HydraNet: Multi-branch Convolution Neural Network Architecture for MRI Denoising

Stephen Gregory[a], Hu Cheng[b], Sharlene Newman[c], and 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 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 issues in deep learning based denoising algorithms is both that the model architectures are not built for the complex noise distributions inherent in MRI, and 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 have 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 complex noise distributions in comparison to the leading deep learning method in our experimentation, in addition to non-local collaborative filtering-based methods, quantitatively in both Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM), and qualitatively upon inspection of denoised MRI samples.

Keywords: Convolutional 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. Often, MRI technicians are faced with a set of mutually exclusive choices when carrying out a scan, and are forced to make compromises. In particular, technicians can choose to optimize for lower scan times while giving up resolution and allowing greater noise, producing images with a lower Signal-to-Noise ratio (SNR), or they can choose to optimize for SNR and subsequently allow for longer scan times, increasing costs. The advances of coil technology, parallel imaging and simultaneous multi-slice acquisition have reduced scan times significantly, but also lead to complicated noise behavior. It is a common mistake to assume that the noise distribution can be modeled by a simple Gaussian; however, it has been shown that the data instead follows a Rician distribution, and is not spatially invariant.[1,2]

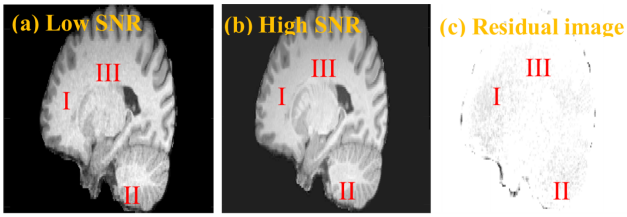

Figure 1. The SNR in MRI images shows a non-uniform pattern. In residual image (c), the SNR in region I is lower than region II and III. Images (a) and (b) are co-registered and contrast-enhanced.

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'

Further author information: (Send correspondence to Stephen Gregory)
Stephen Gregory: E-mail: shgregory@crimson.ua.edu, Telephone: 1-(615)-516-1663
Yu Gan: E-mail: ygan6@eng.ua.edu, Telephone: 1-(205)-348-6105

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. This showcases the issue with using denoising techniques borrowed from natural images: MR images are fundamentally different than those taken from an optical camera, so treating the noise apparent in both as equal is a fallacious assumption. Where natural images often contain noise distributions which are spatially invariant (i.e. the noise is uniform throughout the entirety of the image), MR images contain various levels of noise scattered throughout images. 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 test data sets is different from that of training data sets.[8] This has particularly bottle-necked the performance of deep learning-based solutions due to the aforementioned multiplicity and complexity of noise in MR images.

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 characteristics. Critically, HydraNet has been trained on real, non-synthesized input data (e.g. image pair of Fig. 1(a) and Fig. 1(b)) in order to adequately account for the dynamic noise patterns 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 regions of the brain with complicated, high-magnitude noise distributions (such as centermost region of the brain) while not interfering with the network's ability to denoise less complicated, lower-magnitude noise 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, patch-based denoising, and post-processing. The MRI images are processed on a patch-wise level, with each patch being selectively denoised based on run-time prediction of its noise characteristics. This run-time prediction, explicated further in Sec. oise Characteristic Estimation], allows HydraNet to optimize the task of denoising such that each specialized patch-based denoiser is only responsible for denoising a subset of the space of possible noise types.

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 purposes, and the performance and efficacy of our methods are independent of these steps.

2.3 Patch-based Denoising network(s)

Each patch-based 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. Each layer uses a 1x1 stride and 'same' padding, which results in equivalently sized input and output at each layer. Finally, in the interest of learning the residual distribution,

