Denoising Diffusion Probabilistic Models for DBT data augmentation: preliminary results

Lorenzo D'Errico[1], Lorenzo Pergamo[2], Daniel Riccio[1] and Mariacarla Staffa[2,*]

[1]University of Naples Federico II
[2]University of Naples Parthenope

Abstract

Recent strides in computer vision have led to promising breakthroughs in the realm of image generation. Notably, diffusion probabilistic models such as DALL-E 2, Imagen, and Stable Diffusion have demonstrated the ability to create lifelike images based on textual prompts. Yet, their potential application in the medical domain, where intricate three-dimensional image volumes are commonplace, remains largely untapped. Synthetic imagery presents a compelling avenue in the realm of privacy-preserving artificial intelligence and holds immense potential for enriching datasets with limited samples. This study seeks to assess the effectiveness of diffusion probabilistic models in synthesizing high-fidelity medical imaging data, with a particular focus on Digital Breast Tomosynthesis (DBT) images.

Keywords

Denoising Diffusion Probabilistic Models (DDPMs), Breast Tomosynthesis (DBT), Generative Models

1. Introduction

The success of deep learning across various pattern recognition tasks has ignited widespread excitement and elevated expectations regarding its potential impact on healthcare [1]. Concurrently, Digital Breast Tomosynthesis (DBT) has emerged as a transformative technology in breast cancer screening and diagnosis. Since its clinical debut in 2011, radiologists specializing in breast disease diagnosis nationwide have increasingly adopted this innovative approach for both screening and diagnostic purposes, with its adoption steadily rising [2]. The convergence of DBT and AI presents significant promise, offering opportunities for heightened precision, efficiency, and overall advancements in breast cancer screening and diagnosis. As the healthcare landscape evolves, the fusion of DBT and AI holds the potential to revolutionize breast cancer detection and management [3]. Integrating deep learning algorithms with DBT data could lead to more accurate and timely identification of abnormalities, thereby improving patient outcomes. Despite the optimism surrounding this new era of machine learning, the development and implementation of AI tools in clinical settings encounter numerous challenges [2]. Deep artificial neural networks necessitate substantial training data to effectively learn, a requirement that often proves costly and labor-intensive to fulfill. This challenge is particularly pertinent for digital breast tomosynthesis (DBT), which represents a relatively novel breast cancer screening modality. Data augmentation offers a solution by artificially expanding the training set through label-preserving transformations [4]. This study aims to leverage Denoising Diffusion Probabilistic Models (DDPMs), a class of generative models, to generate synthetic samples for Digital Breast Tomosynthesis (DBT). DDPMs have garnered significant attention across various domains for their ability to produce synthetic data of exceptional quality. These models function by iteratively introducing noise to an input signal, such as an image, text, or audio, and then learning the denoising process to generate novel samples. In the realm of image synthesis, DDPMs have demonstrated success in generating authentic and high-quality images, bolstered by competitive log-likelihoods that attest to their effectiveness in diverse generative tasks [5]. The overarching objective is to address the scarcity and imbalance in existing datasets, thereby enhancing the quality of deep learning algorithms, particularly those related to segmentation and detection. The proposed methodology entails using synthetic DBT samples as a form of data augmentation to mitigate the constraints associated with current dataset availability. This augmentation strategy is expected to contribute to the refinement of deep learning algorithms, ultimately driving advancements in segmentation and detection algorithms within the context of DBT. The study is structured as follows: in Section 2 an overview of the DBT technology as well as of DDPM models it's given; Section 3 delves into the results obtained and finally in Section 4 future perspectives are discussed.

