Supplementary materials

List of all tracts

This is a list of all 72 tracts supported by our model: Arcuate fascicle (AF), Anterior thalamic radiation (ATR), Anterior commissure (CA), Corpus callosum (Rostrum (CC 1), Genu (CC 2), Rostral body (CC 3), Anterior midbody (CC 4), Posterior midbody (CC 5), Isthmus (CC 6), Splenium (CC 7)), Cingulum (CG), Corticospinal tract (CST), Middle longitudinal fascicle (MLF), Fronto-pontine tract (FPT), Fornix (FX), Inferior cerebellar peduncle (ICP), Inferior occipito-frontal fascicle (IFO), Inferior longitudinal fascicle (ILF), Middle cerebellar peduncle (MCP), Optic radiation (OR), Parieto-occipital pontine (POPT), Superior cerebellar peduncle (SCP), Superior longitudinal fascicle I (SLF I), Superior longitudinal fascicle II (SLF II), Superior longitudinal fascicle III (SLF III), Superior thalamic radiation (STR), Uncinate fascicle (UF), Thalamo-prefrontal (T_PREF), Thalamo-premotor (T_PREM), Thalamo-precentral (T_PREC), Thalamo-postcentral (T_POSTC), Thalamo-parietal (T_PAR), Thalamo-occipital (T_OCC), Striato-fronto-orbital (ST_FO), Striato-prefrontal (ST_PREF), Striato-premotor (ST_PREM), Striato-precentral (ST_PREC), Striato-postcentral (ST_POSTC), Striato-parietal (ST_PAR), Striato-occipital (ST_OCC)

RecoBundles default parameters

Parameters for whole brain streamline-based registration:

- streamline-based linear registration transform: affine
- streamline-based linear registration progressive: True
- maximum iterations for registration optimization: 150
- random streamlines for starting QuickBundles: 50000

Parameters for recognize bundles:

- clustering threshold: 15
- reduction threshold: 10
- reduction distance: mdf
- model clustering threshold: 5
- pruning threshold: 5
- pruning distance: mdf
- local streamline-based linear registration: True
- streamline-based linear registration metric: None
- streamline-based linear registration transform: similarity
- streamline-based linear registration progressive: True
- streamline-based linear registration matrix: small

Table 1. Acquisition parameters of additional test datasets.

Project	**Pathology**	**Resolution***	**b-Values**	**Field strength**
TRACED [1]	healthy	2.5mm	3x $b = 0mm/s^2$ 20x $b = 1000mm/s^2$ 48x $b = 2000mm/s^2$ 64x $b = 3000mm/s^2$	3T
Internal (Healthy)	healthy	2.5mm	1x $b = 0mm/s^2$ 81x $b = 1000mm/s^2$ 81x $b = 2000mm/s^2$ 81x $b = 3000mm/s^2$	3T
BrainGluShi (Bustillo et al., 2017) [2]	healthy	2.0mm	5x $b = 0mm/s^2$ 30x $b = 800mm/s^2$	3T
Stanford_hardi (Rokem et al., 2013) [3]	healthy	2.0mm	10x $b = 0mm/s^2$ 150x $b = 2000mm/s^2$	3T
Sherbrooke_3shell [4]	healthy	2.5mm	1x $b = 0mm/s^2$ 64x $b = 1000mm/s^2$ 64x $b = 2000mm/s^2$ 64x $b = 3500mm/s^2$	3T
Rockland (Nooner et al., 2012) [5]	healthy	2.0mm	9x $b = 0mm/s^2$ 128x $b = 1500mm/s^2$	3T
HCP 7T (Van Essen et al., 2013)	healthy	1.05mm	15 $b = 0mm/s^2$ 64x $b = 1000mm/s^2$ 64x $b = 2000mm/s^2$	7T
IXI [6]	healthy > 80 years	1.75x1.75x2mm	1 $b = 0mm/s^2$ 15x $b = 1000mm/s^2$	3T
HCP lifespan (Tisdall et al., 2012) [7]	healthy < 10 years	1.5mm	10 $b = 0mm/s^2$ 76x $b = 1000mm/s^2$ 75x $b = 2500mm/s^2$	3T
COBRE (Çetin et al., 2014) [2]	schizophrenia, enlarged ventricles	2.0mm	5x $b = 0mm/s^2$ 30x $b = 800mm/s^2$	3T
SoftSigns (Hirjak et al., 2017)	neurological soft signs	2.5mm	1x $b = 0mm/s^2$ 81x $b = 1000mm/s^2$	3T
Internal (Autism)	autism spectrum disorder	2.5mm	5x $b = 0mm/s^2$ 60x $b = 1000mm/s^2$	3T
Internal (Schizophrenia)	schizophrenia	1.7mm	3x $b = 0mm/s^2$ 60x $b = 1500mm/s^2$	3T
CNP (Poldrack et al., 2016) [8]	schizophrenia, bipolar, ADHD	2.0mm	1x $b = 0mm/s^2$ 64x $b = 1000mm/s^2$	3T
Internal (MS)	multiple sclerosis	1x1x2mm	2x $b = 0mm/s^2$ 64x $b = 1000mm/s^2$	3T
ADNI [9]	alzheimer	1.4x1.4x2.7mm	5x $b = 0mm/s^2$ 41x $b = 1000mm/s^2$	3T
OASIS [10]	alzheimer	2.5mm	1x $b = 0mm/s^2$ 64x $b = 1000mm/s^2$	3T

