Improving High-Frequency Details in Cerebellum for Brain MRI Super-Resolution

1st Hanzhi Wang
School of Computer Science
Cardiff University
Cardiff, United Kingdom
wangh69@cardiff.ac.uk

2nd David Marshall
School of Computer Science
Cardiff University
Cardiff, United Kingdom
marshallad@cardiff.ac.uk

3rd Derek K. Jones
School of Psychology
Cardiff University
Cardiff, United Kingdom
jonesd27@cardiff.ac.uk

4th Yuhua Li
School of Computer Science
Cardiff University
Cardiff, United Kingdom
liy180@cardiff.ac.uk

***Abstract*—Deep-learning-based single-image super-resolution models are typically trained using image patches, rather than the whole images, due to hardware limits. Since different brain regions have disparate structures and their size varies, such as the cerebrum and cerebellum, models trained using image patches can be dominated by the structures of the larger brain regions and ignore the fine-grained details in smaller areas. In this paper, we first evaluate several previously proposed models using more blurry low-resolution images than previous studies, as input. Then, we propose an effective approach for the conventional patch-based strategy by balancing the proportion of patches containing high-frequency details. This makes the model focus more on high-frequency information in tiny regions, especially for the cerebellum. Compared with the conventional patch-based strategy, the resultant super-resolved image from our approach achieves comparable image quality in the whole brain. In contrast, it improves significantly on the high-frequency details in the cerebellum.**

***Index Terms*—cerebellum, MRI, super-resolution**

I. Introduction

High-resolution (HR) structural MR images contain rich anatomical details, which is preferred in clinical diagnosis. However, HR images come at the cost of longer scan time and lower signal-to-noise ratio [1], which poses obstacles in the clinical situation due to hardware limits. One solution for this problem is to apply the single image super-resolution (SISR) technique, which requires a low-resolution (LR) image to reconstruct its HR counterpart. It does not require extra scan time or high-cost scanners to generate an HR image and could be used to enhance image quality from low-field scanners.

Training SISR models requires paired LR and HR images for models to learn mappings between images. Due to the rare availability of paired MR images, synthetic LR images, generated from their HR counterpart, have been widely adopted in the literature. To mimic the behaviors of the real-world low-field scanners, HR images are transformed into the k-space using the Fast Fourier Transform (FFT), and resolution degradation is performed on the k-space data [2].

The work of Hanzhi Wang was supported by the China Scholarship Council (CSC) for his PhD (No. 202008060053). The work of Derek K. Jones was supported in part by the Wellcome Trust Investigator Award 096646/Z/11/Z and in part by the Wellcome Trust Strategic Award 104943/Z/14/Z.

In previous studies, synthetic LR images are similar to HR images, even if only keeping the central 25% k-space data [2]–[5], which may face performance degradation on real-world LR images. We first investigate whether these previously proposed models are capable of learning more complex mappings when using more blurry LR images as input. We follow the procedures in generating LR images [2], while only keeping 6.25% of the k-space data. By keeping a smaller proportion of the original k-space data, a more significant difference exists between synthetic LR and authentic HR images.

A more fundamental problem in brain super-resolution is the patch-based training strategy. Typically, image patches sampled uniformly from the whole brain volume, rather than the whole images, are adopted as model input to prevent memory issues [2]–[5]. However, such uniform sampling introduces a data imbalance issue. The cerebrum, containing low-frequency information, generates far more patches than the cerebellum, whereas the latter contains more complex structures. It could result in the model being more focused on the information from larger regions of the brain and ignoring the fine-grained details in tiny areas. In the literature, the reconstruction of the cerebrum has achieved almost indistinguishable performance using relatively simple models [2]. However, there is still a huge gap between the super-resolved cerebellum and the authentic HR one, even using decent models [6], [7].

To solve the data-imbalance issue, we propose a simple yet effective approach via a non-uniform patch sampling for the conventional patch-based training strategy, to treat the whole brain volume and cerebellum equally to derive better reconstruction quality.

