AIGC with Diffusion Model

本课程作业关注基于扩散模型的相关问题。扩散模型是一类概率生成模型，它通过注入噪声逐步破坏数据，然后通过学习这个反向过程来生成样本（图 1）。扩散模型的快速发展主要归功于去噪扩散概率模型（DDPMs，denoising diffusion probabilistic models）[1]和基于得分的生成建模（SGMs，score-based generative models）[2]的提出。与经典的基于生成对抗网络（GANs, Generative Adversarial Networks）[3]的方法相比，扩散模型可以对复杂的分布进行更准确地建模[4]，且训练更为稳定，并不会出现模式崩溃（mode collapse）等问题。

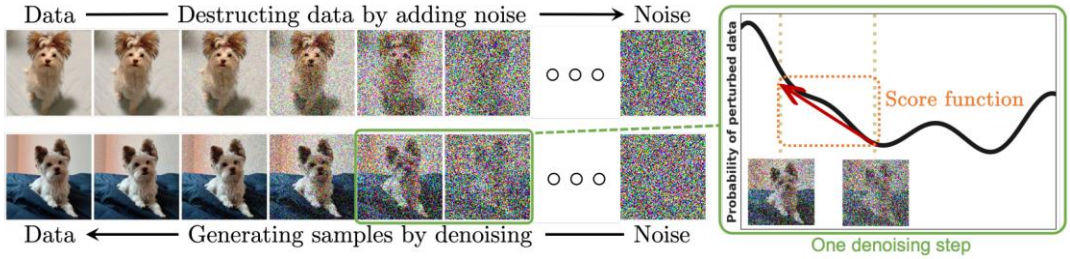

图 1 扩散模型分成加噪和去噪两个过程。每个去噪的步骤需要估计一个如右图所示的得分函数（score function）。图片来自[5]。

基于扩散模型的图像生成方法在计算机领域炙手可热，在多个图像生成问题取得了领先的性能，比如图像合成（image synthesis）[6]，三维模型合成（3D synthesis）[7]，图像超分辨率（image super resolution）[8]，图像补全（inpainting）[9]，图像上色（colorization）[10]，语义分割（semantic segmentation）[11]，视频生成（video generation）[12]等等。

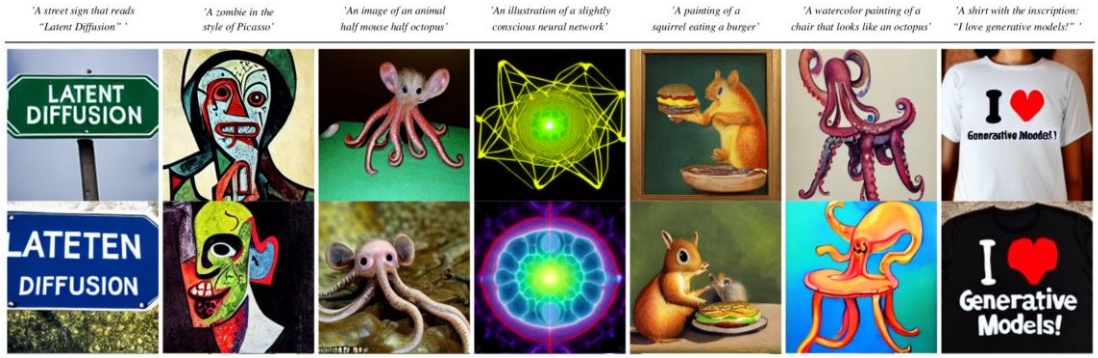

图 2 扩散模型在 Text-to-image synthesis 中的结果。图片来自 https://github.com/CompVis/latent-diffusion。

一个可能的课程作业的内容是去重新实现扩散模型，目前网上有很多公开的实现可以参考，如[13]。另一个方向是去提升计算效率，这是扩散模型的固有缺陷之一，可以参考文章[14]中的做法和分析。

扩散模型的另一个缺陷是无法做到细粒度（fine-grained）的可控输出，因此在已有的问题的基础上，取得更精细的控制，使输出中的不同元素可以定制和组合（composition）也是一个值得去探索的开放问题，最新的一个工作可以参考[15]。

此外，由于效果较好的扩散模型往往需要大量的资源进行训练，利用一个已有的、预训

练的扩散模型去解决一个定制化的问题，也是可以尝试的方向之一。目前有相当一部分工作是基于 Stable Diffusion Model [4]开展的后续工作，包括 fMRI 数据引导的图像生成[16]，文字引导的图像生成（text-to-image synthesis）[17]等。

最后，利用扩散模型去解决除了图像生成以外的问题也是一个有潜力的方向，最近有研究者把扩散模型用在了分子结构生成[18]和抗体设计[19]等问题上。

可以预见，扩散模型的相关研究将会持续引起学术界和工业界的关注。大家可以自由从相关的问题中选取一个自己感兴趣的点作为课程项目作业。更多有关扩散模型的内容可以参考 CVPR2022 的 tutorial[20]和最近的综述文章[5]。

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

[2] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. In Advances in Neural Information Processing Systems, Vol. 33. 12438–12448, 2020.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, Vol. 27. 139–144, 2014.

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

[5] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Yingxia Shao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. "Diffusion models: A comprehensive survey of methods and applications." *arXiv preprint arXiv:2209.00796,* 2022.

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

[7] Ben Poole, Ajay Jain, Jonathan T. Barron, Ben Mildenhall. DreamFusion: Text-to-3D using 2D Diffusion. In International Conference on Learning Representations, 2023.

[8] 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.

[9] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In IEEE Conference on Computer Vision and Pattern Recognition. 11461–11471, 2022.

[10] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. 1–10, 2022.

[11] Emmanuel Asiedu Brempong, Simon Kornblith, Ting Chen, Niki Parmar, Matthias Minderer, and Mohammad Norouzi. Denoising Pretraining for Semantic Segmentation. In IEEE Conference on Computer Vision and Pattern Recognition. 4175–4186, 2022.

[12] William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, and Frank Wood. Flexible Diffusion Modeling of Long Videos. arXiv preprint arXiv:2205.11495, 2022.

[13] https://github.com/lucidrains/denoising-diffusion-pytorch

[14] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. "Tackling the generative learning trilemma with denoising diffusion GANs." arXiv preprint arXiv:2112.07804, 2021.
[15] Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B. Tenenbaum. "Compositional visual generation with composable diffusion models." In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII, pp. 423-439, 2022.
[16] Zijiao Chen, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. "Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding." arXiv preprint arXiv:2211.06956 , 2022.
[17] Weixi Feng, Xuehai He, Tsu-Jui Fu, Varun Jampani, Arjun Reddy Akula, Arjun_Reddy_Akula1, Pradyumna Narayana, Sugato Basu, Xin Eric Wang, and William Yang Wang. Training-Free Structured Diffusion Guidance for Compositional Text-to-Image Synthesis. In International Conference on Learning Representations, 2023
[18] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. In International Conference on Learning Representations, 2021.
[19] Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models. bioRxiv, 2022.
[20] https://cvpr2022-tutorial-diffusion-models.github.io/