Appendix - Improving Out-of-Distribution Detection Using Segmented Images and Cross-View Attention Fusion

A. Dataset Details

We describe the datasets used in our experiments here. In all cases except for our ImageNet subset datasets, we follow the dataset specifications in the well-accepted, state-of-the-art OOD detection benchmark OpenOOD v1.5 [168].

- **ImageNet** - [70] is a large-scale image dataset containing 1000 unique classes across a diverse range of objects and scenes. It contains 1,281,167 training images and 50,000 validation images which are often treated as the testing dataset as labels are not released for the actual ImageNet testing dataset. For the ImageNet-100-Random subset, we choose distinct classes from within ImageNet-1k. We detail all ImageNet-1k class IDs for ImageNet-100-Random mentioned in the main paper in Section A.1.
- **Textures** - [19] compiles images of different textures into a dataset containing classes uniquely describing textures such as "*polka dotted*" and "*bumpy*". We have used the entire dataset of 5,640 images for the OOD dataset.
- **iNaturalist** - [141] is an image classification dataset oriented around pictures of different species. For our OOD dataset, we have used 10,000 images sampled from the test set, following [158].
- **Places** - [176] (or Places365) is a dataset containing exclusive scenes instead of the traditional objects. We have sampled 10,000 images randomly from the test set for the OOD dataset.
- **SUN** - [60, 152], or the Scene Understanding dataset, is a large dataset containing images of scenes. We have used 10,000 images randomly sampled from its test set for the OOD dataset.
- **OpenImage-O** - [144] is a dataset constructed by randomly selecting image from the OpenImage-V3 dataset. OpenImage-V3 is built from the Flickr database and is not annotated with classes to avoid bias and overlap with ImageNet. The OOD dataset contains all 17,632 images from OpenImage-O.
- **Species** - [53] is an image dataset containing thousands of different species, none of which overlap with ImageNet-21k. We use 10,000 images from the OpenOOD v1.5 benchmark [168] as the OOD dataset.
- **NINCO** - [11] is an image dataset curated to challenge OOD detection models trained on ImageNet-1k. It consists of 64 OOD classes for a total of 5,879 object images with no categorical overlap to ImageNet-1k.
- **SSB-Hard** - [143] is a subset of the ImageNet-21K [118] manually curated to be a challenging near-OOD dataset for ImageNet-1k. We use the full 49,000 images as the OOD dataset.
- **Food** - [12] is an image dataset containing 101 classes of various food items and dishes. The validation dataset of 25,250 images as the OOD dataset. We remove classes overlapping with ImageNet-1k: Apple Pie, Breakfast Burrito, Chocolate Mousse, Guacamole, Hamburger, Hot Dog, Ice Cream, Pizza.

Following the implementations if OpenOOD v1.5 [168], we transform batches of data during training using randomized data augmentations to increase generalizability and encourage feature learning. We use the following set of transformations for random data augmentations: random crop, horizontal flip, random erasing, and normalization.

All images are resized to size 256×256 before processing. To ensure consistent foreground-background segmentation across datasets and maintain comparability, all ID and OOD datasets are resized to 512×512 *before* segmentation.

When applying augmentations to a sample, each augmentation is generated once and applied identically to all three views of an image. This ensures that data points maintain consistent features across views when presented to CASOD.

A.1. ImageNet Subset Class IDs

- ImageNet-100-Random: *'n01770081', 'n03450230', 'n02977058', 'n01675722', 'n03017168', 'n02098413', 'n03372029', 'n04418357', 'n01824575', 'n02480495', 'n04008634', 'n03447447', 'n02115913', 'n02051845', 'n02104029', 'n03452741', 'n02129165', 'n02417914', 'n03032252', 'n03590841', 'n04350905', 'n02966193', 'n03956157', 'n03733131', 'n03197337', 'n01882714', 'n02363005', 'n07717410', 'n03457902', 'n01689811', 'n02110185', 'n04357314', 'n04370456', 'n04228054', 'n01641577', 'n09288635', 'n02395406', 'n07583066', 'n07715103', 'n04200800', 'n03958227', 'n03920288', 'n02410509', 'n02483362', 'n02123045', 'n04596742', 'n02229544', 'n02086079', 'n04238763', 'n02086646', 'n02869837', 'n02422699', 'n01665541', 'n02114855',*

'n04483307', 'n03485407', 'n03710721', 'n03998194', 'n03627232', 'n02823428', 'n01795545', 'n12144580', 'n01751748', 'n02999410', 'n02992529', 'n02795169', 'n03602883', 'n02808304', 'n02167151', 'n02504458', 'n03026506', 'n01682714', 'n04515003', 'n02097658', 'n03874599', 'n02107142', 'n03977966', 'n06596364', 'n02007558', 'n02814860', 'n03394916', 'n02840245', 'n02817516', 'n03908714', 'n02093859', 'n02102318', 'n04179913', 'n04591157', 'n04070727', 'n02165456', 'n03594734', 'n04467665', 'n01990800', 'n02086240', 'n04152593', 'n03777754', 'n01945685', 'n02727426', 'n03840681', 'n04606251'

- ImageNet-30 (OOD for ImageNet-100-Random): *'n02012849', 'n02109047', 'n02102040', 'n04389033', 'n04548280', 'n02790996', 'n09193705', 'n02231487', 'n13133613', 'n02096177', 'n03929660', 'n13044778', 'n02113624', 'n02114548', 'n04254120', 'n03594945', 'n07565083', 'n03126707', 'n01871265', 'n02917067', 'n04285008', 'n01806143', 'n02066245', 'n04252077', 'n02841315', 'n04277352', 'n03633091', 'n02843684', 'n03160309', 'n03379051'*
- ImageNet-100 (OOD for ImageNet-200): *'n02109047', 'n07716906', 'n04515003', 'n02089867', 'n02009229', 'n02090622', 'n03623198', 'n01945685', 'n01877812', 'n03633091', 'n01491361', 'n04204347', 'n01742172', 'n04428191', 'n03958227', 'n07579787', 'n03761084', 'n07892512', 'n01622779', 'n03208938', 'n04579432', 'n03617480', 'n02666196', 'n04252077', 'n04525038', 'n02397096', 'n03871628', 'n02115641', 'n02096051', 'n02930766', 'n03706229', 'n01773797', 'n02106382', 'n06785654', 'n02444819', 'n02074367', 'n02966687', 'n03935335', 'n07831146', 'n02107683', 'n02514041', 'n12998815', 'n02096177', 'n02328150', 'n04111531', 'n04367480', 'n03938244', 'n04371774', 'n13040303', 'n02110627', 'n03297495', 'n07590611', 'n04259630', 'n04039381', 'n01689811', 'n04228054', 'n04326547', 'n09468604', 'n02815834', 'n01873310', 'n03160309', 'n03028079', 'n03673027', 'n03063689', 'n01819313', 'n02640242', 'n02676566', 'n03041632', 'n02100236', 'n04371430', 'n03742115', 'n03814906', 'n04330267', 'n03207941', 'n02104029', 'n03770679', 'n02132136', 'n02101006', 'n04074963', 'n02509815', 'n03838899', 'n02389026', 'n02025239', 'n02167151', 'n03532672', 'n04204238', 'n03089624', 'n06596364', 'n02101388', 'n02108422', 'n03729826', 'n02112706', 'n03196217', 'n02097209', 'n02488702', 'n04579145', 'n02090379', 'n03394916', 'n03924679', 'n02783161'*

B. Experiment Details

The codebase for CASOD can be found at https://github.com/alexp205/CASOD. All used datasets are publicly available. The foreground-background separation procedure used to create the multi-view datasets for CASOD is included in the codebase and can be used for re-creation of the data used in the project.

B.1. Hyperparameter Details

The Vision Transformer [28] models for ImageNet [70] data experiments are trained using the configuration described in Section 4.1 of the main paper. Label smoothing [134] is a regularization technique that prevents model overconfidence. We use 0.1 for the degree of label smoothing.

The Vision Transformer configuration used is ViT-B-16 [168]. For the attention modules in our proposed CVCA fusion method, we use the following hyperparameters: 8 heads, head dimension of 96, and dropout of 0.5. Under these settings, T was 197 and d was 768 (see Sec. 3 of the main paper for definitions).

CASOD is trained until convergence with a batch size of 512 for I-1k and 256 for ImageNet-100-Random. We use SGD optimizer, learning rate of 0.001, Nesterov momentum of 0.9, weight decay of 5e-4, and cosine annealing learning rate scheduling after 5 epochs of warmup. We use $B = 2$ cross-view feature fusion blocks. LoRA rank was 16 for ImageNet-100-Random and 128 for I-1k.

The code uses SciPy, PyTorch Lightning, and FAISS [64]. Experiments were run on eight Nvidia A100 GPUs.

B.2. Foreground-Background Separation Tool Details

The specific image segmentation model used for foreground-background separation was YOLOv5 [63]. We used an open-source, pre-trained YOLOv5 processing pipeline[1] as our foreground-background separation toolkit.

C. Additional Results

C.1. Results on ImageNet-200

We additionally consider the experimental setting using the popular ImageNet-200 ID dataset as defined in OpenOOD v1.5 [168]. Building on OpenOOD v1.5, we use Textures, iNaturalist, Places, SUN, OpenImage-O, and Species as far-OOD datasets and NINCO, SSB-Hard, and Food as near-OOD datasets. Similar to our ImageNet-100-Random experiment (see Sec. 4 of the main paper) we additionally include a distinct subset of 100 randomly-selected classes (ImageNet-100) and the remaining 800 classes (ImageNet-800) as near-OOD datasets simulating very-similar semantic distributions to the ID dataset. The classes used for ImageNet-200 ID can be found in the OpenOOD v1.5 implementation [168], and the class IDs for the ImageNet-100 OOD dataset can be found in Section A.1.

For this setting, we train for 20 epochs with a batch size of 256. LoRA rank was 16. All other training and model

[1] https://github.com/ultralytics/yolov5

configurations were identical to the ImageNet-100-Random ID experiment settings.

Results can be found in Table 1 for near-OOD datasets and in Table 2 for far-OOD datasets. All numeric values are percentages and the average of 3 random runs. ImageNet-200 ID experiment standard deviations and additional metrics are in Sec. C.2. We note that performance discrepancies compared to other works originate from our use of a self-supervised pre-trained ViT backbone with a learnable Adapter and classification head. Previous works tune ViT models pre-trained on ImageNet-21k which we believe to be an unfair comparison due to data leakage.

C.2. Results with Error Bars

We include additional OOD detection metrics and standard deviation error bars here for the ImageNet-100-Random experiment presented in the main paper and for the aforementioned ImageNet-200 experiment setting. The results are divided into OOD detection metric: AUC (Table 3), FPR95 (Table4), Area Under the Precision-Recall curve (AUPR-IN and AUPR-OUT) (Tables 5 and 6 respectively), and Detection Error (DE) (Table 7) [158, 168]. Note that ImageNet-100-Random ID results are in the top-half of the tables and ImageNet-200 ID results are in the bottom-half. All numeric values are percentages and the average or standard deviation of 3 random runs.

C.3. Parameter Cost and Runtime Analysis Details

Parameter Cost - Details. As mentioned in the main paper, CASOD uses a *frozen* pre-trained ViT-B-16 as a feature extractor and only requires the LoRA modules and feature fusion component to be trained. The standard multi-view classification technique of training separate backbones for each view requires about 259 million (M) trainable parameters for three views. CASOD only requires 48 M trainable parameters for three views, a parameter efficiency improvement of **81%**. CASOD also improves over fully fine-tuning a single ViT backbone which involves 86 M trainable parameters, an efficiency improvement of **44%**.

We profiled CASOD and a single-view baseline model for comparison. The CASOD model alone consumes 36.68 MiB while the single-view baseline model consumes 2.67 MiB. Training CASOD with a batch size of 128 consumes 52.54 GiB of memory per GPU while a single-view baseline consumes 13.67 GiB. These values scale as expected with batch size. With a batch size of 64, CASOD consumes 26.27 GiB while a single-view baseline consumes 8.19 GiB.

Runtime Analysis - Details. CASOD uses three sets of LoRA adapters for the feature extractors and a feature fusion component (FF, see main paper Section 3). Parallelization collapses the forward pass time requirement of three view feature extractors into an equivalent requirement of a single feature extractor operation.

The FF in CASOD uses B blocks each containing a constant number of CVCA modules. Each CVCA module has a runtime complexity of $O(T^2d + Td^2)$ [142]. Thus, the FF incurs an additional cost of $O(BT^2d + BTd^2)$ in the forward pass, which in practice is dominated by $O(Td^2)$ due to the relative small sizes of B and $T < d$. In the backward pass, CASOD must update the parameters of all three sets of Adapters and the FF.

On our system, we observed training times for CASOD approximately three times that of single-view baselines (which only need to learn a single linear layer). Specifically, under our training settings (see Appendix Section B.1), one epoch of training required 169 seconds walltime for CASOD and 51 seconds walltime for a single-view baseline. With a batch size of 64, CASOD consumed 181 seconds walltime while a single-view baseline consumed 54 seconds walltime. To empirically evaluate the runtime of CASOD in a deployment environment, we evaluated the average time of a foreground-background separation step and forward pass for CASOD and a single-view baseline (i.e. no separation and only one set of Adapters). The additional overhead we observed for the CASOD forward pass (i.e. 0.07 seconds) is due to the FF and movement of multiple view tensors between GPUs. As mentioned in the main paper, CASOD takes more time than single-view baselines but we believe the latency is low enough to still be feasible for many applications.

