	56.92(±1.59)	92.06(±0.23)	66.17(±1.63)	38.22(±3.85)	97.72(±0.75)	45.25(±2.81)	86.37(±0.08)
ASH [10]	82.38(±0.19)	64.89(±0.90)	87.03(±0.06)	93.90(±0.27)	27.29(±1.12)	94.15(±0.32)	86.37(±0.08)
SHE [10]	80.18(±0.25)	66.80(±0.74)	84.20(±0.28)	89.81(±0.61)	42.17(±1.24)	90.05(±0.62)	86.37(±0.08)
ConfBranch [10]	79.10(±0.24)	61.44(±0.34)	82.11(±0.30)	90.43(±0.18)	34.75(±0.63)	88.67(±0.27)	85.92(±0.07)
RotPred [10]	81.59(±0.20)	60.42(±0.60)	84.87(±0.19)	92.56(±0.09)	26.16(±0.38)	90.10(±0.08)	86.37(±0.16)
G-ODIN [10]	77.28(±0.10)	69.87(±0.46)	82.77(±0.16)	92.33(±0.11)	30.18(±0.49)	92.04(±0.10)	84.56(±0.28)
ARPL [10]	82.02(±0.10)	55.74(±0.70)	84.35(±0.08)	89.23(±0.11)	36.46(±0.08)	87.63(±0.19)	83.95(±0.32)
MOS [10]	69.84(±0.46)	71.60(±0.48)	73.38(±0.56)	80.46(±0.92)	51.56(±0.42)	72.79(±1.43)	85.60(±0.20)
VOS [10]	82.51(±0.11)	59.89(±0.47)	85.59(±0.07)	91.00(±0.28)	34.01(±0.97)	90.11(±0.30)	86.23(±0.19)
LogitNorm [10]	82.66(±0.15)	56.46(±0.37)	86.41(±0.08)	93.04(±0.21)	26.11(±0.52)	92.25(±0.32)	86.04(±0.15)
CIDER [10]	80.58(±1.75)	60.10(±0.73)	83.32(±1.76)	90.66(±1.68)	30.17(±2.75)	89.16(±2.38)	-
NPOS [10]	79.40(±0.39)	62.09(±0.05)	84.37(±0.35)	94.49(±0.07)	21.76(±0.21)	94.83(±0.07)	-
UNO (ours)	81.16(±0.65)	61.10(±0.93)	85.31(±0.94)	92.53(±0.48)	32.32(±0.19)	87.24(±0.33)	86.33(±0.36)
OE [10]	84.84(±0.16)	52.30(±0.67)	86.86(±0.22)	89.02(±0.18)	34.17(±0.56)	85.15(±0.26)	85.82(±0.21)
MCD [10]	83.62(±0.09)	54.71(±0.83)	84.44(±0.30)	88.94(±0.10)	29.93(±0.30)	82.90(±0.48)	86.12(±0.17)
UDG [10]	74.30(±1.63)	68.89(±1.72)	78.09(±2.02)	82.09(±2.78)	62.04(±5.99)	81.63(±3.29)	68.11(±1.24)
MixOE [10]	82.62(±0.03)	57.97(±0.40)	84.78(±0.05)	88.27(±0.41)	40.93(±0.29)	86.01(±0.60)	85.71(±0.07)
UNO (ours)	85.07(±0.78)	51.71(±0.31)	85.29(±0.71)	89.63(±0.46)	36.79(±0.21)	87.07(±0.14)	86.42(±0.32)

- [9] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3606–3613, 2014.
- [10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.
- [11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, 2009.
- [12] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6):141–142, 2012.
- [13] Terrance DeVries and Graham W. Taylor. Learning confidence for out-of-distribution detection in neural networks. *CoRR*, abs/1802.04865, 2018. URL <http://arxiv.org/abs/1802.04865>.
- [14] Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. *arXiv preprint arXiv:2209.09858*, 2022.
- [15] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. VOS: learning what you don’t know by virtual outlier synthesis. In *The Tenth International Conference on Learning Representations, ICLR*, 2022.
- [16] Matej Grcić, Petra Bevandić, and Siniša Šegvić. Dense anomaly detection by robust learning on synthetic negative data. *arXiv preprint arXiv:2112.12833*, 2021.
- [17] Matej Grcić, Ivan Grubišić, and Siniša Šegvić. Densely connected normalizing flows. *Advances in Neural Information Processing Systems*, 2021.
- [18] Matej Grcić, Josip Šarić, and Siniša Šegvić. On advantages of mask-level recognition for outlier-aware segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2936–2946, 2023.
- [19] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [21] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

