

# BeFOI: A Novel Method Based on Conditional Diffusion Model for Medical Image Denoising

Huijie Hu<sup>1,2\*</sup>, Zhen Huang<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, Anhui Province, China

<sup>2</sup>Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215000, Jiangsu Province, China

\*Corresponding author: Huijie Hu, [huhuijie123@mail.ustc.edu.cn](mailto:huhuijie123@mail.ustc.edu.cn)

**Copyright:** © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

**Abstract:** The progress in medical imaging technology highlights the importance of image quality for effective diagnosis and treatment. Yet, noise during capture and transmission can compromise image accuracy and reliability, complicating clinical decisions. The rising interest in diffusion models has led to their exploration of denoising images. We present BeFOI (Better Fluoro Images), a weakly supervised model that uses cine images to denoise fluoroscopic images, both DR types. Trained through precise noise estimation and simulation, BeFOI employs Markov chains to denoise using only the fluoroscopic image as guidance. Our tests show that BeFOI outperforms other methods, reducing noise and enhancing clarity and diagnostic utility, making it an effective post-processing tool for medical images.

**Keywords:** Diffusion model; Denoising; Medical images

**Online publication:** March 30, 2024

## 1. Introduction

X-ray Digital Radiography (DR) utilizes digital sensors to capture X-ray images<sup>[1]</sup>, allowing for a non-invasive initial assessment of patients' conditions and lesions. This technology is crucial in the diagnosis of various diseases and is particularly important in cardiac coronary imaging<sup>[2]</sup>, where arterial diseases can lead to serious heart problems. With concerns over the potential cancer risks from ionizing radiation, low-dose imaging has become a prevalent yet challenging approach due to the increased noise<sup>[3]</sup> levels in the images.

Recent advancements in hardware technology and deep learning have led to significant progress in the field of image noise reduction using Convolutional Neural Networks (CNN). CNNs can automatically learn and extract features from images to effectively suppress noise in low-dose X-ray images<sup>[4]</sup>, thereby enhancing their diagnostic quality and minimizing radiation exposure. These high-quality images are essential for complex analyses such as segmentation, Computer-Aided Diagnosis (CAD)<sup>[5]</sup>, and reconstruction.

However, the application of deep learning denoising algorithms in medical imaging presents certain challenges<sup>[6]</sup>. Firstly, large amounts of labeled data are required for training, which can be costly and difficult

to obtain due to privacy constraints. Secondly, the denoising process may inadvertently introduce artifacts that could compromise the diagnostic precision. Therefore, despite its potential, deep learning denoising must continue to evolve to better meet clinical demands and mitigate any adverse impacts.

Our study presents a conditional diffusion model that effectively denoises fluoroscopic images using weakly supervised learning and simulated noise, thus overcoming the need for paired datasets. This approach also incorporates generative models to enhance image quality, preserving details in digital radiography images.

## 2. Related work

### 2.1. Image denoising

Signal denoising is essential due to inevitable noise contamination. Common approaches include neighborhood averaging with Gaussian kernels [7], which can be biased near contours or patterns. Median filtering [8] replaces pixel values with medians, while Fourier transform denoising [9] exploits the separation of images and noise in different frequency bands. Wavelet denoising [10] processes wavelet coefficients using a threshold and reconstructs the signal. Non-local mean denoising [11] calculates a weighted average based on similar neighborhood structures. The BM3D algorithm [12] combines non-local and frequency domain methods for effective denoising. While traditional algorithms are fast, they have limited effectiveness on diverse medical image noise and fail to address low contrast and detailed information in medical images.

