Dream the Impossible: Outlier Imagination with Diffusion Models (Appendix)

A Broader Impact

Our project aims to improve the reliability and safety of modern machine learning models. Our study on using diffusion models to synthesize outliers can lead to direct benefits and societal impacts, particularly when auxiliary outlier datasets are costly to obtain, such as in safety-critical applications i.e., autonomous driving and healthcare data analysis. Nowadays, research on diffusion models is prevalent, which provides various promising opportunities for exploring the off-the-shelf large models for our research. Our study does not involve any violation of legal compliance. Through our study and releasing our code, we hope to raise stronger research and societal awareness towards the problem of data synthesis for out-of-distribution detection in real-world settings.

B Details of datasets

ImageNet-100. We randomly sample 100 classes from IMAGENET-1K [12] to create IMAGENET-100. The dataset contains the following categories: 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.

OOD datasets. Huang *et.al.* [40] curated a diverse collection of subsets from iNaturalist [98], SUN [109], Places [118], and Texture [9] as large-scale OOD datasets for IMAGENET-1K, where the classes of the test sets do not overlap with IMAGENET-1K. We provide a brief introduction for each dataset as follows.

iNaturalist contains images of natural world [98]. It has 13 super-categories and 5,089 sub-categories covering plants, insects, birds, mammals, and so on. We use the subset that contains 110 plant classes which are not overlapping with IMAGENET-1K.

SUN stands for the Scene UNderstanding Dataset [109]. SUN contains 899 categories that cover more than indoor, urban, and natural places with or without human beings appearing in them. We use the subset which contains 50 natural objects not in IMAGENET-1K.

Places is a large scene photographs dataset [118]. It contains photos that are labeled with scene semantic categories from three macro-classes: Indoor, Nature, and Urban. The subset we use contains 50 categories that are not present in IMAGENET-1K.

Texture stands for the Describable Textures Dataset [9]. It contains images of textures and abstracted patterns. As no categories overlap with IMAGENET-1K, we use the entire dataset as in [40].

ImageNet-A contains 7,501 images from 200 classes, which are obtained by collecting new data and keeping only those images that ResNet-50 models fail to correctly classify [34]. In our paper, we evaluate on the 41 overlapping classes with IMAGENET-100 which consist of a total of 1,852 images.

ImageNet-v2 used in our paper is sampled to match the MTurk selection frequency distribution of the original IMAGENET validation set for each class [75]. The dataset contains 10,000 images from 1,000 classes. During testing, we evaluate on the 100 overlapping classes with a total of 1,000 images.

C Formulation of $Z_m(\kappa)$

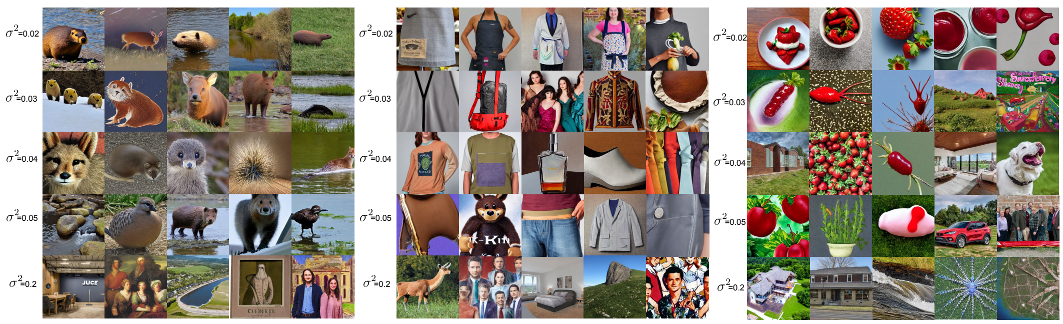
The normalization factor $Z_m(\kappa)$ in Equation (3) is defined as:

$$Z_m(\kappa) = \frac{\kappa^{m/2-1}}{(2\pi)^{m/2} I_{m/2-1}(\kappa)}, \tag{8}$$

where I_v is the modified Bessel function of the first kind with order v. $Z_m(\kappa)$ can be calculated in closed form based on κ and the feature dimensionality m.

D Additional Visualization of the Imagined Outliers

In addition to Section 4.2, we provide additional visualizations on the imagined outliers under different variance σ^2 in Figure 8. We observe that a larger variance consistently translates into outliers that are more deviated from ID data. Using a mild variance value $\sigma^2 = 0.03$ generates both empirically (Figure 7 (b)) and visually meaningful outliers for model regularization on IMAGENET-100.



