



A Segmentation-Aware Deep Fusion Network for Compressed Sensing MRI

Zhiwen Fan¹, Liyan Sun¹, Xinghao Ding¹(✉), Yue Huang¹, Congbo Cai¹,
and John Paisley²

¹ Fujian Key Laboratory of Sensing and Computing for Smart City,
Xiamen University, Xiamen, Fujian, China
dxh@xmu.edu.cn

² Department of Electrical Engineering, Columbia University, New York, NY, USA

Abstract. Compressed sensing MRI is a classic inverse problem in the field of computational imaging, accelerating the MR imaging by measuring less k-space data. The deep neural network models provide the stronger representation ability and faster reconstruction compared with “shallow” optimization-based methods. However, in the existing deep-based CS-MRI models, the high-level semantic supervision information from massive segmentation-labels in MRI dataset is overlooked. In this paper, we proposed a segmentation-aware deep fusion network called SADFN for compressed sensing MRI. The multilayer feature aggregation (MLFA) method is introduced here to fuse all the features from different layers in the segmentation network. Then, the aggregated feature maps containing semantic information are provided to each layer in the reconstruction network with a feature fusion strategy. This guarantees the reconstruction network is aware of the different regions in the image it reconstructs, simplifying the function mapping. We prove the utility of the cross-layer and cross-task information fusion strategy by comparative study. Extensive experiments on brain segmentation benchmark MRBrainS and BratS15 validated that the proposed SADFN model achieves state-of-the-art accuracy in compressed sensing MRI. This paper provides a novel approach to guide the low-level visual task using the information from mid- or high-level task.

Keywords: Compressed sensing · Magnetic resonance imaging
Medical image segmentation · Deep neural network

1 Introduction

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to produce the anatomical images in human body with the advantages of low

Z. Fan and L. Sun—The co-first authors contributed equally.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-01231-1_4) contains supplementary material, which is available to authorized users.

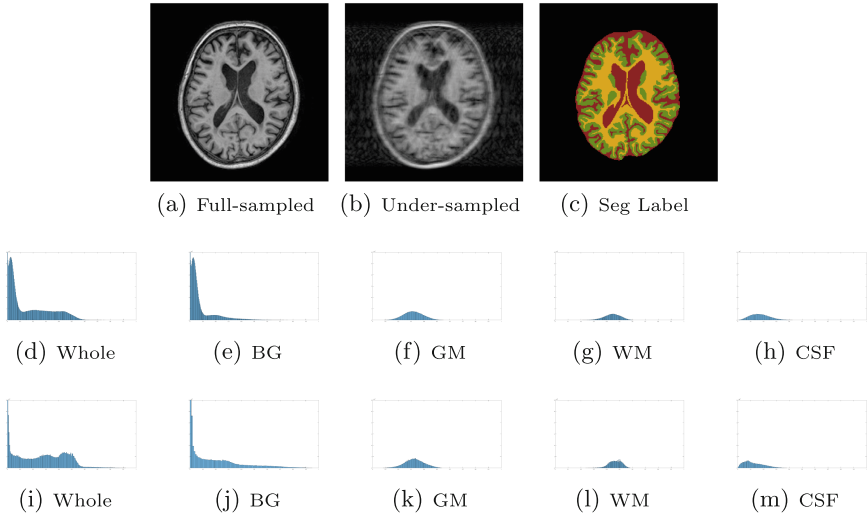


Fig. 1. A full-sampled MR image in Fig. (a), its under-sampled counterpart in Fig. (b) and segmentation labels in Fig. (c). We plot the histograms of under-sampled MRI (second row) and full-sampled MRI (third row) on training MRI datasets. (Color figure online)

radiation, high resolution in soft tissues and multiple imaging modalities. However, the major limitation in MRI is the slow imaging speed which causes motion artifacts [1] when the imaging subject moves consciously or unconsciously. The high resolution in k-t space is also difficult to be achieved in dynamic MRI because of long imaging period [2]. Thus compressed sensing technique is introduced to accelerate the MRI by measuring less k-space samples called compressed sensing MRI (CS-MRI) [3]. The CS-MRI is a classic inverse problem in computation imaging requiring proper regularization for accurate reconstruction.

The standard CS-MRI can be formulated as

$$\hat{x} = \arg \min_x \|F_u x - y\|_2^2 + \sum_i \alpha_i \Psi_i(x), \quad (1)$$

where $x \in C^{P \times 1}$ is the complex-valued MR image to be reconstructed, $F_u \in C^{M \times P}$ is the under-sampled Fourier operator and $y \in C^{M \times 1}$ ($M \ll P$) are the k-space measurements by the MRI machine, Ψ_i represents a certain prior transform, α_i is the parameter balancing the data fidelity term and the prior term. The first data fidelity term ensures consistency between the Fourier coefficients of the reconstructed image and the measured k-space data, while the second prior term regularizes the reconstruction to encourage certain image properties such as sparsity in a transform domain.

In conventional CS-MRI methods, the sparse and nonlocal are common priors for the inverse recovery in situ, which brings three limitations: (1) The common complex patterns hiding massive MRI datasets are overlooked in the

capacity-limited “shallow” prior [4]. (2) The sparse or nonlocal regularization lacks semantic representation ability, which is difficult to distinguish between the image structure details and structural artifacts brought by under-sampling. (3) The optimization for conventional priors requires long time to iterate to reach convergence, which brings long reconstruction time consumption [5].

