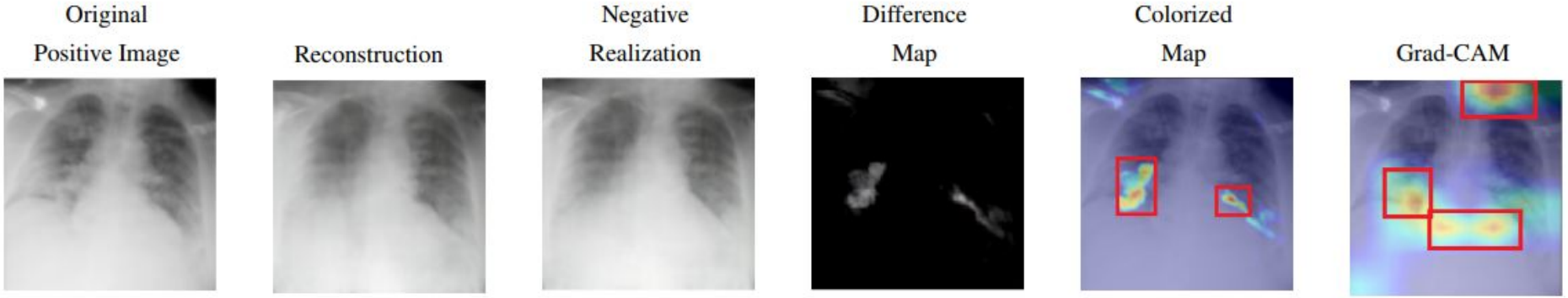
medXGAN

Visual Explanations for Medical Classifiers through a Generative Latent Space
CVPR 2022