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy
*Corresponding author.
†These authors contributed equally.
lorenzo.derrico@unina.it (L. D'Errico); lorenzo.pergamo001@studenti.uniparthenope.it (L. Pergamo); daniel.riccio@unina.it (D. Riccio); mariacarla.staffa@uniparthenope.it (M. Staffa)
0000-0001-8044-8224 (L. D'Errico); 0000-0002-5844-0602 (D. Riccio); 0000-0001-7656-8370 (M. Staffa)

2. Matherial and Methods

2.1. Digital Breast Tomosynthesis

Full-field digital mammography (FFDM) has traditionally been the primary breast cancer screening method, but its effectiveness is hindered by inherent limitations. Visualizing complex breast structures in two dimensions often leads to obscured tumor margins and inaccurate lesion characterization due to overlapping tissue [6]. Standard imaging projections may not capture the full extent of irregular or multifocal tumors [7], further complicating accurate diagnosis. Variations in breast composition, positioning artifacts, and tissue compression during imaging introduce variability into tumor size estimation and localization. In contrast, Digital Breast Tomosynthesis (DBT) (see Fig.1), approved by the FDA in 2011, revolutionizes breast imaging by acquiring a series of low-dose X-ray images from multiple angles and reconstructing them into a 3D dataset [8]. This enables radiologists to

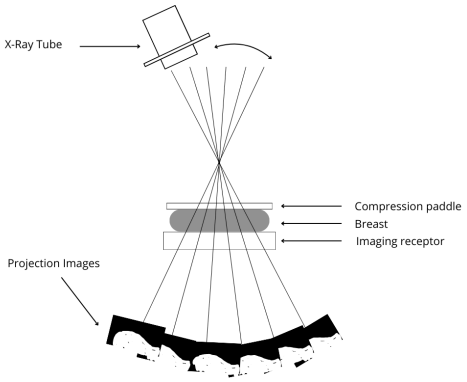

Figure 1: Digital Breast Tomosynthesis procedure.

navigate breast tissue in three dimensions, overcoming the limitations of 2D mammography [9]. DBT enhances lesion visualization, improves diagnostic accuracy, and outperforms FFDM in detecting invasive cancers and architectural distortions [10] (see Fig.2). Advanced reconstruction algorithms and image processing techniques further enhance DBT's diagnostic utility, allowing for the detection of smaller lesions with greater confidence [12]. DBT reduces false positives, minimizes unnecessary recalls, and optimizes patient outcomes by providing clearer, more detailed images. Integration of quantitative imaging biomarkers and machine learning algorithms augments DBT's diagnostic capabilities, ushering in personalized breast cancer screening and management. In conclusion, DBT represents a transformative advancement in breast cancer imaging, promising unparalleled diagnostic accuracy and improved patient outcomes. As research and technology progress, DBT is poised to revolutionize breast healthcare delivery by facilitating early detection and treatment of breast cancer.

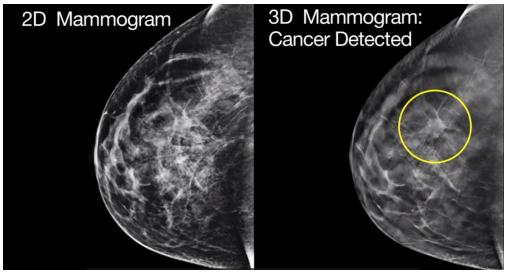

Figure 2: Comparison between a 2D mammogram and a 3D one. In Digital Breast Tomosynthesis (right), tumors are detected, unlike in mammography (left) where tissue overlap obstructs the view of the specialist doctor [11].

2.2. Denoising Diffusion Probabilistic Models

Diffusion models represent an advanced category of generative models renowned for their efficacy in capturing intricate data distributions. Despite being a recent addition to the generative learning field, they have proven valuable across diverse applications. The three dominant generative frameworks are identified as Generative Adversarial Networks (GANs) [13], Variational Autoencoders (VAEs) [14], and normalizing flows [15]. These models, falling under the category of probabilistic generative models, are adept at capturing intricate data distributions, establishing themselves as a formidable tool in various applications. A Denoising Diffusion Probabilistic Model (DDPM) is a parameterized Markov chain trained using variational inference to produce samples matching the data after finite time (see Fig.3). DDPM are composed of two opposite processes, forward and reverse diffusion process. In the forward diffusion process, Gaussian noise is gradually and iteratively introduced to intentionally perturb the images within the training set, aiming to induce a transformation wherein they deviate from their current distribution and align more closely with a normal distribution. In the reverse diffusion process, the objective is to systematically invert the preceding forward diffusion procedure. The reversal is conducted gradually and iteratively to counteract the perturbation applied to images in the forward process. Starting where the forward process concludes, the advantage of initiating from a normal distribution lies in the known methodology for sampling points from this uncomplicated distribution. The primary aim is to discern the means to revert to the original data distribution. Nonetheless, the challenge arises from the potential for an infinite array of trajectories originating from a point in this ostensibly

