Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Q-space-coordinate-guided neural networks for high-fidelity diffusion tensor estimation from minimal diffusion-weighted images*

Maokun ZHENG†, Zhi LI†‡, Long ZHENG†, Weidong WANG†, Dandan LI, Guomei WANG

State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, 550025, China

†E-mail: 13793015018@163.com; zhili@gzu.edu.cn; zhenglong178@163.com; 13051099799@163.com

Received Sept. 03, 2024; Revision accepted Jan. 24, 2025; Crosschecked

Unedited

Abstract: Diffusion tensor imaging (DTI) is a widely used imaging technique for mapping living human brain tissue's microstructure and structural connectivity. Recently, deep learning methods have been proposed to rapidly estimate diffusion tensors (DTs) only using a small quantity of diffusion-weighted (DW) images. However, these methods typically use the DW images obtained with fixed q-space sampling schemes as the training data, limiting the application scenarios of such methods. To address this issue, we developed a new deep neural network called q-space-coordinate-guided diffusion tensor imaging (QCG-DTI), which can efficiently and correctly estimate DTs under flexible q-space sampling schemes. First, we propose a q-space-coordinate-embedded-feature consistency strategy to ensure the correspondence between q-space coordinates and their respective DW images. On this basis, a q-space-coordinate fusion module (QCF) is introduced, which efficiently embeds q-space coordinates into multiscale features of the corresponding DW images by linearly adjusting the feature maps, thus eliminating the dependence on fixed diffusion sampling schemes. Finally, a multiscale feature residual dense module (MRD) is proposed, which enhances the network's feature extraction and image reconstruction capabilities by using dual-branch convolutions with different kernel sizes to extract features at different scales. Compared to state-of-the-art methods that rely on a fixed sample scheme, the proposed network can obtain high-quality diffusion tensors and derived parameters even using DW images acquired with flexible q-space sampling schemes. Compared to state-of-the-art deep learning methods, QCG-DTI reduced the mean absolute error by approximately 15% on fractional anisotropy and around 25% on mean diffusivity.

Key words: Diffusion tensor imaging; Diffusion tractography; Deep learning; Fast diffusion tensor estimation; Q-space coordinate information

https://doi.org/10.1631/FITEE.2400766 **CLC number:** TP

1 Introduction

Magnetic resonance imaging (MRI) is currently the mainstream and commonly used clinical detection and diagnostic method in medical imaging. Diffusion magnetic resonance imaging (dMRI) is a new type of MRI technology that has developed in recent years (Alexander et al., 2007). For imaging, dMRI mainly utilizes the different diffusion motion characteristics exhibited by water molecules in various tissues of organs under the influence of magnetic fields with different diffusion gradient directions. The difference between dMRI and traditional MRI images is that the features of each voxel are not only described by grayscale values but also contain a high-dimensional tensor to describe the diffusion

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 62062023)

ORCID: Maokun ZHENG, http://orcid.org/0009-0008-7777-0042; Zhi LI, http://orcid.org/0000-0001-9813-4979

© Zhejiang University Press 2025

characteristics of water molecules in organs. In clinical applications, it is necessary to collect diffusion-weighted imaging (DWI) images through dMRI to calculate the diffusion tensor (DT) (Le Bihan et al., 2001; Alexander et al., 2007) and use DT to derive a series of parameters, such as mean diffusivity (MD), fractional anisotropy (FA), axial diffusivity (AD), and radial diffusivity (RD). These tensor-derived parameters have been widely used in studies of brain development (Barnea-Goraly et al., 2005; Lebel et al., 2008) and various diseases (Eriksson et al., 2001; Barnea-Goraly et al., 2004; Roosendaal et al., 2009). DT imaging (DTI) is an essential tool for neuroscience research and has many clinical applications.

In the traditional algorithm, only six diffusion-weighted (DW) images and one non-DW image are theoretically needed to estimate the DT. However, DW images obtained in clinical practice usually have low signal-to-noise (SNR) ratios. The traditional DT estimation algorithms cannot accurately estimate the DT for DW images with low SNR ratios. The ordinary least squares (OLS) method was initially proposed to estimate DTs (Pierpaoli et al., 1996), but it is susceptible to noise and variations in local magnetic susceptibility. To enhance the reliability of the estimation results, the weighted least squares (WLS) method was proposed (Basser et al., 2000). WLS adjusts the accuracy of the data and the noise level through a weighting matrix. However, these traditional methods still have certain limitations in DT estimation. Consequently, in practice, it is common to estimate DTs by acquiring a large number of DW images (Jones, 2004). Jones (2004) and Jones et al. (2013) have shown that at least 30 DW images are required along a uniformly distributed direction to achieve statistical rotation invariance. Therefore, to estimate DT more accurately, each patient must undergo approximately 10–30 min of scanning. However, prolonged scanning time can cause discomfort for patients and may lead to motion artifacts. The use of a small amount of DW images for efficient and accurate estimation of DTs has gradually become one of the critical issues studied by scholars in the field of dMRI image processing.

Some scholars leverage the powerful representation learning capability of deep learning to estimate high-quality DT using DW images with six fixed diffusion gradient directions and b-values, along with one non-DW image, thereby reducing the number of DW images required for DT estimation and subsequently decreasing scan time (Tian et al., 2020; Li et al., 2021; Karimi and Gholipour, 2022). DeepDTI (Tian et al., 2020) first uses a neural network to denoise the original (low-quality) DW images to obtain high-quality DW images and then fits a traditional tensor model to obtain high-quality DT. SuperDTI (Li et al., 2021) learns the mapping between DW images and DT-derived parameters through a neural network rather than directly estimating DT. TransDTI (Karimi and Gholipour, 2022) bypasses the traditional tensor model sensitive to noise and errors, applying a more flexible transformer model to achieve end-to-end cross-modal high-quality DT estimation.

Although the above-mentioned deep learning methods can reconstruct high-quality DTs using a small number of DW images, they still have some limitations in practical applications. These methods require DW images with a specific q-space sampling scheme as input, and this q-space sampling scheme must be consistent with the training data. Although the traditional linear least squares (LLS) fitting method has good generalization performance, the accuracy of fitting results is not high when the number of diffusion gradient directions is few. A new method called DIFFnet (Park et al., 2022) has been proposed to address this issue. DIFFnet projects and quantifies the q-space coordinates along with the corresponding diffusion signals to produce a matrix (Qmatrix) as the network input, thereby achieving generalization. However, when the number of diffusion gradient directions decreases, the quality of the DT estimated by DIFFnet correspondingly declines. To address the impact of changes in diffusion gradient directions on tensor estimation quality, FlexDTI (Wu et al., 2024) incorporates gradient direction information through dynamic convolution kernels and adapts to variations in gradient directions by setting a maximum number of input channels, enabling high-quality DT estimation. However, it does not take the b-value input into account, which limits its ability to handle data with varying b-values, potentially affecting the stability and accuracy of the results.

Due to the significant differences among DW images with various q-space coordinates, it is necessary to effectively fuse the q-space coordinates with their corresponding DW images when con-

Unedited

structing a flexible and efficient DT estimation network. In this paper, we propose a network called q-space-coordinate-guided diffusion tensor imaging (QCG-DTI), which presents a q-space-coordinate-embedded-feature consistency strategy to improve the above methods for estimating DTs. Unlike other methods (Tian et al., 2020; Li et al., 2021; Karimi and Gholipour, 2022) that directly concatenate DW images as input, the proposed network treats each q-space coordinate and its corresponding DW image as a set subjected to multiscale embedding and fusion. Specifically, each dataset extracts multiscale features from the DW images using an encoder with shared parameters and integrates these features with the corresponding q-space coordinates. The q-space-coordinate-embedded-feature consistency strategy enables the network to thoroughly learn the correspondence relationship between each q-space coordinate and their respective DW images. To enable efficient fusion of each dataset in the q-space-coordinate-embedded-feature consistency strategy, we propose a q-space-coordinate fusion (QCF) module. This module performs global feature extraction from the DW images, and then the global features are combined with their corresponding q-space coordinates. This module efficiently fuses q-space coordinates with the global feature of their corresponding DW image by linearly adjusting the channel dimension. In order to enhance QCG-DTI's capability in extracting multiscale features and estimating DT, we propose a multiscale feature residual dense (MRD) module. This module extracts multiscale features through a dual-branch structure with convolutions of different kernel sizes and shares features between different convolution layers using long–short skip connections. In summary, the proposed paper combines the q-space-coordinate-embedded-feature consistency strategy, QCF module, and MRD module together to build a flexible and efficient DT estimation network, significantly improving the accuracy and reliability of DT estimation.

In summary, the contributions of this paper are as follows:

1. We propose a flexible and efficient DT estimation network (QCG-DTI) incorporating a q-space-coordinate-embedded-feature consistency strategy. This strategy ensures the correspondence between each q-space coordinate and its respective DW image by efficiently embedding each q-space coordinate with its corresponding DW image.
2. We introduce two modules to implement the q-space embedded feature consistency strategy. The one is the QCF module, which uses the q-space sampling scheme as the condition to linearly adjust the multiscale features extracted from the DW images, thereby achieving efficient embedding of q-space coordinates. The other module is the MRD module, which extracts features at different scales using a dual-branch convolutional layer structure to enhance the ability to extract features from DW images and reconstruct DTs.
3. Experiments on multiple DW image datasets demonstrate that QCG-DTI outperforms state-of-the-art methods and shows greater robustness to data acquired using different q-space sampling schemes. QCG-DTI reduced the mean absolute error by approximately 15% on FA and around 25% on MD.

2 Related works

In DTI, water molecule diffusion is assumed to follow a Gaussian model. Multiple DW images can be obtained by applying at least six diffusion-sensitized gradient fields in different directions. The noise-free amplitude signal in a single voxel is given by the Stejskal–Tanner equation (Stejskal and Tanner, 1965), as shown in Eq. (1). The DW signal can be expressed as follows:

$$s\left(b,g\right)=s_0 exp\left(-bg^T\boldsymbol{D}g\right), \tag{1}$$

where s_0 represents the non-DW signal, b represents the diffusion-weighting intensity, $g=(g_x,g_y,g_z)^T$ represents the gradient direction of the diffusion gradient field, and $\boldsymbol{D}$ represents the DT, which can be expressed as a 3×3 symmetric matrix, as follows:

$$\boldsymbol{D}=\begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}. \tag{2}$$

Here, the diagonal elements represent the diffusion of water molecules along the x, y, and z orthogonal directions. In contrast, the off-diagonal elements represent the correlation of the diffusion process of

Unedited

the water molecule between mutually perpendicular directions.

