Bidirectional Brain Image Translation using Transfer Learning from Generic Pre-trained Models

Fatima Haimour[1], Rizik Al-Sayyed[2], Waleed Mahafza[3] and Omar S. Al-Kadi[2]

[1]*Faculty of Information Technology, Zarqa University, Zarqa 13110, Jordan*

[2]*King Abdullah II School for Information Technology,*
University of Jordan, Amman 11942, Jordan

[3]*Department of Diagnostic Radiology, Jordan University Hospital, Amman 11942, Jordan*

Abstract—Brain imaging plays a crucial role in the diagnosis and treatment of various neurological disorders, providing valuable insights into the structure and function of the brain. Techniques such as magnetic resonance imaging (MRI) and computed tomography (CT) enable non-invasive visualization of the brain, aiding in the understanding of brain anatomy, abnormalities, and functional connectivity. However, cost and radiation dose may limit the acquisition of specific image modalities, so medical image synthesis can be used to generate required medical images without actual addition. CycleGAN and other GANs are valuable tools for generating synthetic images across various fields. In the medical domain, where obtaining labeled medical images is labor-intensive and expensive, addressing data scarcity is a major challenge. Recent studies propose using transfer learning to overcome this issue. This involves adapting pre-trained CycleGAN models, initially trained on non-medical data, to generate realistic medical images. In this work, transfer learning was applied to the task of MR-CT image translation and vice versa using 18 pre-trained non-medical models, and the models were fine-tuned to have the best result. The models' performance was evaluated using four widely used image quality metrics: Peak-signal-to-noise-ratio, Structural Similarity Index, Universal Quality Index, and Visual Information Fidelity. Quantitative evaluation and qualitative perceptual analysis by radiologists demonstrate the potential of transfer learning in medical imaging and the effectiveness of the generic pre-trained model. The results provide compelling evidence of the model's exceptional performance, which can be attributed to the high quality and similarity of the training images to actual human brain images. These results underscore the significance of carefully selecting appropriate and representative training images to optimize performance in brain image analysis tasks.

Keywords—Image Translation, Transfer learning, Pre-trained model, Brain tumor, Magnetic resonance imaging, Computed Tomography, CycleGAN

I. INTRODUCTION

Brain images play a critical role in the understanding of brain anatomy and assessment of its function and abnormalities. Techniques such as magnetic resonance imaging (MRI) and computed tomography (CT) provide valuable diagnostic information for various neurological conditions, aiding in treatment planning and monitoring patient's progress [44]. Each modality has its own unique strengths and limitations [3]. The choice of modality depends on the clinical problem, patient's condition, practitioner's medical expertise, cost, radiation dose, patient age, and the limited availability of images [5, 2].

To address these concerns, medical image synthesis can assist with the generation of necessary medical images without the need for physical scanning [32, 15]. This technique can be especially helpful when imaging is not readily available or when there are concerns about the risks associated with radiation exposure, such as in the case of pregnant women or young children. CT imaging, for example, emits radiation, while MR imaging does not [33]. However, in some cases, both modalities may be required to get a more complete picture of the patient's condition. Additionally, CT-MR image synthesis is used as a preliminary step in medical image segmentation [58]. Therefore, it is essential to develop an accurate CT-MR synthesis model to support these critical applications [39], [38]. The term "CT-MR image synthesis" is used in this paper to refer to the bidirectional translation between CT and MR images.

CycleGAN is an unsupervised machine learning technique designed for unpaired image-to-image translation tasks across different domains [2]. Unlike traditional supervised learning that relies on paired datasets, CycleGAN operates without the need for corresponding pairs of images. It achieves this through a Generative Adversarial Network (GAN) architecture consisting of a generator and a discriminator [48]. The innovation lies in its use of cycle-consistency loss, which ensures that translated images maintain content integrity when moving between source and target domains. Through adversarial training, CycleGAN learns to generate images that are indistinguishable from real ones in the target domain, enabling versatile cross-domain translations without paired data [51, 15].

CT-MR image synthesis is a challenging task due to several issues. Firstly, obtaining CT and MR images separately is often time-consuming, costly, and burdensome to the patient. As a result, most available training datasets are unpaired. Secondly, the limited number of available training datasets for CT-MR image synthesis is a significant issue, with most datasets being small. This problem is critical because the number of training images is essential for generating realistic images. To overcome this issue, pre-trained models is a commonly used technique that can improve the GAN model's performance and generate more realistic images [31]. It enables the GAN model to leverage knowledge gained from training on other datasets, which can then be applied to the CT-MR image synthesis task. It can reduce the amount of time and computational resources required to train deep learning models, which is particularly advantageous in medical imaging

applications that have time constraints and require real-time processing. Furthermore, pre-trained models have been shown to improve deep learning models' generalization capabilities, lowering the risk of overfitting and improving their ability to handle unknown data [2]. However, pre-trained models may have some limitations, such as the possibility that they may not be appropriate for the new task or that they may not have been trained on similar data, which can result in poor performance or over-fitting [6]. Additionally, they may have been trained on data that contains biases or errors, which can negatively impact the model's performance on the new task [17]. Despite these limitations, pre-trained models are a powerful technique that can improve model performance in medical imaging and other fields. It is important to assess the suitability of the pre-trained models for the new task and to use appropriate data and regularization techniques to avoid over-fitting and biases [37]. There have been several studies in recent years that have applied pre-trained models to MR-CT image translation [58]. These studies have shown promising results, improving the performance and efficiency of MR-CT image translation models.

This work aims to explore the effect of applying transfer learning from generic pre-trained models on MR-to-CT image synthesis. This would assist in investigating the factors that contribute to the success of applying a pre-trained model in medical image translation tasks. To achieve this goal, a paired MR-CT dataset was used with 18 different pre-trained models. The image translation efficacy will be evaluated using various image quality metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), universal quality index (UQI), and visual information fidelity (VIF). The paper investigates the outcome of the different pre-tained models on the translated image characteristics. The framework applied in this work relies on the general CycleGAN approach [59]. Figure 1 shows the applied architectures, including CycleGAN. The paper aims to highlight the importance of understanding generic pre-trained models when used across domains, particularly for improving MR-to-CT synthesis, which may lead to better development of trustworthy artificial intelligence tools in medical imaging. The main contributions of this work are:

1) Analyze the effect of transfer learning from generic pre-trained models on bi-directional MR-CT image translation.
2) Evaluate the performance of various generic pre-trained models comprehensively using four widely used image quality metrics which highlights different aspects of the generated image.
3) Assess the clinical significance of the translated MR-CT images through a perceptual study by two experienced radiologists.
4) Investigate the effect of image structure and discontinuities of the pre-trained model on the performance of medical image translation.

The rest of the paper is organized as follows: Section II presents a review of related work of image synthesis in the medical imaging domain. Section III provides coverage of image-to-image translation using GAN models. Section IV explains the proposed paired-unpaired unsupervised learning model with transfer learning. Section V illustrates experimental results and performance. Section VI analyses and interprets main findings. Finally, Section VII concludes and gives future directions.

II. RELATED WORK

In the field of medical imaging there has been an amount of research dedicated to achieving accurate and efficient brain tumor detection [25, 26, 23], classification and prediction of tumor grade [13, 20]. A major advancement, in this area has been the use of pre-trained models especially in the context of translating MR images to CT images. The ultimate goal is to make the process of translating images more precise. In this section we will explore the studies that have utilized pre-trained models to address important tasks such, as detecting tumors classifying them accurately and determining their grade.

A. Brain Tumor Detection

Deep transfer learning was employed by Chelghoum et al. [7] to classify three types of brain tumors with contrast-enhanced magnetic resonance images (CE-MRI) and achieved a high classification accuracy of 98.71%, demonstrating the reliability of transfer learning for small datasets. However, one potential disadvantage is the dependence on a specific dataset. The model's performance and reliability may be limited to the characteristics of that dataset, such as imaging protocols, patient demographics, and tumor variations. The high accuracy achieved on the current dataset does not guarantee similar results on different datasets. Arbane et al. [4] use transfer learning and various CNN architectures, including ResNet, Xception, and MobilNet-V2, for brain tumor classification from MRI images. The MobilNet-V2 architecture achieved the highest accuracy of 98.24% and an F1-score of 98.42%. The main drawback is the lack of in-depth understanding regarding the specific factors that contribute to MobilNet-V2's superior performance compared to other architectures. While the results are impressive, gaining insights into the underlying reasons for this performance disparity can aid in model optimization and generalization to different datasets. while Kaur et al. [27] replaced the last three layers of eight pre-trained CNNs for brain tumor classification using MRI images, achieving a classification accuracy of 96.05%. While this transfer learning approach demonstrated effectiveness in the domain, one potential disadvantage is the limited exploration of alternative modifications to the network architecture. By solely replacing the last three layers, the study may not have fully optimized the network for brain tumor classification. Important features or patterns in the earlier layers might not have been fully utilized or adapted. Also, Saxena et al. [40] The study aimed to classify brain tumor cells as cancerous or non-cancerous using pre-trained VGG16, ResNet50, and Inceptionv3 networks. The researchers employed preprocessing and augmentation techniques to improve the datasets and achieved the highest accuracy of 95% with the ResNet50 network. Considering alternative network architectures or fine-tuning the
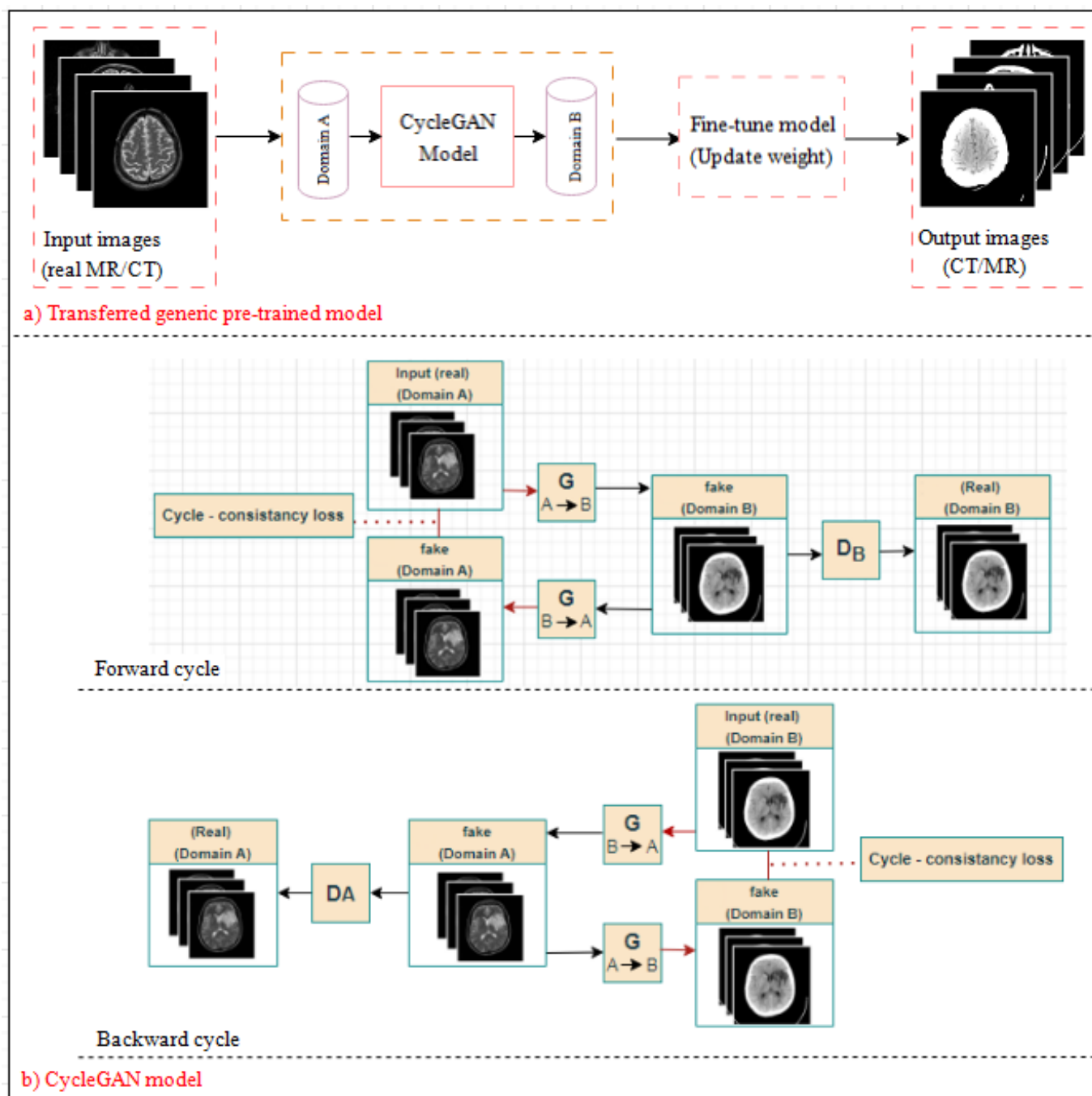
Figure 1: Transfer learning from generic pre-trained models.