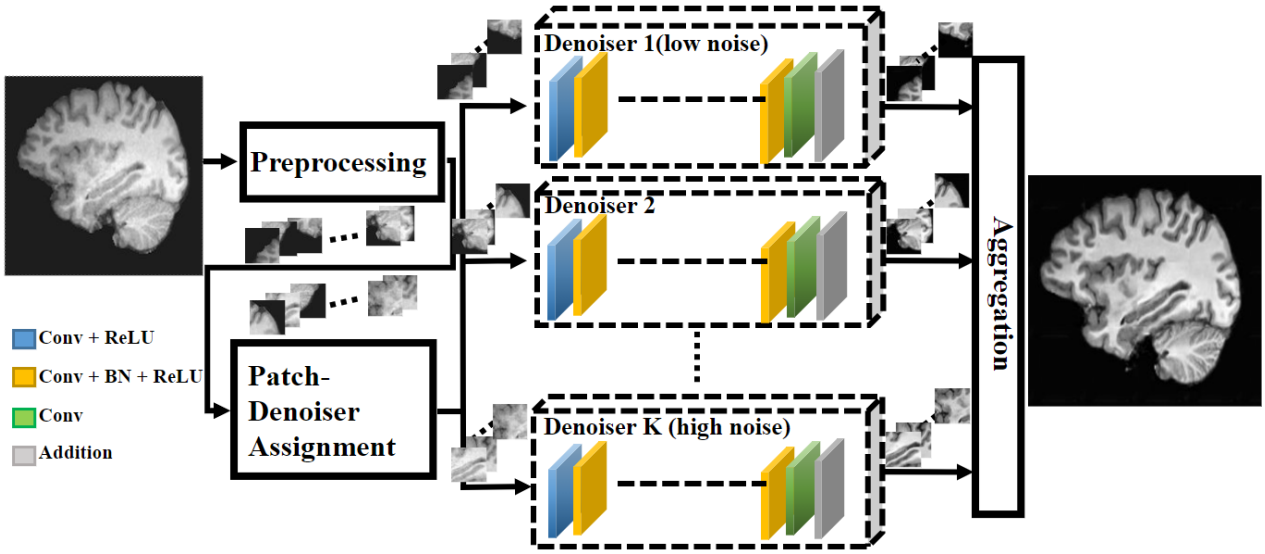

Figure 2. Overall framework. A multi-branch framework is proposed to denoise individually denoise MRI patches based on their structural/noise distribution

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 with an Adam[12] optimizer for backpropogation.

2.4 Training and Testing

To learn our patch-based denoiser networks, our objective is to divide the training set(s) into multiple groups based on noise characteristics and subsequently train each denoiser network separately on the respective training dataset groups. We employ three metrics for estimating the noise characteristics for each image patch: residual standard deviation, peak signal-to-noise ratio (PSNR), and slice location.

2.4.1 Residual Standard Deviation / PSNR

In 2 of our 3 training/testing regimes, we rely upon residual standard deviation and PSNR of individual patches, respectively, to set thresholds for each noise category. We first calculate the *residual standard deviation* based on the measurement of PSNR of each low SNR image for every pair of 40x40 image patches, which provides a robust estimate of the noise level. In the PSNR experiment, we then separate our training data into 3 groups based upon these PSNR values: low-noise patches with a PSNR in the range $(30.0, inf)$, medium-noise patches with a PSNR in the range $(15.0, 40.0)$, and high-noise patches with a PSNR in the range $(0.0, 30.0)$. In the residual standard deviation experiment, we then separate our training data into 3 groups based upon these residual standard deviation values: 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, inf). We use an overlap of the values of both of these sets of ranges for a very specific reason: it is found that for each noise category, utilizing training data that covers a slightly larger range than that which is used for inference allows each CNN to become a "specialist" at denoising a specific noise category, while retaining competency at other patches pertaining to other categories. This is important as our noise category estimates are approximate by definition, and we often are faced with edge cases or patches which are not routed to the appropriate denoisers. We trained each network with a batch size of 128 image patches for 80 epochs, as the models began to approximately converge at this time. Unsurprisingly, the low-noise patch-based denoisers converged to a significantly lower loss value than did the medium-noise denoisers, which, in turn, reached a lower loss value still than the high-noise denoisers. The mean final loss value for low-noise denoisers is just 61% of the mean final loss value for medium-noise denoisers, which is itself just 84% of the mean final loss for high-noise denoisers. This supports our hypothesis that patches with noise of a lower magnitude or complexity are generally easier to denoise than those containing noise with high magnitude or complexity.