In practical applications, the acquired DW images are often affected by noise, and as the diffusion coefficient increases, the SNR ratio of the obtained data decreases. Pierpaoli et al. (1996) proposed a method for estimating DTs and fiber bundle tracking using OLS, which laid the foundation for subsequent research. However, the OLS method cannot describe complex fiber structures and is susceptible to interference from noise and local magnetic susceptibility variations. To address these issues, the weighted least squares (WLS) method (Basser et al., 2000) provides a way to adjust the accuracy and noise level of the data through a weighting matrix, thereby improving the reliability of the estimates. To further enhance the accuracy of the estimates, the matrix diagonalization method (Westin et al., 2002) introduces information on the eigenvalues and eigenvectors of the DT for fiber bundle tracking and anisotropy calculation. The echo signal reconstruction method (Wu and Alexander, 2007) reconstructs the DT by inverting the echo signals and performs well under low SNR conditions. However, this method has high computational complexity and is affected by scanning parameters and noise, making the estimates susceptible to error interference.

Recently, deep learning has shown great promise in improving the accuracy and robustness of dMRI parameter estimation (Golkov et al., 2016; Li et al., 2019; Nath et al., 2019; Almeida Martins et al., 2021). Several recent studies (Golkov et al., 2016; Tian et al., 2020; Li et al., 2021; Karimi and Gholipour, 2022) have demonstrated that deep learning can significantly reduce the number of measurements, thereby decreasing scan time. The pioneering work by Golkov et al. (2016) introduced the concept of deep learning to dMRI and established the q-space deep learning (q-DL) framework, dramatically advancing subsequent work in this field. The q-DL method proposes using a multilayer perceptron (MLP) to synthesize more q-space samples from a small number of acquired q-space samples without imposing any diffusion model. The synthesized q-space samples can fit any preferred diffusion model, thereby improving the results. q-DL and other subsequent research works have demonstrated that deep learning has significant potential in reducing the amount of q-space data required for diffusion imaging (Poulin et al., 2017; Aliotta et al., 2019; Gibbons et al., 2019; Tong et al., 2020).

The introduction of convolutional neural networks (CNNs) has enabled the exploitation of spatial correlations in signals from neighboring voxels, leading to more accurate and robust estimates. Previous works have applied CNNs to estimate fiber directions from DW (Koppers et al., 2017; Li et al., 2019, 2021; Lin et al., 2019; Tian et al., 2020; Park et al., 2022; Wu et al., 2024). Li et al. (2019) demonstrated that simple CNNs can achieve more accurate diffusion kurtosis estimates than the q-DL framework. The DeepDTI (Tian et al., 2020) method utilizes raw DW images and T1/T2-weighted images as inputs to a CNN to enhance the quality of DW images. These high-quality DW images are then used to estimate the DT through traditional methods. While DeepDTI has demonstrated the feasibility of using deep learning for rapid DT reconstruction, it still relies on traditional methods to estimate the DT, which lacks robustness to noise and errors. SuperDTI (Li et al., 2021) avoids the noise-sensitive tensor fitting step by directly using a deep CNN to model the nonlinear relationship between the acquired DW images and the desired DT-derived parameters, e.g. FA, MD, and eigenvectors. This approach improves speed by reducing the number of DW images. However, this method requires training a separate network for each derived parameter, significantly increasing the time cost. Recent literature has proposed a new network structure, Transformer, which leverages more profound mapping (Vaswani et al., 2017), sequence-to-sequence model design (Sutskever et al., 2014), and an adaptive self-attention mechanism with an extended receptive field (Qin et al., 2021). In this context, TransDTI (Karimi and Gholipour, 2022) introduced a transformer-based DT estimation method that uses a two-stage training approach to directly estimate the DT, utilizing the correlation between diffusion signals and DT parameters in neighboring voxels. However, these methods require DW images obtained using a fixed q-space sampling scheme as input.

During the inference phase, when the input data have diffusion gradient directions or b-values different from those in the training data, the network often needs to be retrained, which consumes significant time and effort. To address this issue, DIFFnet (Park et al., 2022) first normalizes the diffusion sig-

Unedited

nals in q-space and then projects and quantizes the corresponding q-space coordinates onto an orthogonal plane, generating a matrix (Qmatrix) as the network's input to achieve generalization. However, when the number of diffusion gradient directions decreases, the projection causes the signals to become sparse and scattered on the orthogonal plane, reducing the quality of the estimated DT-derived metrics. Moreover, DIFFnet does not fully consider the spatial correlations between the signals of neighboring voxels during processing. Given the inherent continuity and structural features within DW images, ignoring spatial correlation information affects the robustness and accuracy of the estimated DT-derived parameters. To address the impact of variations in diffusion gradient directions on the quality of DT estimation, FlexDTI (Wu et al., 2024) embeds gradient direction information into diffusion signal feature maps using dynamic convolution kernels, thereby enhancing sensitivity to changes in gradient directions. It also adapts to variations in the number of gradient directions by setting a maximum number of input channels. However, FlexDTI does not incorporate b-value as an input, limiting its adaptability when processing data with different b-values, which affects the stability and accuracy of the reconstruction results.

To efficiently estimate high-quality DTs from a small number of DW images using flexible q-space sampling schemes, the proposed network combines the q-space-coordinate-embedded-feature consistency strategy, the QCF module, and the MRD module. This can effectively align the q-space coordinates with the DW images and allow for rapid and flexible estimation of the DTs even with a limited number of DW images, to significantly enhance the model's generalization and practicality.

3 Method

3.1 Network architecture

Our primary purpose is to develop a rapid DT estimation model using flexible q-space sampling schemes. To ensure that different q-space coordinates correspond correctly to their respective DW images within the network, we propose a q-space-coordinate-embedded-feature consistency strategy. To implement this strategy, we present a QCF module for effectively embedding q-space coordinates and an MRD module to enhance feature extraction and image reconstruction capabilities. The overall architecture is shown in Fig. 1.

The input data consist of one non-DW image S_0 and six DW images $\boldsymbol{S} = [S_1, S_2, S_3, S_4, S_5, S_6]$, along with their q-space sampling schemes $\boldsymbol{b} = [b_1, b_2, b_3, b_4, b_5, b_6]$, where each q-space coordinate b_i corresponds to the DW image S_i, $i \in \{1, 2, 3, 4, 5, 6\}$. Here, $b_i = (x_i, y_i, z_i, l_i)$, where (x_i, y_i, z_i) represents the diffusion gradient direction and l_i represents the b-value. The network's output consists of the six different elements of the DT $(D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz})$.

3.2 Q-space-coordinate-embedded-feature consistency strategy

The proposed method uses six q-space coordinates along with their corresponding DW images and one non-DW image as inputs. Each q-space coordinate corresponds to a specific DW image, meaning that setting a q-space coordinate can yield the corresponding DW image via an MRI scanner. Suppose all DW images are directly concatenated along the channel dimension as model input. In that case, it will be challenging for the model to learn the correspondence between q-space coordinates and DW images, thereby reducing the quality of DT estimation. We propose a q-space-coordinate-embedded-feature consistency strategy, which is as follows: First, each q-space coordinate is paired with its corresponding DW image to constitute six sets of input data. Then, a shared-parameter encoder extracts multiscale features from the DW images S_i in each input set. Subsequently, the q-space coordinates b_i from each input set are redundantly embedded with the multiscale features of the corresponding DW images, which can more effectively learn the relationships between different q-space coordinates and DW images. Finally, the multiscale features embedded with the q-space coordinates from all six input sets are fused at the same level, and the fused features are fed into a symmetric decoder to reconstruct high-quality DTs. The specific approach is as follows:

First, concatenate the six DW images $\boldsymbol{S}$ with the same non-DW image S_0 along the channel dimension to obtain six sets of inputs $F_{in}^{i} \in R^{B\times 2\times H\times W\times D}$ for the shared parameter encoder. Here, i is the number of DW images, $i \in 1, 2, 3, 4, 5, 6$; B represents

Unedited

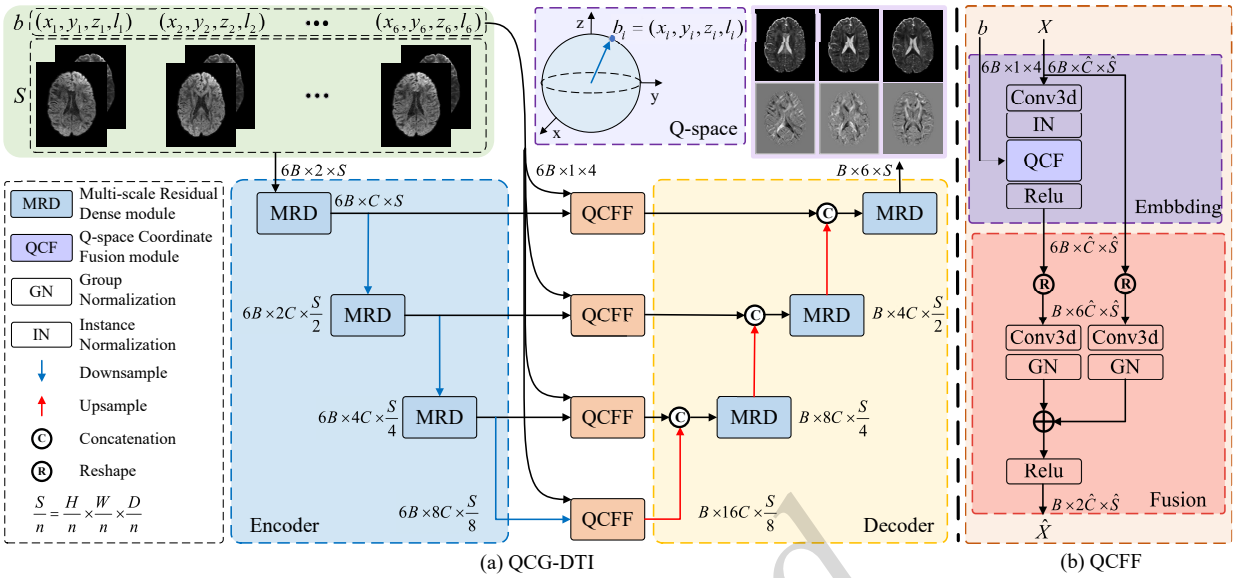

Fig. 1 The architecture of our proposed QCG-DTI. The input data consists of one non-DW image and six DW images with corresponding q-space coordinates. Each DW image, along with the non-DW image and its q-space coordinates, is treated as a set, resulting in six sets. Each of the six sets is processed through an SE to extract multiscale features from the DW images, and the q-space coordinates are embedded into the multiscale features of the corresponding DW images using the QCF module. Finally, the six sets of embedded multiscale features are fed into a decoder to estimate the DT. Downsampling is achieved using 3D convolutions with a kernel size of four and a stride of two, while upsampling is performed using a combination of nearest-neighbor interpolation and convolution-based ResizeConv. DT, diffusion tensor; DTI, DT imaging; DW, diffusion-weighted; QCF, q-space-coordinate fusion; SE, shared encoder; QCFF, QCF module fusion; GN, group normalization; IN, instance normalization.

