Dual Conditioned Diffusion Models for Out-Of-Distribution Detection: Application to Fetal Ultrasound Videos

Divyanshu Mishra[1], He Zhao[1], Pramit Saha[1], Aris T. Papageorghiou[2], J.Alison Noble[1]

[1] Institute of Biomedical Engineering, University of Oxford
[2] Nuffield Department of Women's and Reproductive Health, University of Oxford

Abstract. Out-of-distribution (OOD) detection is essential to improve the reliability of machine learning models by detecting samples that do not belong to the training distribution. Detecting OOD samples effectively in certain tasks can pose a challenge because of the substantial heterogeneity within the in-distribution (ID), and the high structural similarity between ID and OOD classes. For instance, when detecting heart views in fetal ultrasound videos there is a high structural similarity between the heart and other anatomies such as the abdomen, and large in-distribution variance as a heart has 5 distinct views and structural variations within each view. To detect OOD samples in this context, the resulting model should generalise to the intra-anatomy variations while rejecting similar OOD samples. In this paper, we introduce dual-conditioned diffusion models (DCDM) where we condition the model on in-distribution class information and latent features of the input image for reconstruction-based OOD detection. This constrains the generative manifold of the model to generate images structurally and semantically similar to those within the in-distribution. The proposed model outperforms reference methods with a 12% improvement in accuracy, 22% higher precision, and an 8% better F1 score.

1 Introduction

Existing out-of-distribution (OOD) detection methods work well when the in-distribution (ID) classes have low heterogeneity (low variance) but fail when in-distribution classes have high heterogeneity [23] or high spatial similarity between ID and OOD classes [9]. Fetal ultrasound (US) anatomy detection is one such application where both the challenges co-exist.

In this paper, we propose a Dual-Conditioned Diffusion Model (DCDM) to detect OOD samples when in-distribution data has high variance and test the performance by detecting heart views in fetal US videos as an example application. Specifically, an Ultrasound (US) typically comprises 13 anatomies and their views. However,analysis models are usually developed for anatomy-specific tasks. Hence, to separate heart views from other 12 anatomies (head, abdomen, femur etc) we develop an OOD detection algorithm. Our in-distribution data

comprises five structurally different heart views captured across different cardiac cycles of a beating heart during obstetric US scanning. We develop a diffusion-based model for reconstruction-based OOD detection, which extends [14] with a novel dual conditioning mechanism that alleviates the influence of high inter- and intra-class variation within different classes by leveraging in-distribution class conditioning (IDCC) and latent image feature conditioning (LIFC). These conditioning mechanisms allow our model to generate images similar to the input image for in-distribution data. The primary contributions of our paper are summarized as follows: 1) We introduce a novel conditioned diffusion model for OOD detection and demonstrate that the dual conditioning mechanism is effective in tackling challenging scenarios where in-distribution data comprises multiple heterogeneous classes and there is a high spatial similarity between ID and OOD classes. 2) Two original conditions are proposed for the diffusion model, which are in-distribution class conditioning (IDCC) and latent image feature conditioning (LIFC). IDCC is proposed to handle high inter-class variance within in-distribution classes and high spatial similarity between ID and OOD classes. LIFC is introduced to counter the intra-class variance within each class. 3) We demonstrate in our experiments that DCDM can detect and separate heart views from other anatomies in fetal ultrasound videos without needing any labelled data for OOD classes. Extensive experiments and ablations demonstrate superior performance over existing OOD detection methods. Our approach is not fetal ultrasound specific and could be applied to other OOD applications.

