HydraNet: A Multi-branch Convolution Neural Network Architecture for MRI denoising

Stephen Gregory[a], Hu Cheng[b], Sharlene Newman[c,d], Yu Gan[d]
[a]Department of Computer Science, The University of Alabama
[b]Department of Psychological and Brain Sciences, Indiana University, Bloomington
[c]Department of Psychology, The University of Alabama
[d]Department of Electrical and Computer Engineering, The University of Alabama

ABSTRACT

The current state-of-the-art methods of Magnetic Resonance Imaging (MRI) denoising technologies have improved significantly in the past decade, particularly those based in deep learning. However, the major issue in deep learning based denoising algorithms is that the data given to these algorithms is typically synthetic, and thus, they fail to generalize to spatially variant noise distributions. The noise varies greatly dependent upon such factors as pulse sequence of the MRI sequence, reconstruction method, coil configuration, physiological activities, etc. To overcome these issues, we've created HydraNet, a multi-branch deep neural network architecture that learns to denoise MR images at a multitude of noise levels, and which has *critically* been trained using only real image pairs of high and low signal-to-noise ratio (SNR) images. We prove the superiority of HydraNet at denoising a multitude of noise levels in comparison to other deep learning methods in our ablation study, in addition to non-local collaborative filtering-based methods, in both Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

Keywords: Convolution Neural Network, Denoising, MRI, Patch-based, Residual

1. INTRODUCTION

Magnetic resonance imaging (MRI) denoising is a mission-critical problem in clinical applications, and the demand for better techniques has steadily increased over the past few decades. The advances of coil technology, parallel imaging and simultaneous multi-slice acquisition have reduced the scan time significantly, but have also led to complicated noise behavior. Typically, the noise distribution is no longer spatially invariant [1, 2]. For instance, Fig.1 shows a comparison between a low signal-to-noise ratio (SNR) scan ('noisy' image) acquired with a short acquisition time of 1'14" using a Wave-CAIPI technique [3] and a high SNR scan ('clean' image) acquired with the standard MP-RAGE sequence in 7'42''. The residual image Fig.1(c), calculated as the difference between Fig. 1(a) and Fig. 1(b), highlights the differences in noise levels among region I, II, and III. Therefore, special consideration needs to be taken into account to address these issues when applying conventional deep learning denoising architectures from natural images to MRI images. However, most of the existing methods in MRI denoising [4, 5] are based upon synthetically generated training data[6] with a known, static noise level, and thus fail to achieve this goal satisfactorily. Furthermore, existing networks such as DnCNN [7] will fail to generalize to the task of denoising images when the noise distribution in testing datasets is different from that of training datasets [8]. This has particularly bottlenecked the performance of deep learning-based solutions due to the fact that single MR images often contain a multitude of noise levels.

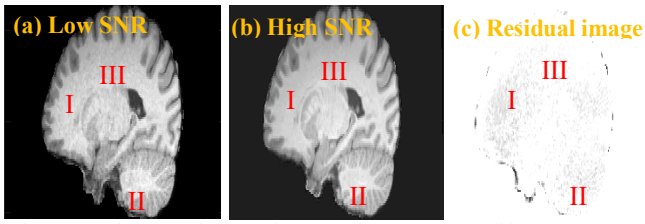

Figure 1. The SNR in MRI images shows a non-uniform pattern. In residual image (c), the SNR in region I is lower than region II and III. Images (a) and (b) are co-registered and contrast-enhanced.

To address these issues, we propose HydraNet, a multi-branch network architecture that uses dynamic network selection to denoise local patches using separately trained CNNs that consider various noise levels. Critically, HydraNet has been trained on *real, non-synthesized input data* (e.g. image pair of Fig. 1(a) and (b)) in order to adequately account for the dynamic noise levels of MR images. In this study, we implemented this framework by using a three-branch structure, each branch of which is a neural network trained to learn the distribution of the residual of input images and denoise them accordingly. Our framework learns to simultaneously denoise high-noise regions that require the largest degree of denoising (such as the deep brain region) while not interfering with the network's ability to denoise lower noise level regions. Our experimentation demonstrates the effectiveness of this architecture

on real data and shows promising results in terms of Peak Signal to Noise Ratio (PSNR) and Structural Similarity index (SSIM).

2. METHODS

2.1. Overall framework

The overall framework of HydraNet is shown in Fig. 2. HydraNet is composed of 4 steps in serial: preprocessing, patch-denoiser assignment, denoising, and post-processing. The MRI images are processed on a patch-wise level, with each patch being selectively denoised based on its structural information, which is correlated to noise level.

