C-DM: Sinh ảnh X-quang Ngực Tổng hợp sử dụng Mô hình Khuếch tán trên Đám mây ứng dụng trong Chẩn đoán Bệnh

Nguyễn Thành Công, Nguyễn Long Thiên, Phạm Trường Quốc Huy, Phạm Văn Thắng, Trương Tiến Dương

Người hướng dẫn: Nguyễn Duy Tân

Machine Learning & IoT Lab (ML IoT)
Ho Chi Minh City University of Technology (HCMUT)

Bối cảnh

Tập dữ liệu ảnh X-quang ngực quan trọng cho trí tuệ nhân tạo trong y tế:

- Mô hình chẩn đoán bệnh chính xác hơn[1].
- Tập dữ liệu lớn giảm hiện tượng quá khớp[1].
- Phân vùng giữa phổi và tim hiệu quả hơn[2].



(a) X-quang ngực bệnh xẹp phổi



(b) X-quang ngực bệnh xơ hóa

Figure: Ảnh X-quang ngực có bệnh[3]

[1] Farzeen Khalid Khan et al. "Leveraging Large-Scale Public Data for Artificial Intelligence-Driven Chest X-Ray Analysis and Diagnosis". In: *Diagnostics* 16.1 (2026). ISSN: 2075-4418. DOI: 10.3390/diagnostics16010146. URL: https://www.mdpi.com/2075-4418/16/1/146

[2] Adnane Ait Nasser and Moulay A. Akhloufi. "A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography". In: *Diagnostics* 13.1 (2023). ISSN: 2075-4418. DOI: 10.3390/diagnostics13010159. URL: https://www.mdpi.com/2075-4418/13/1/159

[3] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471

Bối cảnh

Hạn chế của tập ảnh X-quang ngực:

- Tập dữ liệu chưa đủ lớn[1].
- Mất cân bằng giữa các bệnh[1].
- Quyền riêng tư của bệnh nhân[2].

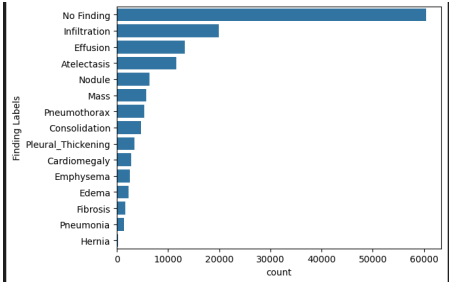


Figure: Biểu đồ số ảnh theo lớp của tập dữ liệu NIH Chest X-ray[3]

[1] Shobhita Sundaram and Neha Hulkund. *GAN-based Data Augmentation for Chest X-ray Classification*. 2021. arXiv: 2107.02970 [eess.IV]. URL: https://arxiv.org/abs/2107.02970

[2] Adnane Ait Nasser and Moulay A. Akhloufi. "A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography". In: *Diagnostics* 13.1 (2023). ISSN: 2075-4418. DOI: 10.3390/diagnostics13010159. URL: https://www.mdpi.com/2075-4418/13/1/159

[3] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471

Lí do chọn đề tài

- Sinh thêm ảnh X-quang ngực để tăng cường dữ liệu (data augmentation).
- Giải quyết vấn đề quyền riêng tư.

Nghiên cứu liên quan

Malik & Humair (2023)

Paper: Evaluating the feasibility of using Generative Models to generate Chest X-Ray Data[a]

Stable Diffusion v2.1

- Text Prompt
- Bounding Box

→ Sinh ảnh theo **bệnh** và **vị trí tổn thương**

$\Rightarrow$

C-DM

Stable Diffusion v1.5 + Fine-tuning

- Label Encoder
- Multi-hot Conditioning
- LoRA

→ Sinh ảnh có điều kiện và **điều khiển mức độ bệnh**

[a] Muhammad Danyal Malik and Danish Humair. "Evaluating the feasibility of using Generative Models to generate Chest X-Ray Data". In: (2023)

