INTERNSHIP REPORT: BRAIN TUMOR SEGMENTATION

Hakan Ak
Department of Electrical and Electronics Engineering
Bogazici University
Istanbul, Turkey 34342
hakan.ak@boun.edu.tr

October 25, 2023

ABSTRACT

During my summer internship, I worked on the field of medical image segmentation, focusing on image segmentation and image translation with generative models. My experience began with familiarizing myself with tools and structrues such as UNet, nnUNet, GAN and Diffusion models. and transitioned into exploration of image translation models, Pix2Pix, and CycleGAN.

In the context of image segmentation, my work included reproducing results with nnUNet for brain tumor segmentation, training models with both 2D and 3D images, and research on MRI modalities. Shifting to the domain of generative models, I explored unpaired image translation between MRI modalities. This involved understanding CycleGAN and Pix2Pix models, addressing dataset-related challenges, and utilizing methods like position-based strategy and progressive translation. I have worked with BraTS21, HCP and UCSF brain datasets.

Results included successful model training, evaluation using metrics like FID scores, and insightful visualizations. Throughout, I collaborated with my research group closely and attended meetings to share progress. This internship allowed me to gain knowledge in medical image analysis, generative models, image segmentation, and image translation. Also, I have gained priceless research skills working with the team. I had the experience of being in a large project and observing how the process worked.

1 Introduction

Tumor segmentation is a crucial problem for healthcare. It is one of the main biomarkers for medical doctors to evaluate, decide and do operations. The project I was involved was focused on brain tumor segmentation, moreover, it was focused on the BraTS challenge.

One of the main challenges in this field is the scarcity of data. While there are state-of-the-art models developed for image segmentation, we still need a good amount of segmented brain tumor data to adapt the models for this purpose. Developing datasets are a hard and consuming job for brain tumors. Apart from it being a expensive process to get a MRI scan, and get it segmented by a doctor, it is also buero bureaucratically time consuming process.

To tackle this problem, we were focused of translating available brain-tumor datasets into the contrast of BraTS dataset, allowing us to enrich our training data.

In the first phase of the project, I was doing literature review on segmentation, GAN and diffusion networks. I have tried to implement available architectures and tried to reproduce them. Then, we were split in the groups for different task. I worked with image translation and modality expansion groups. Where we tried to enrich our dataset with translating from other datasets, as well as we were focused on enriching our available modalities by also doing image translation to extended modalities.

2 Literature Review on Medical Image Segmentation

2.1 Review of MRI & Modalities, Orientations

2.1.1 MRI Modalities

In the context of MRI, a "modality" refers to a specific imaging technique or method used to capture information about the internal structures and functions of the human body. Every modality is based on distinct physical principles and technologies, and it provides unique types of information that are valuable for medical diagnosis and treatment.

Although there are a few more, down below it is provided an overview of key MRI modalities in BraTS dataset and their relevance:

- T1-weighted (T1W):
  T1-weighted images highlight differences in the longitudinal relaxation time of tissues, offering excellent contrast between gray and white matter. These images are often used for anatomical reference and tissue classification in brain tumor segmentation.
- T2-weighted (T2W):
  T2-weighted images emphasize variations in the transverse relaxation time of tissues. They are valuable for identifying edema, necrosis, and other pathological changes.
- T1-Contrast Enhanced (T1CE):
  T1CE MRI is used to visualize blood-brain barrier disruption, enhancing the visualization of contrast-enhancing tumor regions. This modality plays a crucial role in identifying the extent of tumor growth and response to treatment.
- FLuid-Attenuated Inversion Recovery (FLAIR):
  FLAIR MRI suppresses cerebrospinal fluid (CSF) signals, enhancing the detection of lesions near CSF-filled spaces. It is invaluable for identifying peritumoral edema and tumor boundaries.

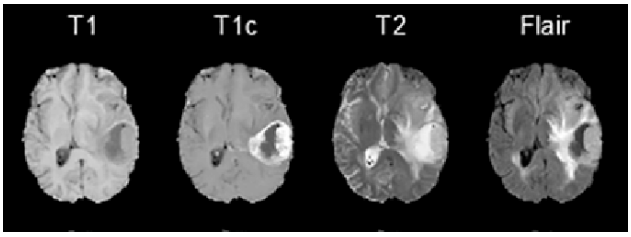


Figure 1: A sample of different MRI modalities in Axial orientation. [Hu et al., 2021]

In the context of brain tumor segmentation, the combined information from multiple MRI modalities is often used to create a comprehensive picture of tumor boundaries and characteristics. Each modality provides unique insights, and the integration of these data sources enhances the accuracy of segmentation and diagnosis.

2.1.2 Orientations in Medical Imaging

There are 3 different orientations that are widely being used in Medical Imaging. Those are Axial, Sagittal and Coronal axes.

- Axial Orientation:
  In this orientation, the body is divided into slices horizontally. It provides a view as if you were looking above from the head of a person.
- Sagittal Orientation:
  It slices a body onto left and right parts on vertical axis. It provides a view as if you were looking from the right or left side of a body.
- Coronal Orientation:
  It also slices a body verticaly, but on front and back. It provides a view as if you were facing the body.

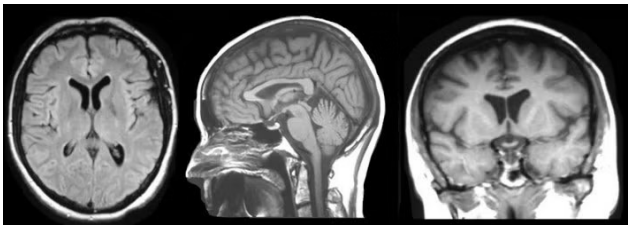


Figure 2: Axial, sagittal and coronal planes in an MRI image.

2.2 Review of U-Net Architecture

The U-Net was proposed by Ronneberger et al. [2015] and it has become a cornerstone in the field of medical image segmentation. The name of the architecture comes from its U-shaped nature.

The main idea of U-Net is to add additional layers to a typical contracting network, replacing pooling operations with upsampling operators. Thus, the resolution of the output is increased by these layers. Based on this information, a subsequent convolutional layer can then learn to put together a precise output.