Recently, the deep neural network models are introduced in the field of CS-MRI to overcome the limitations of conventional CS-MRI methods. Where the information from massive training MRI datasets can be encoded in the network architecture in training phase with large model capacity. Once the network is well-trained, the forward reconstruction for test MRI data is much faster compared with methods based on conventional sparse priors because no iteration is required. More importantly, the deep neural network models enjoy the benefit of modeling the semantic information in the image, providing an appropriate approach to integrate information for different visual tasks, however, which is rarely considered in the existing models for inverse problem, leaving high-level supervision information poorly utilized, causing negative effect on the later automatic analysis phase.

We take segmentation information for example to prove the benefits of introducing high-level supervision information into reconstruction. Usually different tissues in the MR image not only have different diagnostic information, but also show different statistical properties. In Fig. 1(a) and (b), we show a full-sampled and corresponding under-sampled T1-weighted brain MR image which contains three different labeled tissues: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The corresponding GM, WM and CSF labels are shown in green, yellow and red in the segmentation label map in Fig. 1(c). Clearly, different regions show different intensity scales. To further quantify this phenomenon, we give the statistical histograms of the three tissues, back ground (BG) and the whole images of the under-sampled/full-sampled MRI data in the second/third row in Fig. 1 on all the training MRI data. We observe each of the GM, WM and CSF tissues has simple single-mode distribution on the full-sampled and under-sampled MRI data. Since the deep neural network usually learns the function mapping from the under-sampled MR images to their full-sampled counterparts. The function mapping can be significantly simplified by learning the corresponding relations between the single-mode distributions. However, the distributions of the whole under-sampled and full-sampled MRI in Fig. 1(d) and (i) are much more complicated, making the learning of function mapping more difficult.

In this paper, we propose a segmentation-aware deep fusion network (SADFN) architecture for compressed sensing MRI to fuse the semantic supervision information in the different depth from the segmentation label and propagate the semantic features to each layer in the reconstruction network. The main contribution can be summarized as follows:

- The proposed SADFN model can effectively fuse the information from tasks and depths in different levels. Both the MRI reconstruction and segmentation accuracies are significantly improved under the proposed framework.
- The semantic information from the segmentation network is provided to reconstruction network using a feature fusion strategy, helping the reconstruction network be aware of the content it reconstructs and simplifying the function mapping.

- We adopt the multilayer feature aggregation to effectively collect and extract the information from different depth in the segmentation network.

2 Related Work

2.1 Compressed Sensing MRI

In the study of CS-MRI, the researches focus on proposing appropriate regularization. In the pioneer work SparseMRI [3], the fixed transform operator wavelets and total variation is adopted for regularization in Eq. 1. More methods [6–8] are proposed to address the same objective function efficiently. The variants of wavelet are proposed to exploit the geometric information in MR images adaptively in [9–11]. Dictionary learning techniques are also utilized in situ to model the MR images adaptively [5, 12, 13]. Nonlocal prior also can be introduced as regularization [14] or combined with sparse prior in [10].

Recently, the deep neural network models are introduced in CS-MRI. A vanilla deep convolutional neural network (CNN) is used to learn the function mapping from the zero-filled MR images to the full-sampled MR images [15]. Furthermore, a modified U-Net architecture is utilized to learn the residual mapping in [17]. The above deep-based CS-MRI models overlooks the accurate information on the sampled positions in the compressive measurements. In [4], a deep cascaded CNN (DC-CNN) is proposed to cascade several basic blocks to learn the mapping with each block containing the nonlinear convolution layers and a nonadjustable data fidelity layer. In data fidelity layers, the reconstructed MR images are corrected by the accurate k-space samples. Despite the state-of-the-art reconstruction quality has been achieved using the DC-CNN model, the high-level supervision information from the manual labels in MRI datasets hasn't been taken into consideration, still leaving room for further improvement on model performance.

2.2 MR Image Segmentation

With the segmentation labels in MRI datasets, different models are proposed to learn to automatically segment the MR images into different tissues from the test set. Compared with conventional segmentation methods based on manually designed features, the deep neural network models can extract image features automatically, leading to better segmentation performance. Recently, the U-shaped network called U-Net trained in end-to-end and pixel-to-pixel manner is proposed in [18], which can take the input of arbitrary size and produce the output of the same size, achieving the state-of-the-art medical image segmentation accuracy and computational efficiency. Its variant where the 2D operations are replaced with 3D ones is proposed in [19] called 3D U-Net. The residual learning is also utilized in the segmentation model in [20]. The recurrent neural network can efficiently model the relation among different frames in the volumetric MR data can introduced in the medical image segmentation [21, 22].

Throughout the paper, we use the classic 2D U-Net for single-frame MRI segmentation for the single-frame MRI reconstruction, and the proposed model can be easily extended to volumetric MRI data.

2.3 Multilayer Feature Aggregation

The works [23] on visualization of deep CNN has revealed the feature maps at different layers describe the image in different scales and views. In the conventional deep neural network models, the output is produced based on the deep layers or even the last layer of the model, leaving the features in lower layers containing information from different scales underemphasized. In the field of salient object detection, the multilayer feature aggregation is a popular approach to integrate information from different layers in the network [24–26].

