2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Learning to Generate Missing Pulse Sequence in MRI using Deep Convolution Neural Network Trained with Visual Turing Test

Vikas Kumar [1], Manoj Kumar Sharma [2], Ramalingam Jehadeesan [1], Balasubramaniam Venkatraman [1], Garima Suman [3], Anurima Patra [3], Ajit H. Goenka [3], and Debdoot Sheet [4]

***Abstract*— Magnetic resonance imaging (MRI) is widely used in clinical applications due to its ability to acquire a wide variety of soft tissues using multiple pulse sequences. Each sequence provides information that generally complements the other. However, factors like an increase in scan time or contrast allergies impede imaging with numerous sequences. Synthesizing images of such non acquired sequences is a challenging proposition that can suffice for corrupted acquisition, fast reconstruction prior, super-resolution, etc. This manuscript employed a deep convolution neural network (CNN) to synthesize multiple missing pulse sequences of brain MRI with tumors. The CNN is an encoder-decoder-like network trained to minimize reconstruction mean square error (MSE) loss while maximizing the adversarial attack. It inflicts on a relativistic Visual Turing Test discriminator (rVTT). The approach is evaluated through experiments performed with the Brats2018 dataset, quantitative metrics viz. MSE, Structural Similarity Measure (SSIM), and Peak Signal to Noise Ratio (PSNR). The Radiologist and MR physicist performed the Turing test with 76% accuracy, demonstrating our approach's performance superiority over the prior art. We can synthesize MR images of missing pulse sequences at an inference cost of 350.71 GFlops/voxel through this approach.**

I. INTRODUCTION

Magnetic resonance imaging (MRI) is popular for anatomical and pathological investigation of the internals of the human body, offering high contrast delineation of soft tissue and pathology. MRI generally excites hydrogen atoms with a specific pulse sequence to eliminate its density per unit volume in the tissue to visualize it as an image. Each sequence of MRI provides complementary information, such as the T1 weighted sequence delineates gray and white matter in the brain, while the T2 weighted sequence delineates fluid from cortical tissue. Such multi-sequence MR provides rich diagnostic information to physicians. Each additional sequence is associated with an increase in acquisition time, cost, prolonged examination time, which is particularly challenging for the young and elderly patients undergoing such investigation [1]. Noise or artifacts also sometimes render such sequences diagnostically unusable at times [2].

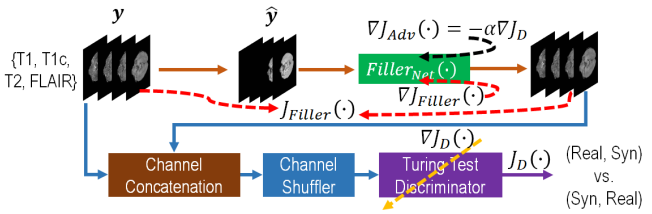

Fig. 1: Framework for learning a convolutional neural network (CNN) to generate missing or corrupt sequences of MR

Challenges: Acquiring all possible MR sequences is challenging for various reasons such as insufficient scan time, artifacts, inappropriate machine parameter settings, and allergies due to specific contrast agents. This leads to inadequate MR image data acquisition, which further adversely affects the quality of diagnosis and treatment. This also affects many downstream data analyses, where it presumes particular pulse sequences to perform the required task. Redoing the scan to acquire the missing or corrupt sequence is impractical due to its expensive nature of the acquisition, longer waiting time with life-threatening cases, and rapid changes in the anatomy of the region in between scan times due to highly active anomalies such as glioblastoma. Hence, it is highly desirable to acquire any missing or corrupt sequences without having to re-scan them completely. In this manuscript, we have implemented a framework (presented in Fig. 1) to simulate missing or corrupt MR pulse sequences from the available MR pulse sequences.