2 Related Work

OOD detection [31] involves identifying samples that do not belong to the training distribution. Such models can be categorized into: (a) unsupervised OOD detection [23] and (b) supervised OOD detection. [34, 5, 11]. Unsupervised OOD detection methods can again be divided into two main categories: (i) likelihood-based approaches [12, 21, 30], and (ii) reconstruction-based [3, 33, 25]. Likelihood-based approaches suffer from several issues, including assigning higher likelihood to OOD samples [19, 4], susceptibility to adversarial attacks [8], and calibration issues [28]. Current reconstruction-based approaches are sensitive to dimensions of the bottleneck layer and require rigorous tuning specific to the dataset and task [10]. Additionally, models trained using a generator-discriminator architecture and optimizing adversarial losses can be highly unstable and challenging to train [1, 2]. Finally, reconstruction-based methods often rely on highly compressed latent representations, which can lead to loss of important low-level detail. This can be problematic when discriminating between classes with high spatial similarity. Recently, diffusion models have been introduced to address these limitations on tasks such as image synthesis [6], and OOD detection [10].

Denoising Diffusion Probabilistic Models (DDPMs) [14] are generative models that work by gradually adding noise to an input image through a forward diffusion process followed by gradually removing noise using a trained neural network in the backward diffusion process [32]. To guide the generative process of a

diffusion model (DM), previous work [18, 24, 22] condition the DDPMs on task-specific conditioning. In image-to-image translation tasks like super-resolution, colourization, *etc.*, previous papers [24] condition the model by concatenating a resized or grayscale version of the input image to the noised image. This concatenation is unsuitable for reconstruction-based OOD detection as the model will generate similar images for ID and OOD samples. In the context of OOD detection using DMs, previous works [10] have trained unconditional DDPMs and, during inference, sampled using a Pseudo Linear Multi Step (PLMS) [16] sampler for varying noise levels. However, their approach generates 5500 samples to detect OOD samples for each input image which is time-consuming and impractical for settings where shorter inference times are needed. AnoDDPM [29] utilises simplex noise rather than Gaussian noise to corrupt the image (t=250 rather than t=1000) for anomaly detection. However, this approach requires data specific tuning,and is outperformed by [10].

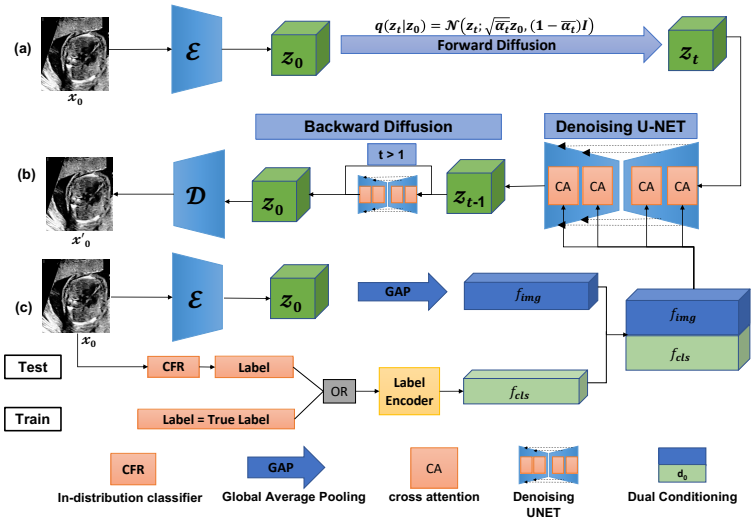

Fig. 1: DCDM architecture where **(*a*)** the input image x_0 is mapped to the latent vector z_0 using a pretrained encoder $\mathcal{E}$ and forward diffusion is applied, **(*b*)** the backward diffusion process denoises the latent vector z_t and the final denoised latent vector z_0 is mapped to pixel space by the decoder $\mathcal{D}$ **(*c*)** the dual-conditioning mechanism. We obtain f_{img} by passing the input image x_0 through the encoder $\mathcal{E}$. f_{cls} is obtained using the true label during training or predicted class label during testing.

