A. Approach

A.1. Code snippet

```
clip.eval() # CLIP model from open_clip
probe.eval() # fully connected layer trained to classify (un-normalized) CLIP image features
classifier.eval() # e.g. Resnet50 trained to classify images

# note: different normalization for CLIP image encoder vs. standard classifier
image_normalized_clip = normalize_for_clip(image)
image_normalized_cls = normalize_for_cls(image)

# 1. get zero-shot CLIP prediction
prompt_template = "a photo of a [cls]"
prompts = [prompt_template.replace("[cls]",f"{class_mapping[idx]}") for idx in range(num_id_classes)]

prompt_tokenized = clip_tokenizer(prompts)
prompt_features = clip.encode_text(prompt_tokenized)
prompt_features_normed = prompt_features / prompt_features.norm(dim=-1, keepdim=True)

clip_image_features = clip.encode_image(image_normalized_clip)
clip_image_features_normed = image_features / image_features.norm(dim=-1, keepdim=True)
text_sim = (image_features_normed @ prompt_features_normed.T)

logits_clip_t100 = 100 * text_sim
softmax_clip_t100 = logits_clip_t100.softmax(dim=1)

# 2. get probe CLIP prediction
logits_probe = probe(clip_image_features)
softmax_probe = logits_probe.softmax(dim=1)

# 3. get classifier prediction
logits_classifier = classifier(image_normalized_cls)
softmax_classifier = logits_classifier.softmax(dim=1)

# 4. combined prediction
softmax_ensemble = torch.stack([softmax_clip_t100, softmax_probe, softmax_classifier]).mean(0)

# class prediction and OOD scores
pred = softmax_ensemble.argmax(dim=1)

msp = softmax_ensemble.max(dim=1)
entropy = torch.distributions.Categorical(probs=softmax_ensemble).entropy()
```

B. Experimental set-up

B.1. Datasets

B.1.1. OpenOOD

For a given ID dataset (e.g. CIFAR100), OpenOOD [59] defines a ID_{train}, ID_{val}, ID_{test} sets and multiple OOD_{test} sets, and we use these without modification. ID_{train} is used for training the classifier and certain OOD detectors (e.g. KNN, RMDS) use it to compute dataset statistics. ID_{val} is used for early stopping and for OOD hyperparameter tuning by certain methods (e.g. TempScale). ID_{test} and OOD_{test} are used for evaluating classification and OOD detection performance.

ID datasets:

- CIFAR100 needs no introduction. ID_{train} is the official training set. ID_{val} (1000 images) and ID_{test} (9000 images) are subsets of the official test set.
- CIFAR100-N [53] contains the same images as CIFAR100 but provides a re-annotated version of CIFAR100's ID_{train} set with real label noise ($\approx 40\%$ noise rate arising from crowdsourced annotations). The ID_{val} and ID_{test} are cleanly labelled and identical to those in CIFAR100. Note that this dataset is not part of the OpenOOD benchmark/leaderboard but we apply identical splits and pre-processing.

- ImageNet-1K corresponds to the ILSVRC2012 version of ImageNet. ID_{train} is the official training set. ID_{val} (5000 images) and ID_{test} (45000 images) are subsets of the official validation set.
- ImageNet-200 is a 200-class subset of ImageNet-1K defined in [20]. ID_{train} is a subset of the official ImageNet-1K training set. ID_{val} (1000 images) and ID_{test} (9000 images) are subsets of the official ImageNet-1K validation set.
- In the covariate shifted versions of ImageNet-1K and ImageNet-200, the original ID_{test} is concatenated with images from ImageNet-R [21], ImageNet-V2 [44] and ImageNet-C [18]. We use the exact same sets as those defined in the OpenOOD full-spectrum benchmark.

Note that our method does not make use of ID_{val}, as we use the last training checkpoint for evaluation, and involves no OOD hyper-parameters.

B.1.2. OODDB

OODDB [1] provides 5 open-set splits for several datasets. We use first split (split 0) for PatternNet and DTD which defines the following set of ID vs OOD classes:

PatternNet:
- ID: 0: 'oil well', 1: 'overpass', 2: 'railway', 3: 'basketball court', 4: 'river', 5: 'wastewater treatment plant', 6: 'christmas tree farm', 7: 'sparse residential', 8: 'chaparral', 9: 'solar panel', 10: 'parking lot', 11: 'airplane', 12: 'golf course', 13: 'bridge', 14: 'freeway', 15: 'transformer station', 16: 'mobile home park', 17: 'nursing home', 18: 'beach'
- OOD: 19: 'ferry terminal', 20: 'storage tank', 21: 'forest', 22: 'coastal mansion', 23: 'swimming pool', 24: 'closed road', 25: 'shipping yard', 26: 'dense residential', 27: 'runway', 28: 'tennis court', 29: 'crosswalk', 30: 'intersection', 31: 'runway marking', 32: 'cemetery', 33: 'baseball field', 34: 'oil gas field', 35: 'parking space', 36: 'football field', 37: 'harbor'