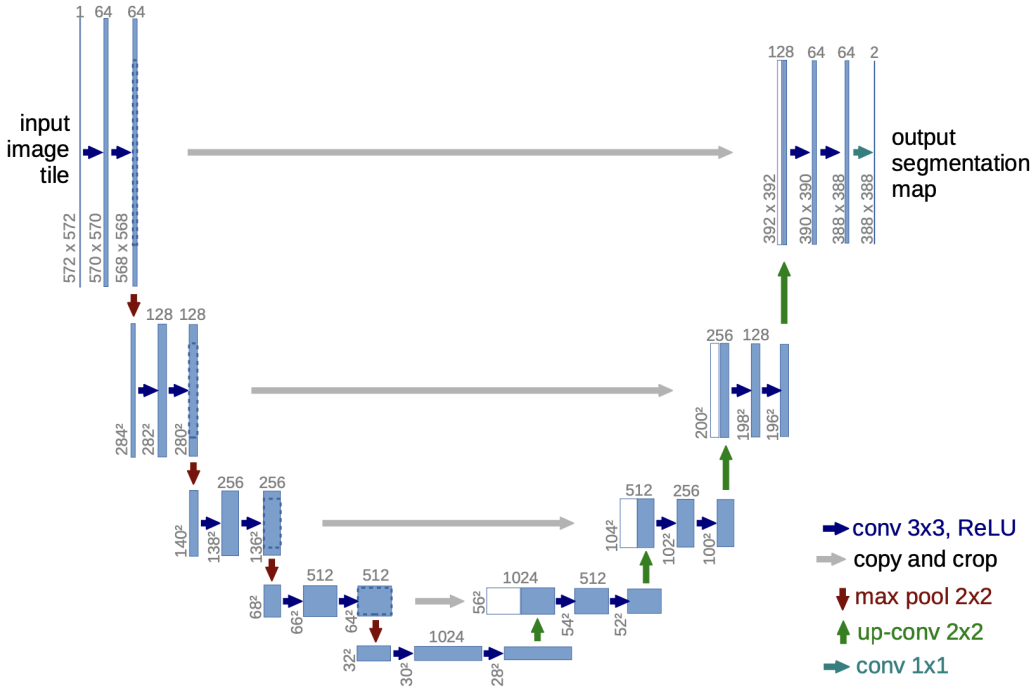


Figure 3: U-net architecture (example for 32x32 pixels in the lowest resolution). Blue boxes correspond to a multi-channel feature map. The numbers on top of the box are number of channels is denoted. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows show different operations as shown in legend [Ronneberger et al., 2015]

The network has a u-shaped architecture because it has both a contracting path and an expansive path. The contracting path is a standard convolutional network that applies convolutions multiple times, followed by rectified linear units (ReLU) and max pooling operations for each one. Spatial information is decreased while feature information is increased during the contraction. Through several up-convolutions and concatenations with high-resolution features from the contracting path, the expansive pathway combines the feature and spatial information. The overall architecture can be seen on ‘3.

U-Net is widely used as a base architecture for a lot of image segmentation and translation models. Over time, a lot of variations of U-Net was developed. One of them, nnUNet is a semantic segmentation method which adapts itself to the

given dataset Isensee et al. [2021]. It provides an easy-to-use pipeline for biomedical segmentation tasks. Over time, it become a baseline in the field.

3 Literature Review on Image Translation

Image translation is a process that involves converting an image from one domain or style into another while preserving its content and structure. It is being used for various purposes, including style-transfer, colorization, domain adaptation. Image translation typically involves the use of deep learning models, such as GANs or CNNs.

The image translation can be done in two different scenarios. With paired translation, we have the corresponding target images for the training images we have. It can also be done as unpaired translation where we don't have the exact target images for training data we have, but we have only two different data in different domains.

3.1 Generative Adversarial Networks (GAN)

GANs are a type of framework that estimate generative models through an adversarial process, in which two models—the generator (G) and discriminator (D) models—are trained concurrently. The discriminator model learns to identify the difference between the real data and the fake data generated by G, while the generator model learns to generate data that resembles the real data. Considering this, the objective function for training the GAN can be described as a two-player min-max game, in which the generator G aims to minimize the objective while the discriminator D aims to maximize it. [Goodfellow et al., 2014]

As training progresses, the generator G estimates or generates sample data that can fool the discriminator D, and the discriminator D eventually becomes terrible at identifying between real and fake data, and its accuracy decreases. The architecture can be seen in Figure 4.

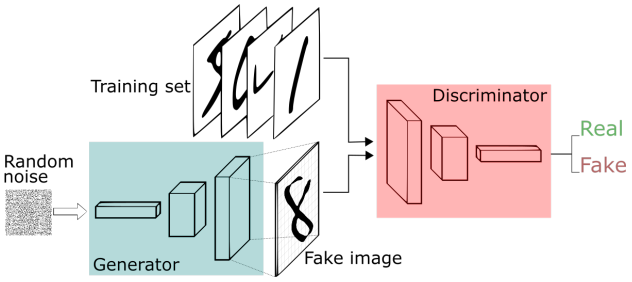


Figure 4: Illustration of GAN [Osterburg, 2019]

3.2 Pix2PixGAN: Paired Translation

Pix2Pix is built on the conditional GAN architecture. Conditional GANs, also known as cGANs, are a type of GAN architecture. To be used for translation, cGANs are trained on paired sets of images or scenes from two domains. Because GANs are trained to generate fake samples from the entire set of training datasets, the results are frequently subpar. As a result, we condition the generator and discriminator on an input image or some assisting information, resulting in targeted image generation from the target domain. As a result, cGANs are well-suited to image-to-image translation tasks, in which we condition on an input image and generate a corresponding output image. [Isola et al., 2017]

Pix2pix's architecture consists of a Generator G and a Discriminator D. Discriminator is a patch-GAN architecture that penalizes at the scale of patches, whereas Generator G is an encoder-decoder net or U-Net with skip connection. The architecture can be seen in Figure 5.

The patch-GAN discriminator is a unique component added to the pix2pix architecture. It works by classifying a patch of (n*n) in an image as real or fake rather than the entire image as real or fake. More constraints are imposed, which encourages sharp high frequency details. This method is faster than classifying the entire image and has fewer parameters. As input, the discriminator accepts two image pairs: input image, target image, and generated image. We join these two input pairs.

