Deep Learning-based Noise Reduction for Fast Volume Diffusion Tensor Imaging: Assessing the Noise Reduction Effect and Reliability of Diffusion Metrics

Hajime Sagawa¹, Yasutaka Fushimi²*, Satoshi Nakajima², Koji Fujimoto², Kanae Kawai Miyake³, Hitomi Numamoto³, Koji Koizumi¹, Masahito Nambu⁴, Hiroharu Kataoka⁵, Yuji Nakamoto², and Tsuneo Saga³

To assess the feasibility of a denoising approach with deep learning-based reconstruction (dDLR) for fast volume simultaneous multi-slice diffusion tensor imaging of the brain, noise reduction effects and the reliability of diffusion metrics were evaluated with 20 patients. Image noise was significantly decreased with dDLR. Although fractional anisotropy (FA) of deep gray matter was overestimated when the number of image acquisitions was one (NAQ1), FA in NAQ1 with dDLR became closer to that in NAQ5.

Keywords: denoising approach with deep learning-based reconstruction, diffusion tensor imaging, diffusion tensor tractography, number of image acquisition

Introduction

Diffusion-weighted imaging (DWI) is widely used in clinical MRI. DWI provides information on tissue microstructure based on measuring the principal random motion of water molecules. With its unique capability to characterize water mobility through tissues, DWI can help to detect subtle lesions, characterize tissue properties, refine the accuracy of diagnosis, and monitor the course of diseases over time. Diffusion tensor imaging (DTI) is based on measuring the diffusion of water molecules along 6 or more gradient directions.¹ By characterizing the preferential direction of water diffusion, DTI can provide voxel-wise information such as fractional anisotropy (FA) and the direction of diffusion.

Generally, DTI requires a long scan time because of the measurement in at least 6 noncollinear directions. Scan time is further extended by increasing the number of diffusion encodings. In addition, high-resolution DWI, such as isotropic imaging (i.e., thinner slice thickness with increased number of slices) suffers from a prolonged TR and low SNR, which necessitates a long acquisition time.² Recently, simultaneous multi-slice (SMS) imaging, a technique to excite multiple slices with a single RF pulse, has enabled the acquisition of several slices simultaneously³ and can considerably shorten acquisition times for DTI. Decreases in SNR can be recovered by various techniques, from simply averaging repeated scans to a more sophisticated model-based reconstructions.⁴

Recently, as an alternative approach to reduced SNR in the MRI, deep learning approaches for image noise reduction have been reported. Kidoh et al.⁵ presented the performance of a denoising approach with deep learning-based reconstruction (dDLR), which reduced image noise while preserving image quality for brain MR images. However, few reports to date have investigated the impact of the deep learning technique on DWI or DTI.⁶,⁷ We have optimized the fast volume SMS-DTI with the number of image acquisitions (NAQ) set at one (NAQ1) and a scan time of about 1 min, and we applied clinically available dDLR for SMS-DTI (dDLR-NAQ1). In the present study, using scan data with 5 averages as a reference, the feasibility and accuracy of fast SMS-DTI with dDLR for the brain were evaluated.
Materials and Methods

Subjects
The present prospective study was approved by the Institutional Review Board and was registered to the UMIN Clinical Trials Registry as UMIN000036700. This study was performed in accordance with the ethical standards of the Declaration of Helsinki (as revised in Fortaleza, Brazil, October 2013). Adult patients who underwent clinical MRI of the brain between October and December 2019 were enrolled. All patients provided written informed consent to participate in this study. Patients were excluded when MRI showed poor image quality due to body movement during MRI data acquisition.