[1] https://my.vanderbilt.edu/ismrmtraced2017/
[2] http://schizconnect.org/
[3] https://purl.stanford.edu/yx282xq2090
[4] http://nipy.org/dipy/reference/dipy.data.html#fetch-sherbrooke-3shell
[5] http://fcon_1000.projects.nitrc.org/indi/enhanced/
[6] https://brain-development.org/ixi-dataset/
[7] https://www.humanconnectome.org/study-hcp-lifespan-pilot
[8] https://openneuro.org/datasets/ds000030/versions/00016
[9] http://adni.loni.usc.edu/data-samples/access-data/
[10] http://www.oasis-brains.org
* if only one value is shown, the resolution is isotropic

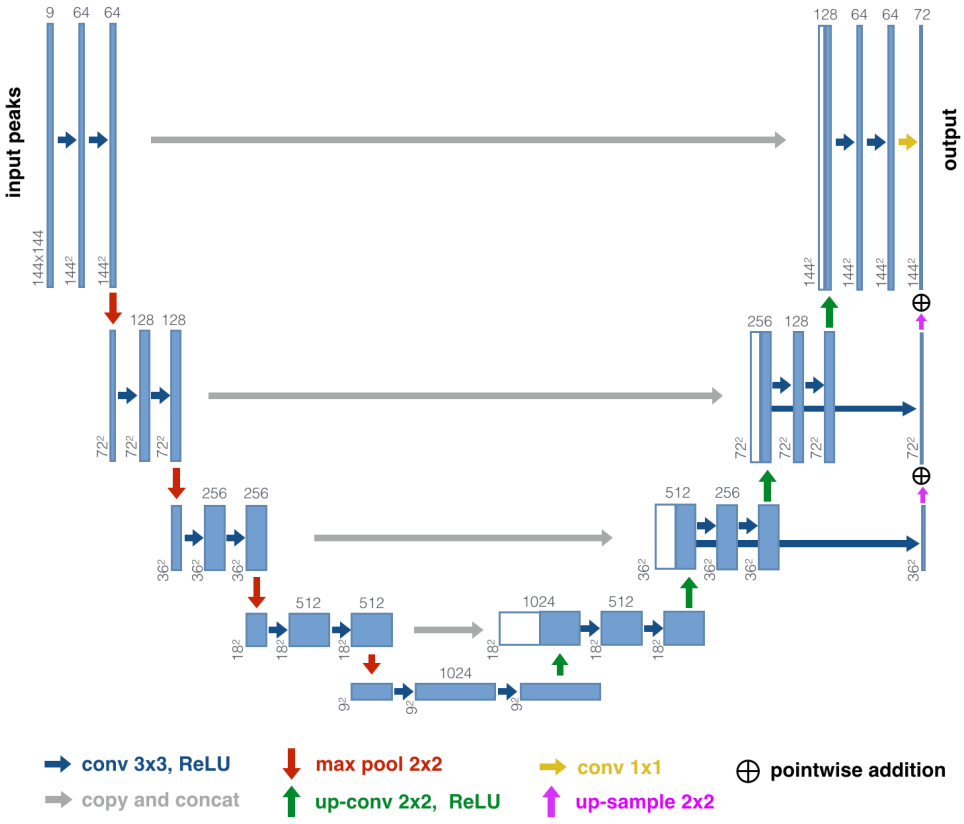

Fig. 1. Proposed U-Net architecture. Blue boxes represent multi-channel feature maps. White boxes show copied feature maps. The gray number on top of each box gives the number of channels, the x-y-size is given at lower left corner of each box. Network operations are represented by differently colored arrows. The main difference to the original U-Net architecture are the extra convolutions layers in the upsampling path allowing better gradient flow (deep supervision).

