q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans

Vladimir Golkov*, Alexey Dosovitskiy, Jonathan I. Sperl, Marion I. Menzel,
Michael Czisch, Philipp Sämann, Thomas Brox, and Daniel Cremers

Abstract—Numerous scientific fields rely on elaborate but partly suboptimal data processing pipelines. An example is diffusion magnetic resonance imaging (diffusion MRI), a non-invasive microstructure assessment method with a prominent application in neuroimaging. Advanced diffusion models providing accurate microstructural characterization so far have required long acquisition times and thus have been inapplicable for children and adults who are uncooperative, uncomfortable, or unwell. We show that the long scan time requirements are mainly due to disadvantages of classical data processing. We demonstrate how deep learning, a group of algorithms based on recent advances in the field of artificial neural networks, can be applied to reduce diffusion MRI data processing to a single optimized step. This modification allows obtaining scalar measures from advanced models at twelve-fold reduced scan time and detecting abnormalities without using diffusion models. We set a new state of the art by estimating diffusion kurtosis measures from only 12 data points and neurite orientation dispersion and density measures from only 8 data points. This allows unprecedentedly fast and robust protocols facilitating clinical routine and demonstrates how classical data processing can be streamlined by means of deep learning.

Index Terms—Artificial neural networks, diffusion kurtosis imaging (DKI), diffusion magnetic resonance imaging (diffusion MRI), neurite orientation dispersion and density imaging (NODDI).

I. INTRODUCTION

OVER the past three decades, diffusion magnetic resonance imaging (diffusion MRI) [1]–[4] has taken on an important role in assessing microstructural tissue and material properties non-invasively based on the diffusion of gases and liquids, primarily water. In radiology, diffusion MRI is a powerful technique, mainly due to its sensitivity to diffusion restriction (e.g., caused by brain ischemia), yet also any other

microstructural tissue rebuilding as found in neoplasms or inflammatory lesions. Its potential as a basis for diagnostic and treatment monitoring markers has been established over the last years [5]–[8]. Advanced diffusion MRI models such as diffusion kurtosis imaging [2], [3] (DKI) and neurite orientation dispersion and density imaging [4] (NODDI) provide more accurate characterization of tissue microstructure [2], [4], [9]–[11] but require long acquisition time. This has so far led to high scan costs and has made advanced diffusion models inapplicable for patients who are uncooperative, uncomfortable or unwell.

A. Model Fitting, Analytical Solutions, Approximation

In diffusion MRI, a number of diffusion-weighted images (DWIs) for different diffusion weightings[1] and directions (constituting the so-called three-dimensional q-space) are acquired [1]. Signal intensity in these images contains information regarding diffusion properties. The task in quantitative diffusion MRI is to find a mapping from a limited number of noisy signal samples to rotationally invariant scalar measures that quantify microstructural tissue properties. This inverse problem is solved in each image voxel. Currently, this problem is addressed by three approaches.

The classical approach of estimating scalar measures is model fitting. Its data processing pipeline consists of fitting [12] a diffusion model and calculating rotationally invariant measures from the fitted model parameters. Prior to model fitting, the q-space data can be obtained by regular acquisition, or using advanced methods such as compressed sensing or dictionary learning (cf. below).

Another approach can be taken if closed-form analytical solutions exist. For the diffusion model of DKI [2], [3]–which requires approximately 150 DWIs [3], [13], [14]–it has recently been shown [15], [16] that for certain DKI-based measures much fewer DWIs (e.g., 13 or 19 DWIs) are sufficient, and that these measures can be analytically calculated from the data in a single step. This has led us to the assumption that for many other scalar measures and tissue properties the most relevant information might as well be recovered from only a few DWIs.

The third approach of calculating scalar measures is approximation, particularly machine learning. Simulations of simplified tissue models with extensive sets of diffusion weightings [17], [18] indicate that standard model fitting procedures can be replaced by approximation methods. It was also mentioned

[1]For simplicity, we also include images with diffusion weighting zero into the definition of "DWIs".

Fig. 1. Possible steps of data processing from scanning a real-world subject (left) to the determination of the tissue properties (right). Standard DKI/NODDI pipeline is shown individually (a) and in comparison to advanced methods (b). Arrows designate possible data processing in the standard pipelines (solid red), state-of-the-art methods based on compressed sensing and machine learning (dash-dotted black) and novel processing possibilities introduced with q-space deep learning (dashed green), see also [24].

[18] that feature selection methods could be applied to identify the most relevant DWIs in order to reduce these extensive sets of diffusion weightings. On the basis of these observations, we apply deep learning [19]–[23] for accurate approximation and present a deep learning framework for different inputs (full and subsampled sets of regular DWIs, non-diffusion contrasts) and different outputs (denoising, missing DWI reconstruction, scalar measure estimation, tissue segmentation). Scalar measure estimation from twelve-fold accelerated acquisition is demonstrated on two advanced models: DKI [2] (using radial kurtosis and fractional kurtosis anisotropy) and NODDI [4] (using orientation dispersion index and intracellular volume fraction). In comparison to most of the well-established models (e.g., diffusion tensor imaging [1]), DKI and NODDI are more elaborate and thus can provide improved sensitivity [2], [4], [9]–[11]; however, they also require considerably longer acquisition times. By shortening the acquisition duration of advanced models by an order of magnitude, we strongly improve their potential for clinical use, and reduce scan costs and motion artifacts caused by long scan durations.

