

Synthetic Data Generation for Small Data Orthopaedic Radiology

Main Point

Address the data scarcity issues faced by computer aided diagnosis tools with synthetic data.

Summary

MS-SSIM and MMD are the most suitable metrics for medical data.
Inpainting diffusion models are better suited for complex medical data generation.
cGANs are better at preserving medical data anonymisation.

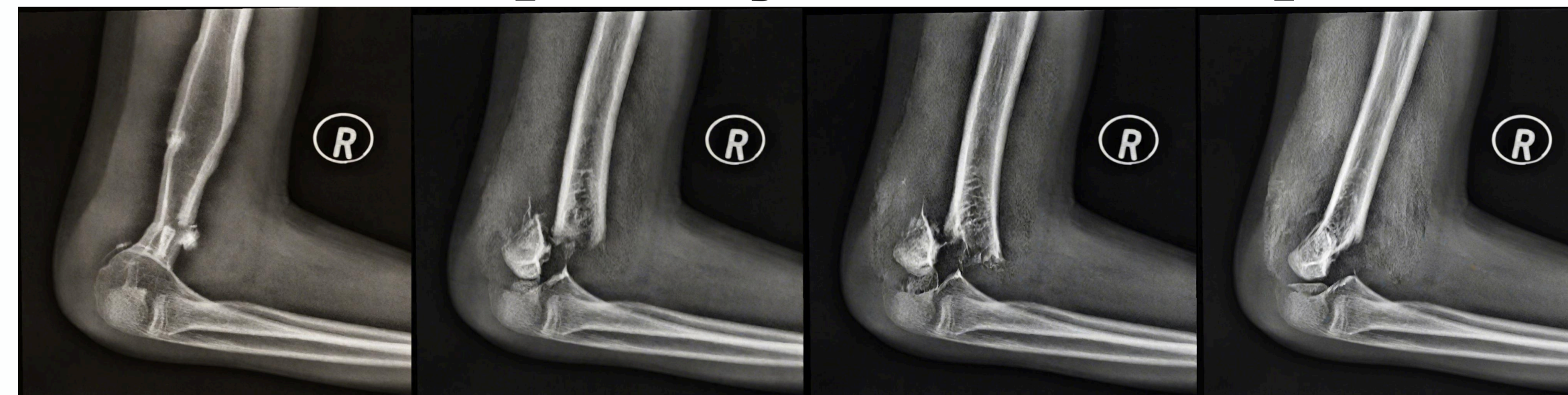
Diffusion Objectives

Compare the Flux.1-dev (FLUX) and Stable Diffusion XL 1.0 (SDXL) latent diffusion models to the GAN models in terms of image generation capabilities by analysing the learning curves for all models using SSIM, MS-SSIM, and MMD.

Diffusion Conclusions

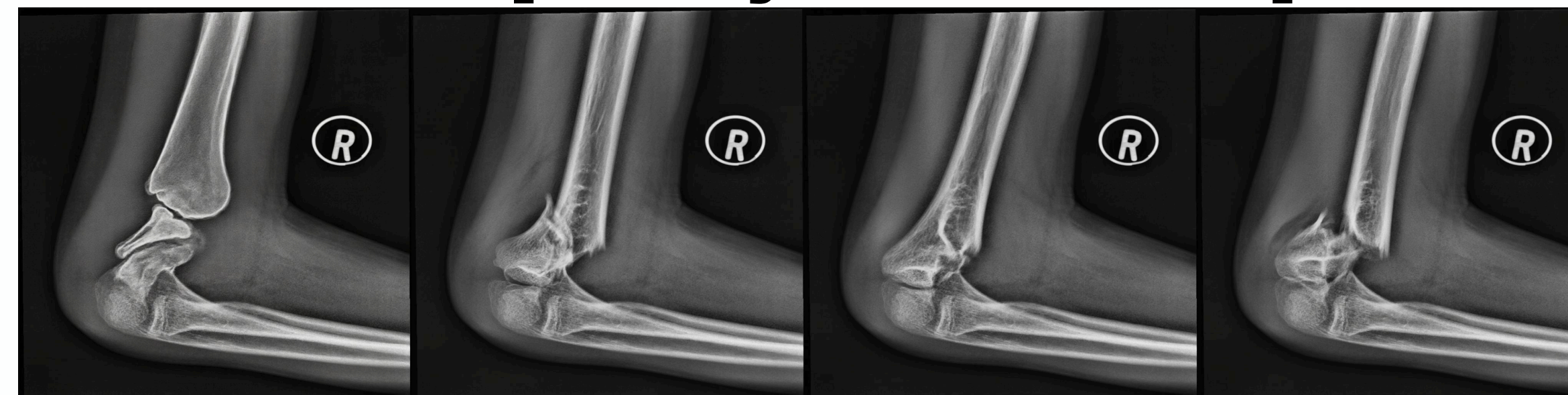
Diffusion models showed better training curves and performed better than the GAN models.
FLUX outperformed SDXL.
Image inpainting allowed the diffusion models to create focused and realistic images.

