SUPPLEMENTARY MATERIALS OF "DECOMPOSING TEXTURE AND SEMANTIC FOR OUT-OF-DISTRIBUTION DETECTION"

Anonymous authors
Paper under double-blind review

A IMPLEMENTATION DETAILS

Training settings. We use Adam (Kingma & Ba, 2014) with a weight decay of 5e-5 and a batch size of 256 for 100 epochs to train all the modules in our framework. Extracted features by the E_T and E_S are passed to the PCA to be reduced the dimension to 56. We run train and test with a single NVIDIA 1080TI GPU.

Dataset. We replicate the color channel by three times for the grayscale datasets such as the MNIST family. For the corrupted CIFAR-10, we follow the CIFAR-10-C setting parameters (Hendrycks & Dietterich, 2019). To perform a fair comparison, we use identical settings to the other methods such as image normalization.

Image resolution. When we are able to utilize the official pre-trained network or samples, we provide the 32×32 resolution images as input. If not the case (such as our benchmark), we prepossess the image as to have 32×32 resolution following the official setup. In our method, except for the resolution change experiment, we resize the images to 112×112 resolution to give enough information to the texture extraction module.

A.1 MODEL ARCHITECTURE

Multi-SVDD. We use ResNet-18 as the encoder network and modify the dimension of the last layer to 512. All the hidden dimensions of semantic modules have the same as the original setup. The γ parameter for the angular distance is set as 250 for all in-distribution datasets.

RealNVP. We use the RealNVP implementation following Izmailov et al. (2020)[1]. In detail, each perspective model has 8 blocks of 8 flows and we use an affine coupling that is defined by the fully connected shift and scale networks each of which 16 dimensions of hidden layers.