B. Advantages of Deep Learning

Deep learning [19]–[23] is a family of algorithms for efficient learning of complicated dependencies between input data and outputs by propagating a training dataset through several layers of hidden units (artificial neurons). Each layer is a data transformation step. The classical diffusion MRI pipeline involving model fitting also consists of several steps. In the example of DKI, approximately 150 measurements [3], [13], [14] are reduced to 22 model parameters in the classical pipeline,

then to a few rotationally invariant measures, and finally (implicitly or explicitly) to one parameter, i.e., the tissue property of interest such as the amount of disease-based microstructural change. (For NODDI, rotationally invariant measures are estimated during model fitting rather than in an additional step, see Fig. 1(a). In every step, information is partly lost by reducing the degrees of freedom. However, the classical pipeline does not provide feedback from the later steps to the earlier steps with regard to what part of the information should be retained or discarded and which transformations should be applied. Thus, the pipeline relies on handcrafting and fixing each step, i.e., the diffusion model and derived scalar measures. Deep learning takes a more flexible approach: the effects of each layer on the final result are propagated back to adjust preceding layers, such that all layers are optimized jointly in terms of the final objective, namely minimizing the output error. This prevents the loss of information during intermediate steps. Advantages of deep learning over handcrafted features have been shown in numerous other applications [23].

The main novelties introduced herein are:
- Using subsampled DWIs as machine learning input *directly*,
- Unprecedented scan time reduction for DKI and NODDI,
- Segmentation without using diffusion models.

Preliminary results presented at a conference [24] are herein extended by additional evaluation, including the influence of neural network parameters, and more[2]. Related applications of

---

[2]This paper has supplementary downloadable material available at http://iee-explore.ieee.org, provided by the authors. This includes additional methods and experiments.

machine learning are tractography [25] and non-diffusion MRI [26].

II. MATERIALS AND METHODS

The relationship between the diffusion-weighted signal and microstructural tissue properties is non-trivial. However, an appropriately chosen, tuned and trained machine learning algorithm can theoretically represent any relationship between inputs and outputs [27] if such a relationship exists. We make use of this fact in order to leverage information contained in very limited numbers of input DWIs. In all experiments presented in this work, training datasets originate from a different human subject than the test datasets. The proposed family of methods is termed "q-space deep learning" (q-DL). In q-DL, we treat each image voxel individually as a data sample.

The task of estimating the vector m of scalar measures from the vector S of signal measurements can be formalized as follows. The analytical solution is as simple as calculating $H(S)$, where H is the closed-form function that maps S to m. Such closed-form solutions are available only for certain measures and certain diffusion weightings [15], [16]. In model fitting, m is estimated as $g(f(S))$, where $\theta = f(S)$ are the estimated diffusion model parameters obtained through model fitting f by solving an optimization problem, e.g., least squares [12], and g calculates rotationally invariant scalar measures from θ. In DKI, the steps of applying f and g are independent and not optimized jointly with respect to the accuracy of estimation of m; in NODDI, f and g are one joint step; in all cases, fitting is susceptible to noise. In contrast, q-DL adjusts the parameters of a multilayer neural network such that the outputs of the network well approximate the target measures m. The measures m are obtained for the training dataset by model fitting, but model fitting is not required for the datasets to which the trained network is subsequently applied.

A. Feed-Forward Neural Networks

A so-called multilayer perceptron is a multilayer artificial neural network that performs a nonlinear data transformation in each layer. Layer 0 is called the input layer, layer L the output layer, intermediate layers are called hidden layers. The transformation in layer $i \in \{1, \ldots, L\}$ follows the rule

$$a_j^{(i)} = s_i \left(W^{(i)} a_j^{(i-1)} + b^{(i)} \right), \tag{1}$$

where $a_j^{(i)}$ is the output vector of layer i for data sample j, the vector $a_j^{(0)}$ is the input of the network, $W^{(i)}$ is called the weight matrix, $b^{(i)}$ the bias vector, and s_i are nonlinearities (see below). The length of the vector $a_j^{(i)}$ corresponds to the number of artificial neurons (hidden units) in layer i. During training, all weight matrices and bias terms are jointly adjusted such that the output vectors $a_j^{(L)}$ for each training sample j (in our case: each image voxel j) well approximate the target output vectors y_j. This adjustment is achieved by using the backpropagation algorithm (implemented in the deep learning toolbox [28]) to solve the optimization problem

$$\arg\min_{W,b} \sum_j \|a_j^{(L)} - y_j\|^2, \tag{2}$$

where the sum of errors is taken over all training samples j, and the outputs $a_j^{(L)}$ recursively depend on the parameters $W^{(i)}$ and $b^{(i)}$ according to the aforementioned recursive transformation rule for the $a_j^{(i)}$ for $i \in \{1, \ldots, L\}$. Once trained, such a neural network works in a deterministic manner.

B. q-Space Deep Learning

The proposed pipelines based on q-space deep learning reduce scan duration and perform the data processing as directly as possible without discarding information at intermediate steps. This is reflected in the comparison of q-DL to the standard pipeline and to other state-of-the-art methods in terms of possible steps of data processing (Fig. 1). Previous methods based on machine learning rely either on extensive acquisitions or on intermediate steps involving model fitting based on diffusion tensor imaging (DTI) and spherical harmonics (SH), whereas q-space deep learning provides the fastest acquisitions and the most direct data processing steps.