Nowadays, deep learning methods are leading in image denoising [13], with supervised CNNs [14] and blind denoising networks like FFDNet [15]. However, this method is far from practical because it is almost impossible to obtain clean and noisy images in pairs. To circumvent this difficulty, unsupervised and self-supervised methods [16-18] have been introduced subsequently. Noise2Self [19] proposes a blind image denoising method based on a generative adversarial network aimed at improving denoising effects without prior knowledge, but real images may not satisfy the requirement of J-invariance [20]. To address the issue, the GAN-based CNN blind denoiser GCBD [21] generates ground truth first and then inputs the obtained ground truth into the GAN to train the denoiser. However, there is a discrepancy between the distributions of the test and training data [22], and the training process for GANs is also prone to instability

### 2.2. DDPM

Denoising Diffusion Probabilistic Model (DDPM) [23] is a probabilistic generative model characterized by a T-step Markov chain, which aims to approximate a given data distribution  $q(x)$  with a model  $p_\theta(x)$ . It contains two processes: the forward diffusion process, where  $q(x)$  is diffused into a Gaussian noise distribution, and the reverse denoise process, where an image  $x \sim p_\theta(x)$  is generated from noise by step-by-step sampling.

#### (1) Forward process

The DDPM process involves progressively adding noise to the same data image until it becomes entirely random noise. Let  $x_0$  denote the data which is uncorrupted, we can derive the expressions for the marginal distribution of  $x_t$  when  $t$  is drawn from a uniform distribution, that is,  $\forall t \sim U(1, \dots, T)$ :

$$q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t} \cdot x_0, (1 - \bar{\alpha}_t) \cdot I), \quad (1)$$

Where Equation (2) demonstrates that we can sample any noisy version  $x_t$  in a single step by having the original image  $x_0$  and specifying a variance schedule  $\alpha_t$ .

## (2) Reverse process

By utilizing neural networks, we can learn parameterized Gaussian transitions  $p_\theta(x_{t-1}|x_t)$ , thereby facilitating the solution of the inverse process of Markov processes  $q(x_{t-1}|x_t)$ :

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 \mathbf{I}) \quad (2)$$

Where  $\mu_\theta(x_t, t)$  refers to the learned mean. Practically, one can relate  $x_t$  and  $x_0$  by decomposing  $\mu_\theta$  into a linear combination of  $x_t$  and the noise approximation  $\varepsilon_\theta$ . During sampling, we can use simple substitution to derive  $\mu_\theta(x_t, t)$  from network prediction  $\varepsilon_\theta(x_t, t)$

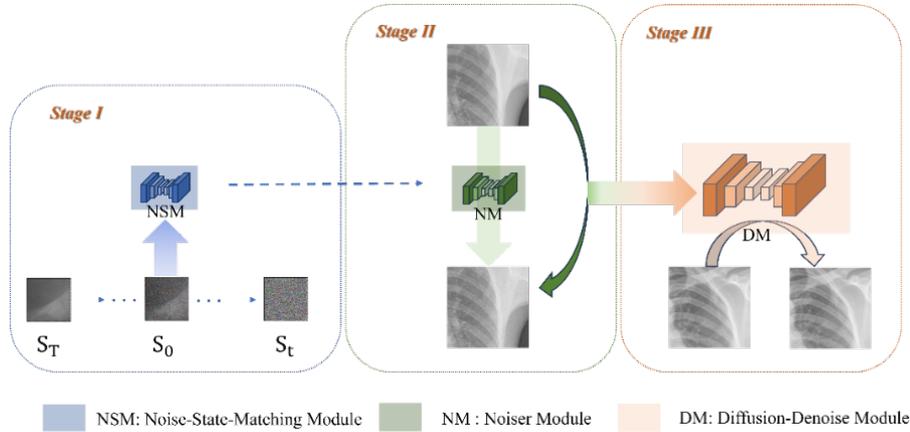
$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \varepsilon_\theta(x_t, t) \right) + \sigma_t z \quad (3)$$

Where  $z \sim N(0, \mathbf{I})$ . Since the model learns the reverse Markov Chain from  $x_t$  to  $x_0$ , which estimates clean image  $x_0$  from partially noisy image  $x_t$ , we refer to this as the reverse process.

## 3. Methodology

### 3.1. Framework

As shown in **Figure 1**, the overall framework of our study is built around the conditional diffusion model as the backbone. The framework first processes parameters through a Noise State Matching Module (NSM) and passes them to the Noiser Module (NM). The NM module performs noise simulation operations on cine images based on this. After completing the training of the conditional diffusion model, we input fluoro images as prior conditions into the diffusion model to guide it to generate the corresponding denoised images.