(a) Imagined outliers for ID class **Beaver** (b) Imagined outliers for ID class **Apron** (c) Imagined outliers for ID class **Strawberry**

Figure 8: **Visualization of the imageined outliers** for the *beaver, apron, strawberry* class with different variance values σ^2.

E Visualization of Outlier Generation by Embedding Interpolation

We visualize the generated outlier images by interpolating token embeddings from different classes in Figure 9. The result shows that interpolating different class token embeddings tends to generate images that are still in-distribution rather than images with semantically mixed or novel concepts, which is aligned with the observations in Liew *et.al.* [51]. Therefore, regularizing the model using such images is not effective for OOD detection (Table 2).

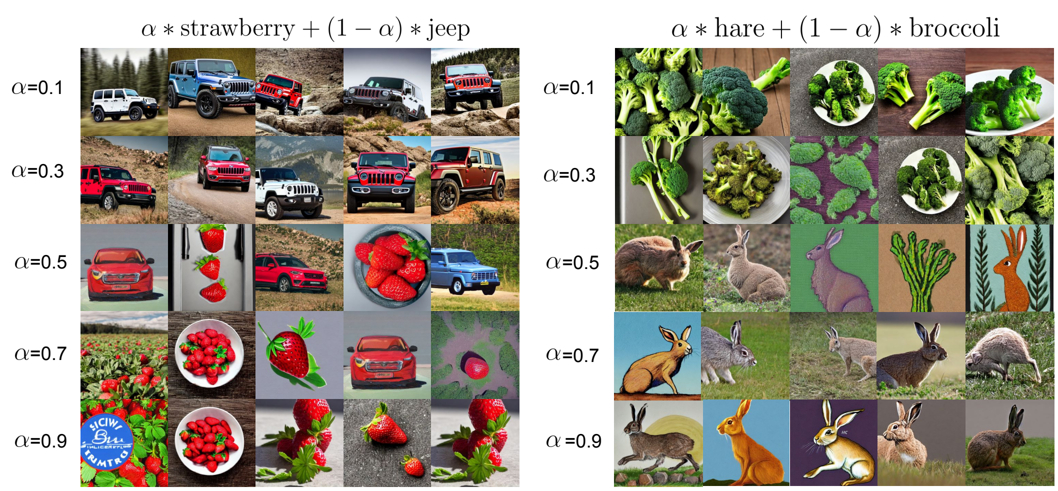


Figure 9: **Visualization of the generated outlier images** by interpolating token embeddings from different classes. We show the results with different interpolation weight α.

F Visualization of the Outlier Generation by Adding Noise

As in Table 2 in the main paper, we visualize the generated outlier images by adding Gaussian and learnable noise to the token embeddings in Figure 10. We observe that adding Gaussian noise tends to generate either ID images or images that are far away from the given ID class. In addition, adding learnable noise to the token embeddings will generate images that are completely deviated from the ID data. Both of them are less effective in regularizing the model's decision boundary.

(a) Generated outliers by adding Gaussian noise for ID class **Beaver** (b) Generated outliers by adding Gaussian noise for ID class **Apron** (c) Generated outliers by adding Gaussian noise for ID class **Strawberry**

(d) Generated outliers by adding learnable noise for ID class **Beaver** (e) Generated outliers by adding learnable noise for ID class **Apron** (f) Generated outliers by adding learnable noise for ID class **Strawberry**

Figure 10: **Visualization of the generated outlier images** by adding Gaussian and learnable noise to the token embeddings from different classes.

G Comparison with Training w/ real Outlier Data.

We compare with training using real outlier data on CIFAR-100, *i.e.,* 300K Random Images [32], which contains 300K preprocessed images that do not belong to CIFAR-100 classes. The result shows that DREAM-OOD (FPR95: 40.31%, AUROC: 90.15%) can match or even outperform outlier exposure with real OOD images (FPR95: 54.32%, AUROC: 91.34%) under the same training configuration while using fewer synthetic OOD images for OOD regularization (100K in total).

H Visualization of Generated Inlier Images

We show in Figure 11 the visual comparison among the original IMAGENET images, the generated images by our DREAM-ID, and the generated ID images using generic prompts "A high-quality photo of a [cls]" where "[cls]" denotes the class name. Interestingly, we observe that the prompt-based generation produces object-centric and distributionally dissimilar images from the original dataset. In contrast, our approach DREAM-ID generates inlier images that can resemble the original ID data, which helps model generalization.

