Perceptual cGAN for MRI Super-resolution

Sahar Almahfouz Nasser1*, Saqib Shamsi2*, Valay Bundele1, Bhavesh Garg1, and Amit Sethi1

Abstract—Capturing high-resolution magnetic resonance (MR) images is a time consuming process, which makes it unsuitable for medical emergencies and pediatric patients. Low-resolution MR imaging, by contrast, is faster than its high-resolution counterpart, but it compromises on fine details necessary for a more precise diagnosis. Super-resolution (SR), when applied to low-resolution MR images, can help increase their utility by synthetically generating high-resolution images with little additional time. In this paper, we present an SR technique for MR images that is based on generative adversarial networks (GANs), which have proven to be quite useful in producing sharp-looking details in SR. We introduce a conditional GAN with perceptual loss, which is conditioned upon the input low-resolution image, which improves the performance for isotropic and anisotropic MRI super-resolution.

Clinical relevance—MR image super-resolution has the potential for improving image acquisition speed to save the time of the clinicians, while guaranteeing high-quality images.

I. INTRODUCTION

A magnetic resonance imaging (MRI) scan is very helpful for non-invasive medical diagnoses. However, an MRI is not an option for medical emergencies and is very difficult to do for pediatric patients. Moreover, MRI equipment cost goes up with its resolution. Furthermore, the scan time is also directly related to the resolution of an MRI, which makes it susceptible to patient motion. A reduction in the spatial resolution would result in faster imaging time at the cost of the fine-grained details in the images that aid in making better diagnostic decisions. AI-based single image super-resolution (SISR) methods can reduce the time and cost of MRI acquisition by using low-resolution (LR) input images and quickly producing synthetic high-resolution (HR) images. The challenge is to produce the finer details that are close to those of the underlying HR images.

We propose a 3D SR method based on deep neural networks and generative adversarial networks (GANs) for MRI volumes that outperforms bicubic interpolation and state-of-the-art methods. We improve various individual components in the GAN framework to get an overall boost in the perceptual quality of super resolved MRIs. We study the impact of various architectural and loss function choices for both isotropic and anisotropic SR cases. In summary, our contribution is threefold. Firstly, we constructed a 3D VGG-style network [1] for the MRI classification task, then we used this pretrained network, which has already learned to encode the perceptual and semantic information from the classification task, to compute the perceptual loss [2] between the real and super resolved MRI volumes. Additionally, we proposed a conditional GAN that discriminates between the real and the super resolved images conditioned on the low-resolution MR images. Lastly, our ablation study revealed that a combination of the SRRResNet generator [8], projection discriminator [12], and perceptual loss [2] is the best for the given problem. The projection discriminator preserves the stylistic features, while the perceptual loss preserves the semantic ones.

II. RELATED WORK

Super-resolution is an ill-posed inverse problem as it might have more than one solution for the high resolution image $x$ given its low-resolution counterpart $y$. Currently, methods based on deep learning are among the state-of-the-art in terms of reconstruction accuracy and inference speed. In 2015 Dong et al. [3] proposed super-resolution convolutional neural network (SRCNN). This was the first deep learning-based method for single image super-resolution which has led to a dramatic leap in this domain. SRCNN is a very shallow network; it adds details to an interpolated version of the LR image.

Unlike SRCNN, fast super-resolution convolutional neural network (FSRCNN) [6] upsamples the feature maps at the end using a nearest-neighbor interpolation-based deconvolutional layer, which speeds the training and reduces the blur in the output image. Kim et al. [7] demonstrated that using a very deep super-resolution network (VDSR) further improves the quality of the output image. To boost the convergence, the authors used a high learning rate and gradient-clipping.

However, the aforementioned methods suffer from the problem of vanishing gradient. To address this issue, many researchers have proposed architectures based on ResNet and DenseNet, such as the super resolution ResNet (SRResNet) [8], and the super resolution deep net (SRDenseNet) [9]. Zhang et al. [10] proposed residual channel attention network (RCAN), which exploits the existence of the short and long skip connections of the residual-in-residual (RIR) blocks [10] to train the very deep neural networks.