Our main contributions in this paper are:

1) We evaluate the capacity of several renowned models using more blurry LR images as input than previous studies. By keeping 6.25% of the k-space data rather than 25% in the literature, synthetic LR images look more similar to real-world LR ones, which may improve model generalization ability on real-world images.
2) We propose a direct and effective method using a non-uniform sampling for the patch-based training strategy to improve the reconstruction quality for high-frequency details, especially for the cerebellum. To the best of

our knowledge, the proposed method is the first work that treats the brain volume as two separate regions, the cerebellum and non-cerebellum regions, for brain MRI super-resolution tasks. The implementations are available in the GitHub repository [1].

The paper is organized as follows. Section II reviews the development of SISR on natural and MR images. Section III introduces our proposed approach for improving the high-frequency details for brain super-resolution. Section IV explains the settings of the experiments, including the dataset and models. Section V demonstrates quantitative experiment results. Section VI provides a use case of the proposed approach using brain age prediction. Section VII summarizes the overall approach to conclude this paper.

II. Related work

A. Deep-learning-based super-resolution

Super-resolution (SR) using deep learning has experienced rapid progress over the past decade. Dong *et al.* proposed SRCNN in 2014, establishing the basic structure for SR models [8]. Kim *et al.* proposed VDSR that learns the residual information between LR and HR images [9], reducing the learning difficulty. Motivated by the ResNet [10], ResNet-based SR models have been proposed [11], [12], achieving superior performance using the residual connections. Inspired by the DenseNet [13], Tong *et al.* applied dense blocks and added dense connections between different blocks to build the SRDenseNet [14]. Zhang *et al.* took advantage of both residual and dense connection, proposing a residual dense network to further improve the reconstruction quality [15]. With the help of the Transformer models [16], Lu *et al.* proposed ESRT, which consists of CNN and transformer backbones, achieving competitive results with low computational costs [17].

Generative adversarial network (GAN) [18] has also been applied to SR problems. Ledig *et al.* proposed SRGAN, generating more realistic images compared with CNN-based models [12]. Bell-Kligler *et al.* introduced an unsupervised model named KernelGAN [19], using real LR images for model training. In addition, various model designs, such as the learning-based upsampling [20] and recursive learning [21] have also been proposed for SR.

B. Super-resolution on MR images

Structural MR images contain four dimensions, increasing the model complexity and learning difficulty. Therefore, models using 2D slices or 3D volumes have both been explored.

2D models are more stable and faster to train, whereas they normally fail to learn the information from the third dimension. Zhao *et al.* introduced a channel splitting block to incorporate different information from different receptive fields to increase the model learning capacity [22]. Du *et al.* built a 2D CNN with residual connections for SR reconstruction of single anisotropic MR images [23]. Lyu *et al.* established an ensemble learning framework using 2D GANs to integrate complementary SR results from each GAN model [24].

On the contrary, 3D models can utilize the additional information from the volumetric images and outperform the 2D models. Pham *et al.* first demonstrated that for brain MRI, 3D models outperform 2D models by a large margin [25]. Chen *et al.* applied the dense connection structure and proposed the 3D mDCSRN and mDCSRN-WGAN models [2], which contain relatively few parameters. Li *et al.* proposed a lightweight network, called VolumeNet, using separable 2D cross-channel convolutions [5], which has fewer parameters but still achieves superior performance. Zhang *et al.* adopted the squeeze and excitation architecture [26] and attention mechanism to make the model learn from the more informative regions of the brain [6]. With the help of the Transformer models [16], Li *et al.* proposed a multi-scale contextual matching method that can learn information from different scales [7]. Zhou *et al.* applied a domain correction network that can super-resolve unpaired MR images [27].

There are also lots of works focusing on different aspects of MRI super-resolution, such as models handling multi-scale LR images [28], [29], image quality transfer [30]–[32], models trained using real-world LR images [32], [33].

Existing works on MRI super-resolution have achieved superior performance. However, even the most advanced models, such as attention-based [6] or transformer-based model [7], fail to handle the whole brain volume and cerebellum altogether. The attention mechanism applied in such models [6], [7] offers an indirect way of learning from more informative areas, whereas the performance is poor on the cerebellum. The non-uniform sampling proposed in this paper provides a naive yet straightforward way to solve this problem efficiently.

III. Method

A. Non-uniform sampling for brain image super-resolution

As we previously mentioned in Section I, the conventional uniform patch sampling results in a data imbalance issue between the over-represented region (cerebrum) and under-represented region (cerebellum), whereas the latter has more complex details and requires much more data to learn.

