A Wide Multimodal Dense U-Net for Fast Magnetic Resonance Imaging

Antonio Falvo, Danilo Comminiello, Simone Scardapane, Michele Scarpiniti, and Aurelio Uncini
Department of Information Engineering, Electronics and Telecommunications
Sapienza University of Rome, Italy

***Abstract*—In this paper, a deep learning method for accelerating magnetic resonance imaging (MRI) is presented, which is able to reconstruct undersampled MR images obtained by reducing the k-space data in the direction of the phase encoding. In particular, we focus on the reconstruction of MR images related to patients affected by multiple sclerosis (MS) and we propose a new multimodal deep learning architecture that is able to exploit the joint information deriving from the combination of different types of MR images and to accelerate the MRI, while providing high quality of the reconstructed image. Experimental results show the performance improvement of the proposed method with respect to existing models in reconstructing images with an MRI acceleration of 4 times.**

***Index Terms*—Fast MRI, MR Image Reconstruction, Deep Neural Network, Multimodal Dense U-Net, Multiple Sclerosis**

I. Introduction

The acceleration of magnetic resonance imaging (MRI) analysis has become an attractive research topic in the last years due to its potential advantages, such as higher availability of MRI scanners and lower healthcare costs for patients. The MRI process needs to sample as much spatial frequencies as possible in the k-space to yield detailed and high-quality images [1]. Therefore, one of the possibilities to reduce the MR scanning time is to capture only a subset of the frequency lines of the k-space and then reconstruct the undersampled image.

The most popular methods proposed in the literature for fast MRI rely on parallel imaging [2]–[4], in which the k-space sampling depends on the simultaneous acquisition from multiple coils, and on compressed sensing [5]–[8], which allows to capture as few k-space frequency lines as necessary. Recently, the acceleration of MRI has been addressed by using deep learning methods, which have shown significant capabilities in reconstructing undersampled MR images [9], [10]. In particular, in the very last years, several deep neural models have been proposed, e.g., based on convolutional neural networks [11]–[13] or on generative adversarial networks [14]–[16], among others.

Some medical diagnoses often require the analysis of different kinds of MR images to be accurate. One of the most significant examples is represented by the multiple sclerosis (MS) diagnosis that is often based on the analysis of the conventional MR sequences, like T1 Weighted (T1W) and T2 Weighted (T2W) images, and also of particular sequences, such as FLAIR and DIR, which allow to stand out the lesions due to the disease. In this way, the MRI analysis is able to detect the MS lesions, which will have different characteristics, e.g., depending on whether they are recent or not.

In order to exploit the different kinds of MR images, a multimodal approach can be adopted to exploit complementary information, thus enhancing the reconstruction of the full MR image. Recently, deep learning methods based on multimodal approaches have been proposed in the literature. One of the most popular tasks is to estimate T2W images (T2WIs) from T1W images, which have lower scan times [17]. Multimodal approaches can be also used to improve the quality of the undersampled images from high resolution images with different contrast [18]. In our preliminary work [19], for the first time an attempt has been made to reconstruct T2WIs from undersampled T2WIs and FLAIR images. In [20], it has been also proved that modifying the dense block appropriately leads to an improvement of the results while maintaining high image quality in the area of lesions.

In this work, we propose a new multimodal deep neural network for accelerating MRI that receives the two kind of MR images as input, T2WIs and FLAIR, separately. Compared with existing multimodal architectures [19], [20], the new model processes the two input images separately in the contracting path in order to exploit the correlation between images of the same kind and optimize the information extraction. Moreover, the new architecture involves a dense block that does not increase the computational complexity. Results prove that the proposed method is able to enhance the reconstruction quality of T2WIs with respect to existing methods, with an acceleration factor of 4 times for the MRI analysis.

The rest of the paper is organized as follows. In Section II, we introduce the problem and then present the proposed framework for the MR image reconstruction. The proposed Wide Multimodal Dense U-Net is described in Section III and experimental results are shown in Section IV. Finally, our conclusion is drawn in Section V.