The increasingly impressive results of the GANs in generating images encouraged Wang et al. to explore their power in generating realistic textures for single image super-resolution [8]. That is, a generator is used to produce the super-resolved images, while a co-trained discriminator...
tries to discriminate between synthesized and real high-resolution images in order to improve the generator. Though the generated images had sharper details compared to the results of the existed super-resolution methods, this method produces unrealistic but sharp-looking artifacts. To counter the generation of unrealistic details, the use of different architectures, adversarial loss, and perceptual loss has been proposed [11].

Our method differs from the rest as it makes use of a GAN framework with a projection discriminator. Additionally, we train a VGG network on a 3D MRI based task on our own for computing perceptual loss rather than using off the shelf weights.

III. PROPOSED METHOD

We propose a deep learning-based method with an end-to-end pipeline that learns the mapping of a low resolution 3D MRI to its high resolution counterpart. This is unlike the method proposed by Lin et al. [7] for the same problem, which was not end-to-end, but won second place in the SuperMUDI challenge for MRI super-resolution [16].

For this study, we used a generator architecture that has been introduced for 2D SISR and adapted it to 3D SR. Unlike the architecture proposed by Lan et al. [5] our generator does not contain self-attention. Not only does this make the architecture simpler to implement, but also simplifies the overall training pipeline as complicated stabilization techniques and loss functions are not required. We also introduced a 3D version of projection discriminator [12], and found that it benefited the training process as well as the overall results. To our knowledge, this is the first time that a 3D projection discriminator has been used for MRI super-resolution.

The following subsections describe the model architectures of the generator and discriminator that we used.

A. Projection Discriminator

Conditional GANs (cGANs) [13] have shown immense usefulness for the task of conditional image generation. Unlike in standard GANs, the discriminator of a cGAN discriminates between the generator’s conditional distribution and the target conditional distribution of generated samples \( x \) conditioned upon a paired input \( y \). A projection based discriminator to incorporate conditional information into the discriminator has also been proposed [12]. They showed the effectiveness of the model over other approaches of feeding conditional information to the discriminator, such as concatenating the conditional information with either the input or the feature map learned by one of the intermediate layers in the network. We adapted their framework for MRI by constructing a projection discriminator using 3D-convolutions.

For super-resolution, the following formulation was used for the discriminator function:

\[
f(x, y; \theta) = \sum_{i,j,k,l} (y_{ijkl} F_{ijkl}(\phi(x; \theta_\phi)))+\psi(\phi(x; \theta_\phi); \theta_\psi),
\]

where \( x \in \mathbb{R}^{H_1 \times H_2 \times H_3 \times H_4} \) is the high resolution MRI and \( y \in \mathbb{R}^{L_1 \times L_2 \times L_3 \times L_4} \) is the low resolution MRI and \( F(\phi(x; \theta_\phi)) = V*\phi(x; \theta_\phi) \) for a convolutional kernel \( V \) and convolutional operator \*.

The four dimensions in the input and output MRI refer to the length, width, depth and channels (volumes) respectively. The architecture of projection discriminator is shown in Fig. 15 of [12].

B. Generator Architecture

Our generator architecture is a fully convolutional one inspired by SRResNet [8]. This generator consists of a 3D-convolutional block followed by eight residual blocks, a point-wise convolution and a \( 3 \times 3 \times 3 \) convolution operation to reduce the number of channels before upsampling so that the architecture becomes memory efficient, followed by an upsampling block and the network ends with another convolutional block. We also use a global skip connection. We use a combination of an upsampling block and the network ends with another convolutional block. We also use a global skip connection. We use a combination of an upsampling block and the network ends with another convolutional block. We also use a global skip connection. We use a combination of an upsampling block and the network ends with another convolutional block. We also use a global skip connection.

C. Perceptual Loss

As the perceptual loss considers the structural and content similarities between the real and generated volumes at different scales, it improves the reconstruction of HR volume [2]. To study its effect on the performance of our proposed method, we constructed a 3D VGG network trained on MR data. This network comprises five convolutional blocks and a classification head. The 3D kernels of all the convolutional layers are of sizes \( 3 \times 3 \times 3 \). The number of filters increases by a factor of two as we go deeper starting from 64 filters at the first layer. Every convolutional block consists of a convolutional layer, a 3D batch normalization layer, and a ReLU activation function. The classification head contains three fully connected layers of dimensions 512, 128, and 3.