Ideally, a second model can be trained using the cerebellum images, which is expected to improve results on the cerebellum. However, it requires additional training resources and introduces problems when merging the two resultant images.

Therefore, a unified approach can be proposed by balancing the number of patches from the cerebrum and cerebellum during the training process. Such non-uniform patch sampling makes the model focus more on under-represented regions. To keep the overall training time unchanged, the number of generated patches stays unchanged, whereas more patches come from the cerebellum compared with conventional patch sampling.

The detailed steps can be summarized as:

1) Extract the cerebellum from the whole brain volume to create a separate cerebellum dataset.

[1] https://github.com/anonymousreviewonly/MRI_SR

2) Generate patches from the whole brain volume and randomly select 50% of the generated patches.
3) Generate patches from the cerebellum dataset until the number of patches is equal to the whole brain volume patches.
4) Combine the whole brain and cerebellum patches randomly for training.
5) Repeat step 2, 3, and 4 at the beginning of each epoch until the end of training.

The intuition of our approach is straightforward. Since the brainstem contains few patterns, the brain super-resolution focuses more on the cerebrum and cerebellum. Since the cerebrum and cerebellum contain disparate structures, it is natural to treat the whole volume super-resolution as two sub-volume super-resolution tasks. Therefore, by keeping a 50/50 distribution of image patches, the model becomes less biased towards the cerebrum, which helps the reconstruction of the cerebellum. Theoretically speaking, such non-uniform sampling might cause performance degradation in the cerebrum, as less information is used for training. However, in Section V, we will demonstrate that our modification results in negligible difference compared with conventional uniform sampling.

IV. Experiments

A. Dataset and data preprocessing

To fairly compare different models and strategies, we chose a public brain structural MRI database, the Human Connectome Project (HCP) [34]. It contains 3D T1-weighted MR images from 1113 subjects acquired via the Siemens 3T platform using a 32-channel head coil on multiple centers. The images come in high spatial resolution as 0.7 mm isotropic in a matrix size of $320 \times 320 \times 256$. These high-quality images serve as the authentic HR images in the following experiments. We randomly chose 800 images as the training set, 100 as the validation set, and the rest as the test set. Brain Extraction Tool (BET) [35] is further performed on the HR images to delete non-brain tissues.

To generate corresponding LR images, we followed the same procedures demonstrated in [2]:

1) Applying the FFT to authentic HR images to convert the original image into the k-space data.
2) Masking (zeroing) outer part of k-space data along two axes representing two MR phase encoding directions.
3) Applying the inverse FFT to the masked k-space data to generate the synthetic LR images.

The LR image synthesis procedure mimics the real MR image acquisition process where a low-resolution MRI is scanned by reducing acquisition lines in phase and slice encoding directions. The synthetic LR images have the same size as HR images.

A hyperparameter, named "scale factor", is introduced in the LR image synthesis process, which determines the proportion of masked k-space data. A larger scale factor masks more k-space data, resulting in more blurry images. Using a scale factor of 2 is common in many studies [2], [4], [5], whereas, in this paper, we set the scale factor to 4, preserving the central 6.25% ($1/4 \times 1/4$) k-space data, to evaluate models on more blurry LR images. An image comparison in different scale factors is shown in Figure 1.

To generate a separate cerebellum dataset for the proposed non-uniform sampling, described in Section III-A, FastSurfer [36], a deep-learning-based brain segmentation tool [2], is adopted. FastSurfer automatically labels the cerebellum regions for extraction. The cerebellum extraction pipeline is demonstrated in Figure 2. It should be stressed that segmentation masks are generated only on HR images, and the corresponding masks are applied on both HR and LR images to ensure both HR and LR cerebellums are voxelwise paired. A diagram of the overall image preprocessing is shown in Figure 3.

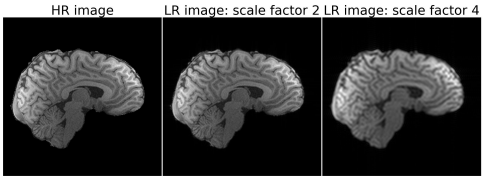


Fig. 1. MR images in different scale factors. From left to right, it shows the HR image, and LR image using a scale factor of 2 and 4. For scale factors 2 and 4, the central 25% and 6.25% k-space data are preserved.