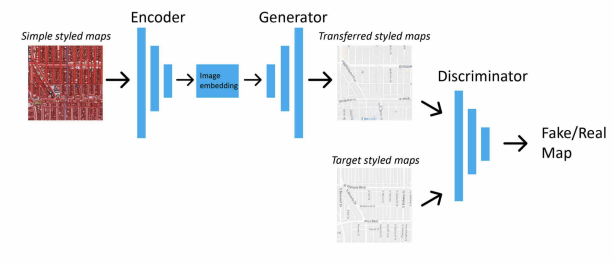


Figure 5: Flow of Pix2Pix architecture [Kang et al., 2019]

3.3 CycleGAN: Unpaired Translation

CycleGAN, like Pix2Pix, is an image-to-image translation model. The main challenge in the Pix2Pix model is that the data required for training must be paired, which means that the images of the source and target domains must be in the same location and have the same number of images. Using an unpaired dataset, CycleGAN network learns mapping between input and output images. [Zhu et al., 2017]

This model is an extension of the Pix2Pix architecture that involves training two generator models and two discriminator models at the same time. In addition to Pix2Pix features, we can use unpaired datasets and convert images in the reverse direction (target to source imagery) with the same model. The overall architecture can be seen in Figure 6.

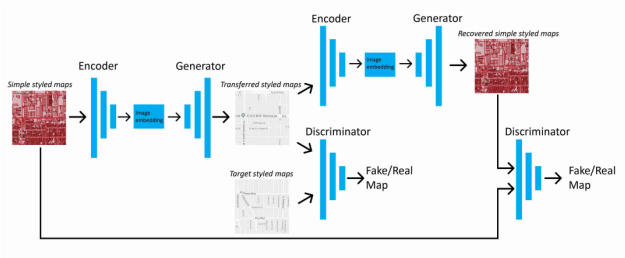


Figure 6: Overview of CycleGAN model architecture [Kang et al., 2019]

The model architecture is made from two generator models: one for generating images for the first domain (Domain-A) and another for generating images for the second domain (Domain-B).

- Domain-B → Generator-A → Domain-A
- Domain-A → Generator-B → Domain-B

There is a discriminator model for each generator (Discriminator-A and Discriminator-B). The discriminator model predicts whether real images from Domain and generated images from Generator are real or fake.

- Domain-A → Discriminator-A → [Real/Fake]
- Domain-B → Generator-A → Discriminator-A → [Real/Fake]
- Domain-B → Discriminator-B → [Real/Fake]
- Domain-A → Generator-B → Discriminator-B → [Real/Fake]

Loss Calculations :

The loss for the model is consisting of three parts:

1. **Adversarial Loss**: We apply Adversarial Loss to both Generators, with the Generator attempting to generate images of its domain and the corresponding discriminator distinguishing between translated and real samples. The Generator seeks to minimize this loss while the Discriminator seeks to maximize it.
2. **Cycle Consistency Loss**: It expresses the idea that if we translate an image from one domain to another and back again, we should end up back where we started. As a result, it computes the L1 loss between the original

image and the final generated image, which should be identical to the original image. It is calculated in two ways:

- Forward: Domain-B → Generator-A → Domain-A → Generator-B → Domain-B
- Backward: Domain-A → Generator-B → Domain-B → Generator-A → Domain-A

3. **Identity loss**: It encourages the generator to keep the color composition of the input and output the same. This is accomplished by feeding the generator an image of its target domain as input and computing the L1 loss between the input and generated images. Such that:

- Domain-A → Generator-A → Domain-A
- Domain-B → Generator-B → Domain-B

4 Methods and Techniques

The goal for both tasks below was similar. To obtain more data to be used for NN training for segmentation network. Their difference comes from that they had different approaches for the solution. Also, they had been utilizing different datasets.

4.1 Image Translation Task

On the image translation side, the goal was to translate HCP or other healthy brain images to BraTS contrast while conserving the brain geometry to help model learn from healthy brain images with no tumors.

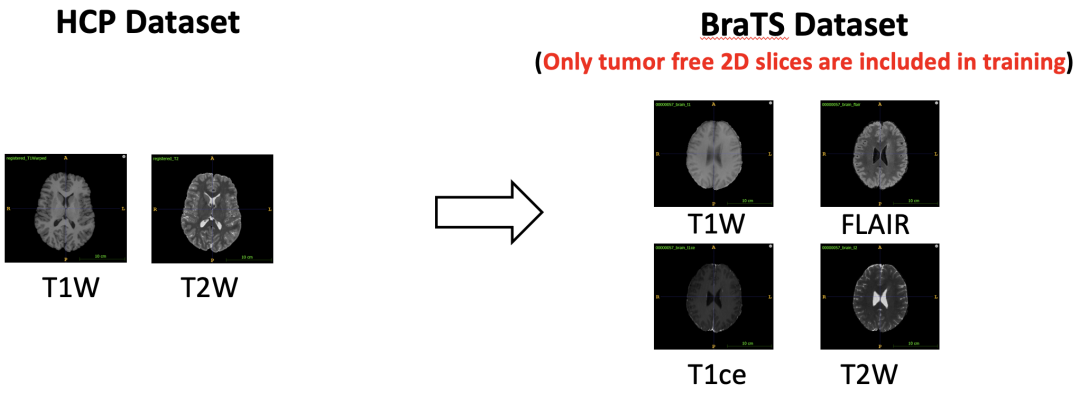


Figure 7: Visualisation of the goal.

Since BraTS data was created out for unhealthy brain data (with tumors), to create the BraTS contrast domain, the slices containing tumors was discarded from data, and only the slices that has **no tumors** were used.

One of the key challenges here was the **unavailability of any paired data**. This problem was emerging from the nature of the issue. Since the datasets was prepared by different institutions, even if they had used the same modalities, the contrast of the scans was different.

To overcome the issue of unpaired translation, the state-of-the-art unpaired image-to-image translation network CycleGAN was considered.

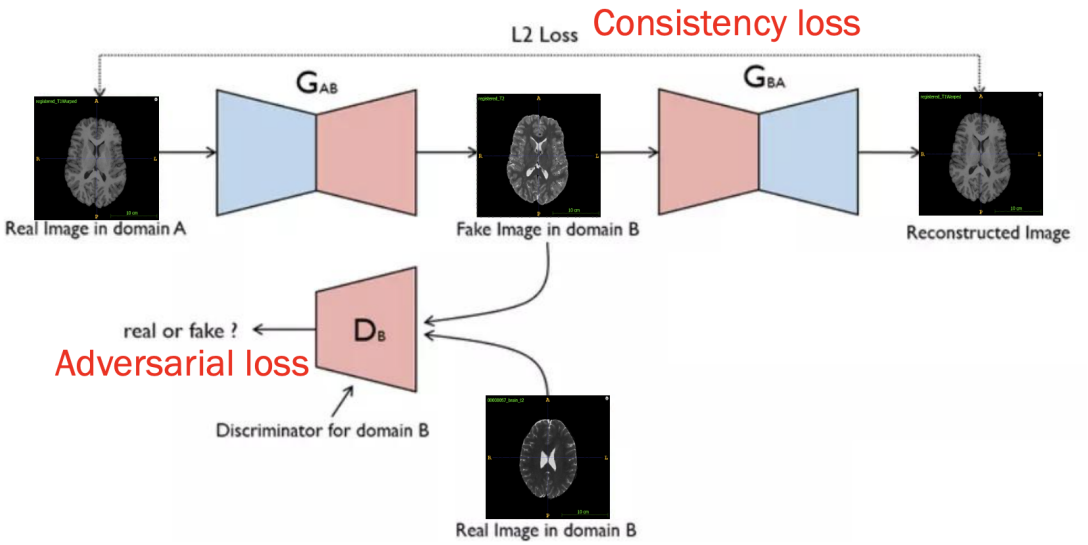


Figure 8: CycleGAN model representing the task.