The task of the VGG network is to classify the input image into one of three classes T1, FLAIR, or diffusion MRI volume to learn the semantic information and later transfer this knowledge to the generator when training it to minimize the perceptual loss.

IV. DATA AND EXPERIMENTS

The Super-resolution of Multi-Dimensional Diffusion MRI (Super MUDI) dataset [16] contains the data of four healthy human subjects with ages range between 19 and 46 years. For each subject 1,344 MRI volumes are provided. The imaging device was clinical 3T Philips Achieva Scanner (Best, Netherlands) with a 32-channel adult head coil.

The Super MUDI Challenge comprises two tasks: isotropic, and anisotropic super-resolution. The names of these tasks were derived from the acquisition strategies of the low-resolution MRI data. The objective of using two down-sampling strategies is to compare the combinations of the down-sampling methods and the super-resolution approaches that can best to be used in a clinical scheme to obtain
simulated high-quality and high-fidelity MRI images while reducing the acquisition time. In the anisotropic subsampling the volume has high in-plane resolution (2.5 mm × 2.5 mm), but thick axial slice (5 mm), while in the isotropic subsampling the volume has low resolution (5 mm) in all the directions. For our experiments, we use one subject each for training and validation, and two for testing.

The preprocessing of the data includes data normalization to the range [0, 1]. We trained on the whole LR volume of size (28 × 46 × 38) for the isotropic task and on a LR patch of size (28 × 70 × 60) for the anisotropic task due to memory limitation. We tested the networks on the whole volume for both tasks. When not including the perceptual loss PL, we trained on a single 12GB GeForce RTX 2080 GPU, else we trained on two 12GB GeForce RTX 2080 GPUs. Additionally, we trained each network for 100 epochs using Adam optimizer with an initial learning rate 10^{-4}. We made the code publicly available at [15].

Our dataset contains 23 T1 volumes and the corresponding FLAIR volumes of 23 subjects from the dataset [17] and 250 diffusion volumes randomly selected from Super MUDI dataset [16]. As our dataset was very imbalanced, we trained the network using class-balanced softmax cross-entropy loss [18], described as follows:

\[ CB_{\text{softmax}}(z, y) = - \frac{1 - \beta}{1 - \beta_{n_y}} \log \left( \frac{\exp(z_y)}{\sum_{j=1}^{C} \exp(z_j)} \right), \quad (2) \]

where \( z = [z_1, z_2, ..., z_C]^T \) is the prediction of the model, \( C \) is the total number of classes, \( n_y \) is the total number of training samples corresponding to class \( y \), \( \beta \in [0, 1) \) is a hyper parameter and we found that the value 0.99 works the best for our problem.

V. RESULTS

Our method (SRResNet+PD+PL) outperformed bicubic interpolation (winner of the Super MUDI challenge) as well as two state-of-the-art SR techniques to which we compared – DCED [19] and 2D-ESRGAN [4] – for both isotropic and anisotropic SR of MRI images. Qualitative results are shown in Figure 1 and quantitative ones in Table I. DCED uses a convolutional encoder-decoder architecture without the GAN framework, unlike our method. 2D-ESRGAN uses a GAN framework, however it operates on 2D slices of the image as opposed to 3D volumes.

We also performed an ablation study with various components introduced in Section III. Projection discriminator (PD) alone was not sufficient to guarantee improved results. Adding the perceptual loss (PL) to train the combined architecture (SRResNet+PD) improved the results for all the evaluation metrics by an appreciable extent for both isotropic and anisotropic super-resolution. Finally, the use of 2D architectures with interpolation creates unwanted artifacts in the slices, which is evident in the Figure 1 for 2D-ESRGAN.

