DOI: 10.21917/ijsc.2025.0504

DYNAMIC CHANNEL-ALIGNED 3D U-NET FOR EXPLAINABLE T2-WEIGHTED INFANT BRAIN MRI SYNTHESIS FROM T1-WEIGHTED MRI

Param Ahir[1] and Mehul Parikh[2]

[1]*Department of Computer and Information Technology, Gujarat Technological University, India,*
[2]*Department of Information Technology, L. D. College of Engineering, India*

Abstract

Magnetic Resonance Imaging (MRI) is a crucial tool in clinical diagnostics, with T1-weighted (T1) and T2-weighted (T2). Acquiring high-quality T2-weighted MRI, especially for infant brains, presents challenges due to lengthy acquisition times, motion artifacts, and scanner variability. This study introduces the Adaptive Dual Domain U-Net, a novel 3D U-Net architecture enhanced with dynamic channel alignment for synthesizing T2-weighted MRI from T1-weighted inputs. The proposed model addresses domain variability, integrates explainability tools using Captum, and employs patch-based training for efficient memory utilization and high-resolution reconstruction. Quantitative evaluations on the iSeg-2019 dataset demonstrate superior performance across key metrics such as Mean Squared Error (MSE), Structural Similarity Index (SSIM), and R^2 compared to baseline methods. Qualitative results highlight the model's ability to generate structurally accurate and clinically interpretable synthetic T2-weighted images, making it a robust tool for both clinical and research applications.

Keywords:

Magnetic Resonance Imaging, Deep Learning, Medical Imaging, Cross-Modality MRI Synthesis, Infant Brain MRI, 3D U-Net

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique widely used in clinical practice to capture detailed anatomical and physiological information. Among various modalities, T1-weighted (T1) and T2-weighted (T2) MRI scans are primarily used, with T1 highlighting structural details and T2 emphasizing fluid-filled regions, making them essential for diagnosing various conditions. Acquiring high-quality T2-weighted images for infant brain MRI, is often challenging due to long acquisition times, increased motion artifacts, and scanner variability.

In infant brain imaging, there are more challenges due to rapid developmental changes, limited annotated datasets, and ethical constraints on extensive data collection. Due to these factors, there is a need for robust methods to generate synthetic T2-weighted images from readily available T1-weighted data. Synthetic MRI generation reduces acquisition time and costs and also enhances dataset diversity. Deep learning-based approaches, such as Convolutional Neural Networks (CNNs) [1] and Generative Adversarial Networks (GANs) [2], are used for medical image synthesis. However, existing methods has drawbacks such as limited adaptability across datasets, domain variability, and a lack of explainability, which are essential requirement for clinical adoption. Particularly in infant brain imaging, the high variability in brain morphology and scanner settings requires architectures that can address these issues.

To address these challenges, we propose a novel approach for T2-weighted infant brain MRI synthesis using a dynamically aligned 3D U-Net [3]. Our architecture uses dynamic channel alignment in the decoder, ensuring seamless feature fusion between the encoder and decoder. This mechanism addresses domain variability, making the model adaptable to diverse imaging datasets. Additionally, our framework integrates patch-based processing with dynamic patch size adjustment for efficient training and reconstruction of high-resolution 3D MRI volumes. By utilizing explainability tools such as Integrated Gradients and Noise Tunnel, our approach highlights the regions in T1-weighted images that contribute most significantly to the generation of corresponding T2-weighted features. This integration of explainability add trust and clinical relevance for the proposed method.

Our contributions can be summarized as follows:

- We introduce a novel dynamic channel-aligned 3D U-Net that addresses domain variability and enhances feature fusion in T2-weighted MRI synthesis.
- We develop a patch-based processing framework with adaptive patch size and stride, allowing for memory-efficient training and robust reconstruction of high-resolution volumes.
- We integrate explainability tools into the pipeline, providing insights into the model's predictions and facilitating clinical trust.
- We demonstrate the effectiveness of our method on the publicly available iSeg-2019 dataset, achieving superior performance in terms of quantitative metrics (MSE, R^2, MAE) and qualitative comparisons with baseline approaches.

The remainder of this paper is organized as follows: Section 2 reviews related work in MRI synthesis and explainable AI. Section 3 details the methodology, including dataset preparation, model architecture, and training pipeline. Section 4 presents experimental results, followed by a discussion of key findings and limitations in section 5. Section 6 concludes with potential future directions.