With CycleGAN, the adversarial loss of the model was guaranteeing the contrast of the brain images while the consistency loss was guaranteeing the fidelity of brain geometry.

One another key issue was to **maintain the 3D consistency**. Since CycleGAN model was conducting the translation on 2D images, only the 2D slices of the 3D brain data could be translated one by one. It was obvious that it would create some inconsistency when the data was observed as a 3D image. To overcome this issue, progessive 2D translation [Yurt et al., 2022] was considered. See Figure 9.

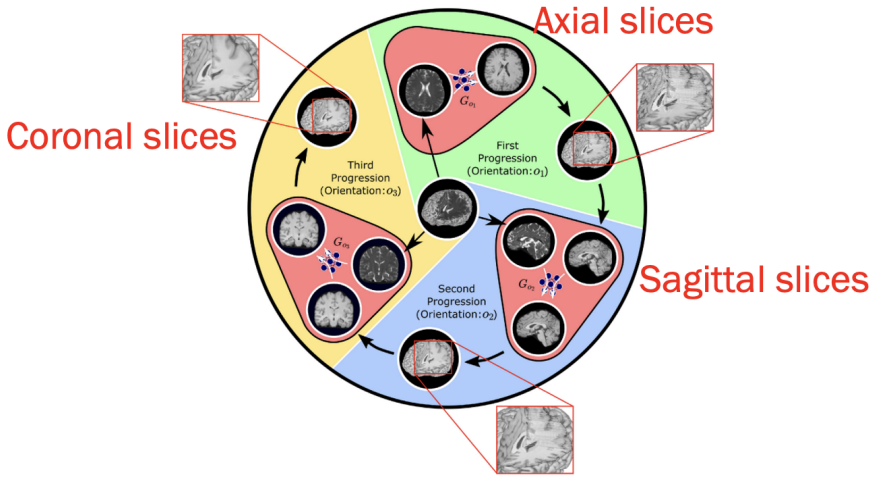


Figure 9: Visualization of progressive 2D image translation. [Yurt et al., 2022]

With this method, the 2D slices of 3D brain data was translated using 3 different neural networks. In the first part, axial slices of the data was first translated onto the target domain using a network trained on axial slices, then those trained slices were concatenated to form a 3D image. After that, the obtained 3D image was sliced onto sagittal axes, and this time, another **paired** pix2pixGAN network was trained from scratch from the same domain, but on sagittal orientation. Here, the paired network was used because now we have the same image for both modalities. Then the translation were executed. This procedure was also repeated for coronal slices.

There are eventually 3 methods for this task we have decided:

4.1.1 Hybrid paired and unpaired image translation

With this strategy, first an unpaired translation was conducted between the HCP T1W and BraTS T1W domains, then utilizing the available data on BraTS, a paired traning was done to the target domain of BraTS T2W.

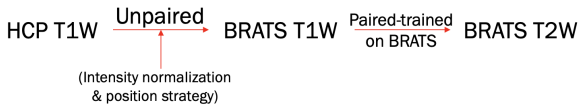


Figure 10: An example of the strategy 1.

4.1.2 Purely unpaired image translation

With this strategy, using only one unpaired model, from one HCP modality to one BraTS modality, the translation was conducted.

Figure 11: An example of strategy 2.

4.1.3 Purely paired-based image translation

In this strategy, a paired model, which was trained purely on BraTS data (since we only have the pairs within the BraTS data) was used to translate the images. Even if the model was solely trained on BraTS data, the hope was that it could perform well on the HCP data as well.

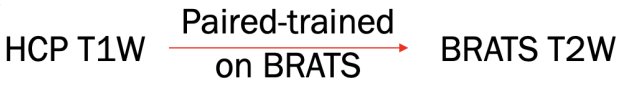


Figure 12: An example of strategy 3.

4.2 Modality Expansion Task

The main objective of this task was to expand the available modalities withing the BraTS dataset to 4 more available MRI modalities to expand the contrasts of our data. The pipeline can be seen in Figure 13.

Here, the model is being trained with 4 different BraTS modalities onto 1 UCSF modality. In total, we need to train 4 different networks for 4 missing UCSF modalities.

After the training of the networks, the conversion is planned to be done with inputting all of the BraTS data and then translate to each modality one by one.

5 Results and Discussion

5.1 Image Translation Task

Effect of Intensity Normalization Throughout the project, it was observed that the intensity differences within the dataset could lead some artifacts in the images. Thus, the intensities within the dataset were normalized. And the observations showed that it improved translation performance. See Figure 14

Effect of Position Based Strategy: The position-based strategy was proposed by Yang et al. [2020]. It proposes that instead of training the model with random selection of image pairs, selecting the corresponding slice in the target domain could increase the performance. It was initially proposed for MR to CT translation, we tried to apply it onto our task.