[1] https://github.com/izmailovpavel/flowgmm

B VISUALIZATION

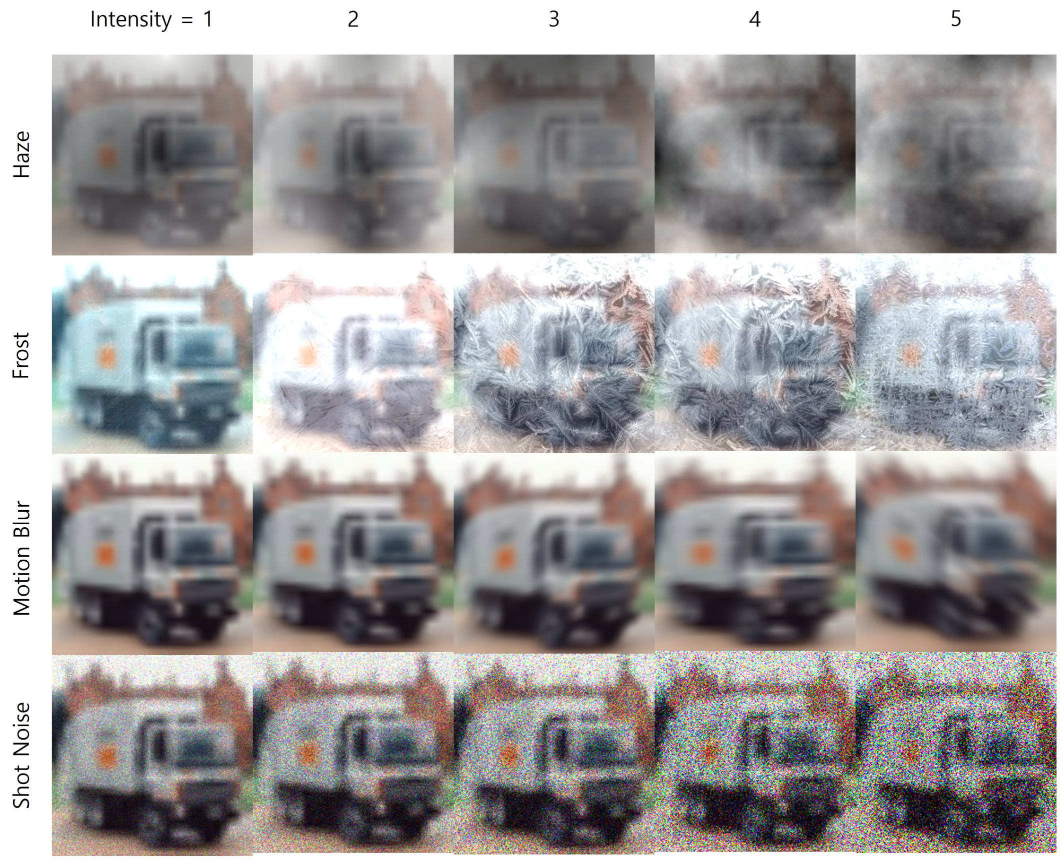

Figure 1: **Examples of the corrupted CIFAR-10 dataset.** We use four types of distortions and five levels of severity in each.

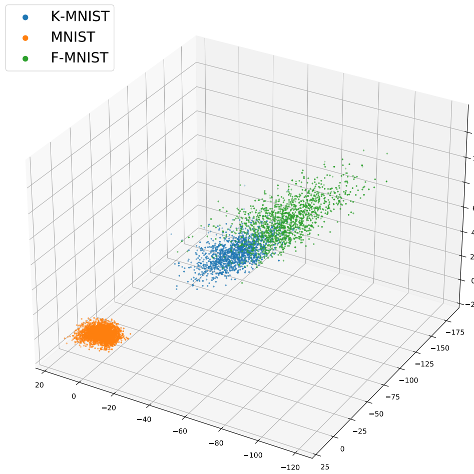

Figure 2: In-distribution: MNIST and F-MNIST, K-MNIST

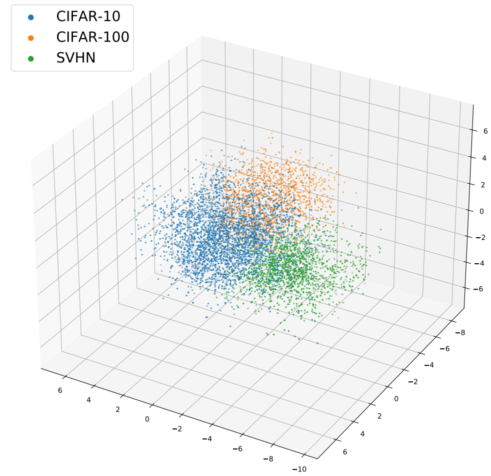

Figure 3: In-distribution: CIFAR-10 and SVHN, CIFAR-100

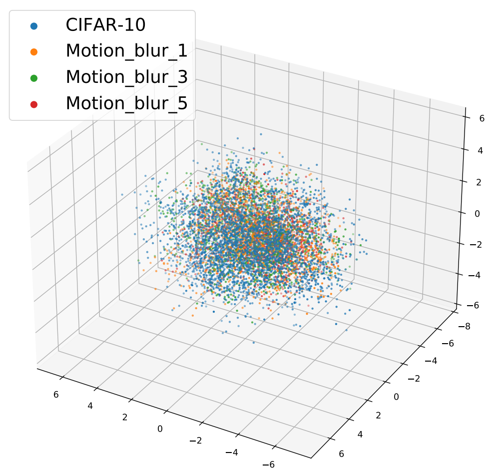

Figure 4: In-distribution: CIFAR-10 and Motion blur corruption.

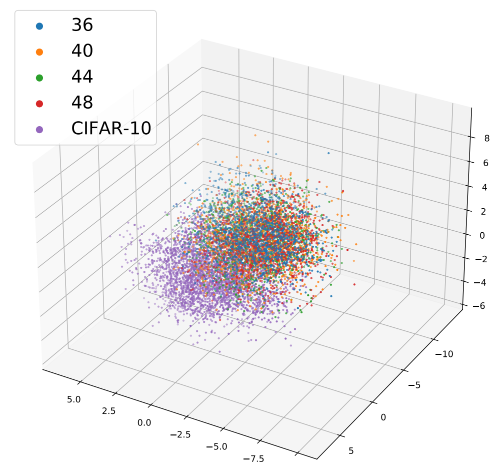

Figure 5: In-distribution: CIFAR-10 and Resolution change.