DTD:
- ID: 0: 'crystalline', 1: 'fibrous', 2: 'spiralled', 3: 'stratified', 4: 'matted', 5: 'porous', 6: 'scaly', 7: 'stained', 8: 'pleated', 9: 'flecked', 10: 'pitted', 11: 'meshed', 12: 'freckled', 13: 'waffled', 14: 'cracked', 15: 'potholed', 16: 'cobwebbed', 17: 'swirly', 18: 'polka-dotted', 19: 'striped', 20: 'studded', 21: 'grooved', 22: 'woven'
- OOD: 23: 'banded', 24: 'blotchy', 25: 'braided', 26: 'bubbly', 27: 'bumpy', 28: 'chequered', 29: 'crosshatched', 30: 'dotted', 31: 'frilly', 32: 'gauzy', 33: 'grid', 34: 'honeycombed', 35: 'interlaced', 36: 'knitted', 37: 'lacelike', 38: 'lined', 39: 'marbled', 40: 'paisley', 41: 'perforated', 42: 'smeared', 43: 'sprinkled', 44: 'veined', 45: 'wrinkled', 46: 'zigzagged'

B.2. Models

B.2.1. Standard classifiers

For standard ImageNet1K classifiers, we use pretrained models from torchvision. Specifically: ResNet18: `resnet18-f37072fd.pth`, ResNet50: `resnet50-0676ba61.pth`, ViT-B-16: `vit_b_16-c867db91.pth`.

For the other ID datasets (CIFAR100, PatternNet and DTD) we train a ResNet18, ResNet50 and ViT-B-16 with the following optimizer and scheduler:

```
optimizer = torch.optim.SGD(
    net.parameters(), lr, momentum=0.9, weight_decay=0.0005, nesterov=True,
)

def cosine_annealing(step, total_steps, lr_max, lr_min):
    return lr_min + (lr_max - lr_min) * 0.5 * (1 + np.cos(step / total_steps * np.pi))

scheduler = torch.optim.lr_scheduler.LambdaLR(
    self.optimizer, lr_lambda=lambda step: cosine_annealing( step, num_epochs * len(train_loader), 1,
                                                              1e-6 / lr )
)
```

for 100 epochs with a learning rate of 0.1 when training from scratch and 0.01 when fine-tuning from a torchvision checkpoint. The batch size is 128 for CIFAR datasets otherwise 256.

B.2.2. Zero-shot CLIP and Probe CLIP

We use the following CLIP models in our experiments:

```
    open_clip.create_model_and_transforms("ViT-B-32", pretrained="openai")
    open_clip.create_model_and_transforms("ViT-B-16", pretrained="openai")
    open_clip.create_model_and_transforms("ViT-L-14", pretrained="openai")
    open_clip.create_model_and_transforms("ViT-H-14", pretrained="laion2b_s32b_b79k")
```

The linear probe is implemented as a single linear/fully connected layer with bias, taking as input the un-normalized image embeddings extracted by CLIP (pre-extracted before training). It is trained for 20 epochs with a learning rate of 0.1, and the same optimizer, scheduler and batch size as for the standard classifier (see above).

B.3. Image pre-processing

To summarize (see below for more details): an input size of 224×224 is used for all datasets. At test-time, images are resized such that the shortest side is 224 and then center-cropped to 224×224.

- Classifier model: input images are normalized based on dataset-specific mean and std. During training, only basic data augmentations from OpenOOD are applied (resize, flip, random crop for CIFAR100, and flip, random resized crop for the rest).
- CLIP models: input images are normalized based on the OpenAI mean and std. The linear classifier is trained on un-normalized CLIP image embeddings which were extracted and stored before training. Thus, no augmentations are applied.

On the standard classifier side, we use the same pre-processing procedure as the OpenOOD codebase. On the CLIP side, we align with previous work on CLIP-based OOD detection.

B.3.1. Image normalization

For standard classifiers, the following normalization paramaters are used:

- CIFAR100: `mean=(0.5071, 0.4867, 0.4408), std=(0.2675, 0.2565, 0.2761)`
- ImageNet-200 and ImageNet-1k: `mean=(0.485, 0.456, 0.406), (0.229, 0.224, 0.225)`
- PatternNet: `mean=(0.5, 0.5, 0.5), std=(0.5, 0.5, 0.5)`
- DTD: `mean=(0.5, 0.5, 0.5), std=(0.5, 0.5, 0.5)`

For CLIP (zero-shot and probe), the same normalization parameters are used across all datasets: `mean=(0.48145466, 0.4578275, 0.40821073), std=(0.26862954, 0.26130258, 0.27577711)`

B.4. Image transformations

For standard classifiers, the following image transformations are used:

```
# training CIFAR100
tvs_trans.Compose([
        tvs_trans.Resize(224),
        tvs_trans.RandomHorizontalFlip(),
        tvs_trans.RandomCrop(image_size, padding=4),
        tvs_trans.ToTensor(),
        normalize_cls
    ])

# training other datasets
tvs_trans.Compose([
        tvs_trans.RandomHorizontalFlip(),
        tvs_trans.RandomResizedCrop(image_size),
        tvs_trans.ToTensor(),
        normalize_cls
    ])

# evaluation
tvs_trans.Compose([
        tvs_trans.Resize(224),
        tvs_trans.CenterCrop(224),
        tvs_trans.ToTensor(),
        normalize_cls
    ])
```

For CLIP (zero-shot and probe), the following image transformations are used:

```
tvs_trans.Compose([
        tvs_trans.Resize(224),
        tvs_trans.CenterCrop(224),
        tvs_trans.ToTensor(),
        normalize_clip
    ])
```

B.5. Class names

Class names are used when constructing prompts for zero-shot CLIP.

