

MRM-24-25421: Responses to Editors and Reviewers

Editor's Comments to the Author:

Thanks for submitting to Magn Reson Med. Firstly, apologies for the delay in getting you an outcome – we struggled to identify willing reviewers. Finally, though, we found two expert referees. As you can see, although there is potential interest from them in the study there are substantial issues raised by both. As such I regret that we must return the current version of the manuscript to you. We would allow a future resubmission if you wished to undertake the considerable work outlined by the reviewers, but we would need to see a substantial improvement in reviewer enthusiasm and accompanying scores. Alternatively, you would be at liberty at this juncture to submit elsewhere.

Deputy Editor: Wang, Shanshan Comments to Author:

Dear Authors,

Thank you for your submission. The paper has been reviewed by two experts, with one recommending rejection and the other suggesting major revisions. The key concerns raised are as follows:

- * The baseline navigated method does not appear to work as expected.*
- * There is no comparison with other state-of-the-art (SOTA) methods.*
- * The results validation is primarily qualitative and subjective.*

Given these issues, we are unable to consider the current version of your manuscript further unless you are willing to make substantial revisions to address these concerns. Please be aware that, should you choose to resubmit, the revised manuscript will typically be sent to the same reviewers.

Thank you again for your submission, and we appreciate your understanding.

We thank both editors for the constructive comments and resubmission invitation. We have thoroughly addressed the raised critiques in this revision. Please note that because this is a resubmission and the manuscript is treated as new, we don't provide the marked version.

Reviewer 1

This manuscript presents a deep learning-based reconstruction method for high-resolution multi-shot DWI. Specifically, the authors employ ADMM unrolling with a network-learned regularization to enable multi-shot DWI reconstruction while mitigating artifacts caused by shot-to-shot phase variations. They report achieving 0.7 mm isotropic resolution DWI on a 7T scanner with good image quality. While the proposed approach is promising, I have several major concerns regarding the presentation and validation of the method.

- (1) *The current review of reconstruction methods in the Introduction feels somewhat mixed. Methods for motion-phase correction (e.g., MUSSELS) and k-q joint reconstruction (e.g., DAE) are discussed together, which might obscure their distinct purposes. I suggest that the authors introduce these methods separately to help readers better understand the core objectives of each category.*

Thank you for the suggestion. We have separated these methods in introduction.

- (2) *Regarding the forward model A used in the proposed ADMM unrolling method (Eq. 2 and Algorithm 1), what was specifically used as the motion phase map Φ ? Was it derived from MUSE? The authors note: "However, it (MUSE) requires small undersampling factors per shot and fully sampled DWI acquisition assembling all shots. Alternatively, undersampled DWI acquisition can be enabled via the acquisition of navigators for shot phase estimation." This gives the impression that navigator-based phase correction should outperform MUSE. However, the results in Fig. 4 suggest otherwise, which seems contradictory.*

Thank you for the constructive comments. The motion phase map Φ can be estimated from either the imaging echo or the navigator echo, and is derived as following: (1) perform SMS-SENSE reconstruction on every shot, (2) extract and smooth phase from the reconstructed images. Yes, these are part of MUSE reconstruction steps. We do expect navigator-based phase correction outperforms MUSE, i.e., self-gated, as the navigator is less undersampled than the imaging echo.

- (3) *The navigator-based phase correction shown in Fig. 4 appears overly artifacted. The authors attribute this to increased scan time leading to greater sensitivity to inter-shot motion: "The main reason for these artifacts is that the acquisition of navigators increases the total scan time, resulting in higher sensitivity to accidental inter-shot motion." However, I find this explanation unconvincing. Since this is a retrospective experiment, the effective scan time for both navigated and self-gated acquisitions should be identical, meaning that the subject would experience the same motion in both cases. The inclusion of a navigator introduces a second echo, but since motion primarily affects the phase during diffusion encoding, the phase should remain static after diffusion encoding, allowing accurate phase correction using the navigator. It seems more likely that the artifacts arise from errors in navigator implementation. To clarify this, I recommend that the authors present the navigator phase data*