During testing, we separate each image into 40x40 patches. We 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. We then refer to the noise level of that most structurally similar image patch as an estimate of the noise characteristics of the training patch. Lastly, we denoise each patch using the denoiser corresponding to the patch's noise category, after which we aggregate all denoised patches and mask out artifacts in the non-brain regions of the image to create a patch-denoised image as output. To evaluate the performance of the algorithm, we compare the SSIM and PSNR between the denoised output slices and their corresponding high-SNR slices.

2.4.2 Slice Location

In our 3rd training/testing experiment, we use slice location as an estimator for the noise characteristics of each image patch. The first step of this experiment is to co-register all MRI scans to a single atlas image using SPM.[10] This enables fair and accurate analysis of the regions of different brain scans with direct slice-to-slice comparison. Next, we split our training data into three separate noise categories via sagittal region cross-correlation, where each of the 3 denoiser networks becomes a *specialist* at a particular region of the brain. Namely, 2 networks are trained to denoise each half of the brain, and the 3rd network is trained to denoise the center of the brain. Again, we utilize strategic overlapping of the noise regions in an effort to generalize the models, so each model is trained on approximately 50% of the total sagittal slices, though is used for approximately 33% of slices at inference time.

During testing while using slice location as our noise estimator, we first co-register the input volume with the same atlas as was used to train each model, after which we separate each image into 40x40 patches. However, we need not cross-reference input patches with a subset of training data, as we do not rely upon similarity to existing patches to provide a noise characteristic. Instead, we compute a robust estimate of the approximate location of each image patch. To accomplish this, we first calculate the boundaries of the brain region of the input volume by finding the first and last image slices containing brain tissue. Then, we use the index of the slice housing each patch in relation to the indices of the boundary slices to calculate the relative position of the patch within the volume. As explained previously, we have trained 3 models for each of the left, middle, and right portions of the brain. Therefore, we send patches located in the left and right 33% of the brain to the left- and right-brain denoisers, respectively, while sending patches located in the middle third to the middle-brain denoiser.

3. EXPERIMENTS

3.1 Noise Characteristic Estimation

The reason for employing multiple denoisers is simple: we hypothesize that accurate estimation of the approximate structural content and noise characteristics of each region of a 2D MRI image slice allows us to direct those patches to denoisers which have been "specialized" to denoise appropriate types of noise distributions. Crucially, the effectiveness of this scheme is completely dependent upon two assumptions:

i. The hypothesis that multiple specialized denoisers can perform better than a single general denoiser is indeed correct.

ii. The metric used to predict noise characteristics is a good estimator of the ground truth.

Proving this first assumption requires that the second assumption be proven as a prerequisite, so we have conducted experiments to measure the efficacy of our noise distribution estimation scheme.

3.1.1 SSIM as Noise Characteristic Estimation

As briefly mentioned in Section 2.4.1, our patch noise estimation scheme when using both PSNR and residual standard deviation is as follows: At inference time, a subset of training data is kept in memory, consisting of a list of pairs of matching low- and high-SNR images. To obtain a list of cross-referencing patches, each image is split up into 40x40 patches, similarly to the scheme used when denoising images. However, in this scheme, we set the stride of the sliding window that captures patches to be a fraction of the "kernel size" of the window itself, such that overlapping patches are captured. Then, we cross-reference each inference-time patch with this list by calculating the SSIM between the input image and each low-SNR reference image in the entire list. Finally, we

hypothesize that the PSNR of the reference patch with the highest SSIM when compared to the input patch is a good estimator for the PSNR of the input patch itself. Therefore, based upon this noise level estimate, we can determine which noise category the input patch is best described by (low, medium or high). Likewise, we also theorize that the residual standard deviation of the reference patch with the highest SSIM when compared to the input patch is a good estimator for the variance of the noise of the input patch self. Shown below in Fig. 3 is a measure of the correlation between true PSNR and the PSNR estimated by this procedure. This is verified by an average R^2 value, also referred to as a ***Coefficient of Determination*** of approximately 0.83, indicating a relatively large positive correlation and quantitatively showing that cross-referencing via maximum SSIM provides accurate estimators of noise level of image patches.

