On the Feasibility of Chest X-ray Reconstruction from Foundation Model Vector Embeddings

Frank Li, PhD, Postdoctoral Research Fellow, Radiology, Emory University
Theo Dapamede, MD, PhD; Mohammadreza Chavoshi, MD; Bardia Khosravi, MD, MPH, MHPE; Janice Newsome, MD; Aawez Mansuri, MS; Rohan Satya Isaac, MS; Hari Trivedi, MD; Judy Gichoya, MS, MD, FSIIM

Introduction

Foundation models are large AI systems pre-trained on vast amounts of data that can be efficiently adapted for diverse tasks through zero/few-shot learning fine-tuning. Their advantages include robust transfer learning capabilities, superior generalizability compared to specialized models, and reduced requirements for task-specific data and resources, offering both enhanced capability and cost-effectiveness. However, the potential for foundation models to encode protected health information (PHI) raises privacy concerns. As an initial investigation, we examined the feasibility of reconstructing original chest X-rays (CXRs) from vector embeddings extracted by a CXR-specific fine-tuned foundation model.

Hypothesis

Vector embeddings, being compact yet information-dense representations, should enable the reconstruction of original CXRs with high fidelity.

Methods

Vector embeddings (n=2,486,502) were extracted from a private CXR dataset using RAD-DINO, a foundation model trained exclusively on medical imaging data (Figure 1). A Wasserstein Generative Adversarial Network (WGAN) was developed for image reconstruction using the vector embeddings, where Wasserstein distance was employed for distribution matching, perceptual loss was utilized for semantic feature preservation, and a hybrid L1+L2 loss function was implemented to maintain both structural integrity and fine details. An additional 14,140 images from 2,000 MIMIC CXR patients were randomly selected for external validation.

Results

The reconstruction quality, measured by Frechet Inception Distance (FID), achieved a high distance of 102.42 between the original and reconstructed MIMIC CXRs, implying challenges to reconstruct images to their original form. Nonetheless, visual inspection of reconstructed images (Figure 2) demonstrated preserved major anatomical features and view positions, while exhibiting slightly reduced contrast and enhanced smoothness. While the reconstruction process partially preserved some burned-in markers and annotations, these elements became indistinct and unidentifiable in the reconstructed images.

Conclusion

Our study demonstrates the feasibility of reconstructing images from vector embeddings, even for images unseen by the WGAN but used in RAD-DINO's training, revealing a potential of exposing training data used to train foundation models. We demonstrate that the burned-in text and markers are partially reconstructed but rendered unrecognizable, which is important to maintain fidelity of data anonymization. The ability to view some form of images can aid in model auditing especially when foundation models are trained on multimodal datasets as well as ablation studies.

Figure(s)

