Transformers for CT Reconstruction From Monoplanar and Biplanar Radiographs

Firas Khader[1], Gustav Müller-Franzes[1], Tianyu Han[1], Sven Nebelung[1], Christiane Kuhl[1], Johannes Stegmaier[2], Daniel Truhn[1]

[1]Department of Diagnostic and Interventional Radiology, University Hospital Aachen, Aachen, Germany
[2]Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany

Summary

• Proposed method leverages vector-quantization [1] and GPT-based [2] autoregressive modelling to synthesize CT images using only biplanar x-rays.
• By framing the underlying task as a language translation problem, the model can be adapted to also synthesize CT images solely from monoplanar x-rays.

Introduction

• X-ray radiographs and CT scans are primary imaging techniques used in clinical practice to diagnose abnormalities and injuries; the former offers 2D projections while the latter provides detailed 3D images, constructed by merging multiple X-rays from different angles.
• CT scans, while offering detailed imagery, come with higher radiation doses due to prolonged exposure times; conversely, radiographs are more radiation-friendly and cost-effective.
• Converting 3D CT scans into 2D radiographs is possible but results in significant data loss, making the reverse process (2D to 3D) challenging.
• Previous attempts to create CT scans from radiographs relied heavily on convolutional neural networks (CNNs) and GAN-based discriminators; however, these methods had limitations, including unsymmetric encoder-decoder networks and mode collapse issues.
• This study leverages recent advances in transformer architecture [3], framing the conversion of radiographs to CT scans as a language translation problem

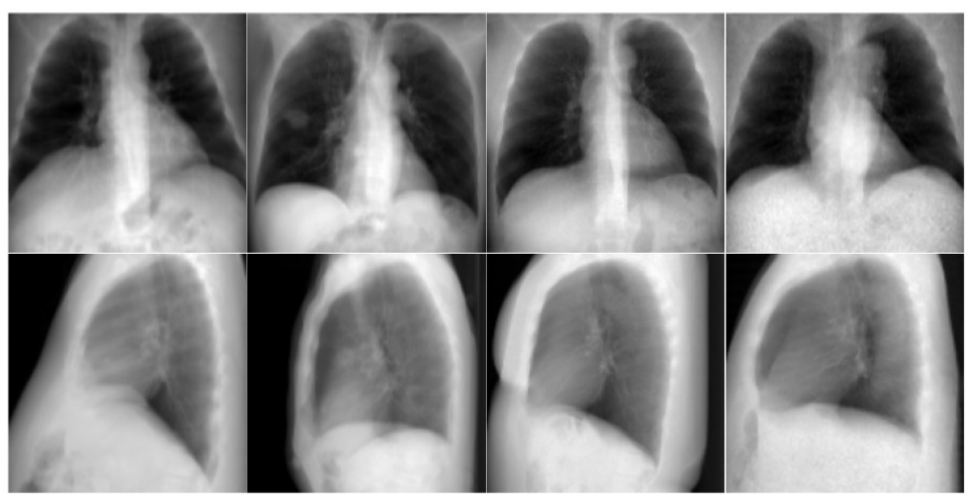
Materials & Methods

Dataset:

- Used LIDC-IDRI dataset comprising 1,018 diagnostic and lung-screening thoracic CT scans from 1,010 patients.
- Data was divided into training (70%, 707 patients), validation (20%, 202 patients), and testing (10%, 101 patients) subsets to facilitate the model training and evaluation.
- The CT scans from the dataset underwent several pre-processing steps including conversion into Hounsfield units, resampling into an isotropic voxel spacing of 1mm in all directions, and resizing to a standard shape of 120×120×120 while normalizing the value range between -1 and 1.

Digitally Reconstructed Radiographs:

- Synthetic chest radiographs in both lateral and posterior-anterior views were generated using digitally reconstructed radiographs technique, which involved converting the voxel values from Hounsfield units to linear attenuation coefficients.