Our approach: It is classically assumed that on account of linear independence between sequences, it is impossible to get one MR sequence from the other MR sequences. Here, we implemented a learning-based method to simulate missing or corrupt synthetic MR sequences from the other available MR sequences. Since tissues are physiologically and pathologically the same, resulting in different MR sequences, we may generate the missing signal from the other signals assuming a higher-order common embedding space between them. We provide hierarchical encapsulation of such higher-order latent entanglement to simulate a missing MR signal assuming that the model can perform the one-to-one mapping from the signals. This manuscript explores an adversarial learning approach to synthesize images corresponding to unacquired or corrupt MR pulse sequences

*This work was not supported by any organization

[1]Vikas Kumar, Ramalingam Jehadeesan and Balasubramaniam Venkatraman are with Indira Gandhi Center for Atomic Research, HBNI, Kalpakkam, Tamilnadu, India-603102 vranjan90@gmail.com {jeha,bvenkat}@igcar.gov.in

[2]Manoj Kumar Sharma is with the Department of Electrical Engineering, Indian Institute of Technology Kharagpur, India-721302 manojsharma.net@gmail.com

[3]Garima Suman, Anurima Patra and Ajit H. Goenka are with the Department of Radiology, Mayo Clinic, Rochester, MN, USA {Suman.Garima,Patra.Anurima, Goenka.Ajit}@mayo.edu

[4] Debdoot Sheet is with the Department of Electrical Engineering, Centre of Excellence in Artificial Intelligence, Indian Institute of Technology Kharagpur, India, debdoot@ee.iitkgp.ac.in

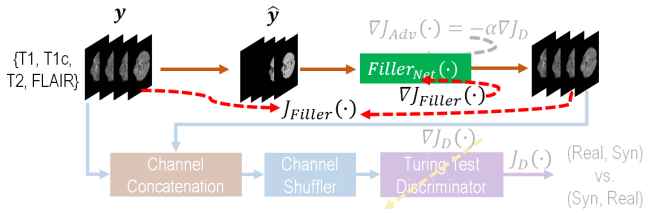

Fig. 2: Training Filler network to minimize distortion loss between the real and synthetic MR sequence

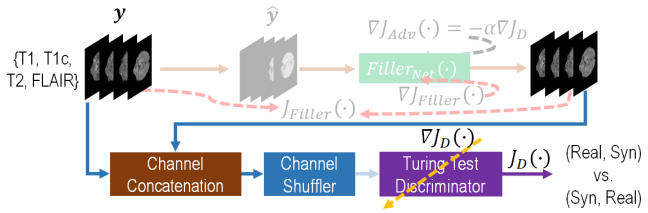

Fig. 3: Training of rVTT (D) to discriminate between the channels of real and generated images

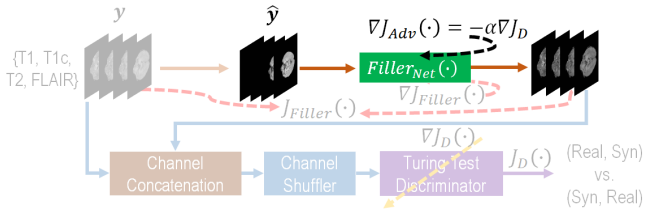

Fig. 4: Adversarial training to enhance the quality of S_{MRI} as close to R_{MRI}

from a given set of acquired sequences. We experimentally demonstrate the ability of the network to synthesize missing pulse sequences (maximum three missing sequences) using the available MR sequences (at least one available). The outcome is anticipated to be useful for treatment planning, reconstruction prior estimation, magnetic field inhomogeneity estimation for image quality correction [3].

II. Prior Art

Single pulse sequence based synthesis: The atlas-based method has been proposed [4] to synthesize T2 weighted image of the head using the T1 weighted image, with the objective of correcting distortion in diffusion-weighted MR image. Local patch-based search on a dictionary of paired-pulse sequences has been used for synthesizing images for a missing pulse sequence [5], followed by joint dictionary learning approaches [6]. In REPLICA [7], a method is proposed to predict the intensity of the alternate tissue contrast from the given tissue by learning nonlinear regression. Yu et.al [8] integrate edge information to improve texture and content structure.