VI. CONCLUSION

We proposed a perceptual cGAN for 3D MRI super-resolution, which outperformed bicubic interpolation (the classical winning algorithm of the Super MUDI challenge) for reconstructing HR images from downsampled LR images. Though our testing data is not the same testing data used for evaluating the algorithms on the SuperMUDI challenge leaderboard, our results seem to be comparable to the winning models.

We showed the benefit of adding the perceptual loss (PL) to projection discriminator and 3-D SRResNet generator for recovering the finer details and producing perceptually more plausible images. We found that the perceptual and adversarial losses preserve the semantic information (structure and content), while per-pixel losses such as MSE preserve the style information (color, texture, and common patterns). Thus, training the generator on a combination of MSE, adversarial loss, and PL drives the generator to focus on the semantic information at the cost of the style information. However, the projection discriminator penalizes the differences in style to preserve the stylistic features, and forms a good combination for the PL that focuses on semantic information.

SR strategies that can reconstruct clinically-relevant details in HR images are very promising for reducing MRI acquisition time. The clinical validity of the details generated by SR methods for medical images remains to be further studied, as the popular quantitative metrics – PSNR and SSIM – are clinically agnostic.

In our future work, we are planning to study the risks of a GAN adding high-frequency details for super-resolution.

VII. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available by the organizers of the SuperMUDI challenge [16]. Ethical approval was not required as confirmed by the license attached with the data.

REFERENCES

[1] K. Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2015.
[2] J. Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in ECCV, 2016.
Fig. 1. Visualization of the results of the isotropic and the anisotropic tasks. Res. stands for the residual image between the high resolution and the reconstructed images.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Image super-resolution using deep convolutional networks,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp.295 –307, 2015.

[4] Zhang Hongtao, Yuki Shinomiya, and Shinichi Yoshida, “3d brain mri reconstruction based on 2d super-resolution technology,” in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2020, pp. 18–23.

[5] Haoyu Lan, Arthur W. Toga and Farshid Sepehrband. “Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis.” Magnetic Resonance in Medicine 86 (2021): 1718 - 1733.

[6] Chao Dong, Chen Change Loy, and Xiaoou Tang, “Accelerating the super-resolution convolutional neural network,” in European conference on computer vision. Springer, 2016, pp. 391–407.

[7] Lin, Hongxiang, et al. "Generalised Super Resolution for Quantitative MRI Using Self-supervised Mixture of Experts." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2021.

[8] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.

[9] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao, “Image super-resolution using dense skip connections,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4799–4807.

[10] Yulu Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, “Image super-resolution using very deep residual channel attention networks,” 2018.

[11] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, “Esrgan: Enhanced super-resolution generative adversarial networks,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 0–0.

[12] Takero Miyato and Masanori Koyama, “cgans with projection discriminator,” in 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. 2018, OpenReview.net.

[13] Mehdi Mirza and Simon Osindero, “Conditional generative adversarial nets,” ArXiv, vol. abs/1411.1784, 2014.

[14] Andrew Aitken, Christian Ledig, Lucas Theis, Jose Caballero, Zehan Wang, and Wenzhe Shi, “Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and resize resolution,” arXiv preprint arXiv:1707.02937, 2017.

[15] https://github.com/SaharAlmahfouzNasser/MRI-SuperResolution.

[16] Marco Pizzolato, Marco Palombo, Jana Hutter, Vishwesh Nash, Fan Zhang, and Noemi Gyori, “Super-resolution of Multi Dimensional Diffusion MRI data,” Mar. 2020.

[17] Yiming Xiao, Maryse Fortin, Geirmund Unsgard, Hassan Rivaz, and Ingerid Reinertsen, “Re-rospective evaluation of cerebral tumors (resect): A clinical database of pre-operative mri and intra-operative ultrasound in lowgrade glioma surgeries,” Medical physics, vol. 44, no. 7, pp. 3875–3882, 2017.

[18] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie, “Class-balanced loss based on effective number of samples,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9268–9277.

[19] Jinglong Du, Lulu Wang, Yulu Liu, Zexun Zhou, Zhongshi He, and Yuanyuan Jia, “Brain mri super-resolution using 3d dilated convolutional encoder-decoder network,” IEEE Access, vol. 8, pp. 18938–18950, 2020.