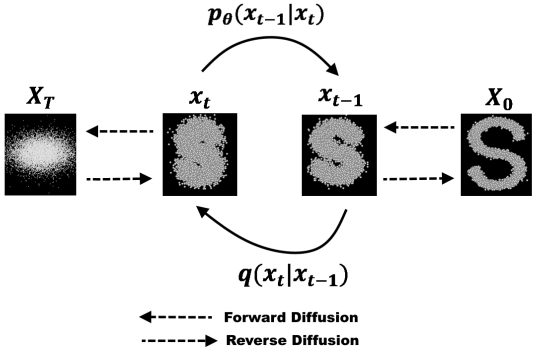


Figure: Quá trình thêm nhiễu và giảm nhiễu của mô hình khuếch tán[1]

[1] PhD J. Rafid Siddiqui. *Diffusion Models Made Easy | Towards Data Science — towardsdatascience.com.* https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da/. 2022

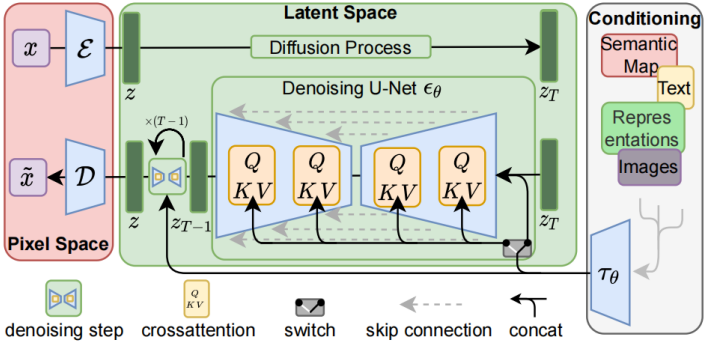


Figure: Kiến trúc của mô hình Stable Diffusion[2]

[2] Robin Rombach et al. "High-Resolution Image Synthesis with Latent Diffusion Models". In: *CoRR* abs/2112.10752 (2021). arXiv: 2112.10752. URL: https://arxiv.org/abs/2112.10752

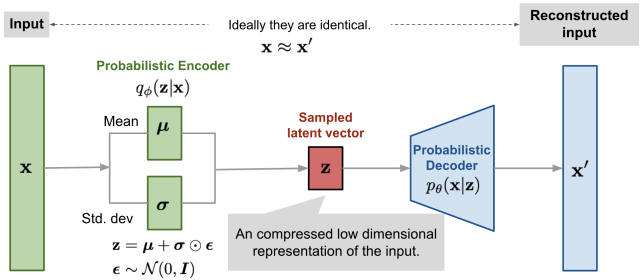


Figure: Kiến trúc của VAE thường[1][2]

[1] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes.* 2022. arXiv: 1312.6114 [stat.ML]. URL: https://arxiv.org/abs/1312.6114

[2] Lilian Weng. *From Autoencoder to Beta-VAE — lilianweng.github.io.* https://lilianweng.github.io/posts/2018-08-12-vae/. 2018

U-Net khử nhiễu

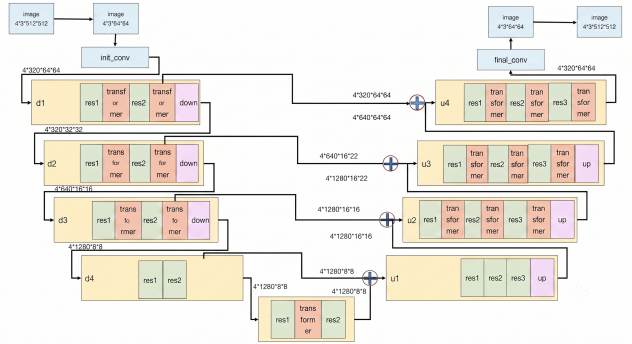


Figure: Kiến trúc của U-Net[3]

[3] Suozhi Wu. *A Detailed Breakdown of the Stable Diffusion Model Architecture*. 2023. URL: https://zhuanlan.zhihu.com/p/638867353

Label Encoder thay cho CLIP

Thiết kế token *present* / *absent*

Mỗi nhãn k được cấp hai bộ token học được. Token của nhãn k:

$$\mathbf{t}_k \;=\; y_k \cdot \mathbf{e}_k^{\text{present}} \;+\; (1 - y_k) \cdot \mathbf{e}_k^{\text{absent}}$$