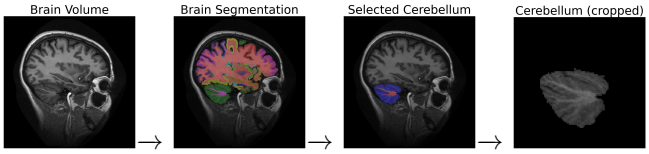


Fig. 2. An illustration of the cerebellum extraction pipeline. (a) represents the authentic HR image. (b) is obtained by applying FastSurfer to (a). (c) is obtained by selecting the cerebellum-related mask of (b) and (d) is generated by applying the cerebellum mask on (a) and removing the useless background. FastSurfer labels the cerebellum automatically for delineation.

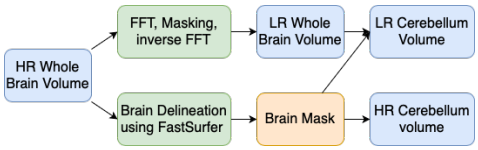


Fig. 3. A diagram of MR image preprocessing.

B. Models

We chose two widely recognized models in this field, mDCSRN and mDCSRN-WGAN [2] to evaluate the proposed approach. We chose these two models since they both adopt the DenseNet [13] as the backbone, which is still popular in various model designs. Also, they represent two distinct model designs, CNN-based and GAN-based models.

It should be noted that there are no limitations in the choice of model architectures since the proposed approach focuses

[2]https://github.com/Deep-MI/FastSurfer

primarily on data sampling. We chose DenseNet-based models for their stability and effectiveness. The mDCSRN model adopts a DenseNet [13] architecture with 4 dense blocks. The mDCSRN-WGAN model uses the same architecture of mDCSRN as the generator and uses the Wasserstein GAN [37] to guide the model training.

C. Training and testing

In the training process, Adam optimizer [38] is used as the default optimizer for all models. The initial learning rate is set to 0.001 and then multiplied by 0.5 every 10 epochs. The batch size is set to 64 and the total number of epochs is set to 50. We use the L1 loss for mDCSRN and the generator of the mDCSRN-WGAN model since L1 loss tends to generate more realistic images than L2 loss [39]. The size of image patches is set to $32 \times 32 \times 32$ for all patches.

To evaluate the super-resolution performance, we used the structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and normalized root mean squared error (NRMSE) to measure the similarity between super-resolved (SR) images and HR images. SSIM uses the mean, variance, and covariance to estimate the similarity of the two images. PSNR is used to further quantify the recovered image quality using the mean squared loss. NRMSE is a more direct way to measure the pixel-wise similarity between the original and super-resolved images. In general, lower NRMSE, higher PSNR, and higher SSIM values represent better super-resolution results. Although these metrics have been criticized that a high score does not represent a better image quality [40], they are the most commonly used metrics, and no other perceptual quality metrics for 3D MR images have been applied in the literature.

V. Results

In this section, we first demonstrate the SISR model performance using synthetic LR images, generated under a scale factor of 4. Then, we compare the model performance between the conventional patch sampling and the proposed non-uniform sampling. Due to space constraints, we only include sample output images from the mDCSRN model. When conducting the significance test, the paired t-test is adopted by default.

A. Model performance under a scale factor of 4

Figure 4 demonstrates the model performance of the whole brain volume using the mDCSRN model from a sagittal view. It can be observed that the SR image is similar to the HR image in most regions of the brain, especially in the cerebrum. It indicates that relatively simple models, such as mDCSRN, are still capable of handling more blurry patterns, especially in the cerebrum region. The quantitative analysis is demonstrated in Table I. From Table I, both mDCSRN and mDCSRN-WGAN significantly improve the image quality in all three metrics compared with the baseline ($p < 0.01$).

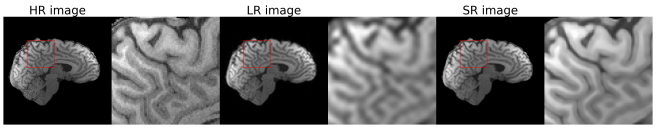


Fig. 4. HR, LR, and SR brain image from sagittal view using the mDCSRN model.