2. RELATED WORK

2.1 CROSS-MODALITY MRI SYNTHESIS

Cross-modality MRI synthesis has become a prominent area of research in recent years, due to advancements in deep learning and generative modelling. Generating one MRI modality (e.g., T1-weighted) from another (e.g., T2-weighted) is helpful as it reduces scanning time, lower costs, and enhances diagnostic utility. Earlier methods for cross-modality MRI synthesis were

focused on atlas-based registration techniques and patch-based synthesis. For example, [4] proposed a patch-based approach that compared patches from the source modality to an atlas of paired modalities to generate the target modality. These methods could not generate complex anatomical variations and often required time-consuming preprocessing steps.

In cross-modality synthesis deep learning methods like convolutional neural networks (CNNs) and generative adversarial networks (GANs) were used earlier. [5] introduced a 3D CNN for synthesizing computed tomography (CT) scan from T1-weighted images. GAN-based approaches, such as [6], further enhanced synthesis quality by adversarial training to produce sharper and more realistic MRI. Cycle-consistent GANs (CycleGANs) have also been used for this task extensively. These methods use cycle-consistency loss, to ensure the synthesized image closely resembles the original input. [7] demonstrated the utility of CycleGANs for multi-site, cross-protocol MRI synthesis, achieving robust results across varied datasets.

Recent studies have extended cross-modality synthesis to multi-modal and multi-domain settings. [8] proposed a unified framework that simultaneously synthesizes multiple modalities using a shared latent space representation. These approaches address inter-modality dependencies more effectively but often require extensive training data to generalize well across different scanners and acquisition protocols. More recently, transformer-based architectures are used for cross-modality synthesis to capture long-range dependencies. For example, [9] introduced a vision transformer (ViT)-based approach that outperformed traditional CNNs and GANs in generating anatomically consistent images.

There are certain challenges in this field. First, existing models often require large datasets, that are scarce in medical imaging. Second, domain variability such as differences in scanner types, acquisition settings, and patient demographics adversely impacts the generalizability of these models. Third, most deep learning models operate as black boxes, providing limited interpretability, which is crucial for clinical adoption. In this study, we address these gaps by proposing a robust, explainable cross-modality synthesis framework. Our method leverages domain adaptation techniques to enhance generalizability across varied acquisition protocols. Additionally, we incorporate explainability measures to ensure transparency so, clinicians can better trust and interpret the synthesized images. Our work aims to bridge the gap between technical advancements and practical clinical applications in MRI synthesis.

2.2 EXPLAINABILITY IN DEEP LEARNING FOR MEDICAL IMAGING

Explainability in deep learning for medical imaging has become an increasingly important area of research, particularly in applications such as disease diagnosis, treatment planning, and outcome prediction. As deep learning models becomes more complex, the lack of transparency in their decision-making processes create a significant challenge for clinical adoption. Explainability methods in medical imaging can be broadly categorized into post hoc and intrinsic techniques. Post hoc methods are applied after model training to interpret predictions, such as saliency maps, Grad-CAM [10], and SHAP [11], which highlight regions in medical images that contribute most to the model's predictions. For example, Grad-CAM has been used extensively to visualize tumour regions in MRI scans, providing clinicians with insights into the model's focus during classification or segmentation tasks. Intrinsic explainability involves designing models with built-in interpretability features. This includes attention mechanisms, self-explanatory networks, and interpretable feature embeddings. Attention-based models [12] enhance model performance and provide spatial and temporal context about how predictions are made. Another area of interest is model disentanglement [13], that is used to separate high-level features such as anatomical structures from domain-specific variations like imaging noise or scanner bias. This disentanglement enhances interpretability and robustness for clinicians to better understand model behaviour under different conditions.

However, Saliency methods often lack consistency and are sensitive to noise, while attention mechanisms can sometimes introduce biases or fail to capture subtle variations in medical images. Furthermore, achieving a balance between model performance and interpretability continues to be a critical area of exploration. In this work, we incorporate explainability measures into our cross-modality MRI synthesis framework to ensure transparency and trustworthiness.

3. METHODOLOGY

The Proposed approach synthesize T2-weighted infant brain MRI from T1-weighted inputs using the Adaptive-Dual-Domain U-Net.