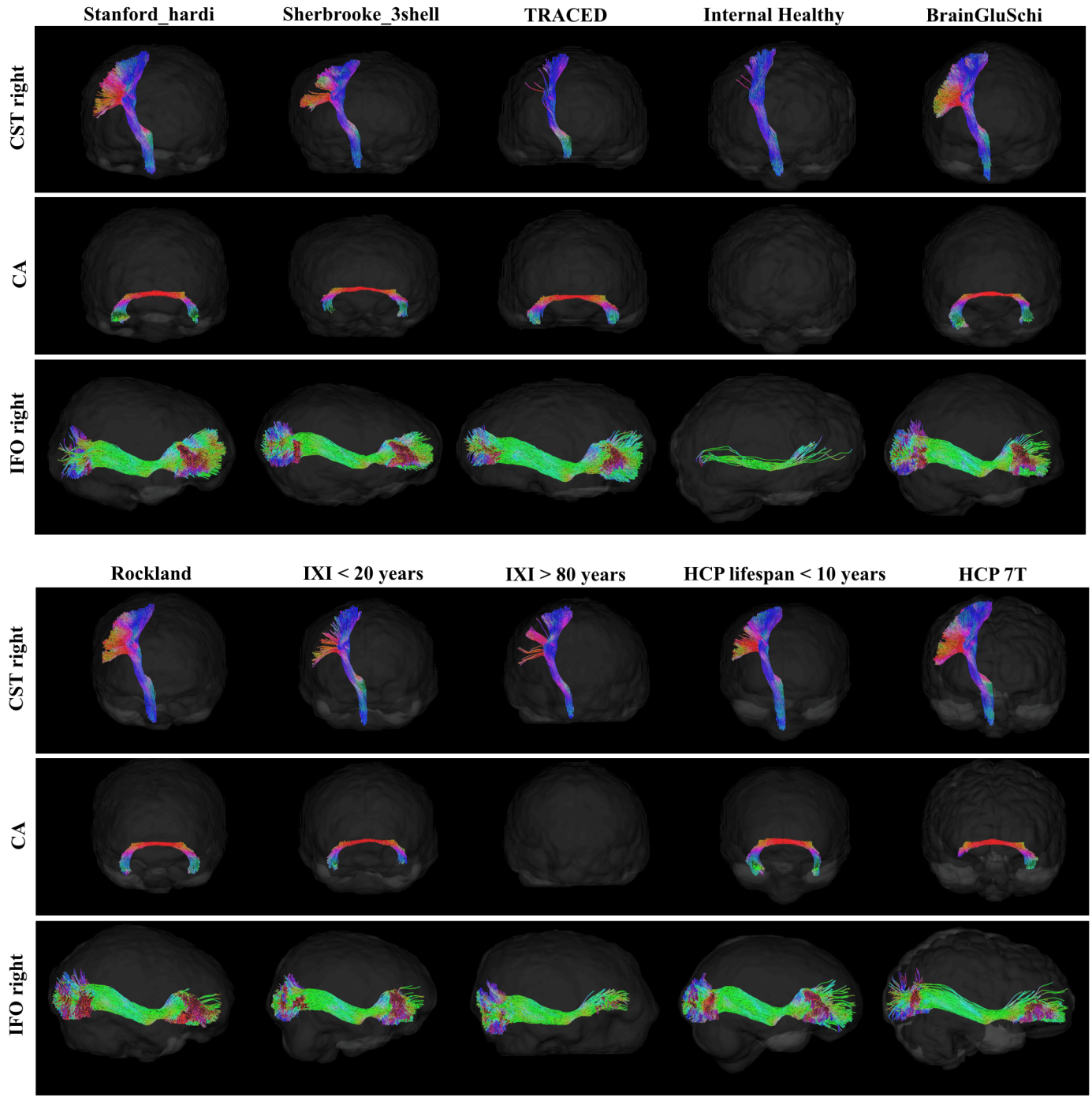

Fig. 2. Qualitative results of our proposed method on 10 healthy subjects from 9 different datasets (see table 1) while being trained on the HCP reference dataset: reconstruction of right corticospinal tract (CST), anterior commissure (CA) and right inferior occipito-frontal fascicle (IFO).

