The Problem of Hidden Noise in MR Image Reconstruction

Jiayang Wang1, Di An1, and Justin Haldar1
1Signal and Image Processing Institute, Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, United States

Synopsis

Keywords: Image Reconstruction, Machine Learning/AI
d

The performance of modern image reconstruction methods is commonly judged using quantitative error metrics like mean squared error and the structural similarity index, where these error metrics are calculated by comparing a reconstruction against fully-sampled reference data. In practice, this reference data contains noise and is not a true gold standard. In this work, we demonstrate that this "hidden noise" can confound performance assessment methods, leading to image quality degradations when typical error metrics are used to tune image reconstruction performance. We also demonstrate that a new error metric, based on the non-central chi distribution, helps resolve this issue.

INTRODUCTION

In the image reconstruction literature, it is commonplace to judge the quality of reconstruction methods using performance metrics like mean-squared error (MSE), mean absolute error (MAE), and structural similarity (SSIM). These metrics are also frequently used to tune reconstruction parameters, with parameters selected to optimize these metrics for training datasets.

Calculation of MSE, MAE, and SSIM requires comparing a reconstructed image with a reference image, and fully-sampled MRI data is typically used as the reference. However, real MRI data will always contain noise. We call this "hidden noise," since conventional error metrics would treat it as part of the reference, rather than properly treating it as error.

Importantly, hidden noise can be substantial, even when it is not visually obvious. To illustrate, Fig. 1 shows a typical multichannel brain MRI dataset from the widely-used fastMRI database. While the root sum-of-squares (rSoS) image (a common choice of reference image in the literature) visually appears to have excellent SNR, closer examination demonstrates that there is substantial noise present in the individual channels. This implies that the pristine appearance of the rSoS image is misleading, and hides a substantial noise bias. (Note that rSoS images should follow the non-central chi distribution, with a biased mean value). Hidden noise is also evident in k-space, as illustrated in Fig. 2, which demonstrates that the SNR is small (<2) for substantial portions of k-space and suggests that high-frequency data is unreliable.

In this work, we investigate how this hidden noise can impact the assessment of image reconstruction performance. Importantly, we observe that the effects of hidden noise can mislead the interpretation of results. This can cause different reconstruction methods to be ranked incorrectly, leading to suboptimal reconstruction quality when standard performance metrics are used to optimize reconstruction performance. However, we also demonstrate that a novel metric based on the non-central chi distribution does not have the same confounds.

METHODS

We performed two separate experiments to assess the impact of hidden noise.

In the first experiment, we acquired a standard fully-sampled MPRAGE dataset. In order to obtain both a typical reference dataset (to emulate standard performance assessment with a noisy reference) and a high-SNR reference dataset (to more accurately assess performance), the data was acquired with five averages. A single average was used to create the typical (noisy) reference, while all averages (with additional SVD-based denoising) were used for the high-SNR reference. To emulate a typical parameter tuning application in accelerated image reconstruction, the single-average k-space data was retrospectively undersampled (acceleration factor 3), and images were reconstructed using Autocalibrated LORAKS with many different regularization parameter choices. We then compared how well the conventional MSE, MAE, and SSIM metrics obtained using the noisy reference compared against metrics obtained using the averaged reference. We also compared against a novel performance metric obtained using the negative log-likelihood of the non-central chi distribution.

In the second experiment, we applied SURE-LET denoising to 3402 slices of multichannel 12-weighted brain images from fastMRI to create synthetic high-SNR reference images. We then added simulated Gaussian noise following the original noise statistics back to these synthetic datasets to create realistic noisy data. The noisy datasets were then retrospectively undersampled (acceleration factor 4), and we trained an image reconstruction neural network (E2E-VarNet) using the noisy data, with a 3000:200:200 split for training, validation, and testing. We created 24 different VarNets using different choices of loss function (MSE or MAE), different numbers of training epochs, and different numbers of cascades. We compared MSE, MAE, and SSIM metrics obtained using noisy reference images against metrics obtained using clean reference images, and also compared against our novel non-central chi metric.

RESULTS

Figure 3 shows the results of the first (Autocalibrated LORAKS) experiment. We observe that error metrics obtained with the averaged reference have poor correspondence with the error metrics obtained with the typical noisy reference. Indeed, the metrics are negatively correlated, implying that tuning the regularization parameter based on the noisy reference would degrade image quality. In contrast, our proposed non-central chi metric had excellent correspondence with the error metrics computed using the averaged reference. The impact this has on optimal parameter tuning is shown in Fig. 4, which demonstrates that choosing the regularization parameter based on the non-central chi metric leads to a sharper-looking image, while choosing the regularization parameter based on conventional metrics leads to a blurrier-looking image.
Figure 5 shows results of the second experiment. We again observe that standard error metrics calculated with a noisy reference have poor correspondence with error metrics calculated with a clean reference, while our non-central chi metric had much stronger correlation with the clean error metrics.

CONCLUSIONS

Our results demonstrate that, although hidden noise is usually neglected in the modern reconstruction literature, it can substantially confound the assessment of image reconstruction performance. Under the conventional paradigm, this will lead to suboptimal ranking of different reconstructions and unnecessary degradations in image quality. These issues can be mitigated by using a novel metric that is cognizant of noise in the reference data. Our findings suggest that there may be substantial advantages to changing the prevailing paradigm for ranking reconstruction methods.

Acknowledgements

This work was supported in part by NIH grants R01-MH116173 and R01-NS074980, the Ming Hsieh Institute for Research on Engineering-Medicine for Cancer, and a USC Annenberg Graduate Fellowship.

References


Figures

Figure 1. A representative multichannel T2-weighted brain image from the fastMRI database. Left: rSoS image. Right: Depiction of the 16 original individual channels that were combined to form the rSoS image. It is visually obvious that the individual channels each contain substantial amounts of noise.

Figure 2. (a) 16-channel k-space data corresponding to the same dataset from Fig. 1. (b) The same data from (a), but normalized by the noise variance of each channel and color-coded to more easily visualize the SNR characteristics in different regions of k-space. In all channels, we observe that there are substantial regions of k-space with SNR<2.
Figure 3. Comparisons of different error metrics for the first experiment (AC-LORAKS with different regularization parameters). The top row shows that the MSE, MAE, and SSIM values obtained using a high-quality (5×-averaged) reference are negatively correlated with the same error metrics obtained using a more-typical noisy reference, while the bottom row demonstrates excellent correlation between the high-quality reference metrics and our proposed non-central chi metric.

Figure 4. Images from the first experiment. The first two images are the noisy and 5×-averaged references. AC-LORAKS 1 was obtained by choosing regularization to optimize the MSE with respect to the noisy reference, while AC-LORAKS 2 optimized the non-central chi metric. When compared to the 5×-averaged reference, the AC-LORAKS 2 reconstruction outperforms the AC-LORAKS 1 reconstruction with respect to all of the MSE, MAE, and SSIM metrics, and is also visually sharper.

Figure 5. Comparisons of different error metrics for the second experiment (different variations of E2E-VarNet). The top row shows that the MSE, MAE, and SSIM values obtained using a clean reference are not well correlated with the same error metrics obtained using a typical noisy reference (notably, the "best-performing" reconstructions identified using a noisy reference are not actually the best), while the bottom row demonstrates excellent correlation between the clean reference metrics and our proposed non-central chi metric.