3.1 DATA PREPROCESSING

The iSeg-2019 dataset [14] contains paired T1- and T2-weighted MRI scans and it was used for training and validation. Each 3D MRI volume was pre-processed to ensure consistency. Voxel intensities were normalized to the range [0,1] using:

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

where, X represents the raw voxel intensities. Rigid registration was applied to align the spatial coordinates of T1 and T2 volumes. All images were resampled to an isotropic resolution of 1 mm^3. To handle the large 3D volumes, each MRI was divided into overlapping patches of size 64×64×64, with a stride of 32×32×32. Dynamic adjustments to the patch size ensured that smaller volumes were fully covered. These patches served as inputs to the network during training and validation.

3.2 ADAPTIVE DUAL DOMAIN U-NET ARCHITECTURE

The Adaptive Dual Domain U-Net network consists of three main components: an encoder, a bottleneck, and a decoder, connected through skip connections. A novel feature of the network is the dynamic channel alignment layer, which addresses channel mismatches between the encoder and decoder. The encoder processes the input T1-weighted MRI, progressively reducing spatial dimensions while increasing feature depth. Each encoder block comprises two 3D convolutional layers with ReLU activation, followed by batch normalization. Down sampling is

achieved through max-pooling. The output of the encoder at layer l is represented as:

$$E_L = f_{enc}^{l}\left(E_{L-1}\right), E_0 = X \tag{2}$$

where, E_l is the output of the l^{th} encoder block, and f_{enc}^{l} denotes its operations. The encoder outputs have progressively smaller spatial dimensions, culminating in a size of 16×16×16 with 128 channels. The bottleneck lies at the centre of the network and operates at the smallest spatial scale. It captures high-level contextual features through additional 3D convolutional layers:

$$B = f_{bottle}\left(E_L\right) \tag{3}$$

where, L denotes the number of encoder layers, and B has dimensions 8×8×8 with 256 channels. The decoder reconstructs the spatial dimensions and synthesizes the T2-weighted image from the bottleneck features. Each decoder block includes an up-sampling operation followed by two 3D convolutional layers with ReLU activation. Skip connections between corresponding encoder and decoder layers preserve spatial details lost during down-sampling. Due to mismatched feature dimensions, dynamic channel alignment layers are applied before concatenation. These alignment layers use 1×1×1 convolutions to harmonize the feature dimensions:

$$D_1 = \left(f_{dec}^{l}\left(f_{align}^{l}U_1 + E_1\right)\right) \tag{4}$$

where, U_{l+1} is the up-sampled output from the previous decoder layer, E_l is the encoder output at layer l, and f_{align}^{l} represents the alignment operation. The final output is generated using a 1×1×1 convolution:

$$Y = f_{out}\left(D_1\right) \tag{5}$$

where, Y is the synthesized T2-weighted MRI.

The Adaptive Dual Domain U-Net architecture is visualized in the Fig.1, which highlights the encoder, bottleneck, and decoder components, as well as the skip connections that link corresponding layers. The encoder is represented on the left, the decoder on the right, and the bottleneck at the centre, with clearly separated sections to emphasize their distinct roles.

3.3 PATCH-BASED TRAINING AND RECONSTRUCTION

Training involves feeding patches of T1-weighted MRI into the network to predict the corresponding patches of T2-weighted MRI. The network is trained to minimize voxel-wise intensity differences between the ground truth and predicted T2-weighted patches, using the Mean Squared Error (MSE) loss:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i + Y_i'\right) \tag{6}$$

In Eq.(6), N is the total number of voxels in a patch, Y_i is the ground truth intensity, and Y_i' is the predicted intensity.

During validation, the full 3D T2-weighted MRI is reconstructed by combining the predicted patches. Overlapping regions are averaged to ensure smooth transitions, using:

$$Y\left(Z:Z+p_d', y:y+p_h, X:X+p_w\right) += p_t \tag{7}$$

$$Y\left(z,y,x\right) = \frac{Y\left(z,y,x\right)}{count\left(z,y,x\right)} \tag{8}$$

In Eq.(7) and Eq.(8), p_i is the predicted patch, and count(z,y,x) tracks the number of overlapping patches contributing to each voxel.