MRI acquisition
All scans were performed on a 3T whole-body scanner (Vantage Galan 3T / ZGO; Canon Medical Systems, Tochigi, Japan) with a 32-channel head coil. DTI was performed using a single-shot spin-echo echo-planar imaging (EPI) sequence with 12 different gradient directions and a b-value of 1000 s/mm². The acceleration factors of in-plane parallel imaging (SPEEDER; Canon Medical Systems) and SMS (Multi-band SPEEDER) were 3 and 2, respectively. Other scan parameters were as follows: TR, 4000 ms; TE, 61 ms; FOV, 220 × 220 mm²; matrix size, 128 × 160; slice thickness, 2 mm; interslice gap, 0 mm; and bandwidth, 1302 Hz/pixel.

To evaluate the denoising effect of dDLR, we acquired images with a single image acquisition (NAQ1) for comparison data, and also acquired 5 NAQ images (NAQ5) for ground-truth data. Acquisition times were 1 min 5 s for NAQ1 and 5 min 45 s for NAQ5. The vendor-supplied dDLR algorithm was applied to b = 0 and 1000 s/mm² images acquired with NAQ1 (dDLR-NAQ1). The denoising level was determined at which the delineation of the basal ganglia and the contrast between the cerebral cortex and white matter on b = 1,000 images were sufficiently preserved. The same denoising level was applied across all patients.

Image analysis

ROI analysis
From DTI, isotropic DWI, apparent diffusion coefficient (ADC) maps and grayscale FA maps as well as color-coded FA maps that contain 3D information on the voxel-wise diffusion orientation were obtained.

Oval ROIs were placed in the following areas: (1) the corpus callosum (genu, splenium); (2) deep white matter (frontal, parietal, occipital); (3) periventricular white matter (parietal, occipital); (4) deep gray matter (GM) (putamen); and (5) cortical GM (posterior parietal, occipital), as described in a previous article.³ ROIs were initially placed on NAQ5 images, then transferred onto NAQ1 and dDLR-NAQ1 images by a radiological technologist with 6 years of experience and approved by a radiologist with 21 years of experience. Mean and standard deviations (SDs) of signal intensity (SI) for isotropic DWI, mean ADC, and mean FA values in each ROI were measured. There were no cases in which lesions were included within the ROIs.

The SNR of isotropic DWI was calculated as follows:

\[ \text{SNR} = \frac{\text{SI}}{\text{SD}} \]  

Fiber-tracking analysis
Diffusion tensor tractography (DTT) of the bilateral pyramidal tract was performed using Ziosation 2 (Ziosoft, Tokyo, Japan), a 3D workstation. Seed ROIs were manually drawn on the anterior pontine area, target ROIs were drawn on the cerebral peduncle and the posterior limb of the internal capsule by a radiological technologist with 11 years of experience. Tracking was initiated at an FA value of 0.2 and was terminated when FA fell to < 0.2 or the angle between two adjacent eigenvectors was > 40°. Fiber volume (VF), a volume of voxels occupied by all streamlines for the pyramidal tract, was measured and compared among NAQ1, NAQ5, and dDLR-NAQ1.

Statistical analysis
All statistical analyses were performed using JMP14 software (SAS Institute, Cary, NC, USA). Normality was examined using the Shapiro–Wilk test. Mean SI, SD, SNR of isotropic DWI, mean ADC, mean FA values, and VF calculated from NAQ1, NAQ5, and dDLR-NAQ1 images were compared using the Steel–Dwass test. All P-values were two-sided and values of P < 0.05 were considered statistically significant.

Correlations of mean ADC and mean FA values between NAQ1 and NAQ5, dDLR-NAQ1 and NAQ5 were assessed by determining the Spearman’s rank correlation coefficient (ρ) for nonparametric correlations.

Results
A total of 20 patients (9 men, 11 women; mean age, 66.5 ± 12.4 years) were enrolled in this study. The purposes of MRI examination were: screening for brain lesions in 4 patients, searching for brain metastasis in 5 patients, follow-up of brain metastasis after treatment in 3 patients, follow-up of benign brain tumors in 5 patients, follow-up of cerebral aneurysm in 2 patients, and follow-up after operation for chronic subdural hematoma in 1 patient. No active lesions such as growing tumors or acute infarction/bleeding that may have affected the present image analysis were detected in these 20 patients, and no patients were excluded from the following analyses.