D. CASOD with Other OOD Detection Methods

CASOD is directly compatible with any post-hoc OOD detection method that makes use of logits (using $\hat{y}_{fused}$) or features (using z^0_{fused}). We present additional OOD detection results (AUC) for ImageNet-100-Random ID, ImageNet-200 ID, and ImageNet-1k ID using OOD detection methods MSP [51] and KNN [133] in Table 8. MSP is a logits-based method and KNN is a recent highly-competitive features-based method [168] like MD (which is presented in the main paper). We also include MD OOD detection with CASOD (the original proposed method) here. CASOD with MD is the best overall performer. Interestingly, CASOD-MSP is slightly better for some near-OOD datasets in the ImageNet-100-Random ID and ImageNet-200 ID settings. We suspect that this is due to model overconfidence benefiting detection of near-OOD samples similar to the ID distribution. Model logits are tuned to recognizing features in ID samples but overconfidence causes slight deviations (i.e. in near-OOD samples) to reduce the MSP score. This is not the case for the ImageNet-1k ID setting which is significantly harder to overfit due to the large number of classes. MD is generally better for far-OOD datasets as it avoids model overconfidence which is known to significantly weaken far-OOD detection [51, 72, 105, 151]. We also note that CASOD

Table 1. **Near-OOD Results - ImageNet-200**: OOD detection comparison (**AUC**) between CASOD and baselines on near-OOD datasets with models **trained only on** ImageNet-200 ID data. All numeric values are percentages and averages of 3 random runs. **Standard deviations** are in Sec. C.2.

	ID data: ImageNet-200					
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Average
MSP	82.16	81.27	77.53	64.13	65.21	74.06
ENERGY	69.56	69.60	59.05	52.80	51.59	60.52
KNN	81.99	82.82	83.72	72.69	72.17	78.68
VIM	79.44	82.44	86.04	77.20	73.27	79.68
MD	79.46	81.66	78.14	78.77	76.77	78.96
GEN	44.64	44.42	55.22	54.40	42.02	48.14
REACT	69.65	69.68	59.14	52.90	51.59	60.59
PCA	76.99	76.31	81.89	70.51	72.69	75.68
ASH	51.79	52.05	46.64	39.35	72.64	52.49
SHE	72.98	71.61	77.11	61.83	59.43	68.59
VRA	39.41	39.76	38.21	54.42	39.03	42.16
NECO	81.13	83.53	82.63	73.00	72.34	78.53
FDBD	84.39	84.40	82.10	70.70	73.46	79.01
CASOD	85.03	86.93	89.96	82.54	87.14	86.32

Table 2. **Far-OOD Results - ImageNet-200**: ImageNet-200 OOD detection (**AUC**) comparison. **Standard deviations** are in Sec. C.2.

	ID data: ImageNet-200						
	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Average
MSP	87.56	78.50	77.58	79.26	79.87	68.97	78.62
ENERGY	79.80	46.01	69.31	66.94	63.78	49.83	62.61
KNN	95.67	86.53	92.56	89.17	89.75	76.90	88.43
VIM	97.87	91.37	97.40	92.36	92.88	82.20	92.35
MD	92.41	78.44	90.48	82.99	82.72	79.42	84.41
GEN	34.85	70.87	39.97	47.54	48.67	64.19	51.01
REACT	79.86	46.13	69.40	67.01	63.87	49.94	62.70
PCA	90.27	79.96	90.48	86.39	86.00	73.70	84.47
ASH	49.76	55.87	35.21	33.89	55.18	56.05	47.66
SHE	91.57	76.31	89.19	86.95	81.26	63.35	81.44
VRA	11.36	31.34	32.09	30.88	26.62	47.06	29.89
NECO	94.30	82.13	90.88	89.24	87.93	77.63	87.02
FDBD	91.97	86.68	87.08	86.84	87.27	74.43	85.71
CASOD	98.48	99.42	98.26	97.81	94.14	86.60	95.79

Table 3. **Results - AUC**: AUC OOD detection results. Avg.N is mean performance for near-OOD. Avg.F is mean performance for far-OOD.

	ID data: ImageNet-100-Random													
	ImageNet-30	ImageNet-900	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	83.55 $\pm$0.05	85.44 $\pm$1.68	89.54 $\pm$0.05	85.16 $\pm$0.06	80.18 $\pm$0.53	84.77	92.71 $\pm$0.09	94.17 $\pm$0.07	88.41 $\pm$0.12	86.96 $\pm$0.06	92.65 $\pm$0.07	89.73 $\pm$0.06	90.77	88.04
ENERGY	79.40 $\pm$0.12	79.99 $\pm$0.62	81.19 $\pm$0.15	84.06 $\pm$0.26	81.86 $\pm$0.74	81.30	91.82 $\pm$0.13	82.03 $\pm$0.55	90.72 $\pm$0.21	83.85 $\pm$0.36	89.11 $\pm$0.17	78.49 $\pm$0.20	86.00	83.87
KNN	77.32 $\pm$0.13	78.33 $\pm$0.81	80.57 $\pm$0.46	73.28 $\pm$0.31	65.76 $\pm$0.86	75.05	96.10 $\pm$0.11	80.39 $\pm$0.69	86.74 $\pm$0.36	80.29 $\pm$0.41	87.80 $\pm$0.37	77.33 $\pm$0.67	84.78	80.36
VIM	78.31 $\pm$0.10	79.55 $\pm$0.98	86.13 $\pm$0.33	80.18 $\pm$0.25	73.17 $\pm$0.80	79.47	98.99 $\pm$0.02	94.05 $\pm$0.36	95.89 $\pm$0.32	88.63 $\pm$0.27	93.83 $\pm$0.16	88.41 $\pm$0.45	93.30	87.01
MD	82.31 $\pm$0.08	83.78 $\pm$1.22	89.38 $\pm$0.29	83.91 $\pm$0.23	78.22 $\pm$0.77	83.52	99.14 $\pm$0.03	95.46 $\pm$0.32	96.09 $\pm$0.25	90.79 $\pm$0.24	95.44 $\pm$0.14	90.60 $\pm$0.40	94.59	89.56
GEN	79.50 $\pm$0.11	80.11 $\pm$0.63	81.35 $\pm$0.14	84.14 $\pm$0.25	81.95 $\pm$0.74	81.41	91.89 $\pm$0.12	82.28 $\pm$0.54	90.76 $\pm$0.19	83.99 $\pm$0.33	89.20 $\pm$0.17	78.73 $\pm$0.18	86.14	83.99
REACT	79.40 $\pm$0.12	79.99 $\pm$0.62	81.19 $\pm$0.15	84.06 $\pm$0.26	81.86 $\pm$0.74	81.30	91.82 $\pm$0.13	82.03 $\pm$0.55	90.72 $\pm$0.21	83.85 $\pm$0.36	89.11 $\pm$0.17	78.49 $\pm$0.20	86.00	83.87
PCA	86.40 $\pm$0.09	83.76 $\pm$2.33	86.42 $\pm$0.30	81.08 $\pm$0.27	80.31 $\pm$0.43	83.59	98.00 $\pm$0.07	90.69 $\pm$0.39	92.27 $\pm$0.43	86.13 $\pm$0.33	91.85 $\pm$0.19	86.88 $\pm$0.43	90.97	87.62
ASH	72.79 $\pm$0.08	73.99 $\pm$1.06	83.13 $\pm$0.31	72.84 $\pm$0.23	73.37 $\pm$0.27	75.22	90.79 $\pm$0.10	91.58 $\pm$0.58	86.15 $\pm$0.27	82.58 $\pm$0.36	90.74 $\pm$0.29	80.28 $\pm$0.45	87.02	81.66
SHE	80.85 $\pm$0.09	83.11 $\pm$1.99	87.91 $\pm$0.13	86.71 $\pm$0.13	68.13 $\pm$0.56	81.34	89.97 $\pm$0.31	95.63 $\pm$0.18	83.66 $\pm$0.37	84.05 $\pm$0.21	89.90 $\pm$0.12	91.49 $\pm$0.18	89.12	85.58
VRA	85.69 $\pm$0.13	87.18 $\pm$1.40	85.08 $\pm$0.19	85.52 $\pm$0.16	77.11 $\pm$0.42	84.11	82.12 $\pm$0.68	83.56 $\pm$0.53	79.35 $\pm$0.61	80.88 $\pm$0.42	83.38 $\pm$0.48	86.60 $\pm$0.11	82.65	83.31
NECO	84.51 $\pm$0.01	86.48 $\pm$1.71	91.62 $\pm$0.05	87.06 $\pm$0.04	87.02 $\pm$0.23	87.34	98.12 $\pm$0.01	95.72 $\pm$0.09	94.60 $\pm$0.14	91.72 $\pm$0.04	95.00 $\pm$0.04	93.27 $\pm$0.01	94.74	91.38
FDBD	83.30 $\pm$0.04	85.29 $\pm$1.72	90.04 $\pm$0.07	85.60 $\pm$0.10	78.19 $\pm$0.16	84.48	95.25 $\pm$0.08	95.01 $\pm$0.09	91.51 $\pm$0.09	88.66 $\pm$0.08	93.74 $\pm$0.08	90.29 $\pm$0.05	92.41	88.81
CASOD	85.42 $\pm$0.09	88.30 $\pm$0.04	93.26 $\pm$0.16	89.66 $\pm$0.15	91.34 $\pm$0.61	89.60	99.02 $\pm$0.07	99.89 $\pm$0.01	98.46 $\pm$0.13	96.40 $\pm$0.06	95.55 $\pm$0.13	94.56 $\pm$0.25	97.31	93.81
	ID data: ImageNet-200													
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	82.16 $\pm$0.18	81.27 $\pm$0.27	77.53 $\pm$0.22	64.13 $\pm$0.41	65.21 $\pm$0.97	74.06	87.56 $\pm$0.24	78.50 $\pm$0.43	77.58 $\pm$0.48	79.26 $\pm$0.43	79.87 $\pm$0.18	68.97 $\pm$0.41	78.62	76.55
ENERGY	69.56 $\pm$0.32	69.60 $\pm$0.40	59.05 $\pm$0.34	52.80 $\pm$0.69	51.59 $\pm$1.92	60.52	79.80 $\pm$0.41	46.01 $\pm$1.34	69.31 $\pm$0.43	66.94 $\pm$0.51	63.78 $\pm$0.45	49.83 $\pm$0.52	62.61	61.66
KNN	81.99 $\pm$0.06	82.82 $\pm$0.04	83.72 $\pm$0.18	72.69 $\pm$0.13	72.17 $\pm$0.47	78.68	95.67 $\pm$0.04	86.53 $\pm$0.22	92.56 $\pm$0.32	89.17 $\pm$0.50	89.75 $\pm$0.22	76.90 $\pm$0.05	88.43	84.00
VIM	79.44 $\pm$0.01	82.44 $\pm$0.03	86.04 $\pm$0.12	77.20 $\pm$0.08	73.27 $\pm$0.54	79.68	97.87 $\pm$0.03	91.37 $\pm$0.07	97.40 $\pm$0.13	92.36 $\pm$0.27	92.88 $\pm$0.14	82.20 $\pm$0.02	92.35	86.59
MD	79.46 $\pm$0.06	81.66 $\pm$0.02	78.14 $\pm$0.30	78.77 $\pm$0.15	76.77 $\pm$0.44	78.96	92.41 $\pm$0.07	78.44 $\pm$0.23	90.48 $\pm$0.23	82.99 $\pm$0.47	82.72 $\pm$0.17	79.42 $\pm$0.40	84.41	81.93
GEN	44.64 $\pm$0.27	44.42 $\pm$0.36	55.22 $\pm$0.57	54.40 $\pm$0.71	42.02 $\pm$1.66	48.14	34.85 $\pm$0.79	70.87 $\pm$1.14	39.97 $\pm$0.94	47.54 $\pm$0.99	48.67 $\pm$0.53	64.19 $\pm$0.72	51.01	49.71
REACT	69.65 $\pm$0.31	69.68 $\pm$0.39	59.14 $\pm$0.35	52.90 $\pm$0.69	51.59 $\pm$1.92	60.59	79.86 $\pm$0.42	46.13 $\pm$1.35	69.40 $\pm$0.43	67.01 $\pm$0.51	63.87 $\pm$0.46	49.94 $\pm$0.53	62.70	61.74
PCA	76.99 $\pm$0.13	76.31 $\pm$0.07	81.89 $\pm$0.27	70.51 $\pm$0.16	72.69 $\pm$0.27	75.68	90.27 $\pm$0.10	79.96 $\pm$0.30	90.48 $\pm$0.13	86.39 $\pm$0.20	86.00 $\pm$0.31	73.70 $\pm$0.25	84.47	80.47
ASH	51.79 $\pm$0.31	52.05 $\pm$0.36	46.64 $\pm$0.99	39.35 $\pm$0.52	72.64 $\pm$0.11	52.49	49.76 $\pm$0.91	55.87 $\pm$1.82	35.21 $\pm$0.80	33.89 $\pm$1.41	55.18 $\pm$0.80	56.05 $\pm$1.36	47.66	49.86
SHE	72.98 $\pm$0.15	71.61 $\pm$0.10	77.11 $\pm$0.32	61.83 $\pm$0.28	59.43 $\pm$0.81	68.59	91.57 $\pm$0.23	76.31 $\pm$0.48	89.19 $\pm$0.56	86.95 $\pm$0.78	81.26 $\pm$0.42	63.35 $\pm$0.21	81.44	75.60
VRA	39.41 $\pm$0.39	39.76 $\pm$0.35	38.21 $\pm$0.84	54.42 $\pm$0.29	39.03 $\pm$0.85	42.16	11.36 $\pm$0.22	31.34 $\pm$0.78	32.09 $\pm$2.96	30.88 $\pm$2.74	26.62 $\pm$0.56	47.06 $\pm$0.49	29.89	35.47
NECO	81.13 $\pm$0.06	83.53 $\pm$0.08	82.63 $\pm$0.16	73.00 $\pm$0.12	72.34 $\pm$0.44	78.53	94.30 $\pm$0.11	82.13 $\pm$0.45	90.88 $\pm$0.14	89.24 $\pm$0.15	87.93 $\pm$0.21	77.63 $\pm$0.16	87.02	83.16
FDBD	84.39 $\pm$0.08	84.40 $\pm$0.11	82.10 $\pm$0.11	70.70 $\pm$0.26	73.46 $\pm$0.52	79.01	91.97 $\pm$0.15	86.68 $\pm$0.25	87.08 $\pm$0.15	86.84 $\pm$0.06	87.27 $\pm$0.08	74.43 $\pm$0.24	85.71	82.67
CASOD	85.03 $\pm$0.95	86.93 $\pm$0.67	89.96 $\pm$0.27	82.54 $\pm$0.68	87.14 $\pm$2.96	86.32	98.48 $\pm$0.49	99.42 $\pm$0.11	98.26 $\pm$0.20	97.81 $\pm$0.13	94.14 $\pm$0.30	86.60 $\pm$0.59	95.79	91.48

Table 4. **Results - FPR95**: FPR95 OOD detection results. Avg.N is mean performance for near-OOD. Avg.F is mean performance for far-OOD.

