Improving Out-of-Distribution Detection Performance using Synthetic Outlier Exposure Generated by Visual Foundation Models: Supplementary Material

Gitaek Kwon*[1]
gitaek.kwon@vuno.co
Jaeyoung Kim*[2]
jaydee.kim@kakaohealthcare.com
Hongjun Choi[1]
hongjun.choi@vuno.co
Byungmoo Yoon[3]
byungmoo.yoon@gachon.ac.kr
Sungchul Choi[4]
sc82.choi@pknu.ac.kr
Kyu-Hwan Jung[5]
khwanjung@skku.edu

[1] VUNO Inc., Korea
[2] KakaoHealthcare Corp., Korea
[3] Gachon University, Korea
[4] Pukyong National University, Korea
[5] Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Korea

A Qualitative Analysis of SHIFT

To provide readers with more information, we show the quality of SHIFT's generation through a qualitative comparison with KIRBY (Figure 1). Although the high-level concept of SHIFT is similar to KIRBY, OOD images generated by SHIFT are more realistic compared to KIRBY. Meanwhile, the outlier images produced by KIRBY have artifacts and there are still regions where the ID class can be inferred (third column in Figure 1).

B Details of Implementations

Super-resolution model. To match the image resolution among datsets, we have resized the ID sample to 128×128 pixels and then the upscaled images are fed into LDM_{sr}[1] that generates the high-resolution images (512×512 pixels). The qualitative result of the OOD samples generated with and without LDM_{sr} is presented in Figure 2. By using an LDM_{sr}, CLIP_{seg} can detect the tight boundary of an ID object, and resultant OOD images are more natural because $\text{LDM}_{\text{inpaint}}$ fills the masked regions using rich features of background.

[1] https://huggingface.co/stabilityai/stable-diffusion-x4-upscaler

Figure 1: Qualitative comparison between SHIFT and KIRBY. The presented images are STL-10.

Baselines The hyperparameters of compared baseline methods are followed the original work as possible for a fair comparison. For the post-hoc methods, excluding parameter-free methods (i.e., MSP and Energy), we report the best OOD detection performance by varying

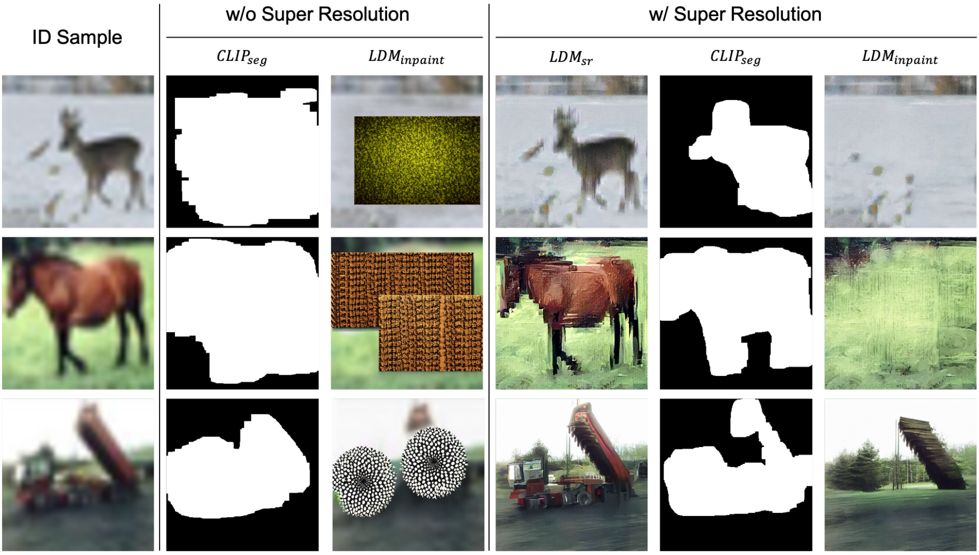

Figure 2: Ablation study with respect to LDM_{sr}. Each ID image is the training sample of CIFAR-10.

their hyperparameters and adopting their best settings on the each ID/OOD pair. The range of hyperparameters used in the experiment is the same as in original papers.

C Additional Result

Impact of the number of K. One of the benefits leveraging $LDM_{inpaint}$ is that diverse OOD samples can be generated from one ID sample (Figure 1). In Table 1, SHIFT shows better OOD detection performance by increasing the number of K.

K	CIFAR-10	CIFAR-100	STL-10
1	0.9826	0.9481	0.9351
2	0.9850	0.9512	0.9420
3	0.9862	0.9521	0.9434
4	0.9872	0.9550	0.9447
5	0.9874	0.9552	0.9460
6	0.9876	0.9557	0.9468
7	**0.9877**	0.9558	0.9464
8	0.9876	**0.9561**	0.9472
9	0.9876	0.9559	**0.9480**
10	0.9876	0.9558	0.9477

Table 1: Comparison of OOD detection results with number of K using WideResNet. Each value is the AUROC and the best results are highlighted in bold.