Paweł Pawlik, Michał Siennicki

Objective

We want to explain CNNs on covid dataset

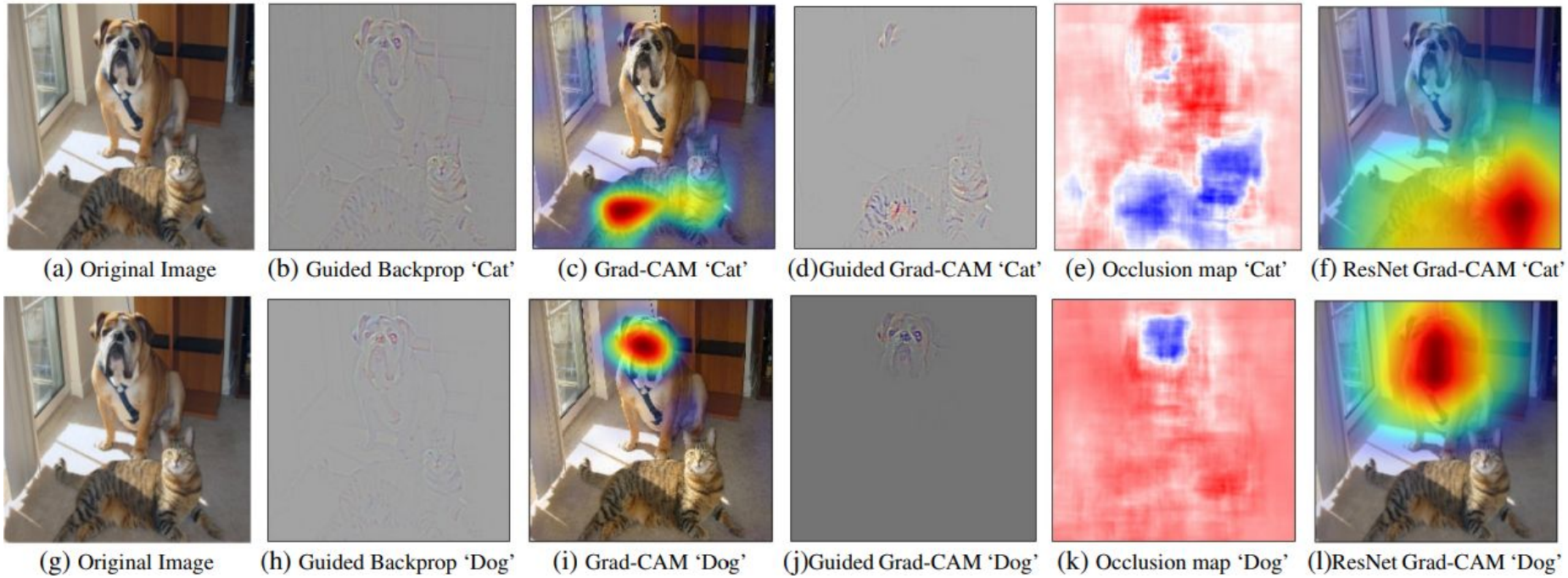


Plan

- Background
  - Grad-CAM
  - GANs
  - GANs - reconstruction
- medXGAN
  - method overview
  - results

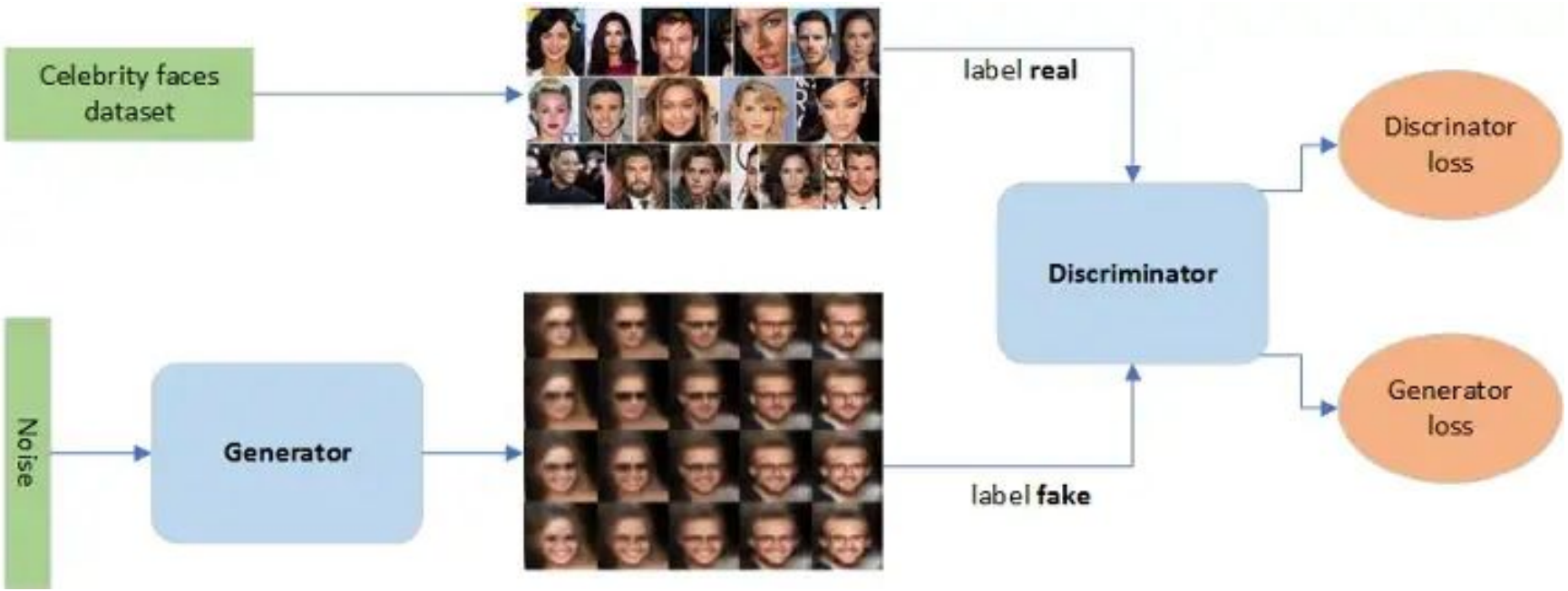
Background

Grad-CAM



Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization
2019, https://arxiv.org/abs/1610.02391