In all experiments, training data originate from different human subjects than test data. The neural networks thus do not "know" the true output vectors of the test data but rather estimate them based on the input-output-mapping learned from training data. Each voxel j is treated individually as a data sample. The algorithm does not know its location in the image. We introduce several input-output-mapping tasks. Different deep networks are trained for different tasks:

1) Estimation of Scalar Measures: A network is trained to predict microstructure-characterizing scalar measures m_j directly from the (reduced set of) DWIs $S_{j,\alpha}$ where α is a pseudorandom subsampling multi-index (such that the q-space sampling is consistent across training and test data). In other words, inputs are $a_j^{(0)} = S_{j,\alpha}$ with length $|\alpha|$, and targets are $y_j = m_j$. The length of the output vector is the number of considered scalar measures. Training targets $y_j = m_j$ are obtained from a fully sampled training dataset S_j (consisting of $|S_j| = n$ DWIs) by model fitting; however, a neural network is trained to predict m_j from the subsampled data $S_{j,\alpha}$. As a consequence, the neural network is able to estimate m_j from α-subsampled datasets. This allows an estimation of m_j at a scan time reduction factor of $n/|\alpha|$ for all subsequent datasets. In our experiments, we use scan time reduction factors of up to $n/|\alpha| = 148/12 \approx 12.3$ for DKI and up to $n/|\alpha| = 99/8 \approx 12.4$ for NODDI.

2) Model-Free Segmentation: Tissue segmentation is achieved by training a neural network to discriminate between several tissue types. We propose modifying the approach [29] of multi-parametric MRI tissue characterization by artificial neural networks such that the DWIs are directly used as inputs rather than using scalar measures obtained from model fitting. Our approach thus allows using the unique information provided by diffusion MRI directly without the information reduction imposed by models. State-of-the-art automatic segmentation [30], [31] (based on non-diffusion images with spatial priors) into healthy white matter (WM), grey matter (GM), cerebrospinal fluid (CSF) and multiple sclerosis lesions was used as ground truth for our proof-of-concept model-free segmentation (based on diffusion images without spatial priors). The q-DL framework allows incorporating additional contrasts other than DWIs as inputs to the learning algorithm. We used fluid-attenuated
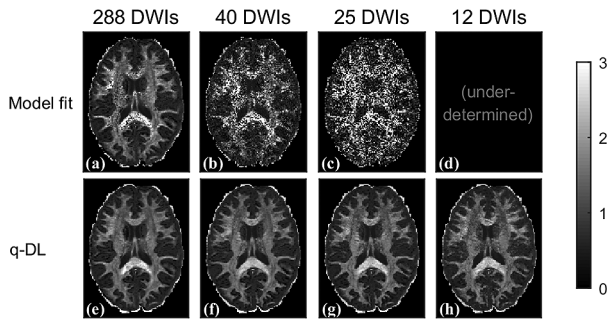
Fig. 2. Maps of radial kurtosis in the human brain for various methods and MRI scan acceleration factors. From left to right: 288, 40, 25 and 12 randomly selected DWIs are used. Model fitting followed by radial kurtosis calculation (a-d), and q-DL for radial kurtosis approximation (e-h) are compared. Model fitting is outperformed by the proposed method.

inversion recovery (FLAIR) signal as an additional input. The length of the output vector is the number of tissue classes (with each output representing a relative class membership "likeliness" using softmax, see below).

C. Details of the Neural Networks

The deep learning toolbox [28] was used for deep learning experiments. The artificial neural network used is a multi-layer perceptron with three hidden layers, each consisting of 150 hidden units with a nonlinearity known as the rectified linear unit [19], [20], i.e., $s_i(z) = \max(0, z)$. This layout, applied to each image voxel independently, can be considered a convolutional neural network with window size 1×1 in each layer, masking out the loss for non-brain voxels. Linear units $s_L(z) = z$ are used in the output layer L for fitting tasks and softmax outputs $s_L(z) = \exp(z)/\|\exp(z)\|_1$ for classification tasks. Each input and output of the neural network is independently scaled to the interval $[0, 1]$ and the same affine transformation parameters are reused for the test datasets. The network is initialized with orthogonal random weights [22]. We use a dropout [21] fraction of 0.1, stochastic gradient descent with momentum 0.9, minibatch size 128, learning rate 0.01 with a warm-up learning rate of 0.001 for the first 10 epochs. The learning rate was decreased by factor 0.9 whenever the training set error stagnated (averaged over 5 epochs) compared to the previous 5 epochs. To prevent overfitting, 10% of the voxels in the training data set were used as a validation set and early stopping was employed when the validation set error (averaged over 10 epochs) increased compared to the average over the previous 10 epochs. These choices of the neural network parameters are based on practical considerations as described in [32]. We use a multilayer perceptron because it is a straightforward and powerful method. Three hidden layers provide acceptable results and short runtime for our purposes. Other network settings are evaluated in Fig. 6. In all experiments, training data originate from different human subjects than test data (except Fig. 12, panels (e,k,q,v)). For different q-space sampling schemes, the values of the network inputs (signal intensities) have a different meaning (and length), therefore a different network must be trained independently for every q-space scheme.
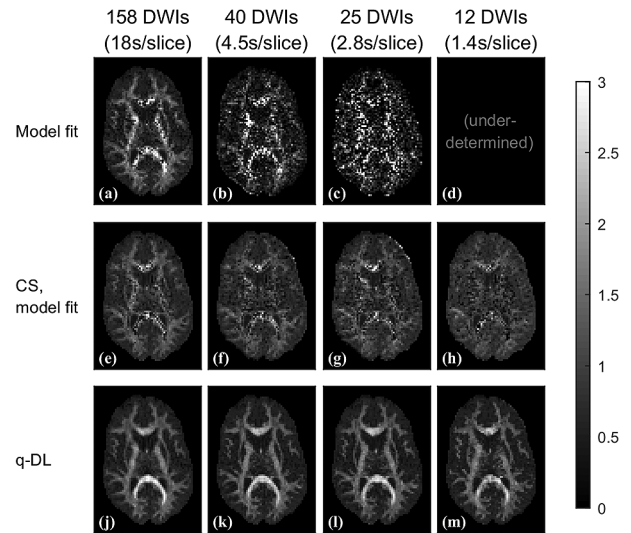

