**Table of Contents**

[**1.** **Introduction** 2](#_Toc129225958)

[**2.** **Literature review** 3](#_Toc129225959)

[***2.1.*** ***GANs*** 3](#_Toc129225960)

[***2.2.*** ***Diffusion models*** 10](#_Toc129225961)

[***2.3.*** ***GANs vs Diffusion models*** 10](#_Toc129225962)

[**2.4.** **Overall Comparison** 11](#_Toc129225963)

[**3.** **Results Analysis** 15](#_Toc129225964)

[4. Conclusion 17](#_Toc129225965)

[5. References 18](#_Toc129225966)

# **Introduction**

Generative models have enabled many difficulties to ease up. For example, with generative models they can generate more data that can be used to provide datasets to train other models or networks. This can be beneficial to aid small datasets and the gathering of more data can be difficult. Generative models have the capability to generate texts, images, and audios. They can also help in denoising images, upscaling its resolution, style transferring, etc.

# **Literature review**

In this section we will compare between various generative models based on multiple metrics, it samples a set of models of each kind of generative models in the recent years and shows how well each model is doing regarding Fréchet Inception Distance (FID) and negative log-likelihood (NLL) in bits-per-dimension (BPD). It also rates the training and testing speed and the number of parameters for each model. Also, many datasets were tested for diversity and generalization purposes.

## ***GANs***

Exploring a survey on Generative Adversarial Networks [1], it explores many GANs developed along the years. We will be exploring the ones developed in the past two years along with their results.

Table shows summary of models used for image generation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Generator | Discriminator | Objective function | Datasets |
| GANsformer [54] | Bipartite Transformer having simplex and duplex attention | Attention CNN-based discriminator | Loss functions of StyleGAN | CLEVR |
| GANformer2 [55] | Generator works in two stages: layout generation and layout to scene translation | One CNN for real vs. fake and one U-Net for semantic matching | Adversarial loss, Semantic-matching loss and Segment-fidelity loss | CLEVR, Bedrooms, CelebA, Cityscapes and COCO |
| TransGAN [49] | A Transformer-based generator that progressively increases feature resolution | Transformer-based discriminator that takes input at multiple scales | WGAN-GP loss | CIFAR-10, STL-10, CelebA, CelebA-HQ and LSUN Church |
| TokenGAN [57] | Visual Transformer with content and style tokens | Discriminator of StyleGAN2 | Non-saturating logistic adversarial loss, R1 regularization to only discriminator | FFHQ and LSUN Church |
| VQGAN [58] | CNN-based image constituents vocabulary and Transformerbased modeling of vocabulary composition within highresolution image | CNN-based discriminator | Adversarial loss, reconstruction loss, commitment loss and perceptual reconstruction loss | ImageNet, ImageNetAnimal, LSUN Churches & Towers, COCO-Stuff, ADE20K, CelebA-HQ and FFHQ |
| Styleformer [59] | Transformer with Styleformer Encoders having Increased Multi-Head Self-Attention | Discriminator of StyleGAN2-ADA | Discriminator of StyleGAN2-ADA | CIFAR-10, STL-10, CelebA, LSUN-Church, CLEVR and Cityscapes |
| StyleSwin[60] | Style-based GAN with Transformer having double attention modules | Wavelet-based discriminator | Non-saturating GAN loss with R1 gradient penalty and spectral normalization on the discriminator | FFHQ, CelebA-HQ and LSUN Church |
| ViTGAN [52] | ViT-based ordered patch generator | ViT-based discriminator | Non-saturating logistic adversarial loss | CIFAR-10, CelebA and LSUN Bedroom |
| Unleashing Transformer [61] | Trained Transformer using Masked Vector-Quantized tokens prediction | Traditional discriminator | Vector-Quantized loss, generator loss and reconstruction loss | FFHQ, LSUN Bedroom and LSUN Churches |
| Swin-GAN [62] | Swin Transformer-based generator | Swin Transformer-based multi-scale discriminator | WGAN-GP loss | CIFAR-10 and Anime images |
| PGTCEGAN [63] | Capsule Embedding based Progressive Growing Transformer | CNN with multi-scale input in different layer | WGAN-GP loss | CIFAR-10, CelebA and LSUN-Church |
| MedViTGAN [64] | ViT Encoder-based generator | ViT Encoder-based discriminator in conditional GAN setup | WGAN-GP loss with adaptive hybrid loss weighting mechanism | Histopathology image dataset: PatchCamelyon (PCam) and BreakHis |
| PTNet3D [65] | U-shape generator with performer encoder, transformer bottleneck and performer decoder | 3D ResNet-18 model pretrained on Kinetics400 dataset | Adversarial loss, Perceptual loss and Mean square error | MRI datasets: Developing Human Connectome Project (dHCP) and longitudinal Baby Connectome Project (BCP) |
| SLATER [66] | Generator uses cross-attention transformers with input from a mapper | CNN-based discriminator | Non-saturating logistic adversarial loss, gradient penalty for discriminator | MRI synthesis: brain MRI data from fastMRI |
| SwinGAN [51] | Swin Transformer U-Net-based frequency-domain and imagedomain generators | CNN-based discriminator | Adversarial loss, k-space loss and image domain loss | MRI reconstruction: IXI brain dataset |
| 3D Face Transformer [67] | Residual blocks followed by a multi-layer transformer encoder-based generator | Traditional discriminator | Adversarial loss, L1 loss, Edge loss, L1 loss on the transformer outputs and Self-supervised reprojection consistency loss | 3D Face reconstruction: 300W-LP, AFLW, AFLW2000-3D, NoW, In-the-wild images |