Unrevised

the batch size of the input data. The number of channels is two: one for the non-DW image, and the other for the DW image. The H, W, and D represent the data's height, width, and depth, respectively. Since each set of input needs to pass through the shared encoder (SE) to extract multiscale features and embed the q-space coordinates, we concatenate the six sets of input data along the batch dimension B during implementation, resulting in the input $F_{in} \in R^{6B\times 2\times H\times W\times D}$ for the SE and the input $F_b \in R^{6B\times 1\times 4}$ for the QCF module.

Next, each input datum's multiscale features $F_{SE}^{j} \in R^{6B\times C\times H\times W\times D}$ are extracted through a shared-parameter encoder. This method fully utilizes shared parameters to enhance the efficiency of feature extraction while ensuring that the feature extraction process for each DW image is interrelated.

$$F_{SE,1} = f_{SE,1}\left(F_{in}\right), \tag{3}$$

$$F_{SE,j} = f_{SE,j}\left(f_{down}\left(F_{SE,j-1}\right)\right), \tag{4}$$

where $f_{SE,j}(\cdot)$ represents the j-th MRD operation in the SE, $f_{down}(\cdot)$ represents the downsampling operation. $F_{SE,j}$ represents the multiscale features of each DW image, and $F_{SE,j} = [F_{SE,j}^{1}, F_{SE,j}^{2}, F_{SE,j}^{3}, F_{SE,j}^{4}, F_{SE,j}^{5}, F_{SE,j}^{6}]$.

Then, through the QCF module, each q-space coordinate is fused with the corresponding multiscale features $F_{SE,j}$ extracted from the DW images to obtain $F_{QCF_j} \in R^{6B\times C\times H\times W\times D}$.

$$F_{QCF_j} = f_{QCF_j}\left(F_{SE,j}, F_b\right), \tag{5}$$

where $f_{QCF_j}(\cdot)$ represents the QCF operation applied to the features obtained through $f_{SE,j}(\cdot)$ operation.

Finally, the same-level multiscale features are fused through channel concatenation after embedding the q-space coordinates. Then, a convolution operation is applied to match the number of feature channels with the corresponding level of the decoder. It ensures that the features of the input decoder are aligned with the features of the encoder, thereby facilitating the subsequent reconstruction process.

$$F_{QCFF_j} = f_{conv_j}\left(f_{reshape}\left(F_{QCF_j}\right)\right), \tag{6}$$

where $f_{conv_j}(\cdot)$ represents the fusion operation

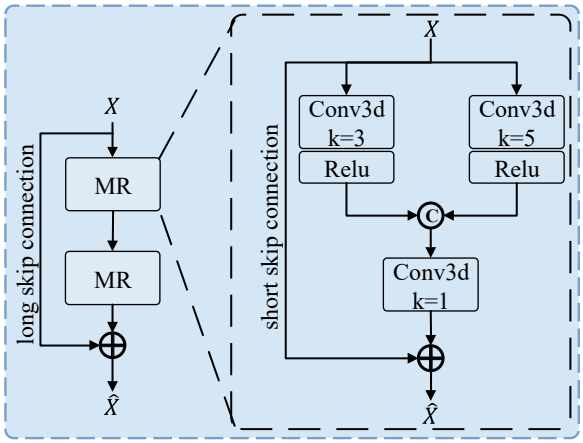

Fig. 2 Multiscale residual dense module: multilevel features are extracted using dual-branch convolutional layers with different kernels, and long-and-short skip connections are utilized to share features of different scales. MR, multiscale residual module; Relu, rectified linear unit.

applied to the features F_{QCF_j} obtained through the $f_{QCF_j}(\cdot)$operation. Specifically, $F_{QCF_j} = [F^1_{QCF_j}, F^2_{QCF_j}, F^3_{QCF_j}, F^4_{QCF_j}, F^5_{QCF_j}, F^6_{QCF_j}]$, where $F^i_{QCF_j}$ denotes the result of applying the QCF operation to the i-th DW image. $f_{reshape}(\cdot)$ denotes the concatenation operation of features embedded with the q-space coordinates.

3.3 MRD module

In CNNs, convolutional layers can automatically learn the weights of convolutional kernels during training to extract local features of the image. Therefore, choosing the size of the convolutional kernel is crucial for feature extraction. Traditionally, small kernels effectively capture an image's fine (local) details but are less effective at capturing structural (global) information; large kernels, on the other hand, are better for capturing structural information. Applying convolutional layers with different kernel sizes within the same layer ensures that the image's structural and detailed information are captured simultaneously.

We propose an MRD module to learn different-scale structural and detail features of DW images, as shown in Fig. 2. MRD consists of three parts: multiscale feature fusion, long skip connections, and short skip connections. Specifically, in the multiscale feature fusion part, MRD first uses two convolutions with kernel sizes of "3" and "5" to extract detailed and structural information from the DW images, respectively. Then, it uses a convolution with a kernel size of "1" to perform channel reduction on these multiscale features. Finally, a convolution with a kernel size of "3" is adopted for feature fusion. MRD incorporates long and short skip connections to share features across different scales, allowing the model to capture local and global information effectively. We combine short skip connections and multiscale feature fusion to form a multiscale residual (MR) module.

$$F_{MR} = X + f_{3conv}(f_{1conv}(f_{concat}(\delta\left(f_{3conv}\left(X\right)\right) + \delta\left(f_{5conv}\left(X\right)\right)))), \quad (7)$$

where F_{MR} represents the features after multiscale feature fusion, X is the input, δ denotes the rectified linear unit (ReLU) activation layer, and $f_{1conv}(\cdot)$, $f_{3conv}(\cdot)$, and $f_{5conv}(\cdot)$ represent convolution operations with kernel sizes of "1," "3," and "5,", respectively. MRD consists of two MR modules and long skip connections.

$$F_{MRD} = X + f_{MR}\left(f_{MR}\left(X\right)\right), \quad (8)$$

where $f_{MR}(\cdot)$ represents the multiscale residual module, X is the input, and F_{MRD} is the output feature.

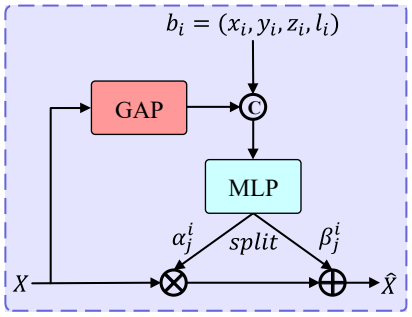

Fig. 3 QCF module. GAP, global average pooling; MLP, multilayer perceptron; QCF, q-space-coordinate fusion.

3.4 QCF module

To efficiently achieve multimodal embedding fusion of q-space sampling schemes with the multiscale features extracted from the corresponding DW images, this paper proposes a QCF module, as shown in Fig. 3. The QCF module is placed on the skip connections between the encoder and the decoder.

The QCF takes as input the q-space coordinate b_i and the j-th layer features $F^i_{SE,j}$, extracted from the corresponding DW image S_i by the SE. It embeds b_i into the features $F^i_{SE,j}$ to obtain $F^i_{QCF_j}$ through

Unedited

a channel-wise linear transformation. Specifically, QCF first extracts the global features $F^i_{avg,j}$ from the extracted multiscale features $F^i_{SE,j}$ using Global Average Pooling (GAP), achieving bidirectional learning between the DW images and the q-space coordinates. Next, the extracted features $F^i_{avg,j}$ are concatenated with b_i after being processed through a shared MLP. The concatenated features are then passed through an MLP, which generates the adjustment coefficients α^i_j and β^i_j, corresponding to the scaling and shifting parameters, respectively. Finally, the features $F^i_{SE,j}$ are linearly transformed using the adjustment coefficients α^i_j and β^i_j, to seamlessly integrate the q-space coordinate b_i into the feature space. The calculation for QCF feature fusion $F^i_{QCF_j}$ is as follows:

$$F^i_{avg,j} = f_{MLP}\left(f_{GAP}\left(F^i_{SE,j}\right)\right), \tag{9}$$

$$\alpha^i_j, \beta^i_j = f_{split}\left(f_{MLP}\left(f_{concat}\left(\sigma\left(F^i_{avg,j}\right), b_i\right)\right)\right), \tag{10}$$

$$F^i_{QCF_j} = \alpha^i_j \times F^i_{SE,j} + \beta^i_j, \tag{11}$$

where $f_{split}(\cdot)$ denotes the average splitting operation, $f_{MLP}(\cdot)$ denotes the MLP, and $f_{GAP}(\cdot)$ denotes the GAP.

3.5 Loss Function

Currently, there is no reliable establishment of more goal-oriented loss functions for dMRI. Therefore, we minimize the mean squared error loss (L2) to optimize the network parameters. The loss is described by Eq. (12).

$$L\left(D_{\text{pred}}, D_{gt}\right) = \| D_{\text{pred}} - D_{\text{gt}} \|_2^2, \tag{12}$$

where D_{pred} represents the DT estimated by the network, and D_{ref} is the corresponding reference DT.

The overall framework algorithm flow is shown in Algorithm 1.

Algorithm 1 QCG-DTI framework

Input: S_0, $\boldsymbol{S}$, $\boldsymbol{b}$

Hyperparameters: L_X,L // The number of input DW images and MRD modules in the shared encoder.

