MedAttentionGAN: A Contrastive Cycle-Consistent Bidirectional GAN for Paired-Unpaired Brain MR-CT Translation

Abdel Rahman Alsabbagh Bushra Alziara Masa Abdallah Rosol Sharairh
Tariq Mansour Tamam Alsarhan
Department of Artificial Intelligence, King Abdullah II School of Information Techonology, The University of Jordan, Amman, Jordan
{abd0200811,bsr0189040,mas0205277,rsl0203085,tar0205865,t_alsarhan}@ju.edu.jo

Abstract

*The acquisition of medical images, particularly through modalities like Magnetic Resonance (MR) and Computed Tomography (CT), is essential for effective disease diagnosis and management. However, limitations such as cost and radiation dose can hinder access to specific imaging modalities. To overcome these constraints and ensure comprehensive diagnostic capabilities, medical image synthesis is needed. Current methods either rely solely on unpaired images or tackle the paired images without mitigating complexity. We propose a contrastive cycle-consistent bidirectional **G**enerative **A**dversarial **N**etwork (**GAN**) for paired-unpaired **med**ical translation, in particular for brain MR-CT and using the **attention** mechanism, called **MedAttentionGAN**. Using a two-phase approach, we leverage both the paired and the unpaired data using CLIP in the first phase and AttentionGAN-V1 in the second phase, respectively. While the overall results were unsatisfactory, the model exhibited the highest visual information fidelity suggesting that the generated images have retained a significant amount of detail and clarity. We also conclude that regardless of the CLIP initialization, the model performs the same, indicating the need for sufficient amounts of data to learn better mappings. Code and data are available at `https://github.com/SabbaghCodes/MedAttentionGAN`.*

1. Introduction

In the healthcare system, medical imaging is a critical component for precise diagnosis and treatment planning. Brain Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) stand as pivotal imaging modalities in medical practice. MRI utilizes radio waves and strong magnets to generate high-resolution images of internal body structures without using ionizing radiation, making it particularly suitable for soft tissue imaging, such as the brain. Conversely, CT imaging, also known as computed axial tomography (CAT) scanning, utilizes X-rays to produce finely detailed images of tissues, organs, and bones, facilitating anatomical visualization [1].

The reliance on multiple scans poses challenges, notably the potential for radiation exposure during CT scans, which can impact patient safety. Although CT scans have become commonplace, the radiation exposure sustained by a patient can be dramatic, impacting patient safety. [2] Moreover, MRI techniques involve patient exposure to static and time-varying magnetic fields, radio frequency electromagnetic fields, and acoustic noise, which may pose health hazards or increased risk. [3] This reliance also places significant financial burdens on individuals and healthcare organizations. Overcoming these challenges necessitates innovative approaches that reduce risks, optimize resource allocation, and enhance data efficiency.

When it comes to data, a paired set of CT and MR is ideal but very rare, making it unreliable to build a model on, while unpaired data is much more available but underutilized and not as efficient as the paired ones when mapping between different domains.

Lately, Generative Adversarial Networks (GANs) have created a huge leap in Artificial Intelligence (AI) [4]. These networks have been able to create realistic-looking images similar to what you see in real life, this is done by using two neural networks, one is called the Generator while the other is called a discriminator, they work in parallel, The generator is trained to produce images that are indistinguishable from real images, while the discriminator is trained to correctly classify the images as real or fake, this process keeps on repeating until we get decent results, one subset of GAN is called the CycleGAN [5], which focuses on Image to Image translation without the usage of a paired dataset. The power of CycleGANs lies in their ability to learn transformations without one-to-one mapping, enabling image translation between unpaired datasets.

On the other hand, contrastive learning plays a pivotal role in the effective acquisition of knowledge from paired data, offering a robust framework for understanding intricate relationships within datasets. By emphasizing the differences and similarities between paired instances, contrastive learning enables models to discern subtle nuances and capture meaningful semantics. In the context of learning from MR-CT paired data, the contrastive learning approach facilitates the extraction of essential features and mappings, contributing to the enhancement of model performance. Particularly, CLIP [6] has shown promising results in medical image analysis due to its generalizability and interpretability.

