

Received February 15, 2022, accepted February 24, 2022, date of publication February 28, 2022, date of current version March 30, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3155226

# Super-Resolution Reconstruction of 3T-Like Images From 0.35T MRI Using a Hybrid Attention Residual Network

JIALIANG JIANG<sup>1</sup>, FULANG QI<sup>1</sup>, HUIYU DU<sup>1</sup>, JIANAN XU<sup>1</sup>, YUFU ZHOU<sup>1</sup>, DAYONG GAO<sup>2</sup>, AND BENSHEG QIU<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Center for Biomedical Imaging, University of Science and Technology of China, Hefei, Anhui 230026, China

<sup>2</sup>Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA

Corresponding authors: Bensheng Qiu (bqiu@ustc.edu.cn) and Dayong Gao (dayong@uw.edu)

This work was supported in part by the National Natural Science Foundation of China under Grant 91859121 and Grant 81627806, and in part by the Fundamental Research Funds for the Central Universities under Grant WK5290000001 and Grant WK5290000002.


This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the First Affiliated Hospital of University of Science and Technology of China, under Application No. 2021 KY205, and performed in line with the Declaration of Helsinki.

**ABSTRACT** Magnetic resonance (MR) images from low-field scanners present poorer signal-to-noise ratios (SNRs) than those from high-field scanners at the same spatial resolution. To obtain a clinically acceptable SNR, radiologists operating the low-field scanners use a much smaller acquisition matrix than high-field scanners. Thus, the current state of the image quality indicates the need for further research to improve the image quality of low-field systems. Strategies based on super-resolution (SR) techniques can be alternatives for image reconstruction. However, predetermined degradation methods embedded in these techniques, such as bicubic downsampling, seem to impose a performance drop when the actual degradation is different from the pre-defined assumption. In this study, we collected a unique dataset by scanning 70 participants to address this problem. The anatomical locations of the scanned image slices were the same for 0.35T and 3T data. Low-resolution (LR) images (0.35T) and high-resolution (HR) images (3T) were the image pairs used for data training. Herein, we introduce a novel CNN-based network with hybrid attention mechanisms (HybridAttentionResNet, HARN) to adaptively capture diverse information and reconstruct super-resolution 0.35T MR images (3T-like MR images). Specifically, the proposed dense block combines variant dense blocks and attention blocks to extract abundant features from LR images. The experimental results demonstrate that our proposed residual network efficiently recovers significant textures while rendering a high peak signal-to-noise ratio (PSNR) and an appealing structural similarity index (SSIM). Moreover, an extensive subjective-mean-opinion-score (SMOS) proves to be promising in the clinical application using HARN.

**INDEX TERMS** MR images, super-resolution, 0.35T MRI, 3T MRI, hybrid attention residual network, subjective-mean-opinion-score.

## I. INTRODUCTION

Nowadays, Magnetic Resonance Imaging (MRI) has been one of the most widely used medical imaging technologies because of its non-invasive examination of the human body's architecture and physiology. The most critical MRI characteristic is spatial resolution. Clinical researchers and hospitals generally prefer high-resolution (HR) images because they

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif .

present a clear anatomy structure and rich texture details, whereas low-resolution (LR) MR images always have fuzzy tissue boundaries with lower contrast. Generally, the image's signal-to-noise (SNR) is proportional to the magnetic field's strength [1], [2]. For instance, a 3T scanner provides 2.5 times higher SNR than a 0.35T scanner for the same spatial resolution. In other words, 3T MRI can provide images in higher resolution with similar SNR. In addition, 3T MRI is more sensitive to tissue changes and anatomical details, presenting more physiological features than 0.35T MRI. Specifically, 3T

MR images obtain more explicit tissue boundaries and higher tissue contrast than 0.35T, with Figure 1 depicting the 3T and 0.35T images of the same person's brain axial slice.

However, 3T MRI scanners are prohibitively costly, limiting their broad adoption. Many hospitals, particularly the township hospitals, still employ 0.35T scanners but wish to improve the images' quality for better disease diagnosis and image-guided intervention. Image post-processing method can be an alternative solution to reconstruct the 3T-like MR images from 0.35T MR images. Notably, though spatial resolution and SNR are not the only differences between images acquired from the low-field and high-field MRI systems, e.g., tissue contrast, using super-resolution (SR) related methods to increase the image's resolution without degrading the SNR is still a prominent approach to improve low-field image quality and make it comparable to the high-field ones.

SR techniques can reconstruct HR images from one or multi LR images without changing the MRI hardware system. The SR methods can be categorized based on the number of input LR images to single image super-resolution (SISR) [3] and multi-image super-resolution (MISR) [4]. Unlike MISR, SISR has a much higher efficiency [5] and lower graphics memory demands. Thus, we only focus on the SISR technique in this study.

Existing SR techniques in MRI can be classified as interpolation-based, reconstruction-based, and learning-based approaches. The interpolation functions are often considered the most straightforward and intuitive SR method [6], [7]. Whereas the interpolation-based methods are computationally simple, the processed images may be over smoothed and usually have visual artifacts such as ringing and fuzzy edges. The present reconstruction-based methods apply the degradation model by utilizing prior information with regularization methods. Bahrami *et al.* [8] used regression random forests and proposed a novel sparse representation method that predicted 7T-like images from 3T MR images. Although their method has high accuracy for brain MRIs, their study has high input requirements, and the sample size limits generalization.

The learning-based methods are the most widely used algorithms because they can generate novel details that do not appear in LR images. The SR methods based on convolution neural networks (CNN) have attracted broad interest, with Dong *et al.* [9] developing the first CNN-based SR method (a simple three-layer architecture called SRCNN) that performed well on super-resolving photographic images. Later they proposed a faster network (FSRCNN) with fewer parameters achieving better performance [10]. Subsequently, researchers focused on the architecture's feature extraction ability, with Kim *et al.* [11] proposing an intensive, very deep SR network (VDSR). To accelerate the VDSR's convergence, researchers put forward residual learning and gradient clipping. Lim *et al.* [12] developed an enhanced deep SR network (EDSR) by removing VDSR's unnecessary modules and expanding the model size. Although VDSR and its variants solve the gradient problems in deep networks

and achieve good performance, a deeper network is harder to train and preserve hierarchical information. To handle this problem, Tong *et al.* [13] leveraged dense skip connections and created a novel super-resolution dense network (SRDenseNet). For MR images, Zheng *et al.* [14] employed variants of dense blocks to enrich the features extracted from the MR slices. Moreover, Pham *et al.* [15] developed a three-dimensional (3D) version of SRCNN for brain MRI. Similarly, Wang *et al.* [16] proposed a 3D feature attention SR network (FASR), which utilized channel and sparse attention operations in parallel.

