Guided Multicontrast Reconstruction based on the Decomposition of Content and Style

Chinmay Rao[1], Laurens Beljaards[1], Matthias van Osch[1], Mariya Doneva[2], Jakob Meineke[2], Christophe Schülke[2], Nicola Pezzotti[3,4], Elwin de Weerdt[5], Marius Staring[1]

[1] *Department of Radiology, Leiden University Medical Center, Leiden, Netherlands*

[2] *Philips Research Hamburg, Hamburg, Germany*

[3] *Cardiologs, Paris, France*

[4] *Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands*

[5] *Philips, Best, Netherlands*

Synopsis

Motivation:

Scans within an MR exam share redundant information due to the same underlying structures. One contrast can hence be used to guide the reconstruction of another, thereby requiring less measurements.

Goals:

Multimodal guided reconstruction to reduce scanning times.

Approach:

Our method exploits AI-based content/style decomposition in an iterative reconstruction algorithm. We explored this concept via numerical simulation and subsequently validated it on *in vivo* data.

Results:

Compared to a conventional compressed sensing baseline, our method showed consistent improvement in simulations and produced sharper reconstructions from undersampled *in vivo* data. By enforcing data consistency, it was also more reliable than blind image translation.

Impact:

In the clinic, this can potentially enable a reduced MR exam time for a given image quality or improve image quality given a scan time budget. The former can reduce strain on the patient, whereas the latter can improve diagnosis.

Introduction

Different contrast-weighted MR images are reflections of the same underlying tissues and share redundant information. Therefore, an existing high-quality reference scan can be used to guide the reconstruction of an undersampled second scan. In this work, we learn an image-to-image model that explicitly represents the shared information and apply it as an artifact removal operator within the proximal gradient algorithm. We studied this concept via numerical simulations and validated it on undersampled *in vivo* data obtaining higher fidelity reconstructions compared to wavelet-regularized compressed sensing and pure image translation.

Methods

Our method is based on MUNIT, an unsupervised technique for learning image-to-image generative models through content/style disentanglement. We describe sets of T1W and T2W images as data distributions and assume that the contrast-independent structural information can be separated as "content" from the contrast-specific variations like acquisition settings. Figure 1 illustrates these modeling assumptions. Plug-and-play (PnP) reconstruction methods
use an off-the-shelf denoiser to model the image prior in an iterative algorithm, applied to the image estimate alternatingly with a data consistency step. Given a spatially aligned T1W reference scan x_{ref}
during a T2W image reconstruction, we observe that a trained MUNIT can function as an image prior in a proximal gradient algorithm, applied as a powerful artifact removal operator:

$$c \leftarrow E_c^1(x_{ref}),$$
$$x^k \leftarrow G^2(c, E_s^2(p^{k-1})),$$

where E_c^1, E_s^2, and G^2 are the T1W content encoder, T2W style encoder, and T2W decoder, respectively, c represents content maps, p^{k-1} is the data-consistent image from iteration $k-1$, and x^k is the image at iteration k with aliasing artifacts removed. This forms the basis of our PnP-MUNITalgorithm, illustrated in Figure 2. If the T1W content perfectly matches the ground truth T2W content, the algorithm will produce high-quality reconstructions. However, in practice, there is always some discrepancy between them, e.g. due to misalignment between images, modeling errors associated with MUNIT, etc. To address this, we initialize the content estimate with the T1W content and iteratively refine it to be increasingly consistent with the measured data:

$$c^k \leftarrow c^{k-1} - \lambda \nabla_c ||p^{k-1} - G^2(c^{k-1}, E_s^2(p^{k-1}))||_2^2$$

where λ is a tunable parameter.