ImageNet-1K and ImageNet-200 Similarly to existing work in zero-shot CLIP OOD detection [25, 31, 36, 38], we use the ImageNet class names provided in the OpenAI CLIP repository, which are "prompt-engineered" and deviate from the raw class names. For example, for class 884, the class name "vaulted or arched ceiling" is used instead of the raw/original class name "vault". Note that using raw vs. prompt-engineered class names for constructing CLIP text prompts has a non-negligible impact on zero-shot performance.

CIFAR100 We use the raw class names, replacing underscores with spaces.

OODDB PatternNet & OODDB DTD We use the raw class names, exactly as listed above.

B.6. Evaluation metrics

- The classification accuracy is computed on $\mathrm{ID}_{\mathrm{test}}$.
- The OOD AUROC and OOD FPR@95 are computed on the ID test set vs. a specific OOD dataset. ID samples are treated as the positive label.

C. Baselines

C.1. CLIP-based approaches

When evaluating previous CLIP-based methods, we used their official implementation for the model, temperature, prompt template and image pre-processing but replaced the data splits with those described in 5.1. ZOC [10] generates candidate negative prompts using a trained BERT encoder which can be used across multiple ID datasets, therefore we use the model weights provided by the authors. NegLabel [25] generates candidate negative prompts based on the ID class names - we therefore obtain a different set of negative prompts for each ID dataset.

C.2. Classifier-based approaches

All methods in our comparison except NAC [33], Weiper-KLD [15] and CombOOD [42] are integrated in the OpenOOD codebase, and we apply them with their default configuration. For Weiper and NAC, we use the official implementation `https://github.com/mgranz/weiper` and `https://github.com/BierOne/ood_coverage/` with the suggested configuration provided for each architecture. We had issues reproducing CombOOD results from its official implementation `https://github.com/rmagesh148/combood/` and therefore did not include it in the final experiment.

D. Results

D.1. COOkeD in action

Figures 6 (classification) and 6 (OOD detection) extend Figure 4 from the main text by showing performance on a per-dataset basis. Figure 6 also provides results in terms of FPR95 and with the Entropy vs. MSP as scoring function.

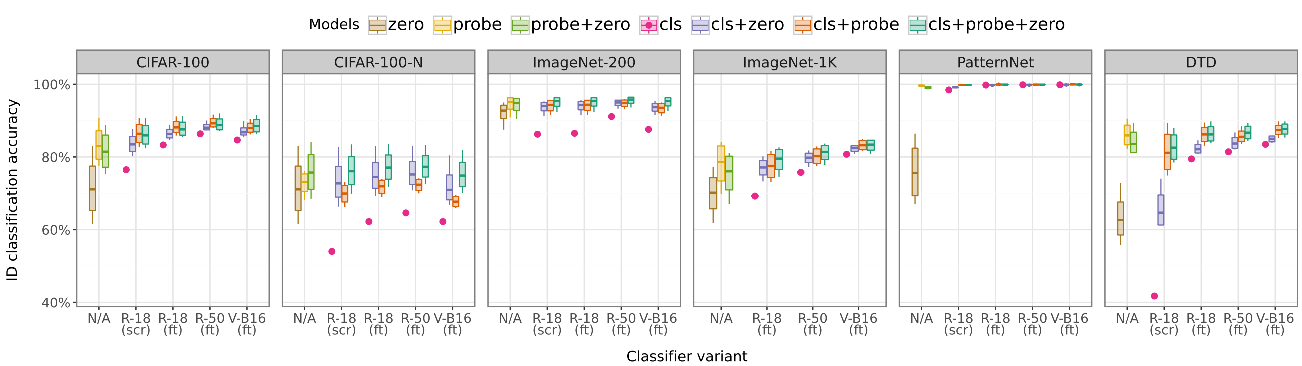
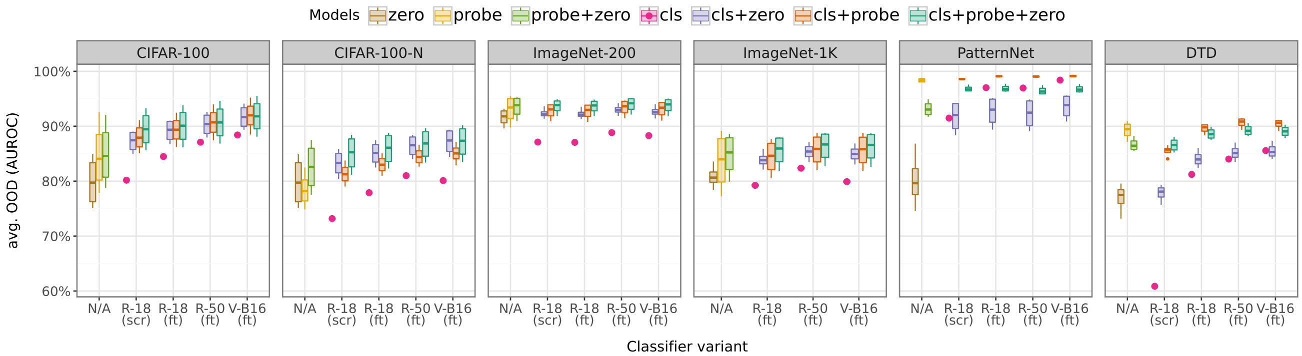
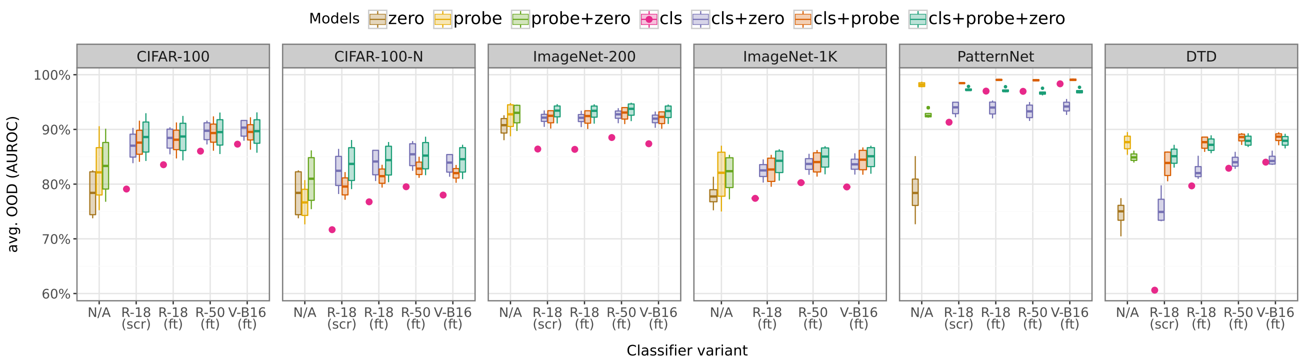