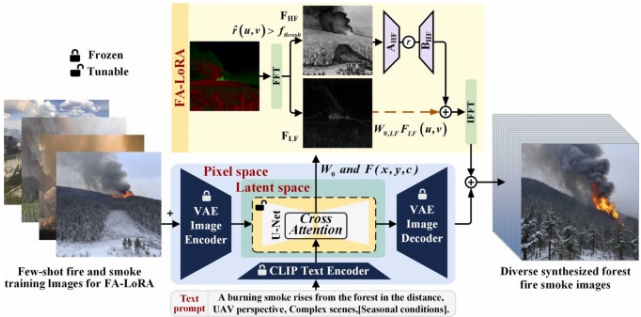


Figure: Minh họa LoRA cho Cross-Attention[4]

[4] Wentao Zhou et al. "Generating diverse data based on FA-LoRA fine-tuning of diffusion models to enhance UAV forest fire detection performance". In: *Applied Soft Computing* 193 (2026), p. 114904. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2026.114904. URL: https://www.sciencedirect.com/science/article/pii/S1568494626003522

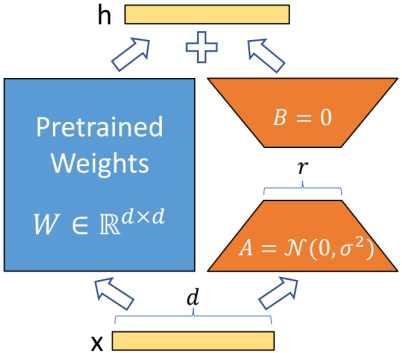


Figure: Minh họa LoRA[5]

[5] Edward J. Hu et al. "LoRA: Low-Rank Adaptation of Large Language Models". In: *CoRR* abs/2106.09685 (2021). arXiv: 2106.09685. URL: https://arxiv.org/abs/2106.09685

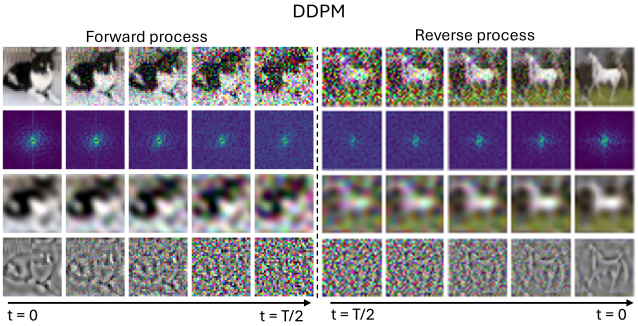
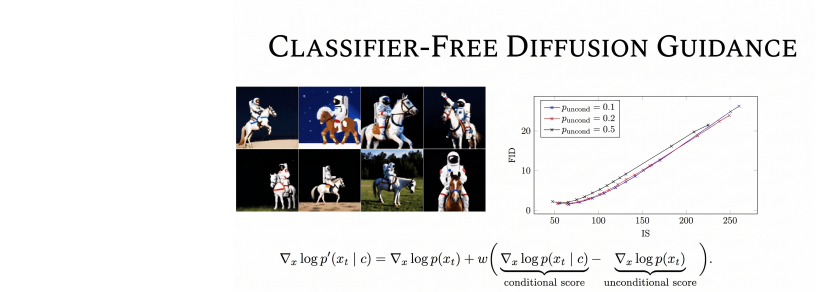


Figure: Minh họa scheduler của mô hình DDPM[6]

[6] Fabian Falck. *Diffusion is not necessarily Spectral Autoregression — fabianfalck.com.* https://www.fabianfalck.com/posts/spectralauto/. 2025



$$\hat{\epsilon}_\theta(x_t|c) = \epsilon_\theta(x_t) + w\,(\epsilon_\theta(x_t|c) - \epsilon_\theta(x_t))$$

Figure: Minh họa Classifier-Free Diffusion Guidance[7]

[7] Nikolas Adaloglou and Tim Kaiser. *An overview of classifier-free guidance for diffusion models | AI Summer — theaisummer.com.* https://theaisummer.com/classifier-free-guidance/. 2024