We analyzed this concept via simulations based on the BrainWeb anatomical models. The 20 brain volumes were split into MUNIT training (18) and validation (1) sets and a reconstruction test set (1). Synthetic T1W/T2W spin-echo scans were simulated using TE/TR values randomly chosen from a fixed set per contrast and were used to train the model. During reconstruction, 2D single-coil T2W k-space with different accelerations and noise levels was simulated from the test brain volume. We additionally validated our method on IRB-approved patient data. The MUNIT training data comprised 277 T1W TSE and 279 T2W TSE 2D-acquired coronal brain scans. As reconstruction test set, we used 10 slices from a separate set of 10 patients, each with 13-coil T2W k-space acquired using 3T Philips Ingenia scanner (with parameters FA=90°, TR=2000-2094ms, TE=90ms, echo train length=17, matrix size=304x263, in-plane resolution 0.39×0.46mm, slice thickness 2mm, 1D random undersampling with R=1.8) and the corresponding spatially registered T1W reference scan. We tested on total acceleration of R=4 and R=8 reached by retrospectively removing portions of the acquired samples. We used two baselines: (a) L1-Wavelet reconstruction implemented with the ISTA algorithm which differs from our method in just the denoising step, and (b) pure T1W-to-T2W image translation with MUNIT which, assuming no knowledge of the acquisition, is defined as the average of synthetic images obtained from randomly sampled style codes.

Results and Discussion

Figure 3 shows the numerical simulation results for a single slice. Note that the learned content need not correspond to brain tissues since the model learns useful arbitrary content representations to solve the learning problem. Interestingly, some correspondence with brain tissues did emerge here, e.g. content maps #3 and #9 seem to encode background and fat tissue, respectively. Given ideal T2W content maps, PnP-MUNITproduced extremely low-error reconstructions, whereas the reconstruction quality dropped with T1W reference content maps due to the T1W/T2W content discrepancy. Introducing the content update step significantly reduced error approaching the ideal content case. PnP-MUNIT consistently improved over L1-Wavelet across the acceleration range and the different noise levels. As shown in Figure 4, PnP-MUNIT produced sharper *in vivo* reconstructions, visible more prominently at R=8. Aggregate metrics for both experiments, shown in Figure 5, agree with the qualitative observations.

Conclusion

The proposed concept was analyzed in numerical simulations and its practicality was demonstrated on undersampled *in vivo* data where it produced substantially improved reconstructions. Future work will be directed towards more extensive analysis on a larger
in vivo test set and benchmarking against state-of-the-art reconstruction algorithms.
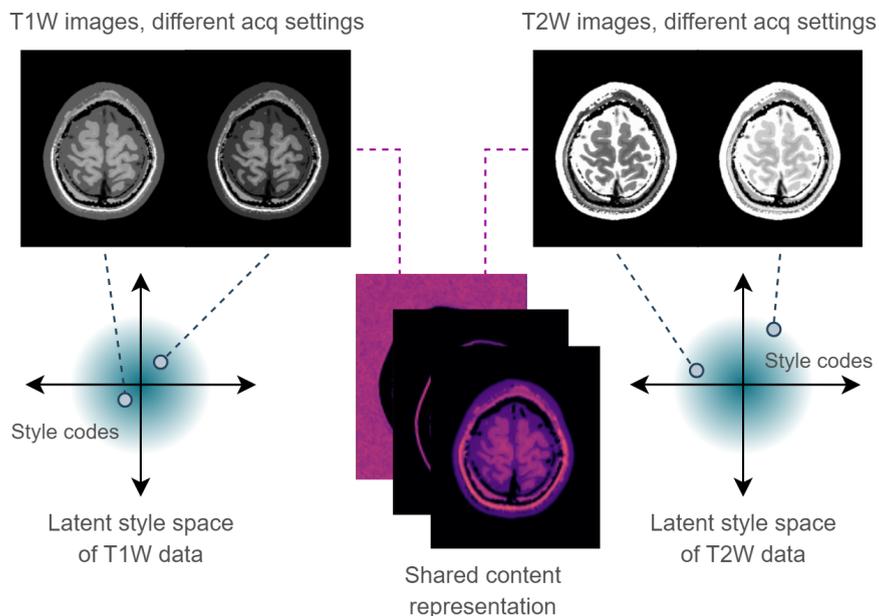
Figure 1: We assume that a pair of MR contrasts of a given anatomy can be decomposed into a shared "content" representation and separate "style" encodings, and seek to model this relationship with MUNIT via unpaired training. Given two distributions of image data, content represents a learned feature space shared by

