Çoklu Kontrast MRG'de Otokodlayıcı ve Öğrenme Aktarımı Kullanarak Görüntü Sentez Kalitesini İyileştirme
Improving Image Synthesis Quality in Multi-Contrast MRI Using Transfer Learning via Autoencoders

2022 30th Signal Processing and Communications Applications Conference (SIU) | 978-1-6654-5092-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SIU55565.2022.9864750

Şahan Yoruç Selçuk†‡ *, Onat Dalmaz†‡ *, Salman ul Hassan Dar†‡, Tolga Çukur†‡§
†Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
‡National Magnetic Resonance Research Center, Bilkent University, Ankara, Turkey
§Neuroscience Program, Sabuncu Brain Research Center, Bilkent University, Ankara, Turkey

***Özetçe* —Manyetik rezonans görüntülemenin (MRG) bir seans içinde çeşitli kontrastları yakalayabilme kapasitesi daha fazla tanı bilgisi elde edilmesini sağlar. Ancak, bu tür çoklu kontrast MRG testleri uzun zaman alır, bu nedenle genelde kontrastların yalnızca bir kısmı elde edilebilir. Sentetik çoklu kontrast MRG, radyolojik gözlemleri ve sonrasındaki görüntü analizlerini iyileştirme potansiyeline sahiptir. Gerçekçi sonuçlar üretebilme yeteneğiyle, çekişmeli üretici ağlar son zamanlarda tıbbi görüntüleme sentezi için en popüler seçim olmuştur. Bu çalışmada çoklu kontrast MRG'de görüntü sentezi kalitesini iyileştirmek için yeni bir üretken çekişmeli yapı önerilmektedir. Yöntemimiz, önceden eğitilmiş otokodlayıcı ağlarını transfer öğrenme yöntemiyle sentez görevine uyarlayarak eğitim sürecini daha optimal ağ parametreleriyle başlattığı için görüntü sentezi kalitesini arttırmaktadır. Yapılan deneyler, önerilen yöntemin iyi bilinen bir çoklu kontrast MRG veri kümesinde rakip sentez modellerinden ortalama olarak 0.95 dB PSNR daha iyi sonuç aldığını göstermektedir.**

Anahtar Kelimeler—Çoklu kontrast MRG, otokodlayıcı, öğrenme aktarımı, çekişmeli üretici ağlar

***Abstract*—The capacity of magnetic resonance imaging (MRI) to capture several contrasts within a session enables it to obtain increased diagnostic information. However, such multi-contrast MRI tests take a long time to scan, resulting in acquiring just a part of the essential contrasts. Synthetic multi-contrast MRI has the potential to improve radiological observations and consequent image analysis activities. Because of its ability to generate realistic results, generative adversarial networks (GAN) have recently been the most popular choice for medical imaging synthesis. This paper proposes a novel generative adversarial framework to improve the image synthesis quality in multi-contrast MRI. Our method uses transfer learning to adapt pre-trained autoencoder networks to the synthesis task and enhances the image synthesis quality by initializing the training process with more optimal network parameters. We demonstrate that the proposed method outperforms competing synthesis models by 0.95 dB on average on a well-known multi-contrast MRI dataset.**

Keywords—Multi-contrast MRI, autoencoder, transfer learning, generative adversarial networks

*Equal contribution

978-1-6654-5092-8/22/$31.00 ©2022 IEEE

I. INTRODUCTION

Multi-modal medical imaging combines several scans on the body to obtain additional tissue information, which improves diagnostic accuracy. Magnetic resonance imaging (MRI) is the dominant modality in clinical neuroimaging because of its superior soft-tissue contrast and lack of ionizing radiations. Its capacity to record the anatomy in a range of different contrasts allows it to collect more diagnostic information during an examination. However, multi-contrast MRI tests require a long time to scan, leading to a collection of insufficient contrast images [1]. As a result, successive scans' time and financial costs severely limit its utilization. This considerable constraint has sparked interest in synthesis approaches for multi-contrast MRI procedures that can reconstruct missing scans from a subset of available images. Following radiological observations and image analysis tasks like segmentation and identification, synthetic multi-contrast MR images are considered helpful [2].