Figure 6: Embedding results of Multi-SVDD models with Angular initialization. For visualization, we apply PCA in three dimensions. The left side of the figure in the first row is an MNIST in-distribution data-based embedding results. In this figure, Other datasets are embedded far from the MNIST. Moreover, these two datasets (KMNIST, FMNIST) show that separately each other. Another figure in-distribution dataset is CIF10. In the case of corrupted CIF10 and CIF100, figures support our argument that our semantics extraction module S(x) focuses on the semantics information, such as labels from CIF10 and CIF100.

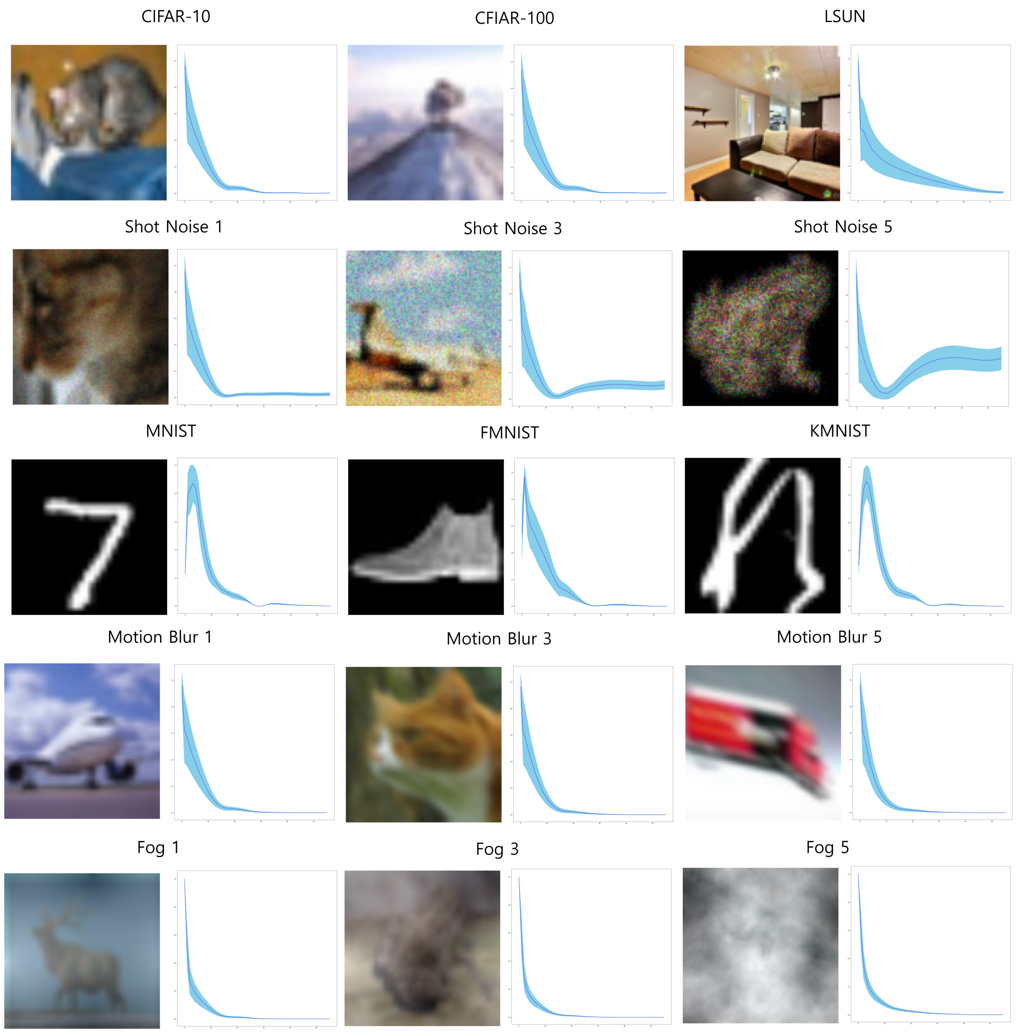

Figure 7: **Visualization of the power spectrum density.** The left side is the example image of the dataset and the right side is the average power spectrum density map of the corresponding dataset.