the two distributions, whereas style is a set of latent variables that encode properties of one image distribution which cannot be explained by the other.
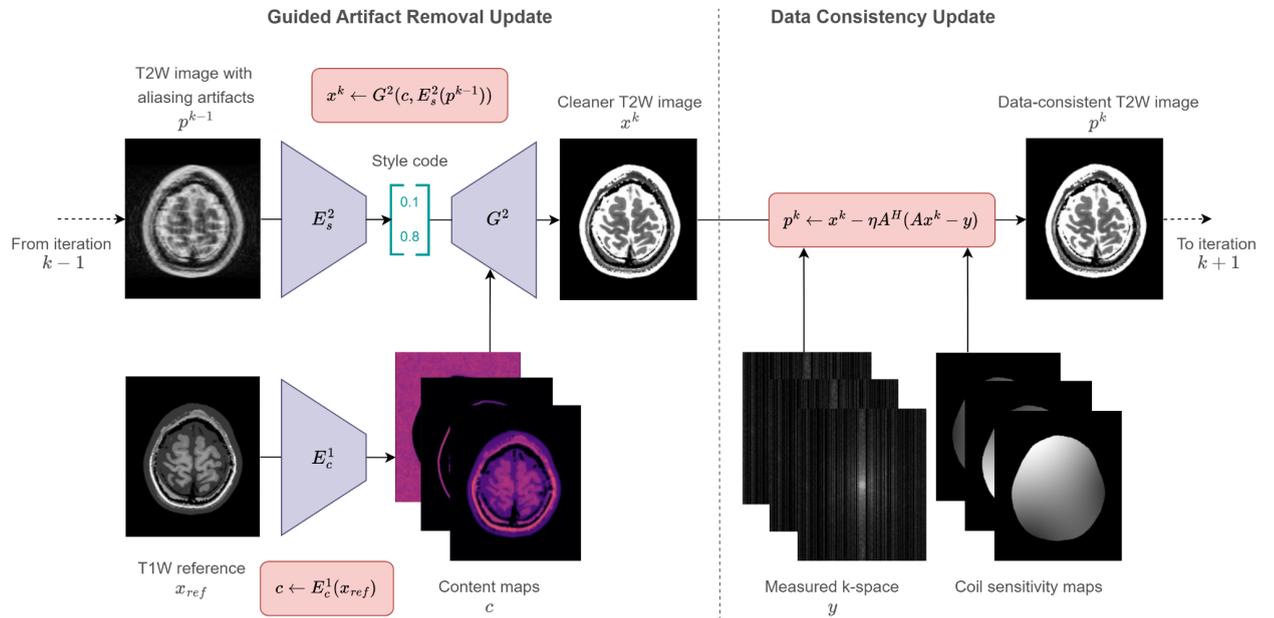
Figure 2: Iteration k of the proposed PnP-MUNIT algorithm. The content update step is not shown here. In the data consistency update, A is the MRI forward operator composed of coil sensitivity encoding, Fourier encoding, and sampling, and η is the size of the gradient descent step.