Figure 1 shows representative images for isotropic DWI, ADC maps, and FA maps calculated from NAQ1, NAQ5, and dDLR-NAQ1 images.
Fig. 1 A 45-year-old woman. Representative isotropic DWI images, ADC maps, and FA maps reconstructed from NAQ1, NAQ5, and dDLR-NAQ1 images. Images from NAQ1 show increased noise, particularly in deep brain regions (left). Noise in dDLR-NAQ1 images is significantly decreased compared to that in NAQ1 images (right). ADC, apparent diffusion coefficient; dDLR, denoising approach with deep learning-based reconstruction; DWI, diffusion-weighted imaging; FA, fractional anisotropy; NAQ, number of image acquisitions.

ROI analysis
No significant difference in mean signal intensity of isotropic DWI was seen between the 3 groups in any regions (Fig. 2). SD of dDLR-NAQ1 was significantly reduced in the corpus callosum, white matter (WM) and deep GM compared to that of NAQ1 (P < 0.001). In deep GM, the SD of dDLR-NAQ1 was significantly lower than that of NAQ5 (P < 0.001). The SNR of isotropic DWI of NAQ5 was significantly higher in the corpus callosum and WM compared to NAQ1 (P < 0.001). The SNR of dDLR-NAQ1 was significantly increased not only in the corpus callosum and WM, but also in the deep GM compared to that of NAQ1 (P < 0.001). Furthermore, in deep GM, the SNR of dDLR-NAQ1 was significantly higher than that of NAQ5 (P < 0.001).

No significant difference in ADC was seen between the 3 groups in any regions (Fig. 3a). No significant difference in FA values were identified between the 3 groups, with the exception of deep GM. In deep GM, significant differences were noted between NAQ1 and NAQ5 (P < 0.001), NAQ1 and dDLR-NAQ1 (P < 0.001), NAQ5 and dDLR-NAQ1.
Fig. 3 ADC and FA values for different brain regions calculated from NAQ1, NAQ5, and dDLR-NAQ1 images. Note that asterisks represent statistically significance ($P < 0.05$). ADC, apparent diffusion coefficient; dDLR, denoising approach with deep learning-based reconstruction; FA, fractional anisotropy; GM, gray matter; NAQ, number of image acquisitions; WM, white matter.

$P = 0.045$). FA values were the highest in the order of NAQ1, dDLR-NAQ1, and NAQ5 (Fig. 3b).

Table 1 shows Spearman’s rank correlation coefficients for ADC and FA values between NAQ1 and NAQ5, and between dDLR-NAQ1 and NAQ5. ADCs showed strong correlations between NAQ1 and NAQ5 and between dDLR-NAQ1 and NAQ5 ($P < 0.001$). In FA values, a weak correlation was observed between dDLR-NAQ1 and NAQ5 in deep GM ($\rho = 0.38$, $P = 0.016$), but no correlation was observed between NAQ1 and NAQ5 ($\rho = 0.23$, $P = 0.15$). Strong correlations were observed between NAQ1 and NAQ5 and between dDLR-NAQ1 and NAQ5 in ROIs other than deep GM ($P < 0.001$).

**Fiber-tracking analysis**

Figure 4 shows representative DTT images of the pyramidal tracts. DTT of NAQ1 was improved with dDLR. FV of the pyramidal tracts (mean ± SD $\times 10^3$ mm$^3$) of NAQ1, NAQ5, and dDLR-NAQ1 were 26.0 ± 7.7, 49.5 ± 15.6, and 44.9 ± 11.8, respectively. FV of NAQ5 and dDLR-NAQ1
Table 1 Spearman’s rank correlation of ADC and FA values between NAQ1 and NAQ5, and between dDLR-NAQ1 and NAQ5