The purpose of MRI synthesis is to generate target-contrast images for a subject using source-contrast images that are already available. Because MR images are high dimensional, target-modality data is lacking during inference. Moreover, tissue contrast changes nonlinearly across different contrasts. These aspects make MRI synthesis an ill-posed inverse problem [3], [4]. On the other hand, deep learning approaches, including autoencoder structures, have yielded significant performance advantages to solve this challenging problem [5]–[11]. With their enhanced capture of detailed tissue structure, Generative Adversarial Networks (GAN) that use an adversarial loss have become the de facto choice for MRI synthesis [5]. Transfer learning methods were also employed in the literature for MRI applications [12], [13], mainly to address the issue of data limitations in deep network training for accelerated MRI.

In this work, we propose a method to improve image synthesis quality in multi-contrast MRI. Motivated by the pGAN model [5], we adopt convolutional autoencoders to represent image features of various MRI contrasts. Different from existing methods in the literature, we pre-train two autoencoder networks to learn the latent representations of

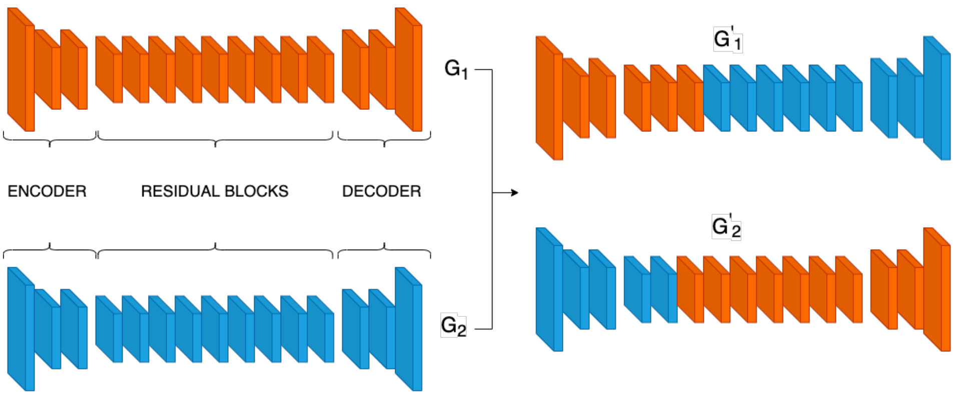

Figure 1: Formation of generators for the two synthesis tasks ($T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$) as proposed in our method. Individual autoencoder networks G_1 and G_2 were firstly pre-trained, then concatenated to form the generators. *Best viewed in color.*

two MRI modalities (i.e., two separate autoencoder networks trained to learn the features of T_1 and T_2 contrasts) and use the combination of these two autoencoders for multi-contrast MRI synthesis task. Since we first pre-train mentioned autoencoder networks on different tasks, we then apply transfer learning to fine-tune the pre-trained networks to the synthesis task. We performed extensive tests for synthesizing missing sequences in multi-contrast MRI, utilizing the IXI MRI dataset (https://brain-development.org/ixi-dataset/) for benchmarking. We illustrate our proposed method's advantage through these experimentations.

II. Theoretical Background

A. Generative Adversarial Networks (GAN)

Generative adversarial networks consist of two adversarial subnetworks, namely the generator (G) and the discriminator (D) [14]. In the original GAN framework, the generator is trained to generate fake images ($\hat{y}$) that imitate the probabilistic distribution of a particular domain from random noise (z). In contrast, the discriminator is trained to discriminate between these fake images and the actual images (y) in the domain. The main idea is to train and optimize both of these subnetworks simultaneously for the generator to generate high-quality fake images that the discriminator cannot distinguish [14]. To this end, the adversarial loss is adopted during the training process:

$$\mathcal{L}_{adv} = -E_y\left[(D(y)-1)^2\right] - E_z\left[D(G(z))^2\right]. \qquad (1)$$

B. Conditional Generative Adversarial Networks (cGAN)

Conditional GAN models are obtained via conditioning classical GAN models with an input instead of random noise. In this case, the generator takes an image from the source domain (x) as the input and tries to generate its corresponding fake image in the target domain ($\hat{y}$), so that $G(x) = \hat{y}$. The discriminator, on the other hand, tries to discriminate the generated fake image and the real image in the target domain, $D(y, \hat{y}) \in [0, 1]$ [15]. Once again, the adversarial loss function is utilized to train both of these subnetworks:

$$\mathcal{L}_{cGAN} = -E_y\left[(D(y)-1)^2\right] - E_x\left[D(G(x))^2\right]. \qquad (2)$$