3 Methods

3.1 Dual Conditioned Diffusion Models

Diffusion models are generative models that rely on two Markov processes known as forward and backward diffusion [14]. To improve efficiency during training

and inference, forward and backward diffusion is applied to the latent space [22]. Autoencoder (AE = $\mathcal{E}$ + $\mathcal{D}$) is pretrained separately on ID heart data and can successfully reconstruct the input heart images (SSIM=0.956). The latent variable z_0 is obtained by passing an input image x_0 through a pretrained encoder $\mathcal{E}$. Given the latent vector z_0 and a fixed variance schedule [14] $\{\beta_t \in (0,1)\}_{t=1}^{T}$, the forward diffusion process, defined by Eqn. 1, gradually adds Gaussian noise to z_0 to give a noised latent vector z_t where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$:

$$q(\boldsymbol{z}_t|\boldsymbol{z}_0) = \mathcal{N}(\boldsymbol{z}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{z}_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{1}$$

In backward diffusion, we aim to reverse the forward diffusion process and predict z_{t-1} given z_t. To predict $(z_{t-1}|z_t)$,we train a denoising U-Net [14] denoted as $\boldsymbol{\epsilon}_\theta\left(z_t, t, d_0\right)$ that takes the current timestep t, noised latent vector z_t and the dual conditioning embedding vector d_0 as input and predicts the noise at timestep t as shown in Eqn. 2.

$$\boldsymbol{z}_{t-1} = \mathcal{N}(\boldsymbol{z}_{t-1}; \frac{1}{\sqrt{\alpha_t}}\Big(\boldsymbol{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, d_0)\Big), (1-\bar{\alpha}_t)\mathbf{I}) \tag{2}$$

The dual embedding vector d_0 is obtained by combining IDCC (f_{cls}) and LIFC (f_{img}) vectors, which we explain in Section 3.2. The output z_{t-1} is again input to ϵ_θ. This process is repeated until z_0 is obtained.

The final model optimisation objective is given by Eqn. 3 where ϵ is the original noise added during the forward diffusion process.

$$\mathcal{L}_{DCDM} := \mathbb{E}_{\mathcal{E}(x),\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta\left(z_t, t, d_0\right)\|_2^2\right] \tag{3}$$

Once we obtain z_0 from the backward diffusion process, it is passed on to the decoder $\mathcal{D}$ and mapped back to the pixel space to give generated image x_0'.

3.2 Dual Conditioning Mechanism

Image features and in-distribution class information are utilised in our proposed dual conditioning mechanism. This guides the DCDM to generate images that are spatially and semantically similar to the input image for in-distribution samples and dissimilar for OOD samples.

Latent Image Feature Conditioning (LIFC): The image conditioning dictates the desired appearance of generated images in terms of shape and texture. In our model, we use the features extracted by a pretrained encoder for conditioning. Empirically, we use the same encoder $\mathcal{E}$ as our feature extractor to obtain latent feature vector z_0 as shown in Fig. 1. Specifically, the input image of dimension $224 \times 224 \times 3$ is passed through the encoder $\mathcal{E}$ and a feature map with the size of $7 \times 7 \times 128$ is obtained which is followed by global average pooling (GAP) resulting in a feature vector (f_{img}) with dimension 128.

In-Distribution Class Conditioning (IDCC): Given an in-distribution dataset comprising n heterogeneous classes, conditioning the model only on image-level

features is insufficient. Therefore we introduce an in-distribution class conditioning (IDCC) that informs the DCDM of the class of the input image and enables it to generate samples belonging to the same class for ID. A label encoder generates a unique class conditional embedding (f_{cls}) of dimension 128 for each class label. The class label is assigned based on the ground truth label during the training phase and to the classifier's prediction during inference, as depicted in Fig. 1. In practice, we train a CNN classifier, freeze its weight and use it as our in-distribution classifier (CFR), as discussed in section 3.3.