Table shows summary of GANs used for image-to-image translation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Generator | Discriminator | Objective Function | Application | Datasets |
| InstaFormer [81] | Generator with ViT encoder blocks consisting of adaptive instance normalization | Traditional discriminator | Adversarial loss, Global and Instance-level content loss, Image and Style reconstruction loss | Image translation | INIT, Domain adaptation: KITTI to Cityscape |
| UVCGAN [53] | UNet-ViT Generator in CycleGAN framework (pre-trained on the image inpainting task) | CycleGAN discriminator with gradient penalty | GAN loss, cycleconsistency loss, and identity-consistency loss | Unpaired image translation | Selfie to Anime, Anime to Selfie, Male to Female, Female to Male, Remove Glasses, and Add Glasses |
| SwinIR [82] | Generator with residual Swin Transformer blocks | Traditional discriminator for super-resolution | Super-resolution: Pixel loss, GAN loss and perceptual loss | Super-resolution | Set5, Set14, BSD100, Urban100, Manga109, RealSRSet |
| RFormer [83] | Transformer-based U-shaped generator with window-based self-attention block | Transformer-based discriminator with window-based self-attention block | Adversarial loss, Edge loss, Charbonnier loss and Fundus quality perception loss | Fundus image restoration | Real Fundus dataset |
| 3D Transformer GAN [84] | Generator consisting of Encoder CNN, Transformer and Decoder CNN | CNN with four convolution blocks | Adversarial loss and L1 loss | PET reconstruction | PET data |
| 3D CVT-GAN [85] | 3D convolutional vision transformer (CVT) based encoder and 3D transposed CVT based decoder | Patch-based discriminator embedded with 3D CVT blocks | Adversarial loss and L1 loss | PET reconstruction | PET data |
| Low-Light TransformerGAN [86] | Transformer using multi-head multi-covariance self-attention and Light feature-forward module structures | Convolutional discriminator | Adversarial loss, smooth L1 loss, perceptual loss, and multi-scale SSIM loss | Low-light enhancement | LOL and SICE |
| LightingNet [87] | Fusion of CNN-based encoderdecoder and ViT-based encoder-decoder | CNN-based discriminator | Adversarial loss, smooth L1 loss, perceptual loss, and multi-scale SSIM loss | Low-light enhancement | LOL, SICE, ExDARK, DICM, LIME, MEF, and NPE |
| ICT [88] | Bi-directional transformer guided CNN | CNN-based discriminator | dversarial loss and L1 loss | Image completion | FFHQ and Places2 |
| BAT-Fill [89] | Bidirectional and autoregressive transformer + CNN-based texture generation | CNN-based discriminator | Adversarial loss, perceptual loss and reconstruction loss | Image inpainting | CelebA-HQ, Places2 and Paris StreetView |
| T-former [90] | U-shaped generator with transformer blocks | Patch GAN discriminator | Adversarial loss, style loss, reconstruction loss and perceptual loss | Image inpainting | CelebA-HQ, Places2 and Paris StreetView |