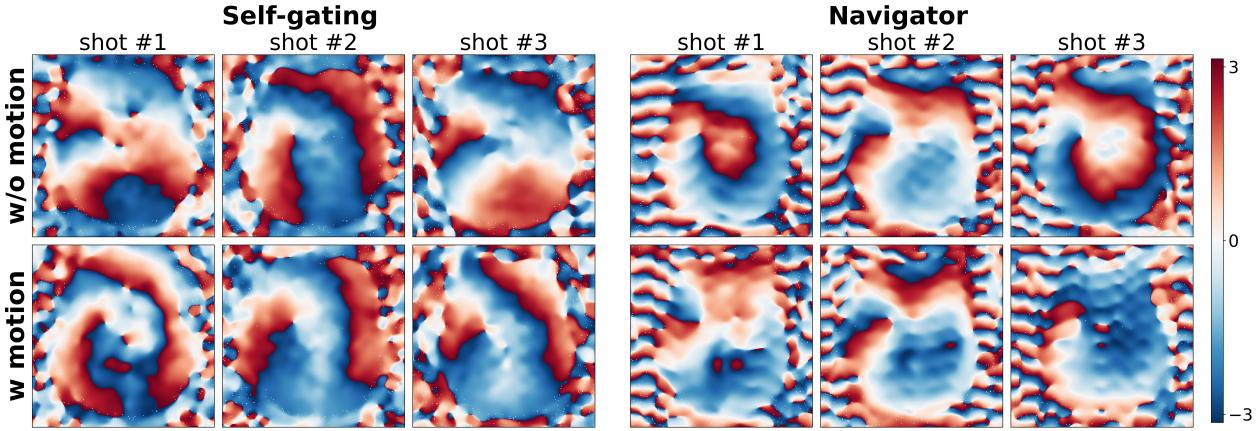


Figure 1: (Left Panel) Shot phases from self-gated reconstruction, i.e., reconstruct and extract phases from each shot of imaging echoes. (Right Panel) Shot phases from navigated reconstruction. (Top Row) A diffusion direction that does not show inter-shot motion. (Bottom Row) A diffusion direction that shows motion in the navigator from the 3rd shot, which appears as rapid phase jumps in the frontal brain region.

and compare it with the MUSE-estimated phase.

Thank you for the valuable insights. Here, Figure 1 shows that the navigator phase from the 3rd shot was indeed corrupted. After further review, we discovered that the volunteer presented here exhibited a very specific motion pattern, which was previously unknown to us. This motion is caused by an underlying disease that leads to abrupt and involuntary movements, which occur over extremely short durations. Consequently, the effect is only visible in isolated diffusion encoding where the navigator scans happen to be corrupted, while the rest show intact navigators and clean DWIs, as shown in Figure 2.

Based on these findings, we have decided to exclude this volunteer's data from our manuscript. We truly appreciate you bringing this to our attention. It led to an important investigation that we otherwise might have missed.

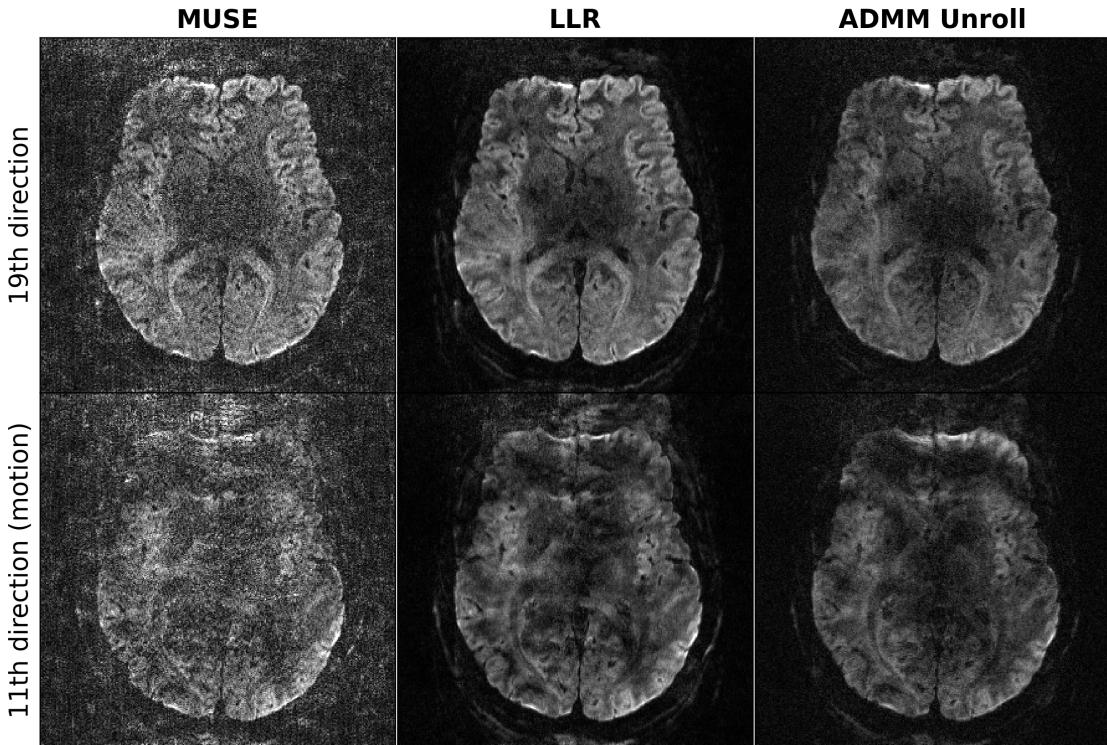
- (4) *Due to the failure of the navigator-based method, a reliable reference for comparison is lacking. Consequently, many of the presented results rely on subjective, qualitative assessments. This may weaken the overall impact of the manuscript. To strengthen the validation, I recommend acquiring a high-quality reference dataset and/or conducting simulations with known ground truth, which would enable more objective and quantitative evaluation.*