Output: D_{pred}

for $i = 1, 2, \cdots, L_X$, **do**
  $F^i_{in} \leftarrow [S_0, S_i]$ // Concatenated along the channel dimension.
  $F^i_{SE,1} \leftarrow f_{SE,1}\left(F^i_{in}\right)$
  $F^i_{QCF_1} \leftarrow f_{QCF_1}\left(F^i_{SE,1}, b_i\right)$ // Embed the q-space coordinate b_i into the extracted image features.
  for $j = 2, 3, \cdots,\ L_{SE}$, **do**
    $F^i_{SE,j} \leftarrow f_{SE,j}\left(f_{down}\left(F^i_{SE,j-1}\right)\right)$
    $F^i_{QCF_j} \leftarrow f_{QCF_j}\left(F^i_{SE,j}, b_i\right)$
  end for
  $F^i_{QCF_{L+1}} \leftarrow f_{QCF_{L+1}}\left(f_{down}\left(F^i_{SE,L}\right), b_i\right)$
end for
for $j = 1, 2, \cdots, L+1$, **do**
  $F_{QCFF_j} \leftarrow f_{conv_j}([F^1_{QCF_j}, F^2_{QCF_j}, F^3_{QCF_j}, F^4_{QCF_j}, F^5_{QCF_j}, F^6_{QCF_j}])$ // Concatenate and fuse along the channel dimension.
end for
$F_{D,L} \leftarrow f_{D,L}\left(\left[f_{up}\left(F_{QCFF_{L+1}}\right), F_{QCFF_L}\right]\right)$ // where $f_{up}(\cdot)$ represents upsampling using resizeconv, and $f_{D,L}(\cdot)$ denotes the MRD operations in the decoder that are symmetric to $f_{SE,L}(\cdot)$.
for $k = L-1, L-2, \cdots, 1$, **do**
  $F_{D,k} \leftarrow f_{D,k}\left(\left[f_{up}\left(F_{D,k+1}\right), F_{QCFF_k}\right]\right)$
end for
$D_{\text{pred}} \leftarrow f_{conv}\left(F_{D,1}\right)$
return D_{pred}

4 Experiments

4.1 Datasets

All algorithms proposed in this paper are experimented on the dMRI data of healthy volunteers from the Human Connectome Project (HCP) (Elam et al., 2021). To further demonstrate the applicability of our proposed QCG-DTI to other datasets, we also performed experimental validation using the Developing Human Connectome Project (dHCP) dataset (Bastiani et al., 2019) and the multicenter dataset of multishell diffusion MRI (MDM) (Tong et al., 2019).

4.1.1 HCP dataset

The dMRI data were collected with three different gradient tables, each including 90 diffusion-weighting directions and six b=0 acquisitions. The diffusion directions were uniformly distributed on multiple q-space shells (Caruyer et al., 2013). The directions were optimized so that every subset of the first M directions is also isotropic. Each dataset includes 18 non-DW images and 270 DW images in three different b-values: (1000 s/mm^2, 2000 s/mm^2, and 3000 s/mm^2) and 90 diffusion directions. In this experiment, we randomly selected DWI data from 80 subjects, of which, the data from 50 subjects were

Unedited

used for training, from 10 subjects for validation, and from 20 subjects for testing.

4.1.2 dHCP dataset

The dHCP dataset consists of dMRI data of neonates scanned at 29–45 gestational weeks. Each dMRI data scan in the dHCP dataset includes measurements at three b-values of 400, 1000, and 2600. Following widely adopted recommendations (Jones et al., 1999), we use $b = 1000$ measurements for DT estimation. Each scan includes 88 measurements in the $b = 1000$ shell, which are approximately uniformly distributed on the sphere. In this experiment, we randomly selected DWI data from 86 subjects aged 40–45 gestational weeks; of these data, 70 sets were used for training, 8 for validation, and 8 for testing.

4.1.3 Multicenter dataset of MDM

The MDM dataset (Tong et al., 2019) includes three healthy traveling subjects (one male, 23 years old, and two females, 26 years old and 23 years old, respectively). Anatomical T1-weighted images were acquired using a 3D magnetization-prepared two rapid acquisition gradient echo (MP2RAGE) sequence. DW images were obtained using a simultaneous multislice (SMS) diffusion Echo-Planar Imaging (EPI) prototype sequence. The diffusion scheme contains 30 vectors with uniform angular coverage on each shell (b-value = 1000, 2000, and 3000, noncolinear between any two shells). Six nondiffusion volumes were equally separated in the scheme for the motion estimation. In this experiment, we use data from all subjects to demonstrate the model's generalizability.

4.2 Data Processing

4.2.1 Data Preprocessing

For the HCP and the MDM datasets, the provided volumetric brain segmentation results from the T1-weighted data of the FreeSurfer software (developed by the Laboratory for Computational Neuroimaging at the Athinoula A. Martinos Center for Biomedical Imaging. `https://surfer.nmr.mgh.harvard.edu`) (Dale et al., 1999; Fischl et al., 1999; Fischl, 2012) were downsampled to the diffusion image space using nearest neighbor interpolation. The brain was segmented using the "recon-all" function in FreeSurfer, followed by obtaining a binary mask of brain tissue excluding the cerebrospinal fluid (CSF) using the "5ttgen" function in MRtrix (Donald Tournier et al., 2012).

4.2.2 Dataset splitting

To obtain the accurate DTI metrics for each subject, the reference DT D_{gt} was obtained by DT fitting to all diffusion data (HCP: 18 non-DW images and 90 DW images; dHCP: 20 non-DW images and 88 DW images; MDM dataset: 6 non-DW images and 30 DW images) using the WLS algorithm from the Dipy library (Garyfallidis et al., 2014). We selected six of all measurements that were closest to the six optimized directions proposed by Skare et al. (2000). Specifically, the unit vectors indicating these directions are as follows: $[0.910, \pm 0.416, 0.000]$, $[\pm 0.416, 0.000, 0.910]$, and $[0.000, 0.910, \pm 0.416]$. These six directions have been derived to minimize the condition number of the DT transformation matrix (Skare et al., 2000). Specifically, the six optimized directions were rotated in a random fashion to six new directions, and the set of the six nearest directions was selected whether the mean absolute angle compared to the rotated directions was $<5^\circ$ and the condition number of the corresponding DT transformation matrix was <2. We choose one non-DW image and six DW images for the rotational variables along the optimization direction as a dataset, i.e., inputs to the model. In this experiment, for the HCP dataset, we select a total of 360 sets of input images (six groups per subject, 60 subjects) for training and validation and randomly select one such image set from among 20 evaluation subjects as the test dataset. For the dHCP dataset, we selected 468 sets of input images (six groups per subject, 78 subjects) for training and validation. We randomly selected one such image set among eight evaluation subjects as the test dataset. For the MDM dataset, we select six sets of input images (two sets per subject, for three subjects) as the evaluation dataset.

4.3 Training Details

The model was implemented using Pytorch, using the Adam optimizer (Kingma and Ba, 2015), to optimize the network parameters using default pa-

Unedited

rameters (except the learning rate), where $\beta_1 = 0.9$ and $\beta_2 = 0.9999$.

To account for subject-to-subject variations in image intensity, the intensities of the input and ground-truth images were standardized by subtracting the mean image intensity and dividing by the standard deviation of image intensities across all voxels within the brain mask from the input images. Input and ground-truth images were brain masked.

The training was conducted using the NVIDIA A100 graphical processing unit (GPU). The learning rate was set based on experience, and the number of epochs for each learning rate was chosen based on tracking of the validation error. For the first eight epochs, the learning rate was initially set to 0.0006. As the network approached convergence, the validation error did not decrease further. Then, for the last eight epochs, the learning rate was set to 0.00003 to fine-tune the network parameters (taking approximately 80 h in total). During the training process, only the network parameters of the latest epoch with the lowest validation error were saved. Due to limited GPU memory, training and validation were performed using blocks of voxel size $64 \times 64 \times 64$. For evaluation purposes, the evaluation was also performed using blocks, and the resulting estimates were subsequently reconstructed to their original positions for further assessment.

4.4 Evaluation Metrics and Comparison Methods

In this paper, we conduct a series of comparative experiments between the proposed method and previous methods that use fixed q-space sampling schemes for reconstructing high-quality DTs from a small number of DW images, specifically DeepDTI (Tian et al., 2020) and TransDTI (Karimi and Gholipour, 2022). DeepDTI denoises DW images and then uses a traditional DT estimation model to obtain the DT. We replicate the data processing and model the same according to the example code provided on their website. TransDTI uses a more flexible Transformer model as the backbone to generate DTs from DW images directly, bypassing the traditional DT estimation model. We reproduce the model according to the data provided in their paper. In addition, to demonstrate the effectiveness of our method for estimating DTs using flexible q-space sampling schemes, we conducted comparative experiments with WLS (Basser et al., 2000), DIFFnet (Park et al., 2022), and FlexDTI (Wu et al., 2024). Due to the lack of publicly available code for FlexDTI, we reproduced its algorithm based on the original paper. Furthermore, since FlexDTI was originally designed for 2D image processing, we made necessary modifications to the algorithm to adapt it for 3D data processing, ensuring fairness in the experiments.

