

DAFTAR PUSTAKA

- Borji, A. (2023). *Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2*. [https://doi.org/https://doi.org/10.48550/arXiv.2210.00586](https://doi.org/10.48550/arXiv.2210.00586)
- Chefer, H., Alaluf, Y., Vinker, Y., Wolf, L., & Cohen-Or, D. (2023). Attend-And-Excite: Attention-Based Semantic Guidance for Text-To-Image Diffusion Models. *ACM Transactions on Graphics*, 42(4). <https://doi.org/10.1145/3592116>
- Chen, M., Liu, Y., Yi, J., Xu, C., Lai, Q., Wang, H., Ho, T.-Y., & Xu, Q. (2024). *Evaluating Text-to-Image Generative Models: An Empirical Study on Human Image Synthesis*. [https://doi.org/https://doi.org/10.48550/arXiv.2403.05125](https://doi.org/10.48550/arXiv.2403.05125)
- Croitoru, F., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(9), 10850–10869. <https://doi.org/10.1109/TPAMI.2023.3261988>
- Croitoru, F., Hondru, V., Tudor, R., & Shah, M. (2024). Reverse Stable Diffusion: What prompt was used to generate this image? *Computer Vision and Image Understanding*, 1–18. <https://doi.org/10.48550/arXiv.2308.01472>
- Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. *International Conference on Learning Representations 2021*. <https://doi.org/10.48550/arXiv.2105.05233>
- Dodge, J., Gururangan, S., Card, D., Schwartz, R., & Smith, N. A. (2019). Show Your Work: Improved Reporting of Experimental Results. *EMNLP-IJCNLP*

2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, *Proceedings of the Conference*, 2185–2194. <https://doi.org/10.48550>

Feng, Y., Wang, X., Wong, K. K., Wang, S., Lu, Y., Zhu, M., & Aug, A. I. (2024). PromptMagician : Interactive Prompt Engineering for Text-to-Image Creation. *IEEE Transactions on Visualization and Computer Graphics*, 30(1), 295–305. <https://doi.org/10.1109/TVCG.2023.3327168>

Ghosal, S. S., Basu, S., Feizi, S., & Manocha, D. (2024). IntCoOp : Interpretability-Aware Vision-Language Prompt Tuning. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 19584–19601. <https://doi.org/10.18653/v1/2024.emnlp-main.1092>

Goodfellow, I. J., Abadie, J. P., Mirza, M., Xu, B., Farley, D. W., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Third Edition*, 375–400. https://doi.org/10.1007/978-3-031-24628-9_17

He, Y., Robey, A., Murata, N., Jiang, Y., Williams, J. N., Pappas, G. J., Hassani, H., Mitsufuji, Y., Salakhutdinov, R., & Kolter, J. Z. (2025). Automated Black-box Prompt Engineering for Personalized Text-to-Image Generation. In *arXiv preprint arXiv:2403.19103*. <https://arxiv.org/html/2403.19103v4?>

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning technologies. *Proceedings of the AAAI*

- Conference on Artificial Intelligence*, 3207–3214. <https://doi.org/https://doi.org/10.1609/aaai.v32i1.11694>
- Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., & Cohen-or, D. (2022). Prompt-to-Prompt Image Editing with Cross Attention Control. *ICLR 2023*, 1–19. <https://doi.org/10.48550/arXiv.2208.01626>
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6629–6640. <https://doi.org/10.18034/ajase.v8i1.9>
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems, 2020-Decem(NeurIPS 2020)*, 1–12. <https://dl.acm.org/doi/abs/10.5555/3495724.3496298>
- Hu, E., Shen, Y., Wallis, P., Zhu, Z. A., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). L O R A : L O W -R A N K A D A P T A T I O N G U A G E M O D E L S O F L A R G E L A N G U A G E M O D E L S . *ICLR 2022 - 10th International Conference on Learning Representations*, 1–13. <https://www.microsoft.com/en-us/research/publication/lora-low-rank-adaptation-of-large-language-models/>
- Hugging Face. (2023). *Diffusers: State-Of-The-Art Diffusion Models*. Hugging Face. <https://huggingface.co/docs/diffusers/index>
- Jiao, Q., Chen, D., Huang, Y., Lin, X., Shen, Y., & Li, Y. (2025). *DetailMaster: Can Your Text-to-Image Model Handle Long Prompts?* 1–22. <http://arxiv.org/abs/2505.16915>

- Kabra, K., & Balakrishnan, G. (2023). FD: On understanding the role of deep feature spaces on face generation evaluation. *ArXiv Preprint ArXiv:2305.20048*. <https://arxiv.org/abs/2305.20048>
- Kale, A., & Altun, O. (2023). Face Age Synthesis: A Review on Datasets, Methods, and Open Research Area. *Pattern Recognition*, 143. <https://doi.org/https://doi.org/10.1016/j.patcog.2023.109791>
- Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *International Conference on Learning Representations (ICLR)*, 1–9. <https://doi.org/10.48550/arXiv.1312.6114>
- Lee, T., Yasunaga, M., Meng, C., Mai, Y., Park, J. S., Bellagente, M., Kang, M., Park, T., Leskovec, J., Zhu, J.-Y., Fei-Fei, L., Wu, J., Ermon, S., & Liang, P. (2023). Holistic Evaluation of Text-to-Image Models. *Advances in Neural Information Processing Systems 36 (NeurIPS 2023) Datasets and Benchmarks Track*, 36(1). https://proceedings.neurips.cc/paper_files/paper/2023/hash/dd83eada2c3c74db3c7fe1c087513756-Abstract-Datasets_and_Benchmarks.html
- Li, H., Wang, Z., Yue, X., Wang, W., & Tomiyama, H. (2023). An architecture - level analysis on deep learning models for low - impact computations. *Artificial Intelligence Review*, 56(3), 1971–2010. <https://doi.org/10.1007/s10462-022-10221-5>
- Liu, V., & Chilton, L. B. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *Conference on Human Factors in Computing Systems - Proceedings*, 1–23. <https://doi.org/10.1145/3491102.3501825>
- Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep Learning Face Attributes in