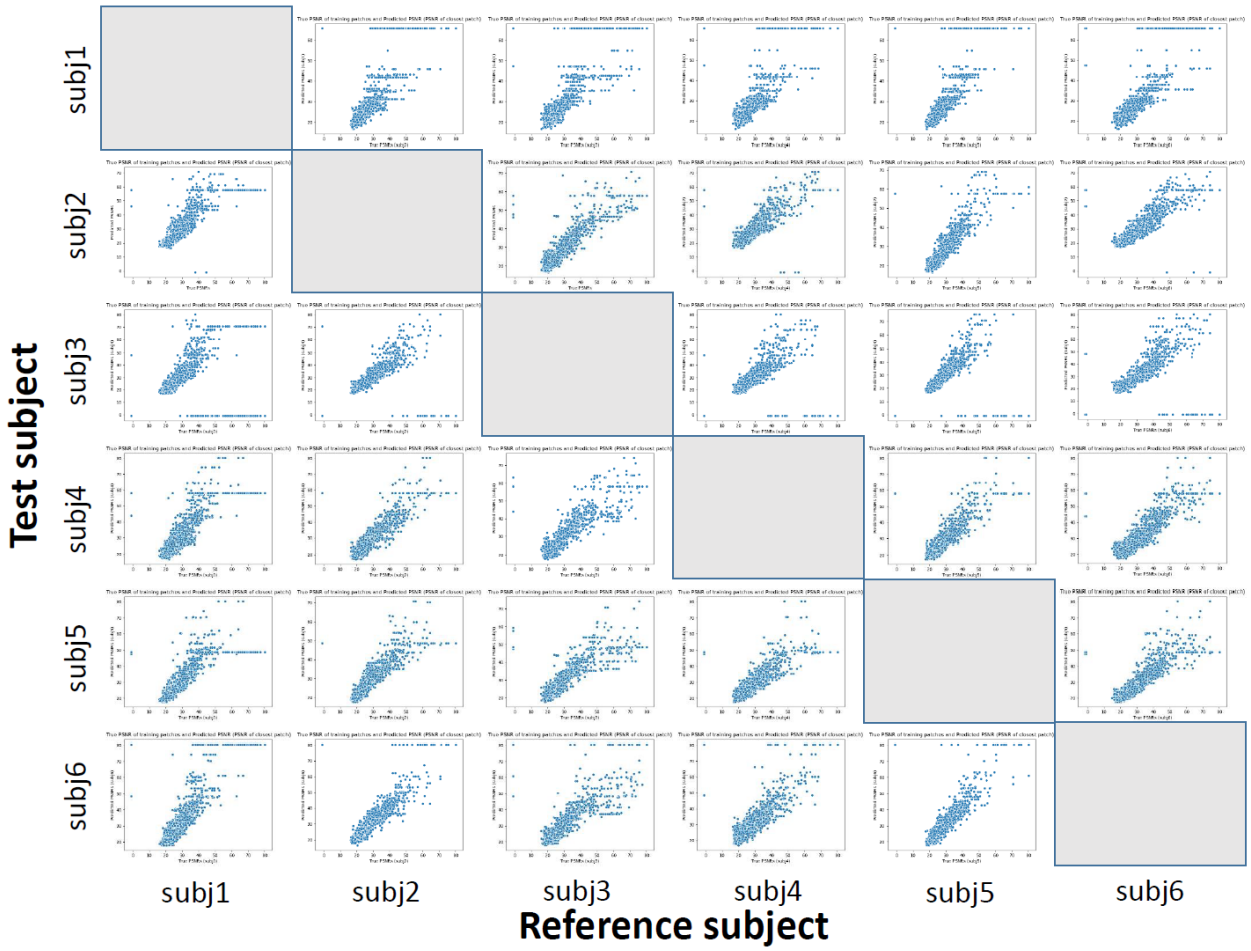

Figure 3. Correlations between True PSNR and Estimated PSNR for combinations of test and reference volumes