| Region          | ADC | FA |
|-----------------|-----|----|
|                 | ρ   | P  |
|                 | ρ   | P  |
| Corpus callosum |     |    |
| NAQ1 vs. NAQ5   | 0.66 | < 0.001 | 0.81 | < 0.001 |
| dDLR-NAQ1 vs. NAQ5 | 0.77 | < 0.001 | 0.84 | < 0.001 |
| Deep WM         |     |    |
| NAQ1 vs. NAQ5   | 0.69 | < 0.001 | 0.94 | < 0.001 |
| dDLR-NAQ1 vs. NAQ5 | 0.76 | < 0.001 | 0.95 | < 0.001 |
| Periventricular WM |     |    |
| NAQ1 vs. NAQ5   | 0.88 | < 0.001 | 0.93 | < 0.001 |
| dDLR-NAQ1 vs. NAQ5 | 0.88 | < 0.001 | 0.94 | < 0.001 |
| Deep GM         |     |    |
| NAQ1 vs. NAQ5   | 0.75 | < 0.001 | 0.23 | 0.15  |
| dDLR-NAQ1 vs. NAQ5 | 0.78 | < 0.001 | 0.38 | 0.016 |
| Cortical GM     |     |    |
| NAQ1 vs. NAQ5   | 0.82 | < 0.001 | 0.79 | < 0.001 |
| dDLR-NAQ1 vs. NAQ5 | 0.78 | < 0.001 | 0.74 | < 0.001 |

ADC, apparent diffusion coefficient; dDLR, denoising approach with deep learning-based reconstruction; FA, fractional anisotropy; GM, gray matter; NAQ, number of image acquisitions; WM, white matter.

Comparison between NAQ1 and dDLR-NAQ1 showed that the SI of isotropic DWI was unchanged, and SD was significantly decreased in all regions except for cortical GM with dDLR to the fast volume SMS-DTI acquisition (NAQ1, 1 min). SNR of dDLR-NAQ1 became equivalent to or better than that of NAQ5. In deep GM, the SD and SNR of dDLR-NAQ1 were significantly better than those of NAQ5. This was because the receiver coil used in this study was a 32-channel phased array coil, and image noise was more severe in deep brain regions. Moreover, T2 value of putamen is shorter than that of the other regions, therefore, the degree of SNR improvement was considered small even with NAQ5 in deep GM. In contrast, since a strong de-noising effect was obtained by dDLR even in low-SNR regions, the SD and SNR of dDLR-NAQ1 seemed significantly different from those of NAQ5. On the other hand, since cortical GM is a region close to the receiver coil, there seemed to be no difference in SD and SNR among the 3 groups.

In this study, ADCs did not differ significantly among the 3 groups in any regions. This was consistent with a report

Fig. 4 A 59-year-old man. Representative DTT of pyramidal tracts reconstructed from NAQ1, NAQ5, and dDLR-NAQ1 images. The depiction ability of DTT was significantly improved with dDLR-NAQ1 compared to NAQ1. DTT, diffusion tensor tractography; dDLR, denoising approach with deep learning-based reconstruction; NAQ, number of image acquisitions.
by Farrell et al.,9 which found no bias trend at low SNR for ADC. Regarding FA values, no significant difference was seen among the 3 groups in corpus callosum or WM, while a significant difference was identified in deep GM. FA values were reportedly overestimated in GM due to increases in $\lambda_1$ and decreases in $\lambda_3$ at low SNR.9 In the present study, FA values did not differ significantly among the three groups in cortical GM, but showed significant differences in deep GM. This was attributed to differences in image noise between the cortical and deep GM. As a result, FA values in deep GM were more affected at low SNR than those in cortical GM. By reducing image noise using dDLR, the FA value of dDLR-NAQ1 in deep GM became much closer to that of NAQ5 compared to that of NAQ1, and correlation between dDLR-NAQ1 and NAQ5 was stronger than that between NAQ1 and NAQ5. However, the FA value of dDLR-NAQ1 was significantly higher than that of NAQ5. Considering relatively low SNR of the DWI at the deep GM, sufficient SNR may not have been achieved to calculate true FA at deep GM in NAQ1. To calculate accurate FA values for deep GM, further studies with more NAQs may be necessary. Nevertheless, FA value of dDLR-NAQ1 was closer to that of NAQ5, compared to that of NAQ1. This might be interpreted that dDLR not only reduced the image noise but, more importantly, improved the reliability of FA values. 