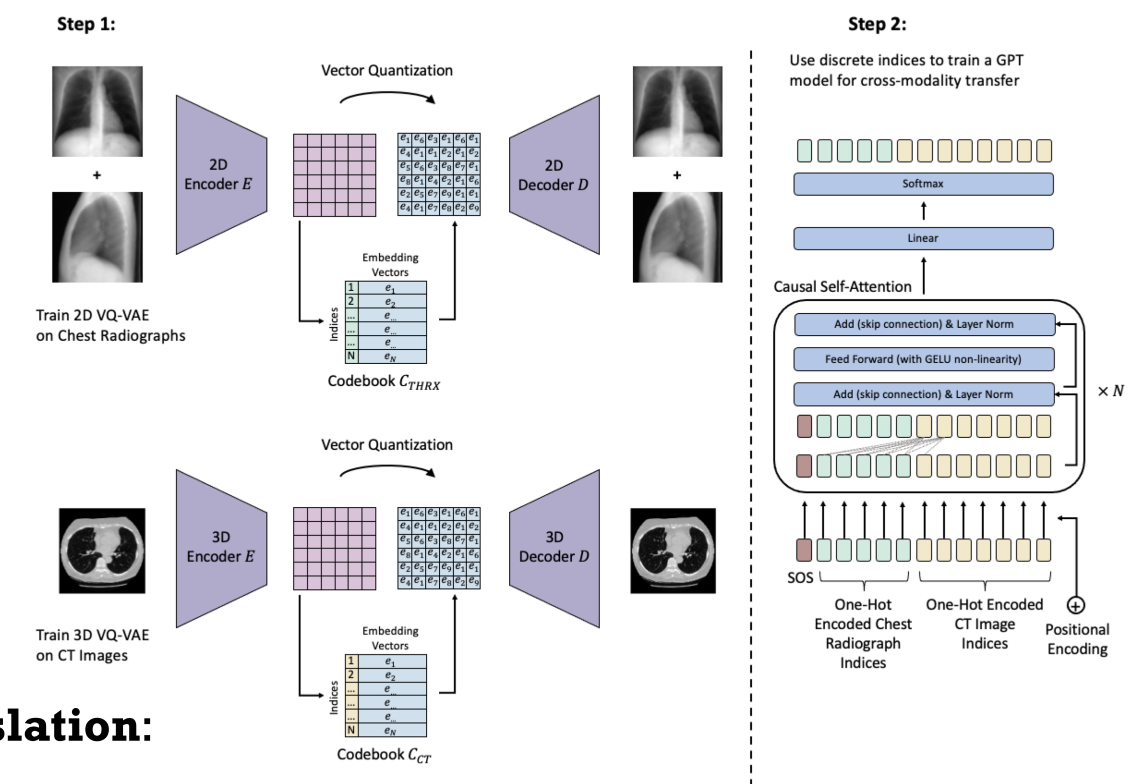
Sample Images of Digitally Reconstructed Radiographs.

$$x_{projection} = I_0 \exp(-\sum_{i=1}^{n} u_i d_i)$$

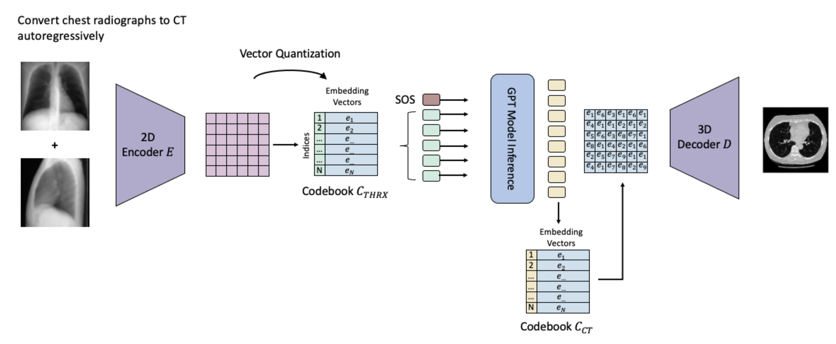
I_0: Signal Intensity of X-ray photons (here, 1kV)
n : # Voxels that the X-ray photons pass through
u_i : Linear attenuation coefficient
d_i : Voxel depth

Vector-Quantization: Utilizes VQ-GAN models for vector quantization to create discrete representations of 2D radiographs and 3D CT volumes.