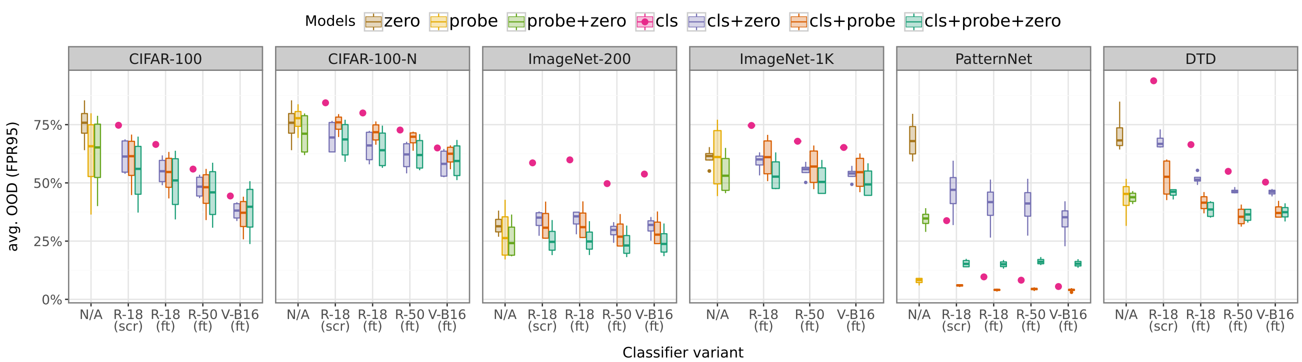
Figure 6. Classification accuracy with different combinations of Zero-shot / Probe / Standard classifier models. Boxplots show the distribution across different CLIP and classifier variants.



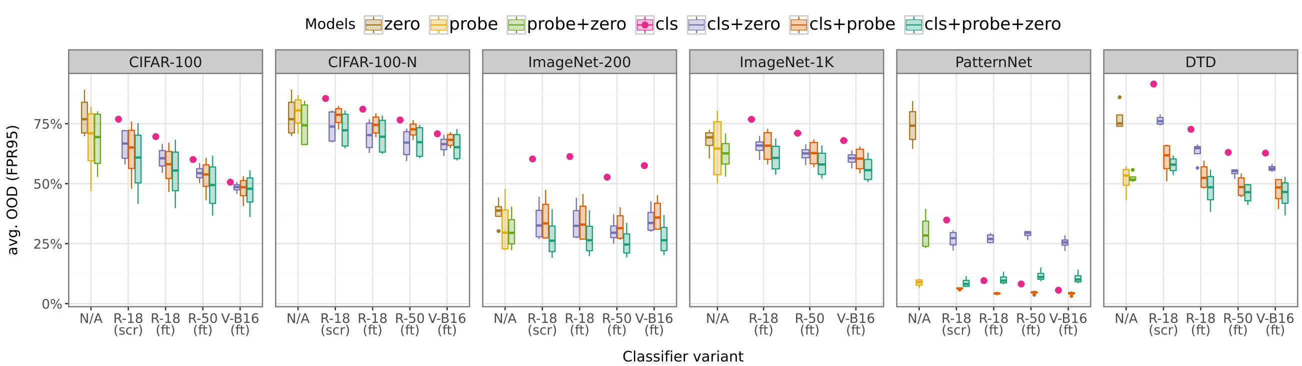
(a) AUROC - Entropy as OOD scoring function



(b) AUROC - MSP as OOD scoring function



(c) FPR95 - Entropy as OOD scoring function



(d) FPR95 - MSP as OOD scoring function

Figure 7. OOD detection performance with different combinations of Zero-shot / Probe / Standard classifier models. Boxplots show the distribution across different CLIP and classifier variants.