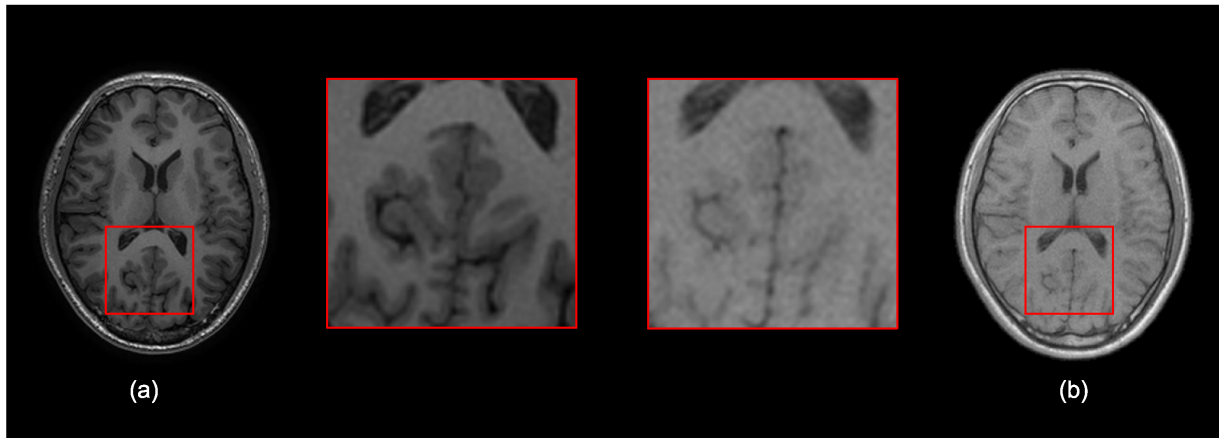
However, the disadvantage of the above-mentioned learning-based SR algorithms is that they assume the degradation from HR to LR is fixed and known. Thus, the LR images could be generated using bicubic or other average-type methods for the models to learn the mapping relationship from the fixed LR and HR images and estimate the weights. The weighted model is then exploited to create the desired HR image. Nevertheless, for a large distribution gap between the LR and HR images, the reconstruction performance of these methods may be unsatisfactory.

To adapt the degradation uncertainty, in this paper, we create a dataset by scanning 70 volunteers with both 0.35T and 3T machines (refer to Section II-A) and utilized Advanced Neuroimaging Tools (ANTs) [17] to pairwise register them. This work assumes that the high-frequency information obtained in high-field MR images can be directly predicted from the low-field MR images. Consequently, a low-field 0.35T image can be reconstructed to a 3T-like image by learning the mapping correction between 0.35T and 3T MR images.

In addition, the learning-based SR networks can extract rich frequency information in the channels and spatial regions. To extract abundant features from input LR images efficiently and motivated by recent advances [18], we utilize a dense attention block (DAB) comprising variants of parallel placed densely and hybrid attention blocks. The dense structure assists in the deeper network's gradients backpropagation, while the attention blocks fully utilize the channel and spatial information. Hence we propose a novel hybrid attention residual network, entitled HybridAttentionResNet (HARN), to generate 3T-like MR images by incorporating the mapping relationship of 0.35T and 3T MR images.

The major contributions of this work are:

- Scanning 70 volunteers using 0.35T and 3T machines to collect a particular dataset (Dataset I) for learning the real-world association between LR and HR imagery.
- Introducing a new feature extraction module, the dense attention block (DAB), based on dense connections with an attention mechanism that focuses more on the channel and spatial information.
- Proposing a deep residual neural network relying on DAB, named HARN, using actual LR and HR dataset pairs to estimate 3T-like MR images involving  $2 \times$  upscaling factors. Furthermore, we verify the main parameters through extensive ablation experiments.



**FIGURE 1.** Axial views of 3T MRI (a) and 0.35T MRI (b) of the same slice and zoomed regions. of (a) 3T MRI and (b) 0.35T MRI. 3T MRI has the higher anatomical quality and details compared to 0.35T.

- HARN can generalize high-quality 3T-like MR images with higher quantitative indicators in terms of PSNR/SSIM. Moreover, we test the proposed HARN via an open-source IXI dataset (Dataset II) to validate its robustness and accuracy.

This paper is presented as follows. In Section II, we introduce the data source and propose our 3T-like images reconstruction network. Designed ablation experiments and visual results are given in Sections III. Section IV and V provides the discussion and conclusion of the paper.

## II. MATERIALS AND METHODS

### A. DATA PREPARATION

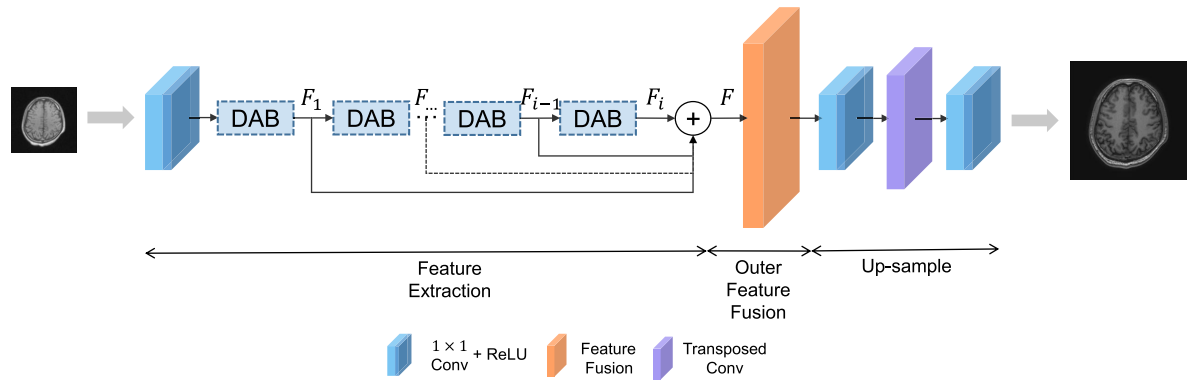
For this work, 70 participants were enlisted, equally divided by gender, and were scanned by 0.35T and 3T MRI scanning systems. Permission was obtained from the Institution Review Board and all subjects provided written informed consent before the scans. A total of 2100 axial slices/images were acquired from the 0.35T scanner (CLIMBER035 designed by Anhui Fuqing Medical Technology Co., Ltd.) with a 2D T1 scanning sequence SE weighted with parameters: TR = 400ms, TE = 16ms, FOV = 24cm × 24cm, and a matrix of 128 × 128. Regarding the 3T system (GE MEDICAL SYSTEM – DISCOVER MR750), 13160 images were acquired with sequence 3D T1-BRAVO adopting the following parameters: TR = 8.2ms, TE = 3.2ms, TI = 1.0ms, FOV = 24cm × 24cm, and a matrix of 256 × 256 × 188. The 3T scanning system was utilized after 0.35T, with the FOV of the 3T covering that of 0.35T scanning for better alignment. As note above, the scanning parameters of the two MRI systems are deviated, and thus the images of both systems were intrinsically different considering image resolution and contrast. Nevertheless, as this study aims to improve low-field images to be like high-field, image contrast differences were properly handled through our proposed network. Despite the contrast difference, for convenience, the 0.35T images are considered the LR dataset, and the 3T images the HR dataset.

If a patient moves between two subsequent scans, he causes image distortion due to the magnetic field inhomogeneity of each system, and thus we perfectly align the LR and HR dataset by applying a medical image analysis toolkit named Advanced Neuroimaging Tools (ANTs) [17]. The latter toolkit includes the software suite Analysis of Functional Neuro Images (AFNI) [19] to minimize the possible distributions between two sub-datasets in different resolutions. We aligned all 3T images on the 0.35T images, and the choice of the target/reference image was due to the fact that 3T images have higher resolution with smaller slice thickness (1mm < 5mm). After registration, the new HR slices were selected by re-slicing the aligned 3T volume, which was reconstructed from the aligned 3T images that corresponded to the same slice location of each slice in the 0.35T dataset. This ensured that the corresponding slices of both datasets depict the same axial physical slice of the brain and have the same anatomical structures. We marked this unique dataset as Dataset I. Moreover, to evaluate the robustness of the proposed method, we employed the IXI open-source dataset provided by BrainWeb.<sup>1</sup> Specifically, we chose 50 different T1 axial plane images from the 3T IXI dataset as an unseen test dataset and marked them as dataset II. To generate the input LR images, we blurred the original 3T images (HR) using a Gaussian kernel with  $\alpha = 4$  and then downsampled them by averaging every four voxels. In this way, the input LR images have half the resolution of the HR images.