3.1.2 Slice Location as Noise Characteristic Estimation

As explained in Section 2.4.2, the run-time scheme for noise characterization prediction is quite different when using slice location as an estimator than when using PSNR or residual standard deviation. Here, an empirical, qualitative analysis of the noise distribution within separate locations of the brain is conducted. We hypothesize that, when co-registered properly, image slices from similar approximate regions of the brain exhibit similar noise characteristics. This is supported by the distance between areas of tissue and the MRI coil itself: the interior regions of the brain are farther from the MRI coil, and are separated from the exterior of the skull by a greater volume of tissue. Therefore, the spin-lattice-relaxation of the material within the interior of the brain must be detected through more tissue, weakening the signal and introducing the potential for noise. As is shown in Fig. 4, samples which are predicted to lie at the same relative location of the brain in the direction orthogonal to the

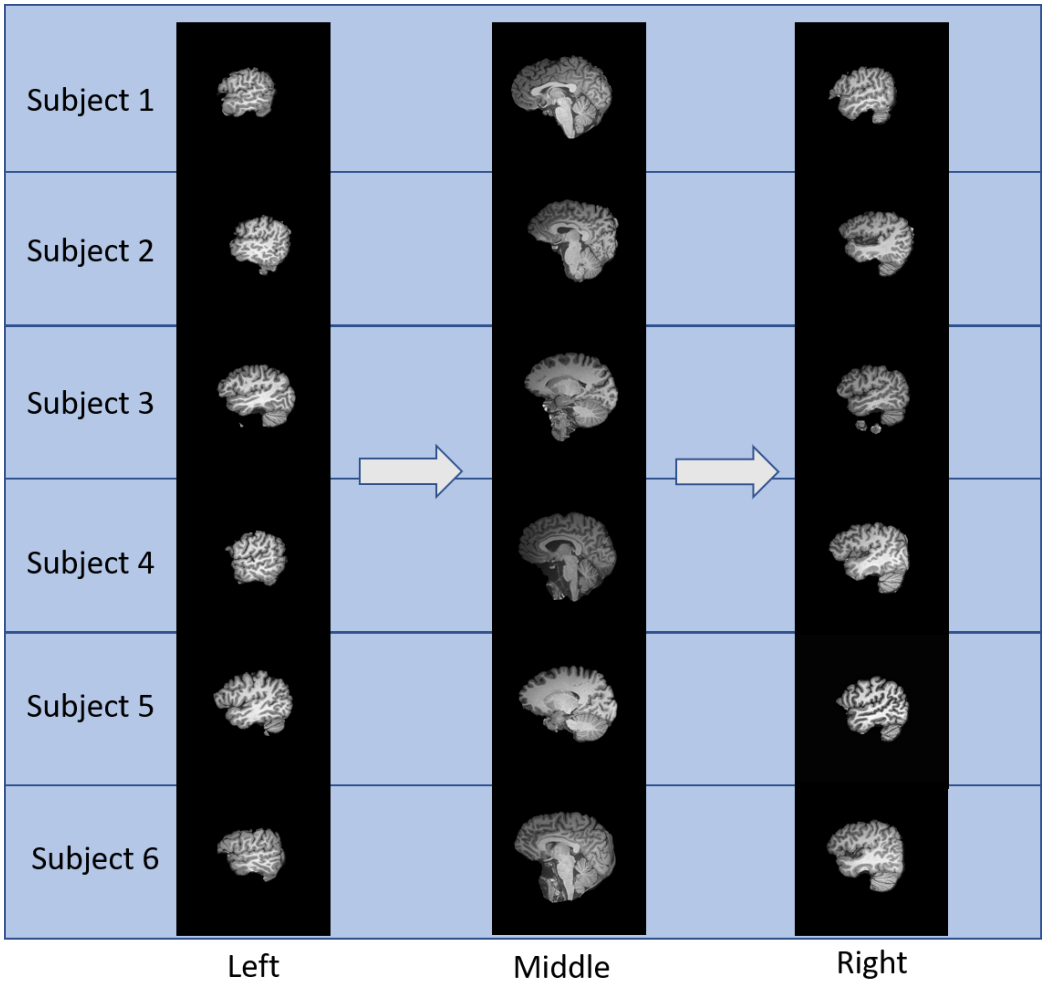

Figure 4. Sample slices from the left, middle, and right portions of each MRI subject, showcasing qualitative correlation between regions.

sagittal plane are indeed shown to be qualitatively similar. Future experiments should be conducted to evaluate the approximate noise levels between these regions of the brain with quantitative metrics, but, as discussed further in Section 3.3, there are inherent difficulties in the existing instantiations of such quantitative measures.