2.4 High-Level Information Guidance for Low-Level Tasks

In [16], the MRI reconstruction and segmentation are integrated into a single objective function, resulting in both improvements on reconstruction and segmentation. However, the sparse-based method is limited by the model capacity and lack of semantic representation. Recently, some works are devoted to combining the low-level task with tasks in higher levels. In the work of [27], a well pre-trained segmentation network is cascaded behind a denoising network, then the loss functions for both segmentation and denoising are optimized to train the denoising network without adjusting the parameters in segmentation network. With this model, the denoising network produces the denoised images with higher segmentation accuracy using automatic segmentation network at the expense of limited improvement in restoration accuracy or even degradation. In the AOD-Net [28], the well-trained dehaze model is jointly optimized with a faster R-CNN, resulting better detection and recognition results.

3 The Proposed Architecture

To incorporate the information from segmentation label into the MRI reconstruction, we proposed the segmentation-aware deep fusion network (SADFN). The network architecture is shown in Fig. 2. The reconstruction network and segmentation network are first pre-trained. Then a segmentation-aware feature extraction module is designed to provide features with rich segmentation information to reconstruction network using a feature fusion strategy.

3.1 The Pre-trained MRI Reconstruction Network

As we introduced above, the DC-CNN architecture achieves the state-of-the-art performance in reconstruction accuracy and computational efficiency. We train a DC-CNN network with N cascaded blocks. Each block contains several convolutional layers and a data fidelity layer. The details of each block in the

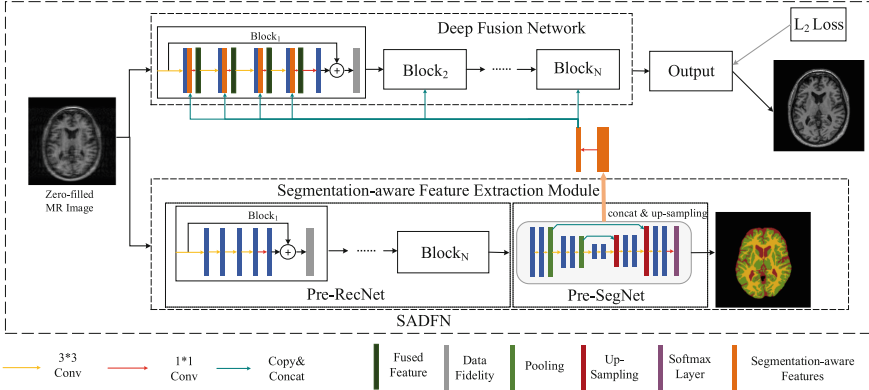


Fig. 2. The network architecture of SADFNet model.

DC-CNN architecture is shown in Table 1. The data fidelity layer enforces consistency between k-space value of the reconstructed image and the measured data. The details can also be found in [4]. Note the identity function is used in last convolutional layer to admit the negative values because of the global residual learning in the blocks. We also refer to the DC-CNN architecture as Pre-RecNet for simplicity. The Pre-RecNet with N blocks are called Pre-RecNet $_N$. We train the Pre-RecNet $_N$ using the under-sampled and full-sampled training data pairs by minimizing the following Euclidean loss function

$$\mathcal{L}_{\text{Rec}} \left(y_i, x_i^{fs}; \theta_r \right) = \frac{1}{L_r} \sum_{i=1}^{L_r} \left\| x_i^{fs} - f_{\theta_r} \left(F_u^H y_i \right) \right\|_2^2. \quad (2)$$

where the x_i^{fs} is the full-sampled MR image, y_i is the under-sampled k-space measurements in the training batch. θ_r denotes the network parameter and L_r is the number of MRI data in the training batch.

Table 1. The parameter setting of a block in the Pre-RecNet.

Layer	Input	Filter size	Stride	Number of filters	Activation	Output
Conv ₁	240 * 240	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₂	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₃	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₄	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₅	240 * 240 * 32	3 * 3	1	1	Linear	240 * 240
Data fidelity	240 * 240	N/A	N/A	N/A	N/A	240 * 240

3.2 The MRI Segmentation Network

To fully utilize the segmentation supervision information, we train a automatic segmentation network. We adopt the popular U-Net architecture as the segmentation model called Pre-SegNet. The parameter setting of the Pre-SegNet is shown in Table 2. The pooling operation can help the network extract the image features in different scales and the symmetric concatenation is utilized to propagate the low-layer features to high layers directly, providing accurate localization. We train the Pre-SegNet using the full-sampled MR images and their corresponding segmentation labels as training data pairs by minimizing the following pixel-wise cross-entropy loss function

$$\mathcal{L}_{\text{Seg}}(x_i^{fs}, t_i^{gt}; \theta_s) = - \sum_{i=1}^{L_s} \sum_{j=1}^R \sum_{c=1}^C t_{ijc}^{gt} \ln t_{ijc}. \quad (3)$$

where the t_i^{gt} is the segmentation label in the training batch and t_i is the corresponding segmentation result produced by Pre-SegNet. θ_s denotes the network parameter and L_s is the number of MRI data in the training batch. C denotes the number of classes of the label. Taking the brain segmentation for example [29], the brain tissues can be classified into white matter, gray matter, cerebrospinal fluid and background. Thus C is 4 for segmentation.

Table 2. The parameter setting of the Pre-SegNet.