This study aims to form a solution by developing a contrastive cycle-consistent bidirectional **G**enerative **A**dversarial **N**etwork (**GAN**) for paired-unpaired **med**ical translation, in particular for brain MR-CT and using the **attention** mechanism, namely **MedAttentionGAN**). Employing a dual-phase strategy, we utilize CLIP for the paired data in the initial phase and AttentionGAN-V1 for the unpaired data in the subsequent phase. More specifically, in the first phase we train two image encoders representing the generators in the subsequent cycle-consistent GAN. Despite the scarcity of paired data, they play a pivotal role in learning initial weights for the cycle-consistent GAN. For leveraging unpaired data, we employ AttentionGAN-V1 in the cycle-consistent GAN due to its proficiency in accurately dissecting and learning attention areas within input images.

The paper starts by deliberating the related works in section 2, after that, we propose the datasets with their description and links, and how they were preprocessed in section 3, then we explain the methodology and experimental setup in section 4, and to sum we show the results in section 5. We also discuss the limitations in section 6, and finally, we present the conclusion in section 7.

2. Related Works

Contrastive learning has been widely used in various computer vision tasks, including image-to-image translation [7] [8], medical image understanding [9], and image representation. In particular, contrastive learning has been employed to measure the perceptual similarity between images in line with human judgment. Existing work commonly relies on fine-tuning weights transferred from ImageNet pretraining, which is sub-optimal due to drastically different image characteristics [9]. Meanwhile, several recent studies show exciting results from unsupervised contrastive learning from natural images [7]. CLIP (Contrastive Language-Image Pre-Training) is a recent multimodal model that combines knowledge of English-language concepts with semantic image understanding [6]. CLIP has been shown to be more robust at feature extraction and image representation than any of its predecessors. In our work, we plan to use the method in CLIP to map CT to MR

GANs were first introduced by Goodfellow et al. in 2014 [4]. GANs consist of two neural networks, a generator and a discriminator, that are trained simultaneously. The generator generates fake images, while the discriminator tries to distinguish between the fake and real images. The generator is trained to produce images that are indistinguishable from real images, while the discriminator is trained to correctly classify the images as real or fake.

In the medical field, GANs have been used for various purposes, including brain segmentation, stroke imaging, neuropsychiatric disorders, breast cancer, chest imaging, imaging in oncology, and medical ultrasound, among others. For example, Ghassemi et al. proposed a DCGAN to produce MR images of the brain [10]. The authors trained the GAN so the discriminator can detect fake MR images and extract their main features. Then, the fully connected layer of the discriminator was used to extract features from the real MR images, which were then used to train a classifier. The results showed that the generated images were of high quality and the classifier achieved high accuracy.

CycleGANs were introduced by Zhu et al. in 2017 [5]. CycleGANs are a type of GAN that can be used for image-to-image translation without the need for paired training data. CycleGANs consist of two generators and two discriminators. The generators are trained to translate images from one domain to another, while the discriminators are trained to distinguish between the translated images and the real images. The cycle consistency loss is used to ensure that the translated images can be translated back to the original domain without loss of information.

Many variations of CycleGANs have been developed over the years. UNIT [11] was introduced in 2017 and is an unimodal-to-unimodal image translation model that can be used for tasks such as style transfer, object transfiguration, and domain adaptation. It uses a shared-latent space to learn the mapping between input and output images. After that MUNIT [12] was introduced in 2018 and is a multimodal-to-multimodal image translation model. DRIT [13] was also introduced in 2018 that uses a domain-specific encoder and a shared decoder to learn the mapping between input and output images. Following that, U-GAT-IT [14] was introduced in 2019, it uses a self-attention mechanism and a cycle-consistent loss function to learn the mapping between input and output images. In the same year, BiGAN [15], CoGAN [16], and SimGAN [17] were introduced. BiGAN uses an encoder-decoder architecture and an adversarial loss function to learn the mapping between input and output images, CoGAN uses a shared-latent space and a joint training procedure to learn the mapping between input and output images, and SimGAN uses a similarity loss function and a joint training procedure to learn the mapping between input and output images.