existing models could enhance accuracy and adaptability to the unique characteristics of brain tumor datasets. Çinar et al. [9] employed the ResNet50 architecture for tumor detection in MRI images. They replaced the last 5 layers of ResNet50 with 8 other layers and experimented with other pre-trained deep networks. The MRI dataset was split into 60% for training and 40% for testing, achieving an impressive accuracy of 97.2%. However, it is unclear whether the high accuracy is solely attributed to the specific dataset used or if it can be replicated across different datasets. Abu-Srhan et al. [2] introduce a method called uagGAN which stands for Paired Unsupervised Attention Guided Generative Adversarial Network. This approach combines transfer learning with the goal of synthesizing brain MR and CT images, in both directions. The uagGAN model takes advantage of paired and unpaired datasets using a process of pre-training and retraining. Notably, attention masks are utilized to improve the accuracy and sharpness of the generated images. Additionally, the model incorporates transfer learning from a trained nonmedical model to enhance the learning process and improve image translation performance. The proposed methodology has shown results in evaluations as well as qualitative assessments by radiologists demonstrating its superior ability to generate accurate brain MR and CT images bi-directionally compared to existing image, to image translation techniques.

B. Brain Tumors Classification

In the context of brain tumor classification, a non-invasive MRI-based approach was introduced by Tandel et al. [46]. Their method is designed to differentiate between LGG and HGG tumors and relies on the amalgamation of five pre-trained CNNs via a majority voting mechanism, which significantly augments the classification accuracy. This approach exhibits enhanced accuracy when applied across three distinct datasets. Nevertheless, it's worth noting that the decision-making process underpinning this algorithm may present challenges in terms of comprehensibility and explication. Also, Rehman et al. [36] utilized three deep networks, AlexNet, VGG16, and GoogLeNet, with transfer learning, pre-processing, and data augmentation to classify brain tumors. The highest accuracy of 98.69% was achieved by the VGG16 network. Training and fine-tuning multiple networks like AlexNet, VGG16, and GoogLeNet can be computationally intensive and time-consuming. while Deepak et al. [11] used transfer learning with a pre-trained deep neural network, GoogLeNet, to classify brain tumors, achieving a mean accuracy of 98%. They only modified the last three layers of GoogLeNet and trained it with a SoftMax classifier. Additionally, they used GoogLeNet as a feature extractor for SVM and KNN classifiers. While this yielded high accuracy, it restricts the ability to adapt the model to dataset-specific characteristics or incorporate domain-specific knowledge. Swati et al. [45] focused on utilizing deep learning for extracting features from medical images, specifically MRIs. They employed a pre-trained VGG19 model and applied block-wise fine-tuning transfer learning to adapt the learned features from general images to medical images. The study achieved an accuracy of 94.82%. The VGG19 model

was trained on a different dataset, likely containing different image types, potentially leading to suboptimal performance in capturing specific features and patterns present in medical images, particularly MRIs. Also, Hao et al. [18] proposed a transfer learning-based active learning framework that aims to reduce annotation costs while maintaining model stability and robustness for brain tumor classification. It achieves an AUC of 82.89% on a balanced dataset of 82%. The lower AUC value indicates that the model's ability to distinguish between positive and negative cases may not be optimal. Additionally, the balanced dataset of 82% suggests a potential class imbalance issue, which could result in a bias towards the majority class and impact the model's performance on minority classes.

C. Tumor Grade Prediction

In the domain of glioma tumor grade prediction from brain MR images, Yang et al. cite19 applied deep learning models, including AlexNet and GoogLeNet. Their investigation yielded compelling results, as both models demonstrated a remarkable level of accuracy in tumor grade identification. These findings underscore the promising role of deep learning in advancing medical image analysis. However, it is imperative to recognize that the study's assessment was limited to a singular dataset, prompting considerations regarding the broader applicability of these results across diverse datasets and clinical scenarios. In the same way, Tenghongsakul et al. [47] compared different deep transfer learning methods (InceptionResNet-V2, ResNet50, MobileNet-V2, and VGG16) for brain tumor prediction using a public MRI dataset. Image enhancement technique CLAHE was applied to improve image quality. The suggested method achieved a remarkable prediction accuracy of up to 100%. while Khan et al. [28] investigate the application of CNNs for image-based brain cancer prediction to enhance treatment reliability. Using transfer learning with pre-trained VGG19 and MobileNetV2 models, the research focuses on improving accuracy. The experiment achieved 97% accuracy with MobileNetV2 and 91% with VGG19. Table I presents a summary of the studies discussed above and provides a quick reference guide for each study.

In this work, we investigate 18 different pre-trained models to translate 400 brain MRI images to CT images and vice versa. The goal of the study was to explore the ability of these models to perform image translation between different modalities and generate images that are comparable to real images. Models are chosen to be comprehensive and based on their performance in previous studies.

III. MATERIALS AND METHODS

Pre-trained models that are used in this work are compatible with the CycleGAN framework [59] and cover diverse domains such as natural scenes, animals to urban and satellite images.

A. Cycle GAN

CycleGAN is a type of Generative Adversarial Network (GAN) that can translate one image type to another. Its primary

Table I: A Review of Related Works

	Network backbone	Training type	Unsupervised	Bi-directional	Dataset	Metrics
Yang et al. [55]	AlexNet and GoogleNet	Cross-validation (5-fold CV)			Brain MR Images, 113 Glioma Patients	Accuracy, AUC
Tenghongsakul et al. [47]	InceptionResNet-V2, ResNet50, MobileNet-V2, VGG16	Not specified	✓		Brian Tumor Dataset from kaggle	Accuracy, Precesion, Recall, F1- score
Khan et al. [28]	VGG19 , MobileNetV2	Transfer learning	✓	✓	Brian Tumor Dataset from kaggle	Accuracy, F1
Tandel et al. [46]	AlexNet, VGG16, ResNet18, GoogleNet, ResNet50	Not specified		✓	Three different datasets: RSM, SSM, and WBM data	Accuracy
Rehman et al. [36]	AlexNet, GoogLeNet, VGGNet	Transfer Learning (Fine-Tune and Freeze)		✓	Brain tumor MRI slices from Figshare	Accuracy
Deepak et al. [11]	CNN with SVM	Fivefold Cross-Validation	✓		Figshare open dataset containing MRI images	Accuracy
Swati et al. [45]	Pre-trained deep CNN model	Transfer Learning (Block-Wise Fine-Tuning)	✓	✓	T1-weighted contrast-enhanced magnetic resonance images (CE-MRI) benchmark dataset	Accuracy
Hao et al. [18]	Pre-trained model as a backbone (2D slice-based approach)	Transfer Learning (Active Learning)			MRI training dataset of 203 patients and a validation dataset of 66 patients	AUC
Chelghoum et al. [7]	Nine deep pre-trained CNN architectures	Cross-validation (5-fold CV)	✓	✓	Brain CE-MRI benchmark dataset	Accuracy
Arbane et al. [4]	ResNet, Xception, MobilNet-V2	Not specified			MRI Images	Accuracy and F1-score
Kaur et al. [27]	AlexNet, Resnet50, GoogLeNet, VGG-16, Resnet101, VGG-19, Inceptionv3, InceptionResNetV2	Transfer Learning using Pre-trained DCNNs	✓		Harvard, Clinical, and Benchmark Figshare Repository	Accuracy, Sensitivity, Specificity, Precision, False Positive Rate, Error, F1-score, AUC, Mathew Correlation Coefficient.
Saxena et al. [40]	Resnet-50, VGG-16, Inception-V3	CNN-based Transfer Learning	✓	✓	Brain MRI images, which contains a total of 253 images	Accuracy
Çinar et al. [9]	Resnet50, Alexnet, Densenet201, InceptionV3, Googlenet	Fivefold Cross-Validation			MRI Images	Accuracy
Abu-Srhan et al. [2]	Paired-unpaired Unsupervised Attention Guided GAN	Pre-training with Paired Data and Retraining on Unpaired Data	✓	✓	The dataset by Han et al. [13] contained 367 paired MR-CT brain images from 18 patients.	PSNR, SSIM, UQI, VIF

objective is to learn how to map between two image domains without the use of paired training data, in an unsupervised manner [59]. CycleGAN consists of two generators, G and F, and two discriminators, Dx and Dy as shown in Figure 2 [59]. The generators are responsible for changing images from one domain to another, while the discriminators learn to differentiate between genuine images and fake ones generated by the generators. The diagram in Figure 1 (b) illustrates the architecture of CycleGAN [56].The network includes both forward and backward cycles. In the forward cycle, there are two generators and one discriminator. Generator-CT is responsible for learning the mapping from MR to CT to produce artificial CT images. Discriminator-CT, on the other hand, is trained to distinguish between real and synthetic CT images and to encourage Generator-CT to produce realistic images. Generator-MR, which learns the CT-to-MR mapping, helps to improve the ability of Generator-CT to synthesize accurate CT images. The Cycle consistency loss is used to ensure that the synthesized CT image can be converted back to the original MR image. In the backward cycle, the structure is identical to the forward cycle, except that CT images are used as input instead of MR images [56]. The loss function
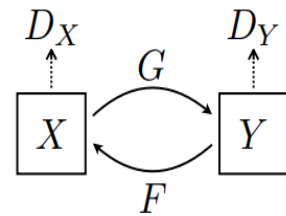

Figure 2: CycleGAN generators and discriminators.

for the CycleGAN includes two components: the adversarial loss and the cycle consistency loss. The adversarial loss is

given by [56]

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data(y)}} [\log D_Y(y)]$$
$$+ E_{x \sim p_{data(x)}} [\log (1 - D_Y G(x))] \quad (1)$$

where G tries to generate images $G(X)$ that look similar to images from domain Y, while D_Y aims to distinguish between translated samples $G(X)$ and real samples y. G aims to minimize this objective against an adversary D that tries to maximize it, i.e., $min_G max_{D_Y} L_{GAN}(G, D_Y, X, Y)$, a similar adversarial loss for the mapping function is introduced $F : Y \rightarrow X$ and its discriminator D_X as well, i.e., $min_F max_{D_X} L_{GAN}(G, D_Y, Y, X)$. The adversarial loss encourages the generator G to generate images that are indistinguishable from real images from domain Y, and the discriminator D_y to correctly distinguish between real and fake images. The cycle consistency loss is given by [59]

$$L_{cyc}(G, F) = E_{x \sim p_{data(x)}} [\|F(G(x)) - x\|_1]$$
$$+ E_{y \sim p_{data(y)}} [\|G(F(y)) - y\|_1] \quad (2)$$

For each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. Similarly, for each image y from domain Y, G and F should also satisfy backward cycle consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

The cycle consistency loss encourages the generator $F(G(x))$ to be like x and $G(F(y))$ to be like y. The objective is to learn mapping functions between the two domains X and Y using training samples with the labels $\{x_i\}_{i=1}^N$, where $x_i \in X$ and $\{y_j\}_{j=1}^N$, where $y_j \in Y$.