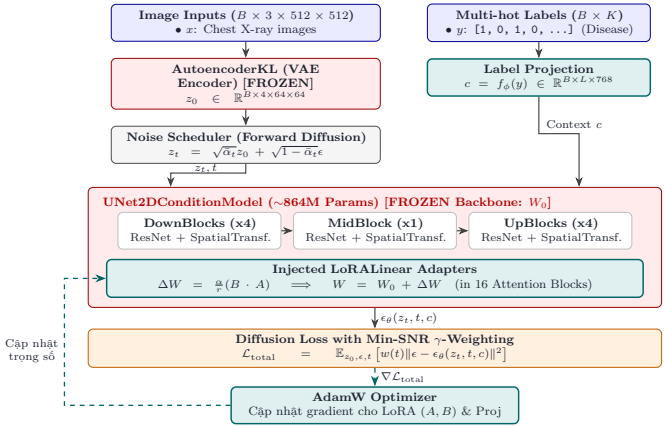
(a) DDIM [19] (b) DPM-Solver (ours)

Figure 1: Samples by DDIM [19] with 10, 15, 20, 100 number of function evaluations (NFE), and DPM-Solver (ours) with only 10 NFE, using the pre-trained DPMs on ImageNet 256×256 with classifier guidance [4].

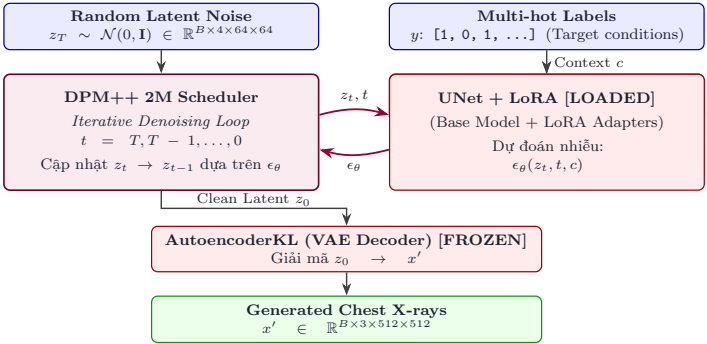
Figure: Minh họa DPM-Solver++ 2M[8]

[8] Cheng Lu et al. *DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps*. 2022. arXiv: 2206.00927 [cs.LG]. URL: https://arxiv.org/abs/2206.00927

Pipeline Quá trình Huấn luyện



Pipeline Quá trình Sinh ảnh (Inference)



Tập dữ liệu: NIH Chest X-ray

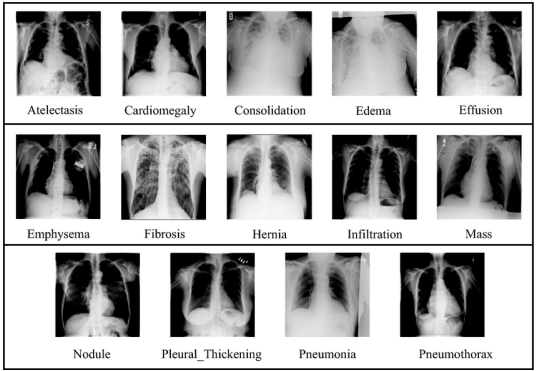
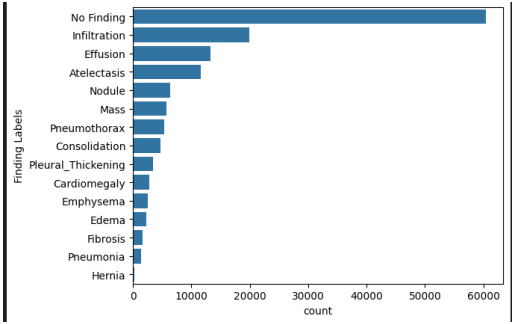


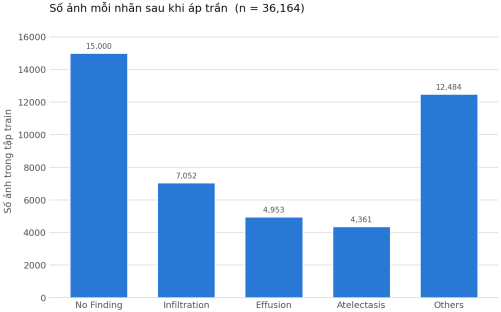
Figure: Ảnh các lớp bệnh của tập dữ liệu NIH Chest X-ray[9]

[9] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471

Tiền xử lý và cân bằng lớp dữ liệu



(a) Trước khi tiền xử lí



(b) Sau khi tiền xử lí

Figure: Biểu đồ số ảnh theo lớp của tập dữ liệu NIH Chest X-ray[10]