A number of approaches have been developed to reduce noise for DTI, e.g. Bayesian method for regularizing the diffusion tensor field,10 transferring the symmetric and positive-definite tensor into Riemannian space for regularization,11 non-local means variants.12 However, these methods were not installed in the commercial software as far, and dDLR algorithm used in this study has developed as a first commercial product that incorporates deep convolutional neural networks restoration process into the reconstruction flow to reduce image noise. Details of dDLR algorithm have been reported in the previous study,5 in which denoising of the high frequency components using the discrete cosine transform layer, which separates the high-frequency components from the zero-frequency component, allow for efficient noise reduction. The use of dDLR removes image noise by learning various noise characteristics using training pairs of different noise level images and the corresponding ground-truth images.5 However, in images with insufficient image quality and/or SNR, dDLR may not work effectively. Therefore, the denoising level has to be determined in consideration of excessive image noise, blurring of anatomical structural boundaries, and artificial texture of the images. In this study, the denoising level of dDLR-NAQ1 was determined at which the delineation of the basal ganglia and the contrast between the cerebral cortex and white matter were sufficiently preserved as with Kidoh et al.5 

Diffusion tensor imaging offers an effective tool for delineating the effects of a tumor on nearby white matter tracts, providing information that may facilitate preoperative planning.13 Measurements of diffusion anisotropy tend to be quite sensitive to image noise, which can also lead to biases in anisotropy estimates.9 The accuracy of DTI measurements may be improved by either increasing the number of encoding directions or increasing the number of acquisitions, increasing the scan time for DTI data collection.14 Therefore, clinically, performing DTI with as many diffusion encoding directions as possible with minimum averaging is desirable. The dDLR approach drastically decreased image noise and generated thin-slice MR images of the brain with sufficient image quality to precisely evaluate fine anatomical details in a relatively short acquisition time.5 Likewise, in this study, dDLR was able to effectively remove the image noise of DWI. By applying dDLR to this short-acquisition DTI combining SMS and minimum averaging, this imaging method can potentially be incorporated into routine protocols as a thin-slice DTI, including acute cerebral infarction screening. 

This study had several limitations. First, single-shot EPI without SMS was not used as the control due to long scan time for the present clinical study, therefore, NAQ5 with SMS was used instead as the control in this study. However, SMS does not cause a decreased SNR theoretically, and the SMS factor of 2 used in this study was not so high. Moreover, no signal leakage15 were apparently observed in all cases. Rician noise introduces a bias into MRI measurements that can have a significant impact on the shapes and orientations of tensors in diffusion tensor MR images.9 Restoration methods such as DWI filtering may be required for DWI with decreased SNR, especially in DWI with higher $b$-values and/or higher-resolution imaging.16 Second, mean value within the ROI may average the effects of noise, therefore, SI, SD, SNR, ADCs, and FA values should have been compared pixel-by-pixel. However, because NAQ1 and NAQ5 images were acquired separately, ROI analysis was performed instead of pixel-by-pixel analysis to avoid the influence of patient’s motion. Third limitation of this study was the small size of the patient cohort. Several studies have shown the utility of DTI parameters in the characterization and grading of brain tumors.17 Further studies for comparing DTI parameters with and without dDLR are desirable, including patients with various neurological diseases. However, the ability of DTT to depict the pyramidal tracts was significantly improved in dDLR-NAQ1 when compared with NAQ1, and no significant difference in FV was apparent between dDLR-NAQ1 and NAQ5. As for NAQ1, more tracking points were terminated due to image noise, on the contrary, visualization of DTI was improved in dDLR-NAQ1 due to denoising with dDLR. Thus, single averaging short-time DTI with dDLR can be used in depicting white matter fibers for preoperative planning. 

