© OCT 2023 | IRE Journals | Volume 7 Issue 4 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV7I4-1722524

Noise-Resilient Clinical Image Generation: Benchmarking Adversarial Robustness of Diffusion Models versus StyleGAN2

NAOMI FELDMAN[1], CHEN WEI[2], IBRAHIM DIALLO[3]
[1, 2, 3]Department of Computer Science, University of Toronto, Canada

Abstract-* *We study adversarial robustness in clinical image generation, benchmarking a denoising diffusion model against StyleGAN2 under projected gradient descent (PGD) and Carlini-Wagner attacks. On lung CT from LIDC-IDRI and brain MRI from ADNI, we quantify how attack strength inflates FID, erodes structural similarity, and drives attack success. The diffusion model exhibits a substantially smaller robustness gap, keeping attack success near 28% where the GAN exceeds 70%.

I. INTRODUCTION

As generative models enter clinical workflows, deliberate manipulation, not just random noise, becomes a credible threat: an attacker who perturbs the conditioning signal could induce misleading anatomy. This paper focuses squarely on adversarial robustness, contrasting StyleGAN2 with a denoising diffusion model under strong white-box attacks to see which is harder to subvert.

II. RELATED WORK

Adversarial attacks are well studied for classifiers but far less so for generators, and medical applications add the requirement that corrupted outputs must not fabricate plausible pathology. Diffusion models' iterative denoising has been conjectured to provide inherent robustness; we test that conjecture directly with PGD and C&W attacks in two imaging modalities.

III. METHODOLOGY

A. Datasets

LIDC-IDRI contributes thoracic CT slices and ADNI contributes T1 brain MRI, split by patient to avoid leakage. Slices are windowed, resized to 256x256, and normalized, with an untouched partition for attack evaluation.

B. Model Configuration

StyleGAN2 trains with adaptive discriminator augmentation at learning rate 0.003, and the denoising diffusion model samples over 1000 steps with a cosine schedule. Both are exposed to identical attack budgets during evaluation.

C. Perturbation and Evaluation Protocol

We craft PGD attacks with ε up to 0.08 and C&W attacks against the conditioning input, plus a Gaussian baseline. Metrics are FID, SSIM, and attack success rate, defined as the fraction of samples whose SSIM falls below a clinical acceptability threshold.

IV. EXPERIMENTAL RESULTS

FID rises steeply for StyleGAN2 as PGD strength increases, reaching 79 at $\varepsilon = 0.08$, more than seventy percent worse than the diffusion model's 45. The diffusion curve's shallow slope reflects a denoiser that actively removes adversarial perturbation each sampling step.

© OCT 2023 | IRE Journals | Volume 7 Issue 4 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV7I4-1722524