The latest development in CycleGANs was the AttentionGAN [18], which is special due to its attention-guided approach, allowing the generator to focus on the most discriminative foreground features while ignoring unwanted background details. This results in significantly better image translation results compared to the aforementioned models. We believe that AttentionGAN is a powerful model for capturing the intrinsic complex features that medical images hold.

Lately, medical CycleGANs have become used. Examples like CT-to-MR synthesis [19] and CBCT-to-CT translation [20] have shown that the synthetic CT images were of high quality and could be used for radiotherapy planning.

Specifically, bidirectional MR-CT Translation GANs are a type of CycleGAN that can be used for translating MR images to CT images and vice versa. These models can be trained with both paired and unpaired images and can be used for bidirectional MR-CT image synthesis. For example, Wang et al. proposed a bidirectional learning model, denoted as dual contrast cycleGAN (DC-cycleGAN), to synthesize medical images from unpaired brain MR-CT synthesis [21]. The authors used a U-Net architecture for the generator and a PatchGAN [22] architecture for the discriminator, leading to great results. While this model was outstanding, it could not handle paired data that hold important semantics. Given that, Abu-Srhan et. al. proposed a paired–unpaired Unsupervised Attention Guided Generative Adversarial Network (uagGAN) model [23] to handle both paired and unpaired data using a transfer learning technique. To the best of our knowledge, no one has leveraged the contrastive learning scheme to learn from the paired data, and that's what we are aiming to do here.

3. Datesets

3.1. Dataset Collection

Our research utilizes several brain MRI and CT datasets taken from different sources that are classified into paired and unpaired CT/MRI images.

3.1.1 Paired CT-MR Datasets

We only leveraged the (MR-CT) Jordan University Hospital (JUH) dataset [23] which consists of 2D image slices extracted using the RadiAnt DICOM viewer software. The extracted images are transformed to jpg image data format with a resolution of 256x256 pixels. There are a total of 179 2D axial image slices referring to 20 patient volumes.

3.1.2 Unpaired MR/CT Datasets

We combined four datasets, starting with brain CT images with intracranial hemorrhage masks from Kaggle (https://www.kaggle.com/datasets/vbookshelf/computed-tomography-ct-images/data) that consist of head CT images in jpg format. There are 2500 brain window images and 2500 bone window images, for 82 patients. There are approximately 30 image slices per patient.

The second dataset we used is the CT of the brain dataset from Kaggle (https://www.kaggle.com/datasets/trainingdatapro/computed-tomography-ct-of-the-brain) which consists of 256 CT brain scans with cancer, tumors, and aneurysm. Each scan represents a detailed image of a patient's brain taken using CT. The data are presented in both .jpg and .dcm formats.

The third is the brain tumor MRI dataset from Kaggle (https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data) that contains 3064 T1-weighted contrast-enhanced images from 233 patients with three kinds of brain tumor: 1. Meningioma (708 slices), 2. Glioma (1426 slices), and 3. A pituitary tumor (930 slices).

The fourth dataset is a rich image dataset provided by the Radiological Society of North America (RSNA) for Intracranial Hemorrhage Detection, we used 2000 CT images from it and dowmloaded then from Kaggle (https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/overview).

Finally, the fifth dataset is the brain tumor MRI dataset https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection that contains two folders: `yes` and `no` containing 3060 brain MRI images, where the `yes` file contains 1500 brain images that are tumorous and The `no` file contains 1500 images that are benign.

3.2. Data Preprocessing

We converted the images, from all datasets, from different extensions to png, to unify the format among images. Moreover, we reshaped all images to 256×256 to allow for an easier pipeline for data processing.

4. Methodology & Experimental Setup

To bridge the gap in the issues mentioned above, we leverage the idea of CLIP [6] to learn the important semantics and mappings between the MR and CT modalities using the paired datasets. We construct a contrastive learning scheme similar to the one in CLIP but using two image encoders representing the generators in the cycle-consistent GAN discussed in the following text. Although paired data exist in small amounts, they play a pivotal role in learning initial weights when then used in the cycle-consistent GAN.