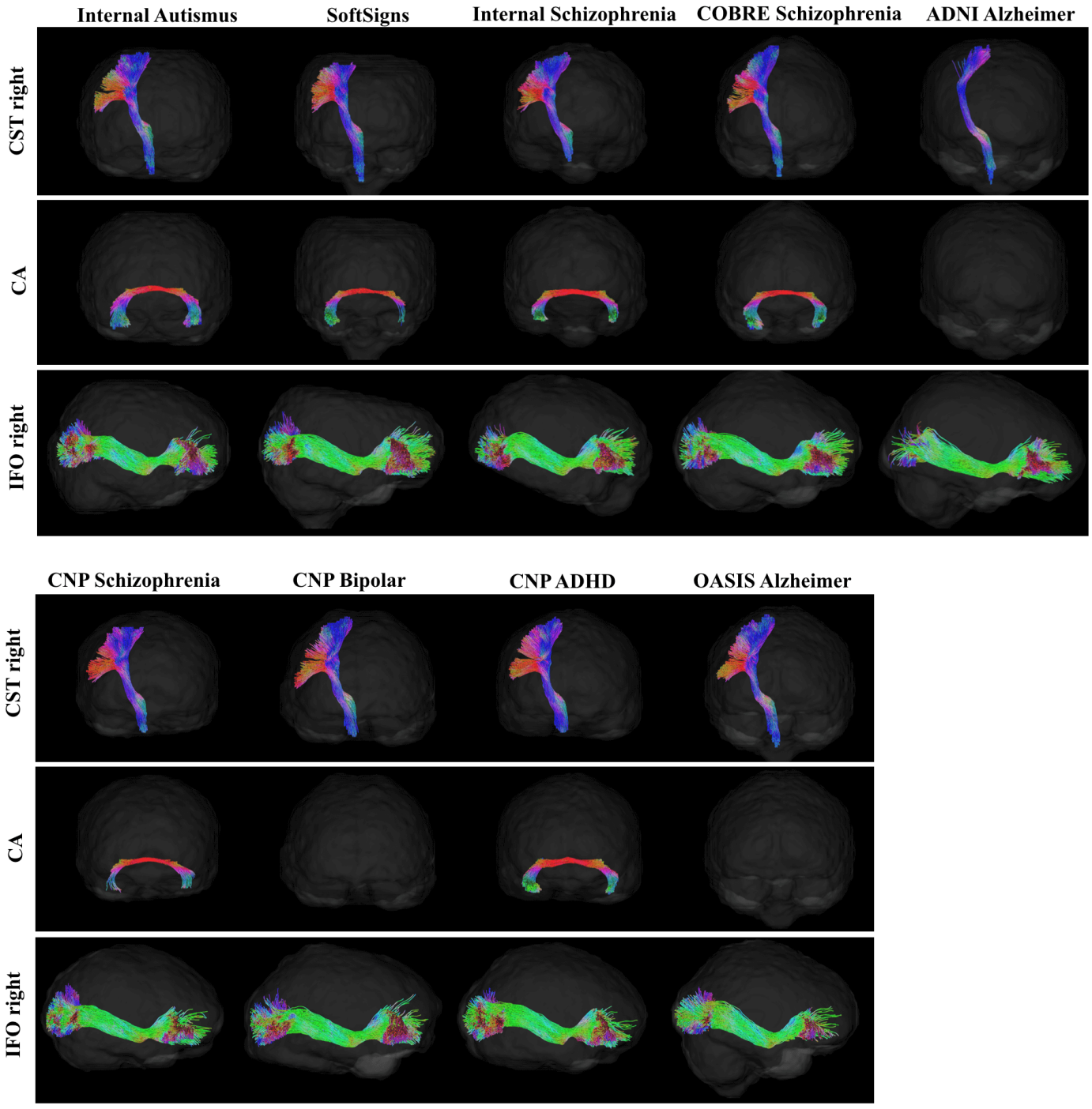

Fig. 3. Qualitative results of our proposed method on 9 subjects with pathologies from 7 different datasets (see table 1) while being trained on the HCP reference dataset: reconstruction of right corticospinal tract (CST), anterior commissure (CA) and right inferior occipito-frontal fascicle (IFO).

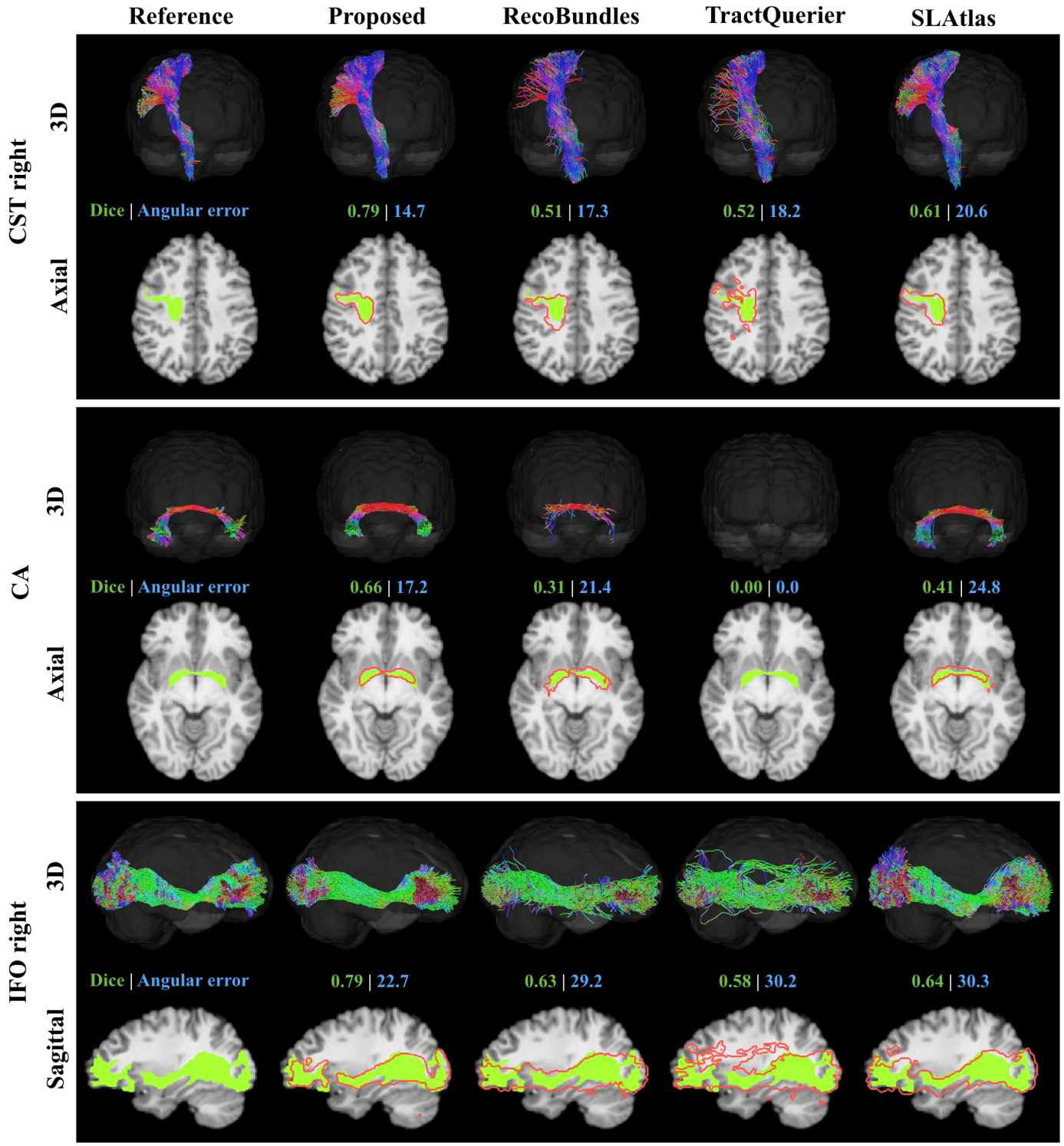

Fig. 4. Qualitative comparison of results on *Phantom* test set: reconstruction of right corticospinal tract (CST), anterior commissure (CA) and right inferior occipito-frontal fascicle (IFO) on subject 623844. Green shows the reference tract and red shows the tract mask of the respective method.