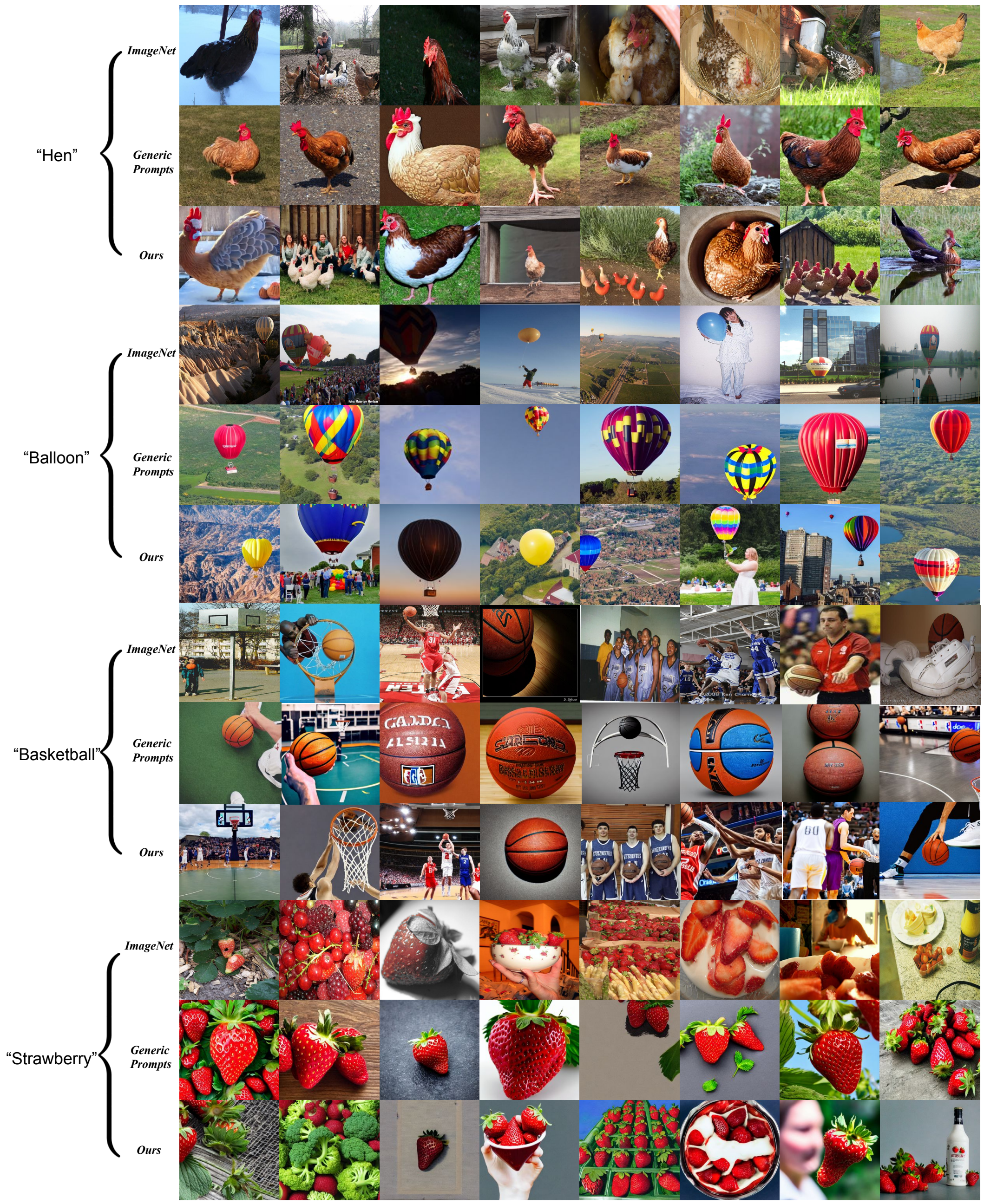


Figure 11: **Visual comparison between our DREAM-ID vs. prompt-based image generation** on four different classes.