To utilize the unpaired data, we use AttentionGAN-V1 as our choice for the cycle-consistent GAN, given its ca-

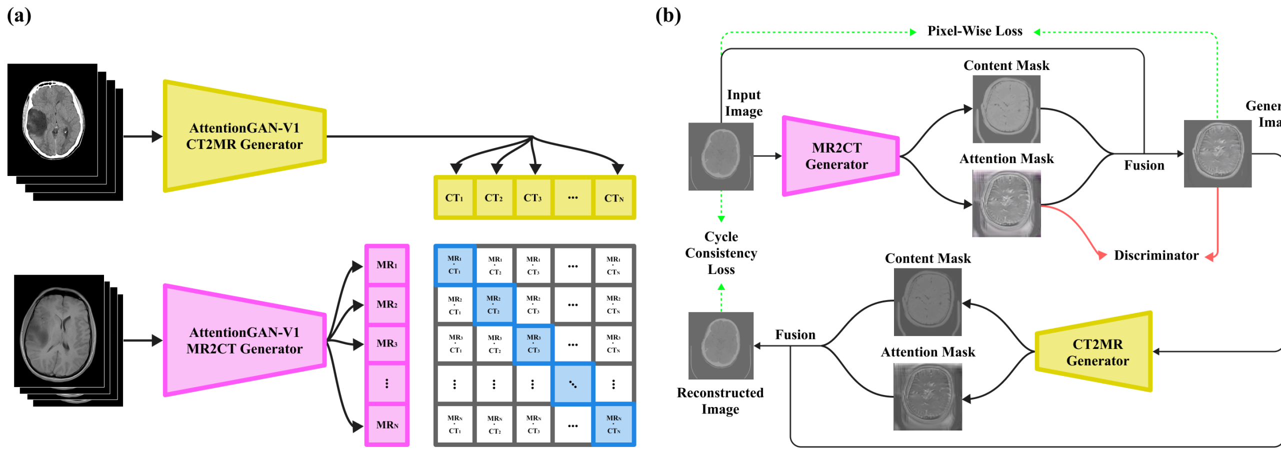

Figure 1. Full framework of MedAttentionGAN with its two phases. In phase (a) mappings between paired MR-CT images are learned using a CLIP-alike model. After that. in phase (b) the learned weights from the first phase are initialized in the AttentionGAN-V1 generators.

pabilities in dissecting and learning accurately the areas of *attention* within the input image. This is crucial in the medical images as they have highly complex details embedded within them.

This creates a full two-phase framework, wherein in the first phase we do a contrastive pertaining using a CLIP-alike model to make paired MR-CT images, and in the second phase, we initialize AttentionGAN-V1 with the weights learned from the first phase. We assume that the first phase will allow the GAN to have a more subtle and faster training process, hence mitigating the modal collapse problem that happens when the generator becomes adept at deceiving the discriminator by producing instances exclusively from a particular class, ignoring the overall variety present in the training dataset. The full framework is depicted in Figure 1.

For the first phase, we train the CLIP-alike model with a batch size of 32 for 5 epochs. We use cosine embedding loss as our objective function and Adam [24] as an optimizer with a learning rate of 30^{-5}, β_1 set to 0.9, and β_2 set to 0.999. The encoders are built on ResNet [25] and we modified the existing convolutions with depthwise-separable convolutions from [26]. We illustrate the encoder/generator in Figure 2. After that, the learned weights are saved to be used in the next phase.

As for the second phase, we initialize the weights of the AttentionGAN-V1 with the ones learned from the previous step. We train the AttentionGAN-V1 with a batch size of 1 for 10 epochs. We used the Mean Squared Error (MSE) as our GAN's objective function and the L1 loss for both the cycle and identity objective functions. Adam was used as an optimizer for both the generators and discriminators with a learning rate of 10^{-4}, β_1 set to 0.5, and β_2 set to 0.999. A learning rate linear decay is set to start at the 8th epoch. We set the lambda weight between GAN loss and reconstruction loss to be 0.01.