simple space, with only a fraction leading to the data distribution. Within the context of DDPM, this is achieved by referencing the incremental steps undertaken in the forward diffusion process. The probability density function (PDF) corresponding to the corrupted images in the forward process exhibits slight variations at each step. Consequently, in the reverse process, a deep-learning model is employed at each step to prognosticate the PDF parameters of the forward process. Subsequent to model training, any point in the simple space can be selected, and the model can be utilized iteratively to navigate back to the data subspace. In reverse diffusion, denoising is systematically performed in small steps, commencing from a noisy image. This method of training and generating new samples is characterized by enhanced stability compared to Generative Adversarial Networks (GANs) and surpasses prior approaches such as Variational Autoencoders (VAE) and normalizing flows. Diffusion models, as outlined in the literature [16], are a category of latent variable models represented by the equation

$$p_\theta(x_0) := \int p_\theta(x_{0:T})\, dx_{1:T},$$

where $x_1, \ldots, x_T$ are latent variables of the same dimensionality as the data $x_0 \sim q(x_0)$. The joint distribution $p_\theta(x_{0:T})$ is denoted as the reverse process, constituting a Markov chain with learned Gaussian transitions that initiate at $p(x_T) = \mathcal{N}(x_T; 0, I)$:

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

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)).$$

Regarding the structural design of the model, it's noteworthy that the dimensions of both the input and output of the model should align. To achieve this objective, Ho et al. [16] utilized a U-Net architecture, thereby ensuring compatibility in size between the input and output components of the model. From the typical UNet architecture, the conventional double convolution at each level was replaced with Residual blocks as employed in ResNet models. In the DDPM implementation, a Wide ResNet block was employed as per Zagoruyko et al. [17]. However, in the adaptation by Phil Wang, the standard convolutional layer was replaced with a weight standardized version, recognized for its improved performance in conjunction with group normalization as outlined by Kolesnikov et al. [18]. Moreover, in order to maintain parameter consistency across various time instances, sinusoidal position embeddings are incorporated, drawing inspiration from the Transformer model [19]. This integration facilitates the neural network in discerning the relevant time step (noise level) for each image within a batch. The SinusoidalPositionEmbeddings module has been used in this work. Finally, an attention module is introduced from the Transformer architecture [19, 20].

3. Experimental Results

This section presents a detailed analysis of the experimental setup employed for training and evaluating the DDPM. The results are presented in a structured manner, showcasing the model's ability to capture and simulate the intricate features present in authentic DBT images. Additionally, we explore the impact of key hyperparameters on the synthesis process.

3.1. Description of Dataset

The dataset comprises patient records from individuals who underwent Digital Breast Tomosynthesis (DBT) examinations at the Duke Health system between January 1, 2014, and January 30, 2018. The acquisition process involved cross-referencing information from radiology reports, pathology reports, and DBT data obtained from the Picture Archiving and Communication Systems (PACS) at Duke. These studies encompassed a total of 13,954 unique patients, each with at least one craniocaudal (CC) and mediolateral oblique (MLO) view available for either the left or right breast (see Fig.4).

The dataset is organized into three sets: a training set comprising 1.42 TB, a validation set comprising 84.71 GB, and a test set comprising 135.14 GB. The images in the dataset are in a DICOM format and were processed using the torchio library for reading. To make them consistent and ready for analysis and research, all the pictures were reshaped to dimensions of 64x64 pixels with 8 slices. The dimensions of the images have been systematically reduced through an iterative process aimed at preserving the utmost quality of the visual content. This iterative approach has been employed with the primary objective of maintaining the highest possible image quality while undergoing size reduction. This change is not just for analysis, but it also enhance computational efficiency.