| APT [91] | Atrous pyramid transformer and dual spectral transform convolution | CNN-based discriminator | Adversarial loss, perceptual loss, style loss and L1 loss for masked and preserved regions | Image inpainting | CelebA-HQ, Places2 and Paris StreetView |
| MAT [92] | A convolutional head, a maskaware transformer body and a convolutional tail | Traditional discriminator | Adversarial loss, perceptual loss and R1 regularization | Large hole image inpainting | CelebA-HQ and Places365- Standard |
| ZITS [93] | Transformer-based structure restorer + CNN-based structure feature encoding and texture restoration | PatchGAN discriminator | Adversarial loss, L1 loss over unmasked region, feature match loss and high receptive field perceptual loss | Image inpainting | Places2, ShanghaiTech, NYUDepthV2 and MatterPort3D |
| HAN [94] | Generator with CNN encoder, hourglass attention structure blocks and CNN decoder | PatchGAN discriminator with spectral norm | Adversarial loss, style loss, reconstruction loss and perceptual loss | Image inpainting | SUN and Beach |
| SRInpainter [95] | Resolution progressive CNN encoder, hierarchical transformer and CNN decoder | SNPatchGAN discriminator | Adversarial loss and superresolved L1 loss | Image inpainting | CelebA-HQ, Places2 and Paris StreetView |
| NDMAL [96] | Nested deformable attention layer mixed with convolution and de-convolution layers | PatchGAN discriminator | Adversarial loss, perceptual loss, edge loss and L1 loss | Image inpainting | CelebA-HQ and Places2 |
| Hint [97] | ViT-based generated hint converts outpaining to inpainting | Traditional discriminator | Adversarial loss, style loss, reconstruction loss and perceptual loss | Image outpainting | SUN and Beach |
| ColorFormer [98] | Generator using transformerbased encoder and a color memory decoder | PatchGAN discriminato | Adversarial loss, perceptual loss and content loss | Image colorization | ImageNet, COCO-Stuff and CelebA-HQ |
| SGA [99] | Generator with stop-gradient attention module between encoder and decoder | Conditional CNN discriminator | Adversarial loss, perceptual loss, reconstruction loss and style loss | Image colorization | Anime portraits and Animal FacesHQ |
| VTGAN [100] | CNN-based generator at different resolution | ViT for discriminator and classification at different resolution | Adversarial loss, mean square error, perceptual loss, embedding feature loss and cross-entropy loss | Retinal image synthesis and disease prediction using fundus and fluorescein angiogram images |  |
| ResViT [101] | Transformer-based generator using aggregated residual transformer blocks | Conditional PatchGAN discriminator | Adversarial loss, reconstruction loss and pixel loss | Multimodal medical image synthesis | IXI brain MRI, BRATS and MRI-CT |