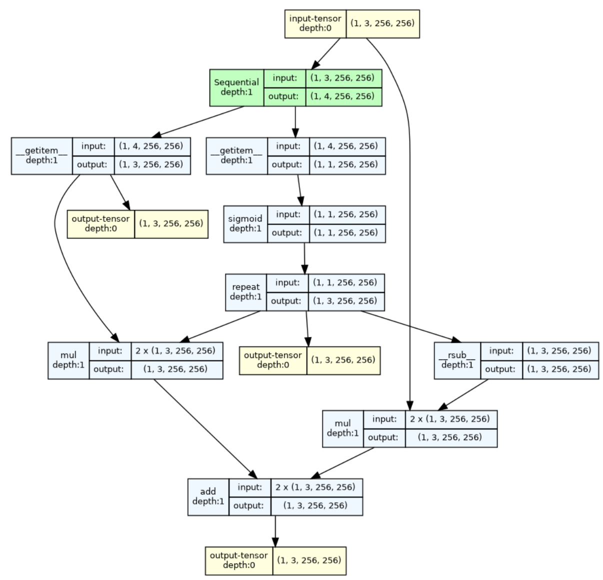

Figure 2. The network of the generator/encoder used in both of the training phases. `Sequential` represents a sequence of ResNet blocks with depthwise-separable convolutions.

The overall model has approximately 28M parameters. We trained the model with the PyTorch library [27] using a single-node NVIDIA V100 GPU.

5. Results & Discussion

To evaluate the performance of the models, we use the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Universal Quality Index (UQI), and

Figure 3. Generator and discriminator losses during the training period.

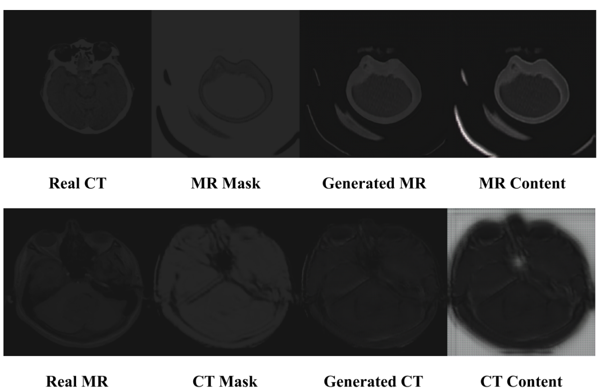

Figure 4. An example of a real image with its mask, result after generation, and content images.

Visual Information Fidelity (VIF) to provide a quantitative measure of the translations' accuracy.

Results after training the model on the test dataset are seen in Tables 1 and 2 for MR-to-CT translation and CT-to-MR translation respectively. When it comes to MR-to-CT translation, the UagGAN boasts the highest performance on all metrics. The same happens in CT-to-MR translation except for two metrics, the UQI and VIF. While the DiscoGAN [28] has the highest UQI, our model has the highest VIF which indicates that the generated images have a high level of fidelity, meaning that they closely resemble the original visual content.

On another note, regardless of the choice of the initialized weights (see Figure 3), the model performed the same. This observation suggests that the inclusion of CLIP weights did not significantly impact the overall loss in the dual-phase strategy employed by MedAttentionGAN. Moreover, the overall results as seen in Tables 1 and 2 suggest that using CLIP has subtle positive effect on the model performance.

The achieved results were deemed unsatisfactory, and we primarily attribute that to the data limitations. We believe that the model tended to overfit due to the constrained dataset, emphasizing the need for a robust and expansive dataset in the intricate domain of medical images. The insufficiency of data adversely affected the model's ability to generalize, necessitating further exploration and expansion of the dataset for improved performance. This can be seen in all models and not only ours.

Additionally, more hyperparameter searching and fine-tuning could've led to better results. We show an example of the outputs of the model in Figure 4.

6. Limitations

The limited amount of the MR/CT datasets that were available to us represented a significant project limitation. We only discovered one paired MR/CT dataset online and faced significant difficulties finding more data, which thus reduced the diversity of training datasets for our model Regarding the unpaired datasets, which are crucial to our GAN model, their small size required many site searches in order to obtain a sufficient amount of data. The lack of paired and unpaired data made it difficult to create a robust and generalizable model. We used data augmentation and contrastive learning hoping to overcome these limitations, which expanded the scope of our dataset and enhanced model performance, but still was not enough as seen in the results above.