Thank you for the constructive suggestion. We provide a full-sampled reference data (listed as Protocol #1 in Table 1) and present the results in Figure 3 in the manuscript.

- (5) *The authors present only raw DWIs, which makes it difficult to evaluate the overall quality of the acquired diffusion datasets, particularly in the absence of a reference. I suggest performing additional diffusion analyses, such as DTI, to provide more convincing evidence of the proposed method's advantages over existing approaches (e.g., LLR).*

Thank you for the constructive suggestion. We perform DTI analysis with the 0.7 mm data

(A) Navigated reconstruction



(B) Selfgated reconstruction

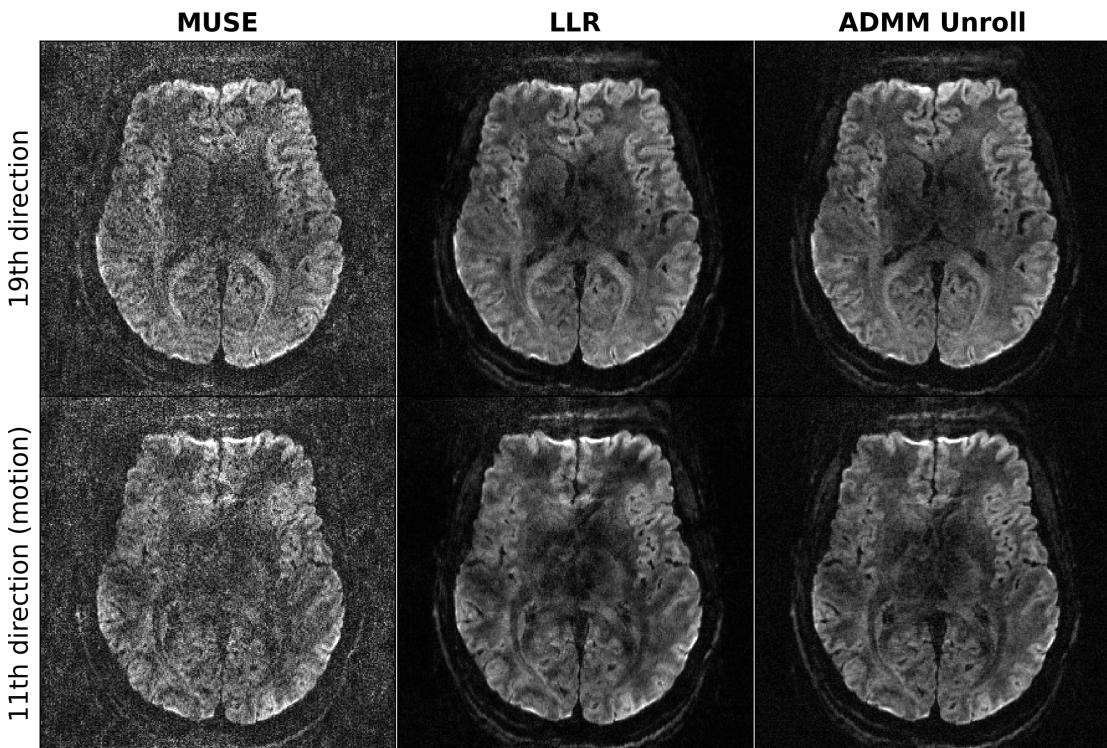


Figure 2: (A) Navigated and (B) selfgated reconstructions: MUSE, LLR, and ADMM Unroll from left to right. For both navigated and selfgated reconstructions, two diffusion directions are displayed. The first one (the 19th direction) shows clear DWI and no motion corruption, whereas the second (the 11th direction) shows artifacts in all reconstruction methods as the navigator data is corrupted by motion. The use of only imaging data, i.e., selfgated DWI reconstruction, shows reduced motion artifacts, as evident especially in the proposed ADMM Unroll reconstruction. Another advantage of selfgated reconstruction is the reduction in scan time.

and present the results in Figure 8 in the manuscript.