D.2. Comparison to classifier-based OOD detection

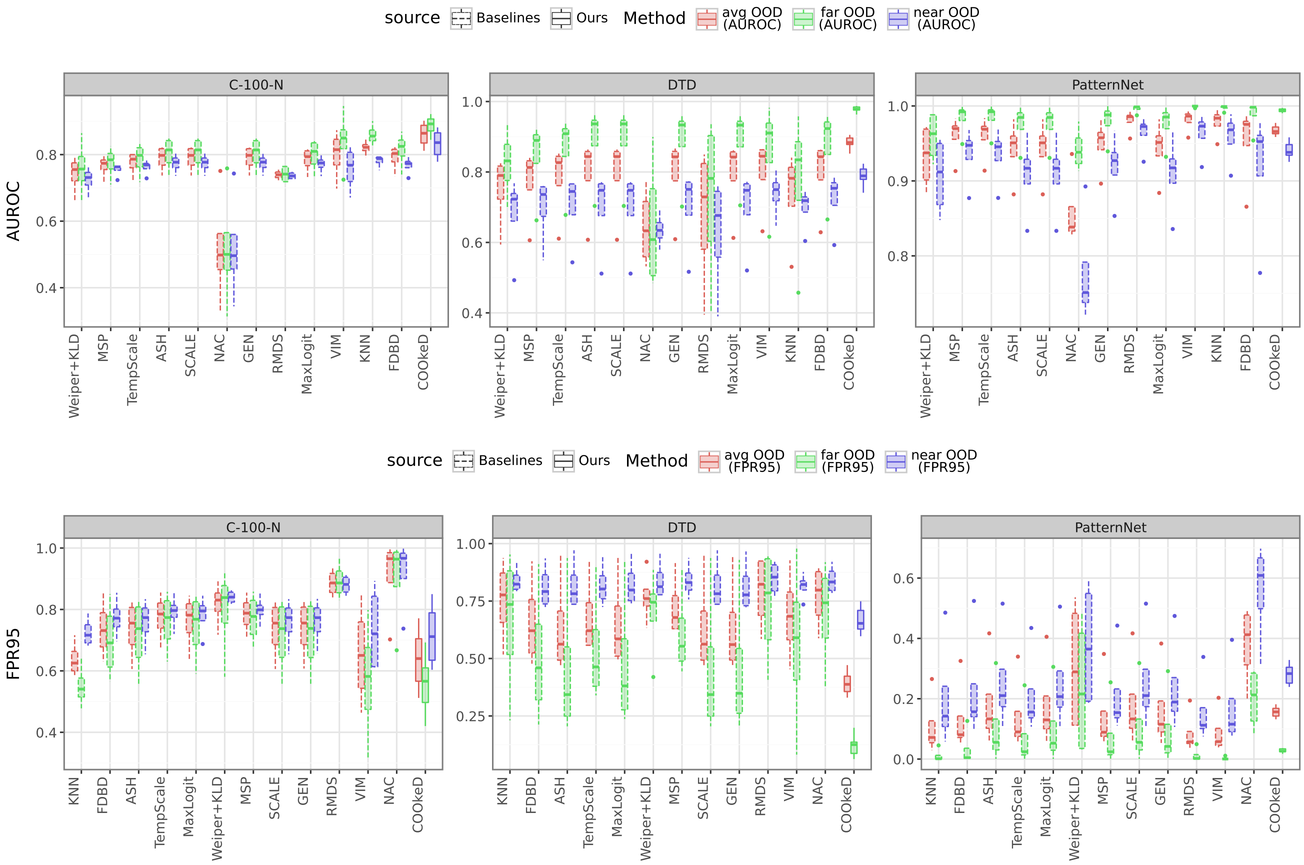


Figure 8. OOD detection performance (AUROC and FPR95) beyond the OpenOOD benchmark. Boxplot distribution is across 4 classifier variants and 4 CLIP variants (for COOkeD only).

D.3. Results for each OOD dataset

The four tables below contain the "raw" results for each ID/OOD dataset pair, classifier variant and CLIP variant, using COOkeD (ensemble of standard classifier, zero-shot CLIP and probe CLIP, with the entropy as OOD score.) All performance is in %. Besides the ID accuracy, the OOD detection performance for each OOD dataset corresponds to the AUROC.

Dataset	Classifier	CLIP	ID Acc.	SVHN	Texture	TIN	Places365	MNIST
CIFAR-100	ResNet-18 (scr)	ViT-B-32	82.38	96.06	82.25	84.74	83.98	89.28
CIFAR-100	ResNet-18 (scr)	ViT-B-16	83.94	95.39	84.35	86.97	84.80	93.48
CIFAR-100	ResNet-18 (scr)	ViT-L-14	87.92	97.74	88.59	90.92	89.99	95.05
CIFAR-100	ResNet-18 (scr)	ViT-H-14	90.68	98.64	91.40	92.50	93.42	94.70
CIFAR-100	ResNet-18 (ft)	ViT-B-32	85.50	96.64	88.24	86.61	81.77	83.78
CIFAR-100	ResNet-18 (ft)	ViT-B-16	86.16	96.22	89.78	88.72	82.77	89.84
CIFAR-100	ResNet-18 (ft)	ViT-L-14	89.01	98.05	92.98	92.49	88.43	92.08
CIFAR-100	ResNet-18 (ft)	ViT-H-14	91.26	98.74	95.27	93.98	92.30	91.06
CIFAR-100	ResNet-50 (ft)	ViT-B-32	87.14	97.41	88.01	87.91	83.60	85.70
CIFAR-100	ResNet-50 (ft)	ViT-B-16	87.57	97.04	89.48	89.83	84.49	91.34
CIFAR-100	ResNet-50 (ft)	ViT-L-14	89.94	98.56	92.93	93.46	89.90	93.18
CIFAR-100	ResNet-50 (ft)	ViT-H-14	91.98	98.99	95.66	95.03	93.86	92.50
CIFAR-100	ViT-B16 (ft)	ViT-B-32	86.23	97.47	93.06	89.38	90.28	83.85
CIFAR-100	ViT-B16 (ft)	ViT-B-16	87.12	96.89	93.80	91.17	90.77	90.53
CIFAR-100	ViT-B16 (ft)	ViT-L-14	89.92	98.52	95.90	94.46	94.33	92.47
CIFAR-100	ViT-B16 (ft)	ViT-H-14	91.58	99.11	97.75	96.11	96.99	92.12