	ID data: ImageNet-100-Random													
	ImageNet-30	ImageNet-900	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	61.69 ±0.28	62.15 ±0.04	55.18 ±0.20	62.83 ±0.22	85.92 ±0.80	65.55	38.71 ±0.49	38.33 ±0.58	59.34 ±0.71	66.01 ±0.57	44.36 ±0.27	49.97 ±0.06	49.45	56.77
ENERGY	70.49 ±0.20	71.47 ±0.70	71.34 ±0.23	59.26 ±0.07	79.08 ±1.44	70.33	40.46 ±0.29	79.16 ±0.40	56.35 ±0.14	81.46 ±0.20	55.20 ±0.45	76.64 ±0.16	64.88	67.36
KNN	69.51 ±0.54	75.71 ±0.66	80.39 ±0.61	90.50 ±0.28	95.73 ±0.37	82.37	19.96 ±0.85	95.63 ±0.43	74.78 ±1.62	91.27 ±0.51	63.27 ±1.05	92.52 ±1.07	72.91	77.21
VIM	67.98 ±0.44	68.90 ±0.47	63.60 ±1.03	71.96 ±0.45	90.82 ±0.51	72.65	4.65 ±0.19	39.01 ±2.39	24.26 ±2.17	61.99 ±1.15	35.01 ±1.20	55.02 ±2.08	36.66	53.02
MD	58.31 ±0.17	60.92 ±0.47	55.62 ±1.39	67.25 ±0.65	85.36 ±0.93	65.49	4.00 ±0.21	31.47 ±2.59	22.42 ±1.86	54.61 ±1.44	26.45 ±1.11	49.46 ±2.59	31.40	46.90
GEN	70.42 ±0.04	71.20 ±0.01	71.18 ±0.06	59.06 ±0.21	78.95 ±1.23	70.16	40.21 ±0.40	78.87 ±0.49	56.03 ±0.39	81.12 ±0.39	54.91 ±0.25	76.28 ±0.33	64.57	67.11
REACT	70.47 ±0.20	71.47 ±0.19	71.35 ±0.23	59.26 ±0.07	79.08 ±1.43	70.32	40.46 ±0.29	79.16 ±0.40	56.35 ±0.13	81.46 ±0.20	55.20 ±0.45	76.64 ±0.16	64.88	67.35
PCA	47.33 ±0.07	60.37 ±0.14	65.31 ±0.88	72.27 ±0.59	87.09 ±0.28	66.48	10.89 ±0.20	60.11 ±1.16	47.47 ±2.36	72.79 ±0.73	47.79 ±0.99	64.32 ±1.51	50.56	57.80
ASH	89.04 ±0.46	88.97 ±0.62	78.15 ±1.50	88.61 ±0.63	95.49 ±0.26	88.05	56.57 ±1.47	60.13 ±4.26	69.40 ±1.84	84.69 ±1.42	58.45 ±2.25	82.43 ±2.05	68.61	77.45
SHE	62.35 ±0.37	61.76 ±0.51	60.05 ±0.72	52.68 ±0.75	91.94 ±0.25	65.76	52.19 ±1.12	27.12 ±1.36	67.92 ±0.96	65.25 ±0.74	53.23 ±0.68	46.93 ±1.44	52.11	58.31
VRA	55.40 ±0.23	52.37 ±0.16	73.51 ±0.13	53.12 ±0.03	87.55 ±0.66	64.39	76.00 ±0.74	81.92 ±0.09	87.78 ±0.36	84.42 ±0.09	80.52 ±0.15	59.17 ±0.54	78.30	71.98
NECO	49.36 ±0.21	50.17 ±0.29	42.21 ±0.13	52.27 ±0.25	64.29 ±0.89	51.66	7.88 ±0.11	27.48 ±0.66	29.01 ±0.72	42.72 ±0.19	27.83 ±0.03	31.58 ±0.12	27.75	38.62
FDBD	61.93 ±0.23	63.53 ±0.21	54.28 ±0.22	64.81 ±0.35	88.49 ±0.53	66.61	27.27 ±0.76	33.32 ±0.49	46.79 ±0.37	60.10 ±0.58	40.01 ±0.49	46.99 ±0.40	42.42	53.41
CASOD	52.93 ±1.01	53.69 ±0.27	35.92 ±0.57	48.49 ±0.70	53.50 ±4.86	48.91	4.85 ±0.13	0.22 ±0.04	7.81 ±0.50	17.58 ±0.62	26.53 ±0.64	23.94 ±1.05	13.49	29.59
	ID data: ImageNet-200													
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	74.32 ±0.80	72.89 ±0.87	79.10 ±0.73	88.54 ±0.48	98.24 ±0.13	82.62	54.04 ±1.72	91.46 ±1.01	85.80 ±0.62	84.81 ±0.87	74.92 ±1.05	92.19 ±0.46	80.54	81.48
ENERGY	80.79 ±0.32	79.87 ±0.51	87.70 ±0.09	70.20 ±0.35	97.32 ±0.48	83.18	64.63 ±0.36	98.72 ±0.20	84.59 ±0.46	86.77 ±0.14	83.67 ±0.04	95.78 ±0.17	85.69	84.55
KNN	72.74 ±0.50	70.46 ±0.66	72.90 ±1.10	87.25 ±0.42	96.36 ±0.12	79.94	13.92 ±0.35	83.04 ±1.23	48.30 ±3.34	77.74 ±2.91	51.45 ±1.24	85.11 ±0.81	59.93	69.02
VIM	66.36 ±0.39	61.31 ±0.29	61.95 ±0.33	74.46 ±0.25	92.98 ±0.36	71.41	5.87 ±0.08	54.31 ±0.24	12.44 ±0.72	51.55 ±1.45	41.18 ±0.60	70.07 ±0.28	39.24	53.86
MD	59.36 ±0.13	54.19 ±0.10	79.18 ±0.22	69.41 ±0.16	91.94 ±0.11	70.82	35.87 ±0.16	88.04 ±0.21	62.24 ±1.42	74.95 ±0.66	74.52 ±0.05	77.24 ±0.37	68.81	69.72
GEN	95.86 ±0.26	95.56 ±0.12	90.47 ±0.46	92.08 ±0.27	97.10 ±0.53	94.21	96.22 ±0.75	85.45 ±1.37	97.67 ±0.26	97.49 ±0.17	94.04 ±0.32	84.06 ±1.08	92.49	93.27
REACT	80.72 ±0.43	79.79 ±0.58	87.61 ±0.13	90.16 ±0.40	97.33 ±0.47	87.12	64.46 ±0.14	98.71 ±0.19	84.51 ±0.36	86.71 ±0.17	83.59 ±0.04	95.74 ±0.12	85.62	86.30
PCA	63.06 ±0.27	66.86 ±0.17	70.89 ±0.27	80.43 ±0.02	90.52 ±0.19	74.35	39.61 ±0.24	79.07 ±0.64	58.09 ±0.73	65.05 ±0.39	64.91 ±0.58	76.56 ±0.09	63.88	68.64
ASH	90.47 ±0.30	90.50 ±0.22	94.04 ±0.19	95.71 ±0.16	98.82 ±0.09	93.91	95.11 ±0.24	91.02 ±0.69	95.14 ±0.35	94.49 ±0.40	93.00 ±0.24	90.07 ±0.53	93.14	93.49
SHE	82.27 ±0.36	81.86 ±0.39	83.53 ±0.84	90.65 ±0.27	97.82 ±0.23	87.23	33.21 ±1.56	88.17 ±0.96	59.64 ±3.05	78.64 ±2.65	66.03 ±0.76	92.78 ±0.33	69.74	77.69
VRA	98.54 ±0.08	97.08 ±0.02	98.05 ±0.12	92.39 ±0.11	99.96 ±0.01	97.20	99.91 ±0.03	99.82 ±0.01	99.43 ±0.11	99.71 ±0.12	99.48 ±0.04	97.00 ±0.18	99.22	98.31
NECO	66.16 ±0.27	61.55 ±0.40	78.47 ±0.49	85.21 ±0.27	94.50 ±0.41	77.18	23.45 ±0.93	95.96 ±0.40	65.38 ±0.39	75.96 ±0.54	66.88 ±0.76	87.80 ±0.50	69.24	72.85
FDBD	67.90 ±0.23	65.91 ±0.38	71.60 ±0.36	82.97 ±0.28	97.02 ±0.31	77.08	32.95 ±1.00	75.68 ±0.93	72.15 ±0.56	75.38 ±0.24	61.28 ±0.23	85.76 ±0.28	67.20	71.69
CASOD	59.39 ±1.17	53.33 ±0.89	50.29 ±1.86	63.95 ±2.37	77.17 ±4.48	60.83	5.44 ±1.22	1.61 ±0.47	7.24 ±0.84	10.55 ±0.67	37.13 ±2.65	50.99 ±0.22	18.83	37.92

Table 5. **Results - AUPR-IN**: AUPR-IN OOD detection results. Avg.N is mean performance for near-OOD. Avg.F is mean performance for far-OOD.

	ID data: ImageNet-100-Random													
	ImageNet-30	ImageNet-900	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	89.65 ±0.70	38.44 ±3.33	72.37 ±3.19	21.67 ±1.93	69.30 ±0.80	58.29	86.03 ±0.81	69.99 ±6.11	72.92 ±1.80	69.83 ±0.25	65.39 ±3.99	62.41 ±4.00	71.09	65.27
ENERGY	92.45 ±0.06	58.74 ±0.21	79.16 ±0.26	66.61 ±1.43	65.82 ±1.47	72.56	92.01 ±0.13	77.49 ±0.71	88.04 ±0.27	77.69 ±0.39	68.68 ±0.11	39.01 ±0.88	73.82	73.25
KNN	91.47 ±1.20	49.89 ±3.19	85.69 ±1.53	40.47 ±0.11	53.13 ±1.10	64.13	94.05 ±2.16	88.43 ±2.32	89.61 ±3.51	84.22 ±5.54	83.45 ±0.32	78.47 ±2.76	86.37	76.26
VIM	91.70 ±1.46	47.69 ±5.17	88.24 ±0.95	43.63 ±0.43	55.78 ±1.15	65.41	97.94 ±0.68	93.93 ±0.07	96.34 ±1.70	88.39 ±3.66	87.72 ±0.26	84.12 ±0.10	91.41	79.59
MD	93.34 ±0.49	40.06 ±2.75	85.56 ±2.71	44.71 ±1.12	62.47 ±1.09	65.23	91.55 ±0.15	84.23 ±2.38	89.67 ±1.45	76.79 ±2.33	70.22 ±3.06	85.56 ±2.53	83.00	74.92
GEN	85.71 ±0.67	18.32 ±3.76	47.35 ±2.32	10.78 ±0.29	66.04 ±1.47	45.64	77.27 ±1.67	31.52 ±3.11	58.37 ±7.15	58.88 ±10.17	37.55 ±2.63	29.80 ±2.11	48.90	47.42
REACT	85.60 ±0.73	17.97 ±3.81	47.00 ±2.39	10.65 ±0.32	65.82 ±1.47	45.41	76.97 ±1.53	31.05 ±3.14	57.97 ±7.01	58.51 ±10.12	37.00 ±2.78	29.45 ±2.08	48.49	47.09
PCA	91.78 ±2.85	42.18 ±8.72	80.92 ±2.61	31.39 ±0.49	60.93 ±0.67	61.44	93.33 ±2.11	82.10 ±0.81	92.90 ±4.37	89.20 ±6.47	80.40 ±2.75	69.82 ±0.32	84.63	74.09
ASH	79.80 ±2.04	10.52 ±1.11	42.41 ±1.88	7.76 ±0.40	63.22 ±0.31	40.74	60.74 ±8.52	31.34 ±2.51	31.64 ±4.07	33.36 ±6.34	27.35 ±2.16	29.95 ±2.25	35.73	38.01
SHE	87.65 ±2.45	35.36 ±7.97	76.22 ±1.80	26.82 ±2.51	45.56 ±0.96	54.32	88.64 ±1.69	77.17 ±3.27	79.03 ±5.51	74.66 ±6.77	70.38 ±2.55	61.63 ±5.37	75.25	65.74
VRA	71.17 ±1.63	8.40 ±0.78	39.89 ±0.36	9.72 ±1.73	61.95 ±0.53	38.23	36.85 ±6.78	35.33 ±7.36	24.77 ±1.32	26.13 ±1.55	16.11 ±1.40	36.25 ±2.37	29.24	33.32
NECO	92.35 ±0.23	52.95 ±0.90	87.40 ±1.38	43.33 ±2.37	73.09 ±0.38	69.82	94.35 ±2.15	90.56 ±1.08	91.20 ±1.66	86.52 ±2.51	85.07 ±2.73	81.42 ±2.05	88.19	79.84
FDBD	90.73 ±0.27	44.13 ±0.86	78.94 ±1.28	29.25 ±0.59	65.38 ±0.22	61.69	90.09 ±0.63	79.26 ±2.82	82.41 ±1.66	78.30 ±1.79	75.31 ±2.38	69.36 ±1.22	79.12	71.20
CASOD	92.49 ±0.13	52.10 ±0.18	92.10 ±0.20	51.91 ±0.41	81.85 ±0.44	74.09	98.84 ±0.08	99.81 ±0.02	97.04 ±0.18	92.35 ±0.14	89.70 ±0.16	88.56 ±0.40	94.38	85.16
	ID data: ImageNet-200													
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	89.90 ±0.08	55.97 ±0.37	84.96 ±0.09	25.48 ±0.16	59.83 ±1.04	63.23	88.60 ±0.16	82.15 ±0.25	80.05 ±0.44	82.52 ±0.33	72.92 ±0.14	69.35 ±0.53	79.27	71.98
ENERGY	81.35 ±0.18	36.54 ±0.11	69.36 ±0.43	18.32 ±0.30	34.22 ±1.77	47.96	82.97 ±0.50	50.45 ±1.14	69.40 ±0.43	68.80 ±0.45	48.49 ±0.76	51.43 ±0.62	61.92	55.58
KNN	90.29 ±0.03	62.68 ±0.03	89.55 ±0.10	36.25 ±0.15	70.27 ±0.59	69.81	95.55 ±0.05	89.43 ±0.20	93.45 ±0.18	92.04 ±0.31	85.85 ±0.28	76.84 ±0.02	88.86	80.20
VIM	87.29 ±0.04	56.08 ±0.07	90.89 ±0.09	41.32 ±0.24	68.04 ±0.69	68.73	98.10 ±0.05	93.01 ±0.07	97.47 ±0.11	93.80 ±0.20	90.23 ±0.16	82.08 ±0.06	92.45	81.67
MD	85.48 ±0.04	48.54 ±0.05	85.74 ±0.25	44.12 ±0.32	73.18 ±0.53	67.41	95.26 ±0.06	82.21 ±0.30	92.45 ±0.19	84.92 ±0.47	77.46 ±0.31	80.88 ±0.42	85.53	77.30
GEN	60.20 ±0.17	15.85 ±0.12	63.47 ±0.48	16.65 ±0.33	26.18 ±0.89	36.47	51.55 ±0.36	69.97 ±1.14	40.61 ±0.58	45.56 ±0.55	32.42 ±0.31	60.08 ±0.62	50.03	43.87
REACT	80.18 ±0.20	34.87 ±0.41	67.67 ±0.45	17.18 ±0.28	34.22 ±1.77	46.83	81.86 ±0.53	48.53 ±1.16	67.83 ±0.45	67.28 ±0.47	46.60 ±0.78	49.47 ±0.64	60.26	54.16
PCA	83.68 ±0.11	42.00 ±0.11	88.39 ±0.20	34.88 ±0.27	66.11 ±0.49	63.01	93.58 ±0.10	82.29 ±0.26	92.01 ±0.10	87.75 ±0.21	80.93 ±0.47	72.85 ±0.30	84.90	74.95
ASH	64.60 ±0.23	18.45 ±0.17	57.86 ±0.66	12.02 ±0.11	72.42 ±0.07	45.07	62.24 ±0.62	52.08 ±1.60	37.60 ±0.29	37.13 ±0.57	36.36 ±0.63	51.67 ±1.16	46.18	45.68
SHE	84.03 ±0.09	39.66 ±0.08	85.05 ±0.22	24.90 ±0.27	47.27 ±1.03	56.18	92.47 ±0.13	77.44 ±0.43	89.43 ±0.33	89.45 ±0.50	71.57 ±0.60	61.91 ±0.09	80.38	69.38
VRA	57.29 ±0.17	14.73 ±0.09	53.05 ±0.56	17.09 ±0.16	31.33 ±0.57	34.70	42.44 ±0.06	37.82 ±0.48	37.54 ±1.44	37.77 ±1.50	23.36 ±0.15	46.50 ±0.27	37.57	36.26
NECO	88.61 ±0.04	59.81 ±0.11	88.91 ±0.03	36.91 ±0.07	67.73 ±0.37	68.39	94.54 ±0.04	87.45 ±0.33	92.51 ±0.11	92.00 ±0.08	85.25 ±0.21	78.16 ±0.23	88.32	79.26
FDBD	91.37 ±0.03	61.71 ±0.20	87.94 ±0.01	30.16 ±0.17	69.00 ±0.48	68.04	92.51 ±0.10	89.09 ±0.19	88.57 ±0.18	89.29 ±0.07	82.57 ±0.11	73.63 ±0.31	85.94	77.80
CASOD	91.33 ±0.72	64.74 ±2.06	93.30 ±0.17	46.75 ±0.90	84.28 ±2.56	76.08	98.84 ±0.49	99.48 ±0.09	98.14 ±0.24	97.82 ±0.16	92.19 ±0.32	83.87 ±1.55	95.06	86.43

