Text-to-Image Synthesis With Generative Models: Methods, Datasets, Performance Metrics,Challenges, and Future Direction

GUIDED BY:

NANDANA S KUMAR

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MOUNT ZION INSTITUTE OF SCIENCE AND TECHNOLOGY,CHENGANNUR

PRESENTED BY:

NANDAKRISHNAN O(MZW21CS013)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MOUNT ZION INSTITUTE OF SCIENCE AND TECHNOLOGY,CHENGANNUR

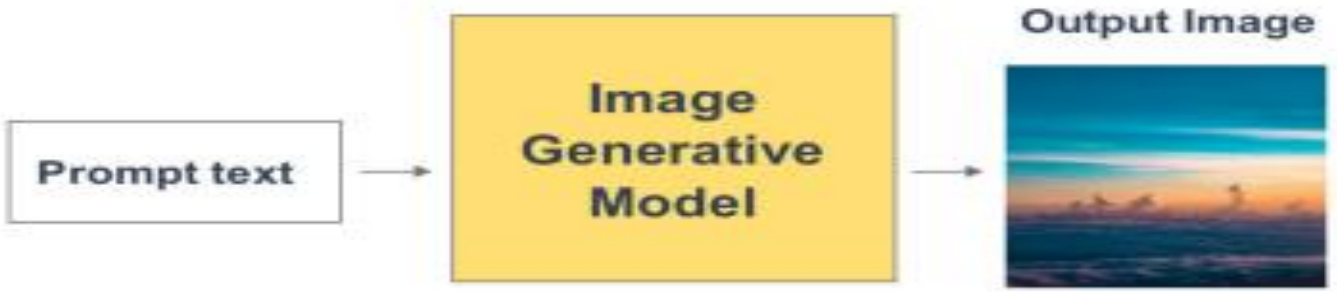
CONTENT

- INTRODUCTIN
- LITERETURE REVIEW
- OBJECTIVES
- CONCLUSION
- REFERENCE

INTRODUCTION

- Text-to-image synthesis is a key area in artificial intelligence and computer vision, aiming to generate realistic images based on textual descriptions.
- The task involves converting human-written text into corresponding visual representations, utilizing advanced computational methods.
- Modern text-to-image synthesis largely relies on generative models like Generative Adversarial Networks (GANs) and diffusion models.
- The field has evolved from early methods using traditional machine learning and supervised approaches to more complex deep learning models.
- Recent advancements in GANs, VAEs, and diffusion models have greatly enhanced the quality and diversity of the generated images.
- These deep learning models are now capable of producing highly realistic and diverse visual outputs from simple textual prompts.

- The research also focuses on exploring the role of datasets in training these models, identifying key challenges in the field.
- The paper highlights the potential future advancements in text-to-image synthesis, particularly in improving model performance and addressing existing limitations.



OBJECTIVES

- Review generative models in text-to-image synthesis.
- Focus on GANs, VAEs, and diffusion models.
- Analyze datasets for training models.
- Study evaluation metrics for model performance.
- Computational complexity.
- Dataset limitations.
- Ethical concerns
- Improve model efficiency
- Address ethical issues.
- Expand model applications across languages and domains

LITRETURE REVIEW

TITLE	AUTHOR	YEAR	JOURNAL/ CONFERENCE	OBJECTIVE
AttnGAN (Attention GAN)	Xu et al	2018	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	AttnGAN used attention mechanisms to focus on specific parts of the text, creating more accurate and detailed images.
MirrorGAN	Qiao et al	2019	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	MirrorGAN used a text-to-image-to-text approach, ensuring images could reconstruct the original text, enhancing text-image consistency.
Diffusion Models in Image Generation	Dhariwal & Nichol	2021	International Conference on Machine Learning (ICML)	Diffusion models surpassed GANs by gradually denoising data, producing higher quality and more complex images.
GLIDE (Guided Language to Image Diffusion)	Nichol et al	2021	OpenAI Research Paper (preprint on arXiv)	GLIDE combined diffusion models with CLIP guidance, improving text-image alignment

CONCLUSION

- Text-to-image synthesis using generative models represents a significant advancement in AI, transforming textual descriptions into coherent images.
- Techniques such as GANs and Transformer-based models have shown remarkable progress in bridging natural language processing and computer vision.
- The ability to generate high-quality, contextually relevant images opens new avenues in fields like art, advertising, and content creation.

- Challenges remain, including image quality, diversity, and ethical concerns surrounding content generation.
- Future research should focus on improving model robustness, exploring multimodal approaches, and addressing ethical implications to enhance the technology's applicability and trustworthiness.

REFERENCES

- [1] S. Frolov, T. Hinz, F. Raue, J. Hees, and A. Dengel, ‘‘Adversarial textto-image synthesis: A review,’’ Neural Netw., vol. 144, pp. 187–209, Dec. 2021.
- [2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative adversarial networks,’’ 2014, arXiv:1406.2661.
- [3] J. Agnese, J. Herrera, H. Tao, and X. Zhu, ‘‘A survey and taxonomy of adversarial neural networks for text-to-image synthesis,’’ WIREs Data Mining Knowl. Discovery, vol. 10, no. 4, Jul. 2020, Art. no. e1345.
- [4] L. Jin, F. Tan, and S. Jiang, ‘‘Generative adversarial network technologies and applications in computer vision,’’ Comput. Intell. Neurosci., vol. 2020, pp. 1–17, Aug. 2020.
- [5] J. Zakraoui, M. Saleh, and J. A. Ja’am, ‘‘Text-to-picture tools, systems, and approaches: A survey,’’ Multimedia Tools Appl., vol. 78, no. 16, pp. 22833–22859, Aug. 2019, doi: 10.1007/s11042-019-7541-4.

- [6] D. Joshi, J. Z. Wang, and J. Li, “The story picturing engine—A system for automatic text illustration,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 2, no. 1, pp. 68–89, Feb. 2006, doi: 10.1145/1126004.1126008.
- [7] X. Zhu, A. Goldberg, M. Eldawy, C. Dyer, and B. Strock, “A text-topicture synthesis system for augmenting communication,” in Proc. 22nd AAAI Conf. Artif. Intell., 2007, p. 1590.
- [8] H. Li, J. Tang, G. Li, and T.-S. Chua, “Word2Image: Towards visual interpreting of words,” in Proc. 16th ACM Int. Conf. Multimedia, 2008, pp. 813–816.
- [9] B. Coyne and R. Sproat, “WordsEye: An automatic text-to-scene conversion system,” in Proc. 28th Annu. Conf. Comput. Graph. Interact. Techn., Aug. 2001, pp. 487–496.
- [10] M. E. Ma, “Confucius: An intelligent multimedia storytelling interpretation and presentation system,” School Comput. Intell. Syst., Univ. Ulster, Coleraine, U.K., Tech. Rep., 2002. [11] Y. Jiang, J. Liu, and H. Lu, “Chat with

THANK YOU