Segmentation-Aware MRI Reconstruction

Mert Acar[1,2(✉)], Tolga Çukur[1,2], and İlkay Öksüz[3]

[1] Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
mert.acar@bilkent.edu.tr
[2] National Magnetic Resonance Research Center, Bilkent University, Ankara, Turkey
[3] Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey

Abstract. Deep learning models have been broadly adopted for accelerating MRI acquisitions in recent years. A common approach is to train deep models based on loss functions that place equal emphasis on reconstruction errors across the field-of-view. This homogeneous weighting of loss contributions might be undesirable in cases where the diagnostic focus is on tissues in a specific subregion of the image. In this paper, we propose a framework for segmentation-aware reconstruction based on segmentation as a proxy task. We leverage an end-to-end model comprising reconstruction and segmentation networks; and leverage backpropagation of segmentation error to devise a pseudo-attention effect to focus the reconstruction network. We introduce a novel stabilization method to prevent convergence onto a local minima with unacceptably poor reconstruction or segmentation performance. Our stabilization approach initiates learning on fully-sampled acquisitions, and gradually increases the undersampling rate assumed in the training set to its desired value. We validate our approach for cardiac MR reconstruction on the publicly available OCMR dataset. Segmentation-aware reconstruction significantly outperforms vanilla reconstruction for cardiac imaging.

Keywords: Cardiac MRI · Reconstruction · Segmentation · Convolutional neural networks

1 Introduction

Magnetic resonance imaging (MRI) is an essential part of routine clinical practice for noninvasive disease detection and monitoring. Yet, its prolonged scan times limit broad utilization. Shortening scan durations can improve patient throughput and lower motion-based image artifacts. However, undersampled k-space acquisitions elicit well-known aliasing artifacts that must be suppressed to obtain diagnostic-quality images. Artifact suppression involves an inverse problem solution where high-quality images are reconstructed from undersampled k-space data [1].

In recent years, deep learning (DL) models have become a preferred approach for accelerated MRI reconstruction [1–4]. DL models are trained to recover

N. Haq et al. (Eds.): MLMIR 2022, LNCS 13587, pp. 53–61, 2022.
https://doi.org/10.1007/978-3-031-17247-2_6

high-quality images consistent with fully-sampled acquisitions given as input undersampled acquisitions. Training is typically performed with a global loss function expressed over the entire image. However, such global loss functions are often dominated by diagnostically-irrelevant background tissues.

Self-attention mechanisms have been proposed to help focus the model's attention on subregions where there is greater tendency for introducing reconstruction errors [5]. Note that vanilla attention mechanisms are not explicitly informed regarding the underlying anatomy. Other studies have considered more direct guidance from segmentation maps to focus on reconstruction performance in regions of interest [6–8]. Pre-trained segmentation models have been transferred to mitigate problems associated with joint training of the reconstruction-segmentation network. Few recent studies have considered joint training of reconstruction-segmentation models, where an alternating optimization is performed between the two networks [9]. However, such alternative optimization is prone to premature stopping of learning where one of the networks performs unacceptably poorly.

In this work, we propose a segmentation-aware reconstruction method. The proposed method is based on a sequential architecture containing two networks for reconstruction and segmentation. To avoid premature stopping of learning, a stabilization approach is introduced for end-to-end training of the model. In particular, the undersampling rate is gradually decreased during the course of training. A composite reconstruction-segmentation loss is used, and errors back-propagated from the segmentation stage are used to focus the reconstruction on critical image regions. Experiments were conducted on a public cardiac MRI dataset [10]. Our results clearly indicate that the proposed segmentation-aware reconstruction improves focal image quality over solo reconstruction methods and unstabilized joint reconstruction-segmentation methods.

2 Methods

2.1 Proposed Framework

We propose an end-to-end training of reconstruction and segmentation networks to enable high reconstruction quality for a target region of interest (ROI). Figure 1 illustrates the proposed framework for segmentation-aware reconstruction for cardiac MRI reconstruction. Our framework is general in the sense that many different state-of-the-art architecture for reconstruction and segmentation modules can be utilized. The specific pairs of architectures that we examine in the current study are described in implementation details.