The data distribution is represented as $x \sim p_{data(x)}$ and $y \sim p_{data(y)}$. This model includes two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. Additionally, The two adversarial discriminators D_X and D_Y are introduced, where D_X aims to distinguish between images x and translated images $F(y)$; in the same way, D_Y aims to discriminate between y and $G(x)$ [59].

B. Transfer learning and Fine-tuning

Transfer learning is a technique in machine learning that involves using a model trained on one task to enhance the performance of a model trained on a different but related task. The primary objective of transfer learning is to make use of the knowledge and experience gained from one problem to improve the model's performance on another [50]. This involves utilizing a pre-trained model as a starting point and fine-tuning it for the new task using a smaller dataset [35]. Transfer learning has been extensively employed in various fields, including computer vision, natural language processing, and speech recognition. It is particularly useful in labelling pre-trained labelled data for the target task is scarce or when the target task is related to the source task [59].

In the medical field, transfer learning is a potent technique that can be used to translate images [8]. Medical image translation is the process of converting an image from one modality to another, such as from an MR image to a CT image or vice versa. The primary objective of medical image translation is to generate images that resemble the originals and contain the same information. Transfer learning can be utilized in medical image translation to enhance the model's performance on a smaller dataset specific to the medical domain by leveraging the knowledge obtained from a more extensive dataset of images [53]. For instance, a model trained on a vast dataset of natural images can be fine-tuned on a smaller dataset of medical images to improve its accuracy on the medical image translation task. According to [16], transfer learning is advantageous in medical image translation because it significantly reduces the amount of labelled data needed for training. This is especially useful in medical imaging, where obtaining labelled data can be challenging and expensive. Additionally, transfer learning can aid models in performing better on unseen data by improving their generalization ability. This is critical in medical imaging, where the model must function effectively on a broad range of patients and conditions [35].

In transfer learning [50], we define a domain as $D = \{X, P(x)\}$, where X is the feature space with $X = \{x_1, \cdots, x_n\} \subset X$ and $P(X)$ is a marginal probability distribution. For example, X could include all possible images derived from a particular MRI protocol, acquisition parameters, and scanner hardware, and $P(X)$ could depend on, for instance, subject groups, such as adolescents or elderly people. P(X) for MR to CT translation represents the probability distribution of intensities or features within the MR images, considered independently of their corresponding CT counterparts. The value can be obtained by the integral sums over all possible values of y, representing the marginalization of the joint distribution over the CT dimension to obtain the marginal distribution of MR. Tasks comprise a label space Y and a decision function f, i.e., $T = \{Y, f\}$. The decision function is to be learned from the training data (X, Y). Tasks in MR brain imaging can be, for instance, the survival rate prediction of cancer patients, where f is the function that predicts the survival rate and Y is the set of all possible outcomes. Given a source domain D_S and task T_S and a target domain D_T and task T_T, transfer learning re-utilizes the knowledge acquired in D_S and T_S to improve the generalization of f_T in D_T. Importantly, D_S must be related to D_T and T_S must be related to T_T [57]; otherwise, transfer learning can worsen the accuracy in the target domain. This phenomenon, called negative transfer, has been recently formalized in [35] and studied in the context of MR brain imaging [56].

In our approach, inspired by the network architectures outlined in [59], we augment each pre-trained model by incorporating two new fully-connected layers. These layers are initialized with random weights and are accompanied by a softmax activation function. Then, we freeze the weights of the remaining layers in the network to prevent them from being updated during the training process. This ensures that the learned features in these layers, derived from the pre-trained model, remain intact while we focus on fine-tuning the added layers.

The fine-tuning of parameters, particularly λ, is crucial and depends on the specific characteristics of the pre-trained model and we set according to its category as the following: for

"Artistic Style Transfer" $\lambda = 9$; "Animal Images", "Natural Landscape Images" and "Photography" $\lambda = 10$; "Satellite and Map Images" and "Urban Scenes" $\lambda = 11$; Experiments have demonstrated the sensitivity of the results to changes in the λ value. Other parameters are set as follows: the batch size is 2, FC1 has 256 neurons, FC2 has 256 neurons (adjusted according to the dimensions of the CT and MR images), training is conducted for 200 epochs, and the activation function employed is softmax. These hyper-parameters were carefully selected based on the outcomes of our experiments, allowing us to tailor the model's performance to meet our objectives and accommodate the dataset's properties. During training, We keep the same learning rate for the first 100 epochs and linearly decay the rate to zero over the next 100 epochs This approach enables smoother convergence and aids in fine-tuning the model effectively. Algorithm 1 illustrates the bidirectional CT-MR translation using transfer learning from generic pre-trained models.

Algorithm 1 Bidirectional CT-MR translation using transfer learning from pre-trained models

---

Require: Generic pre-trained model, dataset (X ← CT, Y ← MR)
Ensure: Fine-tuned model
  for each pre-trained model **do**
    Load the pre-trained Model
    Initialize the new model with the pre-trained weights
    Finetune_model:
    (FC1; FC2;learning rate; activation; epoch; batch)

    Add_layer ($FC1$, activation)
    Add_layer ($FC2$, activation)
    Freeze pre-existing layers
    Set learning rate
    Loss Function:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y)$$
$$+L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (3)$$

    Build a new model
    while not converged **do**
      for each epoch **do**
        for each batch of images from the target domain **do**
          Train new model (X, Y, D_X, D_Y).
        end for
      end for
      Evaluate new model (X, Y, D_X, D_Y).
    end while
  end for

C. pre-trained model

Generic pre-trained models refer to models that are trained on datasets that are not specific to the medical field. These models can be used as a starting point for medical image analysis tasks such as segmentation, classification, and detection.

In this study, we investigate the effect of using non-medical pre-trained models on MR-to-CT synthesis [42].

We used 18 generic pre-trained models for the task of medical bi-directional MR-CT translation. These models were selected from a range of publicly available models, for different types of images, including Artistic Style Transfer, Animal Images, Natural Landscape Images, Photography, Satellite and Map Images, and Urban Scenes, which can be used as a basis for developing models for MR to CT translation. These pre-trained models are trained on large-scale datasets of high-quality images and can learn high-level features that can be transferred to other domains, such as medical imaging. By leveraging pre-trained models, researchers can save significant amounts of time and computational resources and achieve better results than by training models from scratch. For example, the style transfer pre-trained models can learn the underlying style of an image and apply it to another image, which can be useful in developing models for translating MR to CT images with similar styles. Similarly, for animal and natural landscape images, pre-trained models can learn features specific to those domains, which can be used in developing models for medical imaging.

Also, pre-trained models that were trained on satellite and map images can be used to make models for translating images from MRI to CT that were taken with imaging methods like computed tomography (CT) and magnetic resonance imaging (MRI). Satellite and map images often show the edges and shapes of buildings, roads, and other structures, just like medical images do. Also, models that have already been trained for urban scenes can learn details about the buildings and objects that are common in cities. These details can be used to create models for medical imaging of organs and tissues in the body that have similar details.

The models that were used fall into the following categories and Figure 3 shows an example of these categories:

1) **Artistic Style Transfer:** Style transfer is a technique that utilizes pre-trained models to apply the artistic style of paintings to other images. Models like style_monet, style_Cezanne, style_ukiyoe, and style_vanGogh are trained on various artists' paintings. The method optimizes both content and style, creating new images with unique characteristics. It uses content and style loss functions, minimizing them to combine target image content with a reference image style. This technique finds applications in image synthesis, style transfer, and domain adaptation, making it valuable for artists, designers, and social media influencers seeking visually appealing and stylized images.

2) **Animal Images:** Pre-trained models use animal images for object detection, classification, and segmentation. Models like horse2zebra and zebra2horse are trained on labelled datasets to generate images resembling horses or zebras. Variability in animal appearance is crucial. Models are optimized with backpropagation and stochastic gradient descent. Applications include wildlife conservation, ecology, agriculture, and entertainment. Accurate animal detection aids in preserving endangered species and promoting biodiversity.
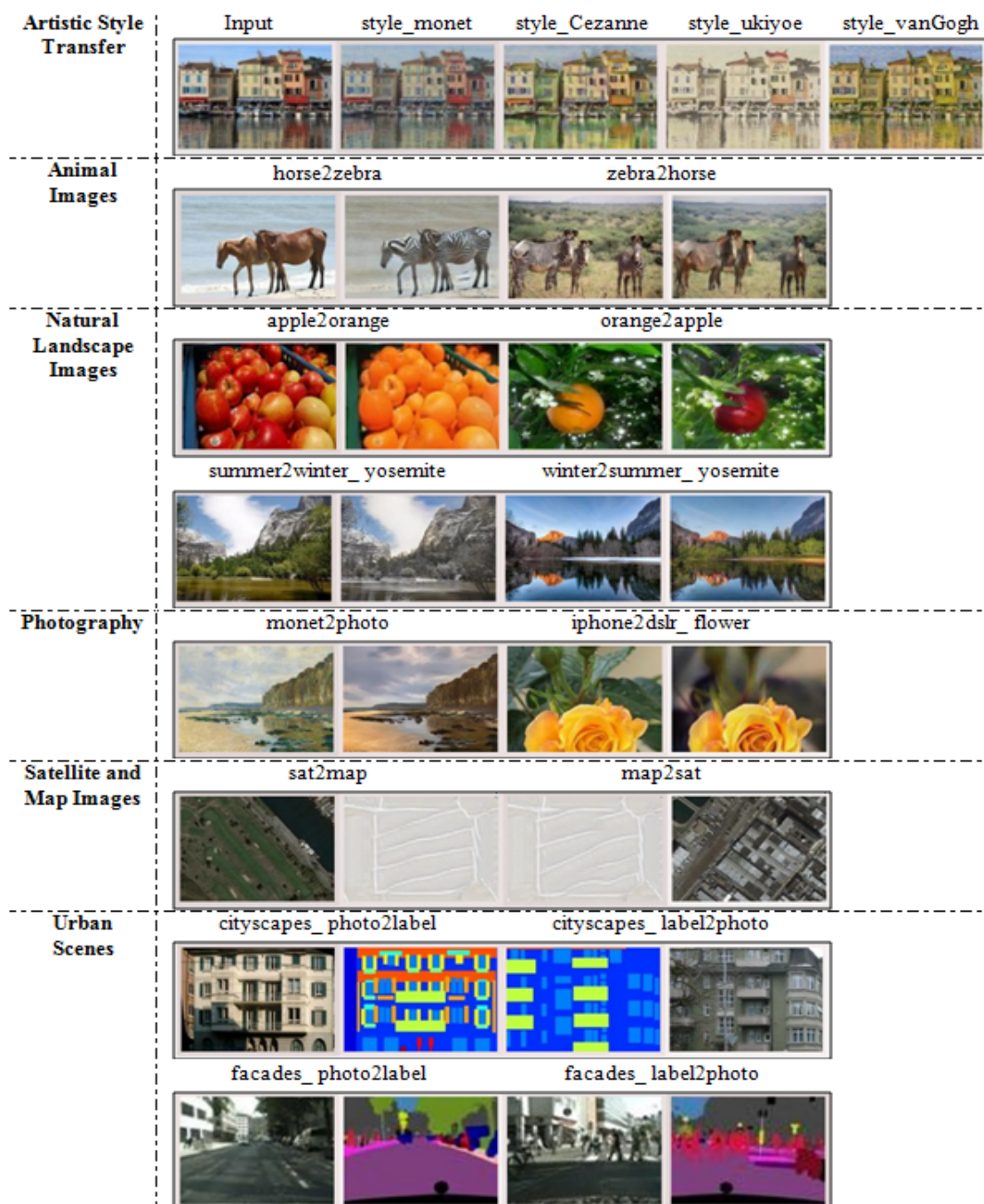
Figure 3: Examples of generic pre-trained models.