Multi pulse sequence based synthesis: Random forest regression-based multi-model synthesis has been proposed [9] for the reconstruction of FLAIR sequence given T1, T2, and P_D weighted images. Auto-encoder-based multi input-output models have been proposed in [10]. Generative Adversarial Network (GAN) based techniques have been proposed in [11] where a loss term is introduced to reproduce vascular structure. To preserve intermediate to high-frequency details and to enhance the quality of the synthesized sequence employing an adversarial loss, multi input-output variants of the GAN model have been proposed in [12], [13].

III. Method

We consider the task of synthesizing images corresponding to missing or corrupt MR pulse sequences. A 2D axial slice from the scan $\hat{\mathbf{y}}$ is fed to the network as a 3D tensor sized $c \times m \times n$ where c is the number of unique MR pulse sequences and $m \times n$ is the spatial size of the slice. This experiment considers $c = 4$ with the pulse sequences being T1, T1c, T2, and FLAIR. In this tensor, any missing pulse sequence is represented as a $m \times n$ sized zero tensor. This manuscript proposes a three-stage learning framework as follow:

Stage 1: The CNN $Filler_{net}(.)$ generates a $c \times m \times n$ sized tensor $\mathbf{o} = Filler_{net}(\hat{\mathbf{y}})$ from a similar sized tensor $\mathbf{y}$ as described in Fig. 2. The goal is to minimize the MSE loss $\mathbf{J}_{Filler}(.)$ between $\mathbf{o}$ and the tensor of real MR images $\mathbf{y}$.

Stage 2: A relativistic visual Turing test discriminator, implemented with a CNN, learn to discriminate between $\mathbf{y}$ and $\mathbf{o}$, when presented as a pair in randomly shuffled order of channels. The objective is to minimize $J_D(.)$ while updating parameters of $D(.)$ with $\nabla J_D(.)$ as shown in Fig. 3, where $D(.)$ is discriminator. This assesses the subtle pairwise difference in visual perception between the original and synthesized image of an MR pulse sequence, which is distinct from the distortion loss described in $J_{Filler}(.)$.

Stage 3: Finally, at stage 3 (Fig. 4), the filler network $Filler_{net}(.)$ parameters are optimized by minimizing the adversarial loss $J_{adv}(.) = -\alpha J_D(.)$ and thus learn to synthesize images of missing MR pulse sequence (S_{MRI}) which closely resembles those from a real MR pulse sequence (R_{MRI}). The impact of incorporating the adversarial loss using visual Turing test discriminator in learning is clearly noticeable in Fig. 5, where high-frequency details are evident in the synthesized images.

Architecture of the networks: An encoder-decoder like CNN architecture [14] is utilized for synthesizing missing pulse sequences $Filler_{net}(.)$. Here, the encoder block has similar architecture as the VGG-11 network [15]. Activation concatenation of matched layer between encoder and decoder is performed, and max-pool indices are transferred (for depth matched unpooling of activation) at the decoder. We have used ReLU and batch normalization after every convolutional layer. Combining the pre-train VGG network for an encoder, activation concatenation techniques, and max-pool index transfer improves the ability to reconstruct fine grain structures.

The rVTT is another CNN with five convolution layers. It contains a convolution kernel of 3×3 and is interleaved with