I Experimental Details for Model Generalization

We provide experimental details for Section 4.3 in the main paper. We use ResNet-34 [27] as the network architecture, trained with the standard cross-entropy loss. For both the CIFAR-100 and IMAGENET-100 datasets, we train the model for 100 epochs, using stochastic gradient descent with the cosine learning rate decay schedule, a momentum of 0.9, and a weight decay of $5e^{-4}$. The initial learning rate is set to 0.1 and the batch size is set to 160. We generate $1,000$ new ID samples per class using Stable Diffusion v1.4, which result in $100,000$ synthetic images. For both the baselines and our method, we train on a combination of the original IMAGENET/CIFAR samples and synthesized ones. To learn the feature encoder h_θ, we set the temperature t in Equation (2) to 0.1. Extensive ablations on hyperparameters σ and k are provided in Appendix K.

J Implementation Details of Baselines for Model Generalization

For a fair comparison, we implement all the data augmentation baselines by appending the original IMAGENET-100 dataset with the same amount of augmented images (*i.e.*, 100k) generated from different augmentation techniques. We follow the default hyperparameter setting as in their original papers.

- For RandAugment [11], we set the number of augmentation transformations to apply sequentially to 2. The magnitude for all the transformations is set to 9.
- For AutoAugment [10], we set the augmentation policy as the best one searched on IMAGENET.
- For CutMix [115], we use a CutMix probability of 1.0 and set β in the Beta distribution to 1.0 for the label mixup.
- For AugMix [33], we randomly sample 3 augmentation chains and set $\alpha = 1$ for the Dirichlet distribution to mix the images.
- For DeepAugment [30], we directly use the corrupted images for data augmentation provided in their Github repo [3].
- For MEMO [116], we follow the original paper and use the marginal entropy objective for test-time adaptation, which disentangles two distinct self-supervised learning signals: encouraging invariant predictions across different augmentations of the test point and encouraging confidence via entropy minimization.

Methods	IMAGENET	IMAGENET-A	IMAGENET-V2
Original (no aug)	87.28	8.69	77.80
RandAugment	87.56	11.07	79.20
AutoAugment	87.40	10.37	79.00
CutMix	87.64	11.33	79.70
AugMix	87.22	9.39	77.80
DREAM-ID (Ours)	**88.46**$\pm$0.1	**12.13**$\pm$0.1	**80.40**$\pm$0.1

Table 5: **Model generalization performance (accuracy, in %), using IMAGENET-100 as the training data.** The baselines are implemented by directly applying the augmentations on IMAGENET-100.

We also provide the comparison in Table 5 with baselines that are directly trained by applying the augmentations on IMAGENET without appending the original images. The model trained with the images generated by DREAM-ID can still outperform all the baselines by a considerable margin.

K Ablation Studies on Model Generalization

In this section, we provide additional analysis of the hyperparameters and designs of DREAM-ID for ID generation and data augmentation. For all the ablations, we use the IMAGENET-100 dataset as the in-distribution training data.

Ablation on the variance value σ^2. We show in Table 6 the effect of σ^2 — the number of the variance value for the Gaussian kernel (Section 3.2). We vary $\sigma^2 \in \{0.005, 0.01, 0.02, 0.03\}$. A small-mild variance value σ^2 is more beneficial for model generalization.

σ^2	IMAGENET	IMAGENET-A	IMAGENET-V2
0.005	87.62	11.39	78.50
0.01	**88.46**	**12.13**	**80.40**
0.02	87.72	10.85	77.70
0.03	87.28	10.91	78.20

Table 6: Ablation study on the variance value σ^2 in the Gaussian kernel for model generalization.

[3] https://github.com/hendrycks/imagenet-r/blob/master/DeepAugment

Ablation on k in calculating k-NN distance. In Table 7, we analyze the effect of k, *i.e.*, the number of nearest neighbors for non-parametric sampling in the latent space. In particular, we vary $k = \{100, 200, 300, 400, 500\}$. We observe that our method is not sensitive to this hyperparameter, as k varies from 100 to 500.

k	IMAGENET	IMAGENET-A	IMAGENET-V2
100	**88.51**	12.11	79.92
200	88.35	12.04	80.01
300	88.46	**12.13**	**80.40**
400	88.43	12.01	80.12
500	87.72	11.78	80.29

Table 7: Ablation study on the k for k-NN distance for model generalization.

L Software and hardware

We run all experiments with Python 3.8.5 and PyTorch 1.13.1, using NVIDIA GeForce RTX 2080Ti GPUs.

References

[1] Hazrat Ali, Shafaq Murad, and Zubair Shah. Spot the fake lungs: Generating synthetic medical images using neural diffusion models. *arXiv preprint arXiv:2211.00902*, 2022.