3) **Natural Landscape Images:** Pre-trained image-to-image translation models transform landscapes between different seasons while preserving key features. Examples include apple2orange, orange2apple, summer2winter_yosemite, and winter2summer_yosemite. Trained on natural Yosemite landscape images, the models depict forests, mountains, rivers, and beaches. Classifying and recognizing objects in such complex landscapes can be challenging. Optimization is done with datasets like Places2 and Open Images. Applications include object recognition, semantic segmentation, scene recognition, and more, benefiting ecology, tourism, and climate change research for analyzing environmental changes and their impact on ecosystems.

4) **Photography:** Image-to-image translation involves using pre-trained models like monet2photo and iphone2dslr_flower for transforming images from one style to another. These models are trained on datasets of Monet paintings and real photographs to create images with similar appearance. They exhibit high resolution, rich texture, and detailed visual features, including lighting, shadows, and color. Object detection, recognition, and segmentation are incorporated to classify and locate objects accurately. Optimization is done with large datasets like ImageNet, COCO, Pascal VOC, and Open Images. These versatile models find applications in surveillance, healthcare, and education, such as disease diagnosis in medical imaging, realistic simulations in education, and object detection in surveillance scenarios.

5) **Satellite and Map Images:** In computer vision, the analysis of aerial or satellite images of the Earth's surface is done. Pre-trained models like sat2map and map2sat perform image-to-image translation of satellite and map images with high spatial resolution, enabling the identification of small-scale surface features. These models incorporate land cover classification, object detection, and change detection for environmental analysis. Optimized through transfer learning, they find applications in urban planning, agriculture, natural resource management, and disaster response. For instance, they aid in urban growth analysis, crop monitoring, land use change detection, and disaster-affected area identification for rapid response and relief efforts.

6) **Urban Scenes:** Pre-trained models like cityscapes_photo2label, cityscapes_label2photo, facades_photo2label, and facades_label2photo are designed for image-to-image translation between urban scene images and their corresponding semantic label maps. Trained on datasets of urban scene images and semantic label maps, these models facilitate object detection, segmentation, and classification. Urban scenes pose challenges with complexity, high density, dynamic elements, and varying lighting conditions, making traditional computer vision techniques less effective. Convolutional neural networks (CNNs) used in these models enable accurate and efficient performance. Optimization with pre-trained models enhances their effectiveness, especially with limited training data.

Applications include urban planning, traffic management, and security, automating urban scene analysis with machine learning and computer vision techniques. Examples include traffic flow analysis, vehicle and pedestrian detection and tracking, and public safety and security monitoring in urban areas.

The pre-trained models mentioned in the provided description utilize the same network architecture for both the generator and discriminator components. However, the key point of differentiation lies in the training dataset used for each model. The network architecture is adapted from [59].

Generator architecture:This architecture employs six residual blocks for training images of size 128×128, while nine residual blocks are used for higher-resolution images such as 256×256 or beyond. The architecture comprises various layers denoted by specific notations: $c_7 s_1 - k$ represents a 7×7 Convolution-Instance Norm ReLU layer with k filters and stride 1, d_k denotes a 3×3 Convolution-Instance Norm-ReLU layer with k filters and stride 2, and R_k signifies a residual block containing two 3×3 convolutional layers with the same number of filters. Additionally, u_k denotes a 3×3 fractional-strided Convolution-Instance Norm-ReLU layer with k filters and stride $^1/_2$. To mitigate artifacts, reflection padding is employed. For a generator network with six residual blocks, the architecture consists of: $c_7 s_1 - 64$, d_{128}, d_{256}, followed by six R_{256} blocks, u_{128}, u_{64}, and finally $c_7 s_1 - 3$. Meanwhile, a generator with nine residual blocks includes $c_7 s_1 - 64$, d_{128}, d_{256}, followed by nine R_{256} blocks, u_{128}, u_{64}, and $c_7 s_1 - 3$.

Discriminator architecture:The discriminator architecture comprises layers denoted by c_k, representing a 4×4 Convolution-InstanceNorm-LeakyReLU layer with k filters and stride 2. Notably, InstanceNorm is excluded for the first c_{64} layer, and leaky ReLUs with a slope of 0.2 are employed. The discriminator architecture sequentially follows a structure of $c_{64} - c_{128} - c_{256} - c_{512}$, with a final convolutional layer producing a 1-dimensional output.

The Table II encapsulates crucial details regarding the images utilized in the training process for each pre-trained model [59]. Each entry delineates the specific dataset employed, offering insights into the diversity and scale of image data utilized in the training regimen. These datasets serve as the foundational bedrock upon which the pre-trained models are built, encompassing a spectrum of visual information necessary for robust and comprehensive training. The effectiveness of these models in MR to CT translation tasks is then evaluated. By scrutinizing the provided dataset information, it is possible to determine which pre-trained model shows the highest efficiency at this specific translation task.

IV. EXPERIMENTS AND RESULTS

Translation from MR–CT brain tumors presented. First, pre-processing was performed on the raw MRI images. The transfer learning method was used to reduce the processing load and achieve successful results with a small amount of data.

Table II: Datasets Used to Train the Generic Pre-Trained Models

Category	Pre-trained model	Photo type	Numbers of training images	Image size	Source of images
Artistic Style Transfer	Style_ monet	monet	1074	256 × 256	Flickr API using the "landscape" tag.
	Style_ Cezanne	cezanne	584		
	Style_ ukiyoe	ukiyoe	1433		
	Style_ vanGogh	vanGogh	401		
Animal Images	horse ↔ zebra	zebra	1177	256 × 256	ImageNet [12]
		horse	939		
Natural Landscape Images	apple ↔ orange	apple	996	256 × 256	ImageNet [12]
		orange	1020		
	summer ↔ winter	summer	1273	256 × 256	Flickr API with the tag "yosemite".
		winter	854		
Photography	monet ↔ photo	monet	1074	256 × 256	Flickr API with the tag "monet".
		photo	6853		
	iphone ↔ dslr flower	iphone_flower	1813	256 × 256	Flickr API with the tag "flower"
		dslr _flower	3326		
Satellite and Map Images	sat ↔ map	sat	400	256 × 256	Google Maps [19]
		map	400		
Urban Scenes	Cityscapes_photo ↔ label	cityscapes	2975	128 × 128	Cityscapes training set [10]
		label	2975		
	facades photo ↔ label	photo	400	256 × 256	CMP Facade Database [49].
		label	400		

A. Dataset

We used a real MR-CT dataset to train, develop and demonstrate the capabilities of our model. The dataset was obtained from [2] and included 367 paired MR-CT brain images from 18 patients. To align the images, a mutual information rigid registration algorithm was used, and the MR images were corrected with the N3 bias field correction algorithm. The dataset contains both normal and abnormal(i.e., contains tumors) MR and CT images. With the RadiAnt DICOM viewer software, 2D image slices were taken from the MR-CT datasets and used in the experiments. The extracted images were then transformed into a PNG image data format with a resolution of 256×256 pixels. Estimation of the effectiveness and efficiency of the proposed model is done by training and testing datasets which are divided into we perform training on the 70% of the given dataset and rest 30% is used for the testing purpose. The model was tested using Google's Colab cloud service, TensorFlow 2.0, and the Python 3 framework. Our dataset is paired dataset which means; Each MR image in the dataset has a corresponding CT image, and vice versa. This pairing is essential for training and evaluating algorithms that aim to translate or convert MR images to CT images, or vice versa.

B. Performance Metrics

Full-Reference Image Quality Assessment (FR-IQA) is an automatic perceptual quality evaluation of a distorted image in comparison with a reference image. Images in FR-IQA are evaluated using PSNR and SSIM, two state-of-the-art evaluation metrics [1]. These measures determine the level of distortion in synthetically created images. PSNR is the quickest and easiest method for gauging image quality. However, PSNR does not always correspond with what humans see and the quality of an image. It was suggested that more factors be used to overcome the limitation of PSNR metrics, specifically SSIM. UQI and VIF were also utilized in our analysis.

1) **Peak Signal-to-Noise Ratio (PSNR):** PSNR is widely utilized by many academics for image comparison and

image synthesis because of its simplicity and ease of implementation, making it the most effective measuring tool to evaluate synthetic images [41]. The PSNR evaluates how far off the generated image pixels are from the ground truth by measuring the amount of pixel loss. The paired images in the test dataset can only be evaluated using pixel loss-based measures like PSNR and SSIM. If the PSNR is high enough, a high-quality image will be produced. The equation for PSNR is represented as [1]

$$PSNR = 20. \log_{10}\left(\frac{R^2}{MSE}\right) \qquad (4)$$

where MSE is a Mean Square Error, and R is the maximum fluctuation in the input image data type. According to (4), if the input image has a double-precision floating-point data type, then R is 1. If it has an 8-bit unsigned integer data type, then R is 255 [1].

2) **Structural Similarity Index (SSIM):** Similarity between two images (x and y) can be evaluated with the use of SSIM. To quantify or predict image quality using the SSIM index, an original uncompressed or distortion-free image must be used as a baseline. SSIM is a perception-based model that takes into account luminance masking and contrast masking as well as other crucial perceptual phenomena to account for how image quality is considered to have changed over time, which is defined as [30]

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (5)$$

where, μ_x and μ_y represent the mean of the particular image; σ_x^2 and σ_y^2 represent the standard deviation of the of the particular image; σ_{xy} represents the covariance between two images, and C_1, C_2 are constants set for avoiding instability.

3) **Universal Quality Index (UQI):** In order to compare the distortion information between the original image and the distorted image, a mathematical metric called UQI is

calculated, bypassing the need for a model of the human visual system. Loss of correlation, illuminance distortion, and contrast distortion all contribute to UQI. This metric is simple to compute and has several potential uses in the field of image processing. The formula for UQI is defined as [52, 1]

$$UQI = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\tilde{x}\tilde{y}}{\tilde{x}^2 + \tilde{y}^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \tag{6}$$

The three components of the equation represent the loss of correlation, distortion of luminance, and distortion of contrast factors, where x is the original image and y is the generated image. $\tilde{x}$ and $\tilde{y}$ are defined as

$$\tilde{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \tilde{y} = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \tilde{x})^2, \quad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \tilde{y})^2,$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \tilde{x})(y_i - \tilde{y}). \tag{7}$$

To perform multiplication and addition operations, x and y images must be square images with $N \times N$.