To compare the proposed algorithm with related research methods, we first use the norm of the difference between the predicted DT and the reference DT as the evaluation metric. Specifically, D_{pred} represents the predicted DT, D_{gt} represents the reference DT, and the estimation error is defined as $\sum_{i=1}^{6} \left| D^i_{\text{pred}} - D^i_{\text{gt}} \right|$, where the index i refers to the six DT elements. We calculate the tensor decomposition for each voxel and use the standard definitions of the eigenvalues to compute FA, MD, AD, and RD (Jones, 2009), as shown in Eqs (14)–(17). For each voxel, the errors in FA, MD, AD, and RD are defined as the absolute errors between the predicted and reference DTs. Additionally, we calculate the angular error between the primary eigenvectors of the predicted and reference DTs as the angular discrepancy between the two DTs. We use Mean Absolute Difference (MAD) to compare FA, MD, AD, and RD quantitatively, and the angles between the primary eigenvectors of the tensors, as shown in Eq. (13).

$$MAD = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| , \tag{13}$$

$$FA = \frac{\sqrt{3\left(\left(\lambda_1 - \bar{\lambda}\right)^2 + \left(\lambda_2 - \bar{\lambda}\right)^2 + \left(\lambda_3 - \bar{\lambda}\right)^2 \right)}}{\sqrt{2\left(\lambda_1^2 + \lambda_2^2 + \lambda_3^2\right)}}, \tag{14}$$

$$MD = \frac{(\lambda_1 + \lambda_2 + \lambda_3)}{3}, \tag{15}$$

$$AD = \lambda_1, \tag{16}$$

$$RD = \frac{(\lambda_2 + \lambda_3)}{2}. \tag{17}$$

For 3D fiber-tracking generation, we used the fiber assigned by continuous tracking (FACT) algorithm (Mori et al., 1999) based on DTI. The FACT algorithm generates streamlines in an iterative process starting at seed points in the white matter (WM) and follows the primary fiber orientation until it terminates; we used the FACT algorithm implemented in MRtrix. The white-matter regions were obtained

Unedited

Table 1 Quantitative comparison of different methods on the HCP dataset*

Dataset	Method	Tensor ↓	FA ↓	MD ↓	AD↓	RD ↓	Angle ↓
Fixed	WLS	0.818±0.102	0.1292±0.0361	0.1614±0.0197	0.1501±0.0299	0.1748±0.0346	41.38±1.93
	DeepDTI	0.357±0.038	0.0502±0.0064	0.0592±0.0055	0.0883±0.0121	0.0627±0.0063	29.17±1.67
	TransDTI	0.341±0.024	0.0551±0.0026	0.0553±0.0068	0.0837±0.0060	0.0622±0.0068	29.39±1.05
	DiFFnet	—	0.1364±0.0036	0.4687±0.0234	0.5609±0.0239	0.3738±0.0225	—
	FlexDTI	0.346±0.030	0.0506±0.0024	0.0581±0.0093	0.0837±0.0073	0.0640±0.0089	28.92±1.04
	Proposed	**0.283±0.016**	**0.0453±0.0028**	**0.0473±0.0046**	**0.0688±0.0040**	**0.0524±0.0050**	**25.79±1.00**
Flexible	WLS	0.817±0.102	0.1289±0.0361	0.1614±0.0197	0.1501±0.0299	0.1748±0.0346	41.38±1.93
	DeepDTI	0.850±0.118	0.1492±0.0272	0.1153±0.0364	0.1739±0.0380	0.1467±0.0373	45.37±2.68
	TransDTI	0.768±0.065	0.0767±0.0034	0.0834±0.0170	0.1409±0.0162	0.0871±0.0145	55.77±3.64
	DiFFnet	—	0.1698±0.0459	0.3148±0.0789	0.3445±0.0864	0.2870±0.0754	—
	FlexDTI	0.373±0.041	0.0520±0.0023	0.0663±0.0124	0.0907±0.0112	0.0728±0.0123	29.33±1.11
	Proposed	**0.289±0.019**	**0.0443±0.0027**	**0.0499±0.0060**	**0.0706±0.0045**	**0.0546±0.0062**	**25.67±0.78**

* Using 20 test subjects from the HCP dataset. Results are presented as mean ± standard deviation. The best results are highlighted in **bold**, and the second-best results are underscored. HCP, human connectome project; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity.

Unedited

using FreeSurfer software. From the results of the whole-brain tractography, we selected the tracts of interest based on the brain parcellation obtained using FreeSurfer. We used the dice score to evaluate the similarity between the reconstructed tracts and the ground-truth labels. Precisely, the similarity was quantified by comparing the binary images generated from streamlines: a voxel was assigned a value of "1" if >5 streamlines traversed it. The dice score was then calculated between the binary ground-truth image and the predicted binary image to measure their overlap quantitatively. The dice score is defined as follows:

$$Dice = \frac{2 \times (pred \bigcap true)}{pred \bigcup true}, \tag{18}$$

where *pred* represents the predicted binary image, and *true* denotes the ground-truth binary image.

4.5 Evaluation Experiments on HCP dataset

4.5.1 Quantitative comparison

Table 1 shows the estimation errors for tensors, FA, MD, AD, RD, and primary eigenvector directions using fixed and flexible q-space sampling schemes on the HCP dataset for the proposed and comparative methods. For each of these variables, we compute the average error for each test subject data, excluding the CSF. As shown in Table 1, our method achieves lower estimation errors than all other methods across all 20 test subjects from the HCP dataset. DIFFnet, which generates DT-derived parameters such as FA, MD, AD, and RD directly from DW images rather than reconstructing the DT, does not provide experimental results or comparisons beyond these derived parameters. Compared to other methods on datasets with fixed q-space sampling schemes, our proposed method achieves optimal performance across all measurement indicators. On datasets with flexible q-space sampling schemes, DeepDTI and TransDTI show significant decreases in all metrics because these methods were trained using fixed q-space sampling scheme datasets. Although DIFFnet is designed based on flexible sampling schemes, it cannot estimate high-quality DTs using only six DW images. FlexDTI demonstrates a certain degree of robustness on datasets with flexible q-space sampling schemes. However, due to the inclusion of data with varying b-values, the estimation of MD, AD, and RD parameters is affected. In contrast, our method not only achieves optimal performance on datasets with random diffusion sampling schemes but also outperforms existing methods on datasets obtained using fixed diffusion sampling schemes. Furthermore, by incorporating b-values as input, our method shows better adaptability on flexible q-space sampling datasets, with minimal impact on the estimation of all parameters, consistently maintaining optimal results.

4.5.2 Qualitative comparison

Fig. 4 shows the visual results of FA, MD, AD, and RD generated using six DW images with WLS, DeepDTI, TransDTI, DIFFnet, and our proposed method. The references were reconstructed by WLS using all 90 DW images and 18 non-DW images. The

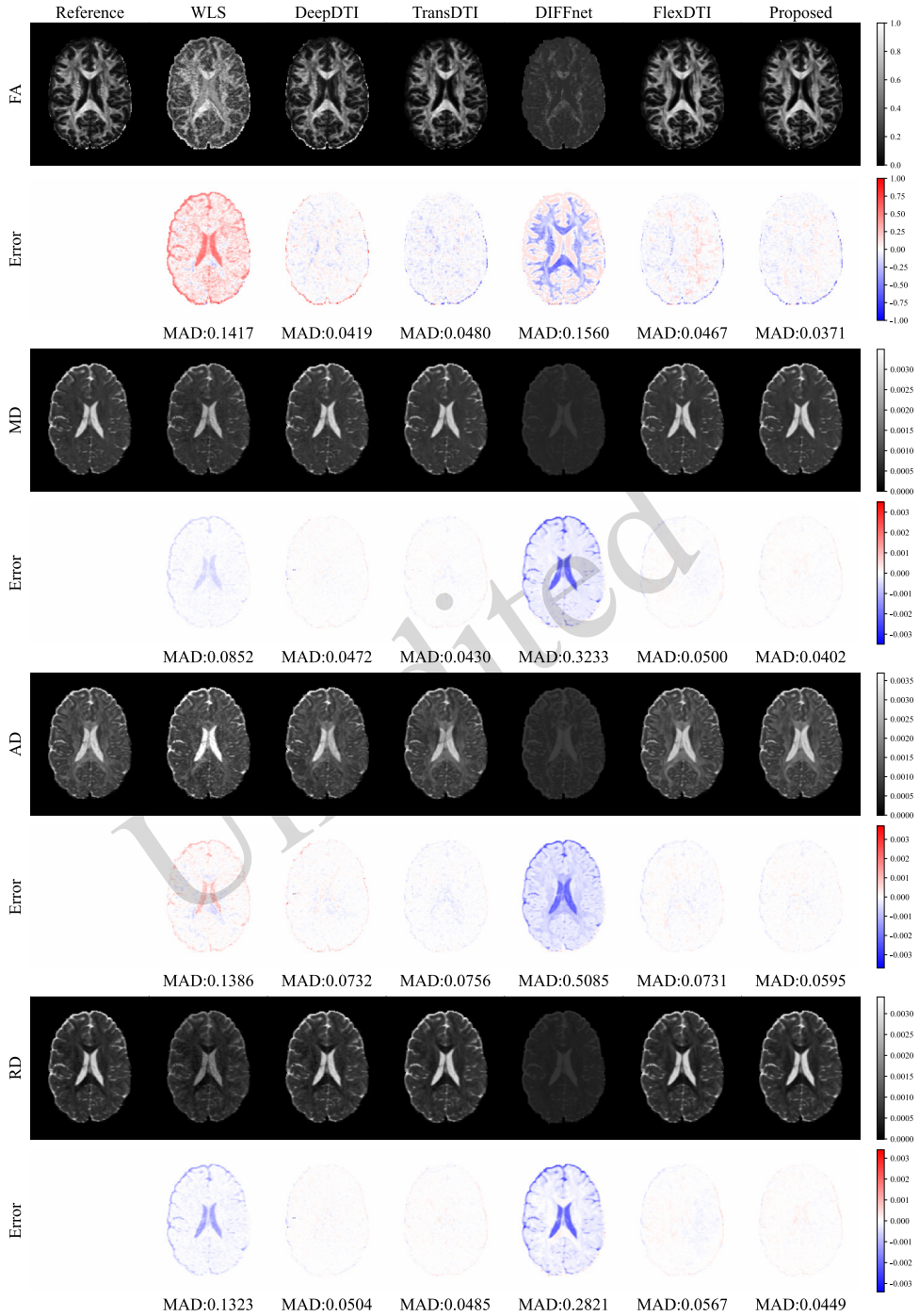

Fig. 4 Qualitative comparison of different methods on FA, MD, AD, and RD. The bottom of the residual maps shows the MAD of each image compared to the reference brain image (excluding CSF). CSF, cerebrospinal fluid; MAD, mean absolute difference; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity.

WLS, being the most affected by noise, is the worst among all methods. DeepDTI also suffers from noise due to its use of traditional methods for DT fitting, which are sensitive to noise. TransDTI, which requires splitting data into tiny blocks for learning, exhibits grid effects that degrade reconstruction quality and visualization. Notably, DIFFnet fails to estimate high-quality DTs with only six DW images.