II. Proposed Approach

A. General Framework

Figure 1 shows the framework implementing the adopted strategy for the reconstruction of undersampled MR images.

Corresponding author's email: danilo.comminiello@uniroma1.it.

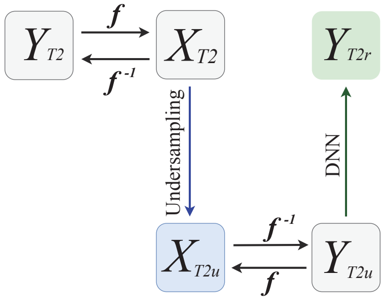

Fig. 1. General framework of the proposed approach.

The fully sampled image, $Y_{T2}\left[k_1, k_2\right]$, represents the reconstruction target for the T2WI, which can be expressed as:

$$Y_{T2}\left[k_1, k_2\right]=\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} S_{T2}\left[n_1, n_2\right] e^{j \frac{2 \pi}{N} k_1 n_1} e^{j \frac{2 \pi}{N} k_2 n_2} \quad (1)$$

where $S_{T2}\left[n_1, n_2\right]$ is the MR signal received in position $\left(2 \pi n_1 / N, 2 \pi n_2 / N\right)$, beeing N_1 the axis of the frequency encoding and N_2 the axis of the phase encoding.

We transform $Y_{T2}\left[k_1, k_2\right]$ into the k-space domain of spatial frequencies by applying a 2D Fourier transform:

$$\begin{aligned} X_{T2}\left[n_1, n_2\right]=\frac{1}{N_1 N_2} & \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} Y_{T2}\left[k_1, k_2\right] \\ & \cdot e^{-j \frac{2 \pi}{N} k_1 n_1} e^{-j \frac{2 \pi}{N} k_2 n_2} \end{aligned} \quad (2)$$

that can be also expressed as:

$$X_{T2}=\left|X_{T2}\right| e^{j \angle X_{T2}}. \quad (3)$$

Then, we undersample X_{T2} by using a masking transform, which is detailed in Subsection II-B, denoted as $M(\kappa, c)$, where κ is the undersampling factor and c the percentage of samples to be considered in the central area. We achieve an undersampled version of the image in the k-space X_{T2u}:

$$X_{T2u}=M(\kappa, c) \cdot\left|X_{T2}\right|. \quad (4)$$

Anti-transforming X_{T2u}, we obtain the undersampled T2W image Y_{T2u} that is used as input to the proposed deep neural network (DNN) model presented in Section III, thus yielding the reconstructed T2W image Y_{T2r}. This last image may be also subject of a further post-processing in the spatial domain to recover the original spatial frequencies (see [20]).

B. Proposed Undersampling Strategy

One of the most critical processes in the described framework involves the choice of the undersampling strategy, i.e., how to set the function $M(\kappa, c)$ in (4) that allows to achieve X_{T2u}. In particular, in [20], we proposed an undersampling strategy consisting in preserving as much as possible the information contained in the k-space of the target image. It allows to concentrate most of the available samples in the central area of the k-space where the low-frequency components responsible for the image contrast information reside. The rest of the samples are distributed in an equidistant manner in the high frequency areas.

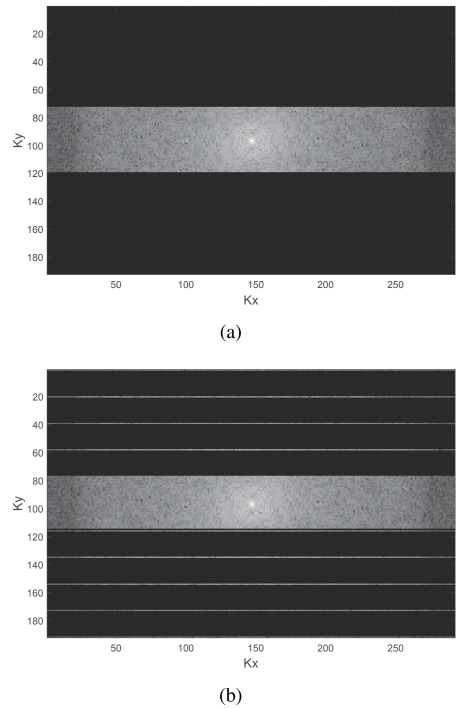

Fig. 2. a) Classic and b) proposed undersampling masks.

