Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070102
ISSN 2371-8412 Vol. 7 Num. 1

Conditional Diffusion Model for X-Ray Segmentation Data Generation

Zehao Fang[1,a,*]

[1]*Shanghai Pinghe School, Shanghai, China*
[a]*fangzehao@shphschool.com*
[*]*Corresponding author*

Keywords: ControlNet, Diffusion Model, Synthetic Medical Images

Abstract: Nowadays training a well-functioning deep learning AI model requires a large amount of data, while in the field of medicine many scenarios lack training data due to privacy issues and legal reasons. In this essay, we propose to use ControlNet, a novel approach that leverages stable diffusion models and conditional control to produce realistic and diverse medical images. ControlNet allows us to specify extra conditions that the diffusion model should follow, such as edge maps, depth maps, segmentation masks, or CLIP image embeddings. These conditions can help us to preserve the structure, shape, and semantics of the target organs or tissues, as well as to manipulate the appearance, style, and context of the generated images. Specifically, we will use ControlNet to generate X-ray of a patient with pulmonary nodules and show the improvement.

1. Introduction

Recently AI has become particularly powerful and the application in other fields such as engineering [1], cybersecurity [2], motion analysis [3], etc. has made significant achievements. However, in the medical field, although AI shows great promise in improving diagnostic accuracy and efficiency [4], a key challenge facing the field is to obtain high-quality annotated datasets for training. In this paper, we propose an approach to address this problem by generating synthetic medical images using the neural network structure [5].

ControlNet enables fine-tuning diffusion models with additional conditions, without causing any distortion to the original model. By leveraging this technology, we can generate a diverse range of synthetic medical images that closely resemble real-world data. These images can serve as a valuable resource for training AI models, thereby overcoming the limitations of traditional datasets.

In general, the use of ControlNet for generating synthetic medical images opens up new avenues in the field of medical imaging. It holds the potential to revolutionize the way we train AI models, ultimately leading to more accurate and efficient diagnostic tools.

2. Literature Review

To generate synthetic medical images, methods like GAN [6] and VAE [7, 8] have been adopted [9, 10, 11, 12]. However, they face two main problems, the instability of training and lack of output

diversity [13]. Diffusion models [14, 15, 16] effectively solve these problems. However, in the field of medicine it lacks structural control to condition the image generation solely by text prompt.

In a recent study, Zhang et al. proposed ControlNet [5] as a modulation architecture, adding spatial conditioning controls to large, pre-trained text-to-image diffusion models [14]. The neural architecture is connected with “zero convolutions” [5] that progressively grow the parameters from zero and ensure that no harmful noise could affect the fine-tuning. The researchers tested various conditioning controls, such as edges, depth, segmentation, human pose, etc., with Stable Diffusion [10], using single or multiple conditions, with or without prompts. It is shown that the training of ControlNet is robust with both small (<50k) and large (>1m) datasets.

ControlNet has been used in various applications such as generating artistic images, medical images. In the study, it was used to generate medical images with greater precision and control. The researchers showed that ControlNet can be used to generate high-quality medical images with greater precision and control than traditional generative models.

3. Method

ControlNet is a neural network architecture which controls diffusion models by adding extra conditions. It can be used to generate images with greater precision and control by allowing users to add conditions such as canny edges and human poses.

The architecture of ControlNet consists of two main components: a diffusion model and a control network. The diffusion model is a pre-trained generative model that generates images from noise. The control network is a neural network that takes in an input image and a prompt and produces a synthesized image that matches the prompt and follows the constraints imposed by the input image.

The control network is trained using a combination of adversarial loss, perceptual loss, and feature matching loss. The adversarial loss ensures that the synthesized image is realistic, while the perceptual loss ensures that the synthesized image matches the prompt. The feature matching loss ensures that the synthesized image follows the constraints imposed by the input image.

4. Results

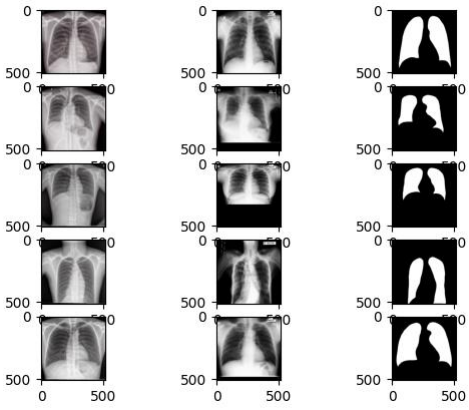

Figure 1: The ground truth, generated image and condition

Finally, the graphs we generated using ControlNet are comparable to the real graphs, the first column in Figure 1 are the ground truth and the second column are the images we generated using ControlNet. The PSNR of the generated image is 19.6, which indicates that the fidelity of the image generation process is sufficient for the initial training phase of the image recognition model in the medical domain. Moreover, the SSIM value of our images reaches 0.65, which indicates that the basic

structural features essential for medical diagnosis are well preserved in the generated images. Furthermore, Figure 2 shows qualitative evaluation of the model. Specifically, image (a) is the initial condition as the input to ControlNet, and images (b)-(l) are the validation images across different training steps. The results show that as the number of training steps increase, the quality of the generated images increase as well.

Our method shows significant improvement in the speed and scalability of image generation when compared to traditional dataset generation methods. The results show that the generated dataset can greatly help in training machine learning models, especially in the case of limited variety of medical images. While the achieved PSNR and SSIM values are promising, there is still room for improvement. Future research will focus on optimizing the control network architecture to enhance these metrics.