2.2 Stabilization

Our framework rests on the key notion of employing segmentation loss to focus the reconstruction process on tissues of high diagnostic interest. Accordingly, we perform end-to-end training of a sequential cascade of reconstruction and

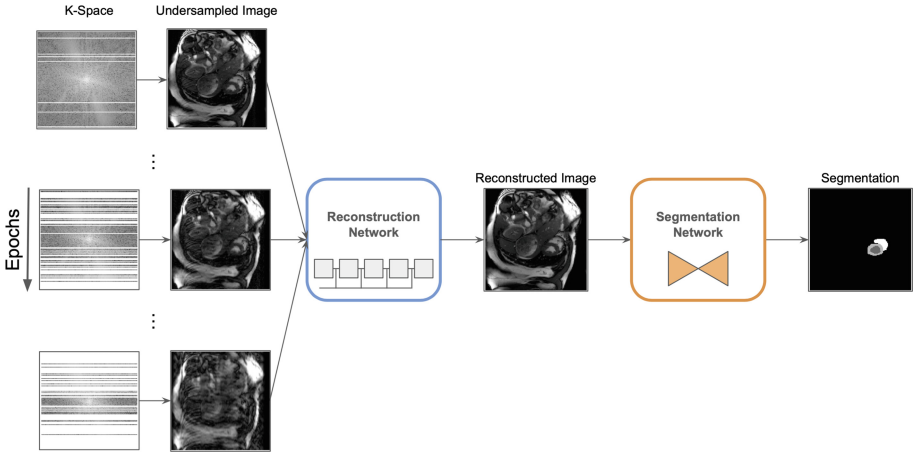


Fig. 1. Segmentation-aware reconstruction framework. During training the network is trained end-to-end with stabilization. During inference an image is recollected from only the reconstruction network, along with an auxiliary segmentation map. Stabilization technique is illustrated by the sample inputs on the left across epochs

segmentation networks. Training of such compound networks is prone to undesirable convergence onto local minima where either network yields undesirably poor performance. During initial stages of training, the segmentation network receives an input computed by an insufficiently trained reconstruction network, so the input will contain a high degree of reconstruction artifacts. This will inevitably compromise the learning process for the segmentation stage, resulting in inaccurate segmentation maps. Note that the segmentation maps are then provided as guidance to the reconstruction network, so a vicious circle can be created where both network are compromised.

To address this critical issue, we propose to use a novel stabilization method where the undersampling rate of the acquisitions are gradually ramped up during the course of training. In the initial stages, the reconstruction network receives lightly undersampled data that is easy to reconstruct with few artifacts. Thus, the segmentation network receives as input high-quality reconstructions that will improve its learning capabilities. Once both network adapt to the instant acceleration rate, then the degree of undersampling can be elevated. Overall, the learning signals generated from the high-quality inputs in earlier epochs are propagated back to "warm-up" the model for the increasingly lower quality samples that are to come in later epochs. In particular, we impose an epoch-specific undersampling rate starting from $1 - \epsilon$ reaching to the desired undersampling rate r in an exponential manner as follows:

$$r_i = \begin{cases} (1-\epsilon)\left(\frac{r}{1-\epsilon}\right)^{(i-1)/P} & 1 \leq i \leq P \\ r & i > P \end{cases} \tag{1}$$

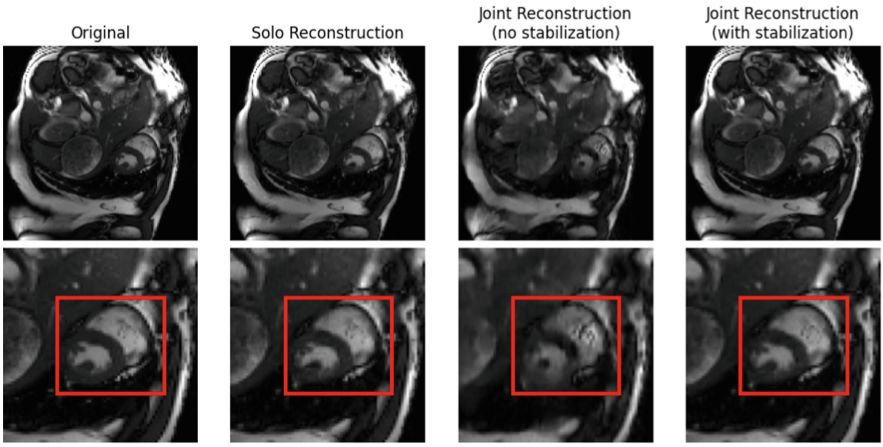
where i denotes the epoch number starting from 1, P governs the stabilization patience and ϵ is a small number close to 0. Therefore, the sub-optimal convergence problem can be mitigated with small perturbations on the task and both networks can be updated for P epochs with healthy gradients to prime the segmentation network to generate meaningful learning signal in focusing the reconstructions around the regions of interest which are dominated by the segmentation maps.