Layer	Input	Filter size	Stride	Number of filters	Activation	Output
Conv ₁	240 * 240	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₂	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Max Pooling ₁	240 * 240 * 32	N/A	2	N/A	N/A	120 * 120 * 32
Conv ₃	120 * 120 * 32	3 * 3	1	64	ReLU	120 * 120 * 64
Conv ₄	120 * 120 * 64	3 * 3	1	64	ReLU	120 * 120 * 64
Max Pooling ₂	120 * 120 * 64	N/A	2	N/A	N/A	60 * 60 * 64
Conv ₅	60 * 60 * 64	3 * 3	1	128	ReLU	60 * 60 * 128
Conv ₆	60 * 60 * 128	3 * 3	1	128	ReLU	60 * 60 * 128
Deconv ₁	60 * 60 * 128	3 * 3	1	64	ReLU	120 * 120 * 64
Conv ₇	120 * 120 * (64 + 64)	3 * 3	1	64	ReLU	120 * 120 * 64
Conv ₈	120 * 120 * 64	3 * 3	1	64	ReLU	120 * 120 * 64
Deconv ₂	120 * 120 * 64	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₉	240 * 240 * (32 + 32)	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₁₀	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₁₁	240 * 240 * 32	3 * 3	1	5	Linear	240 * 240 * 5
Softmax	240 * 240 * 5	N/A	N/A	N/A	N/A	240 * 240

3.3 Deep Fusion Network

With the well-trained Pre-RecNet and Pre-SegNet, we can construct the segmentation-aware deep fusion network with N blocks (SADFN _{N}) by integrating the features from the Pre-RecNet and Pre-SegNet, which involving a cross-layer multilayer feature aggregation strategy and a cross-task feature fusion strategy.

Segmentation-Aware Feature Extraction Module. As we discussed in the related work section, the multilayer feature aggregation can be used to fuse the information from layers in different depth. Here we extract the feature maps from the output of the Conv₁, Conv₂, Conv₃, Conv₄, Conv₅, Conv₆, Conv₇, Conv₈, Conv₉, Conv₁₀ and concatenate them into a single “thick” feature map tensor. Note the smaller size feature maps are up-sampled using bilinear interpolation to the same size of features from the Pre-RecNet_N. Then the “thick” feature maps of the size $240 * 240 * 640$ ($32 + 32 + 64 + 64 + 128 + 128 + 64 + 64 + 32 + 32$) are further compressed into a “thin” feature tensor of the size $240 * 240 * 32$ via the 1×1 convolution with ReLU as activation function.

The Feature Fusion Cross Tasks. The compressed feature tensor obtained by the multilayer feature aggregation strategy contains the supervision information from the Pre-SegNet. We concatenate the feature tensor of the size $240 * 240 * 32$ with the feature maps of the size $240 * 240 * 32$ output by convolutional layers in the Pre-RecNet as shown in Fig. 2. Then the concatenated features of the size $240 * 240 * 64$ are further compressed into a feature tensor of the size $240 * 240 * 32$ via 1×1 convolution with ReLU activation function. The information from feature maps can be efficiently fused via such a concatenation and compression strategy. Note the compressed feature tensor is concatenated to the first four convolutional layers in each Pre-RecNet block, the supervision information from segmentation can guide the reconstruction in different depth. Also, in the Fig. 2, the feature fusion strategy is also utilized in each block of the Pre-RecNet.

To prove the supervision information is effectively fused into the reconstruction, we give some feature maps in the fused feature tensor yielded by the 1×1 convolution in Fig. 3. In Fig. 3(a) we show the segmentation label of a certain MRI data. In Fig. 3(b), (c) and (d), we visualize the feature maps selected from the fused feature tensors in the second layer and fourth layer. We observe the feature maps show clear segmentation information, while no such feature maps are observed in the Pre-RecNet_N model.

The Fine-Tuning Strategy. With the well-constructed deep fusion network, we further fine-tune the resulting architecture. Given a zero-filled MR image in the training dataset, a corresponding high-quality MR image can be yielded by the Pre-RecNet_N in Sect. 3.1. Then the MR image is sent to the Pre-SegNet to extract the segmentation features, which are then utilized for the multilayer feature aggregation in Pre-SegNet and feature fusion. Meanwhile, the zero-filled MR image is also input to the deep fusion network. The ℓ_2 Euclidean distance between the output reconstructed MR image and the corresponding full-sampled MR image in the training dataset is minimized. During the optimization, the parameters in the Pre-RecNet_N and Pre-SegNet are kept fixed, while we only adjust the parameters in the deep fusion network.

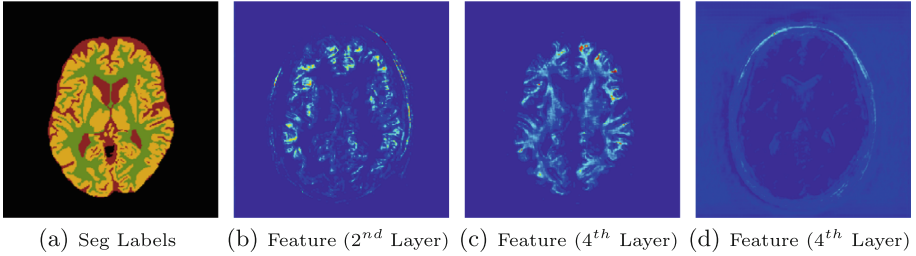


Fig. 3. The selected feature maps from the feature tensors produced by the feature fusion in the deep fusion network.