Moreover, with the corresponding real image from the target domain, the pixel-wise loss function (L_1-distance) can be employed alongside the adversarial loss to obtain higher quality images:

$$\begin{aligned}\mathcal{L}_{cGAN} = &- E_y\left[(D(y)-1)^2\right] - E_x\left[D(G(x))^2\right] \\ &+ E_{xy}[\|y - G(x)\|].\end{aligned} \qquad (3)$$

III. Method

A. Overview

In essence, our proposed method is a conditional and adversarial network model used for image synthesis that consists of generator and discriminator subnetworks. Influenced by the powerful pGAN [5] method, we adopted the architecture of the discriminator subnetwork from [16], referred to as PatchGAN. To obtain the generator subnetwork, we first pre-train two separate autoencoder networks (G_1 and G_2) on learning the distributions and features of T_1 and T_2 modalities of MRI. G_1 and G_2 consist of three components (with the same number of network parameters): 3 encoder layers, 9 residual blocks, and 3 decoder layers, which make a total of 15 layers in each network. The encoder uses strided convolutional layers to extract hidden representations based on the input contrast. Residual blocks use convolutional operations along with additive residual skip connections to further refine the encoder's low-level depictions [17]. In the end, feature maps from residual blocks and the encoder are used by the decoder to merge high-level feature maps with low-level encoder feature maps successfully.

To learn the distributions of differently contrasted images, we provide T_1-contrasted images as the input for G_1 and train it to regenerate the given inputs. Similarly, we feed the second autoencoder (G_2) T_2-contrasted images and train it to generate provided T_2-contrasted images at the output. To train these two autoencoder networks, we use pixel-wise L_1-distance loss function (previously explained in Section II). At the end of this initial pre-training process, two networks that have learned the feature representations of T_1 and T_2 contrasts are acquired. Then, a concatenated combination of these networks is placed to the overall adversarial model as the generator. (Here, it is worth mentioning that, in reality, we train two final generator networks G_1' and G_2' in the same manner, corresponding to two synthesis tasks as shown in Figure 1.) That is, out of the total 15 layers, the first x layers were taken from the G_1

network, and the remaining $15 - x$ layers were taken from the G_2 network. The number of layers used from G_1, x, is a hyperparameter of our method, which was determined by cross-validation during the training process. Next, these parts were concatenated and directly connected to form the generator of the model. This process is shown in Figure 1. Then, we train our conditional and adversarial model based on the loss given in Equation 3, with a slight change. To control the pixel-wise loss prioritization, we use a hyperparameter λ_{L_1}. The final objective is given as:

$$\mathcal{L} = -E_y\left[(D(y)-1)^2\right] - E_x\left[D(G(x))^2\right] + \lambda_{L_1}E_{xy}[\|y - G(x)\|]. \tag{4}$$

As a result of this training, we fine-tune the pre-trained generator G according to the image synthesis task. The main advantage of this transfer learning approach is to start the training process with a more suitable set of weights for the network parameters of the generator.

B. Dataset

We evaluate our proposed method on a well-known multi-contrast MRI dataset IXI. From the IXI dataset, we consider T_1 and T_2 contrasted brain MR images from 53 patients. Out of these 53 patients, we split 25 patients for training, 10 patients for validation, and 18 for testing phases. In both phases, 100 axial brain tissue cross-sections were chosen from each patient. T_2-contrasted images were spatially registered onto T_1-contrasted images before analysis. The registration was accomplished by mutual information, and affine transformation in FSL [18].

C. Competing Methods

We examine the performance of our proposed method against two existing image synthesis methods in the literature.

- We compare our method with pGAN, which is a convolutional GAN model [5]. Its generator is made up of an encoder, nine residual blocks, and a decoder. The encoder is made up of three strided convolutional layers, whereas the decoder is made up of three transposed convolutional layers.
- The other competing method is the pix2pix [19] model using the U-Net [20] generator, which is also a convolutional GAN model.

We used an adversarial setting for the training of all competing models. The hyperparameters of each approach were adjusted using identical cross-validation procedures. We used the same objective function and PatchGAN discriminator architecture for each of the techniques to ensure a fair comparison.