[2] Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J Fleet. Synthetic data from diffusion models improves imagenet classification. *arXiv preprint arXiv:2304.08466*, 2023.

[3] Hritik Bansal and Aditya Grover. Leaving reality to imagination: Robust classification via generated datasets. *arXiv preprint arXiv:2302.02503*, 2023.

[4] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1563–1572, 2016.

[5] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Discriminative out-of-distribution detection for semantic segmentation. *arXiv preprint arXiv:1808.07703*, 2018.

[6] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

[7] Max F Burg, Florian Wenzel, Dominik Zietlow, Max Horn, Osama Makansi, Francesco Locatello, and Chris Russell. A data augmentation perspective on diffusion models and retrieval. *arXiv preprint arXiv:2304.10253*, 2023.

[8] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. *arXiv preprint arXiv:2211.09788*, 2022.

[9] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3606–3613, 2014.

[10] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[11] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[13] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

[14] Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. In *International Conference on Learning Representations*, 2023.

[15] Xuefeng Du, Xin Wang, Gabriel Gozum, and Yixuan Li. Unknown-aware object detection: Learning what you don't know from videos in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[16] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. In *Proceedings of the International Conference on Learning Representations*, 2022.

[17] Zhen Fang, Yixuan Li, Jie Lu, Jiahua Dong, Bo Han, and Feng Liu. Is out-of-distribution detection learnable? In *Advances in Neural Information Processing Systems*, 2022.

[18] Nicola Franco, Daniel Korth, Jeanette Miriam Lorenz, Karsten Roscher, and Stephan Guennemann. Diffusion denoised smoothing for certified and adversarial robust out-of-distribution detection. *arXiv preprint arXiv:2303.14961*, 2023.

[19] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit Haim Bermano, Gal Chechik, and Daniel Cohen-or. An image is worth one word: Personalizing text-to-image generation using textual inversion. In *International Conference on Learning Representations*, 2023.

[20] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the International Conference on Machine Learning*, pages 1050–1059, 2016.

[21] Ido Galil, Mohammed Dabbah, and Ran El-Yaniv. A framework for benchmarking class-out-of-distribution detection and its application to imagenet. In *International Conference on Learning Representations*, 2023.

[22] Yunhao Ge, Jiashu Xu, Brian Nlong Zhao, Laurent Itti, and Vibhav Vineet. Dall-e for detection: Language-driven context image synthesis for object detection. *arXiv preprint arXiv:2206.09592*, 2022.

[23] Yunhao Ge, Jiashu Xu, Brian Nlong Zhao, Laurent Itti, and Vibhav Vineet. Em-paste: Em-guided cut-paste with dall-e augmentation for image-level weakly supervised instance segmentation. *arXiv preprint arXiv:2212.07629*, 2022.

[24] Yonatan Geifman and Ran El-Yaniv. Selectivenet: A deep neural network with an integrated reject option. In *Proceedings of the International Conference on Machine Learning*, pages 2151–2159, 2019.

[25] Sven Gowal, Sylvestre-Alvise Rebuffi, Olivia Wiles, Florian Stimberg, Dan Andrei Calian, and Timothy A Mann. Improving robustness using generated data. *Advances in Neural Information Processing Systems*, 34:4218–4233, 2021.

[26] Mark S Graham, Walter HL Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and M Jorge Cardoso. Denoising diffusion models for out-of-distribution detection. *arXiv preprint arXiv:2211.07740*, 2022.

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[28] Ruifei He, Shuyang Sun, Xin Yu, Chuhui Xue, Wenqing Zhang, Philip Torr, Song Bai, and Xiaojuan Qi. Is synthetic data from generative models ready for image recognition? In *International Conference on Learning Representations*, 2023.

[29] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 41–50, 2019.

[30] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8349, 2021.

[31] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of the International Conference on Learning Representations*, 2017.

[32] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *Proceedings of the International Conference on Learning Representations*, 2019.

[33] Dan Hendrycks*, Norman Mu*, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple method to improve robustness and uncertainty under data shift. In *International Conference on Learning Representations*, 2020.

[34] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15262–15271, 2021.

[35] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*, 2022.

[36] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[37] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47–1, 2022.

[38] 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.

[39] Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. In *Advances in Neural Information Processing Systems*, 2021.