4 Experiments

4.1 Datasets

We train and test our SADFN model on MRBrainS datasets from Grand Challenge on MR Brain Image Segmentation (MRBrainS) Benchmark [29]. The datasets provides well-aligned multiple modalities MRI including T1, T1-IR and T2-FLAIR with segmentation labels by human experts. For simplicity, we only use the T1 weighted MRI data. In the future work, we plan to extend the model on multi-modalities MRI imaging. Total 5 scans are provided public segmentation labels. We randomly choose four scans for training containing total 172 slices. The training MR images are of size 240×240 . We use the remaining MRI scan for testing the model performance containing total 48 slices.

4.2 Implementation Details

We train and test the algorithm on Tensorflow for the Python environment on a NVIDIA Geforce GTX 1080Ti with 11 GB GPU memory and Intel Xeon CPU E5-2683 at 2.00 GHz. The detailed network architectures for Pre-RecNet, Pre-SegNet and SADFN have been introduced in previous section.

The ADAM is used as the optimizer. We train the Pre-RecNet for 32000 iterations using a batch containing four under-sampled and their corresponding full-sampled MR images as training pairs in Eq. 2. The Pre-SegNet is also pre-trained for 32000 iterations using a batch containing 16 randomly cropped fully-sampled 128×128 patches and their segmentation labels. Again, we note that during the fine-tuning of the SADFN model, compressed feature tensor is yielded by multilayer features aggregation (MLFA) and the feature tensor is propagated to the Pre-RecNet before the feature fusion in each block. The SADFN is fine-tuned 12000 iterations using the same training batchsize as the pre-training of Pre-RecNet. We select the initial learning rate to be 0.001 for pre-trained stage and 0.0001 for fine-tune stage, the first-order momentum to be 0.9 and the second momentum to be 0.999 for both stages. We adopt batch normalization (BN) in Pre-SegNet. We also adopt data augmentation for training as implemented in [30].

4.3 Quantitative Evaluation

We use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [31] for the reconstruction quantitative evaluation. We adopt a 30% 1D Cartesian pattern for under-sampling. We compare the proposed SADFN₅ with other state-of-the-art CS-MRI models including transform learning MRI (TLMRI) [12], patch-based nonlocal operator (PANO) [10], fast composite splitting algorithm (FCSA) [8], graph-based redundant wavelet transform (GBRWT) [11], and the deep models such as vanilla CNN [15], U-Net [17] the Pre-RecNet₅ (which is also the state-of-the-art DC-CNN with 5 blocks [4]). For the non-deep CS-MRI methods, we adjust the parameters to their best performance. We also compare the proposed SADFN₅ with the model proposed in [27], where the pre-trained Pre-RecNet₅ and Pre-SegNet are cascaded during fine-tuning and only the parameters in Pre-RecNet₅ are adjusted for optimization. Since no name for the model is provided in the original work, we refer the model as Liu [27]. Besides, we compare the proposed SADFN model with the model without the guidance of segmentation information (SADFN-WOS). For fair comparison, we design the building block of the SADFN-WOS network architecture in Table 3. Note the network architecture is kept unchanged with the only difference is some feature maps in SADFN come from Pre-SegNet while all the features come from the reconstruction network in SADFN₅-WOS. In the model Pre-RecNet₅ and SADFN₅-WOS, no segmentation label is utilize for training, meaning the corresponding supervision information is overlooked.

Table 3. The parameter setting of a block in the SADFN-WOS model

Layer	Input	Filter size	Stride	Number of filters	Activation	Output
Conv ₁	240 * 240	3 * 3	1	64	ReLU	240 * 240 * 64
Conv ₂	240 * 240 * 64	1 * 1	1	32	ReLU	240 * 240 * 32
Conv ₃	240 * 240 * 32	3 * 3	1	64	ReLU	240 * 240 * 64
Conv ₄	240 * 240 * 64	1 * 1	1	32	ReLU	240 * 240 * 32
Conv ₅	240 * 240 * 32	3 * 3	1	64	ReLU	240 * 240 * 64
Conv ₆	240 * 240 * 64	1 * 1	1	32	ReLU	240 * 240 * 32
Conv ₇	240 * 240 * 32	3 * 3	1	64	ReLU	240 * 240 * 64
Conv ₈	240 * 240 * 64	1 * 1	1	32	ReLU	240 * 240 * 32
Conv ₉	240 * 240 * 32	3 * 3	1	32	ReLU	240 * 240 * 32
Conv ₁₀	240 * 240 * 32	3 * 3	1	1	Linear	240 * 240
Data fidelity	240 * 240	N/A	N/A	N/A	N/A	240 * 240

We show the objective evaluation indexes in Fig. 4. Note the deep-based models outperform most non-deep CS-MRI models in reconstruction. We observe the proposed SADFN₅ model achieves the optimal performance in PSNR and SSIM indexes among the compared methods. From the standard deviation of the indexes. We note the improvement of the SADFN₅ is quite steady for different

MRI test data. We observe the model Liu [27] brings little improvement in objective evaluation indexes compared with the Pre-RecNet₅. We also observe the SADFN₅ model outperforms the comparative SADFN₅-WOS around 1dB in PSNR and 0.03 in SSIM in average, which proves the benefits are brought by introducing the supervision information from the segmentation labels instead of merely increasing the network size.