Cross Attention Guidance: To integrate the dual-conditioning guidance into the diffusion model, we use a cross-attention [27] mechanism inside the denoising U-Net rather than just concatenation [24] as it is more effective [13, 20, 17] and allows condition diffusion models on various input modalities [22]. Our LIFC and IDCC are first concatenated to give a feature vector with a dimension of 256. This acts as a side input to each UNet block. The features from the UNet block and the conditional features are fused by cross-attention and serve as input to the following UNet block as shown in Fig. 1.For more details,regarding cross-attention block refer to Rombach *et al* [22].

3.3 In-Distribution Classifier

The in-distribution classifier (CFR) serves two main functions. First, it provides labels for the class conditioning during inference; second, it is utilized as a feature extractor for calculating the OOD score.

Inference Class Guidance. IDCC requires in-distribution class information to generate the class conditional embedding. However, class information is only available during training. To obtain class information during inference, we separately train a ConvNext CNN based classifier (accuracy = 88%) on the in-distribution data and use its predictions as the class information. During inference, the input image x_0 is passed through the classifier, and the predicted label is used to generate the class embedding by feeding to the label encoder as shown in Fig. 1. Moreover, as the classifier is only trained on in-distribution data, it classifies an OOD sample to an in-distribution class. The classifier's prediction is utilised by the DCDM and it tries to generate an image belonging to in-distribution class for the OOD samples. This reduces the structural and semantic similarity between the input and the generated image, as demonstrated by our qualitative results (Fig. 2).

Feature-Based OOD Detection To evaluate the performance of the DCDM, the cosine similarity between features of the input image x_0 and the generated image x'_0 from the in-distribution classifier is calculated and is referred as an OOD score where f_0 and f'_0 are the features of x_0 and x'_0, respectively:

$$\text{OOD score} = \ sim(f_0, f'_0) \ = \frac{f_0 \cdot f'_0}{\|f_0\|_2 \, \|f'_0\|_2}, \tag{4}$$

An input image x_0 is classified as in-distribution (ID) or OOD based on Eqn. 5 where τ is a pre-defined threshold and y_{pred} is the prediction of our feature-based

OOD detection algorithm.

$$y_{pred} = \begin{cases} 0(ID) & if \text{ OOD score} > \tau \\ 1 \text{ (OOD)} & otherwise \end{cases} \tag{5}$$

4 Experiments and Results

Dataset and Implementation For our experiments, we utilized a fetal ultrasound dataset of 359 subject videos that were collected as part of the PULSE project [7]. The in-distribution dataset consisted of 5 standard heart views (3VT, 3VV, LVOT, RVOT, and 4CH), while the out-of-distribution dataset comprised of three non-heart anatomies - fetal head, abdomen, and femur. The original images were of size 1008×784 pixels and were resized to 224×224 pixels.

To train the models, we randomly sampled 5000 fetal heart images and used 500 images for evaluating image generation performance. To test the performance of our final model and compare it with other methods, we used an held-out dataset of 7471 images, comprising 4309 images of different heart views and 3162 images (about 1000 for each anatomy) of out-of-distribution classes. Further details about the dataset are given in **Supp. Fig. 2 and 3**.

All models were trained using PyTorch version 1.12 with a Tesla V100 32 GB GPU. During training, we used T=1000 for noising the input image and a linearly increasing noise schedule that varied from 0.0015 to 0.0195. To generate samples from our trained model, we used DDIM [26] sampling with T=100. All baseline models were trained and evaluated using the original implementation.

4.1 Results

We evaluated the performance of the dual-conditioned diffusion models (DCDMs) for OOD detection by comparing them with two current state-of-the-art unsupervised reconstruction-based approaches and one likelihood-based approach. The first baseline is Deep-MCDD [15], a likelihood-based OOD detection method that proposes a Gaussian discriminant-based objective to learn class conditional distributions. The second baseline is ALOCC [23] a GAN-based model that uses the confidence of the discriminator on reconstructed samples to detect OOD samples. The third baseline is the method of Graham *et al.* [10], where they use DDPM [14] to generate multiple images at varying noise levels for each input. They then compute the MSE and LPIPS metrics for each image compared to the input, convert them to Z-scores, and finally average them to obtain the OOD score.