**Figure 1.** Overview of the network structure of BeFOI.

### 3.2. Conditional diffusion

When applying reference image-guided diffusion model generation techniques, the unconditional DDPM can be transformed into a conditional DDPM <sup>[24]</sup>. This means that a conditioning variable  $c$  is used to guide the generation process, enabling the model to produce images according to given conditional information. The process of obtaining the conditional distribution from the condition  $c$  can be expressed as follows (Equation [4]):

$$p_\theta(x_{0:T} | c) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t, c) \quad (4)$$

Here,  $p_\theta$  represents the probability distribution parameterized by  $\theta$ .  $X_T$  is the intermediate state image at time step  $t$  in the diffusion process,  $x_{t-1}$  is the image state at the previous time step, and  $c$  is the reference or conditional information.

In the conditional diffusion model, a reference image can be used to guide the generation process. Under the DDPM framework, the model first gradually adds noise to a clear image through a forward process of a Markov chain, resulting in a series of increasingly blurred states  $y_T, \dots, y_t, y_{t-1}, \dots$ , where  $T$  typically represents the time step with the maximum noise level.

In the transition from  $x_t$  to  $x_{t-1}$ , we can partially replace a part of the formula, thereby allowing the inverse Markov process to incorporate a conditioning term:

$$x'_{t-1} \sim p_\theta(x'_{t-1}|x_t) \quad (5)$$

$$x_{t-1} = \delta y_{t-1} + (1 - \delta)x'_{t-1} \quad (6)$$

Where  $\delta$  is a hyperparameter that represents the degree of reference of the reference image  $y$ . This means that in the process of transitioning from the current time step  $x_t$  to the previous time step  $x_{t-1}$ , in addition to considering the current noisy image, the information of the reference image  $y$  is also utilized to guide the generation of an image closer to  $y$ . The entire reverse process thus takes place under the condition of the given  $y$ , leading to results that match the desired target image features more closely.

### 3.3. Noise state matching & noise simulation

A specific timestamp  $t$  is obtained through optimization methods that minimize the distance between the state at that moment and a given noisy observation  $x$ :

$$\arg \min_t \|\sqrt{\alpha_t} - \theta\|^p \quad (7)$$

Where  $P$  represents the  $p$ -norm distance. The optimal matching state has been obtained at this point. In other words, under the condition of following a preset noise schedule, for any given noisy input  $x$ , there theoretically exists a posterior state in which characteristics are close to the input  $x$ , that is, it has a sufficiently low distance metric.

The capability of neural networks to perform erosion simulations on cine images is crucial in making them as close as possible to the fluoroscopic image pattern. This process can be understood through the following formula:

$$y_c = G(t, y) \quad (8)$$

Where  $G$  represents the noise simulation network, and  $t$  is obtained from NSM. Building upon this process,  $y_c$  is fed into a conditional diffusion model. From this, the loss is derived as Equation (9).

$$L_{BeFOI} = |F(G(NSM(z), y)) - y|, \quad (9)$$

where  $F$  denotes the diffusion model and  $NSM(z)$  signifies the noise simulation mapping for a given input  $z$ . The training procedure can generally be divided into three stages. Once a stable diffusion model neural network has been achieved, during the inference phase, one only needs to use the fluoroscopic mode image as a condition.