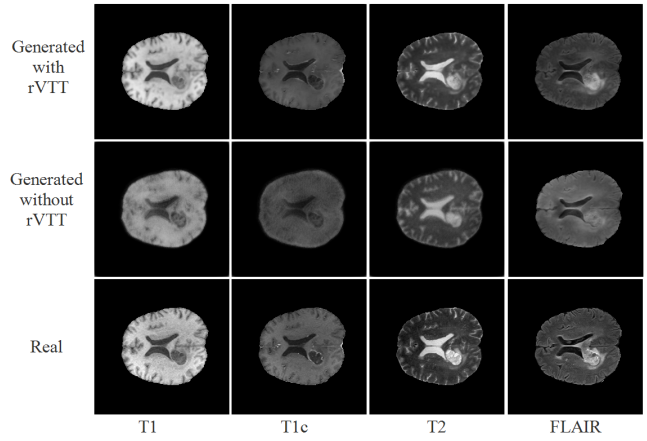

Fig. 5: Qualitative comparison of generated images using with- and without rVTT

a batch-normalization layer followed by non-linear leaky ReLU. Finally, the Sigmoid activation function is added to predict the segmented output from the terminal layer. There are 8 channels at the first layer, and in subsequent layers, channels are multiplied by a factor of 2. The network terminates with a two-neuron fully connected output layer.

IV. Experiments

Dataset: The performance is evaluated on multi-modal brain tumor segmentation challenge 2018 (BratTs 2018) [16] data-set. It consists of 285 volumes of 19 different subjects containing glioblastoma (GBM), lower-grade glioma (LGG), and high-grade glioma (HGG) cohorts. Each volume contains four co-registered MR pulse sequences such as T1, T1c, T2, and FLAIR. The size of individual volume is $240 \times 240 \times 155$. 190 out of 210 volumes of GBM/HGG used for training, 10 for validation, and 10 for testing. The intensity is normalized within the range $[0, 1]$.

Training setup: The filler and discriminator networks are trained for 30 epochs with a learning rate of 1×10^{-5} and Adam optimizer. Experimental analysis are performed with $1, 2$ or 3 missing sequences and summarized in Table I.

V. Results and Discussion

Qualitative: Synthesized images for missing one, two, or three MR pulse sequences are presented in Fig. 6. The figure shows that the synthesized images retain the high-frequency information that delineates the gray and white matter and preserves the tumor regions. It is also worthy to note that the image for the sequence under test is synthesized with equivocal efficacy across the experiments with up to three missing pulse sequences. Further, slight visual superiority is observed when only one sequence is missing compared to two or three missing sequences.

Quantitative: Table I shows the overall statistics of three quantitative metrics MSE, PSNR, and SSIM, for the complete set of experiments. Synthesis of multiple missing sequences is reported in [12] and comparison with our approach is presented in Table I and Table II. From the above

TABLE I: Quantitative comparison of results on different metrics for different combination of missing MRI sequences (✗: missing MR sequence, ✓: available MR sequence)

T1	T1c	T2	FLAIR	MSE	SSIM	PSNR
✗	✓	✓	✓	1.5×10^{-3}	0.97	29.32
✓	✗	✓	✓	2.6×10^{-3}	0.96	28.42
✓	✓	✗	✓	1.5×10^{-3}	0.96	28.34
✓	✓	✓	✗	2.2×10^{-3}	0.95	27.56
✗	✗	✓	✓	7.4×10^{-3}	0.97	31.94
✗	✓	✗	✓	2.4×10^{-3}	0.97	28.25
✗	✓	✓	✗	1.9×10^{-3}	0.96	29.29
✓	✗	✗	✓	4.5×10^{-3}	0.96	28.65
✓	✗	✓	✗	1.7×10^{-3}	0.96	29.64
✓	✓	✗	✗	2.4×10^{-3}	0.95	27.04
✓	✗	✗	✗	3.4×10^{-3}	0.96	27.85
✗	✓	✗	✗	4.7×10^{-3}	0.96	26.62
✗	✗	✓	✗	2.7×10^{-3}	0.95	26.26
✗	✗	✗	✓	3.4×10^{-3}	0.94	26.37
Average (Our method)				**3.2** $\times 10^{-3}$	**0.96**	**28.23**
Sharma et.al [12]				8.2×10^{-3}	0.91	24.79

TABLE II: Quantitative comparison of results with respect to SSIM and PSNR for generation of T2 MR sequence from T1 and vice versa

	pGAN [13]		Replica [7]		Multimodel [12]		Our method	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
T1->T2	0.95	27.19	0.924	24.64	0.94	25.09	**0.96**	**28.62**
T2->T1	0.94	25.80	0.92	24.49	0.93	23.78	**0.95**	**27.26**

tables, we observed that our method outperforms the best prior art by a factor of 5.49% and 13.88% improvement in terms of SSIM and PSNR for T1 and T2 synthesis.