3.2. Hardware

The principal objective of the present research is centered on the generation of synthetic samples in Digital Breast Tomosynthesis (DBT) through the application of a sophisticated Denoising Diffusion Probabilistic Model. To optimize the computational procedures inherent in this complex task, the foundational code underwent a process of parallelization on four Tesla V100 Graphics Processing Units (GPUs). It is imperative to note that the parallelization strategy employed pertained specifically to the data level, signifying that the dataset was effectively partitioned and processed concurrently across all GPUs. This strategic approach played a pivotal role in amplifying both the efficiency and expeditiousness of the model training and synthetic sample generation,

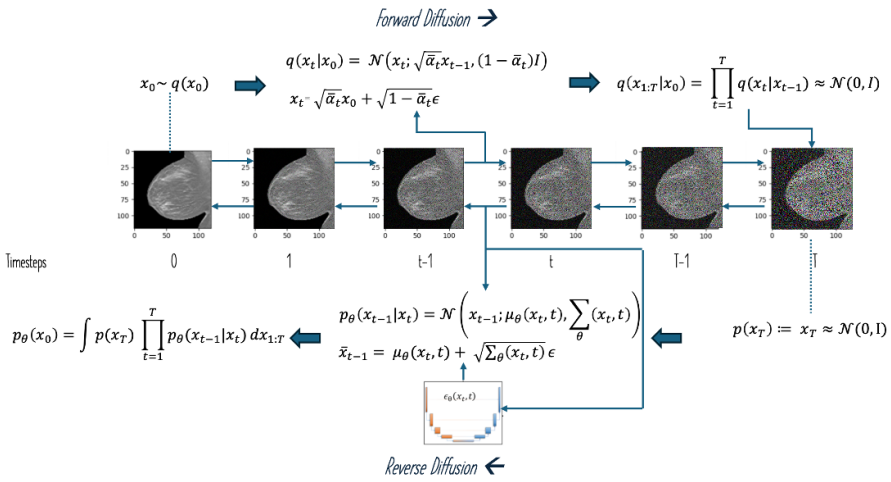

Figure 3: DDPM Architecture.

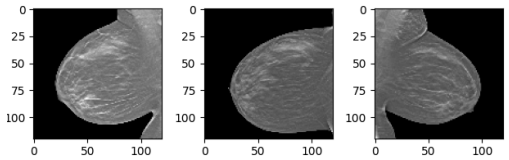

Figure 4: Example slices from the dataset.

thereby significantly augmenting the overall efficacy of the research.

3.3. Model deployed

Hyperparameters

- *Batch Size.* A judiciously determined batch size of 16 was allocated for each of the four Graphics Processing Units (GPUs) employed. This strategic selection was grounded in the quest for an optimal compromise among critical factors such as training speed, output quality, and, notably, memory utilization. The careful consideration of these factors was essential in achieving a harmonious balance that not only facilitated expeditious model training but also ensured the preservation of high-quality outcomes while efficiently managing the computational memory resources. This particular batch size allocation emerged as the most effective compromise, aligning with the overarching objectives of the experiment and contributing to the overall success of the research endeavor.
- *Learning Rate.* It was observed that the adoption of the smallest learning rate, 1e-4, conferred superior outcomes. This discernment underscores the model's susceptibility to nuanced parameter adjustments, wherein diminutive updates within the parameter space correlated with enhanced performance. The adaptive characteristics inherent to the Adam optimizer, which dynamically adjusts learning rates based on historical gradients, likely played a pivotal role in the efficacy of this minimal learning rate.