Quantitative Results The performance of the DCDM, along with comparisons with the other approaches, are shown in Table 1. The GAN-based method ALOCC [23] has the lowest AUC of 57.22%, which is improved to 63.86% by the method of Graham *et al.* and further improved to 64.58% by likelihood-based Deep-MCDD. DCDM outperforms all the reference methods by 20%, 14% and 13%, respectively and has an AUC of 77.60%. High precision is essential for

Table 1: Quantitative comparison of our model (DCDM) with reference methods

Method	AUC(%)	F1-Score(%)	Accuracy(%)	Precision(%)
Deep-MCDD [15]	64.58	66.23	60.41	51.82
ALOCC [23]	57.22	59.34	52.28	45.63
Graham et al. [10]	63.86	63.55	60.15	50.89
DCDM(Ours)	**77.60**	**74.29**	**77.95**	**73.34**

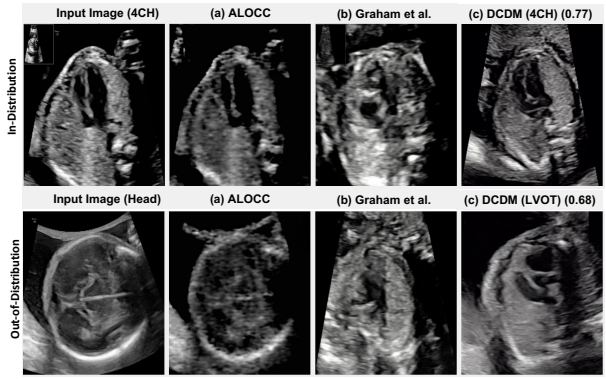

Fig. 2: Qualitative comparison of our method with **(*a*)** ALOCC generates similar images to the input for ID and OOD samples **(*b*)** Graham *et al.* generates any random heart view for a given input image **(*c*)** Our model generates images that are similar to the input image for ID and dissimilar for OOD samples. Classes predicted by CFR and the OOD score ($\tau = 0.73$) are mentioned in brackets.

OOD detection as this can reduce false positives and increase trust in the model. DCDM exhibits a precision that is 22% higher than the reference methods while still having an 8% improvement in F1-Score.

Qualitative Results Qualitative results are shown in Fig. 2. Visual comparisons show ALOCC generates images structurally similar to input images for in-distribution and OOD samples. This makes it harder for the ALOCC model to detect OOD samples. The model of Graham *et al.* generates any random heart view for a given image as a DDPM is unconditional, and our in-distribution data contains multiple heart views. For example, given a 4CH view as input, the model generates an entirely different heart view. However, unlike ALOCC, the Graham *et al.* model generates heart views for OOD samples, improving OOD detection performance. DCDM generates images with high spatial similarity to the input image and belonging to the same heart view for ID samples while structurally diverse heart views for OOD samples. In Fig 2 (c) for OOD sample, even-though the confidence is high (0.68), the gap between ID and OOD classes is wide enough to separate the two.Additional qualitative results can be observed in **Supp. Fig. 4.**

Table 2: Ablation study of different conditioning mechanisms of DCDM.