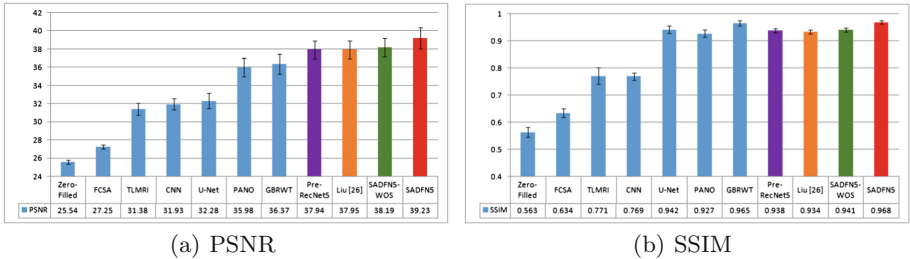


Fig. 4. The comparison in averaged PSNR and SSIM index on the test MRI data.

4.4 Qualitative Evaluation

We give the qualitative reconstruction results produced by compared CS-MRI methods in Fig. 5. We also plot the reconstruction error maps to better observe their differences. The display range for the error maps is [0 0.12]. We observe the Pre-RecNet₅ (DC-CNN [4]) architecture, produce better reconstruction than the conventional sparse- and nonlocal- regularized CS-MRI models. The model in [27] didn't brought significant improvement in reconstruction. The SADFN₅-WOS with larger network size also brought limited improvement. We observe the proposed SADFN₅ achieves much smaller reconstruction errors compared with other models, which is consistent with our observations in objective index evaluations.

4.5 Running Time

We compare the running time of the compared models in Table 4. As we mentioned in the Sect. 1, the CS-MRI models based on sparse or non-local regularization requires a large number of iterations, resulting slow reconstruction speed. Although the running time of the proposed SADFN model is slower than the other deep-based CS-MRI models, it achieves the state-of-the-art reconstruction accuracy, providing the best balance between running time and reconstruction quality.

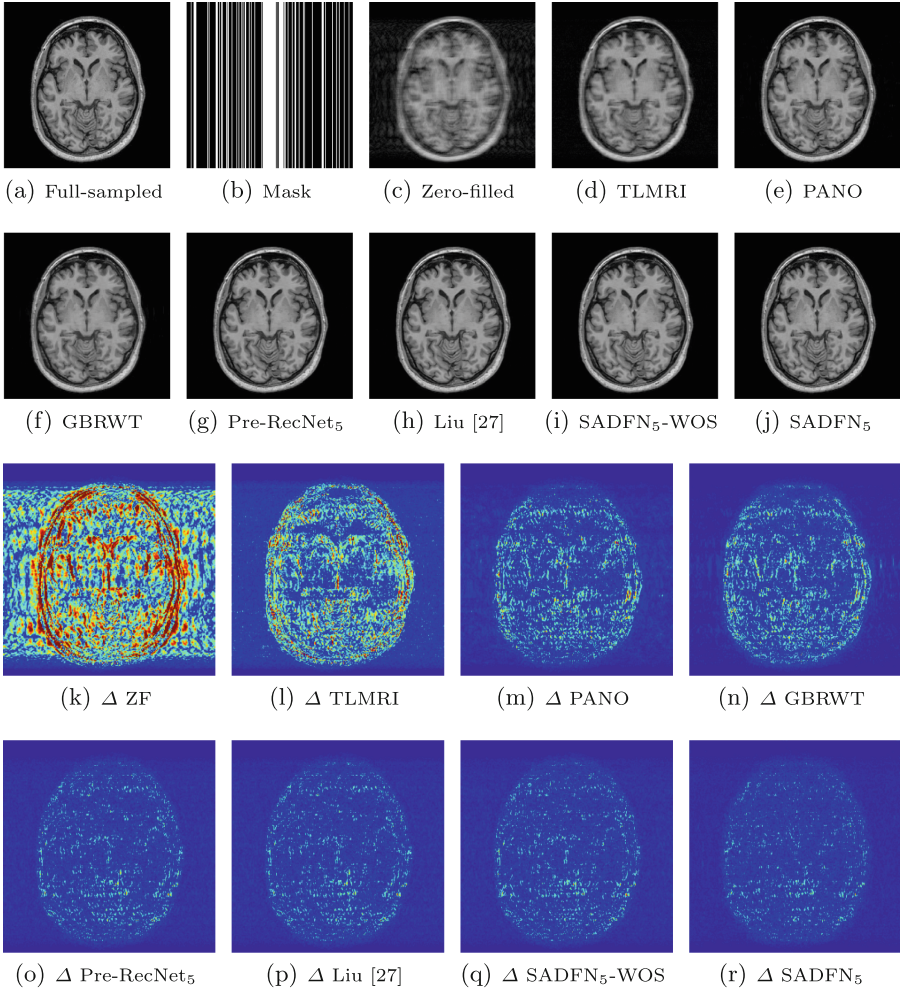


Fig. 5. The reconstruction results of zero-filled (ZF), TLMRI, PANO, GBRWT, Pre-RecNet₅, Liu [27], SADFNet₅-WOS and SADFNet₅. We also give the corresponding reconstruction error maps Δ with display ranges [0 0.12].