### B. NETWORK STRUCTURE

This section introduces HARN, with its overview presented in Figure 2. The HARN network comprises feature extraction, outer feature fusion, and up-sampling modules. The critical phases of HARN are as follows: Initially, a shallow convolution layer with a ReLU function extracts the initial features from the input LR images. Then, the feature extraction mod-

<sup>1</sup><https://brainweb.bic.mni.mcgill.ca/brainweb/>



**FIGURE 2.** The overall architecture of our proposed HARN network. Before the feature fusion layer, a series of DAB outputs are merged using an element-wise summation.

ule recovers the important hierarchical features from the previously constructed feature maps. After that, we simplify the calculations utilizing the outer feature fusion layer (OFFL) scheme that decreases the merged feature maps to a specific size. Finally, the up-sampling module transfers the fused features into the desired 3T MR images.

### 1) DENSE ATTENTION BLOCKS

MR images are fundamentally resembling and redundant. To fully exploit the properties of MR images and capture the delicate local texture information on a small receptive field, we focus more on the feature capturing module and propose a novel architecture named dense attention block (DAB), which can be regarded as a delicate feature encoder. The DAB module is depicted in Figure 3, containing various parallel variant dense blocks (VDB), an inner feature fusion layer (IFFL), and a hybrid attention block (HAB).

### 2) VARIANT DENSE BLOCK

We employ convolution layers with variable kernel sizes to capture enhanced multiscale information combined in a dense structure [20] at the same level. As seen in Figure 3 (a), the blocks adopt two distinct kernel sizes and arrange them in various sequences. For instance, in the  $VDB_1$  the kernel sizes of the two convolution layers are  $1 \times 1$  and  $3 \times 3$ , respectively, arranged alternatively. We employ a small kernel size ( $1 \times 1$  and  $3 \times 3$ ), as such sizes require fewer parameters and use less RAM, speeding up processing.

Each VDB has four layers, each of which implements a composite operation function  $F_l$ , where  $l$  is the layer index. As a consequence, in the  $p^{th}$  path number of VDBs, the  $l^{th}$  layer receives all the previous layers' feature maps  $f_1^p, f_2^p, \dots, f_{(l-1)}^p$ , with the  $l^{th}$  layer's output being:

$$f_l^p = \sigma([f_1^p, f_2^p, \dots, f_{(l-1)}^p]) \quad (1)$$

where  $[f_1^p, f_2^p, \dots, f_{(l-1)}^p]$  represents the concatenated feature maps, and  $\sigma = \max(x, 0)$  refers to the ReLU activation function. Equation (1) indicates that a particular  $f_l^p$  depends on the kernel size of each layer and can extract feature maps

of various sizes. The VDB's hyperparameter growth rate (G) refers to the channels of each layer's output feature maps. Thus, the output channels are the input channels plus four times G, where four indicates the convolution layers numbers inside VDB.

### 3) INNER FEATURE FUSION

We utilize the inner feature fusion layer to concatenate the features and reduce their dimension, preventing excessive model parameter increase as the VDB's path number (P) increases. The output can be defined as:

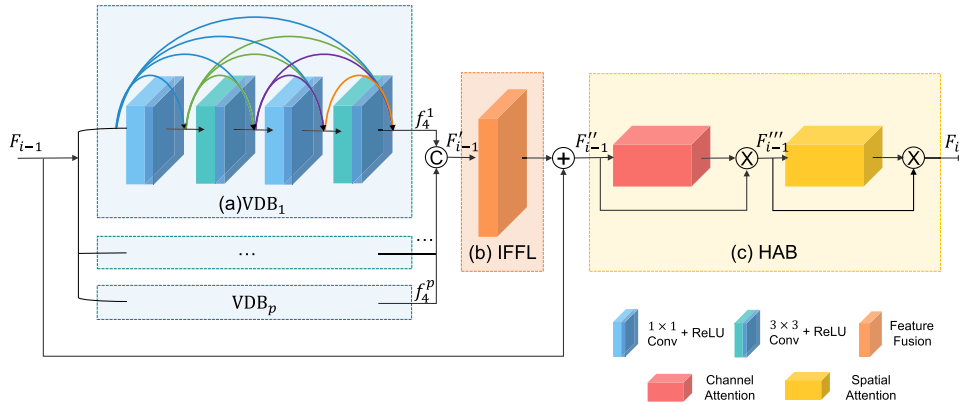
$$F''_{i-1} = \sigma(\text{conv}^{1 \times 1} F'_{i-1}) + F_{i-1} \quad (2)$$

where  $F'_{i-1}$  is  $[f_i^1, f_i^2, \dots, f_i^p]$  refers to the concatenation of p VDB outputs,  $F_{i-1}$  is the DAB input, and  $\text{conv}^{1 \times 1}$  indicates the convolution operation with a  $1 \times 1$  kernel size.

### 4) HYBRID ATTENTION BLOCK

After the IFFL, the scale of the  $F''_{i-1}$  features grows massively and includes much redundant information. Simultaneously, as demonstrated in [16], both channels and spatial areas restore the MRI features during the SR task. Based on these two considerations, we introduce an attention mechanism [18] to augment the network's representation capacity. As illustrated in Figure 3 (b), the Hybrid Attention Block (HAB) comprises two components: spatial attention (SA) and channel attention (CA). Due to the HAB's unique mechanics, the network can be more attentive to informative spatial regions and meaningful cross-channel information. Finally, the HAB's features are multiplied by the input feature maps for adaptive feature refinement.

According to Zeiler et al. [21], each channel in a feature map can act as a feature detector. Thus, CA extracts the global feature information and generates channel weights utilizing inter-channel interaction features. Therefore, we utilize a global max pooling and a global average pooling operation in parallel to capture the global spatial details, generating two different channel information descriptors  $\text{AvgPool}(F)$  and  $\text{MaxPool}(F)$ . The global average pooling function can



**FIGURE 3.** DAB's architecture (the VDB, IFFL, and HAB parts are depicted in blue, orange, and yellow, respectively).

be expressed as:

$$AvgPool(F_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H F_c(i, j) \quad (3)$$

where  $F_c(i, j)$  is the value associated with the position  $(i, j)$  in the  $C^{th}$  channel feature map.  $F_c \in 1 \times 1 \times c$  refers to the channel statistic generated by shrinking the input feature map  $F$  into spatial dimensions  $W \times H$ . Moreover, the max-pooling operation is determined as:

$$MaxPool(F_c) = \max_{i,j} F_c(i, j) \quad (4)$$

After the pooling procedure, we employ a multi-layer perceptron (MLP) to reduce the parameter overhead and set the reduction ratio to 0.5. Then, we use element-wise summing to merge the pooled feature vectors and apply them to a sigmoid gating mechanism. As illustrated in Figure 4 (a),  $F''_{i-1}$  indicates the feature maps of size  $W \times H \times C$ , with the channel attention output  $F_{CA}$  computed as:

$$F_{CA} = \sigma(MLP(MaxPool(F''_{i-1})) + MLP(AvgPool(F''_{i-1}))) \quad (5)$$

The MLP comprises two convolution layers with a ReLU activation function in between, aiming to reduce the network's parameters. Finally, the final output  $F'''_{i-1}$  is obtained by pulsing the spatial input  $F''_{i-1}$  with feature attention weights  $F_{CA}$ .

We supplement CA by utilizing the SA module. SA restores more position-specific information through generating spatial weights by exploiting the feature's inter-spatial relationship, enabling HARN to focus on critical but often neglected spatial areas. The entire procedure, presented in Figure 4 (b), is as follows. Initially, we apply a global average-pooling operation  $AvgPool(F)$  and a global max-pooling operation  $MaxPool(F)$  along the channel axis, which effectively emphasize information regions by reducing the channel's dimension [22]. Following the average pooling, the input feature map  $F'''_{i-1}$  can be regarded as an efficient

feature descriptor. The two feature maps are concatenated, and then are sent to a convolution layer to create a spatial attention map, encoding the regions' weights that are emphasized and suppressed. Mathematically, the complete process is as follows:

$$F_{SA} = \sigma(conv^{7 \times 7}[MaxPool(F'''_{i-1}); AvgPool(F'''_{i-1})]) \quad (6)$$

where  $conv^{7 \times 7}$  denotes the convolution operation with a  $7 \times 7$  kernel size, and  $[MaxPool(F'''_{i-1}); AvgPool(F'''_{i-1})]$  is the concatenation operation involving the pooling feature maps.