**Conclusion**

A deep learning-based noise reduction technique dramatically not only reduces the image noise of fast volume SMS-DTI, but improves the reliability of ADCs and FA values. Furthermore, the ability to depict DTT is also significantly improved with dDLR.
Acknowledgment
This work was supported by an Industry-Academia Collaboration Project, “Research on the development of next-generation medical imaging (No. 151190700001)” provided by Canon Medical Systems Corporation.

Conflicts of Interest
Tsuneo Saga, Kanae Kawai Miyake, and Hitomi Numamoto are the endowed chair of Industry-Academia Collaboration Project between Kyoto University and Canon Medical Systems Corporation. Masahito Nambu is an employee of Canon Medical Systems Corporation. The remaining authors have no conflicts of interest.

References
1. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994; 66:259–267.
2. Dong Z, Dai E, Wang F, et al. Model-based reconstruction for simultaneous multislice and parallel imaging accelerated multishot diffusion tensor imaging. Med Phys 2018; 45:3196–3204.
3. Barth M, Breuer F, Koopmans PJ, Norris DG, Poser BA. Simultaneous multislice (SMS) imaging techniques. Magn Reson Med 2016; 75:63–81.
4. Knoll F, Raya JG, Halloran RO, et al. A model-based reconstruction for undersampled radial spin-echo DTI with variational penalties on the diffusion tensor. NMR Biomed 2015; 28:353–366.
5. Kidoh M, Shinoda K, Kitajima M, et al. Deep learning based noise reduction for brain MR imaging: tests on phantoms and healthy volunteers. Magn Reson Med Sci 2020; 19:195–206.
6. Nath V, Schilling KG, Parvathaneni P, et al. Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI. Magn Reson Imaging 2019; 62:220–227.
7. Lin Z, Gong T, Wang K, et al. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. Med Phys 2019; 46:3101–3116.
8. Fox RJ, Sakaie K, Lee JC, et al. A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. AJNR Am J Neuroradiol 2012; 33:695–700.
9. Farrell JA, Landman BA, Jones CK, et al. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. J Magn Reson Imaging 2007; 26:756–767.
10. Frandsen J, Hobolth A, Ostergaard L, Vestergaard-Poulsen P, Vedel Jensen EB. Bayesian regularization of diffusion tensor images. Biostatistics 2007; 8:784–799.
11. Gur Y, Sochen N. Fast invariant riemannian DT-MRI regularization. IEEE 11th International Conference on Computer Vision, IEEE, Rio de Janeiro, Brazil, 2007; 2449–2455.
12. Wiest-Daesslé N, Prima S, Coupé P, Morrissey SP, Barillot C. Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. Med Image Comput Comput Assist Interv 2007; 10:344–351.
13. Bello L, Castellano A, Fava E, et al. Intraoperative use of diffusion tensor imaging fiber tractography and subcortical mapping for resection of gliomas: technical considerations. Neurosurg Focus 2010; 28:E6.
14. Alexander AL, Lee JE, Lazar M, Field AS. Diffusion tensor imaging of the brain. Neurotherapeutics 2007; 4:316–329.
15. McNabb CB, Lindner M, Shen S, Burgess LG, Murayama K, Johnstone T. Inter-slice leakage and intra-slice aliasing in simultaneous multi-slice echo-planar images. Brain Struct Funct 2020; 225:1153–1158.
16. McGraw T, Yemuri BC, Chen Y, Rao M, Mareci T. DT-MRI denoising and neuronal fiber tracking. Med Image Anal 2004; 8:95–111.
17. Svolos P, Kousi E, Kapsalaki E, et al. The role of diffusion and perfusion weighted imaging in the differential diagnosis of cerebral tumors: a review and future perspectives. Cancer Imaging 2014; 14:20.