3.2 Dataset

Eight sets of T1-weighted 3D MRI images were collected on eight human subjects on a Siemens 3T Prisma scanner with a 64-channel head coil. The high SNR scans were 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 scans were 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 reserve 2 of our 6 cleanest subjects as testing data, and find that our instantiation of the HydraNet architecture learns to robustly denoise either of the 2 subjects with comparable efficacy. The 2 subjects used for testing were chosen as data for inference because they exhibited both the cleanest high-SNR elements, and the greatest discrepancy between their low-SNR and high-SNR pairs, and so are posed as a cardinal example of both a worthy denoising candidate and a trustworthy ground truth. For our first experiment, we use train-test splits of (721 train slices, 142 test slices) and (695 train slices, 168 test slices) corresponding to training on 5

subjects while withdrawing 1 subject for testing in both experiments. Samples are taken from each slice by sliding a $(40, 40)$ pixel window over the image with a stride of $(20, 20)$. Therefore, each image produces 144 patches, providing a training collection for each trial consisting of more than 100,000 image patch samples.

3.3 Results

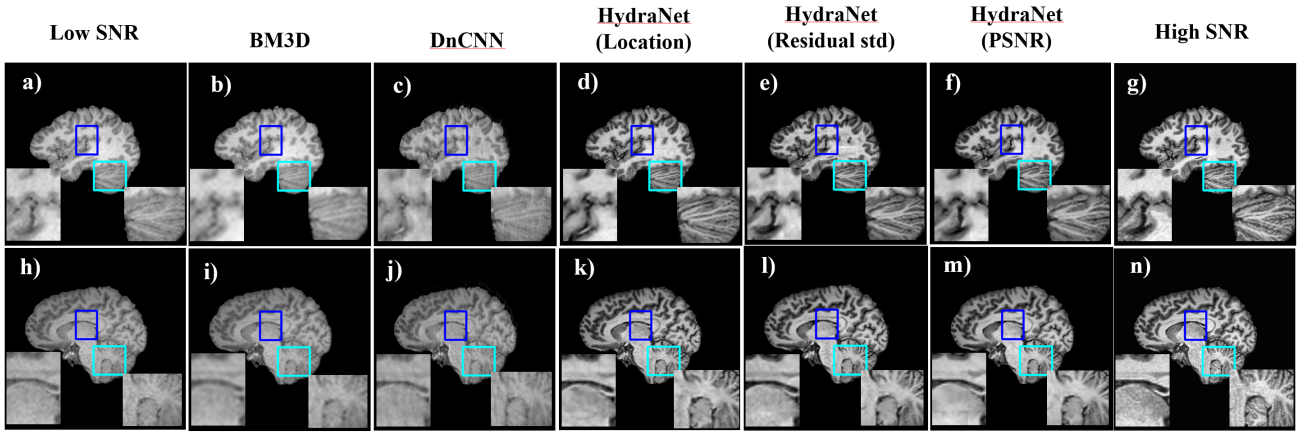

Figure 5. Original low SNR images (a, h) and high SNR images (g, n). Denoising results from instantiations of HydraNet (d, k), (e, l), and (f, m) in comparison with DnCNN (c, j) and BM3D-Brushlet (b, i).

Table 1. **Comparison of image quality measures**

Method	SSIM	PSNR (dB)
HydraNet (PSNR)	0.9540	**28.790**
HydraNet (Residual std)	0.9527	27.855
HydraNet (Location)	**0.9592**	28.420
DnCNN	0.8214	27.520
BM3D-Brushlet	0.9256	24.51