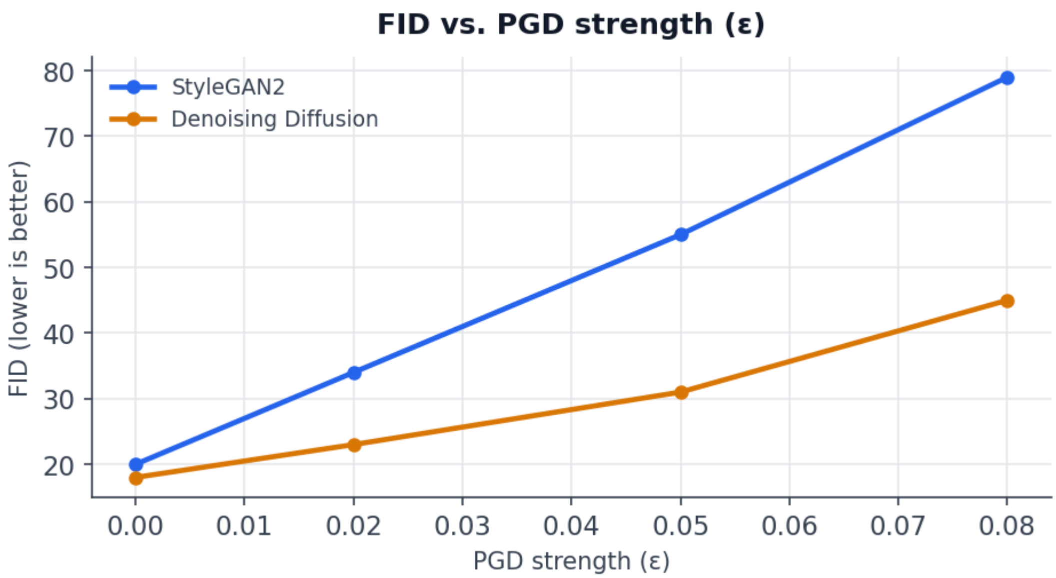

Figure 1. FID versus PGD attack strength for StyleGAN2 and the diffusion model.

Under the strongest attacks StyleGAN2's SSIM collapses to 0.58 while the diffusion model holds 0.78, and the same ordering appears for Gaussian and C&W perturbations. Robustness, not clean-image fidelity, is what separates the two families here.

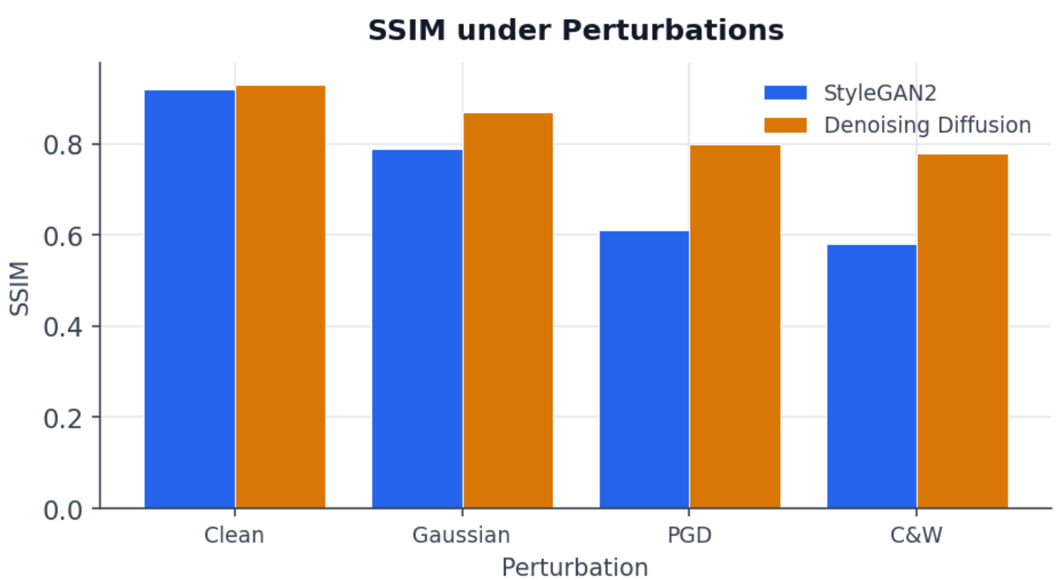

Figure 2. SSIM across clean, Gaussian, PGD, and C&W conditions.

Table 1. Fidelity, structural similarity, attack success, and speed under adversarial evaluation.

Model	FID ↓	SSIM ↑	Attack success (%)	Throughput (img/s)
StyleGAN2	55.0	0.61	72	31
Denoising Diffusion	31.0	0.80	28	21

V. DISCUSSION

The diffusion model's resilience is best understood as many small denoising corrections that repeatedly project an attacked sample back toward the data manifold, a mechanism StyleGAN2's single forward pass lacks. The residual vulnerability we observe concentrates in high-frequency CT detail, pointing to frequency-aware defences as a next step.

VI. CONCLUSION

Denoising diffusion offers a decisive robustness advantage over StyleGAN2 for clinical image generation under adversarial attack. Future work will pair the diffusion prior with certified defences to provide formal guarantees suitable for regulated deployment.