GANs


Celebrity faces dataset
label real
Discrinator loss
Discriminator
Noise
Generator
Generator loss
label fake

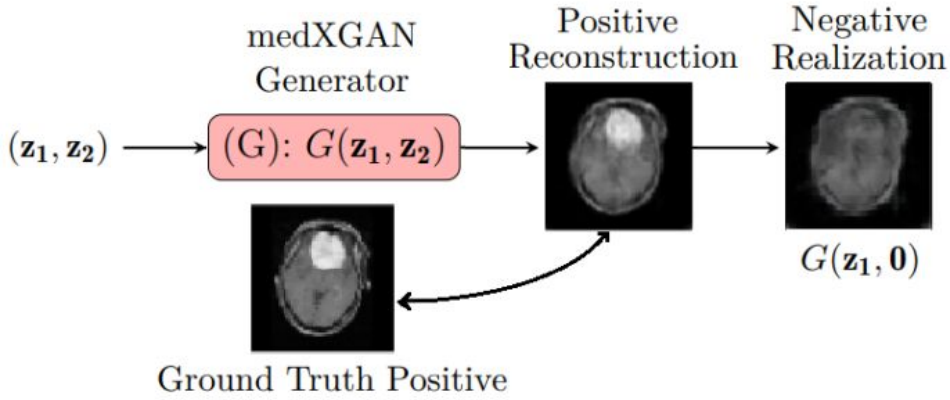
GANs - reconstruction

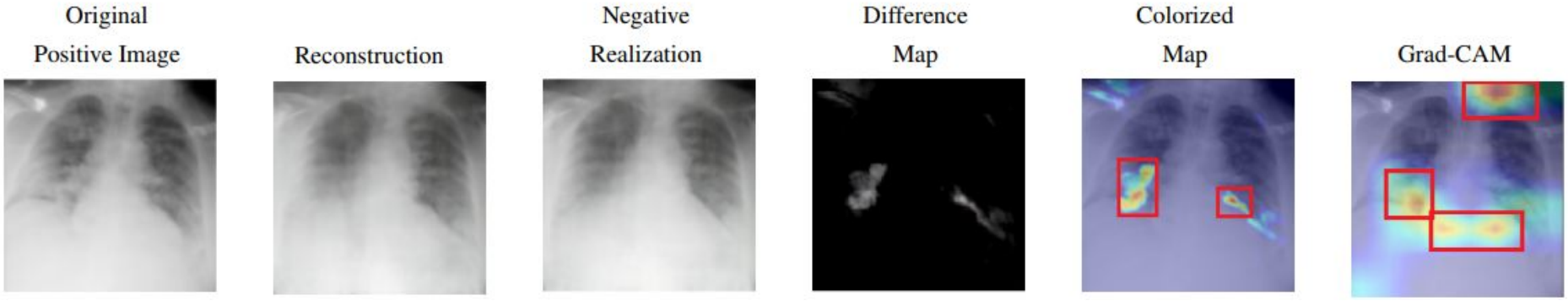
https://openaccess.thecvf.com/content/CVPR2021/papers/Richardson_Encoding_in_Style_A_StyleGAN_Encoder_for_Image-to-Image_Translation_CVPR_2021_paper.pdf

