Super-resolution of portable low-field MRI in real scenarios: integration with denoising and domain adaptation

Sonia Laguna[1] slaguna@ethz.ch
Riana Schleicher[2] rschleicher@mgh.harvard.edu
Benjamin Billot[3] benjamin.billot.18@ucl.ac.uk
Pamela Schaefer[2] pschaefer@partners.org
Brenna McKaig[2] bmckaig@mgh.harvard.edu
Joshua N. Goldstein[2] jgoldstein@mgh.harvard.edu
Kevin N. Sheth[4] kevin.sheth@yale.edu
Matthew S. Rosen[2] msrosen@mgh.harvard.edu
W. Taylor Kimberly[2] wtkimberly@mgh.harvard.edu
Juan Eugenio Iglesias[2,3,5] jiglesiasgonzalez@mgh.harvard.edu
[1] *Swiss Federal Institute of Technology (ETH) Zürich, Switzerland*
[2] *Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[3] *University College London, London, UK*
[4] *Yale New Haven Hospital, New Haven, CT, USA*
[5] *Massachusetts Institute of Technology, Cambridge, MA, USA*

Abstract

Portable low-field MRI has the potential to revolutionize neuroimaging, by enabling point-of-care imaging and affordable scanning in underserved areas. The lower resolution and signal-to-noise ratio of these scans preclude image analysis with existing tools. Super-resolution (SR) methods can overcome this limitation, but: *(i)* training with downsampled high-field scans fails to generalize; and *(ii)* training with paired low/high-field data is hard due to the lack of perfectly aligned images. Here, we present an architecture that combines denoising, SR and domain adaptation modules to tackle this problem. The denoising and SR components are pretrained in a supervised fashion with large amounts of existing high-resolution data, whereas unsupervised learning is used for domain adaptation and end-to-end finetuning. We present preliminary results on a dataset of 11 low-field scans. The results show that our method enables segmentation with existing tools, which yield ROI volumes that correlate strongly with those derived from high-field scans ($\rho > 0.8$).

1. Introduction

Conventional MRI enables evaluation of the brain in health and disease but operates at high field (HF, typically 1.5-3T) and requires dedicated scanner suites. This introduces some limitations, most notably: *(i)* it restricts application to patients that can be transported safely; and *(ii)* the associated costs are large and limit deployment in developing countries and rural areas. The recent advent of portable low-field (LF) scanners like the 64mT Hyperfine Swoop holds great promise to overcome these obstacles (Sheth et al., 2021). However, the limited resolution and signal-to-noise ratio (SNR) that can be achieved in practical scanning times precludes image analysis with existing tools, which is desirable to obtain quantitative morphometric measurements for research and clinical purposes.

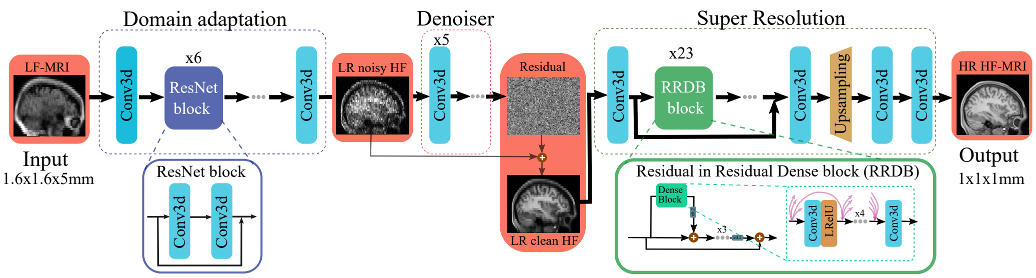


Figure 1: Proposed 3D meta-architecture with building blocks.

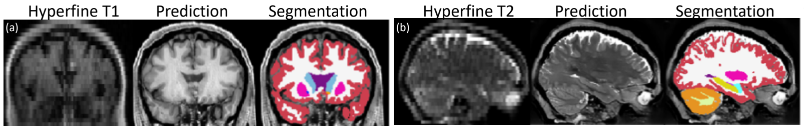
Modern image super-resolution (SR) techniques based on deep CNNs have the potential to bridge the resolution / SNR gap and enable morphometry of LF scans (Wang et al., 2020), but direct application of these methods has two problems. First, they are not designed to cope with strong noise. And second, they often fail to generalize: most existing methods are trained low-resolution (LR) images obtained by downsampling high-resolution (HR) data, and then report high accuracy on LR scans obtained with the same procedure. However, the domain gap between synthetic LR scans and real LF data (due to differences in contrast and noise properties) greatly compromises generalization. An alternative would be supervised training with aligned LF-HF scans, but this is hard due to nonlinear geometric distortions.