Fig. 3. Same as Fig. 2 (different scanner, different volunteer), including a comparison to compressed sensing (e-h). Required scan time for each sampling scheme is shown in seconds per slice. Model fitting and compressed sensing are outperformed by the proposed methods.
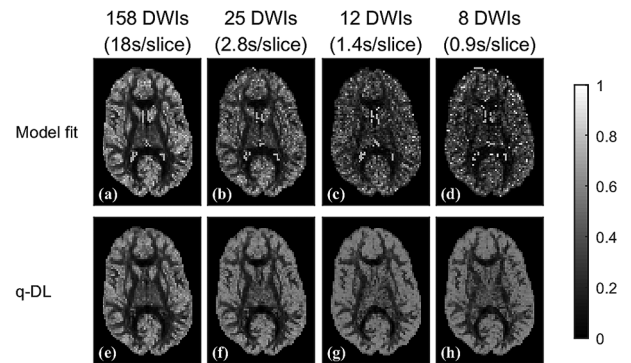

Fig. 4. Same as Fig. 2 for neurite orientation dispersion index based on NODDI. The proposed method better preserves contrast at short scan times.
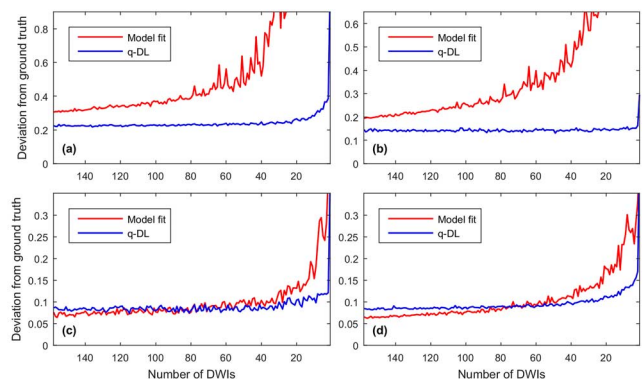

Fig. 5. Root-mean-squared error for different methods and different numbers of DWIs; estimation of radial kurtosis (a), fractional kurtosis anisotropy (b), intra-cellular volume fraction (c), and neurite orientation dispersion index (d); comparison of two different methods: model fitting (red) and q-DL (blue). Reference standard is model fit of nine independent repetitions, i.e., 1422 DWIs, registered to the test data. For DKI measures (a,b), model fitting is always outperformed by q-DL. For NODDI measures (c,d), model fitting is outperformed by q-DL if less than 70 DWIs are used.
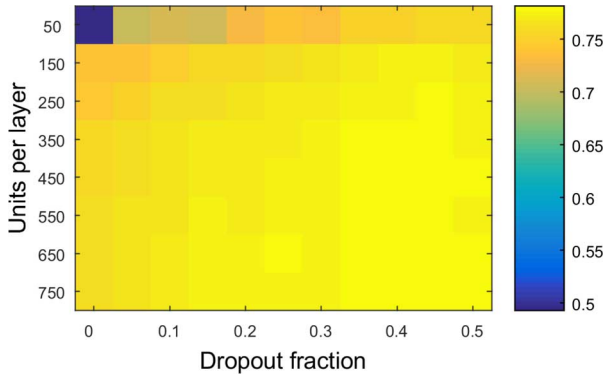
Fig. 6. Correlation of radial kurtosis estimations using different dropout fractions and layer sizes for q-DL from 12 DWIs with radial kurtosis from fully sampled (148 DWIs) model fitting.

D. Data

Approval by the local ethics board for the *in vivo* study protocols and prior informed consent were obtained. In the multiple sclerosis data, datasets from five patients were used for training, and the dataset of the respective sixth patient was used for testing (in all combinations). In all other datasets, data from one healthy volunteer was used for training, and data from another healthy volunteer for testing.

1) Five-Shell and Cartesian Healthy Volunteer Data: Data sets of a total of two healthy volunteers were acquired using the common radial q-space scheme with 30 directions sampled on five shells ($b = 600, 1200, 1800, 2400, 3000$ s/mm^2) and eight $b = 0$ images. Ten repetitions of this scheme were acquired for each volunteer. Besides, Cartesian sampling [33] (515 points, $b_{\max} = 3000$ s/mm^2) was also performed. Echo-planar imaging was performed using a 3T GE MR750 MR scanner (GE Healthcare, Waukesha, WI, USA) equipped with a 32-channel head coil ($TE = 80.7$ ms, $TR = 2$ s, $FOV = 24$ cm $\times$ 24 cm $\times$ 4 cm, isotropic voxel size 2.5 mm, ASSET factor 2). All data underwent FSL topup distortion correction [34], [35]. All DWIs were registered using an affine transformation [36] to compensate for motion. Advanced treatment of motion is subject of future work. Each volunteer data set contained approximately 40,000 brain voxels (i.e., training/test samples).

2) Three-Shell Healthy Volunteer Data: Data sets of a total of four healthy volunteers were acquired using a scheme optimized [13], [14] for DKI and suitable for NODDI [4]: three shells ($b = 750, 1070, 3000$ s/mm^2) with 25, 40, 75 directions, respectively, and eight $b = 0$ images. Acquisition parameters and postprocessing were the same as for the five-shell and Cartesian acquisitions.