Dataset	Classifier	CLIP	ID Acc.	SVHN	Texture	TIN	Places365	MNIST
CIFAR-100-N	ResNet-18 (scr)	ViT-B-32	69.89	96.09	77.18	79.86	76.59	87.89
CIFAR-100-N	ResNet-18 (scr)	ViT-B-16	73.19	94.82	79.97	82.29	76.97	92.23
CIFAR-100-N	ResNet-18 (scr)	ViT-L-14	78.98	97.44	84.19	87.26	82.89	92.96
CIFAR-100-N	ResNet-18 (scr)	ViT-H-14	83.46	98.38	85.32	87.40	86.03	92.09
CIFAR-100-N	ResNet-18 (ft)	ViT-B-32	71.74	96.82	82.54	81.37	75.50	87.17
CIFAR-100-N	ResNet-18 (ft)	ViT-B-16	74.60	95.78	84.63	83.58	75.86	91.34
CIFAR-100-N	ResNet-18 (ft)	ViT-L-14	79.56	97.93	88.07	88.24	81.70	92.29
CIFAR-100-N	ResNet-18 (ft)	ViT-H-14	83.52	98.57	88.74	88.19	84.59	90.26
CIFAR-100-N	ResNet-50 (ft)	ViT-B-32	72.54	97.93	82.38	82.70	77.66	90.78
CIFAR-100-N	ResNet-50 (ft)	ViT-B-16	75.11	96.96	84.47	84.63	78.03	93.98
CIFAR-100-N	ResNet-50 (ft)	ViT-L-14	79.48	98.49	88.15	88.96	83.44	94.57
CIFAR-100-N	ResNet-50 (ft)	ViT-H-14	83.34	98.93	89.34	89.11	86.49	93.17
CIFAR-100-N	ViT-B16 (ft)	ViT-B-32	70.14	97.44	89.07	82.66	81.02	82.57
CIFAR-100-N	ViT-B16 (ft)	ViT-B-16	72.37	96.62	90.46	84.79	81.26	87.95
CIFAR-100-N	ViT-B16 (ft)	ViT-L-14	77.39	98.25	92.92	89.28	86.50	89.22
CIFAR-100-N	ViT-B16 (ft)	ViT-H-14	81.98	98.85	94.18	89.52	89.62	86.04

Table 6. COOkeD performance on CIFAR100 (clean training labels vs noisy training labels setting).

Dataset	Classifier	CLIP	ID Acc.	DTD_{OOD}	SVHN	$PatternNet_{OOD}$	TIN	Places365	MNIST
DTD	ResNet-18 (scr)	ViT-B-32	78.48	74.11	99.53	94.39	94.45	94.85	97.88
DTD	ResNet-18 (scr)	ViT-B-16	79.67	75.08	99.34	92.17	96.27	94.28	99.68
DTD	ResNet-18 (scr)	ViT-L-14	85.43	77.29	99.62	93.98	96.66	97.06	97.51
DTD	ResNet-18 (scr)	ViT-H-14	87.93	78.45	99.88	96.98	96.18	96.94	99.22
DTD	ResNet-18 (ft)	ViT-B-32	83.91	77.50	99.82	96.32	96.77	95.80	98.04
DTD	ResNet-18 (ft)	ViT-B-16	84.57	78.31	99.76	94.46	97.54	95.08	99.64
DTD	ResNet-18 (ft)	ViT-L-14	87.83	80.03	99.71	95.79	98.43	97.91	98.08
DTD	ResNet-18 (ft)	ViT-H-14	89.78	81.10	99.93	97.88	98.36	97.78	99.45
DTD	ResNet-50 (ft)	ViT-B-32	84.35	78.58	99.85	96.66	96.84	96.12	98.56
DTD	ResNet-50 (ft)	ViT-B-16	85.22	79.40	99.83	94.91	97.77	95.43	99.90
DTD	ResNet-50 (ft)	ViT-L-14	88.15	80.93	99.85	96.30	98.63	98.22	98.34
DTD	ResNet-50 (ft)	ViT-H-14	89.35	82.34	99.89	97.96	98.29	97.93	99.35
DTD	ViT-B16 (ft)	ViT-B-32	85.33	78.04	99.90	96.24	97.46	96.74	98.13
DTD	ViT-B16 (ft)	ViT-B-16	86.63	79.19	99.89	94.42	98.16	96.11	99.77
DTD	ViT-B16 (ft)	ViT-L-14	88.80	80.79	99.84	95.83	98.73	98.49	98.18
DTD	ViT-B16 (ft)	ViT-H-14	89.78	81.84	99.93	97.64	98.50	98.18	99.55