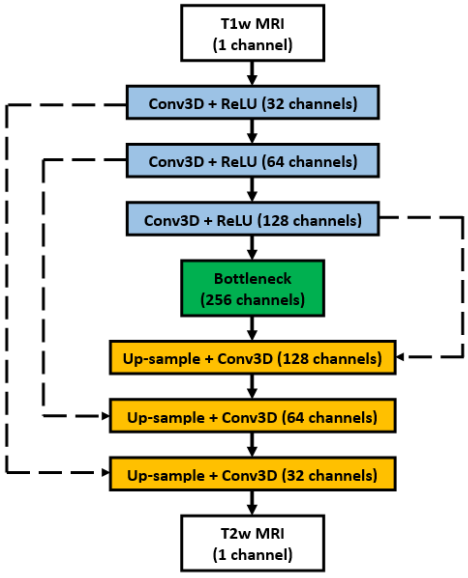

Fig.1 Adaptive Dual Domain U-Net architecture

3.4 EXPLAINABILITY INTEGRATION

Explainability tools were integrated into the framework to provide voxel-wise attributions, identifying regions in the input T1-weighted MRI that are most important for the predicted T2-weighted output. Using Integrated Gradients from the Captum library, the attribution for each voxel iii is computed as:

$$Attr_i = \left(X_i - X_{0,i}\right)\int_{\alpha=0}^{1}\frac{\partial y}{\partial x}\Big|_{X_0+\alpha\left(X+X_0\right)^d\alpha} \tag{9}$$

where, X_0 is a baseline input, such as a zero-filled image, and X is the input T1-weighted MRI. These attributions highlight the regions in the input that are most critical for generating the T2-weighted output, enhancing the interpretability of the network.

4. EXPERIMENTS AND RESULTS

4.1 EXPERIMENTAL SETUP

The experiments were conducted on a system with an NVIDIA Tesla V100 GPU with 32GB of memory, 128GB of system RAM, and an Intel Xeon processor. The training and inference pipelines were implemented in PyTorch. The network was trained with a batch size of 10 and an initial learning rate of 0.001, optimized using the Adam optimizer. The learning rate was reduced using a scheduler based on validation loss. Training was conducted for 60 epochs. Validation was performed after every epoch to monitor model performance, using key metrics such as Loss, Mean Squared Error (MSE), R^2 Score, and Mean Absolute Error (MAE).

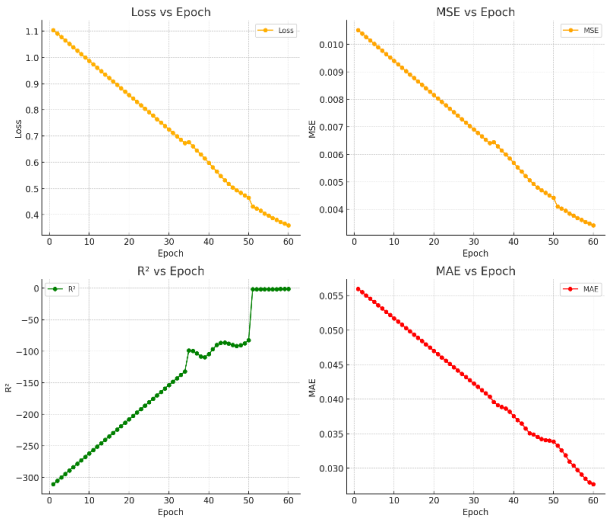

Fig.2. Evaluation of Loss and Error Metrics Across Epochs

The Fig.2 show the model's performance for 60 epochs with metrics like Loss, Mean Squared Error (MSE), R^2 Score, and Mean Absolute Error (MAE).

4.2 QUANTITATIVE RESULTS

The performance of the Adaptive Dual Domain U-Net was evaluated against certain baseline mode as given in table.1for image synthesis. To assess the model performance Structural Similarity Index (SSIM)[15] for structural similarity between synthesized and ground truth images assessment and Peak Signal-to-Noise Ratio (PSNR)[16 to evaluates image quality was used.

The Table.1 below summarizes the results across these metrics for the test set:

Table.1 Model Performance comparison with state-of-art models.

Model	SSIM (↑)	PSNR (↑)
cGAN [17]	0.861	27.535
pix2pix [18]	0.821	26.155
MedGAN [19]	0.805	25.616
pGAN [20]	0.846	27.396
Our Model	0.873	28.557

The Adaptive Dual Domain U-Net consistently outperformed the baselines across all metrics, demonstrating its ability to synthesize high-quality T2-weighted MRI with minimal intensity and structural errors.