Learned Content Maps vs. Brain Tissue Maps
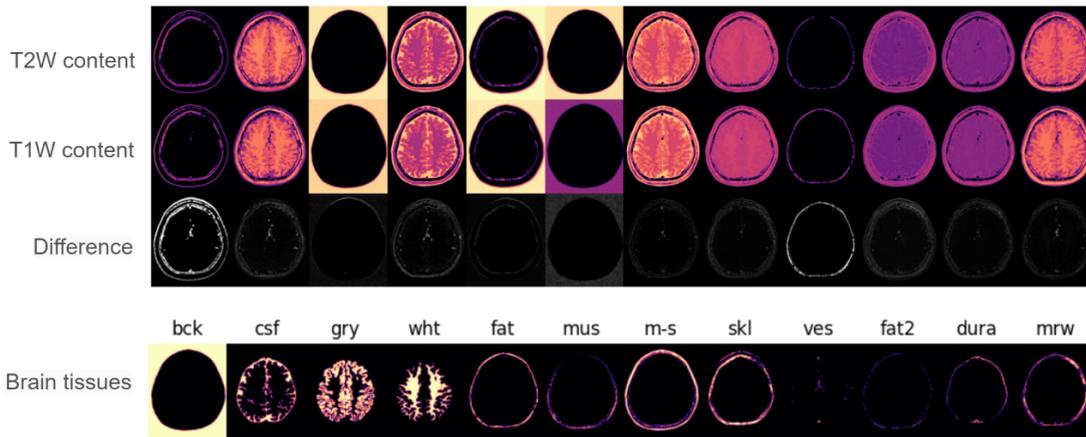
Reconstructions at R=1.5 and Noise Level=0.01
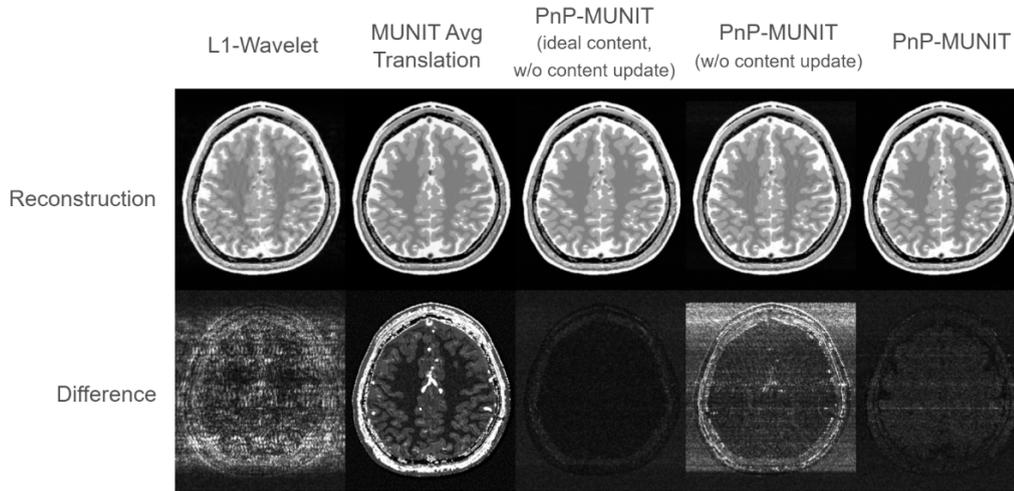
Performance Curves
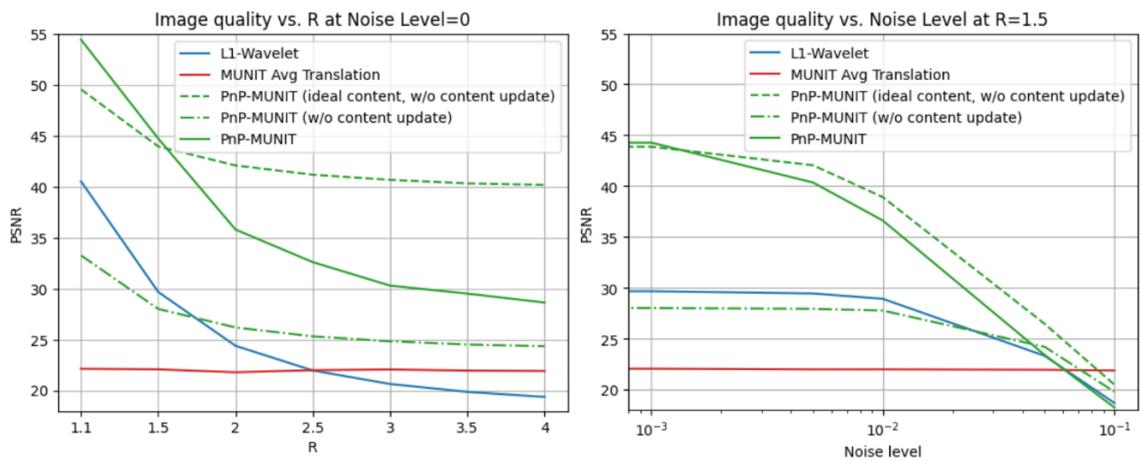
Figure 3: Single-slice result of the numerical simulation. Top: A selection of 12 learned content maps of the T1W/T2W image pair and the error between them. The 12 original BrainWeb tissue maps are shown for