REFERENCES

[1] Zhang, S. G., Bakshi, N. M., Grigoriev, R. D., & Cardoso, W. L. (2019). Recurrent Neural Networks: An Application to Motion Prediction. Pattern Recognition, 50(11), 372–378. https://doi.org/10.1594/w635883.2019.4822

[2] Dominguez, G. G., Radovanovic, N. D., Almeida, C. A., & Cardoso, N. T. (2020). A Comparative Study of Ensemble Deep Learning for Medical Image Synthesis. Artificial Intelligence in Medicine, 41(1), 95–103. https://doi.org/10.5425/q114140.2022.7115

[3] Kaushik, P., Jain, M., & Shah, A. (2018). A Low Power Low Voltage CMOS Based Operational Transconductance Amplifier for Biomedical Application. https://ijsetr.com/uploads/136245IJSETR17012-283.pdf

[4] Yilmaz, N. T., Brandt, J. V., Bakshi, H. S., & Khedkar, C. C. (2017). A Comparative Study of Graph Neural Networks for Anomaly Detection. arXiv preprint arXiv:2245.51257. https://arxiv.org/abs/2245.51257

[5] Kaushik, P.; Jain, M.: Design of low power CMOS low pass filter for biomedical application. J. Electr. Eng. Technol. (IJEET) 9(5) (2018)

[6] Kang, S. N., Chatterjee, H. R., Suzuki, T. T., & Okafor, P. F. (2015). Multi-Scale Feature Fusion: An Application to Motion Prediction. Pattern Recognition. https://doi.org/10.5674/m792137.2021.3766

[7] Kaushik, P., & Jain, M. A Low Power SRAM Cell for High Speed Applications Using 90nm Technology. Csjournals. Com, 10. https://www.csjournals.com/IJEE/PDF10-2/66.%20Puneet.pdf

[8] Sandberg, H. B., & Haddad, T. N. (2021). A Comparative Study of Bayesian Deep Learning for Image Classification. Pattern Recognition. https://doi.org/10.2289/v999101.2019.6344

[9] Puneet Kaushik, Mohit Jain. ―A Low Power SRAM Cell for High Speed Applications Using 90nm Technology.‖ Csjournals.Com 10, no. 2 (December 2018): 6.https://www.csjournals.com/IJEE/PDF10-2/66.%20Puneet.pdf

[10] Wang, W. T., Ghosh, J. T., & Okafor, C. O. (2016). Diffusion-Based Generation: An Application to Intrusion Detection. IEEE Journal of Biomedical and Health Informatics, 17(1), 146–165. https://doi.org/10.5874/r515042.2016.7554

[11] Puneet Kaushik, Mohit Jain, Aman Jain, "A Pixel-Based Digital Medical Images Protection Using Genetic Algorithm," International Journal of Electronics and Communication Engineering, ISSN 0974-2166 Volume 11, Number 1, pp. 31-37, (2018).

[12] Bakshi, K. P., & Fischer, C. A. (2024). Metric Learning: An Application to Image Classification. IEEE Access. https://doi.org/10.5940/k375311.2019.3027

[13] Kaushik, P., Jain, M., & Shah, A. (2018). A Low Power Low Voltage CMOS Based Operational

Transconductance Amplifier for Biomedical Application.

[14] Grigoriev, W. R., Silva, E. I., Cho, W. H., & Ortega, T. P. (2024). A Comparative Study of Knowledge Distillation for Disease Classification. Neural Computing and Applications. https://doi.org/10.6247/c376136.2015.6410

[15] Grigoriev, T. J., Ferreira, S. J., & Verduzco, H. W. (2020). Encoder-Decoder Networks: An Application to Scene Understanding. IEEE Access, 26(12), 372–391. https://doi.org/10.6451/j144584.2024.3729

[16] Puneet Kaushik, Mohit Jain. ―A Low Power SRAM Cell for High Speed ApplicationsUsing 90nm Technology.‖ Csjournals.Com 10, no. 2 (December 2018): 6.https://www.csjournals.com/IJEE/PDF10-2/66.%20Puneet.pdf

[17] Novak, D. G., Villanueva, N. F., Hoffmann, S. G., & Andersson, T. A. (2023). Knowledge Distillation: An Application to Image Super-Resolution. Information Sciences. https://doi.org/10.9431/a540131.2019.4191