## ***Diffusion models***

In recent research, [200] proposes an extension to the denoising diffusion probabilistic model (DDPM) for image generation, which involves adding a learned diffusion process to the model. The authors find that their method produces images that are comparable in quality to those produced by other state-of-the-art GAN-based methods, while also being more computationally efficient. In a recent paper discussing image synthesis using a latent diffusion model, it showed that it has reached quite the lowest FID regarding other models including ADM that was mentioned earlier.

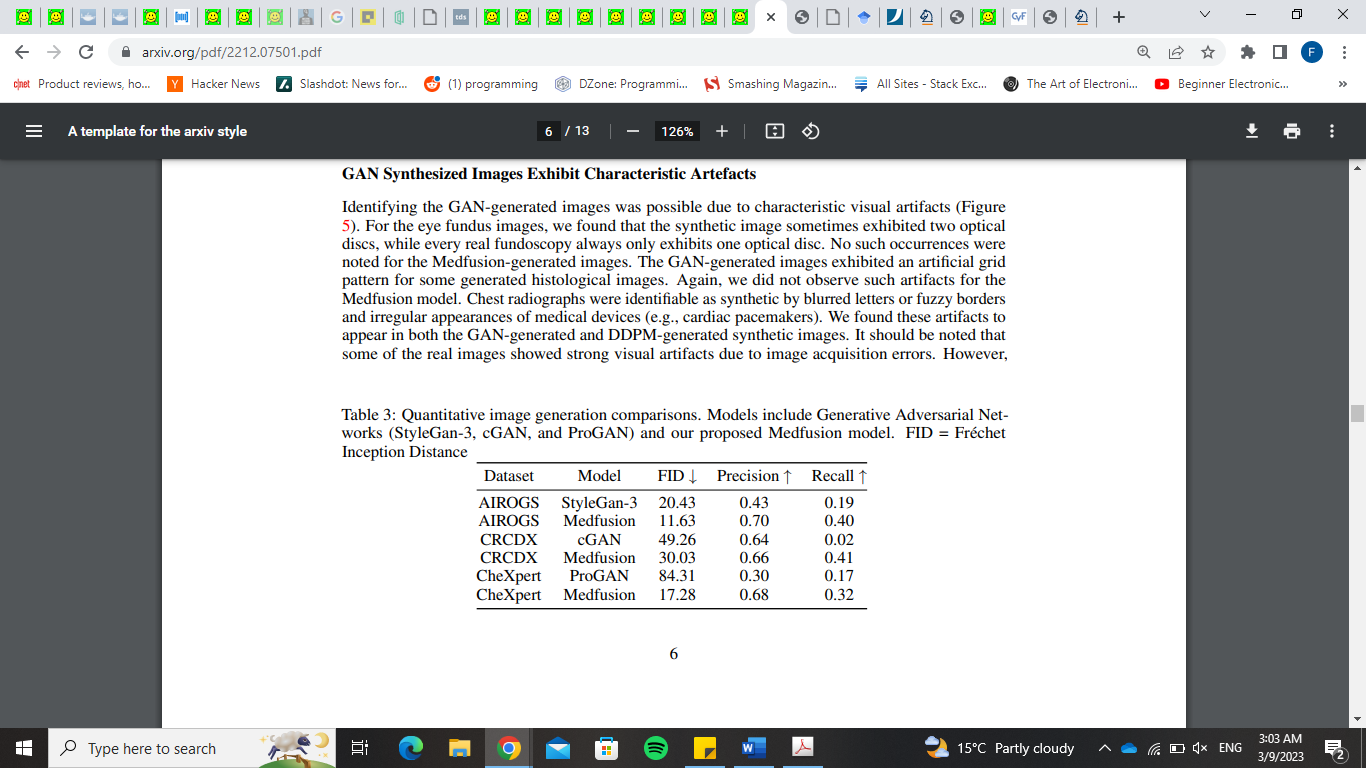
A picture containing text

Description automatically generated

## ***GANs vs Diffusion models***

In several papers, they stated that Diffusion models has defeated GANs in multiple applications.

First, using 2D medical images, the following results show that diffusion models surpassed GANs models.



Additionally, a paper was comparing between various generative models and their new diffusion model ADM and ADM-G on variations of datasets concluding that the models produced had better results.

Graphical user interface, text

Description automatically generated

Table 5: Sample quality comparison with state-of-the-art generative models for each task. ADM refers to our ablated diffusion model, and ADM-G additionally uses classifier guidance.

## **Overall Comparison**

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

Graphical user interface, text

Description automatically generated

Summary for density modeling tasks, in terms of bits per dimension (BPD) on the test set. Model types are autoregressive (AR), normalizing flows (Flow), variational autoencoders (VAE), or diffusion models (Diff).