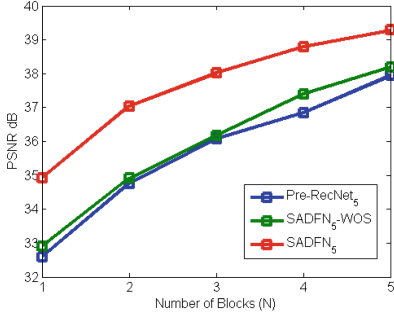
5 Discussions

5.1 The Number of Blocks

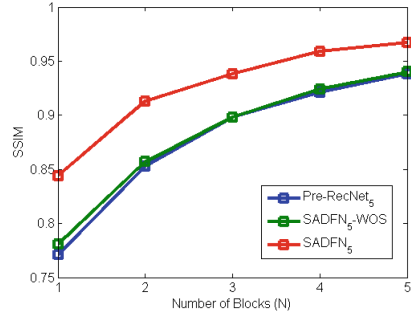
In Fig. 6, we discuss how the model performance varies with the different number of blocks from 1 to 5 in the Pre-RecNet₅, SADFNet₅-WOS and SADFNet₅ models. As expected, the SADFNet₅ model achieves steady improvement to large margins with different model capacity, meaning the supervision information can robustly improve the reconstruction accuracy.

Table 4. The comparison in runtime (seconds) between the compared models.

	TLMRI	GBRWT	PANO	Pre-RecNet ₅	Liu[26]	SADFN ₅ -WOS	SADFN ₅
Runtime	127.67	100.60	11.37	0.03	0.03	0.07	0.07



(a) PSNR



(b) SSIM

Fig. 6. The comparison in averaged PSNR and SSIM index on the test MRI data.

5.2 Different Under-Sampling Patterns

We also test the proposed SADFN model on the 20% Random under-sampling mask shown in Fig. 5. The SADFN₅ achieves the optimal performance, proving it can be well generalized on various kind of under-sampling patterns.

5.3 The Evaluation on the Segmentation Performance

With the reconstructed MR images produced by different CS-MRI models, we input them into the pre-trained automatic segmentation models in Sect. 3.2 to evaluate the effect of different reconstruction models on the segmentation task. We adopt the Dice Coefficient (DC), the 95th-percentile of the Hausdorff distance (HD) and the absolute volume difference (AVD) as objective evaluation indexes for segmentation as recommended in [29]. The higher DC, lower HD and lower AVD values indicate better segmentation accuracy. Details on evaluation of segmentation performance can be referred to [29]. The segmentation results with full-sampled MR image inputs are the performance upper bounds. We show the averaged segmentation results with compared models on the test MRI data set in Table 5. We observe the proposed SADFN₅ achieves the best accuracy on the segmentation task of the compared models.

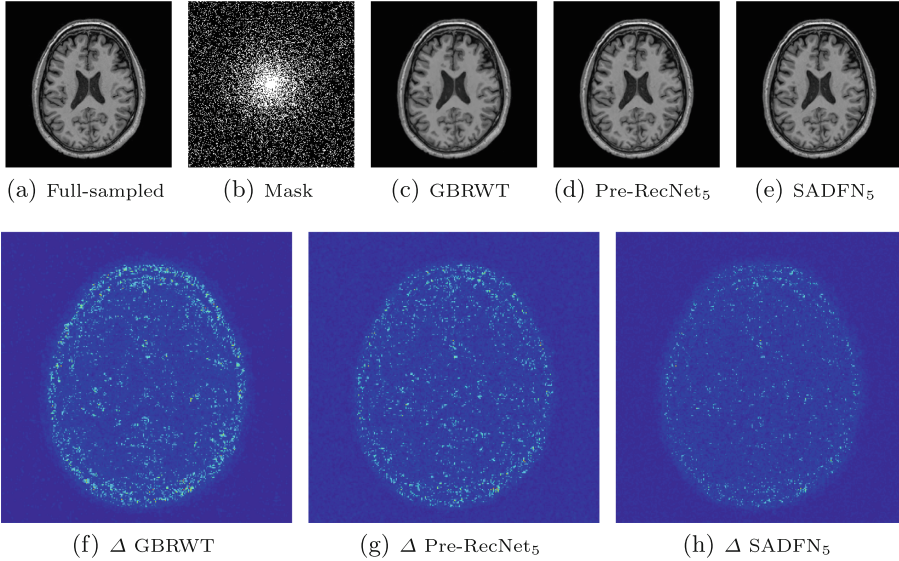


Fig. 7. The reconstruction results of zero-filled (ZF), TLMRI, PANO, GBRWT, Pre-RecNet₅, Liu [27], SADFN₅-WOS and SADFN₅ on the 20% random mask. We also give the corresponding reconstruction error maps Δ with display ranges [0 0.1].

Table 5. The averaged DC, HD and AVD values on the test MRI data.