4.3 QUALITATIVE RESULTS

Visual comparisons between synthesized T2-weighted images and ground truth T2-weighted MRI highlight the effectiveness of the proposed model. As shown in the Fig.3, the Adaptive Dual Domain U-Net produces outputs that closely match the ground truth in terms of structural integrity and intensity distribution. As shown in Fig.3 certain synthetic MRI provides the structure well but lacks the intensity correction compared to ground truth, in future we will try to work on this.

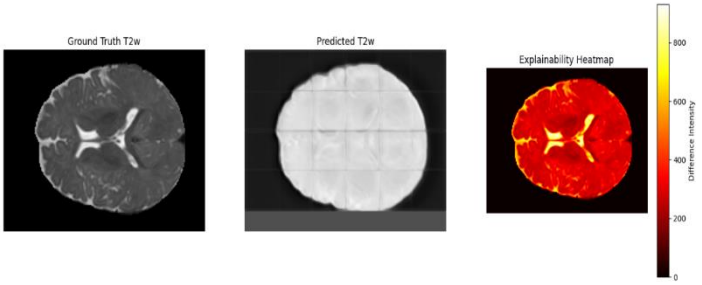

Fig.3. Synthetic MRI with Less accuracy

Additionally, the integration of explainability techniques provides insights into the model's decision-making process. Heatmaps generated using Integrated Gradients reveal regions in the input T1-weighted MRI that are most influential in synthesizing the T2-weighted output. These visualizations demonstrate the network's ability to focus on anatomically relevant regions, such as ventricles and cortical areas.

4.4 ABLATION STUDIES

To understand the contribution of key components in the Adaptive Dual Domain U-Net, ablation studies were conducted. The effect of dynamic channel alignment was analysed by removing the alignment layers and comparing the performance. Without alignment, MSE increased from 0.014 to 0.020, and SSIM dropped from 0.87 to 0.82, highlighting the importance of this component in ensuring smooth feature integration between encoder and decoder. The impact of patch size and stride was studied by varying the patch size and stride during training. Smaller patch sizes (32×32×32) improved local structural details but caused artifacts in reconstructed images due to reduced context. Larger patch sizes (128×128×128) provided global context but resulted in memory constraints and slower convergence. The default patch size of 64×64×64 achieved the best balance between quality and efficiency. The integration of explainability tools was also evaluated. While it does not directly impact synthesis performance, the interpretability provided by heatmaps enhances the usability of the model in clinical and research contexts.

5. CONCLUSION

This work provides an Adaptive Dual Domain U-Net, explainable deep learning framework for synthesizing T2-weighted MRI from T1-weighted inputs. The proposed model effectively addresses challenges associated with domain variability and high-resolution image synthesis by incorporating dynamic channel alignment and patch-based training. Results validate its performance compared to baseline models with SSIM score of 0.873 and PSNR value of 28.557dB. Despite its strengths, limitations such as performance under severe motion artifacts or low-contrast inputs indicate there is a need for future research on domain adaptation and dataset diversity.

REFERENCES

[1] A.F. Osman and N.M. Tamam, "Deep Learning-based Convolutional Neural Network for Intramodality Brain MRI

Synthesis", *Journal of Applied Clinical Medical Physics*, Vol. 23, No. 4, pp. 1-6, 2022.

[2] C.K. Chong and E.T.W. Ho, "Synthesis of 3D MRI Brain Images with Shape and Texture Generative Adversarial Deep Neural Networks", *IEEE Access*, Vol. 9, pp. 64747-64760, 2021.

[3] O. Cicek, A. Abdulkadir, S.S. Lienkamp, T. Brox and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation", *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424-432, 2016.

[4] F. Rousseau, P. Habas, C. Studholme, "A Supervised Patch-based Approach for Human Brain Labeling", *IEEE Transactions on Medical Imaging*, Vol. 30, No. 10, pp. 1852-1862, 2011.

[5] X. Han, "MR-based Synthetic CT Generation using a Deep Convolutional Neural Network Method", *Medical Physics*, Vol. 44, No. 4, pp. 1408-1419, 2017.

[6] A. Chartsias, T. Joyce, M.V. Giuffrida and S.A. Tsaftaris, "Multimodal MR Synthesis Via Modality-Invariant Latent Representation", *IEEE Transactions on Medical Imaging*, Vol. 37, No. 3, pp. 803-814, 2017.

