Creative Chest X-Ray Augmentation Using Cycle GAN

TECHNISCHE UNIVERSITÄT DARMSTADT

Leiter Christoph
Memmel Marius
Wempe Leon
Kubon Philipp

Converting chest X-rays of healthy patients to X-rays with pneumonia

Task Description & Use Cases

Pneumonia is a common disease of the lung, caused by either bacteria or viruses. (Wiki 2019a)

The goal of this project is to write a tool, which will take an X-Ray image of a person without a disease and alter a marked region to show a diseased lung instead.

Input:

X-Ray of a healthy person with a selected area

Output:

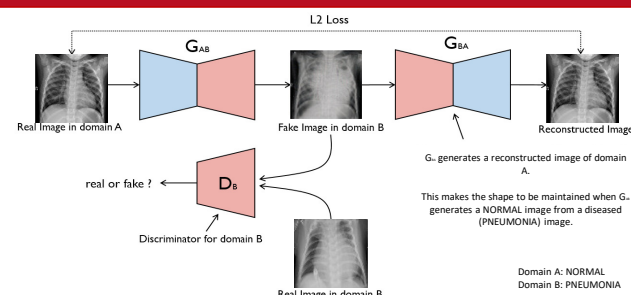X-Ray of same person with Pneumonia in the marked area

Use Cases:
- Simulate diseased lungs for training purposes
- Predict edge cases (how would pneumonia look like in a region no information is known about)
- Training classification models on adversarial examples can prevent attacks
- Work as a POC for similar tasks

Data Preprocessing

Preprocessing Steps:
- Combine train, validation and test set as no differentiation is needed for the GAN setup
- Remove images smaller than a certain threshold in width and in height
- Remove letters in the image by turning pixels with a high white value to black

Modeling – Cycle GAN

L2 Loss

G_{AB} G_{BA}

Real Image in domain A → Fake Image in domain B → Reconstructed Image

real or fake ? ← D_B

Discriminator for domain B

Real Image in domain B

G_{BA} generates a reconstructed image of domain A.

This makes the shape to be maintained when G_{BA} generates a NORMAL image from a diseased (PNEUMONIA) image.
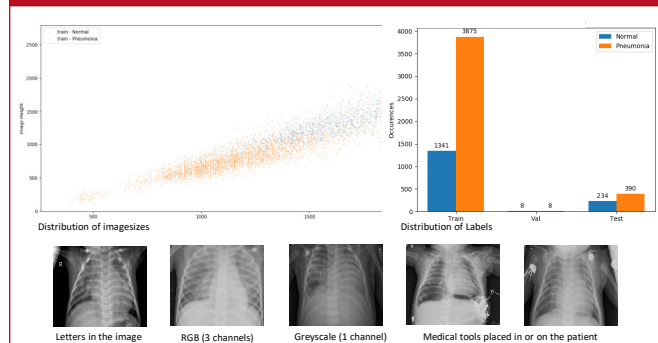
Domain A: NORMAL
Domain B: PNEUMONIA

- Transfers an image from one domain into another and back
- Trained over complete network
- Generator: Encoder-Decoder Architecture
- Discriminator: CNN Architecture
- Loss: Cycle consistency loss

References

- Jason Brownlee, PhD, 2019, https://machinelearningmastery.com/cyclegan-tutorial-with-keras
- Paul Mooney, 2018, https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
- Yunjey Choi, 2017, https://github.com/yunjey/mnist-svhn-transfer (edited)

Dataset Exploration

Distribution of imagesizes

Distribution of Labels
Train: Normal 1341, Pneumonia 3875
Val: 8, 8
Test: 234, 390

Letters in the image | RGB (3 channels) | Greyscale (1 channel) | Medical tools placed in or on the patient
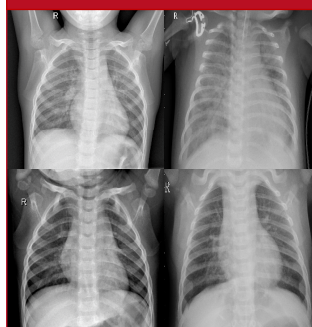
Results:
- Class imbalance towards PNEUMONIA
- Images of healthy lungs tend to be bigger
- Blob of low-resolution images
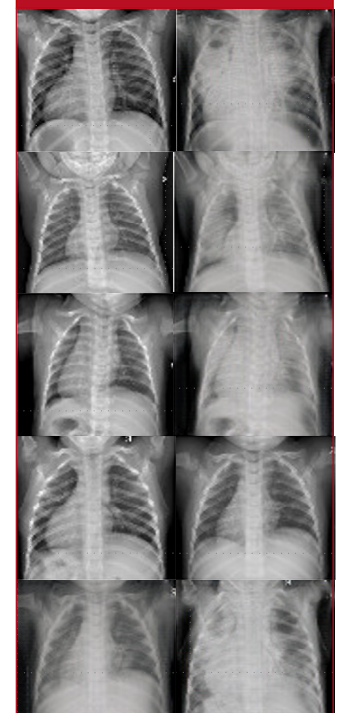- Some images show problematic distinguishing factors

DATASET

NORMAL   **PNEUMONIA**

GENERATED

NORMAL   **PNEUMONIA**

Evaluation & Results

Evaluation:
- Cycle GAN successfully generates diseased lungs from healthy samples
- Objective analysis shows that only very diseased lungs are generated
- No inclusion of mask because training data was missing
- -> no region for pneumonia can be marked in the generation process.