D. Implementation Details

Encoders found in G_1 and G_2 subnetworks (and consequently in the final generator G) are made up of three convolutional layers with kernel sizes of $7, 3$ and 3 in cascade. On the other hand, all decoders are made up of three transposed-convolutional layers in cascade with kernel sizes of $3, 7$, and 7. Our model and the competing models were trained with the same discriminator. During the training process of all models, the final objective function is given in Equation 4, consisting of the adversarial and pixel-wise loss function components. We employ the same optimization setting and pixel-wise loss prioritization during all training processes to ensure a just comparison between the models. Cross-validation is used to select learning rate schedules for all methods and the layer number hyperparameter x for our proposed method only. We saw that these determined settings constantly resulted in near-best performances for all training phases. We utilized the Adam optimizer [21] with momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.99$. The training processes were maintained for 100 epochs in each case. The learning rate for the experiments was set to 0.0002 in the first 50 epochs, and a linear decrement to 0 was employed in the last 50 epochs. Using cross-validation, we fine-tuned λ_{L_1} to 100.

Table I: Synthesis test results for $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ tasks on IXI dataset in terms of PSNR (dB) and SSIM (%). All measurements are presented as mean $\pm$ standard deviation across all test subjects. Bold denotes the best performances for the given task.

Task	Metric	pGAN	pix2pix	Our Method
$T_1 \longrightarrow T_2$	**PSNR**	28.54 ± 1.94	26.96 ± 1.87	**28.85 ± 2.03**
	SSIM	0.925 ± 0.031	0.908 ± 0.037	**0.929 ± 0.036**
$T_2 \longrightarrow T_1$	**PSNR**	28.28 ± 2.32	27.35 ± 2.46	**28.60 ± 2.28**
	SSIM	0.936 ± 0.030	0.926 ± 0.035	**0.937 ± 0.028**

IV. Results & Discussion

To begin with, after the cross-validation analysis, we have found out to be the optimal value for the layer number hyperparameter x is equal to 6 for the $T_1 \rightarrow T_2$ task, and 10 for the $T_2 \rightarrow T_1$ tasks. Meaning that, in the former task, the first 6 layers of G_1 and last 9 layers of G_2 were concatenated to form G. In the second task, the first 5 layers of G_2 and last 10 layers of G_1 were concatenated to form G. This can be explained in two ways: (1) In both tasks, it was observed that more of the end layers of the autoencoder network that have learned the latent representation of the target domain were utilized in the final generator. In general, decoding parts of the autoencoders are the sources of replication error. We can see that our best models try to use more layers of the target domain autoencoder to reduce the error and generate better output images. (2) We have seen that 10 layers of the target domain autoencoder were used for the $T_2 \rightarrow T_1$ task compared to 9 layers used for the $T_1 \rightarrow T_2$ task. It is known that the former task is more difficult to accomplish, so we deduce that our proposed method tries to use even more of the target domain autoencoder to result in a less overall error.

Next, we evaluate our model's multi-contrast MR image synthesis performance compared to the competing methods pGAN and pix2pix. Table I demonstrates the performance results for three mentioned methods. The performance of each model was evaluated using the Peak Signal-to-Noise Ratio (PSNR), the ratio of a signal's maximal power to the level of corrupting noise that influences its fidelity, and the Structural Similarity Index Measure (SSIM), a perceptual method that incorporates image degradation as structural information change, metrics [22]. We have observed that our proposed

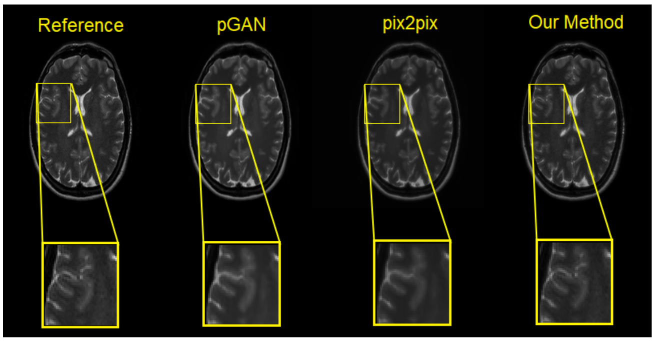

Figure 2: T_2 synthesis

method outperformed competing methods in terms of PSNR and SSIM measurements in all tasks. We find that our method outperforms pGAN by 0.32 dB PSNR and pix2pix by 1.57 dB PSNR, on average across the two tasks. Figure 2 shows images of T_2 synthesis results for all the models. As shown in the figure, we observe that our method depicts tissues better than other competing methods due to starting the training process from a more optimal set of network weights.

