Anatomy-to-Tract Mapping: Inferring White Matter Pathways Without Diffusion Streamline Propagation

Yee-Fan Tan[1,3,4], Siyuan Liu[2], Khoi Minh Huynh[3,4], Raphaël C.-W. Phan[1], Chee-Ming Ting[1], & Pew-Thian Yap[3,4,✉]

[1]*School of Information Technology, Monash University, Subang Jaya, Malaysia*
[2]*Marine Engineering College, Dalian Maritime University, Dalian, China*
[3]*Department of Radiology, University of North Carolina at Chapel Hill, NC, USA*
[4]*Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, NC, USA*

Diffusion tractography, a cornerstone of white matter mapping, relies on point-to-point streamline propagation—a process often compromised by errors stemming from inadequate signal-to-noise ratio and limited spatial-angular resolution in diffusion MRI (dMRI) data. Here, we present Anatomy-to-Tract Mapping (ATM), a novel deep learning framework that, for the first time, generates complete, bundle-specific streamlines directly from anatomical MRI, entirely bypassing the limitations of streamline propagation. ATM leverages the superior quality and minimal distortion of anatomical MRI, learning from multi-subject datasets to deliver robust, subject-specific streamline bundles while accurately preserving structural connectivity. This paradigm-shifting approach overcomes challenges associated with complex configurations, such as crossing, kissing, bending, and bottlenecks, providing anatomically precise reconstructions. Beyond individual-level mapping, ATM facilitates the creation of population-level streamline bundles aligned to average anatomical templates derived from diverse datasets. By offering global connectivity insights less affected by local uncertainties, ATM complements diffusion tractography, advancing white matter pathway reconstruction with a more reliable, anatomy-driven perspective.

Diffusion MRI (dMRI) tractography remains the primary computational technique for mapping long-range white matter pathways in vivo[1], relying on local deterministic[2,3] or probabilistic[4,5] streamline propagation, or global[6] streamline reconstruction techniques. However, its accuracy is hindered by several key limitations. Partial volume effects, which occur when multiple fiber bundles with different orientations intersect within a single voxel, complicate fiber orientation estimation and lead to errors in complex configurations such as bending, fanning, crossing, and kissing fibers[7,8]. Another challenge is the bottleneck problem, where multiple fiber populations converge in a narrow region, causing orientation ambiguities that result in false-positive streamlines[9,10]. Moreover, dMRI-based tractography is highly sensitive to the quality of the data, which is often limited by signal-to-noise ratio and spatial-angular resolution[11]. The acquisition of high-quality dMRI data remains challenging, particularly in clinical settings, where the quality of available neuroimaging data often renders it unsuitable for reliable tractography.

Recent advancements in machine learning (ML) and deep learning (DL) have shown considerable promise in improving tractography. Several DL approaches have been developed to predict fiber tract orientations from diffusion signals. These include methods based on orientation forecasting, which predicts the next location of a fiber tract based on its previous positions[12–14],

✉ Corresponding author: Pew-Thian Yap (ptyap@med.unc.edu)

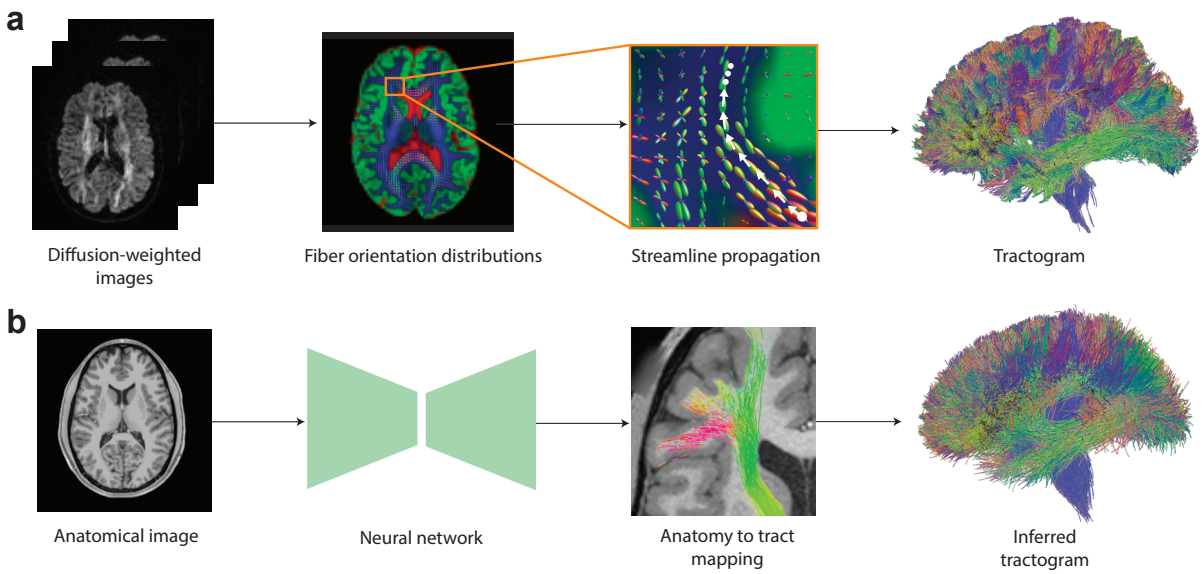

Fig. 1 | Streamline propagation versus Anatomy-to-Tract Mapping (ATM). a, Conventional tractography propagates streamlines point-to-point from seed points using local fiber orientation distributions (FODs), either deterministically or probabilistically. **b**, ATM infers entire bundles from an anatomical T1-weighted (T1w) image.