comparison, though there is no direct correspondence with the content maps. Middle: Reconstructed images. Here, "ideal content" refers to content maps of the T2W ground truth which are, in theory, ideal for PnP-MUNIT. Difference image dynamic range is 10% of ground truth. Bottom: Performance comparison over different acceleration factors and noise levels.
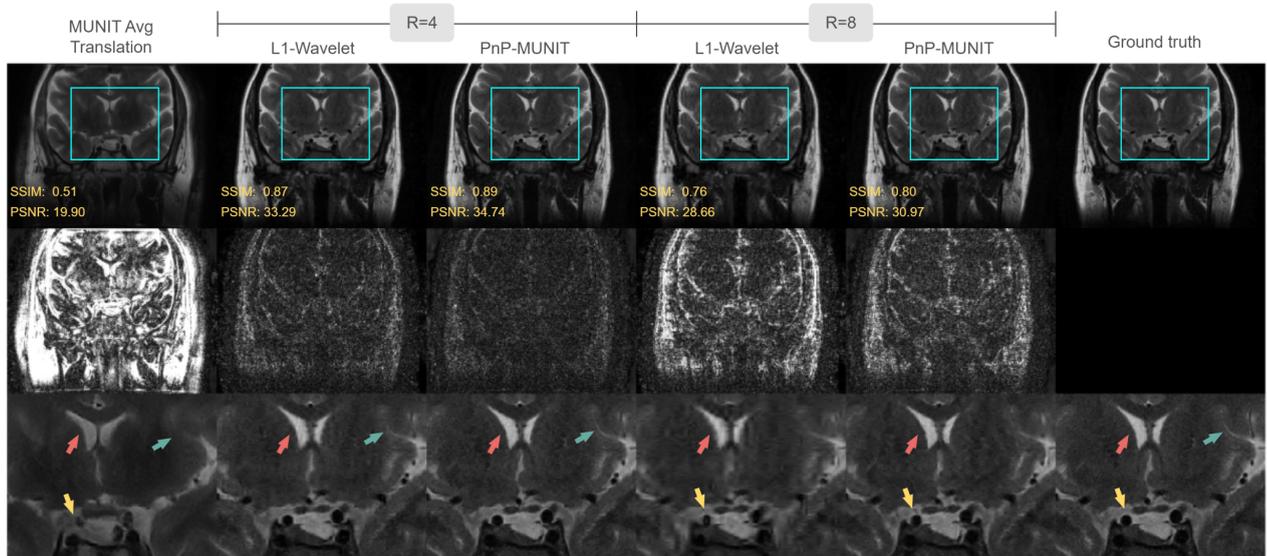
Figure 4: Reconstruction of an undersampled in vivo slice. Difference image dynamic range is 10% of ground truth. A region of interest from the reconstructions is selected and enlarged to show the detail.

Simulation Test Set (R=1.5, Noise Level=0.01)

Method	PSNR
L1-Wavelet	29.08 ± 0.376
MUNIT Avg Translation	19.40 ± 0.360
PnP-MUNIT (ideal content, w/o content update)	*38.44 ± 0.058*
PnP-MUNIT (w/o content update)	28.61 ± 0.158
PnP-MUNIT	**36.65 ± 0.109**

Undersampled *In Vivo* Test Set

Method	R = 4		R = 8	
	SSIM	PSNR	SSIM	PSNR
L1-Wavelet	**0.90 ± 0.021**	35.00 ± 1.226	0.81 ± 0.030	30.53 ± 1.062
MUNIT Avg Translation	0.52 ± 0.019	21.08 ± 0.341	0.52 ± 0.019	21.08 ± 0.341
PnP-MUNIT	**0.90 ± 0.012**	**35.93 ± 1.054**	**0.84 ± 0.018**	**32.62 ± 0.938**

Figure 5: Aggregate metrics over the reconstruction test sets of the numerical simulation (top, 359 slices from 1 test volume) and the undersampled in vivo data (bottom, 10 slices from a test set of 10 patients). The SSIM and PSNR columns show the mean value and 95% confidence intervals.

Acknowledgements

References

1. Ehrhardt MJ and Betcke MM, SIAM Journal on Imaging Sciences, 2016 Aug, 9(3):1084-106.
2. Dar SU et al., IEEE Journal of Selected Topics in Signal Processing. 2020 Jun, 11;14(6):1072-87.
3. Huang X et al., ECCV, 2018, 2018:179-96.
4. Ahmad et al., IEEE signal processing magazine. 2020 Jan, 17;37(1):105-16.
5. Collins DL et al., IEEE TMI, 1998 Jun, 17(3):463-8.