## 4. Experiments and results

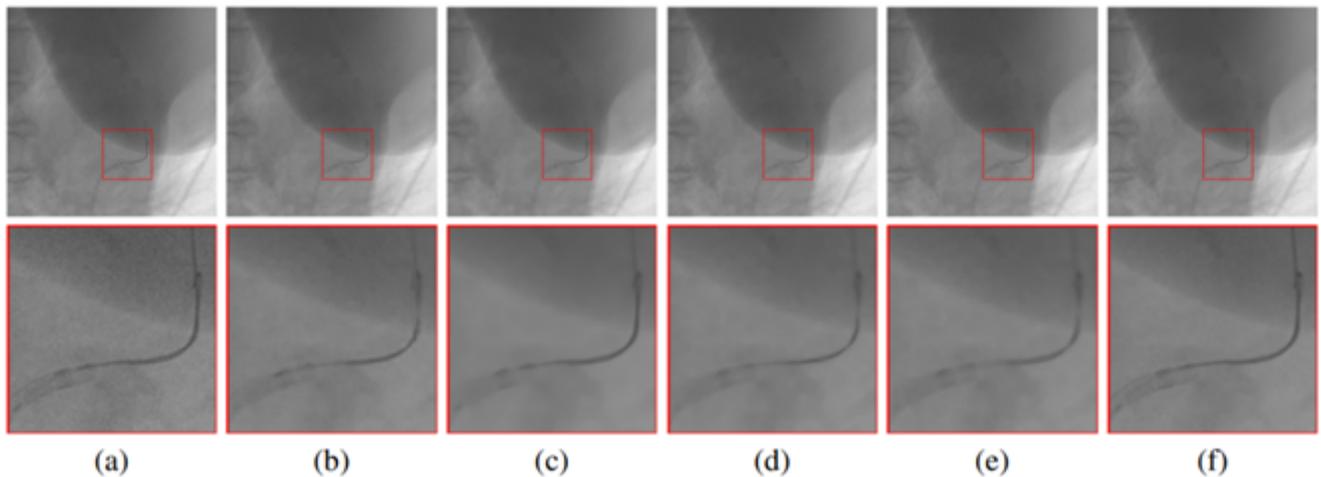
### 4.1. Setup

We tested our denoising method on low-quality X-ray images from the CoroArt dataset, which has 231 coronary artery segment images. We compared it with other methods like BM3D and Noise2Self, using their original codes and recommended settings. Training was done on a GeForce RTX 3090 GPU. We prepared the data by segmenting them into patches and adjusting image window levels for clearer analysis.

### 4.2. Comparisons

Qualitative results are shown in **Figure 2**. While all methods demonstrated considerable denoising effects, the denoising outcomes of certain algorithms manifested excessive smoothing due to the unknown intensity and type of noise, resulting in the loss of fine details and excessive blurring of underlying anatomical structures. Conversely, other algorithms failed to thoroughly eliminate noise, leaving noticeable noise patches in their results. In terms of algorithmic performance, this algorithm deliberately focused on preserving edge structures within images, leading to artifacts that were inconsistent with the true image, despite a more thorough removal of noise.

Quantitative results are listed below in **Table 1**.



**Figure 2.** Visual comparison of qualitative results from different denoising algorithms. The second row magnifies the details within the red box. (a) Noisy fluoroscopic images from the real world; (b)-(e) Results obtained from BM3D, NLM, Noise2self, and GCBD, respectively; (f) Our result.

**Table 1.** Quantitative image dehazing results on the CoroArt datasets

Models	No-GT		Via-GT	
	SNR $\uparrow$	REBLUR $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$
BM3D	27.7346	0.41	36.4513	0.9246
NLM	30.7884	0.37	35.7825	0.9329
Noise2self	33.0014	0.31	36.3523	0.9421
GCBD	34.0925	0.36	36.6117	0.9511
BeFOI (ours)	34.2575	0.30	36.7802	0.9573

Note: The best values for each metric are marked; “ $\downarrow$ ” means the lower the better and “ $\uparrow$ ” means the higher the better.

Due to the lack of ground truth information in our research data, and since we designed the denoising algorithm specifically for such reference-free conditions, we employed reference-free evaluation metrics for performance assessment. Initially, in the calculation of SNR, we selected a rectangular region of interest within the image to be evaluated, measuring the SNR by calculating the ratio of the mean pixel value to its standard deviation within that region. We introduced another reference-free evaluation method – Reblur technique. This method involves applying a secondary blurring process to the image, followed by using a clarity evaluation algorithm to compare the changes before and after the blur.