- [22] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *Advances in neural information processing systems*, 32, 2019.
- [23] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 8759–8773. PMLR, 2022.
- [24] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized ODIN: detecting out-of-distribution image without learning from out-of-distribution data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 10948–10957. Computer Vision Foundation / IEEE, 2020.
- [25] Rui Huang and Yixuan Li. MOS: towards scaling out-of-distribution detection for large semantic space. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 8710–8719. Computer Vision Foundation / IEEE, 2021.
- [26] Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. In Marc’Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 677–689, 2021.
- [27] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 793–802. IEEE, 2021.
- [28] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.
- [29] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 and cifar-100 datasets. URL: <https://www.cs.toronto.edu/kriz/cifar.html>, 6(1):1, 2009.
- [30] L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 2003.
- [31] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.
- [32] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 7167–7177, 2018.

- [33] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.
- [34] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2152–2161, 2019.
- [35] Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. Energy-based out-of-distribution detection. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [36] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [37] Yifei Ming, Yiyu Sun, Ousmane Dia, and Yixuan Li. How to exploit hyperspherical embeddings for out-of-distribution detection? In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. Open-Review.net, 2023.
- [38] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.
- [39] Jie Ren, Stanislav Fort, Jeremiah Z. Liu, Abhijit Guha Roy, Shreyas Padhy, and Balaji Lakshminarayanan. A simple fix to mahalanobis distance for improving near-ood detection. *CoRR*, abs/2106.09022, 2021.
- [40] Chandramouli Shama Sastry and Sageev Oore. Detecting out-of-distribution examples with gram matrices. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 8491–8501. PMLR, 2020.
- [41] Yue Song, Nicu Sebe, and Wei Wang. Rankfeat: Rank-1 feature removal for out-of-distribution detection. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.
- [42] Yiyu Sun and Yixuan Li. DICE: leveraging sparsification for out-of-distribution detection. In Shai Avidan, Gabriel J. Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXIV*, volume 13684 of *Lecture Notes in Computer Science*, pages 691–708. Springer, 2022.
- [43] Yiyu Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. In Marc’Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin,

- Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 144–157, 2021.
- [44] Yiyu Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 20827–20840. PMLR, 2022.
- [45] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. CSI: novelty detection via contrastive learning on distributionally shifted instances. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [46] Leitian Tao, Xuefeng Du, Jerry Zhu, and Yixuan Li. Non-parametric outlier synthesis. In *The Eleventh International Conference on Learning Representations, ICLR*, 2023.
- [47] Grant Van Horn, Oisín Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018.
- [48] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need? *arXiv preprint arXiv:2110.06207*, 2021.
- [49] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4921–4930, 2022.
- [50] Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 23631–23644. PMLR, 2022.
- [51] Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8301–8309, 2021.
- [52] Qing Yu and Kiyoharu Aizawa. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9518–9526, 2019.
- [53] Netzer Yuval. Reading digits in natural images with unsupervised feature learning. In *Proceedings of the NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.
- [54] Jingyang Zhang, Nathan Inkawich, Randolph Linderman, Yiran Chen, and Hai Li. Mixture outlier exposure: Towards out-of-distribution detection in fine-grained environments. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5531–5540, 2023.

- [55] Jingyang Zhang, Jingkan Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Hao-ran Zhang, Yiyu Sun, Xuefeng Du, Kaiyang Zhou, Wayne Zhang, et al. Openood v1. 5: Enhanced benchmark for out-of-distribution detection. *arXiv preprint arXiv:2306.09301*, 2023.
- [56] Jinsong Zhang, Qiang Fu, Xu Chen, Lun Du, Zelin Li, Gang Wang, Xiaoguang Liu, Shi Han, and Dongmei Zhang. Out-of-distribution detection based on in-distribution data patterns memorization with modern hopfield energy. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.
- [57] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017.
- [58] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127:302–321, 2019.