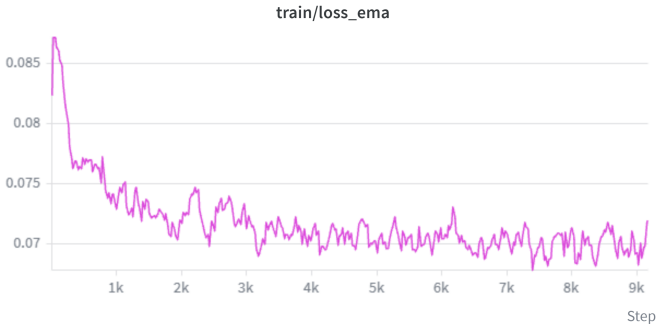
[10] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471

Kết quả tinh chỉnh U-Net



Figure: Đồ thị hàm mất mát của U-Net

Kết quả tinh chỉnh U-Net



$$L_{ema|t} = 0.98 * L_{ema|t-1} + 0.02 * L_t$$

Figure: Đồ thị hàm mất mát EMA của U-Net

Kết quả tinh chỉnh U-Net



(a) Kết quả step 0



(b) Kết quả step cuối

Figure: Kết quả sinh ảnh của U-Net tại các step, với bệnh Infiltration + Effusion + Others

Kết quả tinh chỉnh bộ giải mã VAE



Figure: Đồ thị hàm mất mát MAE

Kết quả tinh chỉnh bộ giải mã VAE

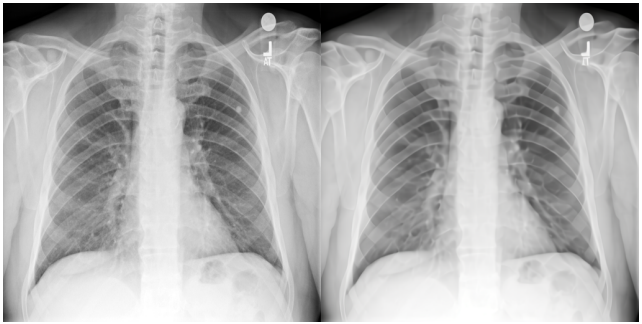


Figure: Kết quả của bộ giải mã VAE step đầu

Kết quả tinh chỉnh bộ giải mã VAE

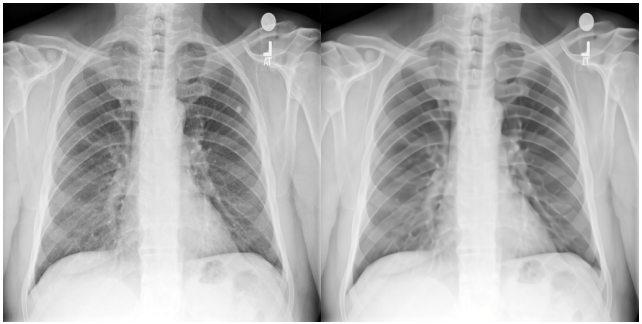


Figure: Kết quả của bộ giải mã VAE step cuối

Ba phương pháp đánh giá:

- **CAS (Classification Accuracy Score)**: Khả năng nhận diện bệnh lý của ảnh[11].
- **TSTR (Train Synthetic Test on Real)**: Huấn luyện mô hình phân loại trên ảnh tổng hợp và kiểm thử trên ảnh thật. So sánh với TRTR (Train on Real, Test on Real)[12].
- **User Study**: Ý kiến chủ quan của con người qua khảo sát.

[11] Augustus Odena, Christopher Olah, and Jonathon Shlens. *Conditional Image Synthesis With Auxiliary Classifier GANs*. 2017. arXiv: 1610.09585 [stat.ML]. URL: https://arxiv.org/abs/1610.09585

[12] Pasquale Zingo and Andrew Novocin. "Introducing the TSTR Metric to Improve Network Traffic GANs". In: Apr. 2021, pp. 643–650. ISBN: 978-3-030-73099-4. DOI: 10.1007/978-3-030-73100-7_46

Classification Accuracy Score (CAS)

Dùng mô hình phân loại để đánh giá chất lượng ảnh được sinh ra.