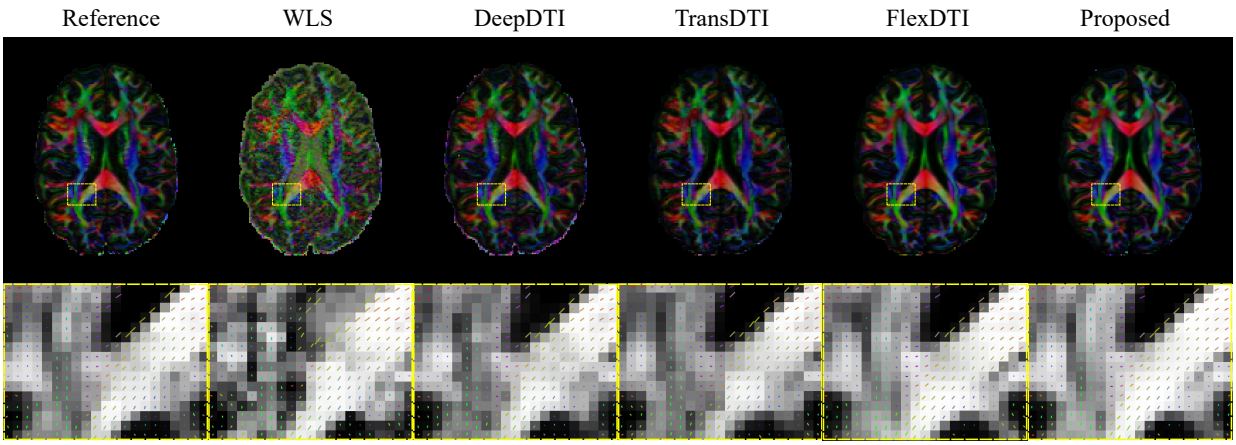

Fig. 5 Detailed anatomical comparison between different methods. The FA maps are color-coded according to the primary eigenvector directions (red: left–right; green: anterior–posterior; blue: superior–inferior) and are derived using all 18 non-DW images and 90 DW images (reference). The results using the original data (including one non-DW image and six DW images) are based on WLS, DeepDTI, TransDTI, and our method. Regions of interest are highlighted with yellow boxes and shown in an enlarged view. The primary eigenvectors are presented as color-coded sticks overlaid on the FA maps. DW, diffusion-weighted; FA, fractional anisotropy; WLS, weighted least squares.

Unedited

Although FlexDTI can estimate high-quality DTs with flexible q-space sampling, it still lags behind our method in visualizing finer details, especially in complex brain regions.

In contrast, our method accommodates flexible diffusion schemes and reconstructs high-quality DTs using just six DW images. Our method shows less noise and is more similar to the reference images in terms of detail than other methods. The residual maps at the bottom display the MAD of the quantitative parameters generated by different methods, demonstrating that our method outperforms others in visual appearance and quality.

Fig. 5 displays the FA maps, which are color-encoded by the primary eigenvector. Regarding visual quality, our method's result is more accurate than those of other methods, as they more closely match the reference and exhibit less noise. A small region is magnified to provide a more detailed comparison, as indicated by the yellow box in Fig. 5. In this magnified view, the primary fiber directions estimated by our method (shown as sticks) are as coherent as the reference. In contrast, the other four methods contain highly unrelated fibers.

In addition to voxel-wise performance comparisons, this study also performed tractography analysis. Fig. 6 displays whole-brain tractography results obtained from diffusion tensors using the FACT fiber-tracking algorithm and 20 representative major white-matter tracts identified using the Automated Fiber Quantification (AFQ) software (`https://github.com/yeatmanlab/AFQ`). Compared to the WLS method, the visual results from DeepDTI, TransDTI, FlexDTI, and the proposed method are closer to the reference. Among these methods, our method exhibits superior visual quality. In the results for major white-matter tracts, the visual quality of each white-matter tract from our method is more similar to the reference. In contrast, other methods miss some tract bundles due to estimation errors. Notably, our method of estimation of the Arcuate white-matter tract is consistent with the reference, whereas DeepDTI, TransDTI, and FlexDTI show estimation errors.

To more intuitively demonstrate the differences between methods in tractography based on estimated DTs, we performed a quantitative analysis using dice scores to measure the similarity between each method and the ground-truth (as shown in Fig. 7). The performance of the WLS method was significantly poorer, revealing substantial errors in its DT estimation. In contrast, DeepDTI, TransDTI, and FlexDTI showed superior performance in certain regions, but their accuracy and stability were still inferior to our method. Our method outperformed these methods across all regions, particularly in the inferior fronto-occipital fasciculus (IFOF), uncinate fasciculus, and arcuate fasciculus, where the

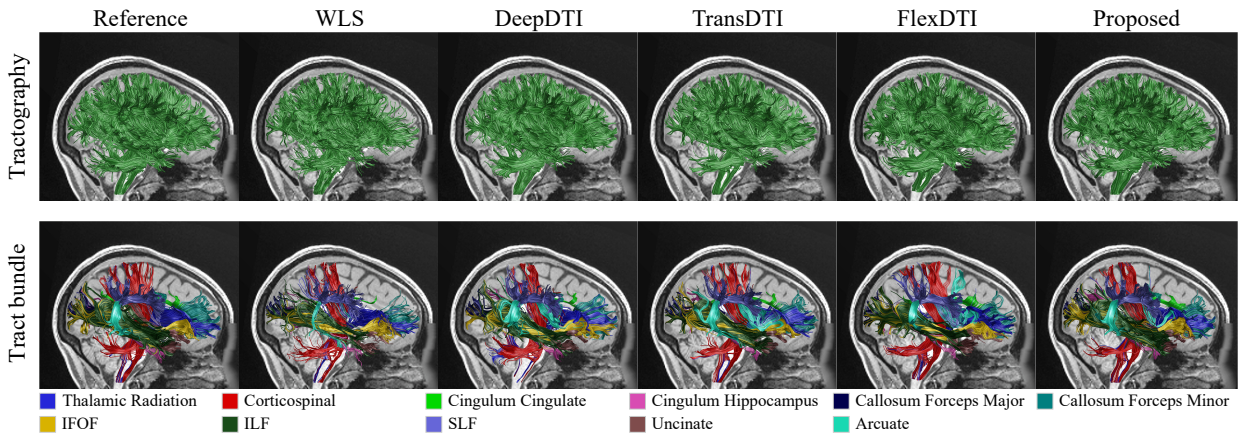

Fig. 6 Whole-brain DTI-based tractography results and major white-matter tracts generated by different methods using six DW images. Whole-brain DTI-based tractography images are generated from DTs using the FACT fiber-tracking algorithm. The results illustrate the left hemisphere of a representative subject. DTI, diffusion tensor imaging; DTs, diffusion tensors; DW, diffusion-weighted; IFOF, inferior fronto-occipital fasciculus; ILF, inferior longitudinal fasciculus; SLF, superior longitudinal fasciculus.

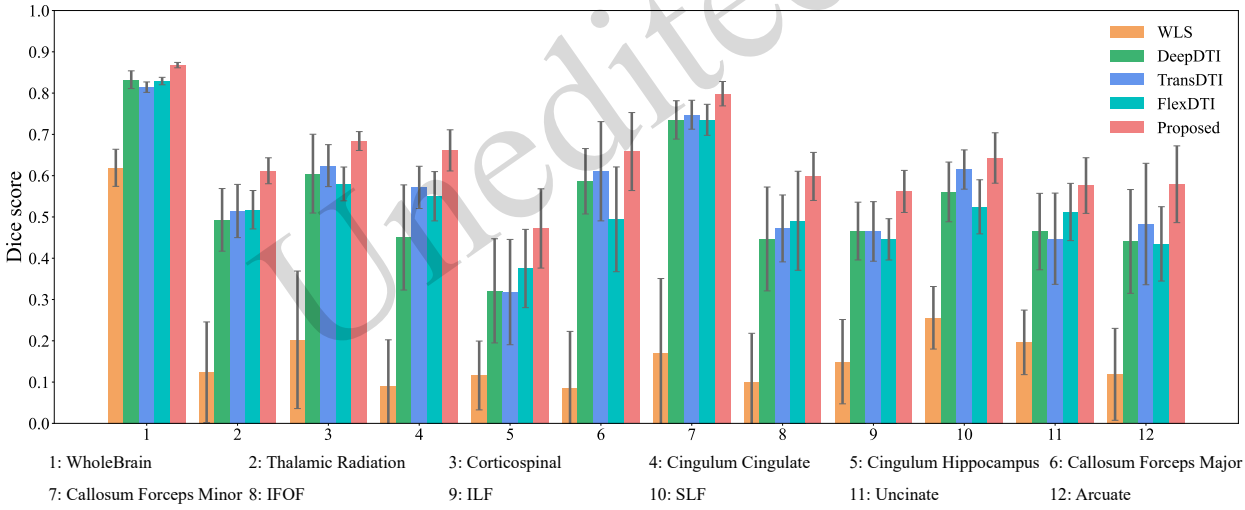

Fig. 7 Quantification of the accuracy of DTI-based tractography. The across-subject mean and standard deviation of the dice score of whole-brain DTI-based tractography results and major white-matter tracts are shown. The similarity metric compared binary images derived from the streamlines by assigning one to a voxel if >5 streamlines pass through it. The similarity between reference and estimated binary images is assessed using dice overlap. DTI, diffusion tensor imaging; IFOF, inferior fronto-occipital fasciculus; ILF, inferior longitudinal fasciculus; SLF, superior longitudinal fasciculus.

Unedited

dice scores were significantly higher, highlighting the clear advantage of our method in DT estimation accuracy. Therefore, our method demonstrated exceptional performance on the flexible q-space sampling dataset, providing strong support for more precise and stable tractography.

4.6 Evaluation Experiments on dHCP dataset

To validate the applicability of our method on other datasets, we conducted a quantitative comparison of DTs and their derived parameters obtained by different methods on the dHCP dataset, as shown in Table 2. DeepDTI and TransDTI were trained and tested using datasets with fixed q-space sampling schemes, while WLS, DIFFnet, FlexDTI, and our method were trained and tested using datasets with flexible q-space sampling schemes. The experimental results indicate that our method outperforms all other methods across all metrics. Notably, DeepDTI fails to estimate high-quality DTs accurately due to

Table 2 Quantitative comparison of different methods on the dHCP dataset*

Method	Tensor ↓	FA ↓	MD ↓	AD↓	RD ↓	Angle ↓
WLS	1.904±0.304	0.2668±0.0340	0.2264±0.0488	0.6030±0.1778	0.3156±0.0569	49.89±1.67
DeepDTI	1.710±0.727	0.1820±0.0951	0.3277±0.1700	0.5326±0.2584	0.3452±0.1866	47.62±4.55
TransDTI	0.760±0.060	0.0799±0.0051	0.1468±0.0163	0.1996±0.0150	0.1554±0.0178	43.86±1.74
DiFFnet	—	0.1034±0.0298	0.4238±0.0712	0.3600±0.0897	0.3570±0.0891	—
FlexDTI	0.695±0.058	**0.0593±0.0029**	0.1420±0.0162	0.1737±0.0168	0.1467±0.0152	36.40±1.53
Proposed	**0.646±0.041**	0.0595±0.0039	**0.1296±0.0103**	**0.1627±0.0104**	**0.1344±0.0116**	**35.72±1.37**

* Using eight test subjects from the dHCP dataset. Results are presented as mean ± standard deviation. The best results are highlighted in **bold**, and the second-best results are underscored. dHCP, developing human connectome project; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity.