In order to maintain consistency with the commonly used reference-based evaluation systems, we also collected additional cine images not used for training and their corresponding noise simulation data as pseudo ground truth to calculate the algorithm’s performance values on the traditional image quality evaluation metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The experimental results show that the BeFOI algorithm proposed in this paper performs optimally across all evaluation metrics. Overall, the deep learning algorithms outperformed traditional denoising methods in this study. However, it is noteworthy that although traditional algorithms approach the performance of deep learning algorithms in terms of PSNR, this usually comes at the expense of sacrificing image detail and producing overly smooth results.

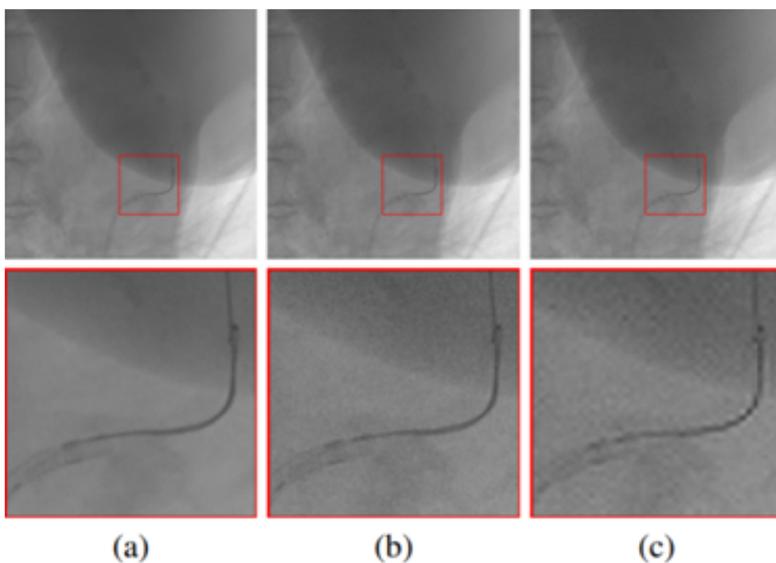
### 4.3. Ablation study

We removed noise estimation from our model and found that it failed to accurately simulate fluoroscopic image noise, leading to poor denoising results, as shown in **Figure 3**. The conditional diffusion model’s reliance on accurate noise makes our module essential. Additionally, using smoothed fluoroscopic images as training conditions did not yield good results, proving the need for another module.

**Table 2.** Results of the ablation study on CoroArt

Models	Distortion	
	PSNR↑	SSIM↑
NSM+NM	36.7802	0.9573
w/o NSM	36.2901	0.9252
w/o NM	35.3531	0.8965

Note: “↓” means the lower the better and “↑” means the higher the better.



**Figure 3.** Comparison of visual results from ablation study. (a) NSM +NM (full model). (b) w/o NSM: Removed noise estimation and matching (Noise-State-Matching Module) from stage I. (c) w/o NM: Removed noise simulation (Noiser Module) from stage II.

## 5. Retrospect and prospect

Our research presents BeFOI, a method for denoising DR images using a weakly supervised conditional diffusion model, aided by cine images for misaligned fluoroscopy. We begin with noise estimation for level mapping and simulation, followed by training the model. Denoising is achieved through Markov chains using the fluoro image. Results indicate notable noise reduction, clarity, and diagnostic enhancement, benefiting postprocessing tasks like vessel extraction. Our tests on a DR dataset show promising outcomes. We intend to apply the processed data to segmentation and lesion classification, sensitive to noise, and explore its use with other modalities like CT and MRI.

### Author contributions

*Conceptualization:* Hujie Hu

*Formal analysis:* Hujie Hu

*Writing – original draft:* Hujie Hu

*Writing – review & editing:* Zhen Huang

### Disclosure statement

The authors declare no conflict of interest.