[7] K. Wu, Y. Qiang, K. Song, X. Ren, W. Yang, W. Zhang and Y. Cui, "Image Synthesis in Contrast MRI based on Super Resolution Reconstruction with Multi-Refinement Cycle-Consistent Generative Adversarial Networks", *Journal of Intelligent Manufacturing*, Vol. 31, pp. 1215-1228, 2020.

[8] T. Zhou, H. Fu, G. Chen, J. Shen and L. Shao, "Hi-Net: Hybrid-Fusion Network for Multi-Modal MR Image Synthesis", *IEEE Transactions on Medical Imaging*, Vol. 39, No. 9, pp. 2772-2781, 2020.

[9] P. Zhang, X. Dai, J. Yang, B. Xiao, L. Yuan, L. Zhang and J. Gao, "Multi-Scale Vision Longformer: A New Vision Transformer for High-resolution Image Encoding", *Proceedings of International Conference on Computer Vision*, pp. 2998-3008, 2021.

[10] M. Xiao, L. Zhang, W. Shi, J. Liu, W. He and Z. Jiang, "A Visualization Method based on the Grad-CAM for Medical Image Segmentation Model", *Proceedings of International Conference on Electronic Information Engineering and Computer Science*, pp. 242-247, 2021.

[11] U.E. Husby, "Exploring Breast Cancer Diagnosis: A Study of SHAP and LIME in XAI-Driven Medical Imaging", *Master's Thesis, Department of Science and Technology, Norwegian University of Life Sciences*, pp. 1-59, 2024.

[12] P. Chen, W. Dong, J. Wang, X. Lu, U. Kaymak and Z. Huang, "Interpretable Clinical Prediction Via Attention-based Neural Network", *BMC Medical Informatics and Decision Making*, Vol. 20, pp. 1-9, 2020.

[13] X. Liu, P. Sanchez, S. Thermos, A.Q. O'Neil and S.A. Tsaftaris, "Learning Disentangled Representations in the Imaging Domain", *Medical Image Analysis*, Vol. 80, pp. 1-6, 2022.

[14] Y. Sun, K. Gao, Z. Wu, G. Li, X. Zong, Z. Lei and L. Wang, "Multi-Site Infant Brain Segmentation Algorithms: The iSeg-2019 Challenge", *IEEE Transactions on Medical Imaging*, Vol. 40, No. 5, pp. 1363-1376, 2021.

[15] V. Mudeng, M. Kim and S.W. Choe, "Prospects of Structural Similarity Index for Medical Image Analysis", *Applied Sciences*, Vol. 12, No. 8, pp. 1-7, 2022.

[16] Y. Tanabe and T. Ishida, "Quantification of the Accuracy Limits of Image Registration using Peak Signal-to-Noise Ratio", *Radiological Physics and Technology*, Vol. 10, pp. 91-94, 2017.

[17] R. Graf, P.S. Platzek, E.O. Riedel, S.H. Kim, N. Lenhart, C. Ramschütz and J.S. Kirschke, "Generating Synthetic High-Resolution Spinal STIR and T1w Images from T2w FSE and Low-Resolution Axial Dixon", *European Radiology*, pp. 1-11, 2014.

[18] J. Yang, X.X. Li, F. Liu, D. Nie, P. Lio, H. Qi and D. Shen, "Fast T2w/FLAIR MRI Acquisition by Optimal Sampling of Information Complementary to Pre-Acquired T1w MRI", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-6, 2021.

[19] K. Pan, P. Cheng, Z. Huang, L. Lin and X. Tang, "Transformer-based T2-Weighted MRI Synthesis from T1-Weighted Images", *Proceedings of International Conference on Engineering in Medicine and Biology Society*, pp. 5062-5065, 2022.

[20] P. Isola, J.Y. Zhu, T. Zhou and A.A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1125-1134, 2017.

[21] K. Armanious, C. Jiang, M. Fischer, T. Kustner, T. Hepp, K. Nikolaou, S. Gatidis and B. Yang, "Medgan: Medical Image Translation using Gans", *Computerized Medical Imaging and Graphics*, Vol. 79, pp. 1-7, 2020.

[22] S.U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem and T. Cukur, "Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks", *IEEE Transactions on Medical Imaging*, Vol. 38, No. 10, pp. 2375-2388, 2019.