with MSP and KNN is also better than the baseline MSP and KNN except for in the ImageNet-1k ID setting where CASOD-KNN is slightly worse. This isolated discrepancy is suspected to be due to the significantly more complicated feature space. The class-specific comparisons in MD may be critical to discriminating between similar data in this setting. Overall, this further validates the importance of principled fusion of foreground-background views in improving OOD detection performance in the proposed CASOD.

Table 6. **Results - AUPR-OUT**: AUPR-OUT OOD detection results. Avg.N is mean performance for near-OOD. Avg.F is mean performance for far-OOD.

	ID data: ImageNet-100-Random													
	ImageNet-30	ImageNet-900	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	55.51 ±2.13	96.89 ±0.04	77.72 ±0.18	95.59 ±0.04	92.86 ±0.24	83.71	88.55 ±2.85	82.94 ±2.14	86.76 ±3.97	85.31 ±3.74	93.73 ±0.42	80.93 ±0.73	86.37	85.16
ENERGY	52.76 ±0.39	81.95 ±0.05	80.35 ±0.19	97.68 ±0.05	93.76 ±0.23	81.30	90.04 ±0.21	85.68 ±0.41	92.19 ±0.16	86.04 ±0.30	95.16 ±0.09	84.83 ±0.21	88.99	85.49
KNN	66.13 ±7.49	97.77 ±0.94	88.19 ±3.66	97.36 ±1.31	86.60 ±0.39	87.21	96.08 ±1.53	94.49 ±3.42	96.38 ±1.95	94.15 ±4.60	97.55 ±0.45	91.80 ±3.65	95.07	91.50
VIM	70.15 ±6.83	98.08 ±0.60	91.63 ±1.92	98.37 ±0.58	90.55 ±0.31	89.76	98.59 ±0.44	97.68 ±0.21	98.86 ±0.56	96.34 ±2.17	98.46 ±0.19	95.70 ±0.48	97.60	94.04
MD	73.72 ±1.55	98.30 ±0.15	88.26 ±2.23	98.69 ±0.18	92.57 ±0.30	90.31	93.57 ±0.13	92.65 ±1.22	94.59 ±1.14	90.34 ±1.48	95.54 ±0.66	95.33 ±0.93	93.67	92.14
GEN	43.74 ±1.99	94.66 ±0.17	63.53 ±0.61	93.28 ±0.49	93.79 ±0.24	77.80	83.37 ±4.88	64.06 ±3.62	78.41 ±8.73	77.08 ±10.16	88.27 ±1.31	65.19 ±3.08	76.06	76.85
REACT	43.50 ±1.95	94.61 ±0.19	63.29 ±0.51	93.22 ±0.48	93.76 ±0.23	77.68	83.24 ±4.82	63.79 ±3.49	78.26 ±8.67	76.95 ±10.12	88.15 ±1.26	64.96 ±2.99	75.89	76.70
PCA	69.22 ±7.08	98.14 ±0.62	87.68 ±2.05	98.05 ±0.54	92.26 ±0.17	89.07	94.33 ±2.08	92.11 ±1.50	96.05 ±3.01	94.68 ±3.67	96.93 ±1.11	90.28 ±0.92	94.06	91.79
ASH	31.76 ±1.12	92.62 ±0.31	56.03 ±0.92	91.09 ±0.75	89.12 ±0.14	72.12	70.14 ±9.21	68.87 ±0.53	65.24 ±5.43	64.53 ±5.33	86.18 ±1.93	67.17 ±0.79	70.36	71.16
SHE	47.20 ±11.66	95.65 ±1.55	78.40 ±5.11	94.78 ±1.79	89.01 ±0.20	81.01	90.58 ±1.62	88.20 ±5.16	91.15 ±4.61	89.58 ±6.84	94.68 ±1.25	80.73 ±6.08	89.15	85.45
VRA	19.85 ±0.56	89.93 ±0.71	50.17 ±0.94	92.62 ±1.30	91.89 ±0.17	68.89	41.04 ±2.97	61.69 ±3.29	57.30 ±1.86	58.05 ±0.96	69.81 ±0.61	67.72 ±1.03	59.27	63.64
NECO	73.63 ±2.55	98.51 ±0.21	90.66 ±0.13	98.45 ±0.21	96.09 ±0.08	91.47	95.89 ±1.32	96.00 ±0.32	96.66 ±0.99	95.24 ±1.41	97.96 ±0.13	94.52 ±0.48	96.05	93.96
FDBD	60.17 ±2.73	97.44 ±0.15	82.62 ±0.69	96.66 ±0.33	92.00 ±0.12	85.78	91.75 ±0.92	88.55 ±2.67	91.73 ±2.47	90.48 ±3.15	95.45 ±0.38	85.56 ±2.16	90.59	88.40
CASOD	72.04 ±0.35	98.45 ±0.01	94.19 ±0.11	98.76 ±0.02	97.24 ±0.31	92.14	99.21 ±0.06	99.94 ±0.00	99.24 ±0.08	98.24 ±0.02	98.57 ±0.05	97.37 ±0.13	98.76	95.75
	ID data: ImageNet-200													
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	67.21 ±0.38	93.83 ±0.12	64.82 ±0.43	89.66 ±0.18	73.77 ±0.64	77.86	79.93 ±0.52	71.23 ±0.59	73.21 ±0.50	74.26 ±0.56	86.07 ±0.18	64.97 ±0.30	74.95	76.27
ENERGY	53.45 ±0.46	89.64 ±0.03	47.49 ±0.17	85.44 ±0.29	69.15 ±1.15	69.03	70.66 ±0.20	45.03 ±0.83	67.45 ±0.47	64.35 ±0.30	76.04 ±0.22	49.30 ±0.36	62.14	65.27
KNN	67.28 ±0.24	94.41 ±0.05	72.30 ±0.38	91.83 ±0.07	78.08 ±0.30	80.78	95.21 ±0.08	79.68 ±0.36	91.04 ±0.53	83.84 ±0.92	93.45 ±0.15	73.10 ±0.22	86.05	83.66
VIM	68.88 ±0.13	94.91 ±0.03	77.45 ±0.23	93.98 ±0.03	80.42 ±0.35	83.13	97.46 ±0.03	88.19 ±0.14	97.14 ±0.17	89.80 ±0.39	95.30 ±0.09	80.55 ±0.02	91.41	87.64
MD	72.38 ±0.06	95.14 ±0.01	65.39 ±0.32	94.76 ±0.04	82.49 ±0.30	82.03	88.14 ±0.03	72.21 ±0.10	86.56 ±0.27	80.16 ±0.46	87.34 ±0.09	77.97 ±0.34	82.06	82.05
GEN	32.57 ±0.18	79.45 ±0.17	45.54 ±0.68	86.54 ±0.24	65.11 ±1.18	61.84	30.92 ±0.73	69.41 ±1.37	45.31 ±0.54	49.28 ±0.77	66.09 ±0.47	66.65 ±0.94	54.61	57.90
REACT	55.52 ±0.47	90.41 ±0.17	49.61 ±0.18	86.48 ±0.27	69.15 ±1.15	70.23	72.29 ±0.20	47.12 ±0.84	69.30 ±0.46	66.23 ±0.30	77.55 ±0.21	51.46 ±0.36	63.99	66.83
PCA	69.88 ±0.11	93.23 ±0.03	70.47 ±0.34	92.07 ±0.05	81.56 ±0.14	81.44	85.62 ±0.07	77.18 ±0.34	88.18 ±0.17	84.89 ±0.20	90.17 ±0.19	74.74 ±0.19	83.46	82.54
ASH	39.89 ±0.30	83.70 ±0.13	38.50 ±0.56	81.02 ±0.22	76.50 ±0.13	63.92	38.19 ±0.52	58.46 ±1.52	45.28 ±0.62	45.53 ±0.93	69.70 ±0.52	59.26 ±0.99	52.74	57.82
SHE	56.41 ±0.39	90.57 ±0.11	62.88 ±0.66	88.65 ±0.13	71.89 ±0.49	74.08	88.69 ±0.56	71.67 ±0.72	87.54 ±0.96	81.75 ±1.36	88.42 ±0.31	61.50 ±0.25	79.93	77.27
VRA	28.74 ±0.21	76.76 ±0.18	31.81 ±0.40	86.29 ±0.09	60.53 ±0.36	56.82	23.22 ±0.03	39.96 ±0.30	40.45 ±1.27	39.72 ±1.09	51.73 ±0.26	49.00 ±0.36	40.68	48.02
NECO	70.46 ±0.18	95.19 ±0.04	70.23 ±0.38	92.25 ±0.07	79.42 ±0.36	81.51	92.32 ±0.30	71.65 ±0.57	86.49 ±0.32	83.79 ±0.32	91.27 ±0.19	73.15 ±0.33	83.11	82.38
FDBD	71.33 ±0.26	94.98 ±0.06	71.39 ±0.32	91.97 ±0.11	78.90 ±0.41	81.71	87.78 ±0.37	81.55 ±0.37	83.44 ±0.22	82.02 ±0.17	91.10 ±0.11	71.27 ±0.24	82.86	82.34
CASOD	75.02 ±0.80	96.24 ±0.14	83.54 ±0.50	95.73 ±0.20	89.96 ±2.88	88.10	98.01 ±0.50	99.28 ±0.20	98.23 ±0.15	97.50 ±0.05	95.98 ±0.20	87.41 ±0.27	96.07	92.45

Table 7. **Results - DE**: DE OOD detection results. Avg.N is mean performance for near-OOD. Avg.F is mean performance for far-OOD.