Method	Accuracy (%)	Precision (%)	AUC (%)
Unconditional	68.16	58.44	69.61
In-Distribution Class Conditioning	74.39	66.12	75.27
Latent Image Feature Conditioning	77.02	70.02	77.40
Dual Conditioning	**77.95**	**73.34**	**77.60**

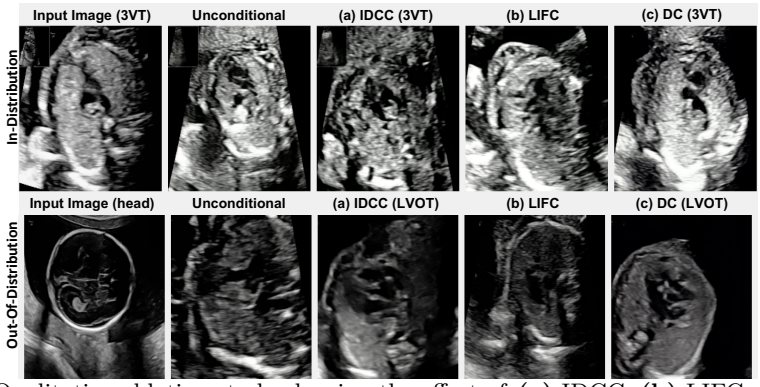

Fig. 3: Qualitative ablation study showing the effect of **(a)** IDCC, **(b)** LIFC and, **(c)** DC on generative results of DM. Brackets in IDCC, DC show labels predicted by CFR.

4.2 Ablation Study

Ablation experiments were performed to study the impact of various conditioning mechanisms on the model performance both qualitatively and quantitatively. When analyzed quantitatively, as shown in Table 2, the unconditional model has the lowest AUC of 69.61%. Incorporating the IDCC guidance or LIFC separately, improves performance with an AUC of 75.27% and 77.40%, respectively. The best results are achieved when both mechanisms are used (DCDM), resulting in an 11% improvement in the AUC score relative to the unconditional model. Although there is a small margin of performance improvement between the combined model (DCDM) and the LIFC model in terms of AUC, the precision improves by 3%, demonstrating the combined model is more precise and hence the best model for OOD detection.

As shown in Fig. 3, the unconditional diffusion model generates a random heart view for a given input for both in-distribution and OOD samples. The IDCC guides the model to generate a heart view according to the in-distribution classifier (CFR) prediction which leads to the generation of similar samples for in-distribution input while dissimilar samples for OOD input. On the other hand, LIFC generates an image with similar spatial information. However, heart views are still generated for OOD samples as the model was only trained on them. When dual-conditioning (DC) is used, the model generates images that are closer aligned to the input image for in-distribution input and high-fidelity heart views for OOD than those generated by a model conditioned on either IDCC or LIFC alone. **Supp. Fig. 1** presents further qualitative ablations.

5 Conclusion

We introduce novel dual-conditioned diffusion model for OOD detection in fetal ultrasound videos and demonstrate how the proposed dual-conditioning mechanisms can manipulate the generative space of a diffusion model. Specifically, we show how our dual-conditioning mechanism can tackle scenarios where the in-distribution data has high inter- (using IDCC) and intra- (using LIFC) class variations and guide a diffusion model to generate similar images to the input for in-distribution input and dissimilar images for OOD input images. Our approach does not require labelled data for OOD classes and is especially applicable to challenging scenarios where the in-distribution data comprises more than one class and there is high similarity between the in-distribution and OOD classes.

6 Acknowledgement

This work was supported in part by the InnoHK-funded Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE) Project 2.1 (Cardiovascular risks in early life and fetal echocardiography), the UK EPSRC (Engineering and Physical Research Council) Programme Grant EP/T028572/1 (VisualAI), and a UK EPSRC Doctoral Training Partnership award.