3) Human Connectome Project Data: To demonstrate feasibility on a different scanner with different acquisition parameters, we used data sets of a total of two healthy volunteers from the Human Connectome Project (HCP) [37]–[44].

4) Multiple Sclerosis Data: For tissue segmentation and lesion detection, six multiple sclerosis patients were scanned using a diffusion spectrum [33] random subsampling pattern with 167 DWIs ($b_{\max} = 3000$ s/mm^2, TE = 80.3 ms, TR = 5.4 s, FOV = 24 cm $\times$ 24 cm $\times$ 12 cm, isotropic voxel size 2.5 mm, ASSET factor 2).

E. Experiments

In all experiments, training data originate from different human subjects than test data. Estimation of scalar measures based on q-DL was performed on the five-shell, three-shell and HCP data for all subsampling sizes $|\alpha|$ from n down to 8 (as well as down to 1 for error evaluation). DKI-based radial kurtosis [45] was estimated for HCP data and five-shell data. Different networks were trained for these different q-space sampling schemes. NODDI-based neurite orientation dispersion index [4] was estimated for three-shell data. State-of-the-art model fitting [4], [12] (own implementation for DKI; NODDI Matlab toolbox for NODDI) and compressed sensing (CS) for Cartesian schemes based on dictionary learning [46] (followed by model fitting) were performed for comparison because they are the currently used approaches to estimate model-based measures (CS was applied to registered Cartesian data of the same volunteer). Model fitting of one fully sampled scan was used on the training set to generate output targets for q-DL training. The quality of the methods on the test data was evaluated in terms of root-mean-squared error:

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{J}\left(\widehat{m}_j - m_{j\cdot\text{gt}}\right)^2}{J}}, \qquad (3)$$

where the sum is taken over all J voxels, the $\widehat{m}_j$ are the results being evaluated, and the model fit of the nine additional independent repetitions of the scan was used for ground truth $m_{j\text{gt}}$ ("reference standard"). The five-shell data were used for this evaluation. The fraction of voxels for which the q-DL value was close to the reference standard value was calculated for the different scalar measures. In addition to the neural network settings described above, different numbers of units per hidden layer (between 50 and 750 in steps of 100) and different dropout fractions (between 0 and 0.5 in steps of 0.05) were compared. Using the three-shell datasets of four volunteers, the influence of three different training datasets on the same test dataset was compared, with reference standard obtained from fully-sampled model fitting.

Model-free segmentation was applied to the multiple sclerosis data. State-of-the-art automatic segmentation [30], [31] into lesions, healthy WM, GM and CSF based on non-diffusion images with spatial priors (see supporting information for details) was used as ground truth for our proof-of-concept model-free segmentation including diffusion images without spatial priors. The ground truth of the training data was used as output targets during training; the ground truth of the test data was used for segmentation quality evaluation. Segmentation quality was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC). The deep learning models presented here cannot be more knowledgeable than the technique used to generate the labels.

III. RESULTS AND DISCUSSION

Figs. 2–5 compare the estimation of scalar measures produced by different methods. We show DKI-based radial kurtosis [45] of HCP data in Fig. 2 and of five-shell data in Fig. 3 (with compressed sensing (CS) [46] applied to Cartesian data of the same volunteer in Figs. 3(e)–3(h) as well

as NODDI-based neurite orientation dispersion index [4] of three-shell data in Fig. 4. State-of-the-art model fitting [4], [12] (Figs. 2(a)–2(d), 3(a)–3(d), 4(a)–4(d)), CS (Figs. 3(e)–3(h)), and q-DL (Figs. 2(e)–2(h), 3(j)–3(m), 4(e)–4(h)) are compared. Several numbers of used DWIs are compared, ranging from full sampling to 12-fold reduced scan time (scan time is shown in seconds per image slice).

Compared with the standard pipeline, results of q-DL exhibit feasibility of scan time reduction by a factor of twelve. Thus, protocols lasting about 30 minutes (Figs. 2–4 panel a) can be reduced to 2.5 minutes, strongly improving clinical feasibility.

Fig. 5 compares the methods in terms of root-mean-squared error. This represents a quantitative evaluation of the results presented in Figs. 2–4. For DKI measures, q-DL always outperforms model fitting (Figs. 5(a), 5(b)). Model fitting of 158 DWIs (error: 0.306 (Fig. 5(a)), 0.195 (Fig. 5(b))) is even outperformed by q-DL of 12 DWIs (error: 0.272 (Fig. 5(a)), 0.150 (Fig. 5(b))). For NODDI measures, q-DL outperforms model fitting when less than 70 DWIs are used (Figs. 5(c), 5(d)).

These curves demonstrate the trade-off between scan duration and quality provided by q-DL. Particularly, twelve-fold reduced scan time provides an error magnitude similar to that of model fitting at full scan time (and for DKI-based measures even lower than that of model fitting at full scan time).

For each number of subsampled DWIs, the subsampling was performed randomly and completely independently (but equally for the three compared methods). Thus, oscillations (amplitude of fluctuation) of the curves in Fig. 5 demonstrate the impact of random subsampling. Not all random subsamplings are equally useful. Among the compared methods, q-DL is most stable with respect to the choice of the samples, whereas model fitting decreases in stability (from very stable to unstable) with decreasing number of DWIs. Analogous variation was observed for repetitions of random subsampling instantiations when the number of DWIs was held constant (not shown).