### 5) OUTER FEATURE FUSION

Section II.B.1 indicates that DAB can have various additional features assisting HR reconstruction. Indeed, we properly align DAB utilizing various parameter setups to exploit fully the hierarchical features it provides. However, the gradient vanishes as the network depth increases, and the loss becomes non-convergent. To solve this matter, we apply a fusion layer to merge all previous transformation feature maps:

$$F_{OFFL} = \sigma(conv^{1 \times 1}[F_1, F_2, \dots, F_{i-1}, F_i]) \quad (7)$$

where  $F_1, F_2, \dots, F_{i-1}, F_i$  represent outputs of different DAB,  $i$  denotes the series number of DAB, and  $F_{OFFL}$  is the output of the outer feature fusion layer, which is then fed to the next up-sampling stage.

### C. LOSS FUNCTION

Several SR techniques utilize a mean square error (MSE)-based loss function to reduce the difference between the input and the reconstructed images. Nonetheless, decreasing MSE typically reduces the reconstructed images' perceptual quality due to over-smoothing. To overcome this problem, we utilize a hybrid loss function comprising an image-domain MSE loss at the pixel level and a VGG loss at the perceptual level:

$$L_{MSE} = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W (I_{ij}^{SR} - I_{ij}^{HR})^2 \quad (8)$$

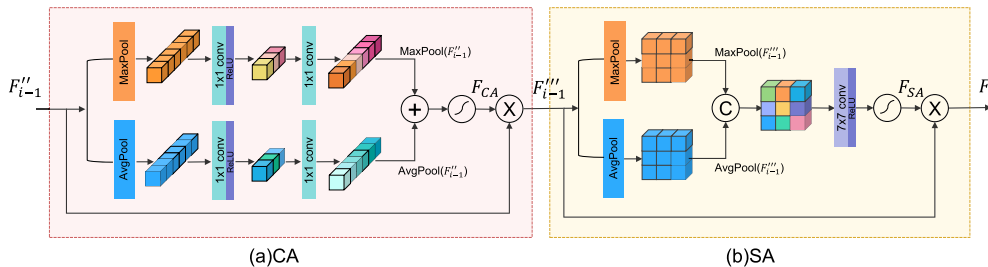


FIGURE 4. The HAB architecture (a) Channel attention, (b) Spatial attention.

where H and W define the image’s height and width, and  $I^{SR}$  and  $I^{HR}$  denote the 3T-like MRI generated by the model and the 3T images acquired from the 3T scanner. Inspired by the content loss [23], we import the VGG loss from the ReLU activation layers of the pre-trained 19-layered VGG network.

$$L_{VGG} = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W \left( S_{VGG} \left( I_{ij}^{SR} \right) - S_{VGG} \left( I_{ij}^{HR} \right) \right)^2 \quad (9)$$

Here,  $S_{VGG}$  indicates the supplied feature map from the VGG19 network. As a result, the total loss function is represented by:

$$L_{TOTAL} = L_{MSE} + \alpha L_{VGG} \quad (10)$$

where  $\alpha$  is a constant coefficient balancing the two losses, heuristically set to 1e-1.

#### D. EVALUATION METRICS

We evaluate the image quality utilizing objective and subjective metrics. Considering the objective metrics, we deploy the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [24]:

$$PSNR = 10 \lg \left( \frac{L^2}{MSE} \right) \quad (11)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12)$$

When an image is normalized using the linear normalization approach, for PSNR, we set  $L = 1$ . In Equation (12),  $C_1$  and  $C_2$  are the stability constants,  $\mu_x$  and  $\mu_y$  are the average values of x and y,  $\mu_x^2$  is the variance of x,  $\sigma_{xy}$  is the covariance of x and y, and  $\sigma_x^2$  is variance of x.

### III. EXPERIMENTS

#### A. IMPLEMENTATION DETAILS

This work regards the original 3T MR images, and as test images as ground truth (GT) HR images ( $256 \times 256$ ), the LR images ( $128 \times 128$ ) acquired from 0.35T scanners having half the GT image resolution. The main changes to the HARN parameters are described in Table 1. During the feature extraction stage, the channel numbers of the first convolution layer’s output features are set to 64. The output channels inside the VDB are determined by the growth rate (G) of the dense connections. Before feeding the features to HAB, we utilize IFFL to reduce the channel numbers from

$((64 + G \times 4) \times P)$  to 64, where P refers to VDB’s path numbers. Then the OFFL receives all DAB’s concatenated features and reduces the channel numbers from  $(64 \times I)$  to 64, where the variant I indicates the number of DABs. Finally, we restore the image size using the transposed convolution layers (convTrans) [25].

A cross-validation strategy divides our dataset into training, validation, and test set with a ratio of 7:2:1. Moreover, we employ the Adam optimizer [26] with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and a learning rate of 1e-4. The proposed model is trained using Pytorch 1.9.0 on an NVIDIA RTX 3090 GPU with 24 GB RAM.

#### B. SUBJECTIVE MEAN OPINION SCORE(SMOS) TESTING

We conducted a subjective mean opinion score (SMOS) test to quantify the reconstruction ability of various approaches. Specifically, we selected ten versions of each image from the test dataset: input LR image, Bicubic, SRCNN [9], FSRCNN [10], VDSR [11], EDSR [12], SRDenseNet [13], HybridNet [14], ours, and the ground truth HR image. Two radiologists with 5 and 7 years of experience, blinded to the acquisition details, were assigned to score for each image from 1 to 5(a higher score indicates better performance). Thus, each rater scored 1350 images (10 versions of 135 images) presented randomly.

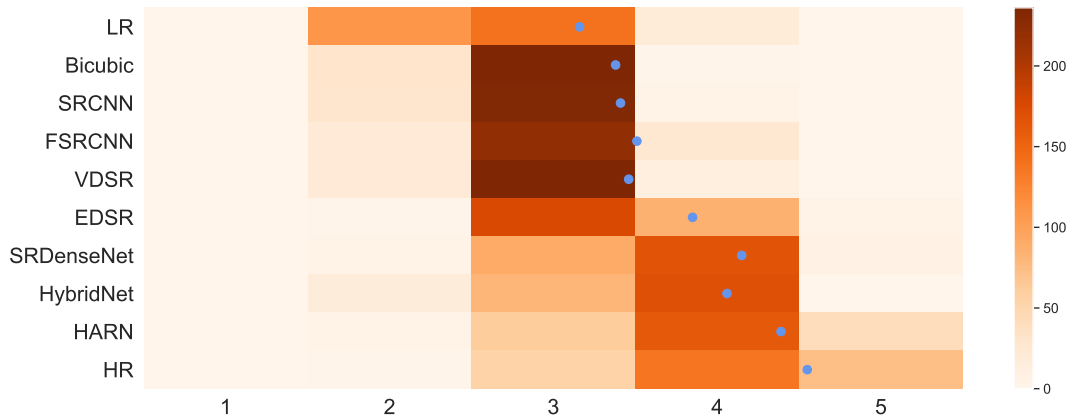
In this testing experiment, we discovered that SMOS has a high degree of dependability because there is no significant discrepancy between the ratings of the identical images. At the start of testing, we collected 20 pairs of different LR and HR images (score 5) for doctors to calibrate the rating criteria. We added the HR and LR images into the test set twice to confirm the raters’ reliability. Interestingly, the two doctors’ ratings for the same image category showed high similarity. Table 2 and Figure 5 describe the experimental results of the SMOS test.