We compare the denoising performance of 3 versions HydraNet with the 2 state-of-the-art methods: DnCNN and BM3D Brushlet. The 3 versions of HydraNet, shown in Fig. 5, utilize all 3 of the proposed patch similarity metrics for training and inference. The first comparison, DnCNN, is trained with the same training data as HydraNet, using high-SNR images with Additive white Gaussian noise (AWGN) added at a fixed noise level as the noisy input, and with the high-SNR image as the target output for training, as outlined in the original DnCNN architecture. The second comparison, BM3D-Brushlet,[13] is a patch-based collaborative filtering method relying solely upon non-local similarity to denoise input images, and is a model-free architecture for denoising. Two low SNR images taken from the testing corpus and their resultant denoised counterparts are shown in Fig. 5. In both images, we can observe that the blue region has a generally lower/less complex noise level than the teal region. Evidently, we can observe that HydraNet is the only method that effectively denoises both regions. DnCNN performs qualitatively poorly here, as it is incapable of generalizing beyond its artificial training noise. 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 from the deep-learning based HydraNet. This qualitative analysis is particularly important when evaluating denoising methods, as metrics such as PSNR[14] and SSIM[15] are fundamentally flawed metrics that give only a limited insight into the real-world efficacy of denoisers. One such issue is that PSNR and SSIM are metrics which are not absolute; they only exist relative to other images. Therefore, if our denoised output is "better" or "less noisy" than our high-SNR ground truth, the PSNR and SSIM of this output is only decreased, even though a thorough qualitative analysis

of the results would indicate that the denoised output is superior. As a result, our visual analysis is supported by SSIM and PSNR, showing superior SSIM and PSNR from HydraNet, but more research is needed to evaluate these real-world denoising results denoising more robustly. Shown in Table 1, variations of HydraNet achieve the highest SSIM and PSNR amongst all four methods. Note that the a significant portion of each MRI slice is covered in black pixels, so, in an effort to accurately report our findings and limitations, we simply omit image patches that are found to contain only black pixels, and choose not to factor such patches into our average SSIM and PSNR results. These experimental results demonstrate that the HydraNet framework is capable of creating models with effective generalization to novel subjects after training on only 5 such subjects. We can see that our deep-learning-based methods take a promising step in improving conventional denoising methods, and critically, we see that the multi-branched, patch-based design and use of real training data is essential to achieving excellent denoising performance, especially considering the dynamic, complex distribution of noise in MRI images.

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. 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. While our model architecture has achieved good results by first slicing 3-dimensional MRI scans into sagittal 2-dimensional images, it seems that the best architecture would instead involve volumetric denoising. In this approach, we would exchange 2-dimensional convolution operations with 3-dimensional convolutions, and the input to the network would be $40 * 40 * c$ image patches, where c is the dimension orthogonal to the sagittal plane. This would further decrease any human biases, and increase the optimality and generality of the system. Furthermore, the system would gain access to a significantly greater amount of non-local information, and the reconstruction of 3-dimensional NIfTI MRI files would be a much more straightforward task with the discarding of 2-dimensional slicing. With such targets in sight, 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] Robson, P. M., Grant, A. K., Madhuranthakam, A. J., 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* **60(4)**, 895–907 (2008).

[2] Barth, M., Breuer, F., Koopmans, P. J., et al., "Simultaneous multislice (sms) imaging techniques," *Magnetic Resonance in Medicine* **75(1)**, 63–81 (2016).

[3] Polak, D., Setsompop, K., Cauley, S. F., et al., "Wave-caipi for highly accelerated mp-rage imaging," *Magnetic Resonance in Medicine* **79(1)**, 401–406 (2018).

[4] Jiang, D., Dou, W., Vosters, L., 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] Manjon, J. V. and Coupe, P., "Mri denoising using deep learning and non-local averaging," (2019).

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

[7] Zhang, K., Zuo, W., Chen, Y., 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] Choi, J. H., Elgendy, O. A., and Chan, S. H., "Optimal combination of image denoisers," *IEEE Transactions on Image Processing* **28(8)**, 4016–4031 (2019).

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

[10] Ashburner, J., Barnes, G., Chen, C., et al., "Spm12 manual." Wellcome Trust Centre for Neuroimaging (2014).

[11] Reza, A. M., "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] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," (2017).
[13] Ahn, B. and Cho, N. I., "Block-matching convolutional neural network for image denoising," (2017).
[14] Horé, A. and Ziou, D., "Image quality metrics: Psnr vs. ssim," in [*2010 20th International Conference on Pattern Recognition*], 2366–2369 (2010).
[15] Zhou, W., Bovik, A. C., Sheikh, H. R., et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13(4)**, 600–612 (2004).