Dataset	Classifier	CLIP	ID Acc.	$PatternNet_{OOD}$	SVHN	DTD_{OOD}	TIN	Places365	MNIST
PatternNet	ResNet-18 (scr)	ViT-B-32	99.68	93.52	99.96	98.77	99.43	97.91	99.98
PatternNet	ResNet-18 (scr)	ViT-B-16	99.71	93.51	99.99	99.05	99.55	98.58	99.99
PatternNet	ResNet-18 (scr)	ViT-L-14	99.89	95.56	100.00	99.13	99.75	99.13	100.00
PatternNet	ResNet-18 (scr)	ViT-H-14	99.89	94.32	100.00	98.91	99.79	99.32	100.00
PatternNet	ResNet-18 (ft)	ViT-B-32	99.91	93.56	99.98	99.29	99.57	97.78	100.00
PatternNet	ResNet-18 (ft)	ViT-B-16	99.86	93.64	100.00	99.29	99.63	98.70	100.00
PatternNet	ResNet-18 (ft)	ViT-L-14	99.89	95.81	99.99	99.63	99.81	99.16	99.99
PatternNet	ResNet-18 (ft)	ViT-H-14	99.93	94.70	100.00	99.52	99.83	99.36	100.00
PatternNet	ResNet-50 (ft)	ViT-B-32	99.89	92.68	99.98	98.74	99.58	98.38	99.98
PatternNet	ResNet-50 (ft)	ViT-B-16	99.86	92.54	99.99	98.96	99.65	98.97	99.97
PatternNet	ResNet-50 (ft)	ViT-L-14	99.88	95.29	100.00	99.55	99.83	99.39	99.98
PatternNet	ResNet-50 (ft)	ViT-H-14	99.96	93.98	100.00	99.16	99.82	99.44	99.99
PatternNet	ViT-B16 (ft)	ViT-B-32	99.92	93.06	99.98	99.42	99.76	98.39	100.00
PatternNet	ViT-B16 (ft)	ViT-B-16	99.87	93.23	100.00	99.36	99.80	98.92	100.00
PatternNet	ViT-B16 (ft)	ViT-L-14	99.92	95.62	100.00	99.73	99.90	99.39	100.00
PatternNet	ViT-B16 (ft)	ViT-H-14	99.95	94.51	100.00	99.61	99.91	99.39	100.00

Table 7. COOkeD performance on DTD and PatternNet.

Dataset	Classifier	CLIP	ID Acc.	iNaturalist	Textures	OpenImage-O	NINCO	SSB-Hard
IN-200	ResNet-18 (scr)	ViT-B-32	92.38	98.36	95.58	96.30	90.60	83.35
IN-200	ResNet-18 (scr)	ViT-B-16	94.47	98.90	96.35	97.37	91.98	85.68
IN-200	ResNet-18 (scr)	ViT-L-14	96.33	99.24	97.32	98.41	94.45	88.52
IN-200	ResNet-18 (scr)	ViT-H-14	96.28	99.00	97.36	97.96	93.97	87.79
IN-200	ResNet-18 (ft)	ViT-B-32	92.42	98.30	95.64	96.29	90.48	83.28
IN-200	ResNet-18 (ft)	ViT-B-16	94.52	98.85	96.35	97.35	91.85	85.63
IN-200	ResNet-18 (ft)	ViT-L-14	96.28	99.21	97.30	98.38	94.36	88.47
IN-200	ResNet-18 (ft)	ViT-H-14	96.33	98.95	97.36	97.94	93.86	87.74
IN-200	ResNet-50 (ft)	ViT-B-32	93.61	98.51	95.99	96.62	91.06	83.34
IN-200	ResNet-50 (ft)	ViT-B-16	95.17	98.92	96.67	97.57	92.38	85.80
IN-200	ResNet-50 (ft)	ViT-L-14	96.57	99.27	97.60	98.52	94.77	88.97
IN-200	ResNet-50 (ft)	ViT-H-14	96.41	99.04	97.79	98.21	94.64	88.47
IN-200	ViT-B16 (ft)	ViT-B-32	92.68	98.37	96.04	96.40	90.61	82.75
IN-200	ViT-B16 (ft)	ViT-B-16	94.49	98.87	96.76	97.46	92.05	85.35
IN-200	ViT-B16 (ft)	ViT-L-14	96.26	99.26	97.69	98.49	94.64	88.54
IN-200	ViT-B16 (ft)	ViT-H-14	96.40	99.03	97.83	98.13	94.37	87.96