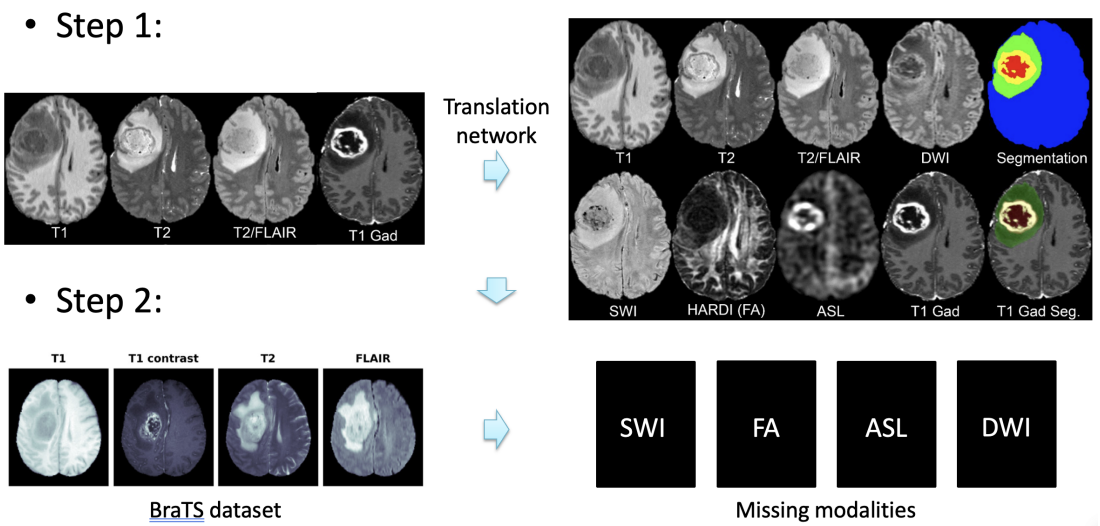


Figure 13: Overall pipeline for modality expansion

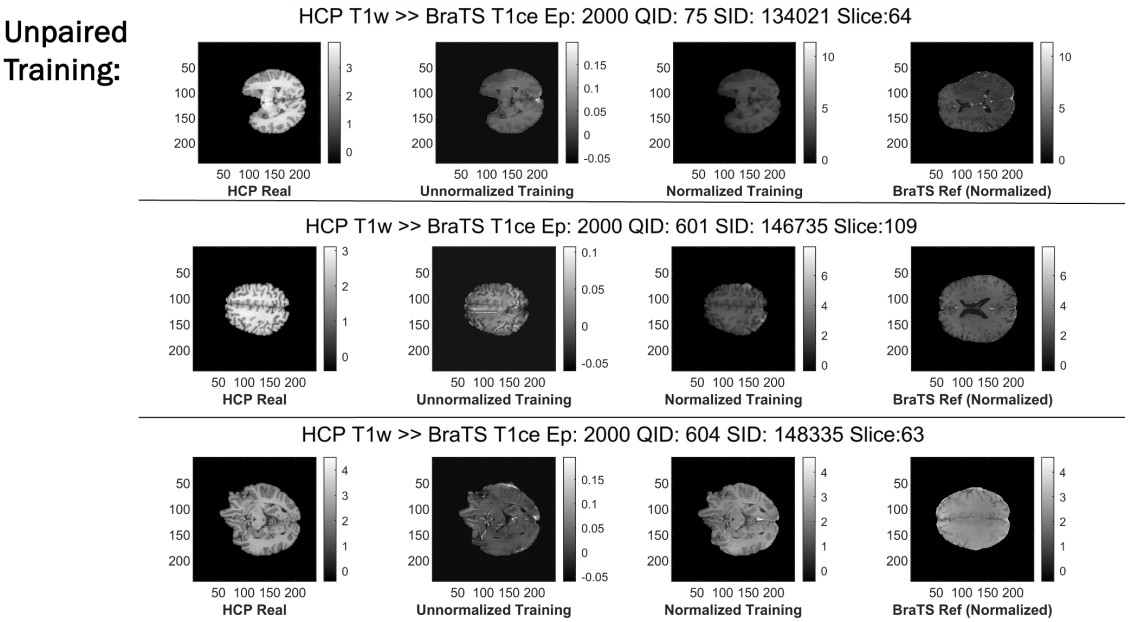


Figure 14: An example showing the effect of intensity normalization. It could be observed that the normalized images were closer to the contrast of the reference images.

I have conducted the implementation of this strategy on PyTorch. Also, we proposed a method of applying the position-based strategy on a range i.e., a random sample was chosen from the corresponding target range in each epoch.

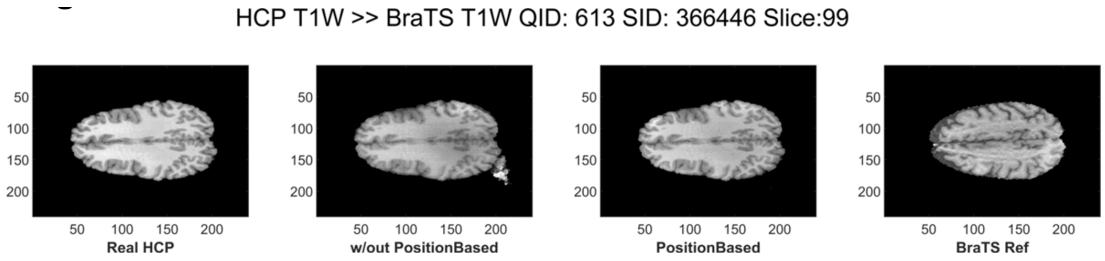


Figure 15: An example showing the effect of position-based strategy. It can be easily observed that the position-based strategy helped removing the artifacts that was causing from the pairing of middle slices with very up/down slices.

Comparison between different strategies: On Figure 17 there are 3 examples showing the comparison of the 3 strategies that was experimented. The translation was experimented from HCP T1W modality to BraTS T2W modality. Figure 16 demonstrates the pipeline.

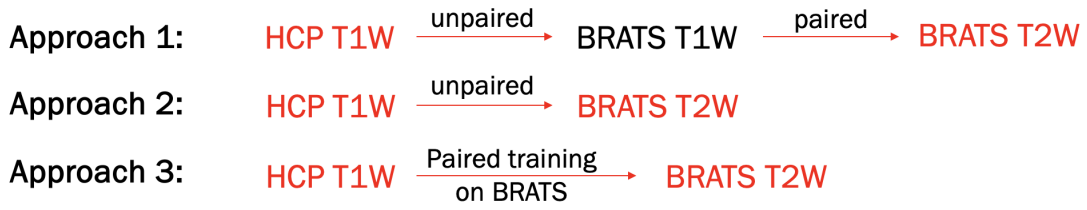


Figure 16: Pipeline in this experiment.

It can be seen by observation that, Strategy 2 was performing below of Strategies 1 and 3. While Strategy 1 and 3 was performing similar to each other.

Also, we trained models and applied the pipeline for other modalities that we wanted to achieve as well. On Figures 18, 19 and 20 there are different results for different modalities.

Also, we have calculated the FID (Frechet inception distance) scores for each strategy, for being able to make a quantitative comparison between the strategies, shown in Table 1

	Strategy 1	**Strategy 2**	**Strategy 3**
HCP T1W to BraTS T2W	102.47	144.18	73.63
HCP T1W to BraTS T1ce	102.61	116.24	90.63
HCP T1W to BraTS FLAIR	109.07	115.09	102.18

Table 1: FID scores for 3 different strategies with 3 different modality translations.