	ID data: ImageNet-100-Random													
	ImageNet-30	ImageNet-900	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	78.52 ±0.90	31.76 ±1.94	53.05 ±1.80	21.38 ±2.10	27.45 ±0.67	42.43	67.38 ±6.38	38.83 ±5.73	43.35 ±7.30	41.49 ±6.85	41.82 ±4.49	37.85 ±3.05	45.12	43.90
ENERGY	79.89 ±0.05	49.84 ±0.18	59.15 ±0.12	45.77 ±0.07	33.16 ±1.20	53.56	76.21 ±0.16	45.56 ±0.27	60.77 ±0.09	44.03 ±0.14	55.89 ±0.34	47.24 ±0.11	54.95	54.32
KNN	81.96 ±2.22	42.09 ±11.77	65.03 ±6.32	36.77 ±14.67	19.27 ±0.31	49.02	84.87 ±4.79	64.56 ±14.75	74.41 ±10.76	65.92 ±19.88	66.41 ±4.71	57.63 ±11.41	68.97	59.90
VIM	83.34 ±2.77	49.36 ±9.48	72.57 ±5.55	52.50 ±13.25	23.36 ±0.42	56.23	93.33 ±1.54	82.47 ±2.41	90.99 ±5.45	76.05 ±14.31	77.99 ±3.24	73.02 ±3.50	82.31	70.45
MD	84.65 ±0.28	57.34 ±1.22	65.81 ±2.63	60.89 ±3.62	27.93 ±0.78	59.32	78.33 ±0.38	57.84 ±2.74	65.09 ±5.04	51.47 ±2.88	55.32 ±2.83	69.14 ±2.98	62.86	61.25
GEN	76.21 ±0.51	22.90 ±1.08	47.40 ±0.06	16.37 ±1.66	33.28 ±1.02	39.23	61.16 ±6.12	31.76 ±1.46	36.90 ±5.11	36.00 ±4.89	32.40 ±3.12	32.35 ±1.61	38.43	38.79
REACT	76.14 ±0.54	22.69 ±1.15	47.30 ±0.12	16.23 ±1.67	33.16 ±1.19	39.11	60.98 ±6.10	31.70 ±1.41	36.78 ±5.03	35.91 ±4.82	32.22 ±3.10	32.30 ±1.62	38.32	38.68
PCA	83.08 ±2.28	52.82 ±8.69	65.29 ±5.31	49.27 ±9.57	26.48 ±0.24	55.39	79.62 ±5.52	55.73 ±5.61	74.33 ±18.58	66.73 ±15.75	63.09 ±10.26	55.37 ±3.69	65.81	61.07
ASH	74.20 ±0.18	17.49 ±0.22	44.49 ±0.30	13.12 ±1.38	19.47 ±0.22	33.75	50.74 ±4.09	34.29 ±0.88	31.90 ±1.22	31.11 ±0.53	29.47 ±2.49	33.36 ±0.52	35.15	34.51
SHE	77.14 ±2.82	25.53 ±7.01	53.61 ±5.04	19.44 ±5.14	22.43 ±0.21	39.63	70.81 ±2.16	48.31 ±11.27	55.76 ±12.09	53.97 ±15.68	48.15 ±6.48	39.52 ±5.91	52.75	46.79
VRA	71.92 ±0.05	12.93 ±0.55	43.03 ±0.20	14.69 ±3.75	26.10 ±0.55	33.73	42.57 ±0.01	29.95 ±0.03	30.28 ±0.07	30.08 ±0.02	19.97 ±0.16	31.97 ±0.83	30.80	32.14
NECO	84.57 ±0.69	55.23 ±3.44	69.74 ±0.12	53.34 ±5.89	45.51 ±0.75	61.68	84.20 ±4.58	70.78 ±2.21	76.43 ±5.43	68.87 ±6.96	71.46 ±1.41	67.12 ±1.86	73.14	67.93
FDBD	79.92 ±1.05	35.72 ±3.15	56.09 ±2.38	26.37 ±3.72	25.31 ±0.45	44.68	72.31 ±3.77	45.50 ±8.47	52.71 ±8.25	50.62 ±10.02	48.40 ±4.69	42.29 ±5.34	51.97	48.66
CASOD	83.20 ±0.25	50.86 ±0.25	77.61 ±0.31	55.27 ±0.65	55.07 ±4.03	64.40	95.08 ±0.08	98.27 ±0.03	93.08 ±0.34	86.38 ±0.43	77.94 ±0.52	82.02 ±0.72	88.79	77.71
	ID data: ImageNet-200													
	ImageNet-100	ImageNet-800	NINCO	SSB-Hard	Food	Avg.N	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg.F	Avg.
MSP	70.59 ±0.27	39.82 ±0.68	65.99 ±0.28	24.59 ±0.40	29.65 ±0.09	46.13	76.30 ±0.66	49.81 ±0.53	52.78 ±0.31	53.33 ±0.44	48.95 ±0.69	49.42 ±0.23	55.10	51.02
ENERGY	69.72 ±0.12	35.06 ±0.17	64.24 ±0.05	24.14 ±0.30	30.30 ±0.33	44.69	73.40 ±0.13	47.96 ±0.09	55.06 ±0.22	53.96 ±0.08	44.66 ±0.02	49.43 ±0.09	54.08	49.81
KNN	71.13 ±0.19	41.76 ±0.54	68.41 ±0.44	25.67 ±0.35	30.97 ±0.09	47.59	91.59 ±0.13	54.21 ±0.64	72.37 ±1.75	57.02 ±1.53	64.41 ±0.82	53.13 ±0.43	65.45	57.33
VIM	73.38 ±0.14	49.21 ±0.23	72.70 ±0.12	36.45 ±0.20	33.33 ±0.25	53.01	94.67 ±0.04	69.24 ±0.14	91.11 ±0.37	70.71 ±0.74	71.18 ±0.39	60.99 ±0.15	76.32	65.72
MD	75.84 ±0.05	55.00 ±0.08	65.94 ±0.08	40.71 ±0.14	34.06 ±0.08	54.31	83.22 ±0.05	51.59 ±0.12	65.09 ±0.75	58.48 ±0.35	49.20 ±0.03	57.24 ±0.19	60.80	57.85
GEN	62.99 ±0.09	21.34 ±0.10	61.52 ±0.17	21.60 ±0.22	30.45 ±0.37	39.58	60.16 ±0.29	52.96 ±0.73	46.55 ±0.14	46.71 ±0.09	36.33 ±0.21	53.68 ±0.57	49.40	44.93
REACT	68.34 ±0.15	34.18 ±0.49	62.65 ±0.05	23.23 ±0.35	30.30 ±0.33	43.74	72.31 ±0.05	46.02 ±0.10	53.45 ±0.18	52.35 ±0.09	43.23 ±0.03	47.58 ±0.06	52.49	48.51
PCA	74.54 ±0.10	44.68 ±0.14	69.18 ±0.11	31.42 ±0.02	35.06 ±0.13	50.98	81.78 ±0.09	56.28 ±0.34	67.25 ±0.38	63.65 ±0.20	55.53 ±0.39	57.59 ±0.04	63.68	57.90
ASH	64.95 ±0.10	25.46 ±0.18	60.18 ±0.11	18.55 ±0.12	29.24 ±0.06	39.68	60.65 ±0.07	50.09 ±0.39	47.93 ±0.20	48.32 ±0.23	37.06 ±0.18	50.58 ±0.30	49.11	44.82
SHE	67.77 ±0.12	32.48 ±0.31	64.24 ±0.34	22.80 ±0.23	29.94 ±0.16	43.45	84.23 ±0.60	51.52 ±0.51	66.45 ±1.60	56.55 ±1.39	54.79 ±0.51	49.11 ±0.17	60.44	52.72
VRA	62.04 ±0.04	20.10 ±0.02	58.55 ±0.05	21.35 ±0.09	28.44 ±0.00	38.10	58.75 ±0.01	45.43 ±0.01	45.63 ±0.06	45.55 ±0.07	32.75 ±0.03	46.91 ±0.09	45.84	42.32
NECO	73.45 ±0.10	49.01 ±0.33	66.22 ±0.20	27.40 ±0.23	32.27 ±0.29	49.67	87.96 ±0.36	47.45 ±0.21	63.46 ±0.22	57.95 ±0.29	54.24 ±0.51	51.72 ±0.26	60.46	55.56
FDBD	72.85 ±0.10	45.49 ±0.33	68.93 ±0.14	29.29 ±0.25	30.50 ±0.21	49.41	84.33 ±0.39	58.06 ±0.49	59.91 ±0.30	58.27 ±0.14	57.94 ±0.17	52.79 ±0.14	61.88	56.21
CASOD	75.83 ±0.41	55.69 ±0.73	77.26 ±0.73	45.31 ±1.99	44.41 ±3.13	59.70	94.83 ±0.47	96.77 ±0.25	93.83 ±0.44	92.10 ±0.35	73.83 ±1.74	70.96 ±0.11	87.05	74.62

E. Further Analysis of CASOD

E.1. Impact of CASOD Views

We empirically analyze the performance of the original (Org), foreground (Fg), and background (Bg) views to evaluate the impact of each view feature on CASOD performance (top 3 rows of Table 4 in the main paper). We can see that using each view individually is insufficient. The original view (Org) performs the best and the foreground view (Fg) performs the second best, both of which are expected. While performance for the background view (Bg) is low, it still outperforms random guessing (i.e. AUC 50%), implying it

Table 8. **OOD Results for CASOD and different OOD methods**: OOD detection comparison (**AUC**) between CASOD using MD (CASOD), CASOD using MSP (CASOD-MSP), and CASOD using KNN (CASOD-KNN).

	ID data: ImageNet-100-Random													
	I-30	I-900	NINCO	SSB-Hard	Food	Avg. - Near	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg. - Far	Avg.
MSP	83.55	85.44	89.54	85.16	80.18	84.77	92.71	94.17	88.41	86.96	92.65	89.73	90.77	88.04
KNN	77.32	78.33	80.57	73.28	65.76	75.05	96.10	80.39	86.74	80.29	87.80	77.33	84.78	80.36
CASOD-MSP	88.06	90.43	91.09	88.81	86.64	89.01	92.63	95.75	93.57	90.38	92.69	90.57	92.60	90.97
CASOD-KNN	85.01	88.34	92.33	86.62	88.53	88.17	96.66	98.32	96.36	94.22	95.00	91.54	95.35	92.08
CASOD	85.42	88.30	93.26	89.66	91.34	89.60	99.02	99.89	98.46	96.40	95.55	94.56	97.31	93.81
	ID data: ImageNet-200													
	I-100	I-800	NINCO	SSB-Hard	Food	Avg. - Near	Textures	iNaturalist	Places	SUN	OpenImage-O	Species	Avg. - Far	Avg.
MSP	82.16	81.27	77.53	64.13	65.21	74.06	87.56	78.50	77.58	79.26	79.87	68.97	78.62	76.55
KNN	81.99	82.82	83.72	72.69	72.17	78.68	95.67	86.53	92.56	89.17	89.75	76.90	88.43	84.00
CASOD-MSP	89.68	89.99	89.00	83.23	88.40	88.06	92.44	95.43	93.94	94.53	92.65	81.28	91.71	90.05
CASOD-KNN	88.14	88.91	89.70	80.27	87.99	87.00	95.77	96.97	96.42	96.44	93.43	83.78	93.80	91.62
CASOD	85.03	86.93	89.96	82.54	87.14	86.31	98.48	99.42	98.26	97.81	94.14	86.60	95.79	91.48
	ID data: ImageNet-1k													
			NINCO	SSB-Hard	Food	Avg. - Near	Textures	iNaturalist			OpenImage-O		Avg. - Far	Avg.
MSP	/	/	78.11	68.94	72.27	73.11	85.52	88.19	/	/	84.86	/	86.19	79.28
KNN	/	/	82.25	65.98	76.62	74.95	91.12	91.46	/	/	89.86	/	90.81	81.75
CASOD-MSP	/	/	82.20	71.40	76.09	76.56	81.56	90.04	/	/	86.71	/	86.10	81.33
CASOD-KNN	/	/	81.34	66.29	77.74	78.46	87.06	88.83	/	/	86.59	/	87.49	82.98
CASOD	/	/	86.35	72.76	77.60	78.97	93.02	98.50	/	/	90.38	/	93.99	86.08

learns discriminative features for some classes.

We additionally investigate naive fusion of pairs of views to evaluate the necessity of using all three views and intelligent fusion. Pairs of views (Org+Fg, Org+Bg, Fg+Bg) were concatenated and processed through a linear layer [123]. Performance is overall worse. This demonstrates that CVCA fusion in CASOD is necessary for learning optimal features from segmented views. We observe that the superior performance of Fg+Bg over the other paired views indicates important feature information in the segmented views, although not sufficient by themselves to outperform the Org view.

E.2. Contrast with Self-Attention Fusion

The alternative method of using self-attention for multi-view fusion is inferior, as evidence by studies on attention mechanisms [49, 77, 140, 173, 179] and our results in Table 4 of the main paper. The self-attention implementation we compared with concatenated view features along the token dimension, passed the concatenated feature through a self-attention block, and used mean-pooling on the resultant feature along the token dimension. Performance was worse, as expected.

Theoretical Analysis We propose the following theoretical analysis. Self-attention applies a softmax to the features during computation of the attention matrix $A_{SA} \approx \text{softmax}(zW_Q W_K^\intercal z^\intercal)$. The resultant attention matrix A_{SA} consists of $3T$ rows of distributions over $3T$ patches $A_{SA} \in \mathbb{R}^{3T\times 3T}$, each with values that must be $[0,1]$, $A_{SA,i} = \{[p_1, p_2, \ldots, p_{3T}]; p \in [0,1]\}$. CASOD uses a sum of two cross-attention matrices in each CVCA block $A_{CVCA} \approx \text{softmax}(z_{v1}W_{Q,v1}W_{K,v2}^\intercal z_{v2}^\intercal) + \text{softmax}(z_{v2}W_{Q,v2}W_{K,v1}^\intercal z_{v1}^\intercal)$. The resultant CVCA attention matrix A_{CVCA} is instead T rows of values over T patches $A_{CVCA} \in \mathbb{R}^{T\times T}$, but each value has magnitude in $[0,2]$, $A_{CVCA,i} = \{[p_1, p_2, \ldots, p_T]; p \in [0,2]\}$.