With respect to the classic uniform undersampling strategies, the proposed method gives higher relevance to the information present at low frequencies by taking a larger number of lines in the center of the k-space, which is the area containing about 95% of the total spectral power, thus guaranteeing the condition of separability (i.e., different images have the same undersampling in the k-space) that may be not guaranteed as well in uniform undersampling.

Based on the above consideration it is possible to define an undersampling strategy by opportunely choosing the values of κ and c. Keeping fixed the value of κ, as c increases, a higher number of samples is taken in the low-frequency area at the expense of high frequencies.

In Fig. 2, two different types of masks are compared, both obtained by setting an undersampling factor $\kappa=4$. In the first mask (Fig. 2(a)), that we denote as central mask, samples are taken exclusively in the central area of the k-space, while the proposed mask in Fig. 2(b) is defined by taking the 80% (i.e., $c=80$) of the total samples from the center and the remaining in an equidistant manner so as to have information even in the high frequencies. The proposed undersampling mask has been also proved to obtain an image reconstruction with a performance improvement of about 15% of accuracy, compared with the central mask [19].

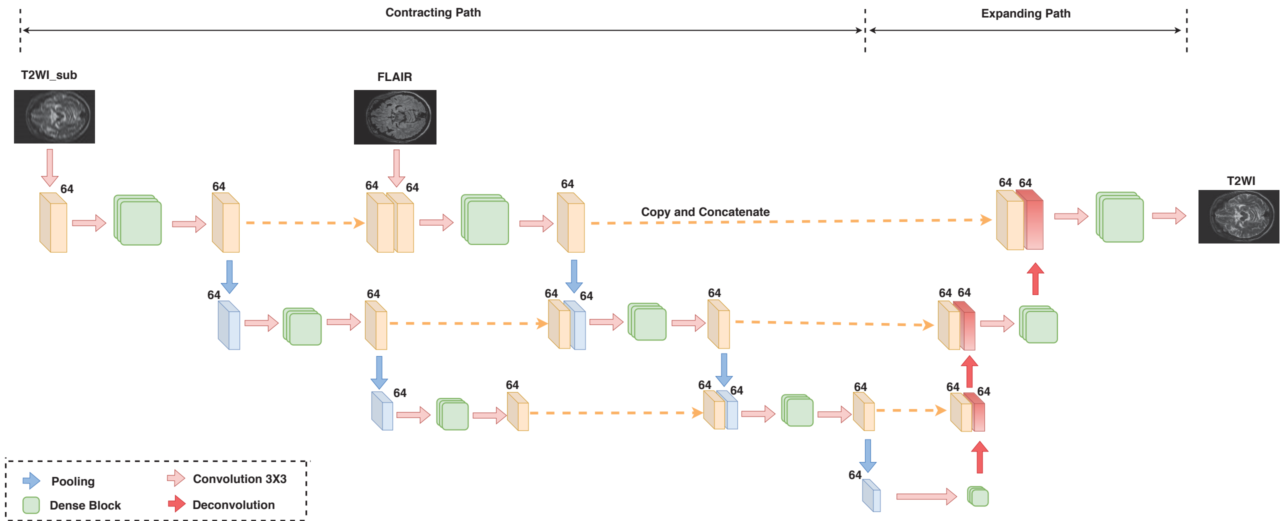

Fig. 3. Scheme of the proposed Wide Multimodal Dense U-Net architecture.

III. Wide Multimodal Dense U-Net

The framework described in Fig. 1 can be implemented by using any DNN model to achieve the predicted image Y_{T2r}. Here we propose an improved method of the Multimodal Dense U-Net (MDU) architecture presented in [20]. The MDU method is motivated by the fact that in the MRI analysis for MS diseases useful information can be obtained by using different type of sequences. In particular, we focus on the T2WIs as input to be reconstructed (i.e., Y_{T2} in Fig. 1), but we also include in input the FLAIR sequences, which provide further information to stand out brain lesions.