4) **Visual Information Fidelity (VIF):** Natural scene statistics and the idea of picture information retrieved by the human visual system form the basis of the Visual Information Feature (VIF). In Figure 4, C represents the source image, D represents the distorted image, E represents the output of the Human Visual System (HVS) for the source image, and F represents the output of the HVS for the deformed image. VIF can be represented as [43, 1]

$$\text{VIF} = \frac{\sum_{k=1}^{M}\sum_{l=1}^{N} I_f(x_{kl}; y_{kl} \mid H_{kl})}{\sum_{k=1}^{M}\sum_{l=1}^{N} I_f(x_{kl}; x_{kl} \mid H_{kl})} \tag{8}$$

where x_{kl} and y_{kl} are the wavelet coefficients of the reference and distorted image, respectively; H_{kl} represents the human visual system (HVS) model; $I_f(x_{kl}; y_{kl} \mid H_{kl})$ is the mutual information between the wavelet coefficients of the reference and distorted images given the HVS model; $I_f(x_{kl}; x_{kl} \mid H_{kl})$ is the mutual information between the wavelet coefficients of the reference image given the HVS model; M and N represent the number of subbands and the number of coefficients in each subband, respectively.
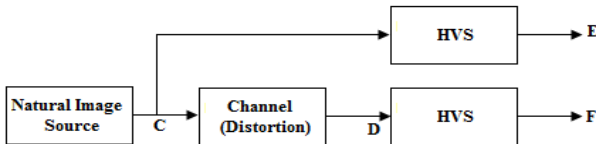

Figure 4: VIF components

C. Comparison of a pre-trained model

In image synthesis tasks, the evaluation of generated images is crucial to assessing the quality of the model's output. The commonly used PSNR metric, which calculates the mean square error between the original and generated images, has limitations as it does not always correlate with human visual perception. Therefore, additional metrics such as SSIM, UQI, and VIF have been proposed to overcome these limitations. In this work, the performance of the model is evaluated using these metrics on both the paired and unpaired datasets. The experiments compared the bidirectional MR-CT synthesis using 18 pre-trained non-medical models, and the iphone2dslr_flower model outperformed the others, resulting in the best score across most of the chosen metrics, with low standard deviation values indicating the stability of the results. Figures 5 and 6 illustrate Bidirectional MR-CT Translation, using a pre-trained model.

Table III and IV present the outcomes of the PSNR, SSIM, UQI, and VIF metrics for the bidirectional MR-CT synthesis, covering both MR-to-CT and CT-to-MR translation using the entire pre-trained model. However, the table shows the evaluation metrics of various pre-trained models on different image-to-image translation tasks. The metrics used are PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), UQI (Universal Quality Index), and VIF (Visual Information Fidelity). PSNR measures the quality of the reconstructed image compared to the original image, where a higher PSNR value indicates better image quality. SSIM measures the similarity between the reconstructed image and the original image based on structural information, where a higher SSIM value indicates better image quality. UQI measures the similarity between the reconstructed image and the original image based on human perception, where a higher UQI value indicates better image quality. VIF measures the information transfer between the reconstructed image and the original image, where a higher VIF value indicates better image quality.

The table shows that different pre-trained models perform differently on different image-to-image translation tasks. For example, the model "iPhone2dslr_flower" has the highest PSNR and UQI values, indicating that it produces high-quality reconstructed images that are similar to the original images. On the other hand, the model "apple2orange" has the lowest PSNR and UQI values, indicating that it produces lower-quality reconstructed images that are less similar to the original images. Based on the CT-MR table, it can be seen that the pre-trained model "iphone2dslr_flower" has the highest PSNR value (34.355), indicating better image reconstruction. The model "summer2winter_yosemite" has the highest SSIM (0.393) and UQI (0.393), indicating a high level of structural similarity and image quality between the two image domains. The model "zebra2horse" has the highest VIF (0.001113), indicating the highest level of visual information fidelity between the two image domains.

In the next section, we conducted a perceptual and validation study to further evaluate the performance of the pre-trained image-to-image translation models. We recognized that while metrics such as PSNR, SSIM, UQI, and VIF are helpful in
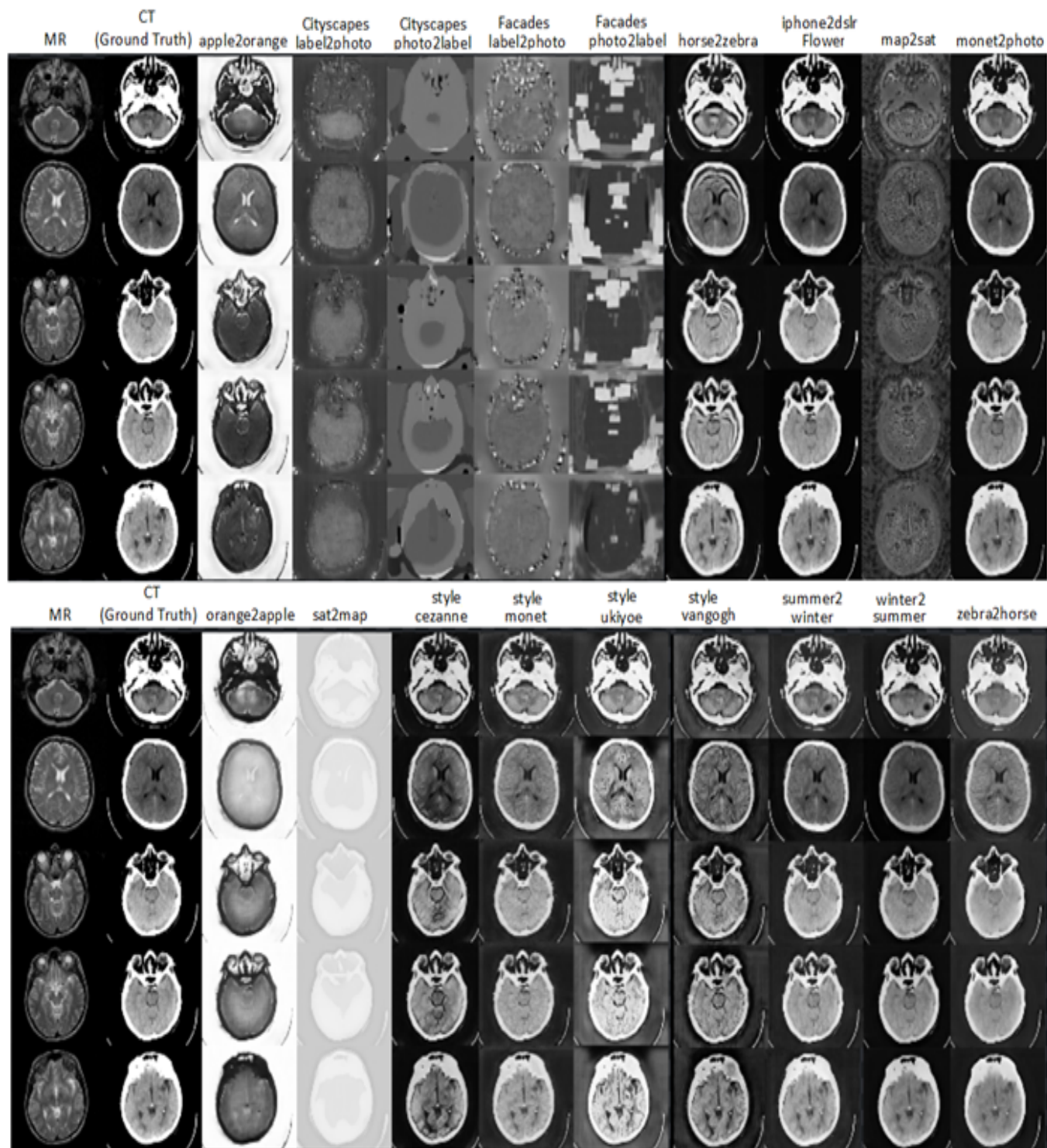
Figure 5: Output results of Magnetic Resonance to Computed Tomography (MR-CT) image translation from different generic pre-trained models.
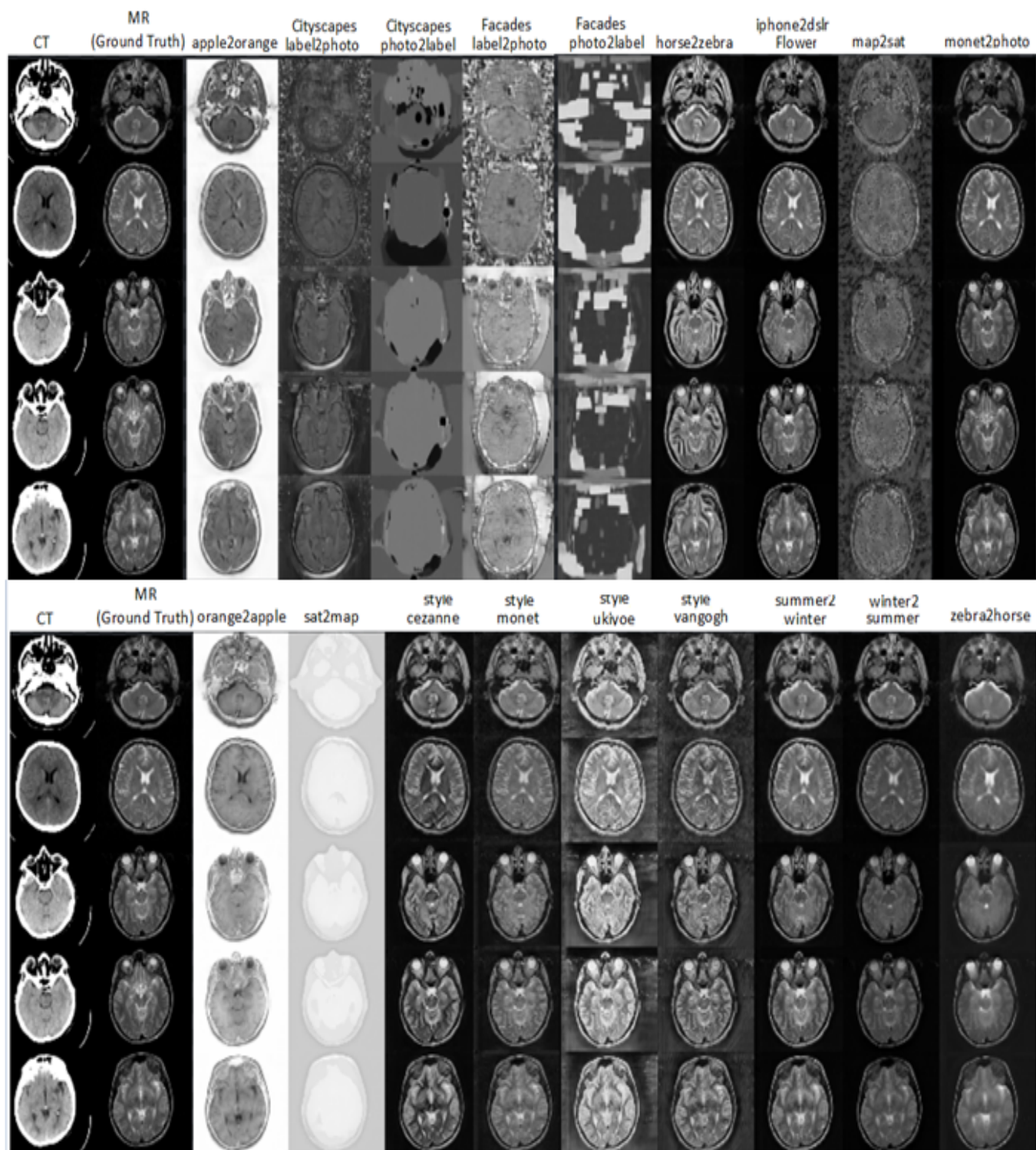
Figure 6: Output results of Computed Tomography to Magnetic Resonance (CT-MR) image translation from different generic pre-trained models.