- the Wild. *2015 IEEE International Conference on Computer Vision (ICCV)*, 3730–3738. <https://doi.org/10.1109/ICCV.2015.425>
- Mañas, O., Astolfi, P., Hall, M., Ross, C., Urbanek, J., Williams, A., Agrawal, A., Romero-Soriano, A., & Drozdal, M. (2024). Improving Text-to-Image Consistency via Automatic Prompt Optimization. *Transactions on Machine Learning Research*, 2024. <https://doi.org/10.48550/arXiv.2403.17804>
- Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J., & Ermon, S. (2022). SDEDIT: GUIDED IMAGE SYNTHESIS AND EDITING WITH STOCHASTIC DIFFERENTIAL EQUATIONS. *ICLR 2022 - 10th International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.2108.01073>
- Mo, W., Zhang, T., Bai, Y., Su, B., Wen, J.-R., & Yang, Q. (2024). Dynamic Prompt Optimizing for Text-to-Image Generation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2024*, 26627–26636. <https://cvpr.thecvf.com/virtual/2024/poster/29217>
- Narayanan, D., Shoeybi, M., Casper, J., Legresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., & Kashinkunti, P. (2021). Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM. *Microsoft Research*. <https://doi.org/10.1145/3458817.3476209>
- Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2024). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In B. Kim and Y. Yue and S. Chaudhuri and K. Fragkiadaki and M. Khan and Y. Sun (Ed.), *International Conference on Representation Learning 2024* (pp. 1862–1874). <https://proceedings.>

iclr.cc/paper_files/paper/2024/hash/081b08068e4733ae3e7ad019fe8d172f-Abstract-Conference.html

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of the 38th International Conference Machine Learning*, 139, 8748–8763. <https://proceedings.mlr.press/v139/radford21a.html>

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/https://doi.org/10.1109/CVPR52688.2022.01042>

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). DreamBooth : Fine Tuning Text-to-Image Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22500–22510. <https://ieeexplore.ieee.org/document/10204880>

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Gontijo-Lopes, R., Salimans, T., Ho, J., Fleet, D. J., & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Advances in Neural Information Processing Systems*, 35. <https://doi.org/https://doi.org/10.48550/arXiv.2205.11487>

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. *NIPS'16: Proceedings of the*

- 30th International Conference on Neural Information Processing Systems, Nips*, 2234–2242. <https://dl.acm.org/doi/10.5555/3157096.3157346>
- Salimans, T., & Ho, J. (2022). Progressive Distillation for Fast Sampling of Diffusion Models. In *ICLR 2022 - 10th International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.2202.00512>
- Saritas, E., & Ekenel, H. K. (2024). Analyzing the Feature Extractor Networks for Face Image Synthesis. *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition, FG 2024*. <https://doi.org/10.1109/FG59268.2024.10581922>
- Schramowski, P., Brack, M., Deiseroth, B., & Kersting, K. (2023). Safe Latent Diffusion : *CVPR (Conference on Computer Vision and Pattern Recognition) 2023*. <https://doi.org/10.48550/arXiv.2211.05105>
- Shoeybi, M., Patwary, M., Puri, R., Legresley, P., Casper, J., & Mar, C. L. (2019). Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism. *SC20 (International Conference for High Performance Computing)*. <https://doi.org/10.48550/arXiv.1909.08053>
- Song, J., Meng, C., & Ermon, S. (2021). Denoising Diffusion Implicit Models. *ICLR 2021*, 1–22. <https://doi.org/10.48550/arXiv.2010.02502>
- Sun, J., Qiyao, D., Qi, L., Muiyi, S., Yunfan, L., & Zhen, S. (2024). AnyFace++: A Unified Framework for Free-style Text-to-Face Synthesis and Manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2023.3345866>
- Sun, Q., Guo, J., & Liu, Y. (2022). Face image synthesis from facial parts. *Eurasip*

- Journal on Image and Video Processing*, 2022(1). <https://doi.org/10.1186/s13640-022-00585-7>
- Vaswani, A. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. <https://doi.org/10.48550/arXiv.1706.03762>
- Wan, Z., Wang, X., Liu, C., Alam, S., Zheng, Y., Liu, J., Qu, Z., Yan, S., Zhu, Y., Zhang, Q., Chowdhury, M., & Zhang, M. (2024). Efficient Large Language Models: A Survey. *Transactions on Machine Learning Research*, 2024, 1–67. <https://doi.org/10.48550/arXiv.2312.03863>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13, 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Deep Learning Approach for Natural Language Processing, Speech, and Computer Vision*, 38–45. <https://doi.org/10.1201/9781003348689-3>
- Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M. H. (2024). Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys*, 56(4). <https://doi.org/10.1145/3626235>