Turing Test: A visual identification task was performed involving MR physicists and Radiologists. A pair of real and synthetic images for the different pulse sequences were presented. Results indicated 76% accuracy in identifying the real one from synthetic images. Particularly, in T1c and FLAIR sequences, 50% of experts predicted real as synthetics whereas 50% predicted synthetic as real MR sequences. This indicates equivocal superiority in the reconstruction of our approach.

VI. Conclusion

The proposed framework utilizes an encoder-decoder-based end-to-end deep convolutional neural network architecture. Here, the network reconstructs the missing or corrupt MR pulse sequences by using available MR pulse sequences. We observed use of a relativistic visual Turing test discriminator network improves reconstruction quality by attacking the filler network adversarially. As a result, we get an enhanced reconstructed MR pulse sequence. The framework further utilizes the pre-trained weight of VGG-11 to initialize the filler network, and this provides a good starting point for the network to optimize faster. In addition, the filler and

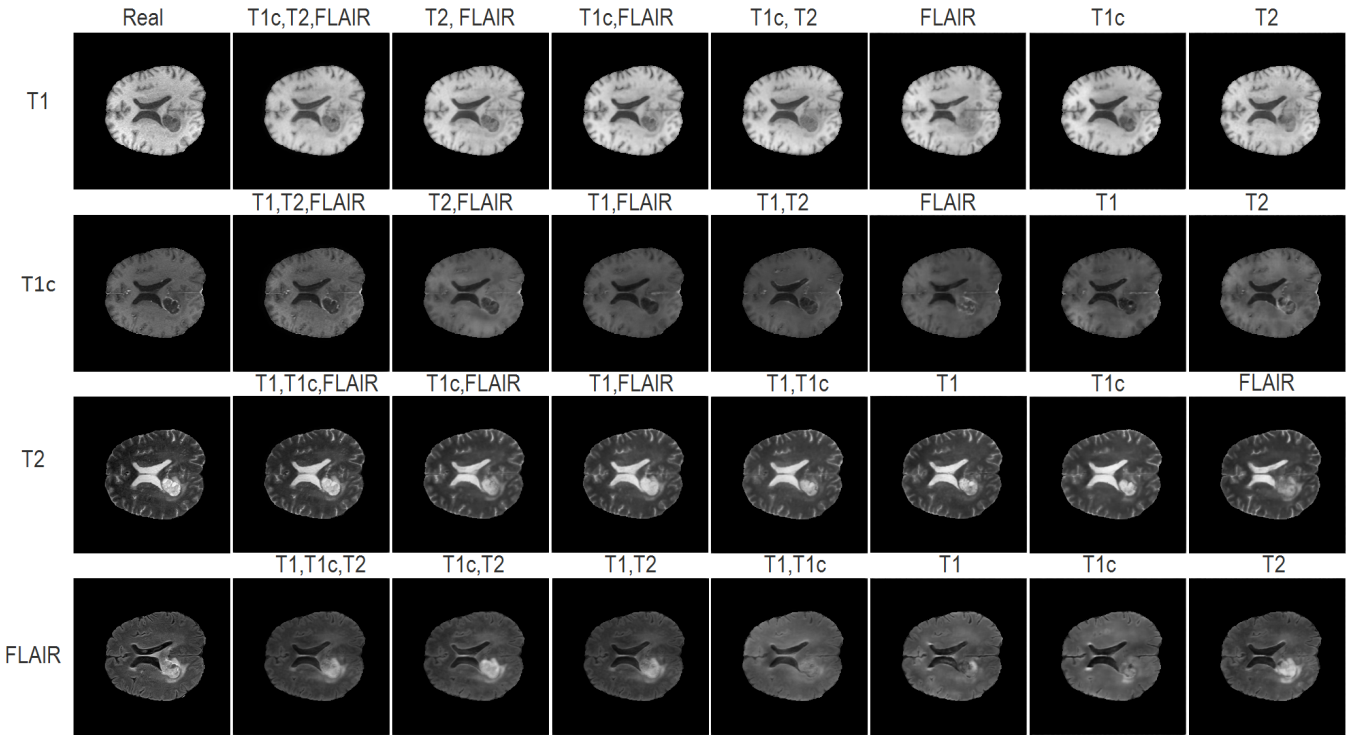

Fig. 6: Qualitative analysis of experiments. First column of each row is real image and remaining columns are generated images. (column names: available sequence, row names: generated missing sequences)

discriminator networks play a min-max game, resulting in a perceptually significant improvement in reconstruction. Here, the filler network utilizes the gradient from both the filler and discriminator network during learning. The experimental results demonstrate that our method outperforms other prior art, and there is significant quantitative improvement in terms of PSNR and SSIM. Also, superiority is proven through a visual Turing test involving practicing experts of radiology reporting. We have presented qualitative results of our method, substantiating the capability to synthesize any combination of missing pulse sequence images.

REFERENCES

[1] B. B. Thukral, "Problems and preferences in pediatric imaging," *Ind. J. Radiol., Imaging*, vol. 25, no. 4, p. 359, 2015.

[2] K. Farahani, U. Sinha, S. Sinha, L. C. Chiu, and R. B. Lufkin, "Effect of field strength on susceptibility artifacts in magnetic resonance imaging," *Comp. Med. Imaging, Graphics*, vol. 14, no. 6, pp. 409–413, 1990.