#### C. ABLATION STUDY

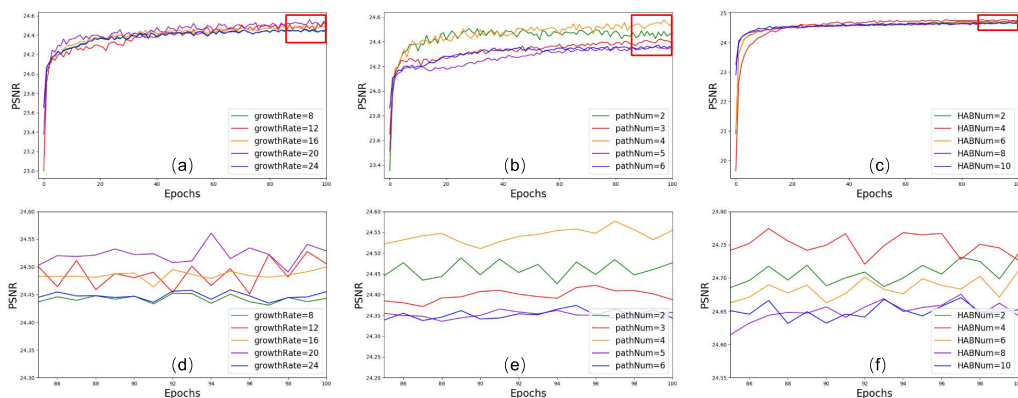
We conduct the following extensive ablation experiments of PSNR and SSIM to explore the best parameter values for HARN’s various components.

##### 1) STUDY OF G, P, AND I

Since growth rate (G) is a hyper-parameter of the dense connections, we performed several ablation studies to explore its influence on HARN’s performance. As visualized in



**FIGURE 5.** Heat map of the SMOS score distribution on Dataset I. For each method, 260 samples (135 images × 2 doctors) were evaluated, and the mean is shown as a blue marker (the bins around the mean value) [2 × upscaling].



**FIGURE 6.** Training convergence changes of PSNR on the growth rate (G), path number (P), and the number of DAB (I). The first row (a-c) shows the influence of G, P, and H in the whole converges of HARN, second row (d-f) zooms the selected rectangle of the first row.

Figure 6 (a) and (d), the PSNR increases first and decreases as quantity increases. Thus, we choose 16 as the final growth rate to balance the computation complexity and network performance.

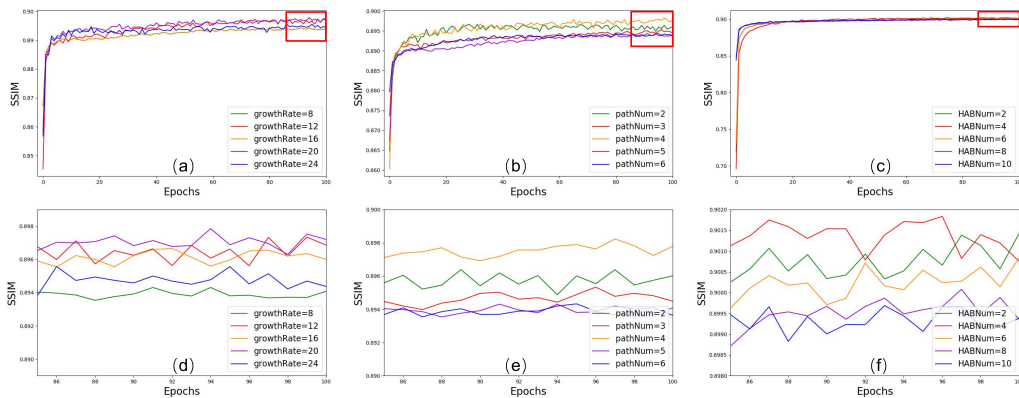
To demonstrate the multipath effect structure in VDB, we perform several contrast experiments, with Figure 6 (b) displaying the HARN’s various training convergences. Limited by the sample size, the PSNR reduces as the path numbers increase. Moreover, Figure 6 (e) shows the detailed convergence changing in the last 15 epochs, illustrating that the increasing path number may not increase PSNR. Finally, after balancing complexity and reconstruction capabilities, we set the path number of VDB to four.

The numbers of HAB affect the entire network depth and complexity. To investigate the effects of HAB’s number on the performance and computational cost, we study parameter I under different HAB numbers. Figure 6 (c) and (f) display the results of HARN’s five training convergences. As the HAB numbers increase, the faster HARN converges, but PSNR becomes lower. To preserve a better balance between computational efficiency and performance, we set the number of HAB to four.

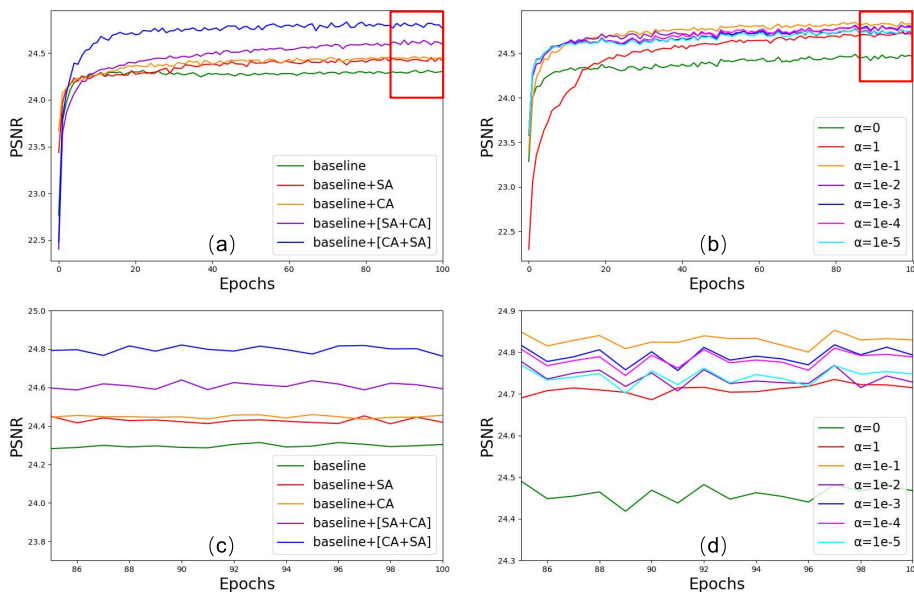
Figure 7 indicates the ablation experiments of G, P and I based on SSIM. Notably, the training convergences of SSIM and PSNR are highly similar. Thus, we set the same parameters as those analyzed based on PSNR.

## 2) STUDY OF ATTENTION MECHANISM AND LEARNING PARAMETERS $\alpha$

To further validate HAB’s effectiveness, we consider a network without HAB as the baseline and investigate the impact of SA and CA at a reduction ratio equal to two. Figure 8 (a) illustrates the convergence curves of several networks, but Figure 8 (c) reveals that the network with CA or SA presents an improved PSNR compared to the baseline. Notably, the cascaded CA and SA network outperform the network solely using CA or SA. Given that CA and SA can generate the weight of each feature map in channel and space, cascading the CA and SA mechanisms combines the channel and spatial information to enhance further the high-frequency features. Furthermore, in this trial, we also verify the effect of the order of CA and SA in the HAB. Figure 9 (a) and (c) show the training convergence changes of SSIM, and the convergences of PSNR and SSIM are almost identical.



**FIGURE 7.** Training convergence changes of SSIM on the growth rate (G), path number (P), and the number of DAB (H). The first row (a-c) shows the influence of G, P, and H in the whole converges of HARN, second row (d-f) zooms the selected rectangle of the first row.