[40] Rui Huang and Yixuan Li. Mos: Towards scaling out-of-distribution detection for large semantic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8710–8719, 2021.

[41] Saachi Jain, Hannah Lawrence, Ankur Moitra, and Aleksander Madry. Distilling model failures as directions in latent space. In *International Conference on Learning Representations*, 2023.

[42] Taewon Jeong and Heeyoung Kim. Ood-maml: Meta-learning for few-shot out-of-distribution detection and classification. *Advances in Neural Information Processing Systems*, 33:3907–3916, 2020.

[43] Julian Katz-Samuels, Julia B. Nakhleh, Robert D. Nowak, and Yixuan Li. Training OOD detectors in their natural habitats. In *Proceedings of the International Conference on Machine Learning*, pages 10848–10865, 2022.

[44] Konstantin Kirchheim and Frank Ortmeier. On outlier exposure with generative models. In *NeurIPS ML Safety Workshop*, 2022.

[45] Agustinus Kristiadi, Matthias Hein, and Philipp Hennig. Being bayesian, even just a bit, fixes overconfidence in relu networks. In *International conference on machine learning*, pages 5436–5446, 2020.

[46] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[47] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, pages 6402–6413, 2017.

[48] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *Proceedings of the International Conference on Learning Representations*, 2018.

[49] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in Neural Information Processing Systems*, 31, 2018.

[50] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *Proceedings of the International Conference on Learning Representations*, 2018.

[51] Jun Hao Liew, Hanshu Yan, Daquan Zhou, and Jiashi Feng. Magicmix: Semantic mixing with diffusion models. *arXiv preprint arXiv:2210.16056*, 2022.

[52] Luping Liu, Yi Ren, Xize Cheng, and Zhou Zhao. Diffusion denoising process for perceptron bias in out-of-distribution detection. *arXiv preprint arXiv:2211.11255*, 2022.

[53] Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In *International Conference on Learning Representations*, 2022.

[54] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33:21464–21475, 2020.

[55] Zhenzhen Liu, Jin Peng Zhou, Yufan Wang, and Kilian Q Weinberger. Unsupervised out-of-distribution detection with diffusion inpainting. In *Proceedings of the International Conference on Machine Learning*, 2023.

[56] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 2022.

[57] Wesley J Maddox, Pavel Izmailov, Timur Garipov, Dmitry P Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems*, 32:13153–13164, 2019.

[58] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in Neural Information Processing Systems*, 31, 2018.

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

[60] Kanti V Mardia, Peter E Jupp, and KV Mardia. *Directional statistics*, volume 2. Wiley Online Library, 2000.

[61] Alexander Meinke and Matthias Hein. Towards neural networks that provably know when they don't know. In *Proceedings of the International Conference on Learning Representations*, 2020.

[62] Yifei Ming, Ying Fan, and Yixuan Li. POEM: out-of-distribution detection with posterior sampling. In *Proceedings of the International Conference on Machine Learning*, pages 15650–15665, 2022.

[63] Sina Mohseni, Mandar Pitale, JBS Yadawa, and Zhangyang Wang. Self-supervised learning for generalizable out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5216–5223, 2020.

[64] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[65] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 427–436, 2015.

[66] Minheng Ni, Zitong Huang, Kailai Feng, and Wangmeng Zuo. Imaginarynet: Learning object detectors without real images and annotations. In *International Conference on Learning Representations*, 2023.

[67] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *International Conference on Machine Learning*, 2022.

[68] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171, 2021.

[69] William Peebles and Saining Xie. Scalable diffusion models with transformers. *arXiv preprint arXiv:2212.09748*, 2022.

[70] Walter HL Pinaya, Mark S Graham, Robert Gray, Pedro F Da Costa, Petru-Daniel Tudosiu, Paul Wright, Yee H Mah, Andrew D MacKinnon, James T Teo, Rolf Jager, et al. Fast unsupervised brain anomaly detection and segmentation with diffusion models. *arXiv preprint arXiv:2206.03461*, 2022.

[71] Walter HL Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M Jorge Cardoso. Brain imaging generation with latent diffusion models. In *MICCAI Workshop on Deep Generative Models*, pages 117–126, 2022.

[72] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022.

[73] 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 *Proceedings of the International Conference on Machine Learning*, pages 8748–8763, 2021.

[74] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.

[75] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? In *International Conference on Machine Learning*, pages 5389–5400, 2019.