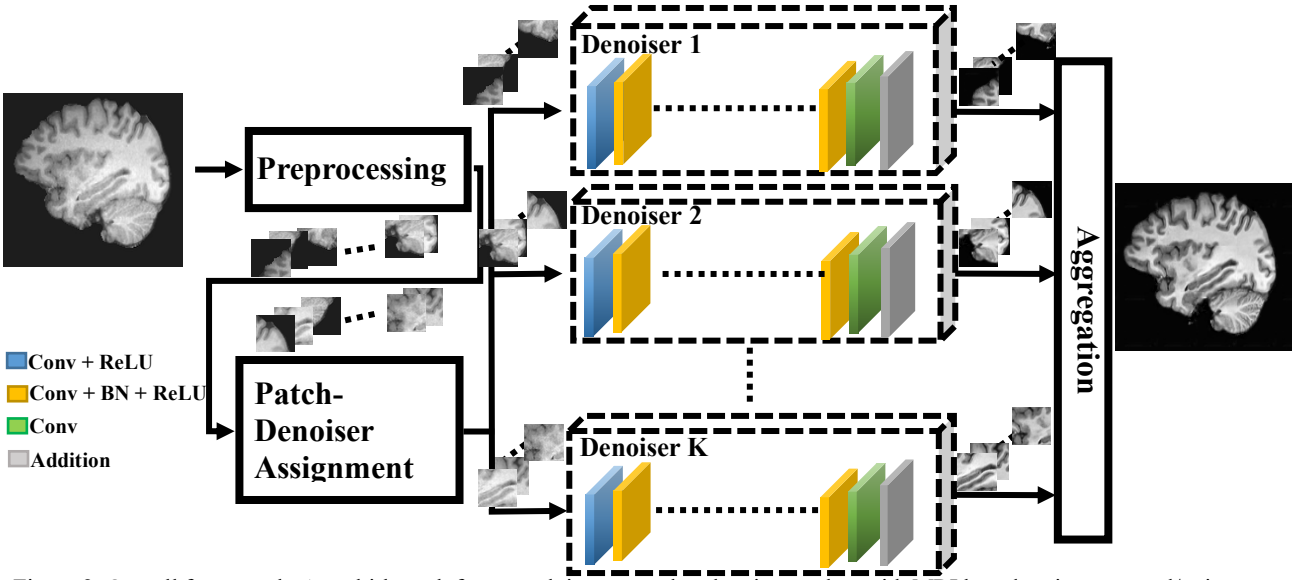

Figure 2. Overall framework. A multi-branch framework is proposed to denoise patches with MRI based on its structural/noise distribution.

2.2. Preprocessing

We prune our training and testing data by first using the Brain Extraction Tool (BET) [9] to isolate brain tissue from T1-weighted anatomical images; we then use an Image Co-registration tool in SPM 12 [10] to produce matching pairs of low-high SNR MRI scans, which we slice in the sagittal plane to produce matching volumes of images. We then perform Contrast Limited Adaptive Histogram Equalization (CLAHE) [11] on the high SNR images to obtain a more uniform level of contrast. Then, we use histogram equalization on each noisy image to match their intensity distributions to their corresponding clean image, thus removing additional unwanted bias on each slice. Finally, before the images are fed to HydraNet, the pixel values are standardized [2], to facilitate easier learning. Co-registration and histogram equalization of testing data to high-SNR images are only performed for validation purpose.

2.3. Denoising network

Each denoiser network in Fig. 2 is built using the DnCNN [7] architecture. In this architecture, each network is composed of a Convolutional layer containing 64 Convolutional kernels with sizes of 7x7 and a ReLU activation layer, followed by 17 Convolutional layers, each containing 64 Convolutional kernels with sizes of 3x3, Batch Normalization and ReLU activation. Finally, in the interest of learning the residual distribution, we subtract our learned feature representation from the input to obtain a denoised image prediction. Here, we use the residual sum of squares as our loss function for backpropagation.