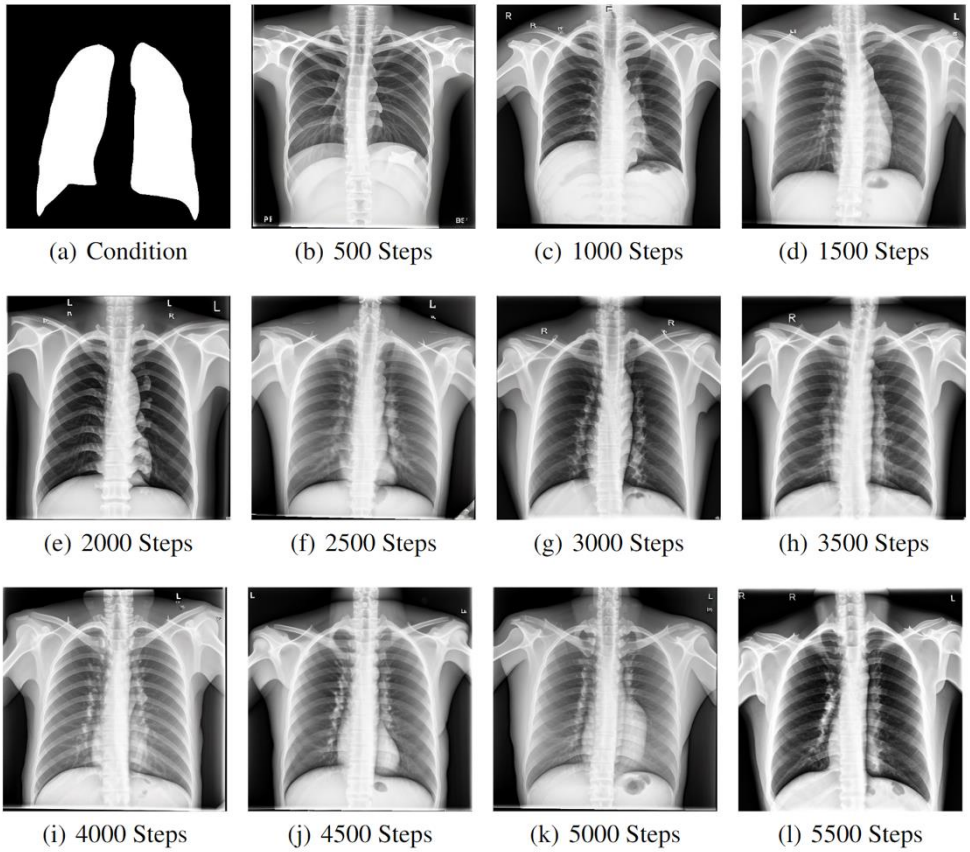

Figure 2: Validation results during training

5. Conclusion

In this essay, we have explored the potential of ControlNet in revolutionizing medical image generation. Our findings suggest that ControlNet can generate images with considerable precision and control, effectively mirroring real-world medical data. The proposed method not only improves the quality and diversity of synthetic medical images but also offers a scalable solution to rapidly expand training datasets for diagnostic AI.

Looking forward, optimizing ControlNet and exploring its full potential in medical imaging will be an essential step. As the technology matures, it could significantly impact various aspects of healthcare, from diagnostic accuracy to treatment planning and medical research. However, it's also critical to address any ethical and technical challenges associated with synthetic data to ensure that

the benefits of such advancements are realized responsibly and equitably.

References

[1] R. W. Blake, R. Mathew, A. George, and N. Papakostas, "Impact of artificial intelligence on engineering: Past, present and future," Procedia CIRP, vol. 104, pp. 1728–1733, 2021, 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827121011896

[2] A. Chakraborty, A. Biswas, and A. K. Khan, "Artificial intelligence for cybersecurity: Threats, attacks and mitigation," in Artificial Intelligence for Societal Issues. Springer, 2023, pp. 3–25.

[3] R. Zhou, H. Zhou, H. Gao, M. Tomizuka, J. Li, and Z. Xu, "Grouptron: Dynamic multi-scale graph convolutional networks for group-aware dense crowd trajectory forecasting," in 2022 International Conference on Robotics and Automation (ICRA), 2022, pp. 805–811.

[4] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," Journal of ambient intelligence and humanized computing, pp. 1–28, 2022.

[5] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3836–3847.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," CoRR, vol. abs/1312.6114, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:216078090

[8] B. Dai and D. P. Wipf, "Diagnosing and enhancing vae models," ArXiv, vol. abs/1903.05789, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:76666188

[9] Z. Ren, S. X. Yu, and D. Whitney, "Controllable medical image generation via generative adversarial networks," Electronic Imaging, vol. 33, no. 11, pp. 112–1–1126, Jan. 2021. [Online]. Available: http://dx.doi.org/10.2352/issn.2470-1173.2021.11.hvei-112

[10] Y.Skandarani,P.-M.Jodoin,andA.Lalande,"Gansformedicalimagesynthesis:Anempirical study," Journal of Imaging, vol. 9, no. 3, p. 69, 2023.

[11] F. Li, W. Huang, M. Luo, P. Zhang, and Y. Zha, "A new vae-gan model to synthesize arterial spin labeling images from structural mri," Displays, vol. 70, p. 102079, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938221000858

[12] I. Cetin, M. Stephens, O. Camara, and M. A. G. Ballester, "Attri-vae: Attribute-based inter- pretable representations of medical images with variational autoencoders," Computerized Med- ical Imaging and Graphics, vol. 104, p. 102158, 2023.

[13] V. Luis, B. A. D. Marques, H. C. Batagelo, and J. P. Gois, "A review on generative adversarial networks for image generation," Computers and Graphics, vol. 114, pp. 13–25, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009784932300064X

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in neural information processing systems, vol. 33, pp. 6840–6851, 2020.

[15] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," ArXiv, vol. abs/2010.02502, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID: 222140788

[16] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image syn- thesis with latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 10 684–10 695.