References

1. Arjovsky, M., Bottou, L.: Towards principled methods for training generative adversarial networks. arXiv preprint arXiv:1701.04862 (2017)
2. Bau, D., Zhu, J.Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., Torralba, A.: Seeing what a gan cannot generate. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4502–4511 (2019)
3. Chen, X., Konukoglu, E.: Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. arXiv preprint arXiv:1806.04972 (2018)
4. Choi, H., Jang, E., Alemi, A.A.: Waic, but why? generative ensembles for robust anomaly detection. arXiv preprint arXiv:1810.01392 (2018)
5. DeVries, T., Taylor, G.W.: Learning confidence for out-of-distribution detection in neural networks. arXiv preprint arXiv:1802.04865 (2018)
6. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in Neural Information Processing Systems **34**, 8780–8794 (2021)
7. Drukker, L., Sharma, H., Droste, R., Alsharid, M., Chatelain, P., Noble, J.A., Papageorghiou, A.T.: Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video. Scientific Reports **11**(1), 14109 (2021)
8. Fort, S.: Adversarial vulnerability of powerful near out-of-distribution detection. arXiv preprint arXiv:2201.07012 (2022)
9. Fort, S., Ren, J., Lakshminarayanan, B.: Exploring the limits of out-of-distribution detection. Advances in Neural Information Processing Systems **34**, 7068–7081 (2021)

10. Graham, M.S., Pinaya, W.H., Tudosiu, P.D., Nachev, P., Ourselin, S., Cardoso, M.J.: Denoising diffusion models for out-of-distribution detection. arXiv preprint arXiv:2211.07740 (2022)
11. Guénais, T., Vamvourellis, D., Yacoby, Y., Doshi-Velez, F., Pan, W.: Bacoun: Bayesian classifers with out-of-distribution uncertainty. arXiv preprint arXiv:2007.06096 (2020)
12. Hendrycks, D., Gimpel, K.: A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv:1610.02136 (2016)
13. Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., Cohen-Or, D.: Prompt-to-prompt image editing with cross attention control. arXiv preprint arXiv:2208.01626 (2022)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
15. Lee, D., Yu, S., Yu, H.: Multi-class data description for out-of-distribution detection. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1362–1370 (2020)
16. Liu, L., Ren, Y., Lin, Z., Zhao, Z.: Pseudo numerical methods for diffusion models on manifolds. arXiv preprint arXiv:2202.09778 (2022)
17. Margatina, K., Baziotis, C., Potamianos, A.: Attention-based conditioning methods for external knowledge integration. arXiv preprint arXiv:1906.03674 (2019)
18. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: Sdedit: Guided image synthesis and editing with stochastic differential equations. In: International Conference on Learning Representations (2021)
19. Nalisnick, E., Matsukawa, A., Teh, Y.W., Gorur, D., Lakshminarayanan, B.: Do deep generative models know what they don't know? arXiv preprint arXiv:1810.09136 (2018)
20. Rebain, D., Matthews, M.J., Yi, K.M., Sharma, G., Lagun, D., Tagliasacchi, A.: Attention beats concatenation for conditioning neural fields. arXiv preprint arXiv:2209.10684 (2022)
21. Ren, J., Liu, P.J., Fertig, E., Snoek, J., Poplin, R., Depristo, M., Dillon, J., Lakshminarayanan, B.: Likelihood ratios for out-of-distribution detection. Advances in neural information processing systems **32** (2019)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
23. Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3379–3388 (2018)
24. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022)
25. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings. pp. 146–157. Springer (2017)
26. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
27. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

28. Wald, Y., Feder, A., Greenfeld, D., Shalit, U.: On calibration and out-of-domain generalization. Advances in neural information processing systems **34**, 2215–2227 (2021)
29. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 650–656 (2022)
30. Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., et al.: Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In: Proceedings of the 2018 world wide web conference. pp. 187–196 (2018)
31. Yang, J., Zhou, K., Li, Y., Liu, Z.: Generalized out-of-distribution detection: A survey. arXiv preprint arXiv:2110.11334 (2021)
32. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796 (2022)
33. Zhou, Y.: Rethinking reconstruction autoencoder-based out-of-distribution detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7379–7387 (2022)
34. Zhou, Z., Guo, L.Z., Cheng, Z., Li, Y.F., Pu, S.: Step: Out-of-distribution detection in the presence of limited in-distribution labeled data. Advances in Neural Information Processing Systems **34**, 29168–29180 (2021)