For model fitting of 158 DWIs, 95.0% of all voxels had a value within the interval $m_{gt} \pm 0.5$ (where m_{gt} is the reference standard value) for radial kurtosis. The ratio was comparably high at 94.7% for q-DL of only 12 DWIs. For fractional kurtosis anisotropy, 81.0% of all voxels in model fit of 158 DWIs had a value in the interval $m_{gt} \pm 0.3$, whereas for q-DL of only 12 DWIs the ratio was as high as 95.9%. Intracellular volume fraction was estimated within $m_{gt} \pm 0.3$ by model fit of 158 DWIs in 88.9% of all voxels, and by q-DL of 12 DWIs in 90.6%. For neurite orientation dispersion index, the ratios were 79.3% and 85.3%, respectively. Thus, q-DL of as few as 12 DWIs provides comparable, and often a better, proximity to the true value compared to model fitting of as many as 158 DWIs.

Table I shows the effects of different random subsampling schemes, training datasets and neural network initializations on the error. All results are very similar; each training dataset leads to good results. Accidental generation of a degenerate subsampling scheme or degenerate network initialization is extremely improbable, has not been encountered in practice, and can be easily checked for (using any qualitative or quantitative experiment).

Fig. 6 shows the effect of neural network settings on the test set quality, indicating that using at least 150 hidden units per

TABLE I
ERROR REPRODUCIBILITY

Sampling Scheme	Training Dataset	Initialization 1	Initialization 2	Initialization 3
1	2	0.331	0.329	0.331
1	3	0.321	0.320	0.321
1	4	0.329	0.332	0.330
2	2	0.337	0.345	0.332
2	3	0.332	0.335	0.334
2	4	0.340	0.340	0.340
3	2	0.334	0.343	0.341
3	3	0.327	0.326	0.329
3	4	0.343	0.341	0.342

Root-mean-squared error of radial kurtosis estimated by q-DL from 12 DWIs of test dataset 1 for 27 experiments with different random subsampling schemes, different volunteer training datasets and different neural network initializations.
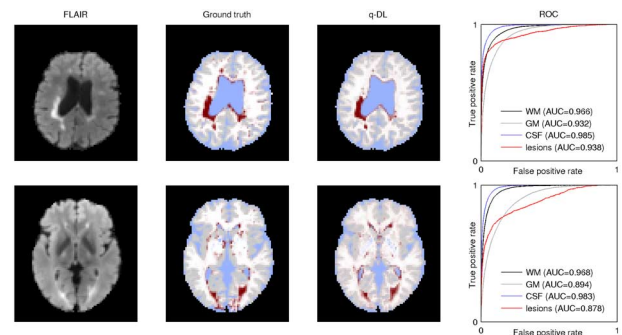

Fig. 7. Direct model-free tissue segmentation and lesion detection. When learning to discriminate multiple sclerosis lesions (red), healthy WM, GM and CSF based on DWIs and FLAIR, the proposed method segments the tissue types well and reliably detects lesions without using any diffusion model. Slices from datasets with the best (upper row, 0.938) and worst lesion AUC (lower row, 0.878) are shown.

layer or a dropout [21] fraction of at least 0.05 improves the performance of q-DL. Results of other quality measures such as root-mean-squared deviation are analogous (not shown). Note that we merely compare the effect of different parameters on the test set, rather than performing definitive hyper-parameter fitting on a validation set.

The final application of q-DL presented here is tissue segmentation and lesion detection. This task is achieved by training the neural network to discriminate between several tissue types based on the diffusion-weighted signal from the DWIs. In a proof-of-concept experiment, we used segmentation into WM, GM, CSF and multiple sclerosis lesions. Segmentation results from q-DL are shown in Fig. 7. The AUC of the ROC for lesions ranged between 0.878 and 0.938 for six different patients. AUC for WM, GM and CSF was consistently above 0.894 for all patients. Thus, DWIs can be used directly for segmentation without a diffusion model, i.e., without the intermediate information loss detailed in Section I-B. Tailoring the protocol to optimal results in specific applications is subject of future research.

Other previously proposed methods, including machine learning methods [17], [18], [47]–[49] as well as state-of-the-art compressed sensing [50] require more DWIs and several intermediate steps (see Fig. 1(b)). For the number of DWIs suggested for different methods, see Table II. Most notably, compressed sensing and machine learning publications suggest using 64 DWIs for DKI [50] and 30 DWIs for NODDI [48],

TABLE II
COMPARISON OF REQUIRED NUMBER OF DWIS

Method	Number of DWIs required for DKI	Number of DWIs required for NODDI	References
Standard Pipeline	150	99	[3], [4]
Compressed Sensing	64	–	[50]
Machine Learning *with* model fitting	–	30	[48]
Analytical Solutions	13-19 (specific measures only)	–	[15], [16]
q-Space Deep Learning	**12**	**8**	proposed

Comparison of suggested protocols and scan time for scalar measure estimation using different methods. q-Space Deep Learning provides the highest scan time reduction for both DKI and NODDI.

whereas our methods work with only 12 DWIs for DKI and 8 DWIs for NODDI. Previous work that uses the DWIs directly as inputs to machine learning for tissue characterization [17], [18] does not only use large numbers of DWIs but is also limited so far to Monte Carlo simulations only, rather than *in vivo* experiments. A related idea is the use of DWIs directly as inputs to machine learning for tractography [25].

When switching to another scanner such that the DWI intensities are not the same anymore, the intensities should either be normalized or the network should be retrained. The same holds for changes in acquisition parameters such as echo time. A network that is able to understand data from different settings is subject of future research.

