Image-to-image translation for Computed Tomography and Magnetic Resonance Images.

Rubén Moreno Aguado

1. Introduction

Medical images come in many modalities that each provide different information for physicians and researchers. Specialized research tools are oftentimes designed for one modality and incompatible with others. Total Segmentator is a powerful Computerized Tomography (CT) image segmentation model. We hope to extend the functionality of Total Segmentator to accept Magnetic Resonance (MR) images or other modalities by training an image-to-image translation model that can change the style of a MR image into the style of a CTs.

There are several existing approaches to image-to-image translation, some of which are designed for paired datasets, and others for unpaired datasets. A paired dataset of CTs and MRIs allows for a simple neural network framework but is difficult to acquire since it would require performing both experiments on a patient and then aligning them with an image registration algorithm that may warp the features of an image. Unpaired datasets are abundant and don’t require significant image processing, however training the neural network will involve a self-supervised approach.

1.1 UNet

The most common approach for a paired dataset is a UNet, which is a deep CNN. It partially compresses and later decompresses the embeddings, while utilizing skip connections.

1.2 Pix2Pix

Pix2Pix is another commonly used framework for paired datasets, it utilizes a Generative Adversarial Network framework.

1.3 Neural Style Transfer

Old approaches to unpaired datasets used Neural style transfer. A pretrained deep CNN is used to extract style and content representations from an image by taking embeddings from an intermediate layer. Styles and contents from different sources can then be combined. Limitations of this approach is that the separation of content and style of an image are not strictly defined where some content information may be mistaken as style and carried over.

1.4 CycleGAN

A more modern approach is CycleGAN. It leverages the requirement of cycle preservation to promote a one-to-one conversion between images, even though the real pair of an image doesn’t exist. They have been previously applied to medical images, with the researchers noting that inequalities in the dataset would manifest in features being added and removed during mapping. For instance, if one dataset contained a greater distribution of tumors, the generator would remove the tumors when converting to the other modality. The inexactness of this approach makes it unreliable for applications in medicine.

2. Proposed model

A possible solution to the shortcomings of the CycleGAN may be to use transformer blocks instead of CNNs. Transformers have a lesser inductive bias than CNNs and therefore may try to reconstruct the features of the image as they are instead of trying to match arbitrary features from the dataset. However, transformers are more costly to train and require larger datasets to match the performance of CNNs.

Using a masked autoencoder, a ViT can learn to reconstruct missing areas of an image, however the synthetic image is usually quite blurry. This is useful as a pretraining approach but not helpful for the task of image generation. A GAN has been used in tandem with a MAE to generate higher quality reconstructions. It leverages weight sharing, where the discriminator shares the encoder’s weights with the generator.

We will extend this application to incorporate the CycleGAN arquitecture, where a single encoder will be shared by the discriminator, a CT generator, and an MR generator.

Two training objectives: cycle preservation, gan loss.

During the GAN stage, real images are masked 80%.

During cycle preservation stage, no masking.

3. Data collection

4. Discussion