GPT-based Language Translation: Employs a trained autoregressive GPT model for modality transfer, using a sequence of token representations derived from biplanar chest radiographs and maintaining causal self-attention within transformer blocks.



Model Architecture.



Inference Procedure.

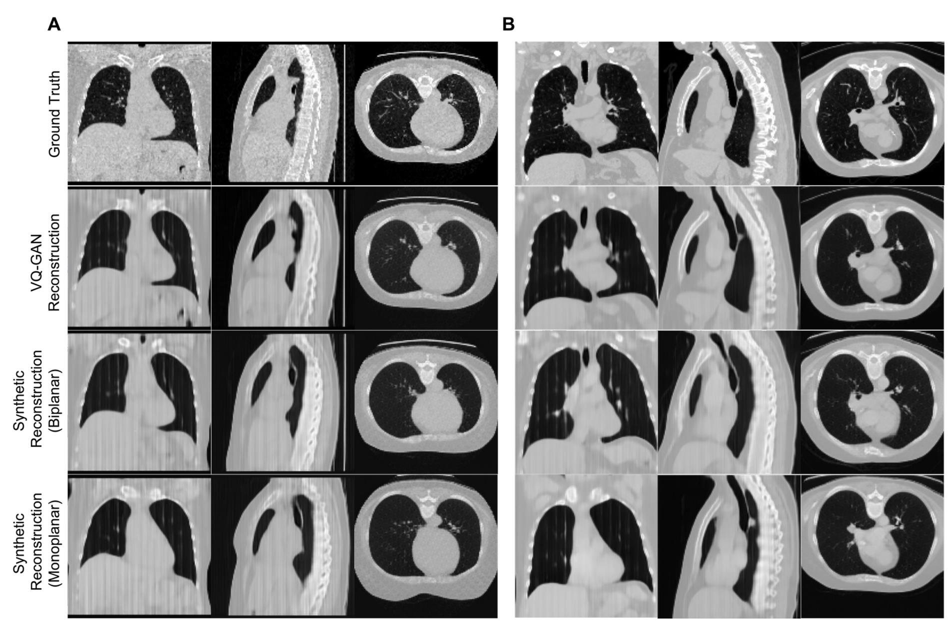
Inference:

In the inference stage, unseen radiographs are transformed into synthetic CT images through a series of transformations utilizing trained GPT and VQ-GAN models, supporting synthetic CT creation from a single radiograph..

Results

Biplanar CT Reconstruction:

The biplanar CT reconstruction yielded realistic general anatomical outlines but struggled with the faithful reconstruction of finer details such as internal structures of organs; ratings from a professional radiologist reflected satisfactory reconstruction of general heart, lung, and bone outlines.



Monoplanar CT Reconstruction:

Monoplanar reconstruction could generate full CT volumes using only posterior-anterior radiographs without model modification, demonstrating the architecture's flexibility to handle missing data and maintain a similar trend in the radiologist's ratings, albeit with slightly decreased scores compared to the biplanar approach.

Conclusion

• Developed a transformer-based method for converting 2D radiographs to 3D CT scans, successfully capturing large organ outlines but facing limitations in reconstructing fine internal details.
• Architecture supports reconstruction from single radiographs and may potentially integrate other modalities; large-scale testing is planned.

References

[1] Esser, Patrick, Robin Rombach, and Bjorn Ommer. "Taming transformers for high-resolution image synthesis." CVPR 2021
[2] Brown, Tom, et al. "Language models are few-shot learners." NIPS (2020)
[3] Vaswani, Ashish, et al. "Attention is all you need." NIPS (2017).

UNIKLINIK RWTHAACHEN

RWTHAACHEN UNIVERSITY