- **Mô hình:** `densenet121-res224-nih`[13].
- **Lí do chọn:** Có sẵn ngưỡng dương tính cho từng nhãn bệnh (op threshs)[14].

Kết quả

Chỉ số	CAS
macro-AUC	**0.7122**
AUC No Finding	0.7486
AUC Infiltration	0.6528
AUC Effusion	0.7273
AUC Atelectasis	0.7075
AUC Others	0.7246

[13] Joseph Paul Cohen et al. *densenet121-res224-nih*. 2020. URL: https://huggingface.co/torchxrayvision/densenet121-res224-nih

[14] Augustus Odena, Christopher Olah, and Jonathon Shlens. *Conditional Image Synthesis With Auxiliary Classifier GANs*. 2017. arXiv: 1610.09585 [stat.ML]. URL: https://arxiv.org/abs/1610.09585

Train Synthetic Test on Real (TSTR)

Đo giá trị sử dụng thực tế của ảnh sinh ra:

- **Mô hình**: DenseNet-121 tiền huấn luyện trên ImageNet.
- **2 phần**: TRTR và TSTR[15].

Kết quả

Chỉ số	TRTR	TSTR
macro-AUC	**0.7223**	**0.6879**
AUC No Finding	0.7283	0.6901
AUC Infiltration	0.6379	0.6229
AUC Effusion	0.8295	0.7775
AUC Atelectasis	0.7384	0.7001
AUC Others	0.6776	0.6490

[15] Pasquale Zingo and Andrew Novocin. "Introducing the TSTR Metric to Improve Network Traffic GANs". In: Apr. 2021, pp. 643–650. ISBN: 978-3-030-73099-4. DOI: 10.1007/978-3-030-73100-7_46

User Study

- **Phép thử Turing**: Cho người khảo sát phân biệt đâu là ảnh AI / ảnh thật.
- **Rating**: Người tham gia khảo sát đánh giá mức độ giống với ảnh thật của từng ảnh.
- **Khảo sát chuyên gia**: Đánh giá độ chân thực cấu trúc giải phẫu y tế.