medXGAN

Objective

Find input image reconstruction and compare it with its negative realization

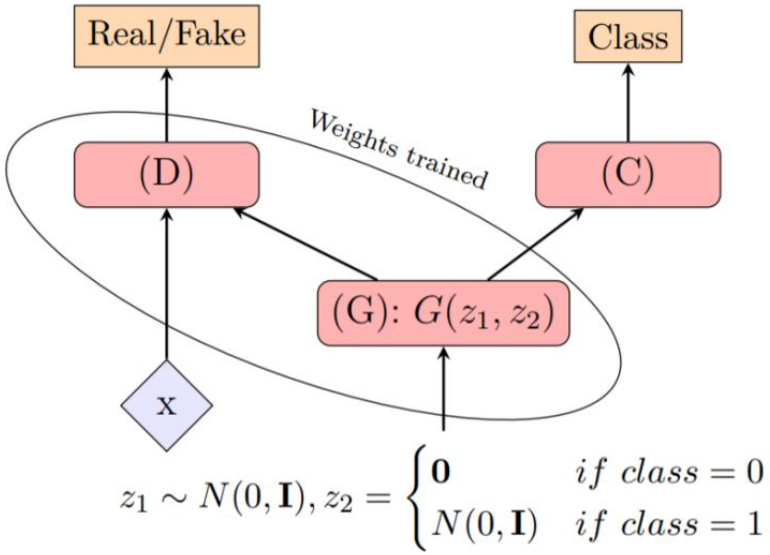


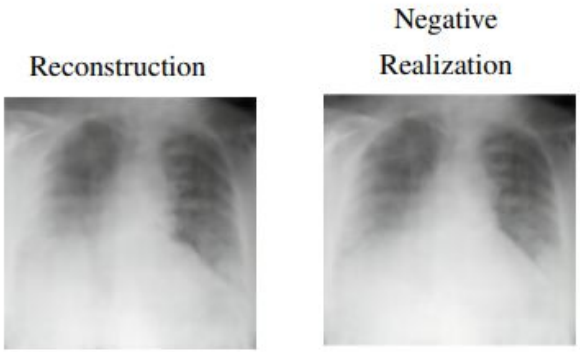


The method

We want to learn the following disantagled representation:

- z_1: lung features
- z_2: covid pathologies
  - **z_2 has high mutual information ($I(z_2, \hat{y})$) with the CNN classifier**

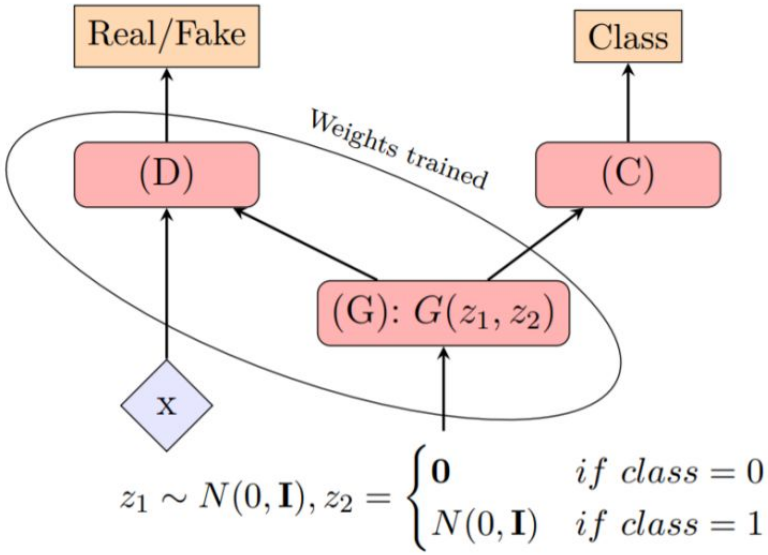




The method

We want to learn the following disantagled representation:

- z_1: lung features
- z_2: covid pathologies
  - **z_2 has high mutual information ($I(z_2, \hat{y})$) with the CNN classifier**



$$\min_G \max_D \mathop{\mathbb{E}}_{x \sim p_x} [\log D(x)] + \mathop{\mathbb{E}}_{z_1 \sim p_{z_1}, y \sim p_y} [\log(1 - D(G(z_1, y)))]$$

normal GAN objective

$$- \mathop{\mathbb{E}}_{z_1 \sim p_{z_1}, y \sim p_y} [\log(p_c(y|G(z_1, y)))]$$

mutual information

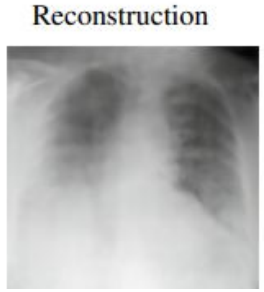
Sampling process

- Input image:

- Find reconstruction latent vector z_1:z_2

$$\arg\min_{z_1, z_2} \text{MSE}(G(z_1, z_2), x) + \text{BCE}(C(G(z_1, z_2)), C(x))$$

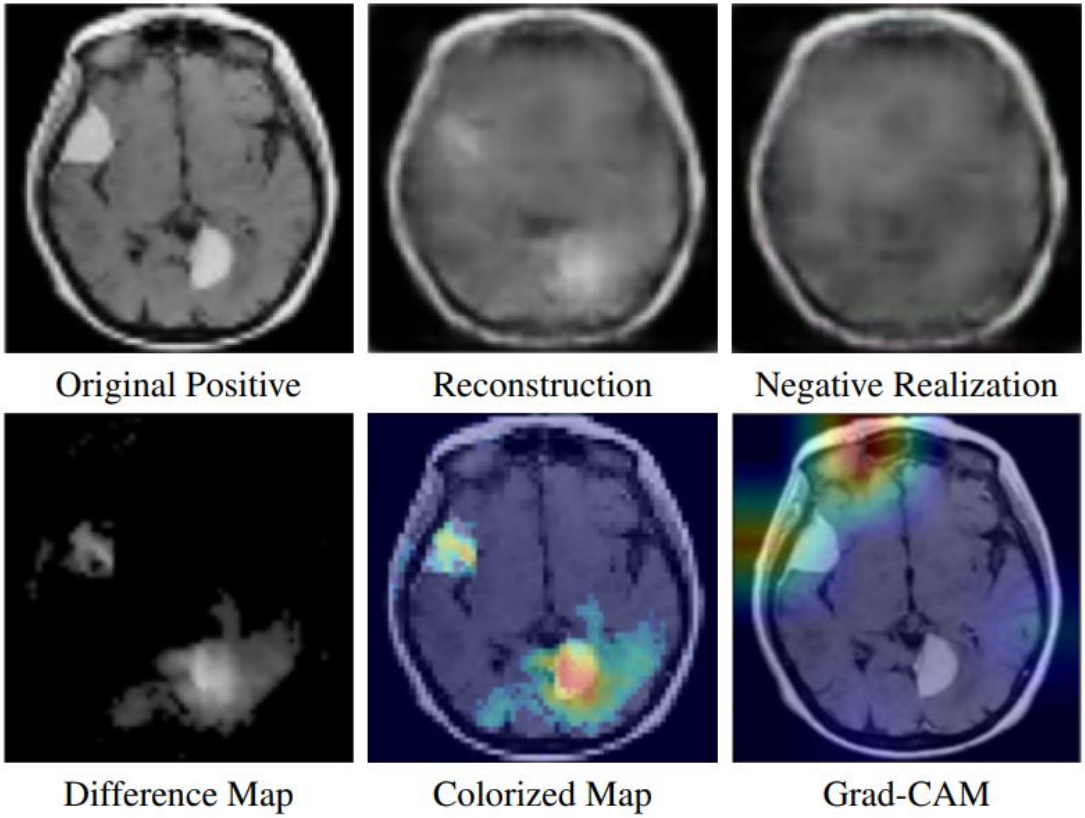
- The reconstruction: G(z_1, z_2)