| Rank | Model | **FID** | Paper | Year | Tags |
| --- | --- | --- | --- | --- | --- |
| 1 | [**EDM-G++** (unconditional)](https://paperswithcode.com/paper/refining-generative-process-with) | 1.77 | [Refining Generative Process with Discriminator Guidance in Score-based Diffusion Models](https://paperswithcode.com/paper/refining-generative-process-with) | 2022 | **Diffusion** |
| 2 | [**StyleGAN-XL**](https://paperswithcode.com/paper/stylegan-xl-scaling-stylegan-to-large-diverse) | 1.85 | [StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets](https://paperswithcode.com/paper/stylegan-xl-scaling-stylegan-to-large-diverse) | 2022 | **GAN** |
| 3 | [**LSGM** (FID)](https://paperswithcode.com/paper/score-based-generative-modeling-in-latent) | 2.10 | [Score-based Generative Modeling in Latent Space](https://paperswithcode.com/paper/score-based-generative-modeling-in-latent) | 2021 | **VAE** |
| 4 | [**LSGM** (balanced)](https://paperswithcode.com/paper/score-based-generative-modeling-in-latent) | 2.17 | [Score-based Generative Modeling in Latent Space](https://paperswithcode.com/paper/score-based-generative-modeling-in-latent) | 2021 | **VAE** |
| 5 | [**StyleGAN2 + DiffAugment + D2D-CE**](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2.236 | [Rebooting ACGAN: Auxiliary Classifier GANs with Stable Training](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2021 | **GAN** |
| 6 | [**INDM** (FID)](https://paperswithcode.com/paper/maximum-likelihood-training-of-implicit) | 2.28 | [Maximum Likelihood Training of Implicit Nonlinear Diffusion Models](https://paperswithcode.com/paper/maximum-likelihood-training-of-implicit) | 2021 | **Diffusion** |
| 7 | [**StyleGAN2 + ADA**](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2.316 | [Rebooting ACGAN: Auxiliary Classifier GANs with Stable Training](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2021 | **GAN** |
| 8 | [**StyleGAN2 + ADA+ D2D-CE**](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2.325 | [Rebooting ACGAN: Auxiliary Classifier GANs with Stable Training](https://paperswithcode.com/paper/rebooting-acgan-auxiliary-classifier-gans) | 2021 | **GAN** |
| 9 | [**BDDM**](https://paperswithcode.com/paper/bddm-bilateral-denoising-diffusion-models-for-1) | 2.38 | [BDDM: Bilateral Denoising Diffusion Models for Fast and High-Quality Speech Synthesis](https://paperswithcode.com/paper/bddm-bilateral-denoising-diffusion-models-for-1) | 2022 | **Diffusion** |
| 10 | [**LeCAM** (StyleGAN2 + ADA)](https://paperswithcode.com/paper/regularizing-generative-adversarial-networks) | 2.47 | [Regularizing Generative Adversarial Networks under Limited Data](https://paperswithcode.com/paper/regularizing-generative-adversarial-networks) | 2021 | **GAN** |
| 11 | [**Styleformer**](https://paperswithcode.com/paper/styleformer-transformer-based-generative) | 2.82 | [Styleformer: Transformer based Generative Adversarial Networks with Style Vector](https://paperswithcode.com/paper/styleformer-transformer-based-generative) | 2021 | **GAN** |
| 12 | [**WaveDiff**](https://paperswithcode.com/paper/wavelet-diffusion-models-are-fast-and) | 4.01 | [Wavelet Diffusion Models are fast and scalable Image Generators](https://paperswithcode.com/paper/wavelet-diffusion-models-are-fast-and) | 2022 | **Diffusion** |

# **Results Analysis**

Table to show results comparison of Transformer-based GANs for image generation in terms of FID score on various datasets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | CIFAR10 | STL10 | CelebA | ImageNet |
| TransGAN [49] | 9.26 | 18.28 | 5.01 | - |
| HiT[56] |  |  |  | 30.83 |
| STrans-G[70] | **2.77** |  | **2.03** | 12.12 |
| Styleformer [59] | 2.82 | 15.17 | 3.92 |  |
| PGTCEGAN [63] | 8.43 |  | 3.59 |  |
| Swin-GAN [62] | 9.23 |  |  |  |
| ViTGAN [52] | 4.92 |  | 3.74 |  |

Table showing results comparison of Transformer-based GANs for high-resolution image generation in terms of FID score on various datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | FFHQ | CelebA-HQ | LSUN | FFHQ | CelebA-HQ |
| VQGAN [58] | 11.40 | 10.70 |  |  |  |
| TransGAN [49] |  | 9.60 | 8.94 |  |  |
| GANsformer [54] | 7.42 |  | 6.51 |  |  |
| GANformer2 [54] | 7.77 |  | 6.05 |  |  |
| TokenGAN [57] | 5.41 |  | 5.56 |  |  |
| STrans-G [70] | 4.84 |  |  |  |  |
| HiT-B [56] | 2.95 | 3.39 |  | 6.37 | 8.83 |
| PGTCEGAN [63] |  |  | 3.92 |  |  |
| StyleSwin [60] | **2.81** | **3.25** | **2.95** | **5.07** | **4.43** |