2.3 Model Architectures

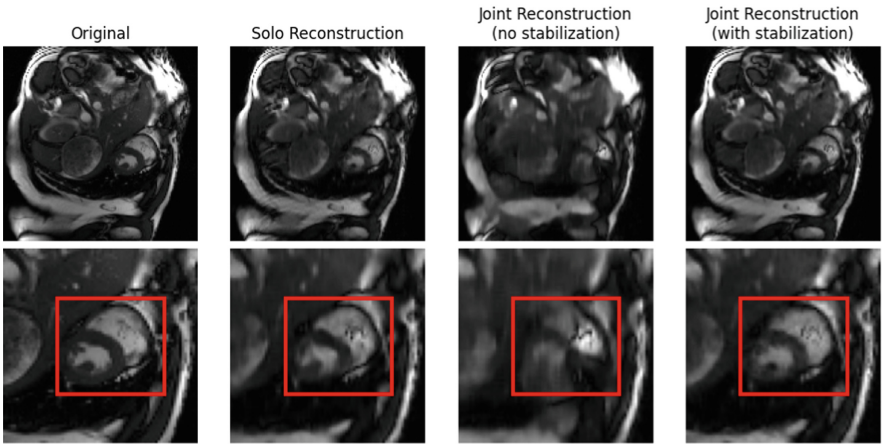
We tested our proposed method with various configurations of reconstruction and segmentation networks. First, we used the U-Net architecture with depth of 4 and filter configuration of (32, 64, 128, 256) with kernel size 3×3 across all layers to implement the reconstruction and segmentation networks [11]. This cascaded model with two sequential U-Nets was used to analyze the influence of end-to-end training and stabilization. For enhanced performance, we then adopted the Cascade Network for reconstruction and Multiscale Attention Network (MANet) for segmentation [12,13]. Cascade Network follows an unrolled architecture with interleaved data consistency and regularization blocks, and progressively suppresses aliasing artifacts in reconstructions. The regularization blocks include residual connections to carry the input signal to the output and force the network to learn required the residual information [12]. In our setup, we used 6 cascades with 5 layers of 64 filters each. MANet improves upon UNet with multiscale information extraction achieved by point-wise and multiscale attention blocks [11,13]. Using dilated convolutions in the decoder distills multiscale information processed with the squeeze-and-excitation attention mechanism to capture dependencies among feature maps [14]. We used ResNet34 in the encoder with a depth of 4 and filter sizes (32, 64, 128, 256) for MANet [15]. Finally, a multi-decoder architecture where the decoder head is split to separately perform reconstruction and segmentation is added to the experiments as an additional baseline [16]. For Multi-Decoder UNet, a common encoder of depth 4 with filter configuration (32, 64, 128, 256) is created for reconstruction and segmentation tasks, taking the undersampled images as input. The encoder is used to obtain disentangled feature representations which are then fed to the first decoder head to reconstruct the underlying image and to the second decoder to create the segmentation map.

2.4 Implementation Details

All networks were trained using the Adam optimizer with parameters $\beta_1 = 0.99$ and $\beta_2 = 0.999$, a learning rate of 10^{-4} and batch size of 16. Models were implemented using PyTorch library and executed on NVIDIA RTX 3090 GPUs. Experiments were conducted on fully-sampled MRI data from the public OCMR dataset containing CINE scans from 74 subjects [10]. Subjects had varying number of slices and frames, yielding a total of 183 slices, which were coil combined to simulate single-coil data acquisition. Data were split into independent training (155 slices) and test (28 slices) sets, with no subject overlap between the two



(a) x4 Acceleration



(b) x8 Acceleration

Fig. 2. Representative reconstructions from competing techniques at (a) 4-fold, (b) 8-fold acceleration. The second row displays the localized reconstruction around the heart. The area on which the local performance calculations are taken is indicated with the red rectangle. First column displays Fourier reconstructions of fully-sampled data. In remaining columns, reconstruction with no help from the segmentation maps, joint reconstruction without stabilization and joint reconstruction with stabilization is shown respectively. (Color figure online)

sets. MRI data were retrospectively undersampled to achieve acceleration rates of 4 and 8. A Gaussian sampling density with an autocalibration region containing 8 lines was used. Magnitude images for the resulting reconstructions are used to generate segmentation map predictions. Ground-truth segmentation maps for