Table III: Image Quality Evaluation Metrics for MR-CT Translation

		MR-CT							
NO.	Pre-trained Model	PSNR		SSIM		UQI		VIF	
		AVG	STD	AVG	STD	AVG	STD	AVG	STD
1	apple2orange	28.63	±0.791	0.07	±0.052	0.02	±0.023	0.02	±0.034
2	cityscapes_photo2label,	27.67	±0.225	0.08	±0.033	0.01	±0.013	0.01	±0.024
3	cityscapes_label2photo	27.61	±0.275	0.14	±0.039	0.00	±0.003	0.00	±0.014
4	facades_label2photo	28.36	±0.389	0.07	±0.035	0.03	±0.033	0.03	±0.044
5	facades_photo2label	27.67	±0.769	0.07	±0.040	0.01	±0.013	0.01	±0.024
6	horse2zebra	27.67	±0.503	0.36	±0.065	0.01	±0.011	0.01	±0.022
7	iphone2dslr_flower	30.98	±0.119	0.65	±0.031	0.00	±0.000	0.01	±0.001
8	map2sat	27.72	±0.142	0.16	±0.082	0.02	±0.021	0.02	±0.033
9	monet2photo	29.54	±1.686	0.50	±0.076	0.03	±0.032	0.04	±0.043
10	orange2apple	30.74	±1.054	0.05	±0.042	0.01	±0.012	0.01	±0.022
11	sat2map	26.82	±0.511	0.14	±0.031	0.00	±0.002	0.01	±0.014
12	style_cezanne	27.55	±0.207	0.38	±0.068	0.02	±0.022	0.03	±0.034
13	style_monet	27.56	±0.160	0.39	±0.068	0.02	±0.024	0.03	±0.036
14	style_ukiyoe	27.79	±0.288	0.33	±0.065	0.02	±0.024	0.03	±0.036
15	style_vangogh	27.80	±0.164	0.35	±0.069	0.03	±0.031	0.03	±0.042
16	summer2winter_yosemite	27.96	±0.461	0.30	±0.041	0.01	±0.012	0.01	±0.024
17	winter2summer_yosemite	29.22	±1.165	0.38	±0.084	0.04	±0.043	0.05	±0.054
18	zebra2horse	27.81	±0.460	0.35	±0.060	0.02	±0.023	0.03	±0.034

Table IV: Image Quality Evaluation Metrics for CT-MR Translation

		CT-MR							
NO.	Pre-trained Model	PSNR		SSIM		UQI		VIF	
		AVG	STD	AVG	STD	AVG	STD	AVG	STD
1	apple2orange	29.55	±0.781	0.16	±0.039	0.02	±0.034	0.12	±0.013
2	cityscapes_photo2label,	27.61	±0.105	0.06	±0.025	0.01	±0.024	0.11	±0.003
3	cityscapes_label2photo	27.27	±0.306	0.09	±0.027	0.00	±0.014	0.10	±0.007
4	facades_label2photo	28.16	±0.272	0.09	±0.035	0.03	±0.044	0.13	±0.023
5	facades_photo2label	27.62	±0.172	0.04	±0.029	0.01	±0.024	0.11	±0.003
6	horse2zebra	29.30	±0.662	0.40	±0.090	0.01	±0.022	0.11	±0.001
7	iphone2dslr_flower	34.36	±0.072	0.83	±0.022	0.00	±0.001	0.10	±0.020
8	map2sat	27.77	±1.619	0.18	±0.036	0.02	±0.032	0.12	±0.011
9	monet2photo	33.49	±1.823	0.72	±0.069	0.03	±0.043	0.13	±0.022
10	orange2apple	31.02	±0.788	0.13	±0.034	0.01	±0.022	0.11	±0.001
11	sat2map	26.70	±0.778	0.08	±0.080	0.00	±0.013	0.10	±0.008
12	style_cezanne	28.22	±0.329	0.43	±0.082	0.02	±0.033	0.12	±0.013
13	style_monet	27.78	±0.269	0.43	±0.085	0.02	±0.035	0.13	±0.015
14	style_ukiyoe	27.54	±0.097	0.29	±0.068	0.02	±0.035	0.13	±0.014
15	style_vangogh	27.74	±0.160	0.36	±0.080	0.03	±0.042	0.13	±0.021
16	summer2winter_yosemite	30.16	±1.356	0.39	±0.061	0.01	±0.023	0.11	±0.002
17	winter2summer_yosemite	30.79	±0.715	0.44	±0.074	0.04	±0.054	0.14	±0.033
18	zebra2horse	27.88	±0.522	0.34	±0.078	0.02	±0.034	0.13	±0.013

quantitatively assessing the quality of the translated images, they do not necessarily capture the full extent of human perception and preferences. Therefore, we designed a study to include human evaluators who would provide qualitative feedback on the translated images, considering factors such as visual realism, image diversity, and overall aesthetic appeal. The validation study would help us to better understand the strengths and limitations of each pre-trained model and provide a more comprehensive evaluation of their performance.

D. Perceptual study and validation

The accuracy of the translated MR and CT pictures is assessed by showing them to two expert radiologists from JUH in a random order alongside the ground truth images. From a pool of 18 non-medical pre-trained models, the best one was selected for this analysis using PSNR, SSIM, UQI, and VIF metrics. With the primary goal of evaluating the images generated by the transfer learning model and comparing them to the ground truth images, the radiologists are shown a total of 240 MR images and 240 CT images with a resolution of 256 x 256 pixels. Radiologists are asked to determine which images are the ground truth and to assess the images' realism on a scale from 1 to 4, with 4 being the most realistic.

Table V presents the results of the perceptual study evaluated by radiologists for MR-to-CT and CT-to-MR translations. The final column of this table shows the percentage of images classified as real by the radiologists over the total number of images.

The given table shows the performance of the iphone2dslr_flower pre-trained model in MR-to-CT translations. It is shown that the model achieved a mean score of 3.79, which reflects its good performance in generating CT images from MR images. Also, the table goes on to state that this pre-trained model also performs well in CT-to-MR translations, achieving a mean score of 3.69. Moreover, the success rate for CT images is 97.91%, while for MR images, it is 97.7%, which indicates that the radiologist was convinced that they were real scans.

Table V: Results of Perceptual Study

MR-CT			
Model	Mean	Std	Real%
iphone2dslr_flower	3.79	0.21	97.91%
Ground truth	3.85	0.23	98.58%
CT-MR			
Model	Mean	Std	Real%
iphone2dslr_flower	3.69	0.21	97.7%
Ground truth	3.54	0.23	98.37%

We compute Lin's concordance correlation coefficients (CCC) for the results of the perceptual study, which quantifies the agreement between the ground truth images and the generated images. Table VI presents Pearson's correlation coefficient (ρC) and accuracy ($C\beta$) for both MR-to- CT and CT-to-MR translation. The results indicate a high or medium correlation between ground truth and the generated images for all radiologists. Details related to the results by the radiologists can be found in Appendix VII-A which shows an example of a form completed by a radiologist for MR-CT images and CT-MR images.

Table VI: Agreement of Generated and Ground-Truth Images

pre-trained model	MR-CT		CT-MR	
	(ρC)	($C\beta$)	(ρC)	($C\beta$)
Radiologist1	0.69	0.99	0.88	0.94
Radiologist2	0.97	1.00	0.60	0.97

E. Latent Space

Latent space, in the context of machine learning and data representation, refers to a lower-dimensional, abstract space where essential information and patterns of data are encoded [34]. It is a transformative concept that allows complex data, such as images or text, to be condensed into a more manageable and informative form. In the case of medical image translation from magnetic resonance imaging (MRI) to computed tomography (CT), the concept of latent space becomes particularly valuable [54]. This process effectively bridges the gap between the two imaging techniques, capturing common features and reducing the impact of modality-specific differences [34]. Latent space, often created through dimensionality reduction techniques like Principal Component Analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE), allows us to represent complex and high-dimensional data in a more comprehensible manner. In various fields, from machine learning to biology to social sciences, latent space plays a pivotal role in revealing hidden patterns, relationships, and clusters within data [29]. In our work, a latent space was created for the pre trained models that performed well, which is :apple2orange, orange2apple, horse2zebra, zebra2horse, style_monet, style_Cezanne, style_ukiyoe, style_vanGogh, summer2winter _yosemite, winter2summer_yosemite, monet2photo and iphone2dslr_flower). Figure 7 provides a visual representation of the latent space, offering a compelling illustration of how the iphone2dslr_flower translation task excelled in correctly separating the distinctive characteristics of the input images. The latent space in Figure 7 has two colors, these colors represent distinct features or characteristics extracted from the input images. The simplification into two colors implies a high-level abstraction or compression of information and the base model was trained for two classes, where red represents CT images and blue represents MR images. However, upon closer inspection, it is clear that colors in most of pre-trained models overlap within the existing space, indicating the models' difficulty in effectively separating or distinguishing input image elements during training. This difficulty may stem from various factors such as variability within the data, complex relationships between MRI and CT features, or limitations within the model structure.

Conversely, in the case of the iphone2dslr_flower model, clear color separation was observed within the latent space. This indicates that the model has successfully learned to distinguish between distinct features or patterns found in MR and CT images, indicating effective training. This analysis indicates that the model has encoded essential information within the latent space in a way that preserves the fundamental differences between MR and CT methods. Thus, when generating cross-sectional images through inverse MR imports, the model translation and underlying features can be better understood and interpreted. Clearer separation within the latent space implies more robust and accurate performance, as the model can more effectively differentiate between the diverse effects of each imaging modality. This underscores the importance of choosing a model that relies not only on visual observations, but on an in-depth analysis of the data it produces.

V. DISCUSSION:

In Fig. 5, a comprehensive analysis of the outcomes reveals a consistent correlation with the findings presented in Fig. 3, which showcases the diverse results obtained from the pre-trained model. Notably, a subset of these models, including facades photo2label and facades label2photo, exhibited erroneous outcomes attributed to their primary specialization in region identification and labeling. Moreover, the sat2map variant also demonstrated significantly divergent results from the desired outcomes. As a result of our meticulous examination, we arrive at the conclusion that the crucial determinant for achieving satisfactory results lies in the judicious selection of a pre-trained model that is specifically tailored to a dataset possessing properties closely resembling the new images used for the model's training. This underscores the pivotal role played by appropriate model selection in ensuring the successful training of the model. The iphone2dslr_flower model stands out as the premier pre-trained non-medical model for image translation, and it has also exhibited exceptional capabilities in the medical field, particularly in translating MRI images into CT scans, achieving top scores in all four evaluation scales. This pre-trained generative model was trained on a vast image
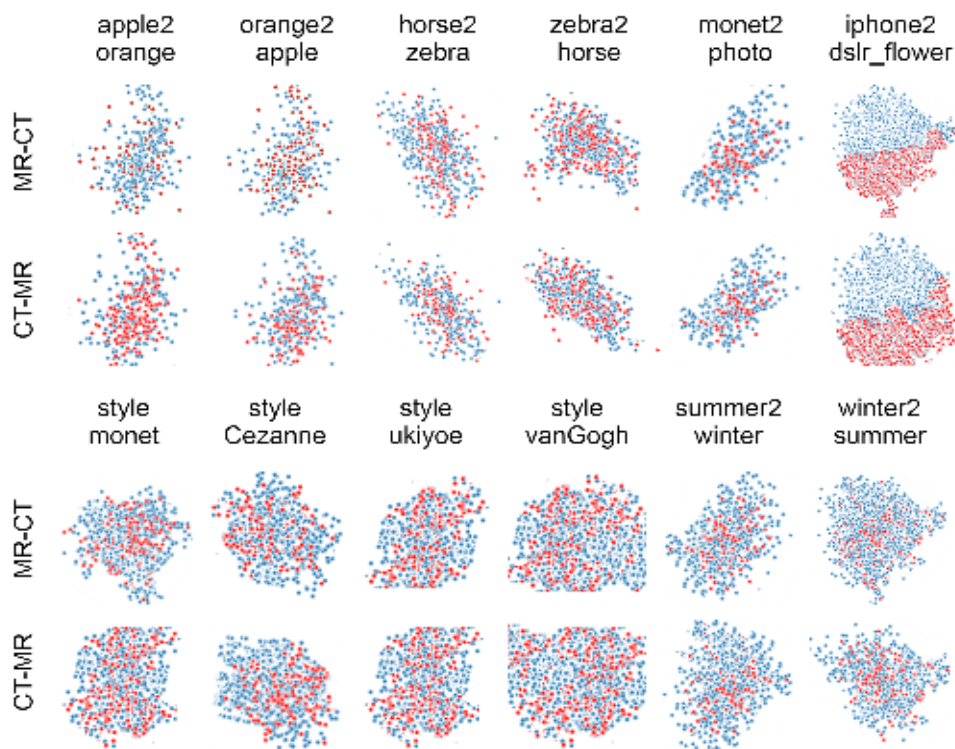
Figure 7: Latent space visualization using different pre-trained generic models for MR-CT and CT-MR image translation.

dataset, the iphone2dslr_flower dataset, which proves to be an ideal choice for transfer learning in photomontage tasks. This dataset comprises images of flowers captured by both iPhone and DSLR cameras, allowing the training of models capable of generating high-resolution flower images that closely resemble the quality of those taken with a DSLR camera. The model's training involved a diverse range of images, including flowers that share visual similarities with certain structures in the human body, such as blood vessels [24]. Additionally, some flower types exhibit compact corals with intricate convoluted surfaces reminiscent of brain structures, making this dataset especially valuable for models aiming to reproduce the intricate details and textures found in these floral specimens. The petals displayed a striking resemblance to the inner tissues of the brain, exhibiting a repeating pattern that mirrors the distinctive torsions and folds present in cerebral structures [21]. This intriguing similarity between the model's results and brain anatomy highlights the potential of this approach to capture complex biological patterns and structures. The model's ability to capture such fine details indicates a close match between the training data and brain features, ultimately contributing to its exceptional performance in the realm of brain image analysis. Figure 8 showcases some of the flowers utilized in the iphone2dslr_flower dataset, further illustrating the dataset's diversity and relevance for the model's training. and Figure 9 Resemblance between Brain Images and iPhone Data. The images showcase striking similarities between brain structures and the data generated by the iPhone model. The intricate

patterns and textures found in certain types of flowers captured by the iPhone camera demonstrate a captivating resemblance to brain tissues [14, 22], highlighting the potential of the model in capturing complex biological structures.