**FIGURE 8.** Convergence analysis on the attention mechanism and learning parameters  $\alpha$  based on PSNR. As (a) indicates, the baseline with attention mechanism can improve the model's reconstruction ability. (b) shows that the  $L_{VGG}$  plays an important role in reconstruction. Second row (c,d) zooms the selected rectangle of the first row.

The model with  $L_{MSE}$  focuses on the loss of each pixel, potentially over-smoothing the image, whereas the model with  $L_{VGG}$  produces distorted details. To balance the hybrid loss, we test several values for the balancing factor  $\alpha$ , with the corresponding results illustrated in Figure 8 (b) and Figure 9 (b), which are the different ablation experiments based on PSNR and SSIM, respectively. From the two figures, the gap between two losses becomes wider when  $\alpha$  decreases. Therefore, the reconstruction performance degrades. According to the results depicted in Figure 8 (d) and Figure 9 (d), we set  $\alpha = 1e-1$  finally.

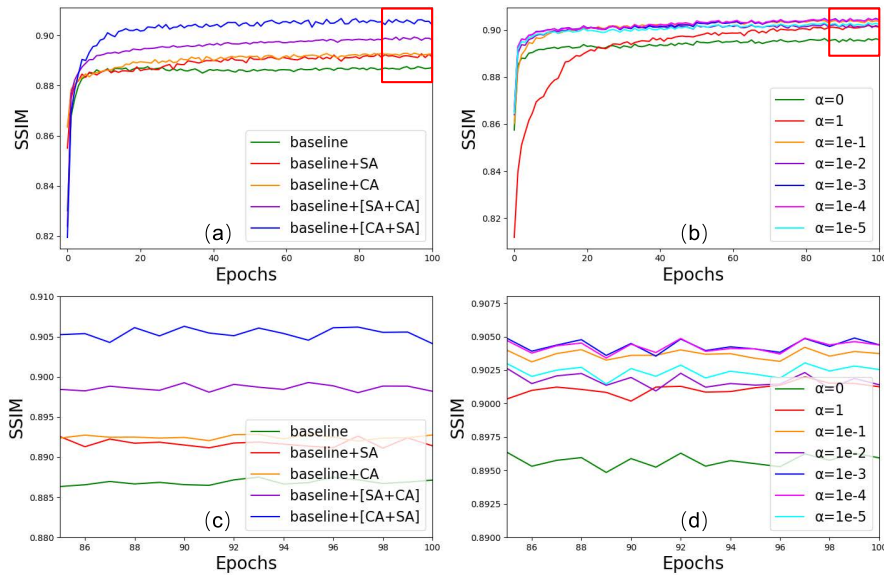
**D. COMPARISONS AGAINST STATE-OF-THE-ART METHODS**

To further evaluate the proposed network's performance, we challenge HARN against bicubic interpolation and six learning-based methods [9]–[14]. Moreover, to analyze the

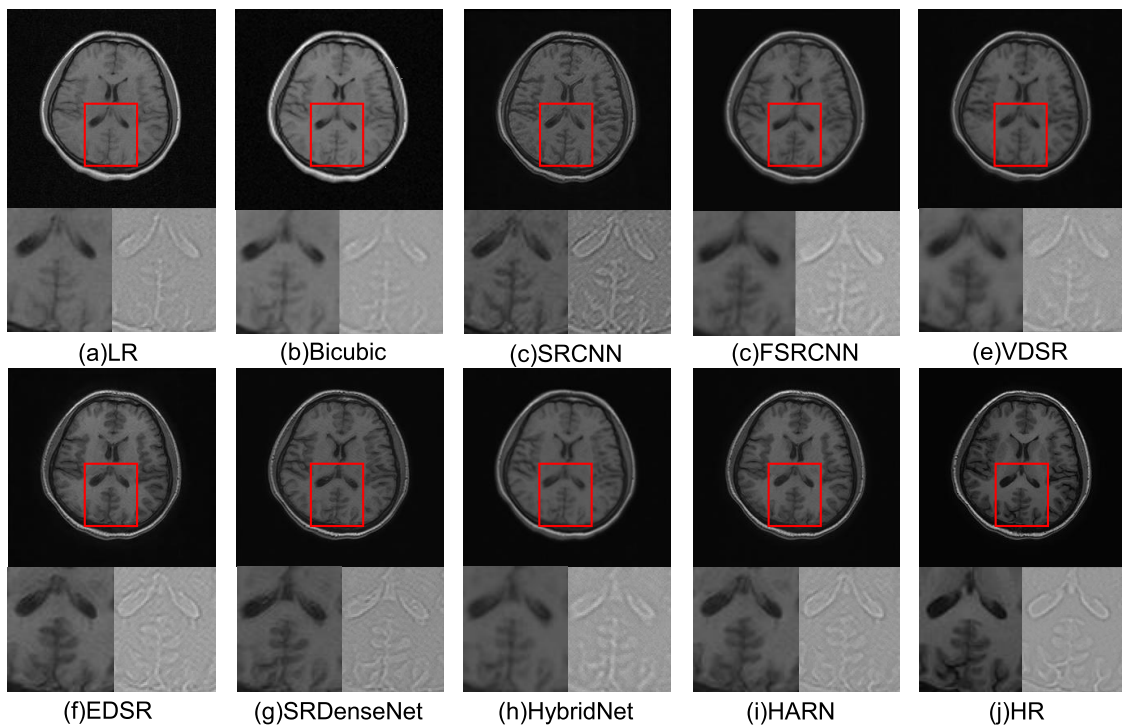
results more precisely, we calculate the mean and variance of PSNR and SSIM. Table 2 lists the corresponding results of the quantitative indicators. From the results on the left side of the table, we conclude that HARN achieves the best performance with  $PSNR = 25.4623 \pm 9.4367$  and  $SSIM = 0.9080 \pm 0.0217$  on Dataset I, significantly outperforming the competitor methods.

Figure 10 depicts a qualitative comparison of the evaluated methods, including two close-up views of selected regions below every reconstruction image: the left image shows the zoomed image of the chosen gyrus region, and the right, the edge information of the left gyrus region. Figure 10 reveals that the competitor algorithms tend to reconstruct fuzzy and over-smoothed details, affecting identifying the depicted details. By comparison, the proposed HARN effectively recovers more contours and minor textures. The zoomed grayscale images show that our algorithm has lower