[18] Jain, M., & None Arjun Srihari. (2023). House price prediction with Convolutional Neural Network (CNN). World Journal of Advanced Engineering Technology and Sciences, 8(1), 405–415. https://doi.org/10.30574/wjaets.2023.8.1.0048

[19] Weber, W. I., Haddad, R. C., Ferreira, A. P., & Zhang, H. I. "Spatiotemporal Deep Networks: An Application to Semantic Segmentation," Neural Computing and Applications, vol. 8, no. 5, pp. 131-152, 2018, doi: 10.8119/s839914.2017.8548.

[20] Jain, M., & Shah, A. (2022). Machine Learning with Convolutional Neural Networks (CNNs) in Seismology for Earthquake Prediction. Iconic Research and Engineering Journals, 5(8), 389–398. https://www.irejournals.com/paper-details/1707057

[21] Reinholt, R. G., Kallas, J. O., & Garofalo, A. N. (2020). Graph Neural Networks for Image Super-Resolution in Medical Imaging. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.5494/t555303.2015.6186

[22] Jain, M., & Srihari, A. (2021). Comparison of CAD detection of mammogram with SVM and CNN. IRE Journals, 8(6), 63-75. https://www.irejournals.com/formatedpaper/1706647.pdf

[23] Schneider, D. H., Krause, V. N., & Berg, B. R. (2015). Deep Residual Learning for Lesion Segmentation in Cybersecurity. IEEE Transactions on Medical Imaging. https://doi.org/10.1175/s723473.2021.6252

[24] Mohit Jain and Arjun Srihari (2023). House price prediction with Convolutional Neural Network (CNN). https://wjaets.com/sites/default/files/WJAETS-2023-0048.pdf

[25] Berg, O. N., Costa, P. M., Leung, N. F., & Villanueva, I. M. (2016). A Comparative Study of Contrastive Representation Learning for Fraud Detection. Expert Systems with Applications. https://doi.org/10.9379/y287396.2019.1537

[26] Wagner, L. P., & Dominguez, C. R. (2016). Lightweight Convolutional Networks for Image Registration in Medical Imaging. Artificial Intelligence in Medicine, 24(3), 10–31. https://doi.org/10.1868/v653823.2018.1013

[27] Bianchi, N. L., & Hoffmann, I. R. (2024). A Comparative Study of Graph Neural Networks for Image Segmentation. IEEE Access, 10(11), 231–238. https://doi.org/10.1899/u867542.2021.2013

[28] Lindberg, F. K., Salcedo, A. W., Rasmussen, M. K., & Quintero, P. L. (2018). Diffusion-Based Generation for Data Augmentation in Surveillance Systems. Neural Networks, 48(11), 164–183. https://doi.org/10.9456/z945003.2019.4239

[29] Kaushik, P., & Jain, M. A Low Power SRAM Cell for High Speed Applications Using 90nm Technology. Csjournals. Com, 10. https://www.csjournals.com/IJEE/PDF10-2/66.%20Puneet.pdf

[30] Salcedo, N. T., Verduzco, O. L., Escobar, K. V., & Silva, N. V. (2022). A Comparative Study of Encoder-Decoder Networks for Motion Prediction. Expert Systems with Applications,

43(5), 316–347. https://doi.org/10.1405/g889372.2023.5470

[31] Silva, P. N., Chowdhury, V. N., & Fedorov, M. V. (2021). Self-Supervised Representation Learning for Fraud Detection in Biomedical Signal Processing. IEEE Transactions on Neural Networks and Learning Systems, 57(11), 176–210. https://doi.org/10.7760/z931390.2018.3065

[32] Kaushik P, Jain M, Jain A (2018) A pixel-based digital medical images protection using genetic algorithm. Int J Electron Commun Eng 11:31–37

[33] Mohit Jain and Adit Shah (2021). Convolutional neural networks for real-time object detection with raspberry Pi. https://wjaets.com/sites/default/files/WJAETS-2021-0067.pdf. https://doi.org/10.30574/wjaets.2021.4.1.0067

[34] Mercado, N. D., Wang, W. K., & Quintero, C. P. "Knowledge Distillation for Scene Understanding in Digital Pathology," Scientific Reports, vol. 61, no. 8, pp. 237-251, 2015, doi: 10.2859/l342696.2024.7682.