The proposed Wide Multimodal Dense U-Net (W-MDU) is depicted in Fig. 3. The network consists essentially of 4 components, namely convolutive layers, pooling layers, deconvolutive layers and dense blocks. The size of the characteristic map decreases along the contraction path through the pooling blocks as it increases along the expansion path by deconvolution. With respect to the MDU, the new architecture receives the two kind of images as input separately, each of which is subjected to a pooling process. Pooling partitions the input image into a set of squares, and for each of the resulting regions returns the maximum value as output. Its purpose is to progressively reduce the size of the representations, so as to reduce the number of parameters and the computational complexity of the network, at the same time counteracting any overfitting. This process does not suffer from any loss of information since at each contraction the feature maps of the first input are concatenated with those of the second input.

The expanding path begins after the third pooling level, in which we have a sequence of deconvolutive levels allowing to obtain an incremental recovering of the original size. Deconvolutive layers act inversely with respect to pooling and aim to increase the spatial dimensions of the inputs. This allows to obtain images of comparable size to that of the input images to the network. In the simplest case, these levels can be implemented as static oversampling with bilinear interpolation.

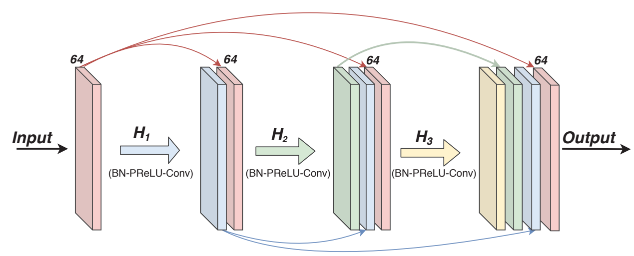

Fig. 4. Scheme of the dense block used in the proposed W-MDU. The network complexity does not increase since we have chosen a null GR.

The dense block, similarly to [21], allows to effectively increase the depth of the entire network while maintaining a low complexity. Moreover, it requires less parameters to be trained. The dense block consists of three consecutive steps: batch normalization (BN), activation functions and 3×3 convolution filters. In particular, the proposed dense block is characterized by parametric rectified linear unit (PReLU) activation functions, which attempt to solve the problem of neuron deactivation of the classic ReLU. The PReLU introduces a small negative slope (α) in the region where the ReLU is zero. α is not a constant but a parameter that is estimated on-par with other network weights. The proposed dense block also involves dropout, which reduces the generalization error. The hyperparameters for the dense block are the growth rate (GR) and the number of convolutional layers (NC). The scheme of the proposed dense block is represented in Fig. 4 with a null growth rate for each layer (i.e., GR $= 0$) and a number of convolutional layers NC $= 4$. The network ends with a reconstruction level consisting of a dense block followed by a 1×1 convolutional layer that yields the reconstructed T2WI.

The proposed W-MDU is based on the minimization of the following loss function:

$$\arg\min\left\{\mathrm{MSE}+\mathrm{DSSIM}\right\} \tag{5}$$

where the MSE denotes the mean-square error, defined as:

$$\mathrm{MSE}\left(Y_{T2},Y_{T2r}\right)=\frac{1}{N}\sum_{i=1}^{N}\left(Y_{T2,i}-Y_{T2r,i}\right)^{2}, \tag{6}$$

and DSSIM is the structural dissimilarity index, defined as:

$$\mathrm{DSSIM}\left(Y_{T2},Y_{T2r}\right)=\frac{1}{2}\left(1-\mathrm{SSIM}\left(Y_{T2},Y_{T2r}\right)\right), \tag{7}$$

being $\mathrm{SSIM}\left(Y_{T2},Y_{T2r}\right)$ the structural similarity index (SSIM):

$$\mathrm{SSIM}\left(Y_{T2},Y_{T2r}\right)=\frac{\left(2\mu_{Y}\mu_{Y_r}+c_1\right)\left(2\sigma_{YY_r}+c_2\right)}{\left(\mu_Y^2+\mu_{Y_r}^2+c_1\right)\left(\sigma_Y^2+\sigma_{Y_r}^2+c_2\right)}. \tag{8}$$

In (8), μ_Y, μ_{Y_r} are mean values, σ_Y^2 and $\sigma_{Y_r}^2$ variances, and σ_{YY_r} a covariance.