Loss function In the work, a deliberate choice was made to employ the Mean Squared Error (MSE) as the primary loss function. This decision was founded upon the premise of calculating the disparity between the noise introduced to images and the corresponding noise predictions generated by the UNet model. However, subsequent empirical investigations, coupled with insights gleaned from alternative implementations, have indicated the potential for the utilization of the Mean Absolute Error (MAE) to yield superior outcomes. These findings prompt a reevaluation of the chosen loss function, necessitating a thorough exploration of the implications associated with the adoption of MAE within the framework of the study's objectives. Such a revision stands to enhance the efficacy and fidelity of the model's predictive capabilities, thereby warranting comprehensive investigation and validation within the context of the research endeavor.

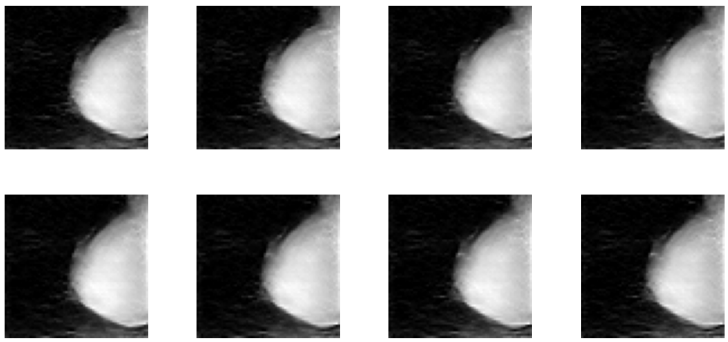

Figure 5: Example of Generated DBT 1.

3.4. Sample quality

While it is evident that the current quality of our samples may not meet the stringent standards required for certain applications (see Fig.5), it is essential to recognize the promising aspects of the results. Despite the imperfections, the dataset produced presents a valuable foundation upon which improvements can be built. Additionally, it is noteworthy that upon examination of the histogram (see Fig.6), the generated samples exhibit lower contrast compared to the desired standards. This contrast deficiency is a significant aspect that requires attention to ensure that the generated images meet the necessary quality thresholds for clinical applications. Contrary to

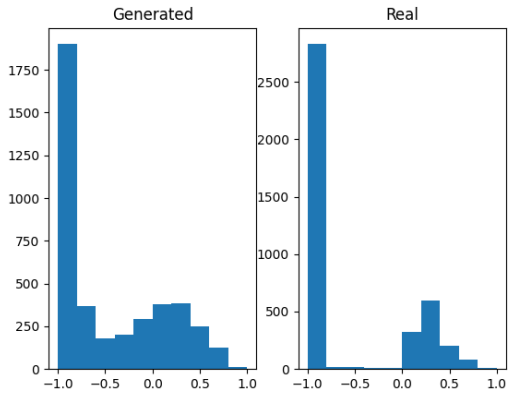

Figure 6: Pixel distribution.

the histogram relative to the generated images, the one related to real images displays a distribution of pixel intensities characterized by the accumulation of values around two distinct modes. This suggests the presence of two dominant intensity regions within the image. The separation and distribution of these modes can significantly influence the visual characteristics of the image.

4. Discussion and Conclusion

Through meticulous experimentation and analysis, the study demonstrated the capability of DDPM to produce artificial DBT images that closely emulate the intricacies of real-world cases. The promising results obtained pave the way for future investigations and applications of DDPM in the realm of medical imaging. The potential for refining and expanding upon these generative models opens avenues for further research, contributing to the ongoing evolution of DBT technology. Future application may involve the conditioning of the sampling procedure, allowing for the deliberate manipulation of the generated samples. In this context, this phenomenon is alternatively denoted as guided diffusion [21], [22]. Moreover, Latent diffusion models could be introduced. In these models an initial step involves projecting the input into a more compact latent space, where the diffusion process is subsequently applied. To elaborate further, Rombach et al. [23] proposed the utilization of an encoder network, denoted as $g(x_t) = z_t$, to encode the input into a latent representation z_t. This strategic choice aims to alleviate the computational demands associated with training diffusion models by conducting processing in a lower-dimensional space. Following this encoding, a conventional diffusion model, specifically a U-Net, is employed to generate new data. The resultant data are then upsampled through a decoder network.