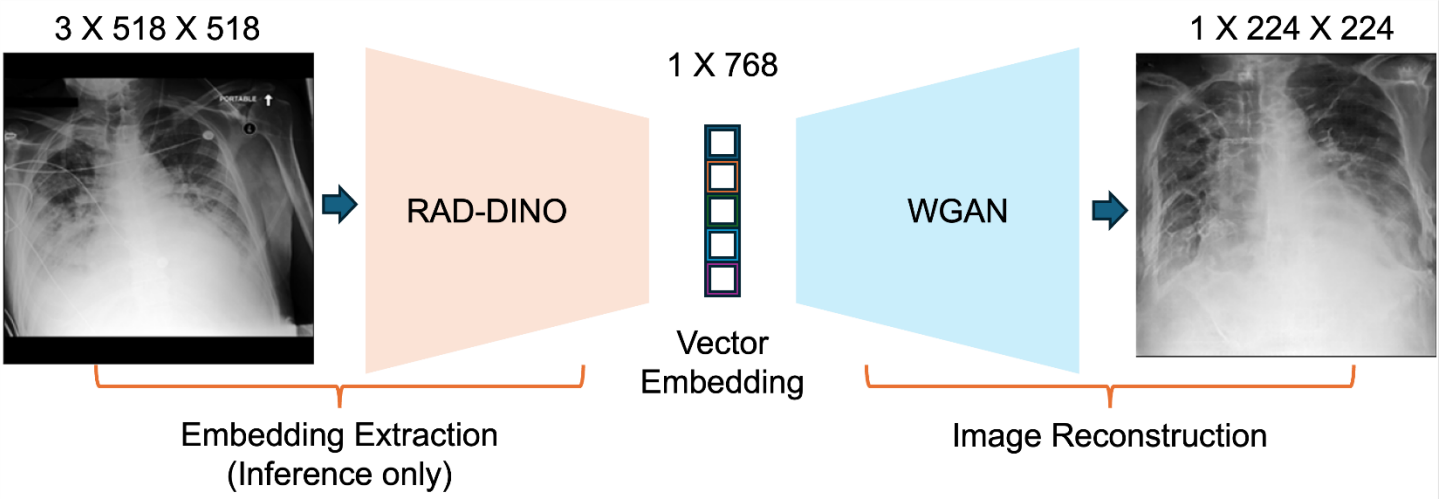

Figure 1. Framework of this study. Vector embeddings firstly were extracted from a private CXR dataset using RAD-DINO and then a WGAN was developed for image reconstruction using the extracted vector embeddings. Note that RAD-DINO was not involved in the training process of WGAN.

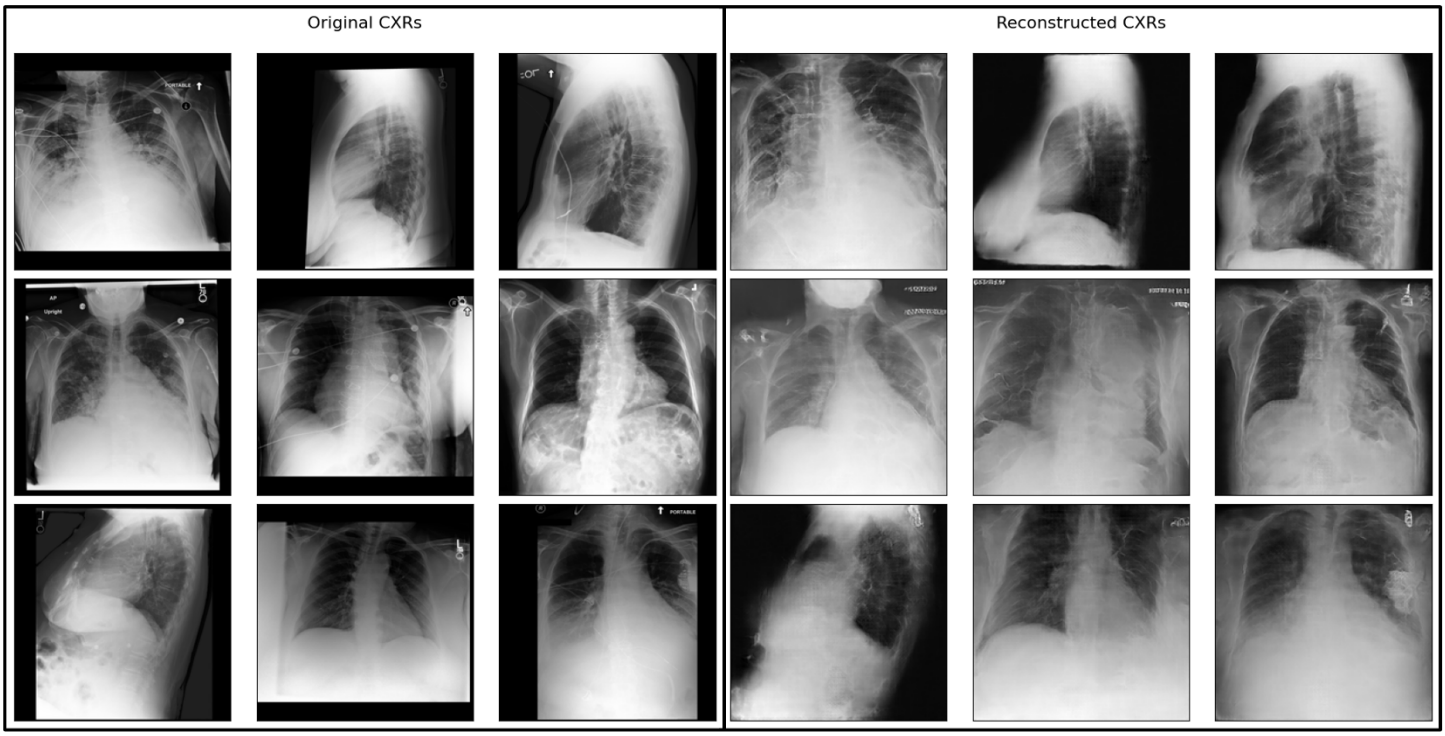

Figure 2. Examples of the reconstructed (right) and matching original (left) MIMIC CXRs showing difference in contrast and radiographic markers but maintaining overall similarity in anatomical structures.

Keywords

Artificial Intelligence/Machine Learning; Imaging Research