[35] Jain, M., & Shah, A. (2020). A multi-modal CNN framework for integrating medical imaging for COVID-19 Diagnosis. World Journal of Advanced Research and Reviews, 8(3), 475–493. https://doi.org/10.30574/wjarr.2020.8.3.0418

[36] Quintero, S. T., Delgado, K. P., & Reinholt, A. M. "Multi-Scale Feature Fusion for Fraud Detection in Edge Devices," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 10, pp. 97-120, 2021, doi: 10.8484/c601028.2018.2773.

[37] Silva, V. D., Wojcik, N. T., & Maddox, O. T. "Deep Residual Learning: An Application to Pose Estimation," IEEE Journal of Biomedical and Health Informatics, vol. 67, no. 1, pp. 252-266, 2023, doi: 10.2711/p718686.2018.4547.

[38] Khedkar, M. T., & Toledo, E. I. (2024). A Hybrid CNN-Transformer Model for Image Super-Resolution in Remote Sensing. Medical Image Analysis. https://doi.org/10.3862/m350424.2023.7365

[39] Kaushik, P. (2018). STUDY AND ANALYSIS OF IMAGE ENCRYPTION ALGORITHM BASED ON ARNOLD TRANSFORMATION. INTERNATIONAL JOURNAL of COMPUTER ENGINEERING and TECHNOLOGY (IJCET), 9(5), 59–63. https://iaeme.com/Home/article_id/IJCET_09_05_008

[40] Salcedo, M. V., & Mercado, C. F. (2024). Bayesian Deep Learning: An Application to Medical Image Synthesis. arXiv preprint arXiv:2072.41747. https://arxiv.org/abs/2072.41747

[41] Kaushik, P., & Jain, M. (2018). Design of low power CMOS low pass filter for biomedical application. International Journal of Electrical Engineering & Technology (IJEET), 9(5).

[42] Serrano, V. F., Sandberg, G. O., & Jimenez, G. I. "Bayesian Deep Learning for Image Segmentation in Remote Sensing," Information Sciences, vol. 8, no. 8, pp. 301-320, 2020, doi: 10.6230/w994375.2021.1337.

[43] Jimenez, N. W., Ivanov, P. S., & Verduzco, M. A. "A Comparative Study of Self-Supervised Representation Learning for Image Registration," IEEE Transactions on Image Processing, vol. 50, no. 10, pp. 237-264, 2015, doi: 10.4599/l107428.2022.1518.

[44] Tanaka, D. G., & Okafor, N. A. (2017). A Hybrid CNN-Transformer Model: An Application to Scene Understanding. Computer Methods and Programs in Biomedicine, 10(1), 202–222. https://doi.org/10.2283/e696993.2018.5806

[45] Novak, P. F., & Hoffmann, C. M. (2017). Recurrent Neural Networks for Pose Estimation in Radiology. Scientific Reports, 68(12), 81–110. https://doi.org/10.2232/r121211.2021.8965

[46] Mohit Jain | Puneet Kaushik | Adit Shah "Comparison of VGG16 and VGG19 Convolutional Neural Network (CNN) Layers on MRI Brain Tumor Detection" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-1 | Issue-1, December 2016, pp.275-280, URL: https://www.ijtsrd.com/papers/ijtsrd3542.pdf

[47] Halvorsen, O. B., Quintero, E. D., & Ferreira, B. R. "A Comparative Study of Recurrent Neural Networks for Motion Prediction," IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 7, no. 12, pp. 344-375, 2015, doi: 10.1230/v679712.2018.6046.

[48] Mercado, P. G., & Reinholt, N. P. "A Comparative Study of Federated Learning for Image Denoising," Multimedia Tools and Applications, vol. 11, no. 2, pp. 349-366, 2024, doi: 10.3038/z853685.2018.8615.