C Other Comparison Results

	ID → OOD	R-Flow	MOOD	1-Dim	G-ODIN	Ours $\lambda =$ 0.0	0.5	1.0
C10	SVHN	98.2	96.4	-	98.8	86.1	99.9	99.9
	TinyImgNet	99.6	-	-	-	65.4	99.9	99.9
	LSUN(c)	-	99.2	99.4	98.3	68.3	90.9	98.6
	LSUN(r)	99.6	93.2	99.3	99.4	97.0	99.4	99.6
	ImgNet(c)	-	-	98.1	98.7	85.2	98.1	98.5
	ImgNet(r)	-	-	98.5	99.1	85.0	98.1	98.9
C100	SVHN	95.1	85.8	-	95.9	80.9	99.9	100
	TinyImgNet	98.1	-	-	-	91.7	100	100
	LSUN(c)	-	96.8	93.8	95.3	65.6	89.9	92.2
	LSUN(r)	98.9	77.6	95.7	98.7	97.0	99.8	99.6
	ImgNet(c)	-	-	88.6	97.6	94.1	100	100
	ImgNet(r)	-	-	93.7	98.6	94.7	97.7	94.2

Table 1: **More comparison results of Conventional OOD detection benchmark:** R-Flow (Zisselman & Tamar, 2020), MOOD (Lin et al., 2021), 1-Dim (Zaeemzadeh et al., 2021) and G-ODIN (Hsu et al., 2020).

D Ablation Study

Method	CIF10		CIF100		SVHN	
	CIF100	SVHN	CIF10	SVHN	CIF10	CIF100
multi	53.7	99.7	67.5	84.2	87.5	84.2
angular	93.5	99.9	84.2	100.	100.	99.9

Table 2: **Effect of our angular initialization. existing Multi-SVDD method could not handle the only subtle change of semantics.**

Method	ODIN	Gram	CSI	Ours
CIFAR10	47.2	44.6	61.9	50.1
CIFAR100	-	-	-	49.8
SVHN	-	-	-	50.0

Table 3: **In-distribution test dataset results. Each model was trained from the training dataset from In-distribution. and tested the test dataset from In-distribution.**

Figure 8: **Example of our ablation analysis for augmentation PSD. We blend the CIF10 and CIF100 dataset half ratio. And we cut the four parts of CIF10 and randomly arranged it (Jigsaw Puzzle).**

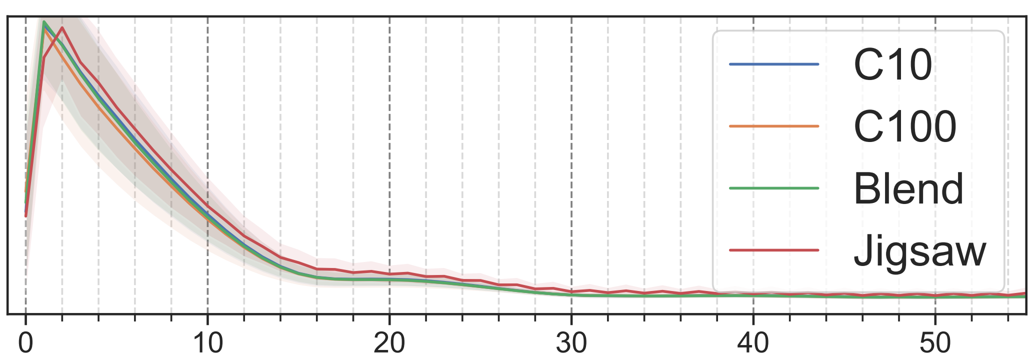

Figure 9: **PSD results from blending and Jigsaw Puzzle. Because Jigsaw Puzzle data generate a sharp line in images, the PSD in Jigsaw has more high frequency.**

REFERENCES

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019.

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*, pp. 10951–10960, 2020.

Pavel Izmailov, Polina Kirichenko, Marc Finzi, and Andrew Gordon Wilson. Semi-supervised learning with normalizing flows. In *International Conference on Machine Learning*, pp. 4615–4630. PMLR, 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Ziqian Lin, Sreya Dutta Roy, and Yixuan Li. Mood: Multi-level out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15313–15323, 2021.

Alireza Zaeemzadeh, Niccolò Bisagno, Zeno Sambugaro, Nicola Conci, Nazanin Rahnavard, and Mubarak Shah. Out-of-distribution detection using union of 1-dimensional subspaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9452–9461, 2021.

Ev Zisselman and Aviv Tamar. Deep residual flow for out of distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13994–14003, 2020.