- (6) *While the authors compare slice-by-slice and single-slice training in Fig. 3, this comparison remains largely qualitative. The single-slice training result appears to exhibit diminished diffusion contrast, which was not reflected in the signal plot—likely because the selected voxel did not fall within a region of strong diffusion contrast. A more compelling evaluation could be plotting the signal variance across diffusion encoding directions to assess whether single-slice training introduces angular smoothing.*

Thank you for the valuable comment. Figure 4 in the manuscript presents both the difference and the signal variance in two regions of interest (mean and standard deviation) across diffusion directions between the two training strategies.

Reviewer 2

This paper proposes a self-supervised unrolled reconstruction algorithm for high-resolution and motion-robust diffusion-weighted imaging. Experimental results on a clinical 7T scanner demonstrate the effectiveness of the proposed method. Some issues remain to be addressed:

Method:

- (1) *The authors partitioned the sampling masks into 12 repetitions. Are the masks consistent across these repetitions? Each mask is subdivided into three disjoint subsets. What advantages does this approach offer compared to partitioning the masks into only two disjoint subsets?*

Thank you for the questions.

The masks are not consistent across repetitions. This is also reflected in the code: https://github.com/ZhengguoTan/DeepDWI/blob/main/examples/run_zssl.py#L45. For each repetition, the function ‘uniform_samp’ uses ‘torch.multinomial’ to sample a new mask (<https://github.com/ZhengguoTan/DeepDWI/blob/main/src/deepdwi/recons/zssl.py#L34>).

The zero-shot learning approach as is done in this work requires three disjoint subsets for training, training loss and validation loss (refer to Figure 2). In this way, a model can be trained and tested on a single dataset, without the need of large-scale datasets. In the case of only two disjoint subsets, you will need extra data for validation, as is done in SSDU.

- (2) *Dividing the sampling masks into three disjoint subsets reduces the amount of original effective data input for each part of the network. Does this approach allow the network to learn the overall data distribution effectively? Have the authors considered this potential limitation?*

Thank you for the insightful comment. As shown in Figure 9 in the manuscript, we train the model with 100 epoches and each epoch consists of 12 repetitions, both training and validation loss show convergent behavior. This approach allows the network to learn the overall data distribution.

- (3) *The authors proposed two distinct training strategies. I want to know whether the model was trained on a subset of data from a single volunteer and tested on another subset from the same volunteer.*

Yes, the model was trained on a subset of data (i.e., one multi-band slice k -space data) from a single volunteer and tested on another subset (i.e., the remaining multi-band slices) from the same volunteer.

- (4) *I would appreciate it if the authors could clarify the aspect in which motion robustness is manifested. Is it achieved through phase shift correction, or does it involve another technique?*

In our observation and from the comparison with other methods such as LLR, the motion robustness is achieved by the unrolled reconstruction utilizing spatial-diffusion convolutions.

Experiments:

- (1) *The experiment utilized data from only three volunteers, representing a very small sample size.*

Thank you for the comment. We put this as a limitation in the Discussion.

- (2) *Would it be possible to include deep learning-based self-supervised methods, such as SSDU, and multi-mask SSDU, as additional comparison methods?*

Thank you for the suggestion. As mentioned above, SSDU requires several datasets for training. Given the small sample size in this study, we are not able to compare our approach with SSDU type methods. However, we compare our approach with MUSE and LLR, which we believe are competitive and representative methods.

- (3) *In Figure 4, the authors assert that the unrolled ADMM offers an advantage in preserving clear tissue boundaries; however, the region marked by the red arrows in the figure is not clearly defined.*

Thank you for the insightful comment, which is inline with (3) from Reviewer 1. We provide detailed explanation above and update the manuscript with more compelling figures.

- (4) *The experiments in this paper are conducted using data from a single volunteer. Did the authors attempt to train on one part of the volunteer's data and test on another part from the same volunteer?*

Thank you for the questions. We clarify in the manuscript that this work recruited three volunteers. We tested both slice-by-slice training and single-slice training, both of which illustrate quantitatively similar results (refer to Figure 4). Yes, the single-slice training strategy is done by training on one part of the volunteer's data and testing on another part from the same volunteer.

- (5) *In Figure 7, could the authors clarify why the validation loss is lower than the training loss in the early stages of training?*

Thank you for the insightful comment. We explain in the Results section that "this is because more data is split into the training mask than the validation mask". In the function 'uniform_samp', the parameter 'rho' is used to control the split ratio. Please refer to <https://github.com/ZhengguoTan/DeepDWI/blob/main/src/deepdwi/recons/zssl.py#L14>.