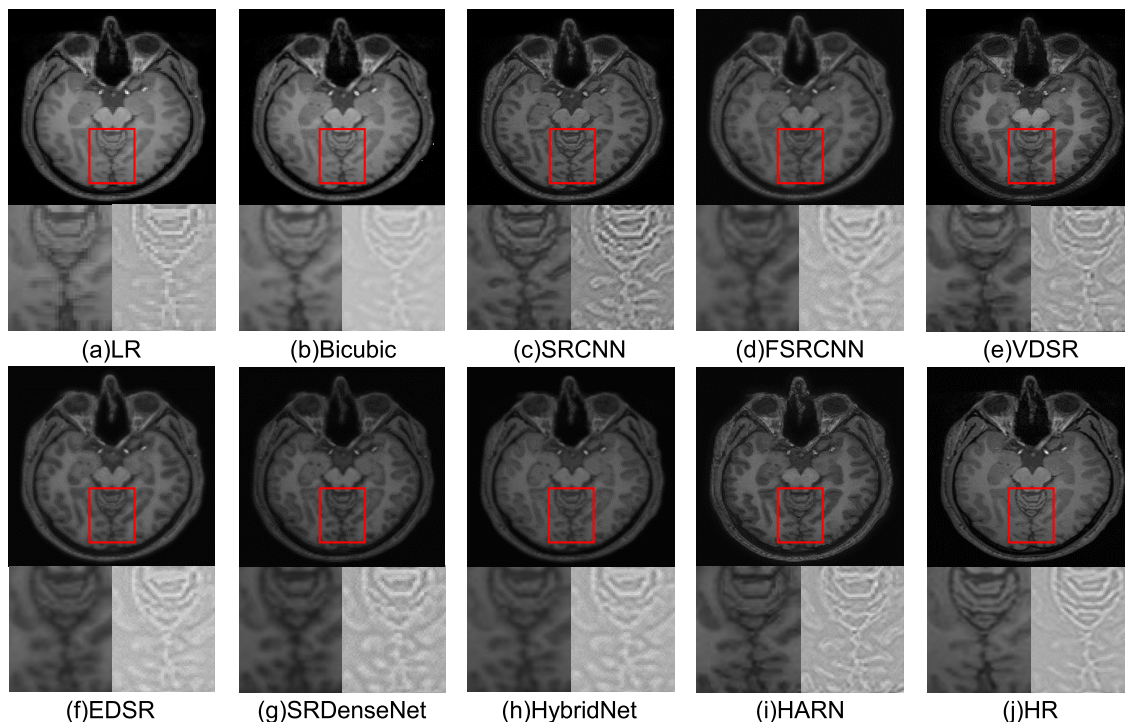
**FIGURE 9.** Convergence analysis on the attention mechanism and learning parameters  $\alpha$  based on SSIM. (a) and (b) show the training convergences of different attention mechanism and learning rates. (c) and (d) are the zoomed images of selected rectangle in (a) and (b), respectively.



**FIGURE 10.** Results of the reconstructed images with various approaches on Datasets I. At the bottom left of each image is the zoomed image highlighted in the red rectangle region, and the right is the Laplace operator's result on the left zoomed image. [2 × upscaling].

noise and more precise edge information. The HARN's ratings are presented in Table 2, highlighting that the SMOS ratings are closer to the original scores than the competitor methods. Figure 5 shows the distribution of all SMOS ratings.

Furthermore, we employ additional open-source datasets (IXI dataset) to incorporate our experiments. The aim is to verify whether the algorithm can produce more realistic images with good generalization ability on other datasets.



**FIGURE 11.** Results of the reconstructed images with various approaches on Datasets II. At the bottom left of each image is the zoomed image highlighted in the red rectangle region, and the right is the Laplace operator’s result on the left zoomed image. [2 × upscaling].

**TABLE 1.** The main changes to the HARN parameters. The variables **G**, **P**, and **I** represent the growth rate of the dense blocks in VDB, VDB path number, and the total number of DABs, respectively.

Network Stage	Layer	Input Channel	Output Channel	Kernel Size
Feature Extraction	conv1	1	64	3±3
	VDB	64	(64+G×4)	-
	IFFL	(64+G×4)×P	64	1×1
	CA	64	64	1×1
Outer Feature Fusion	SA	64	64	7×7
	OFFL	64	64	1×1
Up-sample	conv2	64×1	64	3×3
	convTrans	64	64	2×2
	convTrans	64	1	3×3

As mentioned above, we selected 50 axial images as a new test dataset and marked them as Dataset II. During testing, we exploit the model trained on Dataset I. The right side of Table 2 shows that the HARN does not achieve the best PSNR/SSIM caused by the loss function difference. However, our contrast images are more photo-realistic than the competitor ones. The actual comparison is performed on the chosen 3T axial plane (Figure 11), highlighting that HARN’s reconstructed image has more precise details than input LR images and is more comparable to HR images than the competitor algorithms’ reconstruction outputs. Consequently, the proposed HARN network achieves a good generalization ability and can be applied to other datasets.

#### IV. DISCUSSION

This work demonstrates through SMOS testing that learning-based methods achieve superior clinical performance in generating 3T-like MR images from low-field 0.35T

images. Furthermore, we demonstrate that high-frequency information can be predicted from LR images. Thus, we generate reliable SR images by proposing a CNN-based algorithm named HybridAttentionResNet (HARN), which incorporates dense blocks and attention mechanisms for better feature extraction. We collected a unique dataset by scanning 70 subjects from both 0.35T and 3T MRI systems and aligning the paired images before training to explore the mapping correlation between the LR and HR images. Additionally, we conduct several ablation experiments to determine the best parameters of HARN and employ two datasets for evaluation. The experimental results suggest that HARN performs better than current state-of-the-art SR algorithms and has an appealing generalization ability and accuracy.

In contrast to SRDenseNet [13], the dense blocks exhibit sufficient sensitivity for SR tasks. We speculate that our dense attention block combines the multipath structure of the convolution layers to extract more diverse features for reconstruction. In contrast to Zheng *et al.* [14], our model is optimized for attention mechanisms and content loss, with the proposed attention mechanism having a substantial impact on the network’s performance. Specifically, the CA module generates global features, but the SA module assists the network in focusing more on the local regions.

We only scanned the axial brain slices of 70 healthy volunteers in this work. However, the learning-based methods in SR usually require massive and diverse data for training to afford enhanced robustness. However, the relatively small-sized datasets employed in this work are speculated to

**TABLE 2. Qualitative comparison of contrastive SR algorithms on the unique dataset we scanned (Dataset I) and an open-source IXI dataset (Dataset II). [2 × upscaling].**

	Dataset I			Dataset II	
	PSNR(dB)	SSIM	SMOS	PSNR(dB)	SSIM
Bicubic	28.12±0.858	0.7941±0.012	2.88	28.12±0.858	0.9380±0.00025
SRCNN[9]	23.81±2.100	0.8851±0.0031	2.91	30.33±0.6456	0.9708±0.00022
FSRCNN[10]	24.29±2.516	0.8964±0.0035	3.01	29.77±1.028	0.9664±0.00036
VDSR[11]	24.07±4.069	0.8919±0.0038	2.96	28.33±1.073	0.9504±0.00028
EDSR[12]	24.53±3.004	0.9011±0.0035	3.35	<b>30.66±0.4428</b>	<b>0.9731±0.00064</b>
SRDenseNet[13]	25.12±3.144	0.9010±0.0031	3.65	29.77±2.839	0.9658±0.00015
HybridNet[14]	24.76±2.990	0.9016±0.0030	3.56	29.08±3.72	0.9531±0.00037
HARN(Ours)	<b>25.46±3.341</b>	<b>0.9080±0.0029</b>	<b>3.89</b>	29.91±4.471	0.9622±0.00027
HR	∞	1	4.05	∞	1

be responsible for the low PSNR and SSIM values. Future works could involve a GAN- [27] or Transformer-based [28] method, or a more extensive database, which will be used to solve this problem further. Limited by hardware, the input LR images have some noise and artifacts that are difficult to eradicate. The content loss function is an effective way to characterize spatial contents. Maybe emphasizing the content loss on minimizing rice noise could further enhance the clinical SR findings. Reconstruction with less noise is challenging and is part of future work. Finally, although we evaluated HARN on two brain datasets, applying the same method to other organs is still an open question that will be examined in future works.

## V. CONCLUSION

In this study, we collected a unique dataset by scanning 70 subjects with both 0.35T and 3T MR systems to produce LR and HR images. Instead of utilizing the predetermined known degradations, we use real paired training data to learn the mapping relationship between high field and low field images. Moreover, we proposed a residual network (HARN) with a hybrid attention mechanism based on the convolution neural network. After extensive ablation experiments, we set the best parameters for HARN. The experimental results demonstrate that HARN achieves good performance on the PSNR and SSIM metrics with more photo-realistic results. And via the extensive SMOS testing, HARN is proven to be more reliable in reconstructing HR images over scale ×2 than current state-of-the-art reconstructions methods. We also evaluate HARN on an open-source dataset (IXI dataset), with the experimental results revealing that our network achieves superior performance in robustness and accuracy. Overall, HARN is proved to be an effective approach to improve the image quality of 0.35T MR images. In the future, HARN could be used to apply in clinical applications and other image processing tasks, such as image-guided experiments and lesion segmentation, as it can reconstruct high-resolution images with decent quality and accuracy.