2.4. Training and testing

Our learning objective is to divide the training set(s) into multiple groups based on noise level and subsequently train the each denoiser network separately. We rely upon the distribution of noise to set the threshold for each noise level. We first calculate a residual image for each pair of 40x40 image patches, and then calculate the standard deviation of this residual, which provides a robust estimate of the noise level of each patch. We then separate our training data into

3 groups based upon these residual standard deviations: low-noise patches with a standard deviation in the range (0, 0.15), medium-noise patches with a standard deviation in the range (0.05, 0.2), and high-noise patches with a standard deviation in the range (0.1, ∞). We use an overlap of the regions for a very specific reason: it is found that for each network, utilizing training data that covers a slightly larger noise range than that which is used for inference allows each CNN to become a "specialist" at denoising a specific noise level, while retaining competency at other noise levels. We trained each network with a batch size of 128 for a relatively small number of epochs (~23), as the model began to converge to an approximately fixed loss value close to zero at this number of epochs.

During testing, we separate each image into 40x40 patches. We then cross-reference each input patch with a subset of the patches used for training in order to find the training patch that is structurally closest to the input patch based on SSIM. Then we refer to the noise level of that training image to assign this patch to a specific denoiser. We then denoise each patch using the denoiser corresponding to the patch's noise level, after which we aggregate all denoised patches and mask out artifacts in the non-brain regions of the image to create a final, denoised image as output. To evaluate the performance of testing, we compare the SSIM and PSNR between the output and its corresponding high-SNR image.

3. EXPERIMENTS

3.1. Dataset

Two sets of T1-weighted 3D MRI images were collected on two human subjects on a Siemens 3T Prisma scanner with a 64-channel head coil. The high SNR image was acquired with the standard MP-RAGE pulse sequence (TR/TI/TE = 1800/900/2.7 ms, flip angle = 8°, scan time = 7'42'') while the low SNR image was acquired with a Wave-CAIPI MP-RAGE pulse sequence (TR/TI/TE = 2300/900/3.48 ms, flip angle = 8°, acceleration factor = 3 in phase encoding direction x 3 in 3D direction, scan time = 1'14''). The resolution of the image is 1 mm isotropic.

We use 112 slices from Subject 1 for training and 7 slices from Subject 2 for testing, with each slice being split into 36 patches with sizes of 40x40. In the testing data, we picked slices which are at least 10 slices away from each other, naturally excluded the slices where brain region is too small.

3.2. Results

We compare the denoising performance of HydraNet with three methods: DnCNN (single), DnCNN (synth), and BM3D Brushlet. DnCNN (one branch) uses the same training data as HydraNet, but with a single branch of DnCNN as its only denoiser. DnCNN (synthetic) is a clone of the original DnCNN [7], where synthetic noise with a fixed noise level is used for training. These DnCNN-based methods serve as an ablation study. BM3D-Brushlet **[7]** is a patch-based collaborative filtering method relying solely upon non-local similarity. Two low SNR images taken from the testing corpus and their resultant counterparts are shown in Fig. 3. In both images, we observe that the yellow region has generally lower SNR level than the red region. Evidently, we observe that HydraNet is the only method that effectively denoises both regions, while the two DnCNN-based methods only adequately denoise one region. Noticeably, by comparing Fig. 3(c) with Fig. 3(d) and comparing Fig. 3(i) with Fig. 3(j), we found that the performance

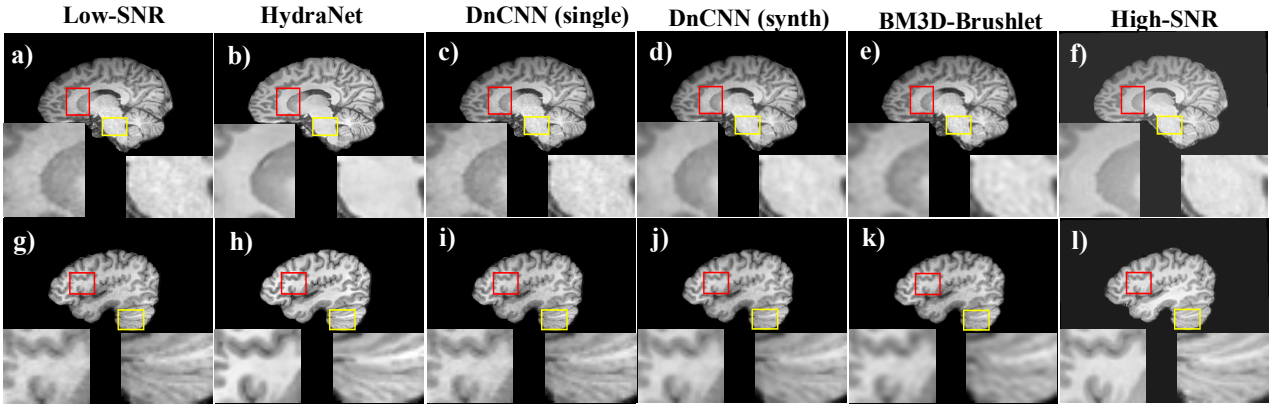

Figure 3. Original low SNR images (a, g) and high SNR images (f, l). Denoising results from HydraNet (b,g) in comparison with DnCNN one branch (c, h), DnCNN with synthetic noise training (d, i), and BM3D-Brushlet (e, j).