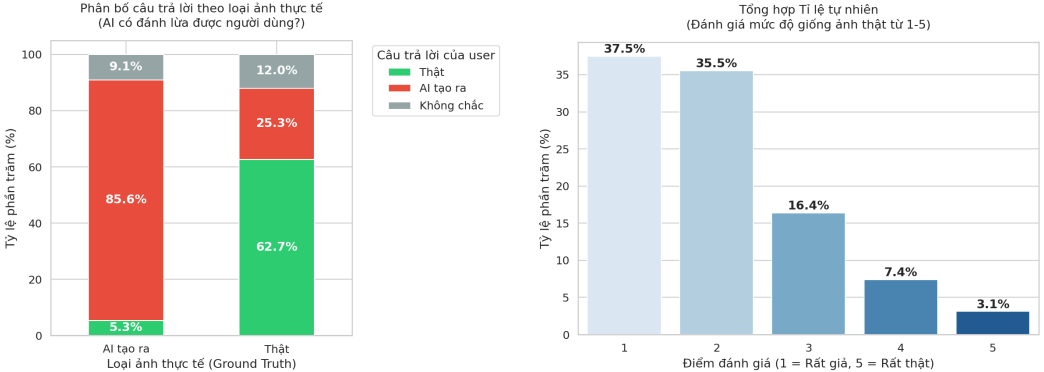


Figure: Kết quả lần 1 với ảnh một lớp

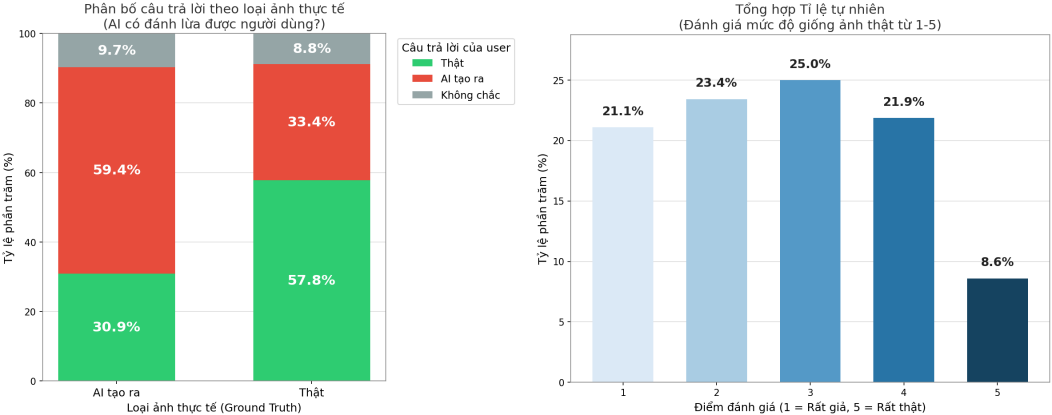


Figure: Kết quả lần 2 với ảnh đa lớp

Demo — Hệ thống Sinh ảnh Trực quan

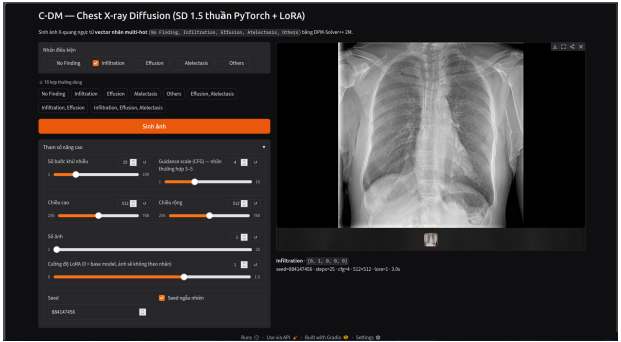


Figure: Giao diện WebUI xây dựng bằng Gradio

Thành phần tương tác

- **Input nhãn:** Checkbox đa nhãn $y \in \{0,1\}^5$.
- **Tham số lấy mẫu:**
  - DPM-Solver++: 25 steps.
  - Guidance Scale (CFG): 4 (dải hợp lí 3-5).
  - Random Seed / Fixed Seed.
- **Tốc độ suy luận:** $\approx 2-3$s / ảnh.

Ảnh hưởng của Guidance Scale (CFG)

CFG (Classifier-Free Guidance) điều khiển mức độ mô hình tuân theo điều kiện đầu vào khi sinh ảnh.



(a) CFG = 1



(b) CFG = 2



(c) CFG = 3



(d) CFG = 4



(e) CFG = 5



(f) CFG = 6



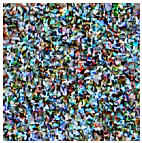
(g) CFG = 7



(h) CFG = 7.5

Ảnh hưởng của số bước DPM-Solver++

DPM-Solver++ sử dụng các bước khử nhiễu để chuyển từ latent nhiễu đến ảnh X-quang cuối cùng.



(a) 1 step



(b) 5 steps



(c) 10 steps



(d) 15 steps



(e) 20 steps



(f) 25 steps

Tổng kết

Ưu điểm:

- Thay đổi từ prompt thành mã hóa nhãn (label encoding), giảm nhiễu điều kiện.
- Tinh chỉnh bằng LoRA giúp giảm tính toán.
- Có phương pháp đánh giá: Các nghiên cứu Weber et al[16] và Malik & Humair[17] chưa có CAS và TSTR.

Nhược điểm:

- Tập dữ liệu nhiều nhiễu: được gán nhãn bằng NLP, không có tiểu sử bệnh của bệnh nhân, một người có thể nhiều bệnh.
- Một số nhãn bệnh còn quá ít ảnh.
- Thời gian huấn luyện và chi phí tính toán còn quá lớn.

[16] Tobias Weber et al. *Cascaded Latent Diffusion Models for High-Resolution Chest X-ray Synthesis*. 2023. arXiv: 2303.11224 [eess.IV]. URL: https://arxiv.org/abs/2303.11224

[17] Muhammad Danyal Malik and Danish Humair. "Evaluating the feasibility of using Generative Models to generate Chest X-Ray Data". In: (2023)

Hướng phát triển

Khắc phục hạn chế

- Tiết kiệm thời gian sinh ảnh bằng Flow matching thay vì DPM-Solver++

Mở rộng tính năng

- Sinh ảnh đa phương thức chẩn đoán (MRI, CT, siêu âm,...).
- Sinh ảnh theo triệu chứng cụ thể (Report-to-Image).