SDXL: Supracondylar Fracture Sample



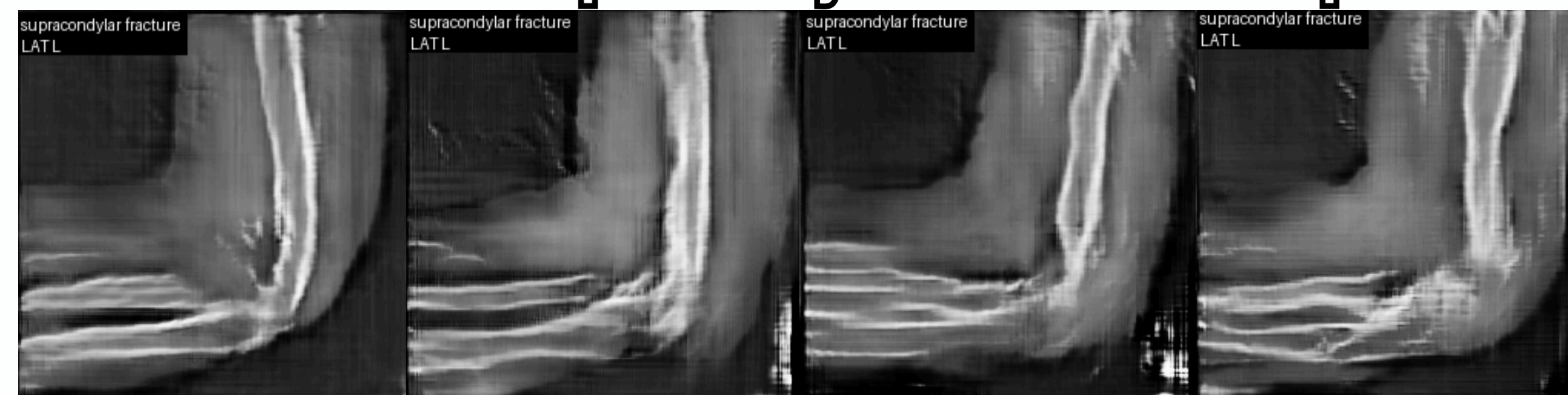
0 Epochs 60 Epochs 120 Epochs 200 Epochs

FLUX: Supracondylar Fracture Sample



0 Epochs 60 Epochs 120 Epochs 200 Epochs

TAC-GAN: Supracondylar Fracture Sample



1K Epochs 2K Epochs 3K Epochs 4K Epochs

AC-GAN: Supracondylar Fracture Sample



1K Epochs 2K Epochs 3K Epochs 4K Epochs

Metrics Objectives

Identify image quality metrics that are reliable at low sample sizes and resilient to common adversarial transformations found in medical data.

Metrics Conclusions

MS-SSIM and MMD were the most robust to transformations.
Only 5 out of 14 pathologies had enough samples for reliable FID and MMD scores!

cGAN Objectives

Compare the Auxiliary Classifier-GAN (AC-GAN) to the Twin AC-GAN (TAC-GAN) in terms of training stability.
Compare the cGANs to SDXL in terms of image diversity.

cGAN Conclusions

Partial TAC-GAN quality improvement over AC-GAN but increased susceptibility to mode collapse.
SDXL tendency towards memorization, cGANs showed greater diversity.