IV. Experimental Results

A. Dataset and Network Setting

We consider a dataset containing MRIs of 30 patients suffering from MS disease [22], including T2W and FLAIR images, among others. The final voxel size of images is $0.46\times0.46\times0.8$ mm^3, which is transformed in isotropic size of $0.8\times0.8\times0.8$ mm^3. We have also performed a further preprocessing to extract slices of size 192×292 and to shrink intensity to the range $[0,1]$.

For each patient we provide the network with 150 FLAIR and T2W undersampled images using the fully sampled T2WI as target. For the dense blocks we set a null GR and NC = 4. The setting of GR = 0 in the dense block allows us to keep a fixed size of the feature maps (i.e., 64, as depicted in Fig. 4). We use the Adam optimizer for training. We ran the network for 100 epochs also using an early stopping with respect to the validation loss. Finally, we use the MSE (6) and SSIM (8) metrics to evaluate quantitatively the reconstruction performance.

B. Network Evaluation

We evaluate the effectiveness of the proposed W-MDU network by comparing the reconstruction quality of the T2WIs yielded by the proposed W-MDU with those achieved by the standard Dense U-Net (DU) [21] and the Multimodal Dense U-Net (MDU) [20].

In particular, using an acceleration factor of $\kappa=4$, the MDU is able to outperform the Dense U-Net [19]. On one hand, using the center mask, the DU gets a 71% reconstruction percentage compared to the target, while using the proposed custom mask the similarity index rises up to 86%, as shown in Fig. 5(a). On the other hand, the most recent MDU [20] using the custom mask is able to outperform the DU and achieve an improvement of the SSIM up to 92% Fig. 5(b). The SSIM improvement, however, is obtained at the expenses of the MSE, which is slightly decreased. Finally, the proposed W-MDU exploits the new network architecture and provides an improvement of the MSE with respect to both the DU and the MDU, as shown in Fig. 5(c). Moreover, the W-MDU also achieves an accuracy of the reconstruction up to the 95%, thus outperforming both the previous networks.

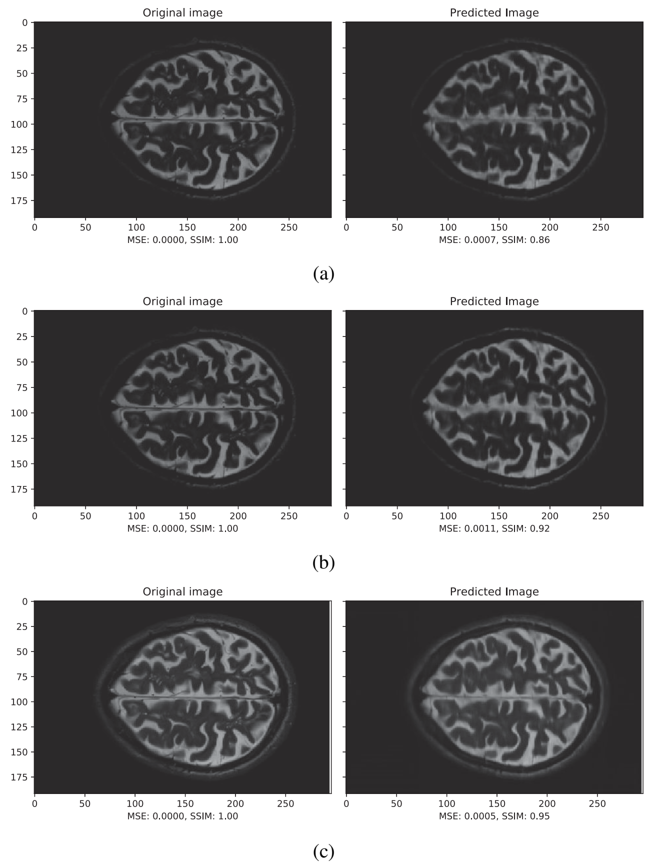

Fig. 5. Results of the image reconstruction with an undersampling factor of $\kappa=4$ using: a) a Dense U-Net [21], b) the Multimodal Dense U-Net [20] and c) the proposed Wide MDU network.