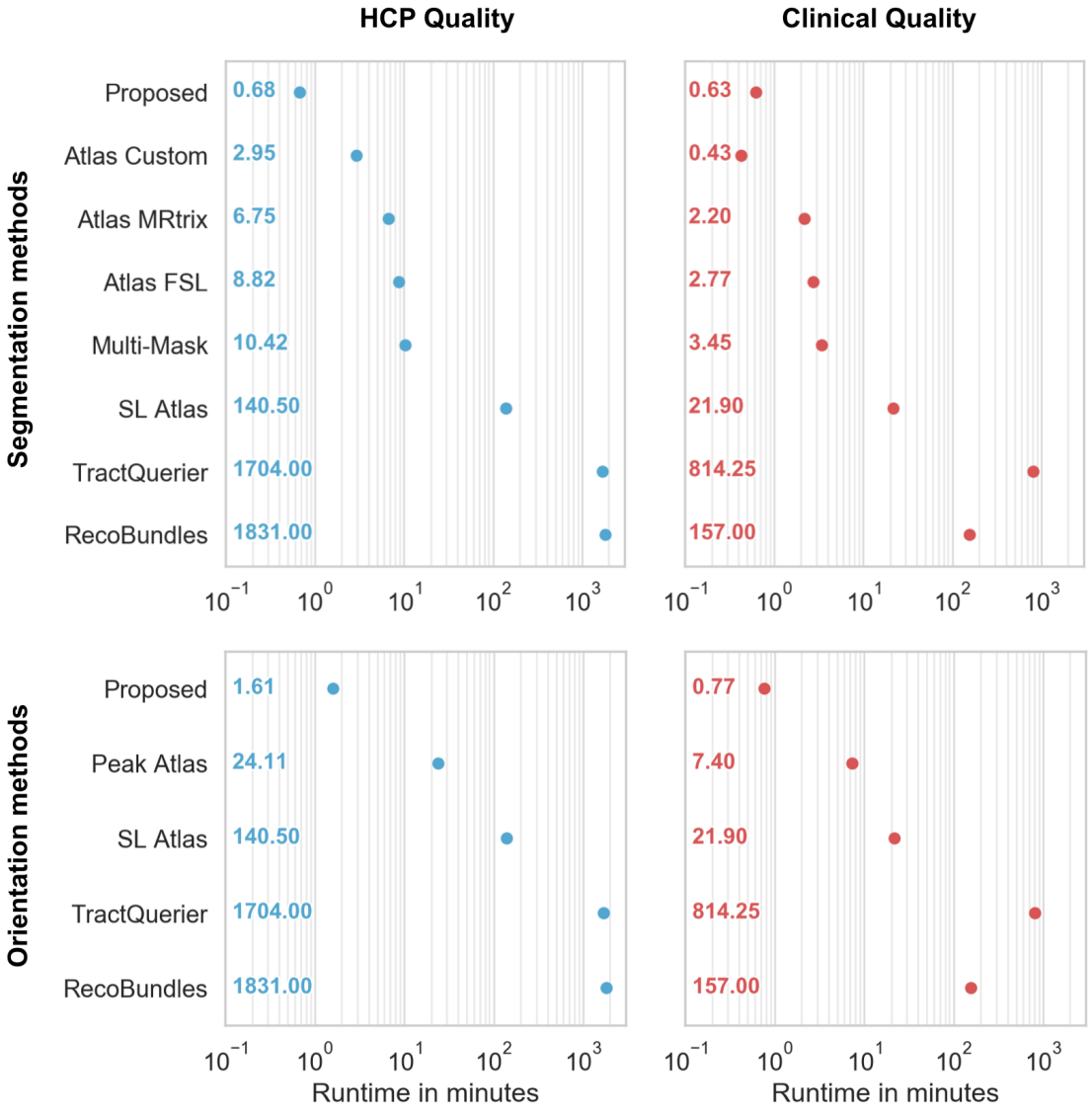

Fig. 5. Runtime (in minutes) of all methods for segmentation and for estimation of orientation shown in the quantitative evaluation.