[3] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. V. Leemput, and B. Fischl, "Is synthesizing mri contrast useful for inter-modality analysis?" in *Proc. Med. Image Comp., Comput. Assist. Interv.*, 2013, pp. 631–638.

[4] S. Roy, Y.-Y. Chou, A. Jog, J. Butman, and D. L. Pham, "Patch based synthesis of whole head mr images: Application to epi distortion correction," in *Int. W. Simulation, Synthesis, Med. Imaging*, 2016, pp. 146–156.

[5] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, "Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization," in *Proc. Med. Image Comp., Comput. Assist. Interv.*, 2013, pp. 606–613.

[6] Y. Huang, L. Beltrachini, L. Shao, and A. F. Frangi, "Geometry regularized joint dictionary learning for cross-modality image synthesis in magnetic resonance imaging," in *Int. W. Simulation, Synthesis, Med. Imaging*, 2016, pp. 118–126.

[7] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Med. Image Anal.*, vol. 35, pp. 475–488, 2017.

[8] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis," *IEEE Trans. Med. Imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.

[9] A. Jog, A. Carass, D. L. Pham, and J. Prince, "Random forest flair reconstruction from t 1, t 2, and p d-weighted mri," in *Proc. Int. Symp. Biomed. Imaging*, 2014, pp. 1079–1082.

[10] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, "Multi-modal mr synthesis via modality-invariant latent representation," *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 803–814, 2017.

[11] S. Olut, Y. Sahin, U. Demir, and G. Unal, "Generative adversarial training for mra image synthesis using multi-contrast mri," in *Int. W. Pred. Intell., Med.*, 2018, pp. 147–154.

[12] A. Sharma and G. Hamarneh, "Missing mri pulse sequence synthesis using multi-modal generative adversarial network," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1170–1183, 2019.

[13] S. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[14] S. Nandamuri, D. China, P. Mitra, and D. Sheet, "Sumnet: Fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes," in *Proc. Int. Symp. Biomed. Imaging*, 2019, pp. 1729–1732.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," pp. 1–14, 2015.

[16] B. Menze, A. Jakab, S. Bauer, J. K-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.