Generating Synthetic Chest X-Rays with generative modeling

Abhishek Kumar, Juan Pablo Triana Martinez.
Stanford University

Background/Introduction

- Chest X-rays is one of the most relevant diagnostic tool, non-invasive in medical imaging.
- Current work: GAN networks, Diffusion models with U net, VAEs, have been used to generate CXR images; yet there is no definitive which is the best performer.
- Network should be able to capture anatomical structures, health conditions, and different angles.

Problem and Related work

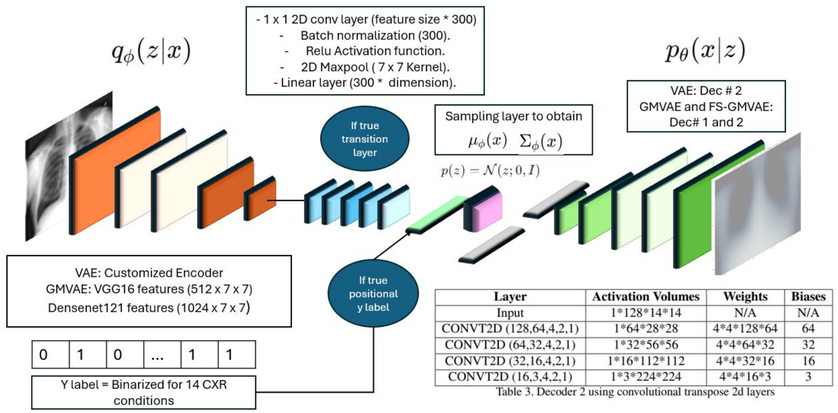
- To create a reliable data augmentation technique, using generative models, to produce CXR images.
- GAN networks have been used most of the time; and then tested with pretrained classification using VGG16; showcasing improvement around 3%.
- During COVID-19 pandemic; latent diffusion models and GANs were used for reliable data augmentation for classification of patients with COVID-19.
- Chambon et. al. developed Roentgen using a latent diffusion model composed of: VAE, conditional U- net, and conditional text encoder for radiologists notes.

Dataset: CXR14

Condition	Count
No Finding	59406
Infiltration	19894
Effusion	13317
Atelectasis	11559
Nodule	6331
Mass	5782
Pneumothorax	5302
Consolidation	4667
Pleural Thickening	3385
Cardiomegaly	2776
Emphysema	2516
Edema	2303
Subcutaneous Emphysema	1991
Fibrosis	1686
Pneumonia	1431
Tortuous Aorta	742
Calcification of the Aorta	455
Pneumoperitoneum	316
Pneumomediastinum	253
Hernia	227

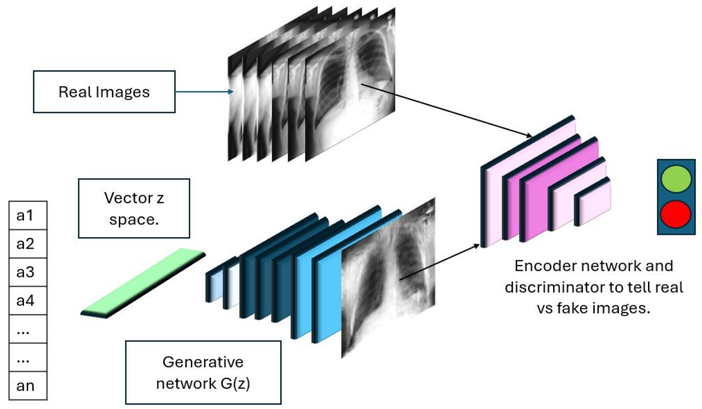
- The dataset includes CXRs in posterior anterior (PA), anterior posterior (AP), and lateral projections (LAT).
- Transformations done: resize images to 224, graychannel conversion, centered to crop, histogram equalization.
- For GMVAEs, a normalization block was used, changing images to blue colour.

Generative Architectures

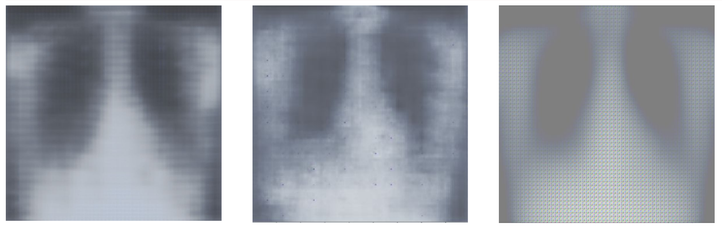


Layer	Activation Volumes	Weights	Biases
Input	1*128*14*14	N/A	N/A
CONVT2D (128,64,4,2,1)	1*64*28*28	4*4*128*64	64
CONVT2D (64,32,4,2,1)	1*32*56*56	4*4*64*32	32
CONVT2D (32,16,4,2,1)	1*16*112*112	4*4*32*16	16
CONVT2D (16,3,4,2,1)	1*3*224*224	4*4*16*3	3

Table 3. Decoder 2 using convolutional transpose 2d layers

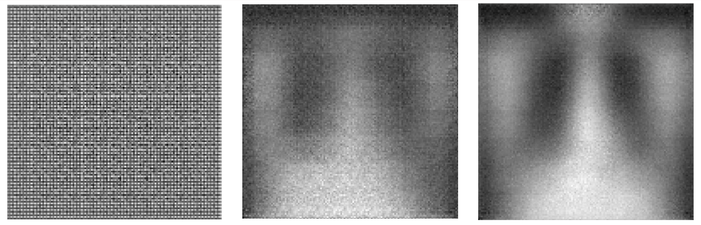


Results GMVAE



Z space length = 1000; K = 50; retrieved at iteration 2500
Results showcase VGG16 encoder where: 1st uses transitional layer and positional y label; 2nd uses transitional layer, without positional y label; 3rd doesn't use transitional layer, and positional y label.

Results VAE



Samples from VAE with Beta = 0.1 at iteration 0, 100 and 250

Results GAN

Samples from GAN at iteration 0, iteration 400, and final iteration after 5 epochs.

Discussion & Future Work

We found that GANs produced qualitatively better results than VAE and GMVAE; as well as using less parameters.

We think that our VAEs did not perform well because of the Gaussian Prior assumption on the distribution of latent space. Future work involves using much more complex ones; such as: Dirichlet or Cauchy.

Once the generative models produce better quality crispier images; quantitative metrics such as frechet distance or kolmogorov smirnov test can be used.