Tài liệu tham khảo I

[1] Nikolas Adaloglou and Tim Kaiser. *An overview of classifier-free guidance for diffusion models | AI Summer — theaisummer.com*. https://theaisummer.com/classifier-free-guidance/. 2024.

[2] Adnane Ait Nasser and Moulay A. Akhloufi. "A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography". In: *Diagnostics* 13.1 (2023). ISSN: 2075-4418. DOI: 10.3390/diagnostics13010159. URL: https://www.mdpi.com/2075-4418/13/1/159.

[3] Joseph Paul Cohen et al. *densenet121-res224-nih*. 2020. URL: https://huggingface.co/torchxrayvision/densenet121-res224-nih.

[4] Fabian Falck. *Diffusion is not necessarily Spectral Autoregression — fabianfalck.com*. https://www.fabianfalck.com/posts/spectralauto/. 2025.

[5] Edward J. Hu et al. "LoRA: Low-Rank Adaptation of Large Language Models". In: *CoRR* abs/2106.09685 (2021). arXiv: 2106.09685. URL: https://arxiv.org/abs/2106.09685.

Tài liệu tham khảo II

[6] PhD J. Rafid Siddiqui. *Diffusion Models Made Easy | Towards Data Science — towardsdatascience.com*. https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da/. 2022.

[7] Farzeen Khalid Khan et al. "Leveraging Large-Scale Public Data for Artificial Intelligence-Driven Chest X-Ray Analysis and Diagnosis". In: *Diagnostics* 16.1 (2026). ISSN: 2075-4418. DOI: 10.3390/diagnostics16010146. URL: https://www.mdpi.com/2075-4418/16/1/146.

[8] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: 1312.6114 [stat.ML]. URL: https://arxiv.org/abs/1312.6114.

[9] Cheng Lu et al. *DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps*. 2022. arXiv: 2206.00927 [cs.LG]. URL: https://arxiv.org/abs/2206.00927.

[10] Muhammad Danyal Malik and Danish Humair. "Evaluating the feasibility of using Generative Models to generate Chest X-Ray Data". In: (2023).

Tài liệu tham khảo III

[11] Augustus Odena, Christopher Olah, and Jonathon Shlens. *Conditional Image Synthesis With Auxiliary Classifier GANs*. 2017. arXiv: 1610.09585 [stat.ML]. URL: https://arxiv.org/abs/1610.09585.

[12] Robin Rombach et al. "High-Resolution Image Synthesis with Latent Diffusion Models". In: *CoRR* abs/2112.10752 (2021). arXiv: 2112.10752. URL: https://arxiv.org/abs/2112.10752.

[13] Shobhita Sundaram and Neha Hulkund. *GAN-based Data Augmentation for Chest X-ray Classification*. 2021. arXiv: 2107.02970 [eess.IV]. URL: https://arxiv.org/abs/2107.02970.

[14] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471.

[15] Tobias Weber et al. *Cascaded Latent Diffusion Models for High-Resolution Chest X-ray Synthesis*. 2023. arXiv: 2303.11224 [eess.IV]. URL: https://arxiv.org/abs/2303.11224.

Tài liệu tham khảo IV

[16] Lilian Weng. *From Autoencoder to Beta-VAE — lilianweng.github.io*. https://lilianweng.github.io/posts/2018-08-12-vae/. 2018.

[17] Suozhi Wu. *A Detailed Breakdown of the Stable Diffusion Model Architecture*. 2023. URL: https://zhuanlan.zhihu.com/p/638867353.

[18] Wentao Zhou et al. "Generating diverse data based on FA-LoRA fine-tuning of diffusion models to enhance UAV forest fire detection performance". In: *Applied Soft Computing* 193 (2026), p. 114904. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2026.114904. URL: https://www.sciencedirect.com/science/article/pii/S1568494626003522.

[19] Pasquale Zingo and Andrew Novocin. "Introducing the TSTR Metric to Improve Network Traffic GANs". In: Apr. 2021, pp. 643–650. ISBN: 978-3-030-73099-4. DOI: 10.1007/978-3-030-73100-7_46.

Thank You

Acknowledgements

mliotlab.github.io | 403.1 H6, BKHCM Campus 2
mlandiotlab@gmail.com | facebook.com/hcmut.ml.iot.lab | youtube.com/@mliotlab