Figure 8: Sample images from generic iphone2dslr_flower pre-trained model.

The model can learn the features and patterns present in the images, which can then be used to generate new images. The process includes loading the pre-trained model, preparing the input image, using the pre-trained model to generate the faked image, post-processing, and evaluating the generated image using metrics such as MSE and SSIM [57]. It is worth

Figure 9: Physiological characteristic similarities between generic iphone2dslr_flower images and brain images.

mentioning that the iphone2dslr_flower model is a conditional generative model that can generate realistic flower images from a given input image, it is not directly designed for medical image translation, but with some adjustments and fine-tuning it could be used to generate images that are similar in appearance and information content to the original images.

VI. CONCLUSION:

It has been demonstrated by our study that the potential of pre-trained models for the task of medical image translation, specifically for MRI-CT and vice versa, can be seen. Eighteen pre-trained models were used from a range of publicly available models, including models trained on natural images and models specifically designed for medical image translation tasks. It was found that these models performed well on all four commonly used image quality metrics: PSNR, SSIM, UQI, and VIF. It was also suggested that models that had been specifically designed for medical image translation tasks and trained on a large dataset of medical images performed better than models that had been trained on general image translation tasks. However, it is essential to note that to obtain the best results for specific medical image translation tasks, such as brain, lung, or cardiac images, it may be necessary to fine-tune the models on a dataset of medical images specific to the task at hand and use a larger dataset. This way, the efficiency and accuracy of diagnosis and treatment in clinical practice can be improved. The best performance across all metrics was achieved by the 'iphone2dslr_flower' model, thus making it the most suitable model for this task. The potential of transfer learning in medical imaging and the effectiveness of the 'iphone2dslr_ flower' model have been demonstrated by our findings. In general, the promising capabilities of pre-trained models for medical image translation tasks have been confirmed by this study, and it has been suggested that these techniques could be implemented in the near future for clinical use.

ACKNOWLEDGMENT

REFERENCES

[1] Alaa Abu-Srhan, Mohammad AM Abushariah, and Omar S Al-Kadi. "The effect of loss function on conditional generative adversarial networks". In: *Journal of King Saud University-Computer and Information Sciences* 34.9 (2022), pp. 6977–6988.

[2] Alaa Abu-Srhan et al. "Paired-unpaired Unsupervised Attention Guided GAN with transfer learning for bidirectional brain MR-CT synthesis". In: *Computers in Biology and Medicine* 136 (2021), p. 104763.

[3] Manar Ahmed Hamza et al. "Optimal and Efficient Deep Learning Model for Brain Tumor Magnetic Resonance Imaging Classification and Analysis". In: *Applied Sciences* 12.15 (2022), p. 7953.

[4] Mohamed Arbane et al. "Transfer learning for automatic brain tumor classification using MRI images". In: *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*. IEEE. 2021, pp. 210–214.

[5] Martin J van den Bent et al. "Adjuvant procarbazine, lomustine and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951". In: *Journal of clinical oncology* 31.3 (2013), pp. 344–350.

[6] Yuda Bi. "Exploring the Power of Generative Deep Learning for Image-to-Image Translation and MRI Reconstruction: A Cross-Domain Review". In: *arXiv preprint arXiv:2303.09012* (2023).

[7] Rayene Chelghoum et al. "Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images". In: *Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part I 16*. Springer. 2020, pp. 189–200.

[8] Veronika Cheplygina, Marleen de Bruijne, and Josien PW Pluim. "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis". In: *Medical image analysis* 54 (2019), pp. 280–296.

[9] Ahmet Cinar and Muhammed Yildirim. "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture". In: *Medical hypotheses* 139 (2020), p. 109684.

[10] Marius Cordts et al. "The cityscapes dataset for semantic urban scene understanding". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 3213–3223.

[11] S Deepak and PM Ameer. "Automated categorization of brain tumor from MRI using cnn features and svm". In: *Journal of Ambient Intelligence and Humanized Computing* 12 (2021), pp. 8357–8369.

[12] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[13] Adrien Depeursinge, Omar S Al-Kadi, and J Ross Mitchell. *Biomedical texture analysis: fundamentals, tools and challenges*. Academic Press, 2017.

[14] Antonio Di Ieva and Omar S Al-Kadi. "Computational Fractal-Based Analysis of Brain Tumor Microvascular Networks". In: *The Fractal Geometry of the Brain*. Springer, 2024, pp. 525–544.

[15] Roa' Al-Emaryeen et al. "Deepfake Image Generation for Improved Brain Tumor Segmentation". In: *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. IEEE. 2023, pp. 6–11.

[16] Liang Ge et al. "On handling the negative transfer and imbalanced distributions in multiple source transfer learning". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 7.4 (2014), pp. 254–271.

[17] Alejandro Gutierrez et al. "Lesion-preserving unpaired image-to-image translation between MRI and CT from ischemic stroke patients". In: *International Journal of Computer Assisted Radiology and Surgery* (2023), pp. 1–10.

[18] Ruqian Hao et al. "A transfer learning–based active learning framework for brain tumor classification". In: *Frontiers in Artificial Intelligence* 4 (2021), p. 635766.

[19] Phillip Isola et al. "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.

[20] Omar S Al-Kadi. "A gabor filter texture analysis approach for histopathological brain tumor subtype discrimination". In: *arXiv preprint arXiv:1704.05122* (2017).

[21] Omar S Al-Kadi. "Fractals for biomedical texture analysis". In: *Biomedical texture analysis*. Elsevier, 2017, pp. 131–161.

[22] Omar S Al-Kadi and Antonio Di Ieva. "Fractal-Based Analysis of Histological Features of Brain Tumors". In: *The Fractal Geometry of the Brain*. Springer, 2024, pp. 501–524.

[23] Omar S Al-Kadi and Oliver Diaz. "Reviews in cancer imaging and image-directed interventions". In: *Frontiers in Oncology* 13 (2023), p. 1183302.

[24] Omar S Al-Kadi, D Watson, et al. "Texture analysis of aggressive and nonaggressive lung tumor CE CT images". In: *IEEE transactions on biomedical engineering* 55.7 (2008), pp. 1822–1830.

[25] Omar S. Al-Kadi et al. "Empowering brain cancer diagnosis: harnessing artificial intelligence for advanced imaging insights". In: *Reviews in the Neurosciences* 35.4 (2024), pp. 399–419.

[26] Omar Sultan Al-Kadi. "Tumour grading and discrimination based on class assignment and quantitative texture analysis techniques". PhD thesis. University of Sussex, 2010.

[27] Taranjit Kaur and Tapan Kumar Gandhi. "Deep convolutional neural networks with transfer learning for automated brain image classification". In: *Machine Vision and Applications* 31.3 (2020), p. 20.

[28] Mohammad Monirujjaman Khan et al. "A novel approach to predict brain cancerous tumor using transfer learning". In: *Computational and Mathematical Methods in Medicine* 2022 (2022).

[29] Zhongyuan Lyu, Dong Xia, and Yuan Zhang. "Latent space model for higher-order networks and generalized tensor decomposition". In: *Journal of Computational and Graphical Statistics* (2023), pp. 1–17.

[30] Vignesh V Menon et al. "Video Quality Assessment with Texture Information Fusion for Streaming Applications". In: *arXiv preprint arXiv:2302.14465* (2023).

[31] Mario Milicevic et al. "Data augmentation and transfer learning for limited dataset ship classification". In: *WSEAS Trans. Syst. Control* 13.1 (2018), pp. 460–465.

[32] Dong Nie et al. "Erratum: Medical Image Synthesis with Deep Convolutional Adversarial Networks (IEEE Trans. Biomed. Eng.(2018) 65: 12 (2720-2730". In: *IEEE Transactions on Biomedical Engineering* 67.9 (2020), p. 2706.

[33] Dong Nie et al. "Medical image synthesis with context-aware generative adversarial networks". In: *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*. Springer. 2017, pp. 417–425.

[34] Savas Ozkan, Mete Ozay, and Tom Robinson. "Conceptual and Hierarchical Latent Space Decomposition for Face Editing". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 7248–7257.

[35] Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning". In: *IEEE Transactions on Knowledge and data engineering* 22.10 (2010), pp. 1345–1359.

[36] Arshia Rehman et al. "A deep learning-based framework for automatic brain tumors classification using transfer learning". In: *Circuits, Systems, and Signal Processing* 39 (2020), pp. 757–775.

[37] Nicolae-clin Ristea et al. "CyTran: A Cycle-Consistent Transformer with Multi-Level Consistency for Non-Contrast to Contrast CT Translation". In: *Neurocomputing* (2023).

[38] Snehashis Roy et al. "MR to CT registration of brains using image synthesis". In: *Medical Imaging 2014: Image Processing*. Vol. 9034. SPIE. 2014, pp. 307–314.

[39] Veit Sandfort et al. "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks". In: *Scientific reports* 9.1 (2019), p. 16884.

[40] Priyansh Saxena, Akshat Maheshwari, and Saumil Maheshwari. "Predictive modeling of brain tumor: a deep learning approach". In: *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020*. Springer, 2020, pp. 275–285.

[41] De Rosal Igantius Moses Setiadi. "PSNR vs SSIM: imperceptibility quality assessment for image steganography". In: *Multimedia Tools and Applications* 80.6 (2021), pp. 8423–8444.

[42] Onur Sevli. "Performance comparison of different pretrained deep learning models in classifying brain MRI images". In: *Acta Infologica* 5.1 (2021), pp. 141–154.

[43] Hamid R Sheikh and Alan C Bovik. "A visual information fidelity approach to video quality assessment". In: *The first international workshop on video processing and quality metrics for consumer electronics*. Vol. 7. 2. sn. 2005, pp. 2117–2128.

[44] Zar Nawab Khan Swati et al. "Brain tumor classification for MR images using transfer learning and fine-tuning". In: *Computerized Medical Imaging and Graphics* 75 (2019), pp. 34–46.

[45] Zar Nawab Khan Swati et al. "Brain tumor classification for MR images using transfer learning and fine-tuning". In: *Computerized Medical Imaging and Graphics* 75 (2019), pp. 34–46.