If we consider a particular high-activation patch $T_j^* \in z$ in the original image that is also in the foreground view $T_{fg,j}^* \in z_{fg}$, self-attention *reduces* the magnitude of both patch activations via the softmax over all $3T$ patches. Meanwhile, our CVCA method *additively reinforces* them. Both SA and CVCA optimize towards the same objective, incentivizing the attention matrices to highlight these high-activation patches. Necessarily, this requires W_Q, W_K, and W_V to also highlight these patches, which is behavior observed in literature [119, 160]. Given this preservation of high-activation patches into the attention matrix, the denominator of the softmax in SA is greater than the denominator of the softmax in each CVCA term

$$\sum_{j=1}^{3T} e^{a_{SA,j}} > \sum_{j=1}^{T} e^{a_{CVCA,j}}. \tag{1}$$

This implies

$$T_{j,CVCA}^* \geq T_{j,SA}^*, \tag{2}$$

where $T_{j,CVCA}^* \in A_{CVCA} z_{v1} W_{V,v1}$ and $T_{j,SA}^* \in A_{SA} z W_V$ are the high-activation patches of interest in the features from the attention blocks.

E.3. Comparison with Equivalent Data and Parameters

Here we detail the experiments evaluating the performance of single-view baseline models when provided data and learnable parameters equivalent to those provided to CASOD.

Equivalent Data. Two experiments were conducted where the baseline model was provided with the original image and two additional augmented images per sample to equalize the amount of training data between baselines and CASOD. The augmentations chosen for the first experiment

were based on findings indicating their benefit for OOD detection and include *rotations* (from CSI [135]) and *noise* (from random augmentation works [22]). For the second experiment, we use the foreground and background views (used in CASOD) as the additional augmented images. The OOD detection results of these experiment on ID dataset ImageNet-100-Random are presented in main paper Table 5 in rows "Aug Base" and "Fg-Bg Base" respectively. We chose the best-performing baseline OOD detection method, ViM, for comparison. CASOD consistently outperforms baselines trained with equivalent data. This implies that the CVCA in CASOD is key to improve OOD detection through fusion of the segmented foreground-background images. Furthermore, CASOD performance benefits do not come from additional data.

Equivalent Parameters. Another claim could be that CASOD outperforms existing baselines due to the increase in parameters. To empirically verify that this is not the case, we conduct an experiment with a version of the single-view baseline model containing learnable parameters equivalent to those in CASOD (i.e. 48M). The results on ID dataset ImageNet-100-Random are shown in main paper Table 5. Even with more parameters, the baselines do not outperform CASOD, indicating that the view fusion mechanisms introduced by CASOD are critical to improved performance instead of the parameter increase.

We note that changing the feature extraction backbone for the single-view baseline (i.e. by increasing the number of learnable parameters) but not for CASOD is an unfair comparison. A significant amount of the learnable parameters (28.67 MiB) in CASOD are in the feature fusion head, which is designed to be modular to patch-based feature extraction backbones. The intuitive decision when presented with a superior feature extraction backbone would be to use it with our proposed feature fusion head, thus re-establishing a fair comparison with *equivalent feature extractors*. Nevertheless, CASOD still outperforms the single-view baseline with equivalent parameters despite using weaker feature extraction backbones.

E.4. Segmented Foreground-Background Quality and Domain Shift

We provide details for our investigation into if CASOD performance benefits originate in properly segmented foregrounds and backgrounds. For this, two questions are asked:

1. Does CASOD require properly separated foregrounds and backgrounds?
2. Can CASOD still perform well across datasets of varying separation quality?

For (1), we generate an additional dataset with poor semantic correlation across views: one view contains off-center crops of the image while the other view contains the rest of the image ("Crops"). Results of this experiment on ID dataset ImageNet-100-Random are presented in main paper Table 5. CASOD with the poor segmentation-quality dataset is weaker than CASOD with foreground-background segmentation. This demonstrates that increasingly semantically meaningful foregrounds and backgrounds directly improve the performance of CASOD.

For (2), we select non-ImageNet datasets with varying foreground-background separation quality as ID data. The Textures dataset does not have clear foreground objects and we use it to evaluate CASOD performance when ID training data foregrounds and backgrounds are poorly separated. Furthermore, we evaluate CASOD with NINCO as ID because it includes complex samples with foregrounds containing multiple objects. With Textures as the ID dataset, we use I-30, iNaturalist, Places, NINCO, SUN, OpenImage-O, Species, SSB-HARD, and Food as OOD datasets. For Textures, CASOD was trained using settings identical to those of ImageNet-100-Random. With NINCO as the ID dataset, we use I-30, SSB-HARD, and Food as near-OOD datasets and Textures, iNaturalist, Places, SUN, OpenImage-O, and Species as far-OOD datasets. We trained CASOD to convergence with a LoRA rank of 16. All other training and model configurations were identical to the ImageNet-100-Random experiment settings. Results in main paper Table 7 indicate that CASOD is robust to foreground-background segmentation quality during training. Despite the varying quality of features in the views, CASOD still outperforms baselines and comparison points. Furthermore, the results reinforce that the proposed cross-view attention is critical to learning discriminative features in images.

Domain shift. Segmentation quality can also be used as a basis for analysis into the performance of CASOD under domain shift. The selected segmentation model for CASOD was YOLOv5, which is trained on an object dataset (i.e. COCO) [63]. However, tasks may not always involve object data. We investigate the performance of CASOD under *domain shift*, that is when the ID and OOD data contain fewer or unclear object features. The results on NINCO and Textures as ID datasets indicates that CASOD is also robust to domain shift. NINCO is similar to COCO and CASOD performs well. Textures contains images of textures, not objects [19], and thus exhibits domain shift from COCO which contains 80 classes of clear objects [83]. CASOD still performs well, although the performance gap is smaller with respect to the baseline.

To evaluate CASOD under severe domain shift, we additionally experiment with a medical dataset that is very different from COCO. Specifically, we use the MEL, NV, BCC, AK, and DF classes from the ISIC 2019 dataset [20] as ID and the BKL, VASC, and SCC classes from the ISIC 2019 dataset as well as 2000 randomly selected samples from the NCT dataset [68] as OOD [103]. The average AUC performance of CASOD is 79.93% while the closest baseline

is 79.78%. CASOD still slightly outperforms the baseline under severe domain shift, likely due to the preservation of the original view ensuring at least the learned knowledge of single-view models.

E.5. Importance of Segmentation Model

To further investigate the importance of segmentation model in CASOD, we experiment with a U2-Net pre-trained on the DUTS-TR [146] dataset. We use ImageNet-100-Random for ID and Textures and NINCO for OOD. To ensure fairness, we remove all overlapping data between DUTS-TR and the ID and OOD datasets. The average AUC and FPR results for CASOD with U2-Net is 83.51% (AUC) and 63.63% (FPR) while CASOD (with YOLOv5) is 96.14% (AUC) and 20.39% (FPR). The older U2-Net is simultaneously slower at segmentation (taking about 0.8 seconds per image) and weaker than YOLOv5 at OOD detection. Qualitative analysis showed that U2-Net often segmented foreground objects correctly but frequently included spurious background features, thus diluting the discriminative features.

Following recent trends in OOD detection literature using pre-trained vision-language models (VLMs) (e.g. CLIP [113]) [62, 76, 100], we investigate using the vision backbone from pre-trained CLIP as the feature extractor in CASOD. The results are in Table 9. CASOD with ViT and LoRA feature extractor backbones outperforms CASOD with CLIP vision feature extractor backbones. CASOD with LoRA adds additional learning capacity to the model, allowing it to learn to extract more discriminative features during training. Additionally, CLIP typically also uses *text data* which is not considered in our domain. This simultaneously shows the reliance on CLIP-based OOD detection methods on text inputs as well as image inputs and demonstrates that VLM-based methods have certain limitations that prevent them from being directly translated to the image-only domain. Overall, we consider this to be an invalid comparison as 1) CLIP relies on text data while we do not, and 2) CLIP has likely seen some of the ID and OOD data before during pre-training, leading to information leakage. There is evidence for this in the literature, with CLIP achieving extremely high zero-shot OOD detection performance [30]. CASOD overcomes this challenge through our proposed view fusion method.

Table 9. Ablation OOD detection results (AUC) for ImageNet-100-Random ID between CASOD with LoRA ViT feature extractor backbones and CASOD with CLIP vision feature extractor backbone.

	OOD		
	Average (Far)	Average (Near)	Average
CASOD-CLIP	93.84	85.90	90.04
CASOD	**97.39**	**89.68**	**93.89**

E.6. Individual View Loss Contribution

We additionally investigated CASOD using *only* the fused feature CE loss $\mathcal{L}_{CE,fused}$, i.e. $\mathcal{L}_{CE} = \mathcal{L}_{CE,fused}$. The OOD detection AUC results for ID dataset ImageNet-100-Random are presented in Table 10. Training with only the final fused feature reduces overall performance. This is primarily because including the individual view CE loss terms encourages learning of more diverse class discriminative features in each of the corresponding views. As the features in these views are different, the resulting individual view features contain more diverse discriminative information for feature fusion. This is particularly noticeable in the increase in near-OOD performance relative to far-OOD performance. Learning more discriminative features improves detection of challenging near-OOD samples, closing the gap for CASOD to **4.21%** from 5.31% for CASOD with only fused feature CE loss.

Table 10. Ablation OOD detection results (AUC) for ImageNet-100-Random ID between CASOD and CASOD when trained without individual view feature losses ($\mathcal{L}_{CE,fused}$ only).

	OOD		
	Average (Far)	Average (Near)	Average
$\mathcal{L}_{CE,fused}$ only	97.05	88.19	93.50
CASOD	**97.39**	**89.68**	**93.89**

F. Additional Related Works

The closed-world assumption has been documented by many researchers and in many different contexts [8, 36, 65, 120]. Expanding ML to open-world settings [65, 85, 120] remains a priority area of research for its many potential applications in real-world settings such as autonomous driving [31, 106], robotics [32, 44, 101], medical diagnosis [13, 122], and industrial automation [10], among many other areas. One form of solution to addressing open-world settings is OOD detection [23, 45, 127, 158, 159, 168], or one of its many related forms. These include rejection [21, 43] which aims to evaluate samples explicitly as "seen" or "unseen", anomaly detection [5, 86, 137] which compares evaluation samples against "ID" data as a whole (rather than via classification with access to labels), open set recognition [7, 8, 42, 104, 120] similarly learns to separate between samples in the seen set and all other unseen samples, and uncertainty estimation [91–93] investigates quantification of ML model uncertainty for use in evaluating the quality of predictions (and thus data samples), among others.

OOD detection has been investigated from many different directions. *Confidence-based methods* use model logits to quantify confidence in a sample for use in determining if the sample is OOD or not [159]. These methods either directly evaluate logits [51, 53, 151] or modify logit representation to be better characterized for separating OOD samples

[17, 56, 57, 81, 87, 89, 90, 98, 132, 150, 156, 167]. Other related directions consider probability density [1, 110, 180], model uncertainty [34, 47, 61, 80, 88, 121, 129, 138], architectural changes for improved confidence estimation [25, 161], decomposed features or logits [58, 60], augmented weights [40, 126, 131], or augmented logits [14, 71, 91, 115, 125, 153, 155, 162]. A related direction to OOD detection is *outlier exposure*, which provides some representative OOD samples to the detector during training [16, 33, 52, 87, 96, 98, 109, 157, 175]. This direction is often considered distinct from OOD detection which does **not** allow the model to see any OOD data during training.

Distance-based methods use some measure of distances, often in the model feature space, to detect OOD samples. Mahalanobis distance is a flagship method [72, 117] which uses covariance to estimate sample distance from ID data. Other works use non-linear spaces to map and compare samples [6, 116, 133, 148, 170]. Additional success has been found by combining distance measurements with logits-based confidence scores [4, 144]. Recently, OOD distance measures have been generated using ML model components besides features [2, 3, 27, 38, 48, 97, 107].

Self-supervised approaches focus on learning better features for use in downstream OOD detection tasks, improving the performance of both confidence-based and distance-based methods [9, 46, 54, 59]. Such methods include contrastive losses [99, 112, 114, 135] or generative [29, 102, 108, 122, 124, 147, 162] methods. Works have found further success by including data augmentation [128, 165], data mixing [55, 138, 163], and reconstruction [24, 178] during self-supervised training and tuning.

Multi-view classification [41, 78, 123, 130, 145, 154, 169, 174] is a rapidly growing area of ML with significant potential, evidence by its expansion in many applications [35, 50, 78, 84, 164, 177]. Transformer-based attention [28, 142] layers have recently been shown to have great success in multi-view approaches [15, 136, 166, 174]. Self-attention and cross-attention methods [39, 140, 166] allow for features to be comprehensively learned while incorporating information from all views [79, 82, 140, 179]. Further performance gains were observed with further refinements to view feature fusion [15, 73, 149] and alignment of learned features [67, 69, 75, 82].

Multi-view OOD detection is relatively uninvestigated. Some applications, such as MRI image lesion detection, have used multi-view autoencoders [37]. Other processes, such as Gaussian Processes, have been improved through multi-view processing [66]. Vision-language models have expanded multi-modal processing to multi-view domains using CLIP [113] to identify image foregrounds through prompt learning [74, 94, 100, 111, 181]. Anomaly detection, a parallel domain to OOD detection, has considered multi-view approaches but typically only considers the foreground [18, 139, 171]. OOD segmentation more directly considers separated foreground and background components in images but does not evaluate the image as a whole [95, 172]. Most similar to ours, one work uses dense prediction to disentangle foreground and background features in an image but does not use segmentation or fusion for improved feature learning [26].