TABLE I
MODEL PERFORMANCE OF THE WHOLE BRAIN VOLUME

Model	Image Pair	PSNR	SSIM	NRMSE
None (Baseline)	HR & LR	31.33 ± 2.92	0.92 ± 0.01	0.148 ± 0.09
mDCSRN	HR & SR	34.23 ± 2.14	0.95 ± 0.01	0.119 ± 0.06
mDCSRN -WGAN	HR & SR	34.19 ± 2.11	0.95 ± 0.02	0.121 ± 0.07

B. Comparisons between conventional and proposed sampling

Table II demonstrates the quantitative analysis of model performance for mDCSRN and mDCSRN-WGAN. For each model, we compare the conventional sampling and the proposed approach on the whole brain volume and the cerebellum respectively. We also add the PSNR, SSIM, and NRMSE values between the HR and LR images as the baseline to demonstrate the improvement of our approach.

From Table II, when considering the reconstruction quality of the whole brain volume for both mDCSRN and mDCSRN-WGAN model, the conventional patch-based approach achieves comparable performance with our sampling approach in all three metrics. There are no significant differences in all three metrics when evaluated on the whole brain volume between the two sampling approaches, whereas both approaches outperform the baseline performance significantly ($p < 0.05$). When evaluated on the cerebellum, our proposed sampling approach significantly outperforms the conventional sampling approach for both mDCSRN and mDCSRN-WGAN models in all three metrics ($p < 0.05$).

The reason for the improved performance on the cerebellum is clear, as we have increased the proportion of image patches from the cerebellum for models to learn. The potential reason for the comparable performance on the whole brain volume between the conventional and the proposed approach is that the number of patches from the whole brain is redundant. Reducing the proportion of patches from the over-represented regions does not affect model performance significantly. Also, at the beginning of each epoch, random patches will be selected until the end of training. Therefore, the model is still able to learn information from all whole brain patches.

Figure 5 and Figure 6 demonstrate the super-resolution performance between the conventional and the proposed approach on both whole brain volume and cerebellum respectively. It can be observed that our approach achieves almost the same performance in the whole brain volume, whereas it reconstructs more details in the cerebellum.

Therefore, compared with the conventional sampling, the

proposed non-uniform sampling approach achieves comparable performance on the whole brain volume, whereas significantly improves the reconstruction quality for the fine structures in the cerebellum.

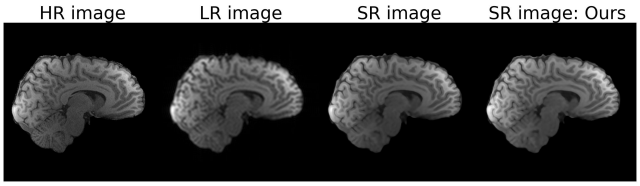


Fig. 5. HR, LR, SR (conventional sampling), and SR (proposed sampling) whole brain volume illustrations from sagittal view by mDCSRN model.

VI. Evaluation via Brain Age Prediction

To further validate the proposed sampling approach, we provide an example use case of super-resolved images via brain age prediction. Brain age prediction has attracted interest in the past few years [41], [42], which predicts a marker measuring brain health. It takes brain MR images as input and uses the chronological ages of participants as output.

A. Experiment settings

DenseNet architecture is selected for brain age prediction considering its performance and speed and the HCP dataset is also adopted for brain age prediction. Then, the HR images are used to generate corresponding LR images, as discussed in Section IV-A.

Afterward, the generated LR images are super-resolved using the mDCSRN models trained with conventional sampling and the proposed sampling approach respectively, generating two super-resolved images for each LR image. Therefore, two super-resolved MR image datasets are created. One adopts the conventional uniform sampling and the other uses the proposed non-uniform sampling.

The two SR image datasets are used to train the brain age prediction model respectively, resulting in two brain age prediction models, each dedicated to a sampling strategy. Two more brain age prediction models are also trained using original HR images and LR images respectively as the contrast. In total, four brain age prediction models are trained using HR images, LR images, SR images (conventional approach), and SR images (our approach) respectively. To decrease the variations in the model training process, each model is trained five times. The Mean Absolute Error (MAE) is adopted as the evaluation metric.