### References

- [1] Li Y, Sixou B, Peyrin F, 2021, A Review of the Deep Learning Methods for Medical Images Super Resolution Problems, *IRBM*, 42(2): 120–133.
- [2] Cheung WK, Bell R, Nair A, et al., 2021, A Computationally Efficient Approach to Segmentation of the Aorta and Coronary Arteries using Deep Learning. *IEEE Access*, 9: 108873–108888.
- [3] Kascenas A, Sanchez P, Schrepf P, et al., 2023, The Role of Noise in Denoising Models for Anomaly Detection in Medical Images. *Medical Image Analysis*, 90: 102963.
- [4] Ilesanmi AE, Ilesanmi TO, 2021, Methods for Image Denoising using Convolutional Neural Network: A Review, *Complex & Intelligent Systems*, 7(5): 2179–2198.
- [5] Li X, Chen L, Rahaman MM, et al., 2022, A Comprehensive Review of Computer-Aided Whole-Slide Image Analysis: From Datasets to Feature Extraction, Segmentation, Classification and Detection Approaches. *Artificial Intelligence Review*, 55(6): 4809–4878.
- [6] Zhou SK, Greenspan H, Davatzikos C, et al., 2021, A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises. *Proceedings of the IEEE*, 109(5): 820–838.
- [7] Wink AM, Roerdink JB, 2004, Denoising Functional MR Images: A Comparison of Wavelet Denoising and Gaussian Smoothing. *IEEE Transactions on Medical Imaging*, 23(3): 374–387.
- [8] Li X, Ji J, Li J, et al., 2021, 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), June 18–20, 2021: Research on Image Denoising Based on Median Filter. *Chongqing*, 528–531.
- [9] Bluestein L, 1970, A Linear Filtering Approach to the Computation of Discrete Fourier Transform. *IEEE Transactions on Audio and Electroacoustics*, 18(4): 451–455.

- [10] Ouahabi A, 2013, 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), May 12–15, 2013: A Review of Wavelet Denoising in Medical Imaging, Algiers, 19–26.
- [11] Xu M, Lv P, Li M, et al., 2016, Medical Image Denoising by Parallel Non-Local Means. *Neurocomputing*, 195: 117–122.
- [12] Danielyan A, Katkovnik V, Egiazarian K, 2011, BM3D Frames and Variational Image Deblurring. *IEEE Transactions on Image Processing*, 21(4): 1715–1728.
- [13] Tian C, Fei L, Zheng W, et al., 2020, Deep Learning on Image Denoising: An Overview. *Neural Networks*, 131: 251–275.
- [14] Tian C, Xu Y, Li Z, et al., 2020, Attention-Guided CNN for Image Denoising. *Neural Networks*, 124: 117–129.
- [15] Zhang K, Zuo W, Zhang L, 2018, FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising, *IEEE Transactions on Image Processing*, 27(9): 4608–4622,
- [16] Lehtinen J, Munkberg J, Hasselgren J, et al., Noise2Noise: Learning Image Restoration without Clean Data. *arXiv*. <https://doi.org/10.48550/arXiv.1803.04189>
- [17] Krull A, Vičar T, Prakash M, et al., 2020, Probabilistic Noise2void: Unsupervised Content-Aware Denoising. *Frontiers in Computer Science*, 2: 5.
- [18] Ulyanov D, Vedaldi A, Lempitsky V, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18–23, 2018: Deep Image Prior. Salt Lake City, 9446–9454.
- [19] Batson J, Royer L, 2019, Proceedings of the 36th International Conference on Machine Learning Noise2self, June 10–15, 2019: Blind Denoising by Self-Supervision, 524–533.
- [20] Xie Y, Wang Z, Ji S, 2020, Optimizing a Self-Supervised Bound for Image Denoising. *arXiv*. <https://doi.org/10.48550/arXiv.2010.11971>
- [21] Chen J, Chen J, Chao H, et al., 2018, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18–23, 2018: Image Blind Denoising with Generative Adversarial Network Based Noise Modeling, 3155–3164.
- [22] Arjovsky M, Bottou L, 2017, Towards Principled Methods for Training Generative Adversarial Networks. *arXiv*. <https://doi.org/10.48550/arXiv.1701.04862>
- [23] Ho J, Jain A, Abbeel P, 2020, Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, 33: 6840–6851.
- [24] Choi J, Kim S, Jeong Y, et al., 2021, ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.02938>

**Publisher's note**

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.