In all presented applications, neural network training takes about one minute on a desktop computer. The network needs to be trained only once and can be applied to any number of datasets, taking 0.03 seconds per dataset, as opposed to several minutes per dataset required by most model fitting methods. Analytical solutions [15], [16] of scalar measure estimation provide acceleration of acquisition and processing comparable to q-DL, but are limited to specific scalar measures and acquisition schemes. With q-DL, the acceleration factor can be freely chosen and all scalar measures can be obtained simultaneously. There is also freedom in the choice of the sampling; in particular, random sampling yields robust results.

IV. CONCLUSIONS

The presented scan acceleration factor twelve sets a new state of the art in DKI and NODDI and thus opens new perspectives for clinical protocols. The results indicate that a considerable amount of information is contained in a limited number of DWIs, and that this information can be better retrieved by deep learning than by model fitting. The number of used DWIs can be freely chosen and represents a better trade-off between scan duration and quality than provided by conventional methods.

Our framework for model-free diffusion MRI can be used to estimate arbitrary tissue properties in various settings where ground truth training datasets are available. Future research may focus on creating ground truth training data from simulations, scanned phantoms and histologically validated data. Moreover, model-free q-space novelty detection (q-ND) without requirement for lesion labels has also been demonstrated [51]. q-ND

and q-DL are the first model-free diffusion MRI segmentation methods, meaning that they use q-space data directly and do not partly discard information at intermediate steps.

Recent work [52] indicates that the complexity of state-of-the-art diffusion models is at the limit of allowing a stable model fit to the noisy diffusion MRI data obtained in an acceptable scan duration. Herein we demonstrate the fact that omitting model fitting allows considerably more stable measure estimation at short scan durations; this might circumvent the fitting stability "bottleneck" when balancing scan duration against model complexity.

Classical quantitative diffusion MRI requires creating a diffusion model that well captures disease-related tissue changes via its associated scalar measures. Subsequently, a set of MRI contrasts needs to be chosen (diffusion-weighted gradient strengths and durations, single-pulsed or other gradient forms, non-diffusion sequences) that allow estimating all parameters of the model. The presented segmentation and abnormality detection method on the other hand is concerned with finding a set of contrasts whose signal "vector" (signal values from all contrasts) is most strongly affected by disease.[3] Simulational tissue models can still drive the design of meaningful gradient forms, but subsequent experiments do not rely on any model—particularly, model parameters do not have to be estimated. This allows future research to explore experiment design using elaborate simulational tissue models with large numbers of microstructural parameters. In this framework, model complexity is not limited by ill-posedness of subsequent model parameter estimation.

A combination of q-DL (requiring twelve times less DWIs than standard methods for estimation of arbitrary scalar measures) with simultaneous multi-slice imaging [39] (three-fold accelerated acquisition of the DWIs) in future applications is straightforward, yielding an unprecedented 36-fold scan time reduction.

Our recommendation in the short term is to use short acquisitions with q-DL instead of long acquisitions with fitting. In the long term, we recommend creating complex tissue models that are not limited by fitting instabilities; as well as using model-free q-DL tissue characterization.

The capability of q-DL to accelerate the acquisition by an order of magnitude and detect tissue changes without a diffusion model opens new perspectives for research in quantitative diffusion MRI and demonstrates the benefits of deep learning for multi-step data processing pipelines.

[3]In other words, we search for a set of contrasts that well captures disease-related variation of the data.

References

[1] D. Le Bihan, "Looking into the functional architecture of the brain with diffusion MRI," *Nat. Rev. Neurosci.*, vol. 4, no. 6, pp. 469–80, 2003.

[2] J. H. Jensen, J. A. Helpern, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging," *Magn. Reson. Med.*, vol. 53, no. 6, pp. 1432–40, 2005.

[3] H. Lu, J. H. Jensen, A. Ramani, and J. A. Helpern, "Three-dimensional characterization of non-Gaussian water diffusion in humans using diffusion kurtosis imaging," *NMR Biomed.*, vol. 19, no. 2, pp. 236–47, 2006.

[4] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander, "NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain," *NeuroImage*, vol. 61, no. 4, pp. 1000–16, 2012.

[5] A. R. Padhani *et al.*, "Diffusion-weighted magnetic resonance imaging as a cancer biomarker: Consensus and recommendations," *Neoplasia*, vol. 11, no. 2, pp. 102–125, 2009.

[6] M. Hori *et al.*, "Visualizing non-Gaussian diffusion: Clinical application of q-space imaging and diffusional kurtosis imaging of the brain and spine," *Magn. Reson. Med. Sci.*, vol. 11, no. 4, pp. 221–33, 2012.

[7] A. Luna, R. Ribes, and J. A. Soto, *Diffusion MRI Outside the Brain*. Berlin, Germany: Springer, 2012.

[8] *Clinical MR Neuroimaging: Physiological and Functional Techniques*, J. H. Gillard, A. D. Waldman, and P. B. Barker, Eds., 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[9] M. M. Cheung *et al.*, "Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study," *NeuroImage*, vol. 45, no. 2, pp. 386–92, 2009.

[10] G. P. Winston, "The physical and biological basis of quantitative parameters derived from diffusion MRI," *Quant. Imag. Med. Surg.*, vol. 2, no. 4, pp. 254–65, 2012.

[11] L. Qi, Y. Wang, and E. X. Wu, "D-eigenvalues of diffusion kurtosis tensors," *J. Comput. Appl. Math.*, vol. 221, no. 1, pp. 150–157, 2008.

[12] J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen, "Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls," *NeuroImage*, vol. 81, pp. 335–46, 2013.