MR images were created in-house via manual labeling with experts under the guidance of a senior radiologist. In all experiments, a composite loss function with an $\ell_1 - \ell_2$ term for the reconstruction task, and a Dice term for the segmentation task [17] is employed with equal weights. Global reconstruction quality was assessed by measuring peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and mean-squared error (MSE) between the reconstructed and ground-truth images. Local reconstruction quality was also measured via the same metrics, albeit the measurement regions containing the target tissues were selected based on the segmentation maps. Local measurements are denoted with the 'F' (for 'Focused') prefix in Table 2.

Table 1. Performance comparisons of various baselines on cardiac MRI data across ×4 and ×8 acceleration rates. (S) suffix signifies the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training aids Multi-Decoder Unet and UNet → UNet in focused metrics indicating improved reconstruction quality for cardiac cavity.

Method	MSE	PSNR	SSIM	F-PSNR	F-MSE	F-SSIM
×4 acceleration						
UNet	1.7693	26.9204	0.6247	25.5002	1.9009	0.5926
Multi-Decoder UNet	1.9693	26.3084	0.6651	25.3396	2.0209	0.6318
Multi-Decoder UNet (S)	1.7530	**27.1221**	**0.7082**	25.8338	**1.7109**	**0.6735**
UNet → UNet	1.9509	25.4176	0.5227	24.6887	2.5709	0.5488
UNet → UNet (S)	**1.7474**	27.0973	0.6729	**26.4023**	1.8174	0.6475
×8 acceleration						
UNet	**2.2378**	**25.4062**	**0.6583**	24.3128	2.7243	0.6033
Multi-Decoder UNet	2.4312	25.1512	0.6532	24.0712	2.9125	0.6219
Multi-Decoder UNet (S)	2.2441	25.2193	0.6646	24.7312	2.6217	**0.6422**
UNet → UNet	3.1743	23.2683	0.5491	23.0352	3.4719	0.5428
UNet → UNet (S)	2.2782	25.2931	0.6529	**24.9023**	**2.4174**	0.6375

3 Experimental Results

Figure 2 illustrates reconstructions for a representative test subject at ×4 and 8x acceleration along with the fully-sampled ground truth. PSNR, SSIM and MSE of all tested methods are presented in Tables 1 and 2 along with the localized measurements around the heart. Table 1 underlines the results of the experiments done on the UNet architecture which is taken as a baseline for both the reconstruction and segmentation tasks. Additionally Multi-Decoder UNet is introduced to the experiments to compare against the segmentation-aware reconstruction framework. We see that in a stabilized setting, aided by segmentation, reconstruction of the diagnostically-relevant areas improve compared to single UNet. Multi-Decoder UNet seems to surpass the reconstruction-segmentation

network performance at ×4 acceleration task in MSE and SSIM measures around the regions of interest. However, when tasked with harsher undersampling rates, Table 1 highlights the effect of an end-to-end architecture as it yields better quality reconstructions in terms of PSNR and MSE metrics around the heart.

Table 2 shows, while Cascade Network performs better at solo reconstruction task in terms of global performance metrics, segmentation-aware reconstruction with stabilization outperforms competing methods in terms of localized metrics. Note that the jointly trained compound model for segmentation-aware reconstruction greatly suffer in the absence of stabilization. When supplied with high quality reconstructions, segmentation network is able learn the mapping to create accurate segmentation maps. However, in the case of an insufficiently trained reconstruction network in the model, the segmentation network is exposed to heavy undersampling artifacts which in turn misguide the resulting segmentation output. Therefore, inaccurate segmentation information propagating into the reconstruction network damages overall quality of the reconstructions, leading to poor performance on both networks. As loss functions typically utilized for reconstruction is expressed over the entire image the learning signal coming directly from the reconstruction output largely governs the performance for the overall image in the global setting. However, such loss functions are mainly dominated by bright bone structures, diagnostically irrelevant body parts and overall low-frequency information. Back-propagating the errors from the segmentation network into the reconstruction network, indirectly emphasizes the areas of interest during training since the segmentation loss is concentrated around such diagnostically-relevant areas. Therefore, end-to-end training of a reconstruction-segmentation network creates a pseudo-attention effect to focus the efforts of the reconstruction network which in turn improves the localized performance around regions of interest.