Table 3 Quantitative comparison of different methods on the MDM dataset*

Method	Tensor ↓	FA ↓	MD ↓	AD↓	RD ↓	Angle ↓
WLS	0.730±0.130	0.1335±0.0229	0.1391±0.0423	0.1248±0.0161	0.1716±0.0374	40.73±2.96
DeepDTI	0.394±0.052	0.0573±0.0046	0.0626±0.0163	0.0952±0.0167	0.0629±0.0130	33.84±2.08
TransDTI	0.367±0.018	0.0578±0.0025	**0.0551±0.0028**	0.0897±0.0044	**0.0606±0.0028**	**32.12±1.36**
DiFFnet	—	0.2238±0.0617	0.3932±0.0165	0.3709±0.0723	0.3741±0.0301	—
FlexDTI	0.402±0.009	0.0592±0.0011	0.0644±0.0049	0.1021±0.0064	0.0690±0.0026	32.91±1.03
Proposed	**0.362±0.015**	**0.0536±0.0014**	0.0601±0.0037	**0.0887±0.0041**	0.0626±0.0039	33.03±0.76

* Using three test subjects from the MDM dataset. Results are presented as mean ± standard deviation. The best results are highlighted in **bold**, and the second-best results are underscored. MDM, multishell diffusion MRI; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity.

Uncorrected

its reliance on noise-sensitive traditional methods for DT fitting. Although TransDTI can estimate high-quality DTs, its performance on DW images with flexible q-space sampling schemes is suboptimal. DIFFnet can estimate DTs from DW images with flexible q-space sampling schemes but does not account for the correlation between adjacent voxels, reducing the estimation quality. FlexDTI outperforms all other methods except ours across all metrics, primarily due to the incorporation of diffusion gradient directions. However, its neglect to account for the influence of b-values prevents it from surpassing our method in overall performance. To more intuitively demonstrate the performance differences between different methods on the dHCP dataset, we present the qualitative comparison results in Fig. S1 of the Supplementary materials. Our method exhibits the most minor errors on all metrics' residual maps compared to other methods and provides more accurate detail reconstruction.

4.7 Evaluation experiments on MDM dataset

To verify the generalizability of our method across different datasets, we performed a quantitative comparison of DTs and their derived parameters obtained from various methods trained on the HCP dataset when applied to the MDM dataset. The results are shown in Table 3. Here, DeepDTI and TransDTI were trained using the fixed q-space sampling scheme of the HCP dataset, while WLS, DIFFnet, FlexDTI, and our method were trained and tested using the flexible q-space sampling scheme of the HCP dataset. The experimental results indicate that our method achieves optimal or near-optimal performance for DTs, FA, AD, and RD metrics. Although our method did not reach the best performance in some metrics, its effectiveness on datasets with flexible q-space sampling is almost comparable to methods trained with the fixed q-space sampling scheme. In addition to the quantitative results mentioned earlier, qualitative comparisons are also provided in the Supplementary materials (as shown in Fig. S2) to visually demonstrate the detail reconstruction capability of the proposed method. Notably, our method shows more pronounced details in the FA metric.

4.8 Computational Cost Analysis

Table 4 compares the computational costs among different methods, including model FLOPS,

Table 4 Comparative results of computational costs*

Method	Model FLOPS (G)	Trainable parameters (M)	Inference time (ms)
WLS	—	—	69811.68
DeepDTI	941.47	3.59	253.90
TransDTI	159.95	39.08	205.47
FlexDTI	60.99	17.24	400.50
Proposed	488.30	19.28	524.37

* All evaluations were conducted on the A100 GPU.

trainable parameters, and inference time. Although the inference time of our method is slightly longer than that of FlexDTI, this discrepancy mainly arises from the more complex model structure used in our approach, which facilitates flexible DT estimation and, in turn, achieves higher accuracy and stability.

4.9 Ablation experiments

We conducted a comprehensive ablation experiment on QCG-DTI using the HCP dataset to validate the effectiveness of our method, as shown in Table 5. We use the model without the QCF and q-space-coordinate-embedded-feature consistency strategy as the baseline model.This baseline model concatenates DW images along the channel dimension as input and uses MRD as the basic block.

In the ablation experiment of the QCF module, we removed QCF from the model. The results showed a significant drop in model performance, particularly in terms of FA, due to the absence of q-space-coordinate embedding. Additionally, the network's robustness was affected.

In the ablation study of the q-space-coordinate-embedded-feature consistency strategy, we used input data concatenated along the channel dimension. Although q-space coordinates were included, the performance did not improve significantly compared to the baseline. This indicates that the q-space-coordinate-embedded-feature consistency strategy helps the model learn the correspondence between each q-space coordinate and its corresponding DW image, thereby enhancing the quality of DT estimation.

In the ablation study of the MRD module, we replaced MRD with residual network (ResNet) blocks having the same parameters. The results demonstrated that MRD significantly outperformed the ResNet blocks. This is because MRD effectively utilizes multiscale information from the images, enhancing both the feature extraction capability from DW images and the DT reconstruction ability. This further confirms the effectiveness of QCF and the q-space-coordinate-embedded-feature consistency strategy, even when using ResNet blocks as the base blocks.

4.10 Discussion and Future Work

The experimental results demonstrate that the QCG-DTI method outperforms traditional tensor model fitting methods and other state-of-the-art tensor estimation algorithms. Our model excels in terms of accuracy, stability, and high reconstruction quality when handling complex fiber bundle structures. These advantages are attributed to the well-designed q-space-coordinate-embedded-feature consistency strategy, the QCF module, and MRD module.

Although the method performs excellently in the context of DT estimation, computational efficiency remains a challenge that needs to be addressed. To further promote its application in clinical and neuroscience fields, future research will focus on enhancing the model's computational efficiency by exploring more lightweight network architectures. Additionally, improving the model's robustness is crucial for future work. Image moments and spherical harmonics, as alternative feature representations of images, have been widely applied in deep learning and medical image processing. As stable image features, image moments can effectively enhance the model's adaptability to complex environments (Wang et al., 022a,b; Liu et al., 2024). At the same time, spherical harmonics (Ha and Lyu, 2022), with their unique advantages in dMRI data processing, have been successfully applied in DT estimation tasks (Chen et al., 2025). Future research will explore effectively integrating these two image features into the existing framework to improve the model's estimation accuracy and stability while enhancing its robustness across diverse datasets.

5 Conclusion

This paper proposes a QCG-DTI method for efficient DT estimation using DW images with flexible q-space sampling schemes. To overcome the dependence on fixed q-space sampling schemes, we first ensure the correspondence between each q-space coordinate and its corresponding DW image through

Unedited

Table 5 Ablation experiments*

Method	Tensor ↓	FA ↓	MD ↓	AD↓	RD ↓	Angle ↓
Baseline	0.357±0.034	0.0531±0.0036	0.0624±0.0117	0.0877±0.0078	0.0677±0.0121	29.48±0.97
w/o QCF	0.326±0.022	0.0509±0.0032	0.0513±0.0054	0.0790±0.0044	0.0568±0.0059	29.79±1.61
w/o strategy	0.354±0.032	0.0524±0.0033	0.0600±0.0112	0.0842±0.0075	0.0667±0.0106	29.61±0.96
w/o MRD	0.313±0.021	0.0479±0.0031	0.0525±0.0061	0.0784±0.0051	0.0568±0.0068	27.15±0.97
Proposed	**0.289±0.019**	**0.0443±0.0027**	**0.0499±0.0060**	**0.0706±0.0045**	**0.0546±0.0062**	**25.67±0.78**

* A lower value indicates better performance. Results are presented as mean ± standard deviation. The best results are highlighted in bold, and the second-best results are underscored. QCF, q-space-coordinate fusion module; MRD, multiscale feature residual dense module; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity.

a q-space-coordinate-embedded-feature consistency strategy, thereby improving the quality of DT estimation. Next, through the QCF module, q-space coordinates are efficiently embedded into DW images to estimate DTs. Finally, the proposed MRD module effectively extracts multiscale features from DW images, enhancing feature extraction and DT reconstruction capability. Extensive quantitative and qualitative experiments demonstrate that the proposed algorithm outperforms traditional tensor model fitting methods and other state-of-the-art tensor estimation algorithms. The proposed algorithm is expected to be applied in various clinical and neuroscience research settings where flexible, rapid, and efficient DT estimation is required.