[13] D. H. J. Poot, A. J. den Dekker, E. Achten, M. Verhoye, and J. Sijbers, "Optimal experimental design for diffusion kurtosis imaging," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 819–29, Mar. 2010.

[14] J. Veraart, W. Van Hecke, and J. Sijbers, "Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model," *Magn. Reson. Med.*, vol. 66, no. 3, pp. 678–86, 2011.

[15] B. Hansen, T. E. Lund, R. Sangill, and S. N. Jespersen, "Experimentally and computationally fast method for estimation of a mean kurtosis," *Magn. Reson. Med.*, vol. 69, no. 6, pp. 1754–60, 2013.

[16] B. Hansen, T. E. Lund, R. Sangill, and S. N. Jespersen, "A fast and robust method for simultaneous estimation of mean diffusivity and mean tensor kurtosis," in *Proc. Joint Annu. Meet. ISMRM-ESMRMB*, 2014, p. 2602.

[17] M. Nilsson *et al.*, "Evaluating the accuracy and precision of a two-compartment Kärger model using Monte Carlo simulations," *J. Magn. Reson.*, vol. 206, pp. 59–67, 2010.

[18] G. Nedjati-Gilani, M. G. Hall, C. A. M. Wheeler-Kingshott, and D. C. Alexander, "Learning microstructure parameters from diffusion-weighted MRI using random forests," in *Joint Annu. Meet. ISMRM-ESMRMB*, 2014, p. 2626.

[19] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2146–2153.

[20] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, p. 432.

[21] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv:1207.0580*, 2012.

[22] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2014.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] V. Golkov *et al.*, "q-Space deep learning for twelve-fold shorter and model-free diffusion MRI scans," in *Proc. MICCAI*, 2015, pp. 37–44.

[25] P. F. Neher, M. Götz, T. Norajitra, C. Weber, and K. H. Maier-Hein, "A machine learning based approach to fiber tractography using classifier voting," *Proc. MICCAI*, pp. 45–52, 2015.

[26] S. M. Plis *et al.*, "Deep learning for neuroimaging: A validation study," *Front. Neurosci.*, vol. 8, no. 229, 2014.

[27] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 2924–2932.

[28] R. B. Palm, "Prediction as a candidate for learning deep hierarchical models of data," M.S. thesis, Tech. Univ. Denmark, Lyngby, Denmark, 2012.

[29] H. Bagher-Ebadian *et al.*, "Predicting final extent of ischemic infarction using artificial neural network analysis of multi-parametric MRI in patients with stroke," *PLoS One*, vol. 6, no. 8, 2011.

[30] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.

[31] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 677–688, Aug. 2001.

[32] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K. Müller, Eds., 2nd ed. New York: Springer, 2012.

[33] V. J. Wedeen, P. Hagmann, W.-Y. I. Tseng, T. G. Reese, and R. M. Weisskoff, "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging," *Magn. Reson. Med.*, vol. 54, no. 6, pp. 1377–86, 2005.

[34] J. L. R. Andersson, S. Skare, and J. Ashburner, "How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging," *NeuroImage*, vol. 20, no. 2, pp. 870–88, 2003.

[35] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. S208–19, 2004.

[36] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[37] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62–79, 2013.

[38] D. A. Feinberg *et al.*, "Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging," *PLoS One*, vol. 5, no. 12, 2010.

[39] K. Setsompop *et al.*, "Improving diffusion MRI using simultaneous multi-slice echo planar imaging," *NeuroImage*, vol. 63, no. 1, pp. 569–80, 2012.

[40] J. Xu *et al.*, "Characterizing tumor response to chemotherapy at various length scales using temporal diffusion spectroscopy," *PLoS One*, vol. 7, no. 7, p. E41714, 2012.

[41] S. N. Sotiropoulos *et al.*, "Advances in diffusion MRI acquisition and processing in the human connectome project," *NeuroImage*, vol. 80, pp. 125–143, 2013.

[42] M. F. Glasser *et al.*, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, 2013.

[43] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[44] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.

[45] E. S. Hui, M. M. Cheung, L. Qi, and E. X. Wu, "Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis," *NeuroImage*, vol. 42, no. 1, pp. 122–34, 2008.

[46] B. Bilgic *et al.*, "Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries," *Magn. Reson. Med.*, vol. 68, no. 6, pp. 1747–54, 2012.

[47] G. Nedjati-Gilani, T. Schneider, M. G. Hall, C. A. M. Wheeler-Kingshott, and D. C. Alexander, "Machine learning based compartment models with permeability for white matter microstructure imaging," *Proc. MICCAI*, pp. 257–264, 2014.

[48] D. C. Alexander, D. Zikic, J. Zhang, H. Zhang, and A. Criminisi, "Image quality transfer via random forest regression: Applications in diffusion MRI," *Proc. MICCAI*, pp. 225–232, 2014.

[49] T. Schultz, "Learning a reliable estimate of the number of fiber directions in diffusion MRI," *Proc. MICCAI*, pp. 493–500, 2012.

[50] M. Paquette, S. Merlet, G. Gilbert, R. Deriche, and M. Descoteaux, "Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging," *Magn. Reson. Med.*, vol. 73, pp. 401–416, 2015.

[51] V. Golkov *et al.*, "Model-free novelty-based diffusion MRI," *Proc. Int. Symp. Biomed. Imag.*, pp. 1233–1236, 2016.

[52] U. Ferizi *et al.*, "White matter compartment models for in vivo diffusion MRI at 300 mT/m," *NeuroImage*, vol. 118, pp. 468–483, 2015.