A picture containing graphical user interface

Description automatically generated

Graphical user interface, text

Description automatically generated

A picture containing graphical user interface

Description automatically generated

# Conclusion

Many GANs models have proved their high accuracy and quality of generated results yet diffusion models proved that they can be slightly better and can get rid of many of the GANs’ disadvantages. GANs are challenging to train due to the constant competition between the generator and discriminator networks, which can lead to unstable and slow training. Furthermore, generating good results with GANs often necessitates a significant amount of training data, which may not be readily available or may be insufficiently large. Finally, mode collapse is a potential issue with GANs, as it can restrict the generator's output to a limited number of options instead of a desired variety.

# 5. References

[54] D. A. Hudson and L. Zitnick, “Generative adversarial transformers,” in International conference on machine learning. PMLR, 2021, pp. 4487–4499.

[55] D. Arad Hudson and L. Zitnick, “Compositional transformers for scene generation,” Advances in Neural Information Processing Systems, vol. 34, pp. 9506–9520, 2021.

[49] Y. Jiang, S. Chang, and Z. Wang, “Transgan: Two pure transformers can make one strong gan, and that can scale up,” Advances in Neural Information Processing Systems, vol. 34, pp. 14 745–14 758, 2021.

[57] Y. Zeng, H. Yang, H. Chao, J. Wang, and J. Fu, “Improving visual quality of image synthesis by a token-based generator with transformers,” Advances in Neural Information Processing Systems, vol. 34, pp. 21 125–21 137, 2021.

[58] P. Esser, R. Rombach, and B. Ommer, “Taming transformers for high-resolution image synthesis,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 12 873–12 883.

[59] J. Park and Y. Kim, “Styleformer: Transformer based generative adversarial networks with style vector,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8983–8992.

[60] B. Zhang, S. Gu, B. Zhang, J. Bao, D. Chen, F. Wen, Y. Wang, and B. Guo, “Styleswin: Transformer-based gan for high-resolution image generation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11 304–11 314.

[52] K. Lee, H. Chang, L. Jiang, H. Zhang, Z. Tu, and C. Liu, “Vitgan: Training gans with vision transformers,” in International Conference on Learning Representations, 2022.

[61] S. Bond-Taylor, P. Hessey, H. Sasaki, T. P. Breckon, and C. G. Willcocks, “Unleashing transformers: parallel token prediction with discrete absorbing diffusion for fast high-resolution image generation from vector-quantized codes,” in Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIII. Springer, 2022, pp. 170–188.

[62] S. Wang, Z. Gao, and D. Liu, “Swin-gan: generative adversarial network based on shifted windows transformer architecture for image generation,” The Visual Computer, pp. 1–11, 2022.

[63] X. Jiang, T. Zhang, Y. Xiao, T. Qiu, and Q. Yang, “A novel gan based on progressive growing transformer with capsule embedding,” in 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2022, pp. 1401–1406.

[64] M. Li, C. Li, P. Hobson, T. Jennings, and B. C. Lovell, “Medvitgan: End-to-end conditional gan for histopathology image augmentation with vision transformers,” in 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022, pp. 4406–4413.

65] X. Zhang, X. He, J. Guo, N. Ettehadi, N. Aw, D. Semanek, J. Posner, A. Laine, and Y. Wang, “Ptnet3d: A 3d high-resolution longitudinal infant brain mri synthesizer based on transformers,” IEEE transactions on medical imaging, vol. 41, no. 10, pp. 2925–2940, 2022.

[66] Y. Korkmaz, S. U. Dar, M. Yurt, M. Ozbey, and T. Cukur, “Unsuper- ¨ vised mri reconstruction via zero-shot learned adversarial transformers,” IEEE Transactions on Medical Imaging, 2022

[51] X. Zhao, T. Yang, B. Li, and X. Zhang, “Swingan: A dual-domain swin transformer-based generative adversarial network for mri reconstruction,” Computers in Biology and Medicine, p. 106513, 2022.