Here, we present a method that overcomes these limitations and effectively estimates a 1 mm isotropic scan from a LR, LF scan (1.6×1.6×5mm axial, i.e., the standard resolution of Hyperfine). This is achieved by combining domain adaptation (DA), denoising and SR modules with artificially corrupted data (SR, denoiser) and unpaired LF-HF scans (DA).

2. Methods

Architecture. Our proposed architecture is shown in Fig. 1. It consists of three modules: DA, denoiser, and SR. For the DA module, we have explored two approaches based on unpaired translation: the widespread CycleGAN (Zhu et al., 2017) and FastCUT (Park et al., 2020), which builds on CycleGAN and combines a 6-block ResNet generator with contrastive losses. The SR module is a 3D adaptation of ESRGAN (Wang et al.), with a generator of 23 residual dense blocks at LR, and no VGG loss. The denoiser uses 5 convolutional layers to predict the residual between the noisy and clean LR data.

Learning. The three modules were first trained independently. The SR and denoiser modules can be effectively trained in a supervised fashion with large amounts of existing HF data: one can simply take an HF scan and artificially corrupt it (downsample, add noise) to obtain training pairs. The SR block was trained with a combination of an L1 loss and an adversarial loss (relative weight 0.01). The denoiser minimized an L1 loss. The DA module is trained in an unsupervised manner, using unpaired data consisting of real LF scans and synthetically corrupted (noisy, downsampled) HF images. The relative weight of the adversarial loss with respect to the cycle consistency loss was set to 0.1. Finally, the whole model was finetuned end-to-end with just the SR adversarial loss. Training used the Adam optimizer and aggressive augmentation, including: linear and nonlinear deformation; brightness and contrast; bias field; and segmentation-guided intensity variations.



(c) DA / Mod.	FastCUT		CycleGAN	
	Denois	No den	Denois	No den
T1	**0.88**	0.81	0.78	0.71
T2	0.80	**0.86**	0.80	0.74

Figure 2: (a,b) Sample outputs with FastCUT as DA. (c) Median correlation for amygdala, pallidum, cortex, hippoc., putamen, caudate, thalamus, white mat., ventricles.

3. Experiments and results

Experimental setup. We used two datasets for training, one HF and one LF. The HF dataset consists of HR T1/T2 scans of 898 subjects from the Human Connectome Project (HCP). The LF dataset consists of T1/T2 Hyperfine scans of 11 subjects, which are part of an ongoing study at Massachusetts General Hospital and had neurological symptoms but no large pathological lesions. For evaluation, we used the same LF dataset – which is fine, as unsupervised finetuning of our model with a new scan before inference is always possible. As ground truth, we used HF clinical scans that were also available for the same 11 subjects: we segmented the synthetic HR output and the HF scans with a robust approach (Billot et al., 2020), computed ROI volumes, and used the correlation between the estimated (from enhanced LF scans) and ground truth volumes (from HF scans) as a proxy for performance.

Results. Despite the low resolution and SNR of the input, our method recovers images with enough detail to produce usable segmentations (Fig. 2a,b), which yield ROI volumes with strong correlations with the ground truth, both for T1 and T2 (Fig. 2c).

4. Discussion and Conclusion

Our method maximizes the amount of training data that can be used by each module and yields images with enough quality for volumetry. Our results also show that FastCut outperforms CycleGAN for DA, and that explicit inclusion of a denoiser improves the results in most scenarios (future analysis with larger datasets is necessary). As LF MRI becomes increasingly widespread, developing methods than can cope with the domain gap is crucial.

References

B Billot et al. A learning strategy for contrast-agnostic MRI segmentation. *MIDL*, 2020.

T Park et al. Contrastive learning for unpaired image-to-image translation. *ECCV*, 2020.

KN Sheth et al. Assessment of brain injury using portable, low-field magnetic resonance imaging at the bedside of critically ill patients. *JAMA Neurology*, 2021.

X Wang et al. ESRGAN: Enhanced super-resolution generative adversarial networks.

Z Wang et al. Deep learning for image super-resolution: A survey. *IEEE-PAMI*, 2020.

J Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ICCV*, 2017.