of choosing one branch can be no better than directly using synthetic data. Similar to our multi-branch scheme, BM3D-brushlet uniformly denoises regions with different noise levels; however, the results for each region prove to be systematically inferior to those the deep-learning based HydraNet. This visual observation is further validated by the two quantitative metrics, SSIM [12] and PSNR [13], shown in Table 1: HydraNet has the highest SSIM *and* PSNR amongst all four methods. Note that the grossly large SSIM values result from the large amount of purely black surface area present in our MR images. Overall, the experimental results demonstrate that our deep-learning method takes a promising step in improving conventional patch-based collaborative filtering methods, and critically, we see that our multi-branched design and use of real training data is essential to achieve premium denoising performance, especially considering the many noise levels that constitute MR images.

Table 1. Comparison of image quality measures.

Method	SSIM	PSNR (dB)
HydraNet	**0.9689**	**28.9390**
DnCNN [7] **(single)**	0.9647	27.9831
DnCNN [7] **(synth)**	0.9684	28.1132
BM3D-Brushlet [7]	0.9640	27.7017

4. DISCUSSION AND CONCLUSION

We propose HydraNet, a novel multi-branch image denoising framework, to denoise MR images with spatially variant noise distributions. Our experimentation shows that our method outperforms existing collaborative filtering methods as well as deep learning methods using one-branch or synthetic data. Importantly, our MRI denoising method can be generalized to any number of complex denoising tasks in which the noise level is variant across the entire image. In the future, we plan to improve the efficiency, and hence viability, of the HydraNet architecture by optimizing the Patch-Denoiser Assignment stage, distilling the denoiser structure, and implementing parallel computing and explore the feasibility of volumetric denoising. Our methods forego the usage of hand-crafted mechanisms and their latent biases toward any particular domains, and learn to robustly denoise a wide range of noise levels in medical images.

REFERENCES

[1] P. M. Robson, A. K. Grant, A. J. Madhuranthakam *et al.*, "Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions," Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 60(4), 895-907 (2008).

[2] M. Barth, F. Breuer, P. J. Koopmans *et al.*, "Simultaneous multislice (SMS) imaging techniques," Magnetic resonance in medicine, 75(1), 63-81 (2016).

[3] D. Polak, K. Setsompop, S. F. Cauley *et al.*, "Wave-CAIPI for highly accelerated MP-RAGE imaging," Magnetic resonance in medicine, 79(1), 401-406 (2018).

[4] D. Jiang, W. Dou, L. Vosters *et al.*, "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," Japanese Journal of Radiology, 36(9), 566-574 (2018).

[5] J. V. Manjón, and P. Coupe, "MRI denoising using Deep Learning and Non-local averaging," arXiv, arXiv: 1911.04798 (2019).

[6] D. L. Collins, A. P. Zijdenbos, V. Kollokian *et al.*, "Design and construction of a realistic digital brain phantom," IEEE transactions on medical imaging, 17(3), 463-468 (1998).

[7] K. Zhang, W. Zuo, Y. Chen *et al.*, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," IEEE Transactions on Image Processing, 26(7), 3142-3155 (2017).

[8] J. H. Choi, O. A. Elgendy, and S. H. Chan, "Optimal combination of image denoisers," IEEE Transactions on Image Processing, 28(8), 4016-4031 (2019).

[9] S. M. Smith, "Fast robust automated brain extraction," Human brain mapping, 17(3), 143-155 (2002).

[10] J. Ashburner, G. Barnes, C. Chen *et al.*, "SPM12 manual," Wellcome Trust Centre for Neuroimaging, London, UK, 2464 (2014).

[11] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," Journal of VLSI signal processing systems for signal, image and video technology, 38(1), 35-44 (2004).

[12] W. Zhou, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, 13(4), 600-612 (2004).

[13] A. Hore, and D. Ziou, "Image quality metrics: PSNR vs. SSIM." 2366-2369.