B. Results

Results in Table III demonstrate the model performance of brain age prediction using HR images, LR images, SR images from conventional patch sampling, and SR images from our approach respectively.

It can be observed that SR images from our approaches achieve comparable predictive accuracy compared with SR images from conventional patch sampling, and both approaches surpass LR images significantly ($p < 0.05$). The HR images show slightly better performance compared with super-resolved images from both approaches, whereas the difference is not significant.

The model performance is expected. Because, in brain age prediction, the model predicts age based on the whole brain volume rather than the cerebellum. The improvements in the cerebellum, discussed in Section V, are likely to be ignored by the model. However, it also validates that our approach achieves almost the same reconstruction quality on the whole brain volume, compared with the conventional approach. Therefore, we have validated that the super-resolved MR images derived from the proposed non-uniform sampling strategy can be applied for subsequent clinical analysis and it may demonstrate more significant improvements in some cerebellum-specific analysis.

VII. Conclusion

We have proposed a simple yet effective approach, via a non-uniform sampling strategy, for the conventional patch-based training in brain MRI super-resolution. We provide a different perspective to treat brain structures differently in super-resolution. By maintaining a 50/50 distribution of patches, our approach achieves comparable performance on the whole brain volume, whereas significantly surpasses the conventional approach on the cerebellum. Our approach restores more high-frequency details in the cerebellum.

We have demonstrated the capacity of two renowned models using more blurry LR images as input in Section V. In the literature, a scale factor of 2 has been widely adopted to generate LR images. In this paper, we select a scale factor of 4 to generate LR images and the models are still capable of learning the mappings between LR and HR images.

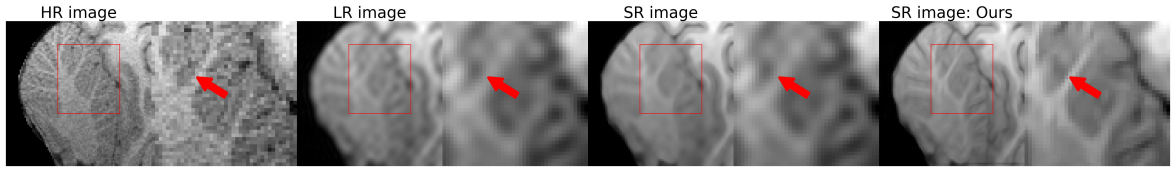
We have shown that our non-uniform sampling strategy across the whole brain volume does not result in significant reconstruction degradation in the non-cerebellum regions. The most probable reason behind this is that in the conventional approach, the number of patches from the over-represented regions is already redundant, and non-cerebellum regions contain simpler patterns. As we demonstrated in Figure 4 and Table II, the mDCSRN model can achieve indistinguishable reconstruction quality over non-cerebellum regions.

We have provided a use case for MRI super-resolution via brain age prediction. From Table III, the super-resolved MR images demonstrate significantly better predictive accuracy in predicting brain age compared with LR images ($p < 0.05$). Although the HR images achieve the highest accuracy, the difference between HR images and super-resolved images is not significant. In terms of the two sampling approaches, since the proposed approach focuses primarily on the cerebellum, there is no significant difference between the two approaches for age prediction using the whole brain volume.

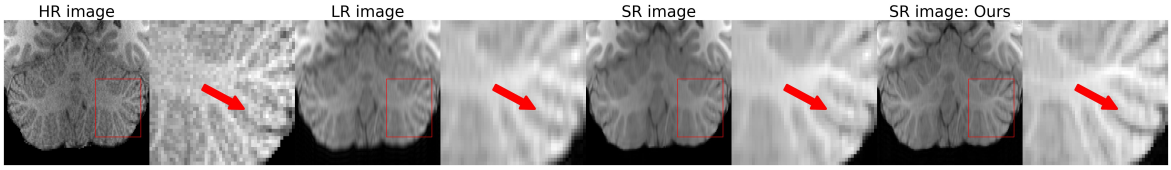
Some aspects still require further improvements. First, although our approach does not depend on model architectures, it could be worthwhile to adopt more advanced architectures.