## REFERENCES

- [1] A. G. Van der Kolk, J. Hendrikse, J. J. Zwanenburg, F. Visser, and P. R. Luijten, "Clinical applications of 7 T MRI in the brain," *Eur. J. Radiol.*, vol. 82, no. 5, pp. 708–718, 2013.
- [2] B. L. Schmitz, A. J. Aschoff, M. H. K. Hoffmann, and G. Gron, "Advantages and pitfalls in 3t mr brain imaging: A pictorial review," *Amer. J. Neuroradiol.*, vol. 26, no. 9, pp. 2229–2237, 2005. [Online]. Available: <http://www.ajnr.org/content/ajnr/26/9/2229.full.pdf>
- [3] L. Pan, W. Yan, and H. Zheng, "Super-resolution from a single image based on local self-similarity," *Multimedia Tools Appl.*, vol. 75, no. 18, pp. 11037–11057, Sep. 2016. [Online]. Available: <https://WOS:000382679900011>
- [4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 44–1327, Sep. 2004. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/15462143>
- [5] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8723565/>
- [6] T. M. Lehmann, C. Gönnér, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999. [Online]. Available: <https://WOS:000085030400001>
- [7] D. Su and P. Willis, "Image interpolation by pixel-level data-dependent triangulation," *Comput. Graph. Forum*, vol. 23, no. 2, pp. 189–201, Jul. 2004. [Online]. Available: <https://WOS:000222579000006>
- [8] K. Bahrami, F. Shi, I. Rekik, Y. Gao, and D. Shen, "7T-guided super-resolution of 3T MRI," *Med. Phys.*, vol. 44, no. 5, pp. 1661–1677, May 2017.
- [9] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2014, pp. 184–199.
- [10] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. ECCV*, vol. 9906, 2016, pp. 391–407. [Online]. Available: <https://WOS:000389383900025>
- [11] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654. [Online]. Available: <https://WOS:000400012301074>
- [12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
- [13] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.
- [14] Y. Zheng, B. Zhen, A. Chen, F. Qi, X. Hao, and B. Qiu, "A hybrid convolutional neural network for super-resolution reconstruction of MR images," *Med. Phys.*, vol. 47, no. 7, pp. 3013–3022, Jul. 2020. [Online]. Available: <https://WOS:000528737900001>
- [15] 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 Med. Imag. Graph.*, vol. 77, Oct. 2019, Art. no. 101647. [Online]. Available: <https://WOS:000493216400002>
- [16] L. Wang, J. Du, H. Zhu, Z. He, and Y. Jia, "Brain MR image super-resolution using 3D feature attention network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 1151–1155.
- [17] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *NeuroImage*, vol. 54, no. 3, pp. 2033–2044, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065962/pdf/nihms238501.pdf>
- [18] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[19] R. W. Cox, "AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages," *Comput. Biomed. Res.*, vol. 29, no. 3, pp. 162–173, 1996. [Online]. Available: <https://WOS:A1996UV56700002>

[20] T. Tong, G. Li, X. J. Liu, and Q. Q. Gao, "Image super-resolution using dense skip connections," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817. [Online]. Available: <https://WOS:000425498404093>

[21] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 8689, 2014, pp. 818–833. [Online]. Available: <https://WOS:000345524200047>

[22] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," 2016, *arXiv:1612.03928*.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. [Online]. Available: <https://WOS:000220784600014>

[25] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*.

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

[27] W. Ueki, T. Nishii, K. Umehara, J. Ota, S. Higuchi, Y. Ohta, Y. Nagai, K. Murakawa, T. Ishida, and T. Fukuda, "Generative adversarial network-based post-processed image super-resolution technology for accelerating brain MRI: Comparison with compressed sensing," *Acta Radiologica*, vol. 2022, Feb. 2022, Art. no. 02841851221076330.

[28] C.-M. Feng, Y. Yan, H. Fu, L. Chen, and Y. Xu, "Task transformer network for joint MRI reconstruction and super-resolution," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*. Cham, Switzerland: Springer, 2021, pp. 307–317.



**JIANAN XU** received the B.E. degree in medical information engineering from the Hefei University of Technology, Hefei, China, in 2019. He is currently pursuing the M.E. degree in biomedical engineering with the University of Science and Technology of China, Hefei. His current research interest includes deep learning fast MR reconstruction.



**YUFU ZHOU** received the B.E. degree in electronic science and technology from Anhui University, Hefei, China, in 2012. He is currently pursuing the Ph.D. degree in biomedical engineering with the University of Science and Technology of China. His current research interest includes design of MR gradient coil based on stream function.



**DAYONG GAO** received the bachelor's degree in mechanical engineering from the University of Science and Technology of China, Hefei, China, in 1983, and the Ph.D. degree in mechanical engineering and biomedical engineering from Concordia University, Montreal, QC, Canada, in 1991.

He is currently a Full Professor of mechanical engineering and bioengineering and the Director of the Center for Cryo-Biomedical Engineering and Artificial Organs, University of Washington, Seattle, WA, USA. He is also the Distinguished Chair Professor (Adjunct Position) at the University of Science and Technology of China. Prior to joining as a Faculty Member at the University of Washington, he was a Baxter Healthcare Chair of engineering and an Alumni Professor of the University of Kentucky, Lexington, KY, USA. He has published more than 200 journal articles and more than 300 chapters/manuscripts in scientific books (17 books) or conference proceedings, such as cryo-biomedical engineering, fundamental cryobiology, biopreservation, artificial organs (artificial kidney and liver), and bioinstruments (biosensors and bioMEMS).

Dr. Gao received numerous awards/grants, as well as have been serving international scientific societies/associations and organizing/chairing scientific conferences worldwide.



**BENSHENG QIU** (Member, IEEE) received the bachelor's degree in electrical engineering from the College of Electronic Engineering, in 1987, the master's degree in acoustic engineering from Northwestern Polytechnic University, Fremont, CA, USA, in 1990, and the Ph.D. degree in computer science from the Hefei University of Technology, Hefei, China, in 1995.

From 1997 to 2001, he worked as an Associate Professor of radiology at the PLA General Hospital. From 2001 to 2005, he worked at the Department of Radiology, The Johns Hopkins University School of Medicine. From 2006 to 2012, he was an Assistant Professor at the University of Washington. He is currently a Professor and the Vice Chairperson of the Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, and the Director of the Medical Imaging Center, University of Science and Technology of China. He has conducted many projects on MRI-guided gene/stem cell therapy and interventions supported by NIH, RSNA, NSF, and the National Basic Research Program of China.

Dr. Qiu received many awards from the Radiology Society of North America, the American Heart Association, and the International Society for Magnetic Resonance in Medicine.

...



**JIALIANG JIANG** received the B.E. degree in medical information engineering from the Hefei University of Technology, Hefei, China, in 2019. He is currently pursuing the M.E. degree in biomedical engineering with the University of Science and Technology of China, Hefei. His current research interest includes MR images processing.



**FULANG QI** received the B.E. degree in electronic information of science and technology from the University of Science and Technology of China, Hefei, China, in 2014. He is currently pursuing the Ph.D. degree in biomedical engineering with the University of Science and Technology of China. His current research interest includes fast MR imaging.



**HUIYU DU** received the B.E. degree in mechanical engineering from the China University of Mining and Technology, Xuzhou, China, in 2018. He is currently pursuing the Ph.D. degree in biomedical engineering with the University of Science and Technology of China. His current research interest includes MRI hardware.