References

[1] Davide Abati, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Latent space autoregression for novelty detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 481–490, 2019. 10

[2] Vahdat Abdelzad, Krzysztof Czarnecki, Rick Salay, Taylor Denounden, Sachin Vernekar, and Buu Phan. Detecting out-of-distribution inputs in deep neural networks using an early-layer output. *arXiv preprint arXiv:1910.10307*, 2019. 10

[3] Yong Hyun Ahn, Gyeong-Moon Park, and Seong Tae Kim. Line: Out-of-distribution detection by leveraging important neurons. *arXiv preprint arXiv:2303.13995*, 2023. 10

[4] Mouïn Ben Ammar, Nacim Belkhir, Sebastian Popescu, Antoine Manzanera, and Gianni Franchi. NECO: NEural collapse based out-of-distribution detection. In *ICLR*, 2024. 10

[5] Jerone Andrews, Thomas Tanay, Edward J Morton, and Lewis D Griffin. Transfer representation-learning for anomaly detection. JMLR, 2016. 9

[6] Sima Behpour, Thang Long Doan, Xin Li, Wenbin He, Liang Gou, and Liu Ren. Gradorth: A simple yet efficient out-of-distribution detection with orthogonal projection of gradients. *Advances in Neural Information Processing Systems*, 36, 2024. 10

[7] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1893–1902, 2015. 9

[8] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016. 9

[9] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. In *International Conference on Learning Representations*, 2019. 10

[10] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 9

[11] Julian Bitterwolf, Maximilian Mueller, and Matthias Hein. In or out? fixing imagenet out-of-distribution detection evaluation. *arXiv:2306.00826*, 2023. 1

[12] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In *ECCV*, 2014. 1

[13] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1721–1730, 2015. 9

[14] Chao Chen, Zhihang Fu, Kai Liu, Ze Chen, Mingyuan Tao, and Jieping Ye. Optimal parameter and neuron pruning for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 36, 2024. 10

[15] Chun-Fu Richard Chen, Quanfu Fan, and Rameswar Panda. Crossvit: Cross-attention multi-scale vision transformer for image classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 357–366, 2021. 10

[16] Jiefeng Chen, Yixuan Li, Xi Wu, Yingyu Liang, and Somesh Jha. Atom: Robustifying out-of-distribution detection using outlier mining. In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part III 21*, pages 430–445. Springer, 2021. 10

[17] Jinggang Chen, Junjie Li, Xiaoyang Qu, Jianzong Wang, Jiguang Wan, and Jing Xiao. Gaia: delving into gradient-based attribution abnormality for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 36:79946–79958, 2023. 10

[18] Xiaolu Chen, Haote Xu, Chenghao Deng, Xiaotong Tu, Xinghao Ding, and Yue Huang. Implicit foreground-guided network for anomaly detection and localization. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2970–2974. IEEE, 2024. 10

[19] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *CVPR*, 2014. 1, 8

[20] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018. 8

[21] Corinna Cortes, Giulia DeSalvo, and Mehryar Mohri. Learning with rejection. In *Algorithmic Learning Theory: 27th International Conference, ALT 2016, Bari, Italy, October 19-21, 2016, Proceedings 27*, pages 67–82. Springer, 2016. 9

[22] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *CVPR*, 2020. 8

[23] Peng Cui and Jinjia Wang. Out-of-distribution (ood) detection based on deep learning: A review. *Electronics*, 11(21):3500, 2022. 9

[24] Taylor Denouden, Rick Salay, Krzysztof Czarnecki, Vahdat Abdelzad, Buu Phan, and Sachin Vernekar. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. *arXiv preprint arXiv:1812.02765*, 2018. 10

[25] Terrance DeVries and Graham W Taylor. Learning confidence for out-of-distribution detection in neural networks. *arXiv:1802.04865*, 2018. 10

[26] Choubo Ding and Guansong Pang. Improving open-world classification with disentangled foreground and background features. In *ACM Multimedia*, 2024. 10

[27] Xin Dong, Junfeng Guo, Ang Li, Wei-Te Ting, Cong Liu, and HT Kung. Neural mean discrepancy for efficient out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19217–19227, 2022. 10

[28] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020. 2, 10

[29] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. *arXiv preprint arXiv:2202.01197*, 2022. 10

[30] Sepideh Esmaeilpour, Bing Liu, Eric Robertson, and Lei Shu. Zero-shot out-of-distribution detection based on the pre-trained model clip. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6568–6576, 2022. 9

[31] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1625–1634, 2018. 9

[32] Fabio Falcini and Giuseppe Lami. Deep learning in automotive: Challenges and opportunities. In *Software Process Improvement and Capability Determination: 17th International Conference, SPICE 2017, Palma de Mallorca, Spain, October 4–5, 2017, Proceedings*, pages 279–288. Springer, 2017. 9

[33] Ke Fan, Tong Liu, Xingyu Qiu, Yikai Wang, Lian Huai, Zeyu Shangguan, Shuang Gou, Fengjian Liu, Yuqian Fu, Yanwei Fu, et al. Test-time linear out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23752–23761, 2024. 10

[34] Kun Fang, Qinghua Tao, Kexin Lv, Mingzhen He, Xiaolin Huang, and Jie Yang. Kernel pca for out-of-distribution detection. *arXiv preprint arXiv:2402.02949*, 2024. 10

[35] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 643–650, 2015. 10

[36] Geli Fei and Bing Liu. Breaking the closed world assumption in text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 506–514, 2016. 9

[37] Alvaro Fernandez-Quilez, Linas Vidziunas, Ørjan Kløvfjell Thoresen, Ketil Oppedal, Svein Reidar Kjosavik, and Trygve Eftestøl. Out-of-distribution multi-view auto-encoders for prostate cancer lesion detection. In *ISBI*, pages 1–5. IEEE, 2023. 10

[38] Stanislav Fort, Jie Ren, and Balaji Lakshminarayanan. Exploring the limits of out-of-distribution detection. *NeurIPS*, 2021. 10

[39] Stella Frank, Emanuele Bugliarello, and Desmond Elliott. Vision-and-language or vision-for-language? on cross-modal influence in multimodal transformers. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9847–9857, Online and Punta Cana, Dominican Republic, 2021. Association for Computational Linguistics. 10

[40] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 10

[41] Zan Gao, DY Wang, YB Xue, GP Xu, Hua Zhang, and YL Wang. 3d object recognition based on pairwise multi-view convolutional neural networks. *Journal of Visual Communication and Image Representation*, 56: 305–315, 2018. 10

[42] Zongyuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. In *British Machine Vision Conference*. BMVA Press, 2017. 9

[43] Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. *Advances in neural information processing systems*, 30, 2017. 9

[44] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 9

[45] Navid Ghassemi and Ehsan Fazl-Ersi. A comprehensive review of trends, applications and challenges in out-of-distribution detection. *arXiv preprint arXiv:2209.12935*, 2022. 9

[46] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. *Advances in neural information processing systems*, 31, 2018. 10

[47] Xiaoyuan Guan, Zhouwu Liu, Wei-Shi Zheng, Yuren Zhou, and Ruixuan Wang. Revisit pca-based technique for out-of-distribution detection. In *CVPR*, 2023. 10

[48] Manuel Gunther, Steve Cruz, Ethan M Rudd, and Terrance E Boult. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 71–80, 2017. 10

[49] Yutong He, Ruslan Salakhutdinov, and J Zico Kolter. Localized text-to-image generation for free via cross attention control. *arXiv preprint arXiv:2306.14636*, 2023. 7

[50] Ziqiang He, Shaohua Wan, Marco Zappatore, and Hu Lu. A similarity matrix low-rank approximation and inconsistency separation fusion approach for multi-view clustering. *IEEE Transactions on Artificial Intelligence*, 2023. 10

[51] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2017. 3, 9

[52] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018. 10

[53] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. *arXiv:1911.11132*, 2019. 1, 9

[54] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *NeurIPS*, 2019. 10

[55] Dan Hendrycks, Andy Zou, Mantas Mazeika, Leonard Tang, Bo Li, Dawn Song, and Jacob Steinhardt. Pixmix: Dreamlike pictures comprehensively improve safety measures. In *ICLR*, 2022. 10

[56] Claus Hofmann, Simon Schmid, Bernhard Lehner, Daniel Klotz, and Sepp Hochreiter. Energy-based hopfield boosting for out-of-distribution detection. *arXiv preprint arXiv:2405.08766*, 2024. 10

[57] Claus Hofmann, Simon Lucas Schmid, Bernhard Lehner, Daniel Klotz, and Sepp Hochreiter. Energy-based hopfield boosting for out-of-distribution detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2025. 10

[58] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10951–10960, 2020. 10

[59] Wenpeng Hu, Mengyu Wang, Qi Qin, Jinwen Ma, and Bing Liu. Hrn: A holistic approach to one class learning. *NeurIPS*, 33:19111–19124, 2020. 10

[60] Rui Huang and Yixuan Li. Mos: Towards scaling out-of-distribution detection for large semantic space. In *CVPR*, 2021. 1, 10

[61] Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. *NeurIPS*, 2021. 10

[62] Xue Jiang, Feng Liu, Zhen Fang, Hong Chen, Tongliang Liu, Feng Zheng, and Bo Han. Negative label guided ood detection with pretrained vision-language models. *arXiv preprint arXiv:2403.20078*, 2024. 9

[63] Glenn Jocher, Alex Stoken, Jirka Borovec, Liu Changyu, Adam Hogan, Laurentiu Diaconu, Jake Poznanski, Lijun Yu, Prashant Rai, Russ Ferriday, et al. ultralytics/yolov5: v3. 0. *Zenodo*, 2020. 2, 8

[64] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. 2

[65] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *CVPR*, 2021. 9

[66] Myong Chol Jung, He Zhao, Joanna Dipnall, Belinda Gabbe, and Lan Du. Uncertainty estimation for multi-view data: The power of seeing the whole picture. *NeurIPS*, 2022. 10

[67] Andrej Karpathy, Armand Joulin, and Li F Fei-Fei. Deep fragment embeddings for bidirectional image sentence mapping. *Advances in neural information processing systems*, 27, 2014. 10

[68] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue. *(No Title)*, 2018. 8

[69] Guanzhou Ke, Bo Wang, Xiaoli Wang, and Shengfeng He. Rethinking multi-view representation learning via distilled disentangling. *arXiv preprint arXiv:2403.10897*, 2024. 10

[70] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 1, 2

[71] Gerhard Krumpl, Henning Avenhaus, Horst Possegger, and Horst Bischof. Ats: Adaptive temperature scaling for enhancing out-of-distribution detection methods. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3864–3873, 2024. 10

[72] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *NeurIPS*, 31, 2018. 3, 10

[73] Seungik Lee, Jaehyeong Park, and Jinsun Park. Crossformer: Cross-guided attention for multi-modal object detection. *Pattern Recognition Letters*, 2024. 10

[74] Yuxiao Lee, Xiaofeng Cao, Jingcai Guo, Wei Ye, Qing Guo, and Yi Chang. Concept matching with agent for out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4562–4570, 2025. 10

[75] Hui Li, Xiao-Jun Wu, and Josef Kittler. Rfn-nest: An end-to-end residual fusion network for infrared and visible images. *Information Fusion*, 73:72–86, 2021. 10

[76] Jinglun Li, Kaixun Jiang, Zhaoyu Chen, Bo Lin, Yao Tang, Weifeng Ge, and Wenqiang Zhang. Synthesizing near-boundary ood samples for out-of-distribution detection. *arXiv preprint arXiv:2507.10225*, 2025. 9

[77] Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, et al. Grounded language-image pre-training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10965–10975, 2022. 7

[78] Rui Li, Dong Gong, Wei Yin, Hao Chen, Yu Zhu, Kaixuan Wang, Xiaozhi Chen, Jinqiu Sun, and Yanning Zhang. Learning to fuse monocular and multi-view cues for multi-frame depth estimation in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21539–21548, 2023. 10

[79] Shaochen Li, Zhenyu Liu, Guifang Duan, and Jianrong Tan. Mvhanet: multi-view hierarchical aggregation network for skeleton-based hand gesture recogni-

tion. *Signal, Image and Video Processing*, pages 1–9, 2023. 10

[80] Yixia Li, Boya Xiong, Guanhua Chen, and Yun Chen. Setar: Out-of-distribution detection with selective low-rank approximation. *arXiv preprint arXiv:2406.12629*, 2024. 10

[81] 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. 10

[82] Tao Liang, Guosheng Lin, Lei Feng, Yan Zhang, and Fengmao Lv. Attention is not enough: Mitigating the distribution discrepancy in asynchronous multimodal sequence fusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8148–8156, 2021. 10

[83] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 8

[84] An-An Liu, Nian Hu, Dan Song, Fu-Bin Guo, He-Yu Zhou, and Tong Hao. Multi-view hierarchical fusion network for 3d object retrieval and classification. *IEEE Access*, 7:153021–153030, 2019. 10

[85] Bing Liu, Sahisnu Mazumder, Eric Robertson, and Scott Grigsby. Ai autonomy: Self-initiated open-world continual learning and adaptation. *AI Magazine*, 2023. 9

[86] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008. 9

[87] Kai Liu, Zhihang Fu, Sheng Jin, Chao Chen, Ze Chen, Rongxin Jiang, Fan Zhou, Yaowu Chen, and Jieping Ye. Rethinking out-of-distribution detection on imbalanced data distribution. *arXiv preprint arXiv:2407.16430*, 2024. 10

[88] Litian Liu and Yao Qin. Fast decision boundary based out-of-distribution detector. In *ICML*, 2024. 10

[89] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *NeurIPS*, 2020. 10

[90] Xixi Liu, Yaroslava Lochman, and Christopher Zach. Gen: Pushing the limits of softmax-based out-of-distribution detection. In *CVPR*, 2023. 10

[91] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018. 9, 10

[92] Andrey Malinin and Mark Gales. Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness. *Advances in Neural Information Processing Systems*, 32, 2019.