[81] S. Kim, J. Baek, J. Park, G. Kim, and S. Kim, “Instaformer: Instanceaware image-to-image translation with transformer,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 18 321–18 331

[53] D. Torbunov, Y. Huang, H. Yu, J. Huang, S. Yoo, M. Lin, B. Viren, and Y. Ren, “Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 702– 712.

[82] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[83] Z. Deng, Y. Cai, L. Chen, Z. Gong, Q. Bao, X. Yao, D. Fang, W. Yang, S. Zhang, and L. Ma, “Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 9, pp. 4645–4655, 2022.

[84] Y. Luo, Y. Wang, C. Zu, B. Zhan, X. Wu, J. Zhou, D. Shen, and L. Zhou, “3d transformer-gan for high-quality pet reconstruction,” in International Conference on Medical Image Computing and ComputerAssisted Intervention. Springer, 2021, pp. 276–285.

[85] P. Zeng, L. Zhou, C. Zu, X. Zeng, Z. Jiao, X. Wu, J. Zhou, D. Shen, and Y. Wang, “3d cvt-gan: A 3d convolutional vision transformer-gan for pet reconstruction,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 516–526.

[86] S. Yang, D. Zhou, J. Cao, and Y. Guo, “Rethinking low-light enhancement via transformer-gan,” IEEE Signal Processing Letters, vol. 29, pp. 1082–1086, 2022.

[87] ——, “Lightingnet: An integrated learning method for low-light image enhancement,” IEEE Transactions on Computational Imaging, 2023.

[88] Z. Wan, J. Zhang, D. Chen, and J. Liao, “High-fidelity pluralistic image completion with transformers,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4692–4701.

[89] Y. Yu, F. Zhan, R. Wu, J. Pan, K. Cui, S. Lu, F. Ma, X. Xie, and C. Miao, “Diverse image inpainting with bidirectional and autoregressive transformers,” in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 69–78.

[90] Y. Deng, S. Hui, S. Zhou, D. Meng, and J. Wang, “T-former: An efficient transformer for image inpainting,” in Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 6559–6568.

[91] M. Huang and L. Zhang, “Atrous pyramid transformer with spectral convolution for image inpainting,” in Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 4674–4683.

[92] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, “Mat: Maskaware transformer for large hole image inpainting,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10 758–10 768.

[93] Q. Dong, C. Cao, and Y. Fu, “Incremental transformer structure enhanced image inpainting with masking positional encoding,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11 358–11 368.

[94] Y. Deng, S. Hui, R. Meng, S. Zhou, and J. Wang, “Hourglass attention network for image inpainting,” in Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII. Springer, 2022, pp. 483–501.

[95] F. Li, A. Li, J. Qin, H. Bai, W. Lin, R. Cong, and Y. Zhao, “Srinpaintor: When super-resolution meets transformer for image inpainting,” IEEE Transactions on Computational Imaging, vol. 8, pp. 743–758, 2022.

[96] S. S. Phutke and S. Murala, “Nested deformable multi-head attention for facial image inpainting,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 6078–6087.

[97] D. Kong, K. Kong, K. Kim, S.-J. Min, and S.-J. Kang, “Image-adaptive hint generation via vision transformer for outpainting,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3572–3581.

[98] X. Ji, B. Jiang, D. Luo, G. Tao, W. Chu, Z. Xie, C. Wang, and Y. Tai, “Colorformer: Image colorization via color memory assisted hybridattention transformer,” in Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI. Springer, 2022, pp. 20–36.

[99] Z. Li, Z. Geng, Z. Kang, W. Chen, and Y. Yang, “Eliminating gradient conflict in reference-based line-art colorization,” in Computer Vision– ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23– 27, 2022, Proceedings, Part XVII. Springer, 2022, pp. 579–596.

[100] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, and S. A. Baker, “Vtgan: Semi-supervised retinal image synthesis and disease prediction using vision transformers,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3235–3245.

[101] O. Dalmaz, M. Yurt, and T. C¸ ukur, “Resvit: Residual vision transformers for multimodal medical image synthesis,” IEEE Transactions on Medical Imaging, vol. 41, no. 10, pp. 2598–2614, 2022.

[200] N. Ho, A. Poklonskaya, J. H. Metzen, and A. L. Maas, "Improved Denoising Diffusion Probabilistic Models for Image Generation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 8381-8390.