TABLE II
MODEL PERFORMANCE ON THE WHOLE BRAIN VOLUME AND CEREBELLUM

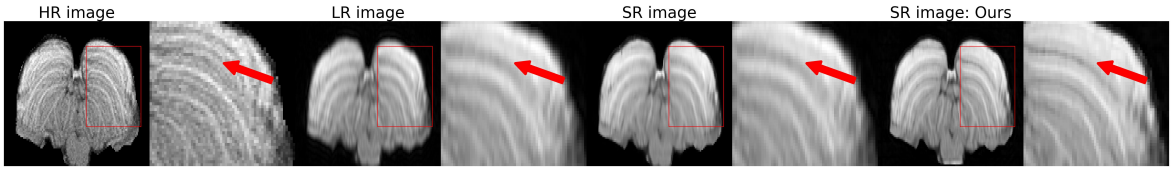
Brain Region Evaluated	Model	Image Pair	PSNR	SSIM	NRMSE
Whole Brain Volume	None (baseline)	HR & LR (baseline)	31.33 ± 2.92	0.92 ± 0.01	0.148 ± 0.09
	mDCSRN	HR & SR (conventional sampling approach)	34.23 ± 2.14	0.95 ± 0.01	0.119 ± 0.06
		HR & SR (our approach)	34.18 ± 1.98	0.95 ± 0.02	0.120 ± 0.05
	mDCSRN-WGAN	HR & SR (conventional sampling approach)	34.19 ± 2.11	0.95 ± 0.02	0.121 ± 0.07
		HR & SR (our approach)	34.13 ± 2.17	0.95 ± 0.03	0.122 ± 0.05
Cerebellum	None (baseline)	HR & LR (baseline)	27.37 ± 2.66	0.88 ± 0.01	0.15 ± 0.10
	mDCSRN	HR & SR (conventional sampling approach)	28.89 ± 1.81	0.905 ± 0.01	0.135 ± 0.06
		HR & SR (our approach)	29.62 ± 1.55	0.93 ± 0.01	0.127 ± 0.02
	mDCSRN-WGAN	HR & SR (conventional sampling approach)	28.91 ± 1.73	0.902 ± 0.02	0.139 ± 0.04
		HR & SR (our approach)	29.47 ± 1.81	0.918 ± 0.02	0.124 ± 0.05



(a)



(b)



(c)

Fig. 6. HR, LR, SR (conventional sampling), and SR (proposed sampling) cerebellum image by mDCSRN. (a), (b), and (c) show the sagittal, coronal, and axial view.

TABLE III
MODEL PERFORMANCE OF BRAIN AGE PREDICTION

Images used for training models	MAE
HR images	**2.25 ± 0.27**
SR images (conventional sampling)	2.38 ± 0.31
SR images (our approach)	2.33 ± 0.34
LR images	3.01 ± 0.28

Second, the proposed sampling strategy is mainly based on intuitions. Given that the brain mainly contains the cerebrum, cerebellum, and brainstem, whereas the latter contains very few structures, it is reasonable to treat the whole brain volume into two sub-volume SR problems and apply a 50/50 data distribution for the cerebellum and non-cerebellum regions. However, more sophisticated sampling could be explored to evaluate the optimal combination of the two regions.

To conclude, we propose a more effective sampling method for the conventional patch-based training strategy for brain MRI super-resolution problems. Using a non-uniform sampling across the whole brain volume, the model tends to learn more high-frequency details from the cerebellum, rather than being focused on non-cerebellum regions due to data imbalance. With the help of the proposed sampling approach, we improve the model performance on the cerebellum significantly and also achieve comparable performance on the whole brain volume.

REFERENCES

[1] E. Plenge, D. H. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, L. van der Weerd, W. J. Niessen, and E. Meijering, "Super-resolution methods in mri: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?" *Magnetic resonance in medicine*, vol. 68, no. 6, pp. 1983–1993, 2012.

[2] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 91–99.

[3] Q. Lyu, H. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, and G. Wang, "Multi-contrast super-resolution mri through a progressive network,"

IEEE transactions on medical imaging, vol. 39, no. 9, pp. 2738–2749, 2020.

[4] J. Wang, Y. Chen, Y. Wu, J. Shi, and J. Gee, "Enhanced generative adversarial network for 3d brain mri super-resolution," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3627–3636.