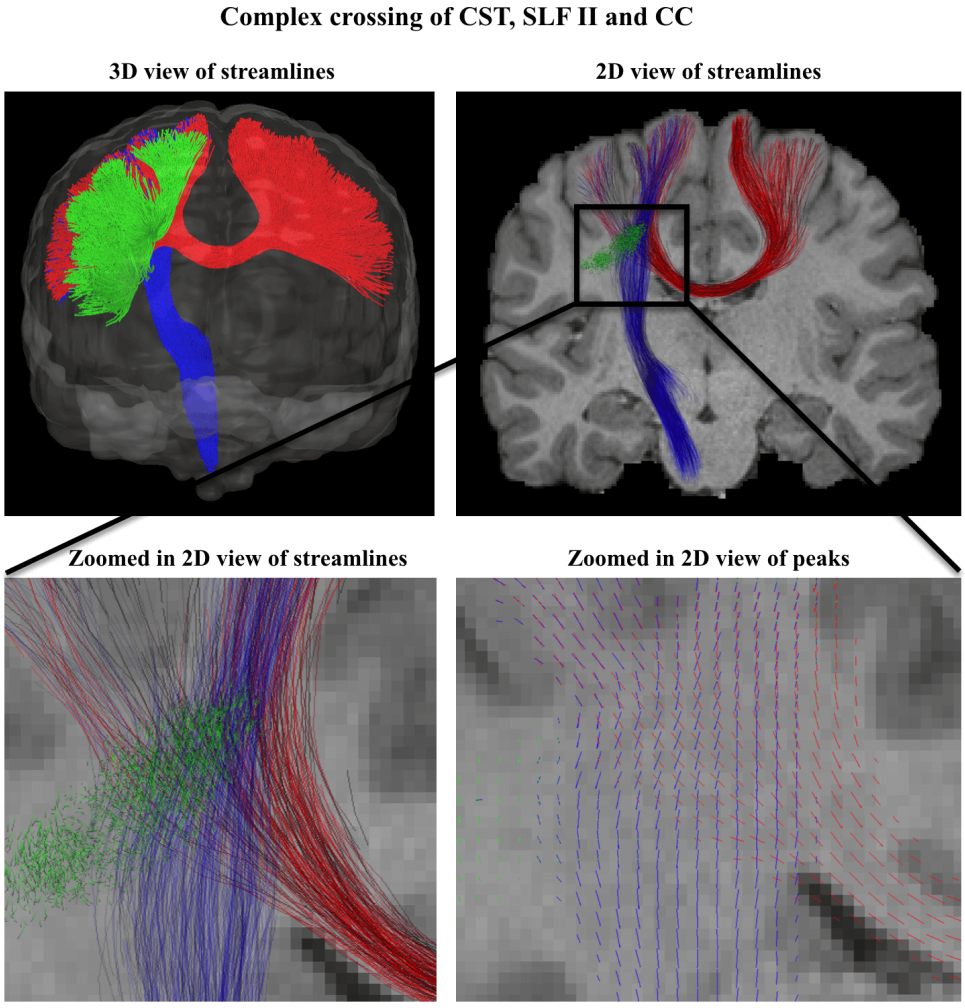

Fig. 6. The figure shows the results for the proposed method on the crossing of CST (blue), SLF II (green) and CC 4 (blue) on subject 623844 of the *HCP Quality* dataset. The zoomed in peak image shows the TOM peaks which the tracking is based on. The proposed method manages to properly resolve this complex crossing of three major tracts.

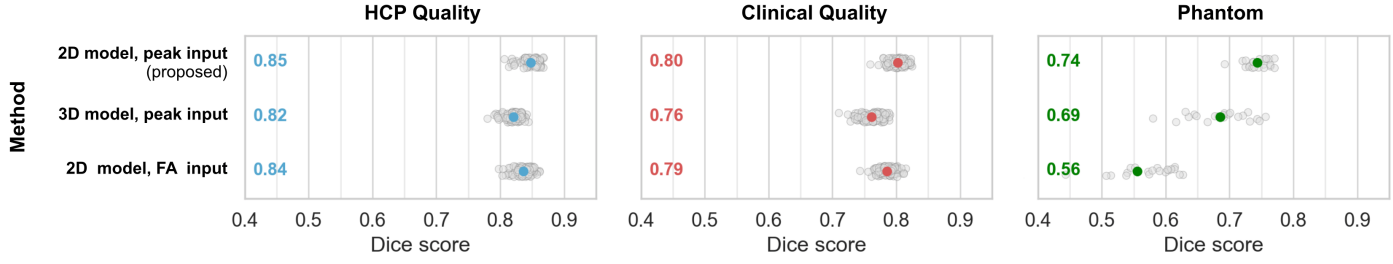

Fig. 7. Comparison of segmentation results for 2D and 3D model as well as for peak input and FA input. For the 3D model to fit into the memory we had to reduce the batch size to 1, the number of filters by a factor of 8 and the number of downsampling levels from 4 to 3. The 2D model with peak input showed the best results. In terms of Dice score the differences on the *HCP Quality* and *Clinical Quality* dataset are only minor. But differences on the *Phantom* dataset are more severe. Also when looking at results on the 17 non-HCP datasets the 3D model as well as the FA input showed clearly worse results, missing major parts of several bundles.