It was seen that according to the FID scores, the strategy 3 was the best in all experiments, and the strategy 2 was the worst.

After that point, the strategy 2 was discarded.

3D Consistency Problem: It was also investigated the 3D performances of the models. As mentioned earlier. The translations made on axial orientation was concatenated and the 3D images was obtained again.

It was observed that, even though the translation on a single axis (Axial) was performing well enough, when we have checked on other planes, the results were showing artifacts, this can be seen in Figure 21. Thus, it was decided to apply the progressive 2D translation method.

After the the second iteration in 2D progressive training, on sagittal slices. The results on Figures 22 and 23 was obtained.

Here, it was observed that even one iteration of the progressive training helped to increase the performance of translation and it reduced the artifacts.

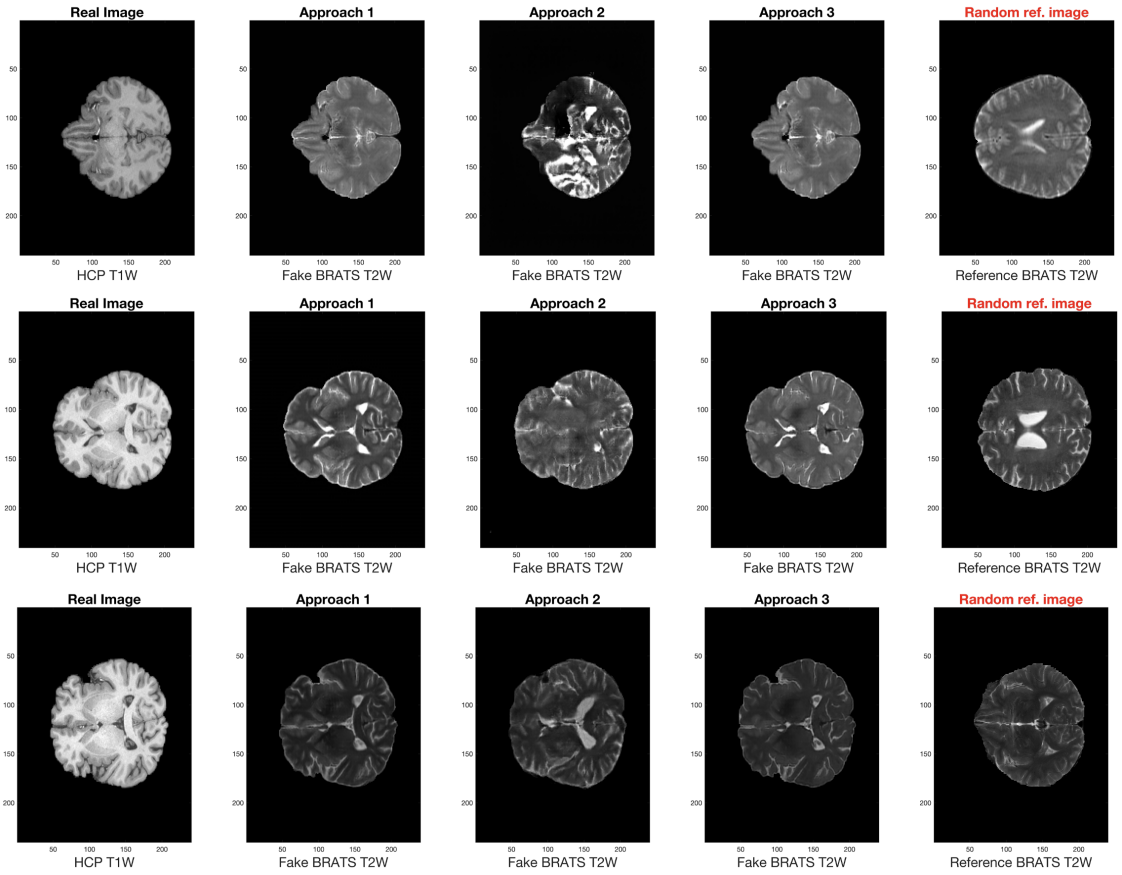
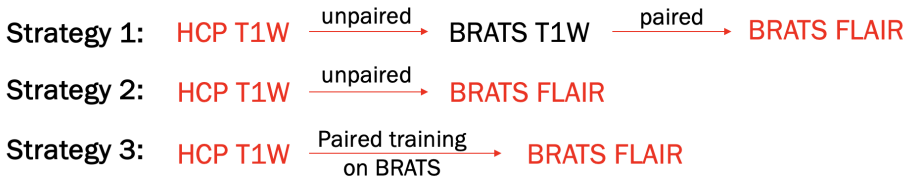


Figure 17: Three different example slices for the inspection of the performance of the strategies.



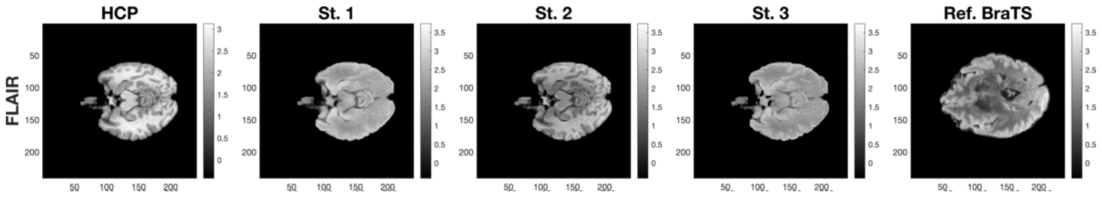


Figure 18: Example of HCP T1W to BraTS FLAIR

This was the end of my internship; I left before the implementation of the last part of the progressive training and the evaluation with the tissue intensity distribution.