5. Acknowledgments

The work was supported by the PNRR MUR project PE0000013-FAIR.

References

[1] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, C. Zhou, Deep learning in medical image analysis, in: Advances in Experimental Medicine and Biology, Springer International Publishing, 2020.

[2] S. Destounis, Role of digital breast tomosynthesis in screening and diagnostic breast imaging (2018).

[3] F. D. Marco, A. A. Citarella, L. D. Biasi, L. D'Errico, R. Francese, G. Mettivier, M. Staffa, G. Tortora, Ai-based solutions for the analysis of biomedical images and signals, in: Proceedings of the Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023, volume 3486 of *CEUR Workshop Proceedings*, 2023, pp. 171–176.

[4] L. Taylor, G. Nitschke, Improving deep learning with generic data augmentation, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, pp. 1542–1547.

[5] A. Q. Nichol, P. Dhariwal, Improved denoising diffusion probabilistic models, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 8162–8171.

[6] D. Förnvik, S. Zackrisson, O. Ljungberg, T. Svahn, P. Timberg, A. Tingberg, I. Andersson, Breast tomosynthesis: accuracy of tumor measurement compared with digital mammography and ultrasonography, Acta radiologica 51 (2010) 240–247.

[7] B. D. Fornage, O. Toubas, M. Morel, Clinical, mammographic, and sonographic determination of preoperative breast cancer size, Cancer 60 (1987) 765–771.

[8] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, et al., Digital tomosynthesis in breast imaging, Radiology 205 (1997) 399–406.

[9] N. P. Tirada, G. Li, D. Dreizin, L. Robinson, G. R. Khorjekar, S. A. Dromi, T. S. Ernst, Digital breast tomosynthesis: Physics, artifacts, and quality control considerations., Radiographics : a review publication of the Radiological Society of North America, Inc 39 2 (2019) 413–426.

[10] I. Andersson, D. M. Ikeda, S. Zackrisson, M. Ruschin, T. Svahn, P. Timberg, A. Tingberg, Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammographic findings, European radiology 18 (2008) 2817–2825.

[11] M. Staffa, L. D'Errico, R. Ricciardi, P. Barra, E. Antignani, S. Minelli, G. Mettivier, How to increase and balance current dbt datasets via an evolutionary gan: preliminary results, in: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 2022, pp. 913–920.

[12] R. Moore, D. Kopans, E. Rafferty, D. Georgian-Smith, R. Hitt, E. Yeh, Initial callback rates for conventional and digital breast tomosynthesis mammography comparison in the screening setting, in: Radiological Society of North America 93rd Scientific Assembly and Annual Meeting, Nov, 2007.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144.

[14] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, S. Lima Netto, Variational autoencoder (2021) 111–149.

[15] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, D. Merhof, Diffusion models in medical imaging: A comprehensive survey, Medical Image Analysis 88 (2023) 102846.

[16] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, arXiv preprint arxiv:2006.11239 (2020).

[17] S. Zagoruyko, N. Komodakis, Wide residual networks, 2016. URL: http://arxiv.org/abs/1605.07146, cite arxiv:1605.07146.

[18] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, N. Houlsby, Big transfer (BiT): General visual representation learning (2019).

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need (2017).

[20] M. Staffa, P. Barra, Multi-monitor system for adaptive image saliency detection based on attentive mechanisms, in: H. Degen, S. Ntoa (Eds.), Artificial Intelligence in HCI - 4th International Conference, AI-HCI 2023, Held as Part of the 25th HCI International Conference, HCII, volume 14051 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 607–617.

[21] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, M. Chen, Glide: Towards photorealistic image generation and editing with text-guided diffusion models., in: ICML, volume 162 of *Proceedings of Machine Learning Research*, PMLR, 2022, pp. 16784–16804.

[22] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, M. Norouzi, Photorealistic text-to-image diffusion models with deep language understanding., CoRR abs/2205.11487 (2022).

[23] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models., in: CVPR, IEEE, 2022, pp. 10674–10685.