[46] Gopal S Tandel, Ashish Tiwari, and OG Kakde. "Performance enhancement of MRI-based brain tumor classification using suitable segmentation method and deep learning-based ensemble algorithm". In: *Biomedical Signal Processing and Control* 78 (2022), p. 104018.

[47] Kasi Tenghongsakul et al. "Deep transfer learning for brain tumor detection based on MRI images". In: *Journal of Physics: Conference Series*. Vol. 2497. 1. IOP Publishing. 2023, p. 012015.

[48] Dmitrii Torbunov et al. "Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023, pp. 702–712.

[49] Radim Tyleček and Radim Šára. "Spatial pattern templates for recognition of objects with regular structure". In: *Pattern Recognition: 35th German Conference, GCPR 2013, Saarbrücken, Germany, September 3-6, 2013. Proceedings 35*. Springer. 2013, pp. 364–374.

[50] Juan Miguel Valverde et al. "Transfer learning in magnetic resonance brain imaging: A systematic review". In: *Journal of imaging* 7.4 (2021), p. 66.

[51] Jiayuan Wang, QM Jonathan Wu, and Farhad Pourpanah. "Dc-cyclegan: bidirectional ct-to-mr synthesis from unpaired data". In: *Computerized Medical Imaging and Graphics* (2023), p. 102249.

[52] Zhou Wang and Alan C Bovik. "A universal image quality index". In: *IEEE signal processing letters* 9.3 (2002), pp. 81–84.

[53] Zirui Wang et al. "Characterizing and avoiding negative transfer". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 11293–11302.

[54] Chen Henry Wu and Fernando De la Torre. "A Latent Space of Stochastic Diffusion Models for Zero-Shot Image Editing and Guidance". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 7378–7387.

[55] Yang Yang et al. "Glioma grading on conventional MR images: a deep learning study with transfer learning". In: *Frontiers in neuroscience* 12 (2018), p. 804.

[56] Haihang Zhang et al. "Seismic impedance inversion based on geophysical-guided cycle-consistent generative adversarial networks". In: *Journal of Petroleum Science and Engineering* 218 (2022), p. 111003.

[57] Jing Zhang et al. "Unsupervised conditional consensus adversarial network for brain disease identification with structural MRI". In: *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*. Springer. 2019, pp. 391–399.

[58] Zizhao Zhang, Lin Yang, and Yefeng Zheng. "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network". In: *Proceedings of the IEEE conference on computer vision and Pattern Recognition*. 2018, pp. 9242–9251.

[59] Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

VII. APPENDIX

A. Perceptual study and validation details

In our work, we devised a questionnaire to gather feedback on the accuracy and perceptions of CT-MR translation in real-world clinical settings. Our goal is to improve the quality of CT-MR translation by identifying any existing issues or limitations. Additionally, we aim to assess healthcare professionals' satisfaction with the translation output and its practical utility. Radiologists participating in the study are tasked with identifying ground truth images and using a four-point scoring system to evaluate image accuracy, with '4' indicating strong agreement on realism (4 strongly agree, 3 somewhat agree, 2 somewhat disagree, 1 disagree).

Figure 10 shows an example of a form completed by a radiologist for MR-CT images, while Figure 11 presents a similar form for CT-MR images. In these evaluations, the radiologists used a degree scale to rate the quality or diagnostic clarity of the images. If the degree given by the radiologist is above 2, the result is denoted as "Yes." Conversely, if the degree is 2 or below, the result is marked as "No." This binary classification simplifies the comparison of the imaging modalities, allowing for a straightforward assessment of their relative effectiveness based on the radiologists' ratings.

image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N
1	2	N	31	3	Y	61	4	Y	91	3	Y	121	4	Y	151	4	Y	181	4	Y	211	3	Y
2	2	N	32	4	Y	62	4	Y	92	3	Y	122	4	Y	152	4	Y	182	3	Y	212	3	Y
3	2	N	33	3	Y	63	4	Y	93	3	Y	123	4	Y	153	4	Y	183	4	Y	213	3	Y
4	3	Y	34	4	Y	64	4	Y	94	3	Y	124	4	Y	154	4	Y	184	4	Y	214	3	Y
5	4	Y	35	4	Y	65	2	N	95	3	Y	125	4	Y	155	3	Y	185	4	Y	215	4	Y
6	4	Y	36	4	Y	66	2	N	96	4	Y	126	4	Y	156	3	Y	186	4	Y	216	4	Y
7	4	Y	37	4	Y	67	3	Y	97	4	Y	127	4	Y	157	3	Y	187	3	Y	217	4	Y
8	4	Y	38	4	Y	68	3	Y	98	4	Y	128	4	Y	158	4	Y	188	4	Y	218	4	Y
9	3	Y	39	2	N	69	3	Y	99	4	Y	129	4	Y	159	4	Y	189	3	Y	219	4	Y
10	3	Y	40	2	N	70	4	Y	100	4	Y	130	4	Y	160	4	Y	190	4	Y	220	4	Y
11	3	Y	41	3	Y	71	3	Y	101	3	Y	131	4	Y	161	4	Y	191	4	Y	221	3	Y
12	4	Y	42	2	N	72	3	Y	102	4	Y	132	4	Y	162	3	Y	192	4	Y	222	4	Y
13	4	Y	43	3	Y	73	3	Y	103	4	Y	133	4	Y	163	4	Y	193	4	Y	223	4	Y
14	4	Y	44	3	Y	74	3	Y	104	4	Y	134	3	Y	164	3	Y	194	4	Y	224	4	Y
15	4	Y	45	2	N	75	4	Y	105	3	Y	135	3	Y	165	4	Y	195	4	Y	225	4	Y
16	4	Y	46	3	Y	76	4	Y	106	4	Y	136	3	Y	166	4	Y	196	3	Y	226	2	N
17	4	Y	47	4	Y	77	3	Y	107	3	Y	137	4	Y	167	4	Y	197	3	Y	227	2	N
18	4	Y	48	4	Y	78	4	Y	108	4	Y	138	4	Y	168	4	Y	198	4	Y	228	3	Y
19	4	Y	49	3	Y	79	3	Y	109	4	Y	139	4	Y	169	4	Y	199	4	Y	229	3	Y
20	4	Y	50	2	N	80	3	Y	110	4	Y	140	4	Y	170	4	Y	200	4	Y	230	3	Y
21	4	Y	51	2	N	81	4	Y	111	4	Y	141	4	Y	171	4	Y	201	3	Y	231	3	Y
22	3	Y	52	2	N	82	4	Y	112	4	Y	142	4	Y	172	3	Y	202	4	Y	232	3	Y
23	3	Y	53	4	Y	83	2	N	113	3	Y	143	4	Y	173	3	Y	203	3	Y	233	3	Y
24	3	Y	54	2	N	84	4	Y	114	3	Y	144	4	Y	174	4	Y	204	3	Y	234	3	Y
25	3	Y	55	2	N	85	3	Y	115	4	Y	145	3	Y	175	4	Y	205	3	Y	235	3	Y
26	4	Y	56	2	N	86	3	Y	116	4	Y	146	3	Y	176	4	Y	206	3	Y	236	4	Y
27	4	Y	57	3	Y	87	3	Y	117	4	Y	147	3	Y	177	4	Y	207	3	Y	237	3	Y
28	4	Y	58	3	Y	88	4	Y	118	4	Y	148	3	Y	178	4	Y	208	4	Y	238	3	Y
29	4	Y	59	2	N	89	4	Y	119	4	Y	149	4	Y	179	4	Y	209	4	Y	239	3	Y
30	4	Y	60	4	Y	90	3	Y	120	4	Y	150	4	Y	180	4	Y	210	3	Y	240	3	Y

Figure 10: Completed form by one radiologist for MR_CT images.

image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N	image	points	Y/N
1	3	Y	31	3	Y	61	4	Y	91	3	Y	121	3	Y	151	3	Y	181	4	Y	211	3	Y
2	2	N	32	3	Y	62	4	Y	92	2	N	122	2	N	152	3	Y	182	4	Y	212	3	Y
3	3	Y	33	3	Y	63	3	Y	93	2	N	123	3	Y	153	3	Y	183	4	Y	213	3	Y
4	3	Y	34	2	N	64	3	Y	94	3	Y	124	4	Y	154	4	Y	184	3	Y	214	3	Y
5	2	N	35	3	Y	65	2	N	95	3	Y	125	4	Y	155	4	Y	185	3	Y	215	4	Y
6	3	Y	36	3	Y	66	4	Y	96	3	Y	126	4	Y	156	4	Y	186	4	Y	216	4	Y
7	4	Y	37	3	Y	67	3	Y	97	3	Y	127	4	Y	157	3	Y	187	3	Y	217	3	Y
8	3	Y	38	3	Y	68	2	N	98	3	Y	128	4	Y	158	4	Y	188	4	Y	218	3	Y
9	3	Y	39	3	Y	69	4	Y	99	3	Y	129	4	Y	159	3	Y	189	4	Y	219	3	Y
10	2	N	40	3	Y	70	4	Y	100	4	Y	130	3	Y	160	3	Y	190	4	Y	220	4	Y
11	3	Y	41	3	Y	71	4	Y	101	4	Y	131	3	Y	161	3	Y	191	4	Y	221	4	Y
12	3	Y	42	4	Y	72	3	Y	102	4	Y	132	4	Y	162	4	Y	192	3	Y	222	4	Y
13	3	Y	43	4	Y	73	4	Y	103	4	Y	133	4	Y	163	4	Y	193	4	Y	223	4	Y
14	2	N	44	3	Y	74	3	Y	104	2	N	134	4	Y	164	4	Y	194	3	Y	224	3	Y
15	4	Y	45	3	Y	75	2	N	105	3	Y	135	4	Y	165	4	Y	195	3	Y	225	3	Y
16	3	Y	46	3	Y	76	2	N	106	3	Y	136	4	Y	166	3	Y	196	3	Y	226	3	Y
17	3	Y	47	4	Y	77	3	Y	107	4	Y	137	3	Y	167	3	Y	197	4	Y	227	3	Y
18	3	Y	48	4	Y	78	3	Y	108	4	Y	138	3	Y	168	4	Y	198	3	Y	228	3	Y
19	2	N	49	4	Y	79	4	Y	109	4	Y	139	3	Y	169	4	Y	199	3	Y	229	4	Y
20	4	Y	50	4	Y	80	3	Y	110	4	Y	140	4	Y	170	4	Y	200	3	Y	230	3	Y
21	2	N	51	4	Y	81	4	Y	111	4	Y	141	3	Y	171	4	Y	201	3	Y	231	4	Y
22	2	N	52	4	Y	82	4	Y	112	4	Y	142	4	Y	172	3	Y	202	4	Y	232	3	Y
23	4	Y	53	4	Y	83	4	Y	113	4	Y	143	4	Y	173	3	Y	203	4	Y	233	3	Y
24	2	N	54	4	Y	84	4	Y	114	3	Y	144	4	Y	174	4	Y	204	4	Y	234	3	Y
25	4	Y	55	4	Y	85	4	Y	115	3	Y	145	4	Y	175	4	Y	205	4	Y	235	4	Y
26	3	Y	56	3	Y	86	3	Y	116	4	Y	146	4	Y	176	3	Y	206	4	Y	236	4	Y
27	3	Y	57	3	Y	87	3	Y	117	4	Y	147	4	Y	177	4	Y	207	4	Y	237	4	Y
28	3	Y	58	2	N	88	3	Y	118	3	Y	148	3	Y	178	3	Y	208	3	Y	238	2	N
29	3	Y	59	4	Y	89	3	Y	119	2	N	149	3	Y	179	4	Y	209	3	Y	239	3	Y
30	3	Y	60	4	Y	90	2	N	120	3	Y	150	3	Y	180	4	Y	210	3	Y	240	4	Y

Figure 11: Completed form by one radiologist for CT-MR images.