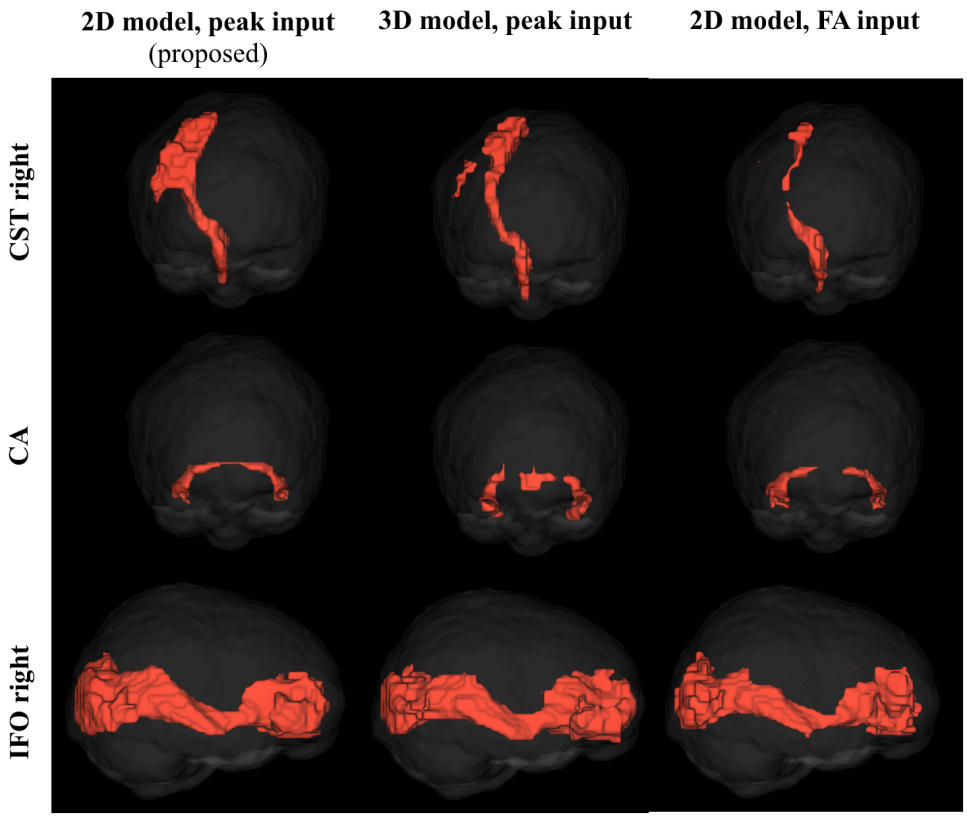

Fig. 8. Segmentation results for 2D and 3D model as well as for peak input and FA input for one subject from the COBRE dataset with enlarged ventricles. The proposed method (2D model with peak input) manages to completely segment all bundles. The 3D model as well as the 2D model using FA as input are missing major parts of several bundles.

Fig. 9. Comparison of the proposed method to results from Wasserthal et al. (2018a) at MICCAI 2018. By using an extra model for segmentation and an extra model for learning of the peak angle we achieved better results than using a model which tries to learn both at the same time (by learning a peak angle and a peak length).

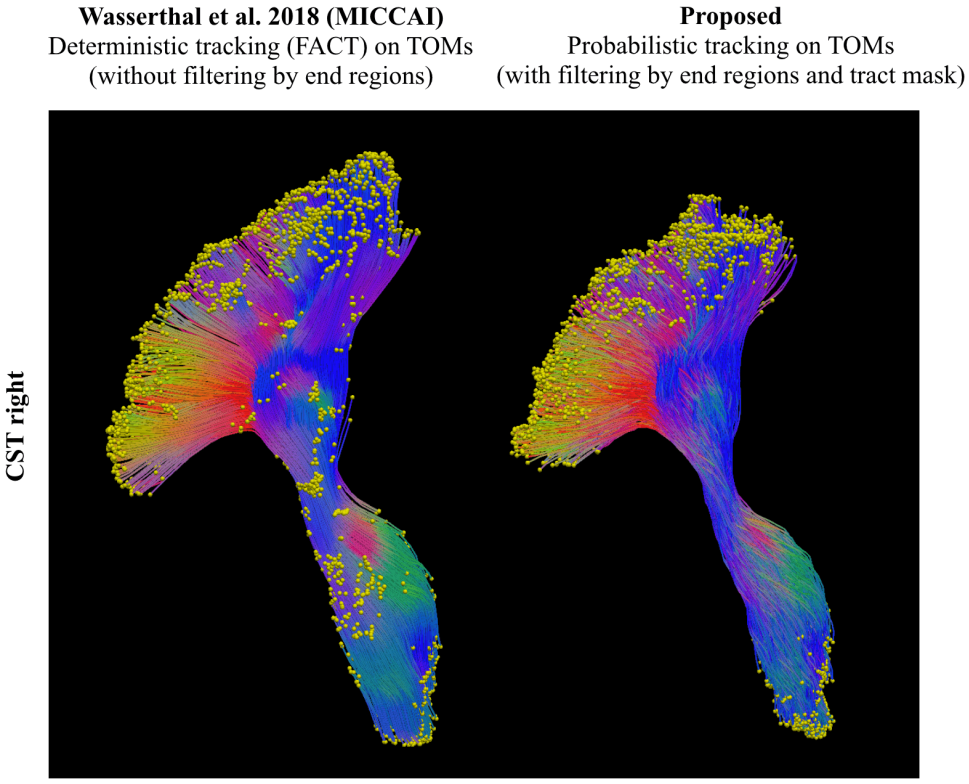

Fig. 10. Comparison of the proposed method to results from Wasserthal et al. (2018a) at MICCAI 2018 for one subject from the BrainGluSchi dataset. Streamline endpoints are marked with yellow dots. By filtering streamlines by start and end region masks streamlines ending prematurely can be removed. This leads to sparser results. By using custom probabilistic tracking complete tractograms can be obtained while still filtering by endpoints.