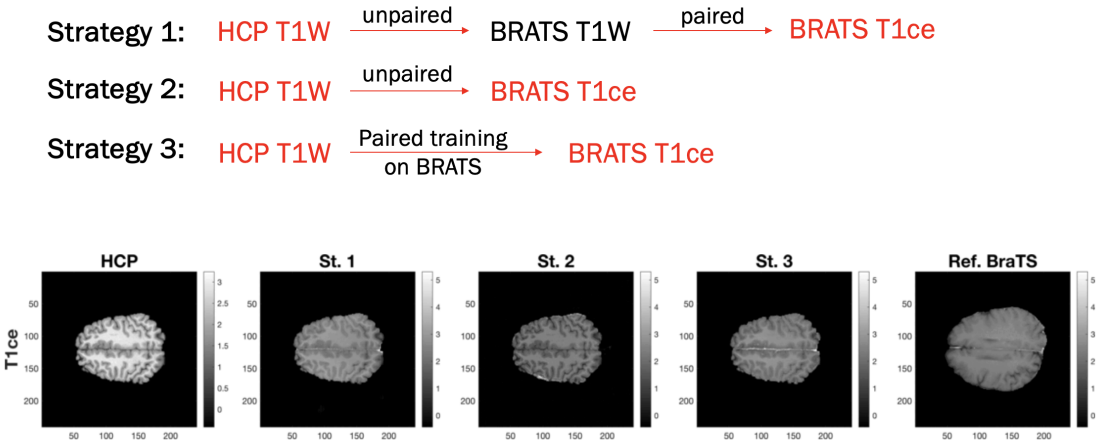


Figure 19: Example of HCP T1W to BraTS T1ce

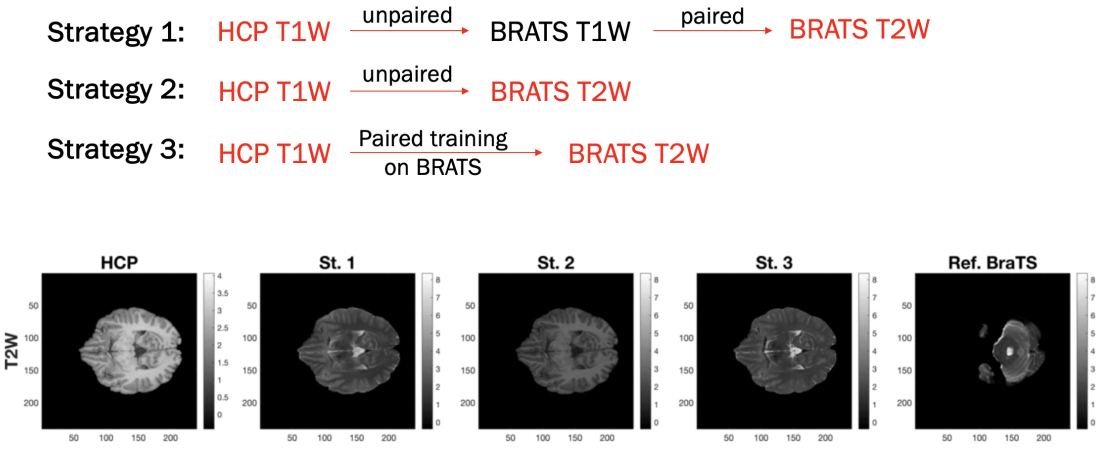


Figure 20: Example of HCP T1W to BraTS T2W

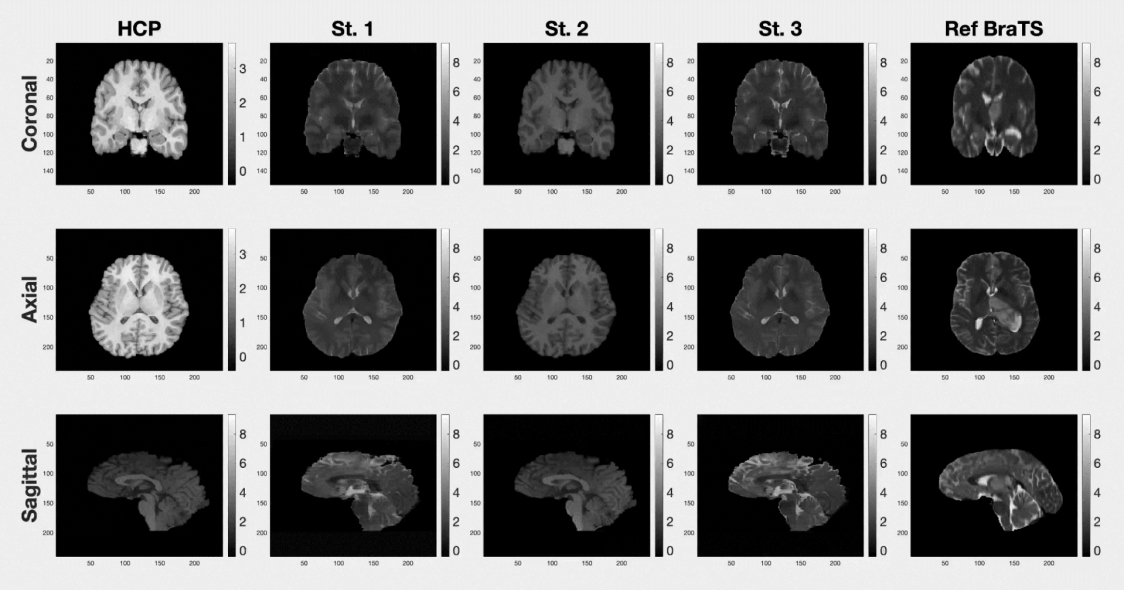


Figure 21: An example of the 3D views of translated images with 3 different strategies.

