A DATASET DETAILS

We follow the exact experimental protocol of Du et al. (2024). The ID datasets are CIFAR-100 and ImageNet-100, which we briefly describe below:

CIFAR-100 (Krizhevsky et al., 2009) contains 50’000 training images and 10’000 testing images belonging to 100 classes.

ImageNet-100 is a subset of the full ImageNet (Deng et al., 2009) dataset. We take the 100 classes sampled by Du et al. (2024) for a total of 129’860 training samples and 5’000 test samples. These classes are: n01498041, n01514859, n01582220, n01608432, n01616318, n01687978, n01776313, n01806567, n01833805, n01882714, n01910747, n01944390, n01985128, n02007558, n02071294, n02085620, n02114855, n02123045, n02128385, n02129165, n02129604, n02165456, n02190166, n02219486, n02226429, n02279972, n02317335, n02326432, n02342885, n02363005, n02391049, n02395406, n02403003, n02422699, n02442845, n02444819, n02480855, n02510455, n02640242, n02672831, n02687172, n02701002, n02730930, n02769748, n02782093, n02787622, n02793495, n02799071, n02802426, n02814860, n02840245, n02906734, n02948072, n02980441, n02999410, n03014705, n03028079, n03032252, n03125729, n03160309, n03179701, n03220513, n03249569, n03291819, n03384352, n03388043, n03450230, n03481172, n03594734, n03594945, n03627232, n03642806, n03649909, n03661043, n03676483, n03724870, n03733281, n03759954, n03761084, n03773504, n03804744, n03916031, n03938244, n04004767, n04026417, n04090263, n04133789, n04153751, n04296562, n04330267, n04371774, n04404412, n04465501, n04485082, n04507155, n04536866, n04579432, n04606251, n07714990, n07745940.

For CIFAR-100, we use the test sets of five different datasets as OOD:

SVHN (Netzer et al., 2011) containing 10’000 images of house numbers.

Places365 (Zhou et al., 2017) is a dataset of large scenes, where we use the 10’000 random images sampled and provided by Sun et al. (2022); Tao et al. (2023).

Lsun (Yu et al., 2015) is a large-scale dataset of scenes and objects. We use the subset of 10’000 images provided by (Sun et al., 2022; Tao et al., 2023).

iSun (Xu et al., 2015) contains images of natural scenes, where we use the subset of 10’000 images provided by Sun et al. (2022); Tao et al. (2023).

Textures (Cimpoi et al., 2014) has 5’640 images of patterns and textures.

For ImageNet-100, we use four datasets where the classes of the test sets do not overlap with the full Imagenet, as provided by Huang & Li (2021):

iNaturalist (Van Horn et al., 2018) has images of plants and animals. We use a 10’000 image subset of 110 plant classes not present in ImageNet (Du et al., 2024).

SUN (Xiao et al., 2010) contains images of natural scenes, where we use a 10’000 image subset of 50 natural objects not present in Imagenet (Du et al., 2024).

Places (Zhou et al., 2017) is a dataset of large scenes, where we use 10’000 images from 50 categories that are not present in Imagenet (Du et al., 2024).

Textures (Cimpoi et al., 2014) has 5’640 images of patterns and textures.

B TOY EXAMPLE

We show how a cVPN works on a toy example in Fig. 6.

C FURTHER ABLATIONS

Effect of β. We ablate β with CIFAR100 as the ID. For $\beta \in [0.5, 1, 1.5, 2, 2.5]$, the AUCs are 95.3, 97.5, 97.7, 97.8, and 98.3, respectively, showcasing the robustness of NCIS to the parameter. The corresponding classification accuracies are 78.82, 79.22, 79.26, 78.62, and 78.63.

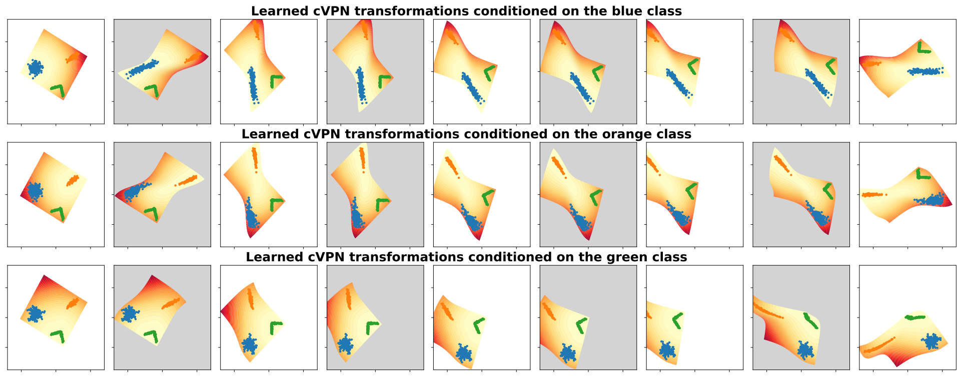

Figure 6: **The class-specific representations learned by the cVPN on toy data with three classes.** Depending on the conditioning, the cVPN transforms the input data such that the current class is transformed into a representation with an invariant (the y-axis). The background color indicates the distance to the nearest training data point from the current class in the original space. Images with a white-shaded background result from rotation layers, and images with a gray background result from the conditional coupling layers.

D NEAR-OOD

Near-OOD experiments have largely been ignored when evaluating outlier synthesis methods. We compare against the strongest publicly available baseline on CIFAR100:CIFAR10 and ImageNet100:SSB-Hard. From Tab. 5, we find NCIS to outperform the official Dream-OOD checkpoints in both cases, demonstrating its effectiveness also in these challenging scenarios.

	CIFAR10		SSB-HARD
	ResNet34	ConvNeXt	ResNet34
Dream-OOD	78.7	-	83.9
NCIS (ours)	**80.5**$\pm$0.3	**91.6**$\pm$0.1	**85.2**$\pm$0.1

Table 5: **Comparison on Near-OOD in AUC (↑).**

E COMPARISON TO BOOD

We compare to concurrent work BOOD (Liao et al., 2025) in Tab. 6 and Tab. 7. We find the methods to achieve similar results on CIFAR100, and NCIS to outperform BOOD on ImageNet-100.

Methods	SVHN		PLACES365		LSUN		iSUN		TEXTURES		Average		Acc
	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	
BOOD	5.42	98.43	40.55	90.76	2.06	99.25	0.22	99.91	5.1	98.74	10.67	97.42	78.03
NCIS (ours)	14.43±3.5	96.76±0.8	8.72±0.5	97.71±0.2	21.72±3.1	95.39±0.5	1.42±0.5	99.56±0.1	7.9±0.5	97.96±0.3	10.84±0.8	97.48±0.2	78.86±0.5

Table 6: **Comparing NCIS to BOOD on CIFAR-100.**

Methods	iNATURALIST		PLACES		SUN		TEXTURES		Average		Acc
	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	FPR95↓	AUC↑	
BOOD	18.33	96.74	33.33	94.08	37.92	93.52	51.88	85.41	35.37	92.44	87.92
NCIS (ours)	20.7±0.2	96.56±0.2	34.66±0.4	94.07±0.2	35.43±0.8	94.13±0.2	44.83±1.8	88.5±0.9	33.89±0.6	93.32±0.2	87.24±0.1

Table 7: **Comparing NCIS to BOOD on IMAGENET-100.**