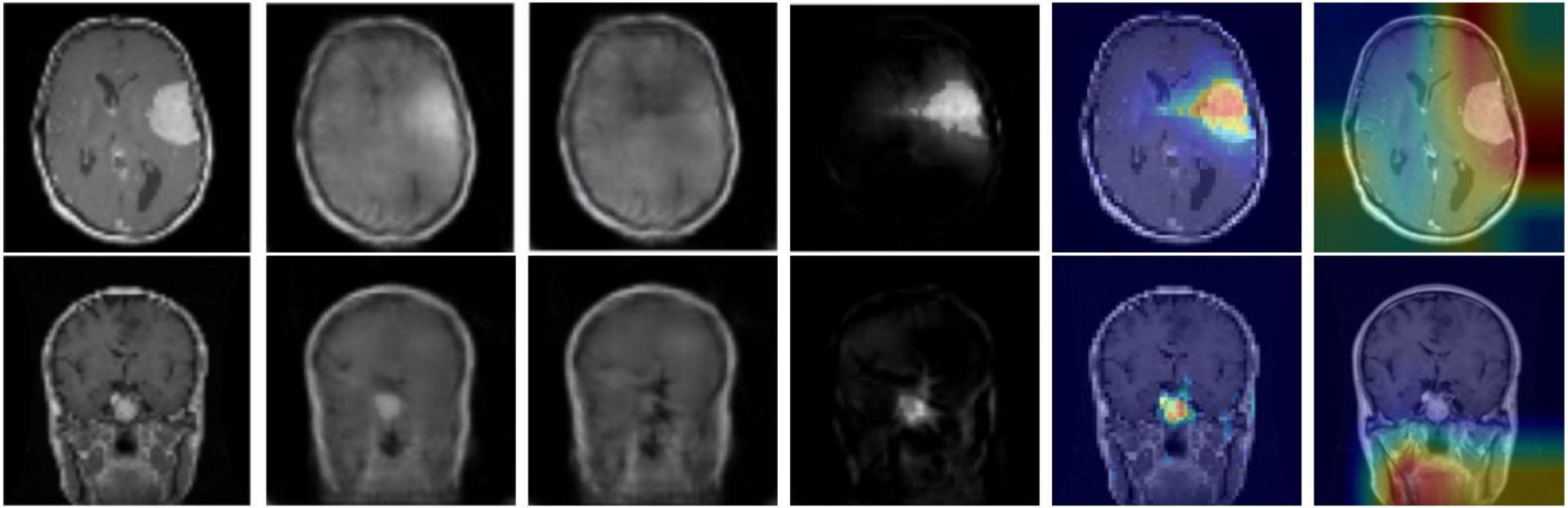
- Negative realization: G(z_1, 0)

medXGAN and Grad-CAM on Brain MRIs

In this example, despite lacking a perfect reconstruction, the medXGAN method localizes two tumors, while Grad-CAM focuses on one tumor and an eye.

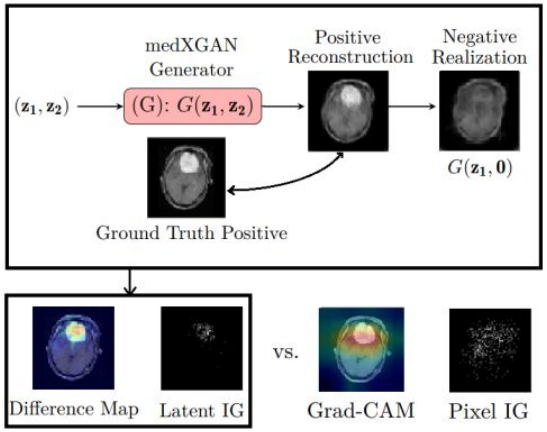


medXGAN and Grad-CAM on Brain MRIs



(a) Original (b) Reconstruction (c) Negative Realization (d) Difference Map (e) Colorized Map (f) Grad-CAM

Summary



- medXGAN can be used to replace
  - Grad-Cam
  - IG (integrated gradients)
- limitations
  - the space along which we are explaining the model must be continuous and independent
  - can only explain positive samples
  - GANs require significant amount of data to train
  - GANs have unstable training

Thank you!