Although the method was novel, using CLIP to translate between MR and CT images using paired data came with certain difficulties. It was necessary to modify CLIP, which was primarily created for picture and text correlation, for the purpose of translating between various kinds of medical images. To make sure this adaptation worked well for our particular use case, it needed to be significantly optimized and fine-tuned.

Additionally, limited computational resources, in particular RAM capacity and GPUs, limited the project. This constraint impacted our model's scalability and efficiency, especially when handling big medical imaging datasets — which are necessary for precise analysis.

Moreover, some medical specialists expressed doubts to us about the viability and efficacy of the methods proposed for bidirectionally translating between different medical imaging modalities. This skepticism could make it

Table 1. Image quality evaluation metrics on various cycle-consistent GAN models on **MR to CT** translation. Results are represented as the mean ±standard deviation. We brought the results from translation from [23] for models we did not run.

Model	PSNR	SSIM	UQI	VIF
CycleGAN [5]	25.404 ±1.648	0.554 ±0.052	0.779 ±0.021	0.361 ±0.070
DualGAN [29]	21.880 ±2.607	0.541 ±0.053	0.688 ±0.038	0.234 ±0.103
DiscoGAN [28]	24.675 ±2.136	0.553 ±0.042	0.743 ±0.034	0.067 ±0.017
ComboGAN [30]	21.219 ±2.143	0.453 ±0.056	0.676 ±0.074	0.221 ±0.058
UNIT [11]	25.572 ±1.821	0.573 ±0.040	0.718 ±0.046	0.370 ±0.0691
UagGAN [23]	**27.599 ±1.769**	**0.595 ±0.043**	**0.781 ±0.042**	**0.387 ±0.073**
MedAttentionGAN W. CLIP (Ours)	15.176 ±1.0821	0.369 ±0.120	0.002 ±0.003	0.015 ±0.009
MedAttentionGAN W.O. CLIP (Ours)	15.761 ±1.053	0.377 ±0.106	-0.014 ±0.318	0.013 ±0.008

Table 2. Image quality evaluation metrics on various cycle-consistent GAN models on **CT to MR** translation. Results are represented as the mean ±standard deviation.We brought the results from translation from [23] for models we did not run.

Model	PSNR	SSIM	UQI	VIF
CycleGAN [5]	30.529± 2.318	0.529 ±0.058	0.607 ± 0.083	0.049 ±0.012
DualGAN [29]	27.292 ± 1.715	0.360 ±0.044	0.360 ±0.046	0.121 ±0.052
DiscoGAN [28]	28.316 ±3.494	0.393 ± 0.083	**0.612 ±0.074**	0.040 ±0.032
ComboGAN [30]	30.320 ±2.301	0.382 ± 0.060	0.462 ±0.072	0.042 ±0.039
UNIT [11]	30.701 ±1.435	0.539 ± 0.021	0.602 ±0.097	0.072 ±0.033
UagGAN [23]	**31.049 ± 1.306**	**0.542 ±0.051**	0.543±0.084	0.178± 0.029
MedAttentionGAN W. CLIP (Ours)	14.220 ±2.731	0.350 ±0.118	-0.004 ±0.020	0.261 ±0.165
MedAttentionGAN W.O. CLIP (Ours)	12.035 ±1.835	0.321 ±0.099	0.001 ±0.035	**0.334 ±0.162**

more difficult for our findings to be accepted and used in actual medical situations, which would call for additional validation and interaction with the medical community.

Furthermore, noise and artifacts are common in medical imaging, which may have an effect on the quality of the images and model outputs.

Also, privacy concerns are raised by the way sensitive patient information is handled in medical imaging. We follow ethical guidelines and ensure compliance with data protection regulations.