[93] Andrey Malinin, Bruno Mlodozeniec, and Mark Gales. Ensemble distribution distillation. In *International Conference on Learning Representations*, 2019. 9

[94] Evelyn Mannix and Howard Bondell. Paws-vmk: A unified approach to semi-supervised learning and out-of-distribution detection. *arXiv preprint arXiv:2311.17093*, 2023. 10

[95] Samuel Marschall and Kira Maag. Multi-scale foreground-background confidence for out-of-distribution segmentation. *arXiv preprint arXiv:2412.16990*, 2024. 10

[96] Nikhil Mehta, Kevin J Liang, Jing Huang, Fu-Jen Chu, Li Yin, and Tal Hassner. Hypermix: Out-of-distribution detection and classification in few-shot settings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2410–2420, 2024. 10

[97] Pedro R Mendes Júnior, Roberto M De Souza, Rafael de O Werneck, Bernardo V Stein, Daniel V Pazinato, Waldir R de Almeida, Otávio AB Penatti, Ricardo da S Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 106(3):359–386, 2017. 10

[98] Wenjun Miao, Guansong Pang, Jin Zheng, and Xiao Bai. Long-tailed out-of-distribution detection via normalized outlier distribution adaptation. *Advances in Neural Information Processing Systems*, 37:132106–132132, 2024. 10

[99] Yifei Ming, Yiyou Sun, Ousmane Dia, and Yixuan Li. How to exploit hyperspherical embeddings for out-of-distribution detection? *arXiv:2203.04450*, 2022. 10

[100] Atsuyuki Miyai, Qing Yu, Go Irie, and Kiyoharu Aizawa. Locoop: Few-shot out-of-distribution detection via prompt learning. *NeurIPS*, 2024. 9, 10

[101] Khan Muhammad, Amin Ullah, Jaime Lloret, Javier Del Ser, and Victor Hugo C de Albuquerque. Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4316–4336, 2020. 9

[102] E Nalisnick, A Matsukawa, Y Teh, D Gorur, and B Lakshminarayanan. Do deep generative models know what they don't know? In *International Conference on Learning Representations*, 2019. 10

[103] Vivek Narayanaswamy, Yamen Mubarka, Rushil Anirudh, Deepta Rajan, and Jayaraman J Thiagarajan. Exploring inlier and outlier specification for improved medical ood detection. In *Proceedings of the*

IEEE/CVF International Conference on Computer Vision, pages 4589–4598, 2023. 8

[104] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 613–628, 2018. 9

[105] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015. 3

[106] Julia Nitsch, Masha Itkina, Ransalu Senanayake, Juan Nieto, Max Schmidt, Roland Siegwart, Mykel J Kochenderfer, and Cesar Cadena. Out-of-distribution detection for automotive perception. In *ITSC*, pages 2938–2943. IEEE, 2021. 9

[107] Bartłomiej Olber, Krystian Radlak, Adam Popowicz, Michal Szczepankiewicz, and Krystian Chachuła. Detection of out-of-distribution samples using binary neuron activation patterns. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3378–3387, 2023. 10

[108] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2307–2316, 2019. 10

[109] Aristotelis-Angelos Papadopoulos, Mohammad Reza Rajati, Nazim Shaikh, and Jiamian Wang. Outlier exposure with confidence control for out-of-distribution detection. *Neurocomputing*, 441:138–150, 2021. 10

[110] Stanislav Pidhorskyi, Ranya Almohsen, and Gianfranco Doretto. Generative probabilistic novelty detection with adversarial autoencoders. *Advances in neural information processing systems*, 31, 2018. 10

[111] Yuehan Qin, Yichi Zhang, Yi Nian, Xueying Ding, and Yue Zhao. Metaood: Automatic selection of ood detection models. *arXiv preprint arXiv:2410.03074*, 2024. 10

[112] Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, and Maja Rudolph. Neural transformation learning for deep anomaly detection beyond images. In *International Conference on Machine Learning*, pages 8703–8714. PMLR, 2021. 10

[113] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 9, 10

[114] Edward T Reehorst and Philip Schniter. Score combining for contrastive ood detection. *arXiv preprint arXiv:2501.12204*, 2025. 10

[115] Sudarshan Regmi. Adascale: Adaptive scaling for ood detection. *arXiv preprint arXiv:2503.08023*, 2025. 10

[116] Sudarshan Regmi, Bibek Panthi, Yifei Ming, Prashnna K Gyawali, Danail Stoyanov, and Binod Bhattarai. Reweightood: Loss reweighting for distance-based ood detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 131–141, 2024. 10

[117] Jie Ren, Stanislav Fort, Jeremiah Liu, Abhijit Guha Roy, Shreyas Padhy, and Balaji Lakshminarayanan. A simple fix to mahalanobis distance for improving near-ood detection. *arXiv preprint arXiv:2106.09022*, 2021. 10

[118] Tal Ridnik, Emanuel Ben-Baruch, Asaf Noy, and Lihi Zelnik-Manor. Imagenet-21k pretraining for the masses. *arXiv preprint arXiv:2104.10972*, 2021. 1

[119] Mohammadreza Salehi, Hossein Mirzaei, Dan Hendrycks, Yixuan Li, Mohammad Hossein Rohban, and Mohammad Sabokrou. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. *arXiv preprint arXiv:2110.14051*, 2021. 7

[120] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012. 9

[121] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2317–2324, 2014. 10

[122] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 9, 10

[123] Marco Seeland and Patrick Mäder. Multi-view classification with convolutional neural networks. *Plos one*, 16(1):e0245230, 2021. 7, 10

[124] Joan Serrà, David Álvarez, Vicenç Gómez, Olga Slizovskaia, José F Núñez, and Jordi Luque. Input complexity and out-of-distribution detection with likelihood-based generative models. *arXiv preprint arXiv:1909.11480*, 2019. 10

[125] Gabi Shalev, Yossi Adi, and Joseph Keshet. Out-of-distribution detection using multiple semantic label representations. *Advances in Neural Information Processing Systems*, 31, 2018. 10

[126] Chandramouli Shama Sastry and Sageev Oore. Detecting out-of-distribution examples with in-distribution examples and gram matrices. *arXiv e-prints*, pages arXiv–1912, 2019. 10

[127] Zheyan Shen, Jiashuo Liu, Yue He, Xingxuan Zhang, Renzhe Xu, Han Yu, and Peng Cui. Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, 2021. 9

[128] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019. 10

[129] Richard L Smith. Extreme value theory. *Handbook of applicable mathematics*, 7(437-471):18, 1990. 10

[130] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *CVPR*, 2015. 10

[131] Yiyou Sun and Yixuan Li. Dice: Leveraging sparsification for out-of-distribution detection. In *European Conference on Computer Vision*, pages 691–708. Springer, 2022. 10

[132] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *NeurIPS*, 2021. 10

[133] Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *ICML*, 2022. 3, 10

[134] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 2

[135] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *NeurIPS*, 2020. 8, 10

[136] Wei Tang, Fazhi He, Yu Liu, and Yansong Duan. Matr: Multimodal medical image fusion via multiscale adaptive transformer. *IEEE Transactions on Image Processing*, 31:5134–5149, 2022. 10

[137] David Martinus Johannes Tax. One-class classification: Concept learning in the absence of counter-examples. 2002. 9

[138] Sunil Thulasidasan, Gopinath Chennupati, Jeff A Bilmes, Tanmoy Bhattacharya, and Sarah Michalak. On mixup training: Improved calibration and predictive uncertainty for deep neural networks. *NeurIPS*, 2019. 10

[139] Long Tian, Hongyi Zhao, Ruiying Lu, Rongrong Wang, Yujie Wu, Liming Wang, Xiongpeng He, and Xiyang Liu. Foct: Few-shot industrial anomaly detection with foreground-aware online conditional transport. In *ACM Multimedia*, 2024. 10

[140] Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the conference. Association for computational linguistics. Meeting*, page 6558. NIH Public Access, 2019. 7, 10

[141] Grant Van Horn, Oisin 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 *CVPR*, 2018. 1

[142] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 2017. 3, 10

[143] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: a good closed-set classifier is all you need? *arXiv*, 2021. 1

[144] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *CVPR*, 2022. 1, 10

[145] Jiao Wang, Bin Wu, Zhenwen Ren, Hongying Zhang, and Yunhui Zhou. Multi-scale deep multi-view subspace clustering with self-weighting fusion and structure preserving. *Expert Systems with Applications*, 213:119031, 2023. 10

[146] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 136–145, 2017. 9

[147] Mengyu Wang, Yijia Shao, Haowei Lin, Wenpeng Hu, and Bing Liu. Cmg: A class-mixed generation approach to out-of-distribution detection. In *ECML*, 2022. 10

[148] Qizhou Wang, Zhen Fang, Yonggang Zhang, Feng Liu, Yixuan Li, and Bo Han. Learning to augment distributions for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 36, 2024. 10

[149] Xixi Wang, Xiao Wang, Bo Jiang, Jin Tang, and Bin Luo. Mutualformer: Multi-modal representation learning via cross-diffusion attention. *International Journal of Computer Vision*, pages 1–22, 2024. 10

[150] Yezhen Wang, Bo Li, Tong Che, Kaiyang Zhou, Ziwei Liu, and Dongsheng Li. Energy-based open-world uncertainty modeling for confidence calibration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9302–9311, 2021. 10

[151] 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*, pages 23631–23644. PMLR, 2022. 3, 9

[152] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, 2010. 1

[153] Chenhui Xu, Fuxun Yu, Zirui Xu, Nathan Inkawhich, and Xiang Chen. Out-of-distribution detection via deep multi-comprehension ensemble. In *Forty-first International Conference on Machine Learning*, 2024. 10

[154] Jie Xu, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Untie: Clustering analysis with disentanglement in multi-view information fusion. *Information Fusion*, 100:101937, 2023. 10

[155] Mingyu Xu, Zheng Lian, Bin Liu, and Jianhua Tao. Vra: Variational rectified activation for out-of-distribution detection. *NeurIPS*, 2024. 10

[156] Guide Yang, Chao Hou, Weilong Peng, Xiang Fang, Yongwei Nie, Peican Zhu, and Keke Tang. Eood: Entropy-based out-of-distribution detection. *arXiv:2504.03342*, 2025. 10

[157] 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. 10

[158] Jingkang Yang, Pengyun Wang, Dejian Zou, Zitang Zhou, Kunyuan Ding, Wenxuan Peng, Haoqi Wang, Guangyao Chen, Bo Li, Yiyou Sun, et al. Openood: Benchmarking generalized out-of-distribution detection. *NeurIPS*, 2022. 1, 3, 9

[159] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, 2024. 9

[160] Catherine Yeh, Yida Chen, Aoyu Wu, Cynthia Chen, Fernanda Viégas, and Martin Wattenberg. Attentionviz: A global view of transformer attention. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):262–272, 2023. 7

[161] Qing Yu and Kiyoharu Aizawa. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *CVPR*, 2019. 10

[162] Yeonguk Yu, Sungho Shin, Seongju Lee, Changhyun Jun, and Kyoobin Lee. Block selection method for using feature norm in out-of-distribution detection. In *CVPR*, 2023. 10

[163] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 10

[164] Bangcheng Zhan, Enmin Song, Hong Liu, Zhenyu Gong, Guangzhi Ma, and Chih-Cheng Hung. Cfnet: A medical image segmentation method using the multi-view attention mechanism and adaptive fusion strategy. *Biomedical Signal Processing and Control*, 79: 104112, 2023. 10

[165] Boxuan Zhang, Jianing Zhu, Zengmao Wang, Tongliang Liu, Bo Du, and Bo Han. What if the input is expanded in ood detection? *Advances in Neural Information Processing Systems*, 37:21289–21329, 2024. 10

[166] Chao Zhang, Zichao Yang, Xiaodong He, and Li Deng. Multimodal intelligence: Representation learning, information fusion, and applications. *IEEE Journal of Selected Topics in Signal Processing*, 14(3): 478–493, 2020. 10

[167] Jinsong Zhang, Qiang Fu, Xu Chen, Lun Du, Zelin Li, Gang Wang, Shi Han, Dongmei Zhang, et al. Out-of-distribution detection based on in-distribution data patterns memorization with modern hopfield energy. In *ICLR*, 2022. 10

[168] Jingyang Zhang, Jingkang Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Haoran Zhang, Yiyou Sun, Xuefeng Du, Kaiyang Zhou, Wayne Zhang, et al. Openood v1. 5: Enhanced benchmark for out-of-distribution detection. *arXiv*, 2023. 1, 2, 3, 9

[169] Le Zhang, Jian Sun, and Qiang Zheng. 3d point cloud recognition based on a multi-view convolutional neural network. *Sensors*, 18(11):3681, 2018. 10

[170] Yonggang Zhang, Jie Lu, Bo Peng, Zhen Fang, and Yiu-ming Cheung. Learning to shape in-distribution feature space for out-of-distribution detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 10

[171] Lijie Zhao, Yuan Chai, Qichun Zhang, and Hamid Reza Karimi. Self-supervised anomaly detection based on foreground enhancement and autoencoder reconstruction. *Signal, Image and Video Processing*, 18(1):343–350, 2024. 10

[172] Wenjie Zhao, Jia Li, Xin Dong, Yu Xiang, and Yunhui Guo. Segment every out-of-distribution object. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3910–3920, 2024. 10

[173] Xuran Zhao, Luyang Yu, and Xun Wang. Cross-view attention network for breast cancer screening from multi-view mammograms. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1050–1054. IEEE, 2020. 7

[174] Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. Cddfuse: Correlation-driven dual-branch

feature decomposition for multi-modality image fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5906–5916, 2023. 10

[175] Haotian Zheng, Qizhou Wang, Zhen Fang, Xiaobo Xia, Feng Liu, Tongliang Liu, and Bo Han. Out-of-distribution detection learning with unreliable out-of-distribution sources. *Advances in Neural Information Processing Systems*, 36, 2024. 10

[176] 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. 1

[177] Jingchun Zhou, Jiaming Sun, Weishi Zhang, and Zifan Lin. Multi-view underwater image enhancement method via embedded fusion mechanism. *Engineering Applications of Artificial Intelligence*, 121: 105946, 2023. 10

[178] Yibo Zhou. Rethinking reconstruction autoencoder-based out-of-distribution detection. In *CVPR*, 2022. 10

[179] Tong Zhu, Leida Li, Jufeng Yang, Sicheng Zhao, Hantao Liu, and Jiansheng Qian. Multimodal sentiment analysis with image-text interaction network. *IEEE transactions on multimedia*, 25:3375–3385, 2022. 7, 10

[180] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018. 10

[181] Shu Zou, Xinyu Tian, Qinyu Zhao, Zhaoyuan Yang, and Jing Zhang. Simlabel: Consistency-guided ood detection with pretrained vision-language models. *arXiv*, 2025. 10