Table 2. Performance comparisons (MSE, PSNR and SSIM) on cardiac MRI data with ×4 and ×8 acceleration rates across experiment setups. MSE is scaled with 10^3. (S) suffix indicates the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training improves the focused metrics indicating improved reconstruction quality for cardiac cavity.

Method	MSE	PSNR	SSIM	F-MSE	F-PSNR	F-SSIM
×4 acceleration						
CascadeNet	**0.6698**	**32.1988**	**0.9113**	1.0342	29.1253	0.8331
CascadeNet → MANet	0.7640	31.5884	0.8563	1.1980	28.3712	0.8131
CascadeNet → MANet (S)	0.7114	31.4328	0.8991	**0.9731**	**30.0000**	**0.8828**
×8 acceleration						
CascadeNet	**1.2825**	**28.4301**	**0.8259**	1.7826	25.5156	0.7728
CascadeNet → MANet	1.7603	25.2837	0.7265	2.4456	23.3219	0.6673
CascadeNet → MANet (S)	1.3226	27.4895	0.8217	**1.5453**	**26.3421**	**0.7931**

4 Conclusion

Here we proposed a segmentation-aware reconstruction framework for cardiac MRI acquisitions. Experiments were conducted to systematically demonstrate the proposed method against solo reconstruction methods. As expected, solo reconstruction with global loss terms yields higher performance in global quality metrics. That said, our results clearly demonstrate that the segmentation-aware reconstruction outperform solo reconstruction in local quality metrics focused on the target ROI. Furthermore, we observe that stabilization of the acceleration rate during the course of joint network training is highly effective in mitigating convergence onto local minima with unacceptably poor reconstruction or segmentation performance.

Acknowledgements. This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C353). However, the entire responsibility of the publication/paper belongs to the owner of the paper. The financial support received from TUBITAK does not mean that the content of the publication is approved in a scientific sense by TUBITAK.

References

1. Dar, S.U.H., Yurt, M., Shahdloo, M., Ildız, M.E., Tınaz, B., Çukur, T.: Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. IEEE J. Sel. Top. Sig. Process. **14**(6), 1072–1087 (2020)
2. Oksuz, I., et al.: Cardiac MR motion artefact correction from K-space using deep learning-based reconstruction. In: Knoll, F., Maier, A., Rueckert, D. (eds.) MLMIR 2018. LNCS, vol. 11074, pp. 21–29. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00129-2_3
3. Wang, S., et al.: Accelerating magnetic resonance imaging via deep learning. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 514–517 (2016)
4. Fuin, N., et al.: A multi-scale variational neural network for accelerating motion-compensated whole-heart 3D coronary MR angiography. Magn. Reson. Imaging **70**, 155–167 (2020)
5. Yan, W., Ma, Y., Liu, J., Jiang, D., Xing, L.: Self-attention convolutional neural network for improved MR image reconstruction. Inf. Sci. **490**, 317–328 (2019)
6. Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D.: MRI reconstruction via cascaded channel-wise attention network. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1622–1626 (2019)
7. Yuan, Z., et al.: SARA-GAN: self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction. Front. Neuroinform. **14**, 611666 (2020)
8. Pramanik, A., Jacob, M.: Reconstruction and segmentation of parallel MR data using image domain Deep-SLR (2021)
9. Huang, Q., Yang, D., Yi, J., Axel, L., Metaxas, D.: FR-Net: joint reconstruction and segmentation in compressed sensing cardiac MRI. In: Coudière, Y., Ozenne, V., Vigmond, E., Zemzemi, N. (eds.) FIMH 2019. LNCS, vol. 11504, pp. 352–360. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21949-9_38

10. Chen, C., et al.: OCMR (v1.0)-open-access multi-coil k-space dataset for cardiovascular magnetic resonance imaging (2020)
11. 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
12. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **37**(2), 491–503 (2018)
13. Fan, T., Wang, G., Li, Y., Wang, H.: MA-NET: a multi-scale attention network for liver and tumor segmentation. IEEE Access **8**, 179656–179665 (2020)
14. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks (2017)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
16. Amyar, A., Modzelewski, R., Li, H., Ruan, S.: Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. Comput. Biol. Med. **126**, 104037 (2020)
17. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28