Unedited

Contributors

Maokun ZHENG designed the research. Maokun ZHENG and Dan-dan LI processed the data. Maokun ZHENG drafted the paper. Zhi LI and Guomei WANG helped organize the paper. Zhi LI, Long ZHENG and Weidong WANG revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Alexander AL, Lee JE, Lazar M, et al., 2007. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316-329. https://doi.org/10.1016/j.nurt.2007.05.011

Aliotta E, Nourzadeh H, Sanders J, et al., 2019. Highly accelerated, model-free diffusion tensor mri reconstruction using neural networks. *Med Phys*, 46(4):1581-1591. https://doi.org/10.1002/mp.13400

de Almeida Martins JP, Nilsson M, Lampinen B, et al., 2021. Neural networks for parameter estimation in microstructural mri: Application to a diffusion-relaxation model of white matter. *NeuroImage*, 244:118601. https://doi.org/10.1016/j.neuroimage.2021.118601

Barnea-Goraly N, Kwon H, Menon V, et al., 2004. White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biol Psychiat*, 55(3):323-326. https://doi.org/10.1016/j.biopsych.2003.10.022

Barnea-Goraly N, Menon V, Eckert M, et al., 2005. White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cereb Cortex*, 15(12):1848-1854. https://doi.org/10.1093/cercor/bhi062

Basser PJ, Pajevic S, Pierpaoli C, et al., 2000. In vivo fiber tractography using dt-mri data. *Magn Reson Med*, 44(4):625-632. https://doi.org/10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O

Bastiani M, Andersson JLR, Cordero-Grande L, et al., 2019. Automated processing pipeline for neonatal diffusion mri in the developing human connectome project. *NeuroImage*, 185:750-763. https://doi.org/10.1016/j.neuroimage.2018.05.064

Caruyer E, Lenglet C, Sapiro G, et al., 2013. Design of multishell sampling schemes with uniform coverage in diffusion mri. *Magn Reson Med*, 69(6):1534-1540. https://doi.org/10.1002/mrm.24736

Chen YW, Li JL, Lu QQ, et al., 2025. Spherical harmonics-based deep learning achieves generalized and accurate diffusion tensor imaging. *IEEE J Biomed Health Inform*, 29(1):456-461. https://doi.org/10.1109/JBHI.2024.3471769

Dale AM, Fischl B, Sereno MI, 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179-194. https://doi.org/10.1006/nimg.1998.0395

Donald Tournier J, Calamante F, Connelly A, 2012. Mrtrix: diffusion tractography in crossing fiber regions. *Int J Imaging Syst Technol*, 22(1):53-66. https://doi.org/10.1002/ima.22005

Elam JS, Glasser MF, Harms MP, et al., 2021. The human connectome project: a retrospective. *NeuroImage*, 244:118543. https://doi.org/10.1016/j.neuroimage.2021.118543

Eriksson SH, Rugg-Gunn FJ, Symms MR, et al., 2001. Diffusion tensor imaging in patients with epilepsy and malformations of cortical development. *Brain*, 124(3):617-626. https://doi.org/10.1093/brain/124.3.617

Fischl B, Sereno MI, Dale AM, 1999. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195-207. https://doi.org/10.1006/nimg.1998.0396

Fischl B, 2012. Freesurfer. *NeuroImage*, 62(2):774-781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Garyfallidis E, Brett M, Amirbekian B, et al., 2014. Dipy, a library for the analysis of diffusion mri data. *Front Neuroinform*, 8:8. https://doi.org/10.3389/fninf.2014.00008

Gibbons EK, Hodgson KK, Chaudhari AS, et al., 2019. Simultaneous noddi and gfa parameter map generation from subsampled q-space imaging using deep learning. *Magn Reson Med*, 81(4):2399-2411. https://doi.org/10.1002/mrm.27568

Golkov V, Dosovitskiy A, Sperl JI, et al., 2016. Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE Trans Med Imaging*, 35(5):1344-1351. https://doi.org/10.1109/TMI.2016.2551324

Ha S, Lyu I, 2022. Spharm-net: Spherical harmonics-based convolution for cortical parcellation. *IEEE Trans Med Imaging*, 41(10):2739-2751. https://doi.org/10.1109/TMI.2022.3168670

Jones DK, Horsfield MA, Simmons A, 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med*, 42(3):515-525. https://doi.org/10.1002/(SICI)1522-2594(199909)42:3<515::AID-MRM14>3.0.CO;2-Q

Jones DK, Knösche TR, Turner R, 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri. *NeuroImage*, 73:239-254. https://doi.org/10.1016/j.neuroimage.2012.06.081

Jones DK, 2004. The effect of gradient sampling schemes on measures derived from diffusion tensor mri: a monte carlo study. *Magn Reson Med*, 51(4):807-815. https://doi.org/10.1002/mrm.20033

Jones DK, 2009. Chapter 3 - Gaussian modeling of the diffusion signal. In: Johansen-Berg H, Behrens TEJ (Eds.), Diffusion MRI: From Quantitative Measurement to in Vivo Neuroanatomy. Elsevier, Amsterdam, p.37-54.

Karimi D, Gholipour A, 2022. Diffusion tensor estimation with transformer neural networks. *Artif Intell Med*, 130:102330. https://doi.org/10.1016/j.artmed.2022.102330

Kingma DP, Ba LJ, 2015. Adam: A method for stochastic optimization. Int Conf on Learning Represent (ICLR), article 13, https://hdl.handle.net/11245/1.505367

Koppers S, Friedrichs M, Merhof D, 2017. Reconstruction of diffusion anisotropies using 3d deep convolutional neural networks in diffusion imaging. In: Schultz T, Özarslan E, Hotz I (Eds.), Modeling, Analysis, and Visualization of Anisotropy. Springer, Cham, p.393-404. https://doi.org/10.1007/978-3-319-61358-1_17

Le Bihan D, Mangin JF, Poupon C, et al., 2001. Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging*, 13(4):534-546. https://doi.org/10.1002/jmri.1076

Lebel C, Walker L, Leemans A, et al., 2008. Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40(3):1044-1055. https://doi.org/10.1016/j.neuroimage.2007.12.053

Li HY, Liang ZF, Zhang CY, et al., 2021. Superdti: Ultrafast dti and fiber tractography with deep learning. *Magn Reson Med*, 86(6):3334-3347. https://doi.org/10.1002/mrm.28937

Li ZW, Gong T, Lin ZC, et al., 2019. Fast and robust diffusion kurtosis parametric mapping using a three-dimensional convolutional neural network. *IEEE Access*, 7:71398-71411. https://doi.org/10.1109/ACCESS.2019.2919241

Lin ZC, Gong T, Wang KW, et al., 2019. Fast learning of fiber orientation distribution function for mr tractography using convolutional neural network. *Med Phys*, 46(7):3101-3116. https://doi.org/10.1002/mp.13555

Liu YN, Wang CP, Lu MY, et al., 2024. From simple to complex scenes: Learning robust feature representations for accurate human parsing. *IEEE Trans Pattern Anal Mach Intell*, 46(8):5449-5462. https://doi.org/10.1109/TPAMI.2024.3366769

Mori S, Crain BJ, Chacko VP, et al., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol*, 45(2):265-269. https://doi.org/10.1002/1531-8249(199902)45:23.0.CO;2-3

Nath V, Schilling KG, Parvathaneni P, et al., 2019. Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted mri. *Magn Reson Imaging*, 62:220-227. https://doi.org/10.1016/j.mri.2019.07.012

Park J, Jung W, Choi EJ, et al., 2022. Diffnet: diffusion parameter mapping network generalized for input diffusion gradient schemes and b-value. *IEEE Trans Med Imaging*, 41(2):491-499. https://doi.org/10.1109/TMI.2021.3116298

Pierpaoli C, Jezzard P, Basser PJ, et al., 1996. Diffusion tensor mr imaging of the human brain. *Radiology*, 201(3):637-648. https://doi.org/10.1148/radiology.201.3.8939209

Poulin P, Côté MA, Houde JC, et al., 2017. Learn to track: deep learning for tractography. In: Descoteaux M, Maier-Hein L, Franz A, et al. (Eds.), Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Springer, Cham, p.540-547. https://doi.org/10.1007/978-3-319-66182-7_62

Qin H, Yang C, Zhu X, et al., 2021. Dynamic receptive field adaptation for attention-based text recognition. In: Lladós J, Lopresti D, Uchida S (Eds.), 16th International Conference on Document Analysis and Recognition, ICDAR 2021. Springer, Berlin, Heidelberg, 12822:225-239.

Roosendaal SD, Geurts JJG, Vrenken H, et al., 2009. Regional dti differences in multiple sclerosis patients. *NeuroImage*, 44(4):1397-1403. https://doi.org/10.1016/j.neuroimage.2008.10.026

Skare S, Hedehus M, Moseley ME, et al., 2000. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with mri. *J Magn Reson*, 147(2):340-352. https://doi.org/10.1006/jmre.2000.2209

Stejskal EO, Tanner JE, 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*, 42(1):288-292. https://doi.org/10.1063/1.1695690

Unedited

Sutskever I, Vinyals O, Le QV, 2014. Sequence to sequence learning with neural networks. Proc 28th Int Conf on Neural Information Proc Syst - Vol 2, p.3104-3112.

Tian QY, Bilgic B, Fan QY, et al., 2020. Deepdti: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage*, 219:117017. https://doi.org/10.1016/j.neuroimage.2020.117017

Tong QQ, He HJ, Gong T, et al., 2019. Reproducibility of multi-shell diffusion tractography on traveling subjects: a multicenter study prospective. *Magn Reson Imaging*, 59:1-9. https://doi.org/10.1016/j.mri.2019.02.011

Tong QQ, Gong T, He HJ, et al., 2020. A deep learning–based method for improving reliability of multicenter diffusion kurtosis imaging with varied acquisition protocols. *Magn Reson Imaging*, 73:31-44. https://doi.org/10.1016/j.mri.2020.08.001

Vaswani A, Shazeer N, Parmar N, et al., 2017. Attention is all you need. Proc 31st Int Conf on Neural Inf Proc Syst, p.6000-6010.

Wang CP, Ma B, Xia ZQ, et al., 2022a. Stereoscopic image description with trinion fractional-order continuous orthogonal moments. *IEEE Trans Circ Syst Vid Technol*, 32(4):1998-2012. https://doi.org/10.1109/TCSVT.2021.3094882

Wang CP, Hao QX, Xu SJ, et al., 2022b. Rd-iwan: Residual dense based imperceptible watermark attack network. *IEEE Trans Circ Syst Vid Technol*, 32(11):7460-7472. https://doi.org/10.1109/TCSVT.2022.3188524

Westin CF, Maier SE, Mamata H, et al., 2002. Processing and visualization for diffusion tensor mri. *Med Image Anal*, 6(2):93-108. https://doi.org/10.1016/s1361-8415(02)00053-1

Wu YC, Alexander AL, 2007. Hybrid diffusion imaging. *NeuroImage*, 36(3):617-629. https://doi.org/10.1016/j.neuroimage.2007.02.050

Wu ZJ, Wang JC, Chen ZQ, et al., 2024. Flexdti: flexible diffusion gradient encoding scheme-based highly efficient diffusion tensor imaging using deep learning. *Phys Med Biol*, 69(11):115012. https://doi.org/10.1088/1361-6560/ad45a5

List of supplementary materials

Fig. S1 Qualitative comparison of different methods on the dHCP dataset. dHCP, developing human connectome project.

Fig. S2 Qualitative comparison of different methods on the MDM dataset. MDM, multishell diffusion MRI; MRI, magnetic resonance imaging.

Unedited