V. Conclusion

In this work, we have proposed a novel method to improve the image synthesis quality in multi-contrast MRI. Our proposed method uses two separate autoencoder subnetworks, which are trained to learn the latent representations of two MRI modalities. Our approach then concatenates different parts of these two individual autoencoders to form the pre-trained generator network for our conditional and adversarial image synthesis framework. Our strategy achieved a good performance in the MRI synthesis task by smartly connecting two pre-trained networks to initiate the training process with favorable network weights for the generator. We showed that our method achieved better performance than two baseline synthesis models in terms of standard evaluation metrics. We anticipate that the proposed method can enhance multi-contrast MRI synthesis in both practice and clinical contexts.

VI. Acknowledgements

This study was supported in part by a TUBITAK BIDEB scholarship awarded to O. Dalmaz, and by TUBA GEBIP 2015 and BAGEP 2017 fellowships awarded to T. Çukur.

References

[1] Brij Bhushan Thukral, "Problems and preferences in pediatric imaging," *Indian Journal of Radiology and Imaging*, vol. 25, no. 04, pp. 359–364, 2015.

[2] Blake E Dewey, Can Zhao, Jacob C Reinhold, Aaron Carass, Kathryn C Fitzgerald, Elias S Sotirchos, Shiv Saidha, Jiwon Oh, Dzung L Pham, Peter A Calabresi, et al., "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes," *Magnetic resonance imaging*, vol. 64, pp. 160–170, 2019.

[3] Ciprian Catana, Andre van der Kouwe, Thomas Benner, Christian J Michel, Michael Hamm, Matthias Fenchel, Bruce Fischl, Bruce Rosen, Matthias Schmand, and A Gregory Sorensen, "Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype," *Journal of nuclear medicine*, vol. 51, no. 9, pp. 1431–1438, 2010.

[4] Yawen Huang, Ling Shao, and Alejandro F Frangi, "Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 815–827, 2017.

[5] Salman UH Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[6] Onat Dalmaz, Mahmut Yurt, and Tolga Çukur, "ResViT: Residual vision transformers for multi-modal medical image synthesis," *arXiv preprint arXiv:2106.16031*, 2021.

[7] Alper Güngör, Baris Askin, Damla Alptekin Soydan, Emine Ulku Saritas, Can Barış Top, and Tolga Çukur, "TranSMS: Transformers for Super-Resolution Calibration in Magnetic Particle Imaging," *arXiv preprint arXiv:2111.02163*, 2021.

[8] Yilmaz Korkmaz, Salman UH Dar, Mahmut Yurt, Muzaffer Özbey, and Tolga Cukur, "Unsupervised MRI reconstruction via zero-shot learned adversarial transformers," *IEEE Transactions on Medical Imaging*, 2022.

[9] Mahmut Yurt, Muzaffer Özbey, Salman Ul Hassan Dar, Berk Tınaz, Kader Karlı Oğuz, and Tolga Çukur, "Progressively volumetrized deep generative models for data-efficient contextual learning of MR image recovery," *arXiv preprint arXiv:2011.13913*, 2020.

[10] Mahmut Yurt, Salman UH Dar, Aykut Erdem, Erkut Erdem, Kader K Oguz, and Tolga Çukur, "mustGAN: Multi-stream generative adversarial networks for MR image synthesis," *Medical image analysis*, vol. 70, pp. 101944, 2021.

[11] Gokberk Elmas, Salman UH Dar, Yilmaz Korkmaz, Emir Ceyani, Burak Susam, Muzaffer Özbey, Salman Avestimehr, and Tolga Çukur, "Federated Learning of Generative Image Priors for MRI Reconstruction," *arXiv preprint arXiv:2202.04175*, 2022.

[12] SU Hassan Dar and Tolga Çukur, "Transfer learning for reconstruction of accelerated MRI acquisitions via neural networks," in *Proceedings of the 26th Scientific Meeting of ISMRM*, 2018.

[13] Salman Ul Hassan Dar, Muzaffer Özbey, Ahmet Burak Çatlı, and Tolga Çukur, "A transfer-learning approach for accelerated MRI using deep neural networks," *Magnetic resonance in medicine*, vol. 84, no. 2, pp. 663–685, 2020.

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[15] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[16] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "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] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

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

[22] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.