Dataset	Classifier	CLIP	ID Acc.	iNaturalist	Textures	OpenImage-O	NINCO	SSB-Hard
IN-200 (CS)	ResNet-18 (scr)	ViT-B-32	71.80	83.65	77.85	76.19	61.08	52.20
IN-200 (CS)	ResNet-18 (scr)	ViT-B-16	78.08	88.34	80.78	80.96	64.57	56.41
IN-200 (CS)	ResNet-18 (scr)	ViT-L-14	88.07	93.07	87.32	88.54	75.38	65.78
IN-200 (CS)	ResNet-18 (scr)	ViT-H-14	89.64	90.46	86.80	85.33	72.10	62.71
IN-200 (CS)	ResNet-18 (ft)	ViT-B-32	72.03	83.30	78.19	76.22	60.86	52.38
IN-200 (CS)	ResNet-18 (ft)	ViT-B-16	78.24	88.00	81.00	80.95	64.36	56.56
IN-200 (CS)	ResNet-18 (ft)	ViT-L-14	88.00	92.84	87.46	88.49	75.15	65.85
IN-200 (CS)	ResNet-18 (ft)	ViT-H-14	89.64	90.17	86.98	85.33	71.83	62.84
IN-200 (CS)	ResNet-50 (ft)	ViT-B-32	72.35	82.92	77.60	76.25	61.46	52.22
IN-200 (CS)	ResNet-50 (ft)	ViT-B-16	78.09	87.54	80.52	80.81	64.81	56.29
IN-200 (CS)	ResNet-50 (ft)	ViT-L-14	87.87	92.51	87.16	88.27	75.29	65.48
IN-200 (CS)	ResNet-50 (ft)	ViT-H-14	89.47	89.74	86.70	85.19	72.23	62.51
IN-200 (CS)	ViT-B16 (ft)	ViT-B-32	70.53	82.85	79.37	76.22	60.72	51.31
IN-200 (CS)	ViT-B16 (ft)	ViT-B-16	76.80	87.63	82.18	80.97	64.23	55.58
IN-200 (CS)	ViT-B16 (ft)	ViT-L-14	87.22	92.65	88.51	88.55	75.27	65.20
IN-200 (CS)	ViT-B16 (ft)	ViT-H-14	89.15	89.92	88.02	85.41	72.01	62.05

Table 8. COOkeD performance on ImageNet-200 (standard and full-spectrum evaluation).

Dataset	Classifier	CLIP	ID Acc.	iNaturalist	Textures	OpenImage-O	NINCO	SSB-Hard
IN-1K	ResNet-18 (ft)	ViT-B-32	74.56	94.63	88.43	90.76	80.12	64.91
IN-1K	ResNet-18 (ft)	ViT-B-16	77.26	95.97	90.16	92.88	82.10	68.23
IN-1K	ResNet-18 (ft)	ViT-L-14	81.83	97.64	91.62	95.62	87.67	73.61
IN-1K	ResNet-18 (ft)	ViT-H-14	82.63	96.84	92.38	94.94	88.28	74.21
IN-1K	ResNet-50 (ft)	ViT-B-32	77.85	95.05	88.93	91.69	81.04	66.21
IN-1K	ResNet-50 (ft)	ViT-B-16	79.76	96.24	90.59	93.61	82.95	69.62
IN-1K	ResNet-50 (ft)	ViT-L-14	82.99	97.76	91.91	96.01	88.37	75.09
IN-1K	ResNet-50 (ft)	ViT-H-14	83.68	97.01	92.76	95.45	89.15	75.96
IN-1K	ViT-B16 (ft)	ViT-B-32	80.85	95.08	89.43	91.83	80.82	65.36
IN-1K	ViT-B16 (ft)	ViT-B-16	82.21	96.20	91.00	93.68	82.85	68.97
IN-1K	ViT-B16 (ft)	ViT-L-14	84.57	97.71	92.19	96.03	88.42	74.69
IN-1K	ViT-B16 (ft)	ViT-H-14	84.73	97.00	93.07	95.44	89.27	75.38

Dataset	Classifier	CLIP	ID Acc.	iNaturalist	Textures	OpenImage-O	NINCO	SSB-Hard
IN-1K (CS)	ResNet-18 (ft)	ViT-B-32	60.71	83.25	74.21	76.53	61.91	47.19
IN-1K (CS)	ResNet-18 (ft)	ViT-B-16	64.67	85.80	76.64	79.75	63.64	49.72
IN-1K (CS)	ResNet-18 (ft)	ViT-L-14	73.73	91.27	80.46	86.48	72.34	56.24
IN-1K (CS)	ResNet-18 (ft)	ViT-H-14	76.41	88.36	81.26	84.25	72.22	56.58
IN-1K (CS)	ResNet-50 (ft)	ViT-B-32	63.03	82.96	74.05	77.20	62.32	48.20
IN-1K (CS)	ResNet-50 (ft)	ViT-B-16	66.17	85.41	76.41	80.27	63.96	50.82
IN-1K (CS)	ResNet-50 (ft)	ViT-L-14	74.33	90.86	80.22	86.68	72.52	57.40
IN-1K (CS)	ResNet-50 (ft)	ViT-H-14	76.93	87.90	81.08	84.54	72.50	57.94
IN-1K (CS)	ViT-B16 (ft)	ViT-B-32	66.73	83.52	75.68	78.34	63.23	48.21
IN-1K (CS)	ViT-B16 (ft)	ViT-B-16	69.30	85.87	77.93	81.29	64.91	50.97
IN-1K (CS)	ViT-B16 (ft)	ViT-L-14	76.08	91.09	81.41	87.34	73.47	57.83
IN-1K (CS)	ViT-B16 (ft)	ViT-H-14	78.16	88.30	82.28	85.30	73.73	58.38

Table 9. COOkeD performance on ImageNet-1K (standard and full-spectrum evaluation).