7. Conclusion

In conclusion, our proposed MedAttentionGAN, a contrastive cycle-consistent bidirectional Generative Adversarial Network (GAN) tailored for paired-unpaired medical translation, particularly for brain MR-CT, incorporates the attention mechanism to enhance performance. Employing a two-phase strategy involving CLIP and AttentionGAN-V1 for paired and unpaired data, respectively, we observed that while the overall results fell short of expectations, the model demonstrated the highest visual information fidelity. This suggests that the generated images retained substantial detail and clarity. Furthermore, our findings indicate that the model's performance remains consistent regardless of CLIP initialization, underscoring the importance of ample data for optimal learning of mappings. The assessment demonstrated that the UagGAN exhibited strong performance in translating MR-to-CT images, while the DiscoGAN showed excellence in CT-to-MR translation.

For future directions, we plan to use a larger dataset to better enhance the generalizability of the model. We also would like to discover different model architectures that are less complex. Finally, we intend to use different contrastive learning schemes other than CLIP.

We hope that our findings will help the research community to find ways to tackle this problem and develop upon it.

References

[1] Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017, Proceedings 2*, pages 14–23. Springer, 2017. 1

[2] B. Giordano, J. Grauer, Christopher P. Miller, T. Morgan, and G. Rechtine. Radiation exposure issues in orthopaedics. *The Journal of bone and joint surgery. American volume*, 93 12:e69(1–10), 2011. 1

[3] Medical magnetic resonance (mr) procedures: Protection of patients. *Health Physics*, 87:197–216, 2004. 1

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. 1, 2

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020. 1, 2, 6

[6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 2, 3

[7] Yao Gou, Min Li, Yu Song, Yujie He, and Litao Wang. Multi-feature contrastive learning for unpaired image-to-image translation. *Complex & Intelligent Systems*, 9(4):4111–4122, August 2023. 2

[8] Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. Modulated contrast for versatile image synthesis, 2022. 2

[9] Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D. Manning, and Curtis P. Langlotz. Contrastive learning of medical visual representations from paired images and text, 2022. 2

[10] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez, R. Romero, A. Seara Vieira, and L. Borrajo. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. *Neural Computing & Applications*, 35(3):2291–2323, 2023. 2

[11] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks, 2018. 2, 6

[12] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation, 2018. 2

[13] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *European Conference on Computer Vision*, 2018. 2

[14] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation, 2020. 2

[15] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning, 2017. 2

[16] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks, 2016. 2

[17] Yifeng Jiang, Tingnan Zhang, Daniel Ho, Yunfei Bai, C. Karen Liu, Sergey Levine, and Jie Tan. Simgan: Hybrid simulator identification for domain adaptation via adversarial reinforcement learning, 2021. 2

[18] Hao Tang, Dan Xu, Nicu Sebe, and Yan Yan. Attention-guided generative adversarial networks for unsupervised image-to-image translation. In *International Joint Conference on Neural Networks (IJCNN)*, 2019. 3

[19] Sampada Tavse, Vijayakumar Varadarajan, Mrinal Bachute, Shilpa Gite, and Ketan Kotecha. A systematic literature review on applications of gan-synthesized images for brain mri. *Future Internet*, 14(12), 2022. 3

[20] Chitchaya Suwanraksa, Jidapa Bridhikitti, Thiansin Liamsuwan, and Sitthichok Chaichulee. CBCT-to-CT Translation Using Registration-Based Generative Adversarial Networks in Patients with Head and Neck Cancer. *Cancers*, 15(7):2017, March 2023. 3

[21] Jiayuan Wang, Q. Wu, and Farhad Pourpanah. Dc-cyclegan: Bidirectional ct-to-mr synthesis from unpaired data, 11 2022. 3

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018. 3

[23] Alaa Abu-Srhan, Israa Almallahi, Mohammad A.M. Abushariah, Waleed Mahafza, and Omar S. Al-Kadi. Paired-unpaired unsupervised attention guided gan with transfer learning for bidirectional brain mr-ct synthesis. *Computers in Biology and Medicine*, 136:104763, 2021. 3, 6

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 4

[26] François Chollet. Xception: Deep learning with depthwise separable convolutions, 2017. 4

[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 4

[28] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks, 2017. 5, 6

[29] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation, 2018. 6

[30] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation, 2017. 6