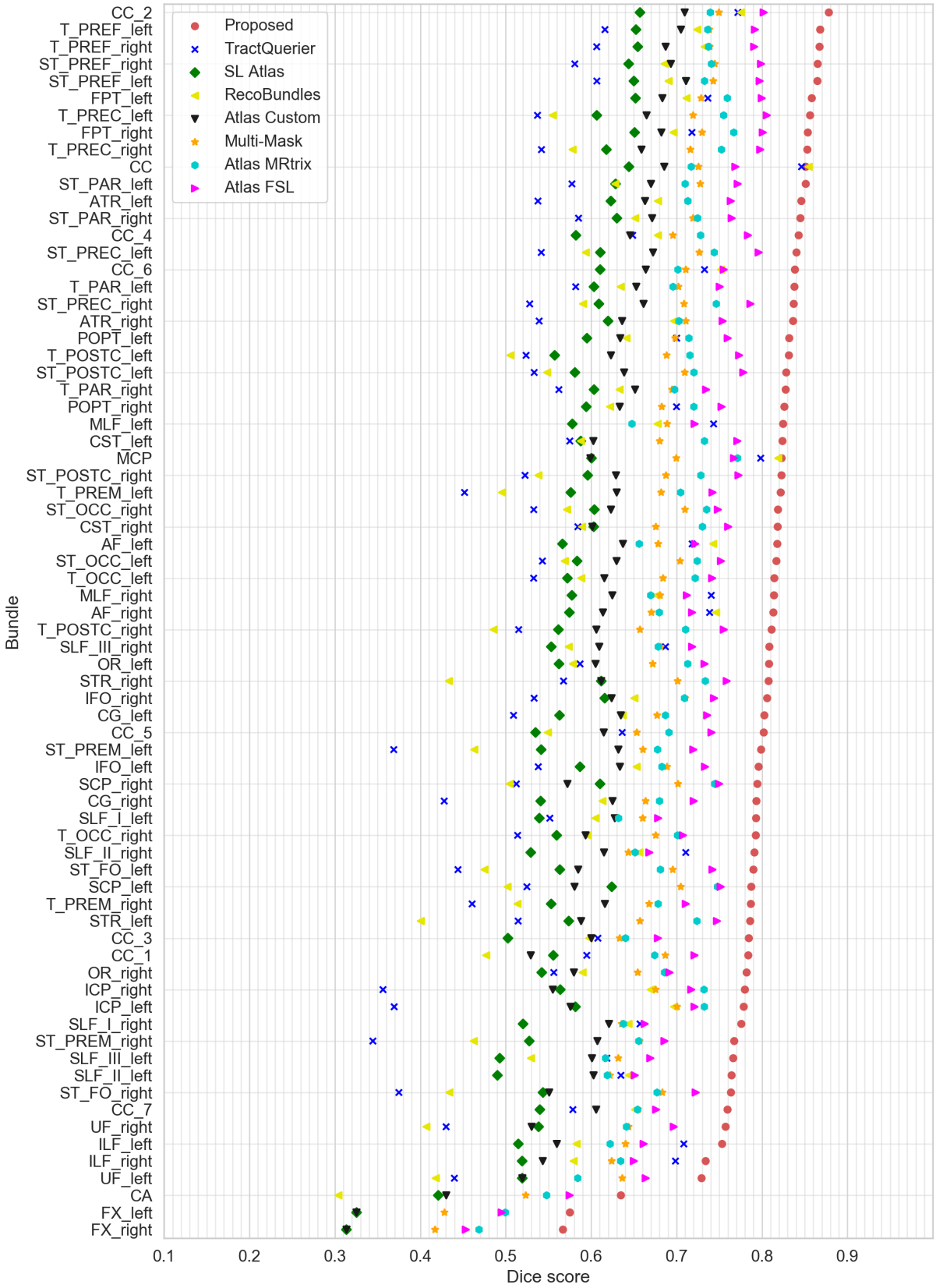

Fig. 11. Dice scores for all 72 tracts on the Clinical Quality dataset for our proposed method and all reference methods sorted by score. The full name of each tract can be seen in the supplementary materials.

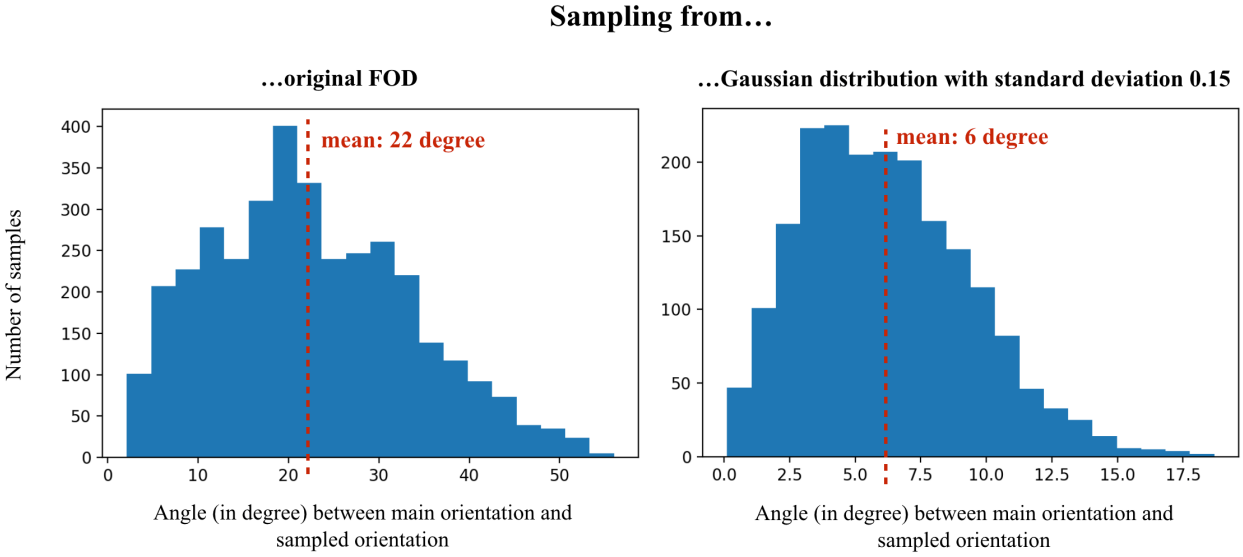

Fig. 12. Histogram of sampled orientations for one voxel in the corpus callosum of one HCP subject. The x-axis shows the angle between the main orientation and the sampled orientation. On the left is the result for the original FOD (generated with constrained spherical deconvolution). On the right is the result for sampling from a fixed Gaussian distribution with standard deviation 0.15. As can be seen when using a standard deviation of 0.15 the fixed distribution has clearly lower dispersion in the angles than the original FOD, thus providing a conservative lower bound.