V. Conclusion

In this work, we proposed a neural model based on a U-Net network, which is able of exploiting the potential of the multimodal approach using a new separate scheme for the initial pooling. Our approach allows to reconstruct T2WIs, undersampled by a factor of 4, exploiting the correlation that exists with FLAIR images. A high quality of image reconstruction is preserved making possible to easily stand out the brain lesions due to MS disease. Comparisons with the state-of-the-art models have shown outperforming performance in terms of perceptive quality of the image reconstruction.

Future work will involve network improvements to enhance the reconstructed accuracy even using higher acceleration factors.

References

[1] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging. A Signal Processing Perspective*. New York, NY: The Institute of electrical and Electronics Engineers, 2000.

[2] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 694–706, 2011.

[3] X. Ye, Y. Chen, and F. Huang, "Computational acceleration for MR image reconstruction in partially parallel imaging," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1055–1063, May 2011.

[4] J. Lyu, U. Nakarmi, D. Liang, J. Sheng, and L. Ying, "KerNL: Kernel-based nonlinear approach to parallel MRI reconstruction," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 312–321, Jan. 2019.

[5] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[6] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing MRI with random encoding," *IEEE Trans. Med. Imaging*, vol. 30, no. 4, pp. 893–903, Apr. 2012.

[7] Y. Wang and L. Ying, "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1109–1120, Apr. 2014.

[8] F. Ong, R. Heckel, and K. Ramchandran, "A fast and robust paradigm for Fourier compressed sensing based on coded sampling," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 5117–5121.

[9] Y. Han, L. Sunwoo, and J. C. Ye, "k-space deep learning for accelerated MRI," *IEEE Trans. Med. Imaging*, pp. 1–1, 2019.

[10] M. Golbabaee, D. Chen, P. A. Gómez, M. M. Menzel, and M. E. Davies, "Geometry of deep learning for magnetic resonance fingerprinting," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 7825–7829.

[11] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 491–503, 2018.

[12] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 280–290, Jan. 2019.

[13] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhiri, J. Qi, and Q. Li, "Iterative PET image reconstruction using convolutional neural network representation," *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 675–685, 2019.

[14] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.

[15] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[16] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[17] L. Xiang, Y. Chen, W. Chang, Y. Zhan, W. Lin, Q. Wang, and D. Shen, "Deep leaning based multi-modal fusion for fast MR reconstruction," *IEEE Trans. Biomed. Eng.*, 2018.

[18] K. H. Kim, W.-J. Do, and S.-H. Park, "Improving resolution of MR images with an adversarial network incorporating images with different contrast," *Medical Physics*, vol. 45, no. 7, pp. 3120–3131, Jul. 2018.

[19] A. Falvo, D. Comminiello, S. Scardapane, M. Scarpiniti, and A. Uncini, "A multimodal deep network for the reconstruction of T2W MR images," in *arXiv preprint arXiv:1908.03009v2*, Jun. 2019.

[20] ——, "A multimodal dense U-net for accelerating multiple sclerosis MRI," in *IEEE Int. Workshop on Machine Learning for Signal Process. (MLSP)*, Pittsburgh, PA, Oct. 2019, pp. 1–6.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 2261–2269.

[22] L. Žiga, A. Galimzianova, A. Koren, M. Lukin, F. Pernuš, B. Likar, and v. Špiclin, "A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus," *Neuroinf.*, vol. 16, pp. 51–63, 2018.