[76] Jie Ren, Jiaming Luo, Yao Zhao, Kundan Krishna, Mohammad Saleh, Balaji Lakshminarayanan, and Peter J Liu. Out-of-distribution detection and selective generation for conditional language models. In *International Conference on Learning Representations*, 2023.

[77] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[78] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *arXiv preprint arXiv:2208.12242*, 2022.

[79] Luke W Sagers, James A Diao, Matthew Groh, Pranav Rajpurkar, Adewole S Adamson, and Arjun K Manrai. Improving dermatology classifiers across populations using images generated by large diffusion models. *arXiv preprint arXiv:2211.13352*, 2022.

[80] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022.

[81] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Raphael Gontijo-Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. In *Advances in Neural Information Processing Systems*, 2022.

[82] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[83] Mert Bulent Sariyildiz, Karteek Alahari, Diane Larlus, and Yannis Kalantidis. Fake it till you make it: Learning transferable representations from synthetic imagenet clones. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[84] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade W Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa R Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. LAION-5b: An open large-scale dataset for training next generation image-text models. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, 2022.

[85] Vikash Sehwag, Mung Chiang, and Prateek Mittal. Ssd: A unified framework for self-supervised outlier detection. In *International Conference on Learning Representations*, 2021.

[86] Vikash Sehwag, Caner Hazirbas, Albert Gordo, Firat Ozgenel, and Cristian Canton. Generating high fidelity data from low-density regions using diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11492–11501, 2022.

[87] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265, 2015.

[88] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

[89] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428, 2021.

[90] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[91] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[92] Yiyou Sun and Yixuan Li. Dice: Leveraging sparsification for out-of-distribution detection. In *Proceedings of European Conference on Computer Vision*, 2022.

[93] Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *Proceedings of the International Conference on Machine Learning*, pages 20827–20840, 2022.

[94] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *Advances in Neural Information Processing Systems*, 2020.

[95] Leitian Tao, Xuefeng Du, Xiaojin Zhu, and Yixuan Li. Non-parametric outlier synthesis. In *Proceedings of the International Conference on Learning Representations*, 2023.

[96] Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective data augmentation with diffusion models. *arXiv preprint arXiv:2302.07944*, 2023.

[97] Joost van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *Proceedings of the International Conference on Machine Learning*, pages 9690–9700, 2020.

[98] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8769–8778, 2018.

[99] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 1999.

[100] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4921–4930, 2022.

[101] Haoran Wang, Weitang Liu, Alex Bocchieri, and Yixuan Li. Can multi-label classification networks know what they don't know? *Proceedings of the Advances in Neural Information Processing Systems*, 2021.

[102] Larry Wasserman. Lecture notes of statistical methods for machine learning. 2019.

[103] Daniel Watson, William Chan, Ricardo Martin Brualla, Jonathan Ho, Andrea Tagliasacchi, and Mohammad Norouzi. Novel view synthesis with diffusion models. In *International Conference on Learning Representations*, 2023.

[104] Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. In *Proceedings of the International Conference on Machine Learning*, pages 23631–23644, 2022.

[105] Yeming Wen, Dustin Tran, and Jimmy Ba. Batchensemble: an alternative approach to efficient ensemble and lifelong learning. In *International Conference on Learning Representations*, 2020.

[106] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. Diffusion models for medical anomaly detection. *arXiv preprint arXiv:2203.04306*, 2022.

[107] Qitian Wu, Yiting Chen, Chenxiao Yang, and Junchi Yan. Energy-based out-of-distribution detection for graph neural networks. In *International Conference on Learning Representations*, 2023.

[108] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 650–656, 2022.

[109] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, 2010.

[110] Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755*, 2015.

[111] 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 International Conference on Computer Vision*, pages 8281–8289, 2021.

[112] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2021.

[113] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

[114] Jianhao Yuan, Francesco Pinto, Adam Davies, Aarushi Gupta, and Philip Torr. Not just pretty pictures: Text-to-image generators enable interpretable interventions for robust representations. *arXiv preprint arXiv:2212.11237*, 2022.

[115] 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.

[116] Marvin Zhang, Sergey Levine, and Chelsea Finn. Memo: Test time robustness via adaptation and augmentation. *Advances in Neural Information Processing Systems*, 2022.

[117] Yifan Zhang, Daquan Zhou, Bryan Hooi, Kai Wang, and Jiashi Feng. Expanding small-scale datasets with guided imagination. *arXiv preprint arXiv:2211.13976*, 2022.

[118] 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.