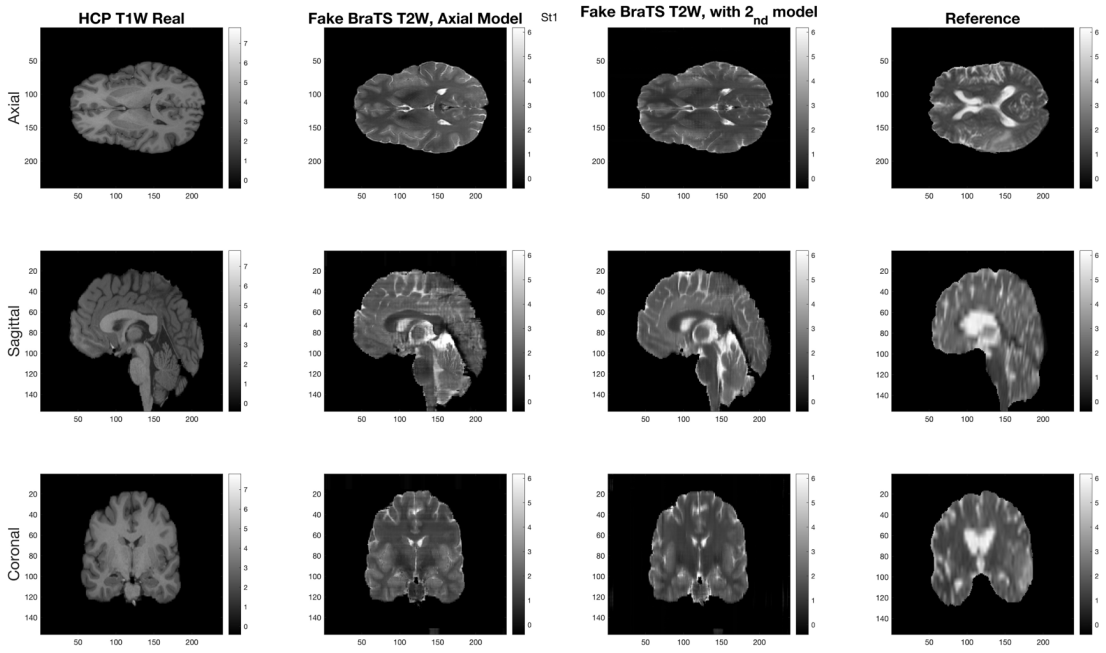


Figure 22: The results for the second step of the 2D progressive training with Strategy 1, after the axial slices was translated, a new model with sagittal slices was trained and it was translated with the second model also.

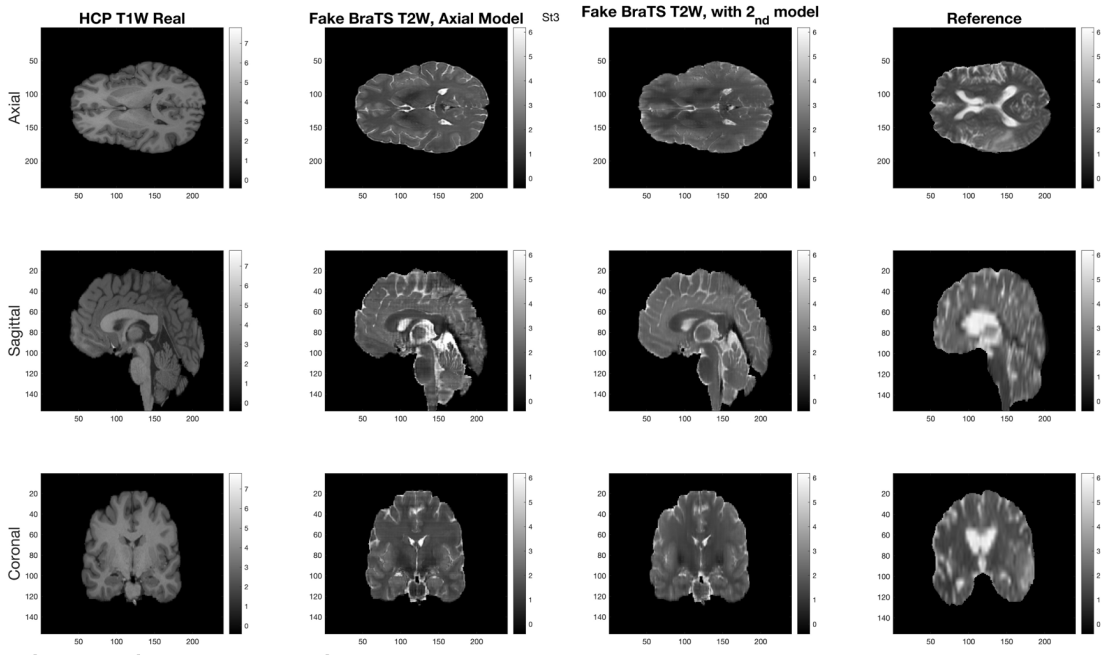


Figure 23: The results for the second step of the 2D progressive training with Strategy 3, after the axial slices was translated, a new model with sagittal slices was trained and it was translated with the second model also. It is the same patient as in Figure 22

6 Conclusion

This internship has been an exciting journey into advanced medical image analysis. We aimed to bridge the gap between MRI datasets by employing novel techniques, with a particular emphasis on image translation. Despite the difficulty of unpaired data, the use of CycleGAN and three distinct strategies resulted in notable success.

Our dedication to 3D consistency, as well as the introduction of progressive 2D translation, indicates well for future refinement. The Modality Expansion Task broadened the scope of our research, providing richer contrasts for medical image analysis.

The importance of deep learning frameworks, data quality, and preprocessing in image translation are among the key takeaways. Collaboration and knowledge sharing among colleagues were critical to our success.

In conclusion, this internship has expanded my understanding of the potential of medical image analysis as well as the iterative, collaborative nature of research.

7 Acknowledgements

Here, I want to thank Prof. Zhi-Pei Liang for making this internship possible for me. Also, I would like to thank Yudu Li, Ruihao Liu, and Wenli Li for their guidance and supervision throughout the project. Finally, I would like to thank my dear companion Berkay Şekeroğlu, throughout the project we have discussed a lot of things and I believe it was priceless.

References

Jingyu Hu, Xiaojing Gu, and Xingsheng Gu. Dual-pathway densenets with fully lateral connections for multimodal brain tumor segmentation. *International Journal of Imaging Systems and Technology*, 31(1):364–378, 2021. doi:https://doi.org/10.1002/ima.22472. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22472`.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Fabian Isensee, Paul Jaeger, Simon Kohl, Jens Petersen, and Klaus Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:1–9, 02 2021. doi:10.1038/s41592-020-01008-z.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL `https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf`.

Stephan Osterburg. Acgan architectural design, 2019. URL `https://stephan-osterburg.gitbook.io/coding/coding/ml-dl/tensorfow/chapter-4-conditional-generative-adversarial-network/acgan-architectural-design`. Accessed: 2023-10-10.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017. doi:10.1109/CVPR.2017.632.

Yuhao Kang, Song Gao, and Robert Roth. Transferring multiscale map styles using generative adversarial networks, 05 2019.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017. doi:10.1109/ICCV.2017.244.

Mahmut Yurt, Muzaffer Özbey, Salman U.H. Dar, Berk Tinaz, Kader K. Oguz, and Tolga Çukur. Progressively volumetrized deep generative models for data-efficient contextual learning of mr image recovery. *Medical Image Analysis*, 78:102429, 2022. ISSN 1361-8415. doi:https://doi.org/10.1016/j.media.2022.102429. URL `https://www.sciencedirect.com/science/article/pii/S1361841522000809`.

Heran Yang, Jian Sun, Aaron Carass, Can Zhao, Junghoon Lee, Jerry L. Prince, and Zongben Xu. Unsupervised mr-to-ct synthesis using structure-constrained cyclegan. *IEEE Transactions on Medical Imaging*, 39(12):4249–4261, 2020. doi:10.1109/TMI.2020.3015379.