Methods	GM			WM			CSF		
	DC %	HD	AVD	DC %	HD	AVD	DC %	HD	AVD
ZF + Pre-SegNet	64.78	2.587	6.202	54.07	2.085	4.294	57.37	2.221	4.689
TLMRI + Pre-SegNet	76.28	2.093	3.985	63.77	1.870	3.185	68.17	2.072	3.796
PANO + Pre-SegNet	83.73	1.819	2.958	75.72	1.348	1.815	78.93	1.653	2.361
GBRWT + Pre-SegNet	83.66	1.821	2.937	76.14	1.353	1.783	79.39	1.647	2.342
Pre-RecNet ₅ + Pre-SegNet	83.63	1.795	2.874	75.16	1.378	1.813	78.99	1.668	2.386
SADFN ₅ -WOS + Pre-SegNet	83.85	1.782	2.838	75.84	1.357	1.762	79.25	1.661	2.364
Liu [27] + Pre-SegNet	84.08	1.776	2.814	76.30	1.335	1.724	79.37	1.661	2.357
SADFN ₅ + Pre-SegNet	85.76	1.690	2.579	81.29	1.143	1.381	80.08	1.649	2.305
Full-sampled + Pre-SegNet	87.30	1.596	2.328	86.89	0.973	1.092	80.76	1.617	2.225

6 Conclusion

In this paper, we proposed a segmentation-aware deep fusion network (SADFN) for compressed sensing MRI. We showed the high-level supervision information can be effectively fused into deep neural network models to help the low-level MRI reconstruction. The multilayer feature aggregation is adopted to fuse cross-layer information in the MRI segmentation network and the feature fusion strategy is utilized to fuse cross-task information in the MRI reconstruction network. We prove the proposed SADFN architecture enables the reconstruction network aware of the contents it reconstructs and the function mapping can be signif-

icantly simplified. The SADFN model achieves state-of-the-art performance in CS-MRI and balance between accuracy and efficiency (Fig. 7).

References

1. Atkinson, D., et al.: Automatic compensation of motion artifacts in MRI. *Magn. Reson. Med.* **41**(1), 163–170 (1999)
2. Jung, H., Sung, K., Nayak, K.S., Kim, E.Y., Ye, J.C.: k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magn. Reson. Med.* **61**(1), 103–116 (2009)
3. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**(6), 1182–1195 (2007)
4. Schlemper, J., et al.: A deep cascade of convolutional neural networks for MR image reconstruction. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) *IPMI 2017*. LNCS, vol. 10265, pp. 647–658. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_51
5. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging* **30**(5), 1028–1041 (2011)
6. Ma, S., Yin, W., Zhang, Y., Chakraborty, A.: An efficient algorithm for compressed MR imaging using total variation and wavelets. In: *CVPR*, pp. 1–8. IEEE (2008)
7. Yang, J., Zhang, Y., Yin, W.: A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. *IEEE J. Sel. Top. Sig. Process.* **4**(2), 288–297 (2010)
8. Huang, J., Zhang, S., Metaxas, D.: Efficient MR image reconstruction for compressed MR imaging. *Med. Image Anal.* **15**(5), 670–679 (2011)
9. Qu, X., Guo, D., Ning, B., Hou, Y., Lin, Y., Cai, S., Chen, Z.: Undersampled MRI reconstruction with patch-based directional wavelets. *Magn. Reson. Imaging* **30**(7), 964–977 (2012)
10. Qu, X., Hou, Y., Lam, F., Guo, D., Zhong, J., Chen, Z.: Magnetic resonance image reconstruction from undersampled measurements using a patch-based non-local operator. *Med. Image Anal.* **18**(6), 843–856 (2014)
11. Lai, Z., et al.: Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform. *Med. Image Anal.* **27**, 93–104 (2016)
12. Ravishankar, S., Bresler, Y.: Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM J. Imaging Sci.* **8**(4), 2519–2557 (2015)
13. Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., Zhang, X.P.: Bayesian non-parametric dictionary learning for compressed sensing MRI. *IEEE Trans. Image Process.* **23**(12), 5007–5019 (2014)
14. Dong, W., Shi, G., Li, X., Ma, Y., Huang, F.: Compressive sensing via nonlocal low-rank regularization. *IEEE Trans. Image Process.* **23**(8), 3618–3632 (2014)
15. Wang, S., et al.: Accelerating magnetic resonance imaging via deep learning. In: *ISBI*, pp. 514–517. IEEE (2016)
16. Caballero, J., Bai, W., Price, A.N., Rueckert, D., Hajnal, J.V.: Application-driven MRI: joint reconstruction and segmentation from undersampled MRI data. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014*. LNCS, vol. 8673, pp. 106–113. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_14

17. Lee, D., Yoo, J., Ye, J.C.: Deep residual learning for compressed sensing MRI. In: ISBI, pp. 15–18. IEEE (2017)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
20. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* **170**, 446–455 (2018)
21. Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J.: Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: NIPS, pp. 2998–3006 (2015)
22. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: NIPS, pp. 3036–3044 (2016)
23. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
24. Li, G., Yu, Y.: Deep contrast learning for salient object detection. In: CVPR, pp. 478–487 (2016)
25. Zhang, P., Wang, D., Lu, H., Wang, H., Ruan, X.: Amulet: aggregating multi-level convolutional features for salient object detection. In: ICCV, October 2017
26. Hou, Q., Cheng, M.M., Hu, X., Borji, A., Tu, Z., Torr, P.: Deeply supervised salient object detection with short connections. In: CVPR, pp. 5300–5309 (2017)
27. Liu, D., Wen, B., Liu, X., Huang, T.S.: When image denoising meets high-level vision tasks: a deep learning approach. arXiv preprint [arXiv:1706.04284](https://arxiv.org/abs/1706.04284) (2017)
28. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: AOD-Net: all-in-one dehazing network. In: ICCV, October 2017
29. Mendrik, A.M., et al.: MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Comput. Intell. Neurosci.* **2015**, 1 (2015)
30. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 506–517. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_44
31. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Process.* **13**(4), 600–612 (2004)
32. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: CVPR, pp. 2790–2798 (2017)