[5] Y. Li, Y. Iwamoto, L. Lin, R. Xu, R. Tong, and Y.-W. Chen, "Volumenet: A lightweight parallel network for super-resolution of mr and ct volumetric data," *IEEE Transactions on Image Processing*, vol. 30, pp. 4840–4854, 2021.

[6] Y. Zhang, K. Li, K. Li, and Y. Fu, "Mr image super-resolution with squeeze and excitation reasoning attention network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 425–13 434.

[7] G. Li, J. Lv, Y. Tian, Q. Dou, C. Wang, C. Xu, and J. Qin, "Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 636–20 645.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[9] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[12] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[14] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4799–4807.

[15] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, "Transformer for single image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 457–466.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[19] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-gan," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[20] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3262–3271.

[21] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 11, pp. 2599–2613, 2018.

[22] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, "Channel splitting network for single mr image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5649–5662, 2019.

[23] J. Du, Z. He, L. Wang, A. Gholipour, Z. Zhou, D. Chen, and Y. Jia, "Super-resolution reconstruction of single anisotropic 3d mr images using residual convolutional neural network," *Neurocomputing*, vol. 392, pp. 209–220, 2020.

[24] Q. Lyu, H. Shan, and G. Wang, "Mri super-resolution with ensemble learning and complementary priors," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 615–624, 2020.

[25] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain mri super-resolution using deep 3d convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 197–200.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[27] H. Zhou, Y. Huang, Y. Li, Y. Zhou, and Y. Zheng, "Blind super-resolution of 3d mri via unsupervised domain transformation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1409–1418, 2022.

[28] C.-H. Pham, C. Tor-Díez, H. Meunier, N. Bednarek, R. Fablet, N. Passat, and F. Rousseau, "Multiscale brain mri super-resolution using deep 3d convolutional networks," *Computerized Medical Imaging and Graphics*, vol. 77, p. 101647, 2019.

[29] S. Dong, G. Hangel, W. Bogner, G. Widhalm, K. Rössler, S. Trattnig, C. You, R. de Graaf, J. A. Onofrey, and J. S. Duncan, "Multi-scale super-resolution magnetic resonance spectroscopic imaging with adjustable sharpness," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 410–420.

[30] D. C. Alexander, D. Zikic, A. Ghosh, R. Tanno, V. Wottschel, J. Zhang, E. Kaden, T. B. Dyrby, S. N. Sotiropoulos, H. Zhang *et al.*, "Image quality transfer and applications in diffusion mri," *NeuroImage*, vol. 152, pp. 283–298, 2017.

[31] R. Tanno, D. E. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion mri," *NeuroImage*, vol. 225, p. 117366, 2021.

[32] H. Lin, M. Figini, F. D'Arco, G. Ogbole, R. Tanno, S. B. Blumberg, L. Ronan, B. J. Brown, D. W. Carmichael, I. Lagunju *et al.*, "Low-field magnetic resonance image enhancement via stochastic image quality transfer," *Medical Image Analysis*, vol. 87, p. 102807, 2023.

[33] V. Lau, L. Xiao, Y. Zhao, S. Su, Y. Ding, C. Man, X. Wang, A. Tsang, P. Cao, G. K. Lau *et al.*, "Pushing the limits of low-cost ultralow-field mri by dual-acquisition deep learning 3d superresolution," *Magnetic Resonance in Medicine*, 2023.

[34] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[35] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[36] L. Henschel, S. Conjeti, S. Estrada, K. Diers, B. Fischl, and M. Reuter, "Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline," *NeuroImage*, vol. 219, p. 117012, 2020.

[37] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[39] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[40] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.

[41] H. Wang, M. S. Treder, D. Marshall, D. K. Jones, and Y. Li, "A skewed loss function for correcting predictive bias in brain age prediction," *IEEE transactions on medical imaging*, vol. 42, no. 6, pp. 1577–1589, 2023.

[42] D. A. Wood, S. Kafiabadi, A. Al Busaidi, E. Guilhem, A. Montvila, J. Lynch, M. Townend, S. Agarwal, A. Mazumder, G. J. Barker *et al.*, "Accurate brain-age models for routine clinical mri examinations," *NeuroImage*, 2022.