as well as reinforcement learning models that trace fibers by optimizing a reward function for accurate tract reconstruction[15]. Additionally, DL models have been used to predict fiber orientation distribution functions (fODFs), representing the likelihood of fiber directions at each voxel[16–18], and Fischer-von-Mises (FvM) distributions to improve tractography results[19].

A notable recent innovation is a DL-based approach utilizing a convolutional-recurrent neural network architecture to approximate tractography directly from T1-weighted (T1w) MRI, bypassing the need for dMRI data[11]. This method attempts to infer brain connectivity solely from anatomical features, offering a promising alternative to traditional dMRI-based techniques. While this model has shown encouraging results, it still relies on point-to-point tracking, which remains vulnerable to the same tractography challenges, such as fiber crossing and bending, leading to artifacts and inaccuracies in reconstructed pathways.

In this work, we introduce a novel deep learning framework for Anatomy-to-Tract Mapping (ATM), which completely eliminates the reliance on diffusion MRI. Unlike previous ML and DL approaches[11–17,19], ATM represents a pioneering shift by avoiding point-to-point tracking and instead learning to generate complete, subject-specific white matter bundles directly from anatomical MRI. By leveraging the high-quality, low-distortion properties of T1w MRI, ATM effectively circumvents the limitations of complex fiber configurations such as crossing fibers and bottlenecks (Figure 1). Specifically, ATM captures anatomical context from T1w MRI to guide the reconstruction of full white matter bundles, employing a probabilistic approach where random vectors are drawn from a probability density function generated via kernel density estimation (KDE), inspired by recent work using variational autoencoders (VAE) for streamline generation[20].

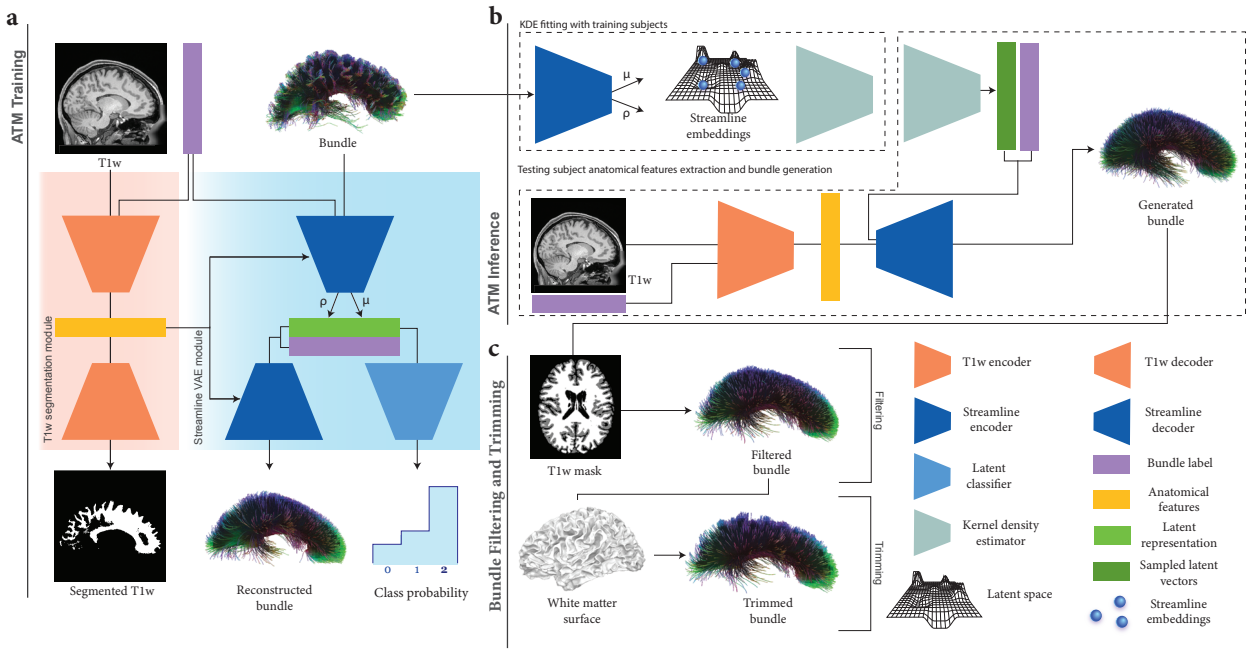

Fig. 2 | ATM overview. a, ATM model training. ATM takes as inputs anatomical T1w images and streamlines from bundle c, project them into different latent spaces through T1w encoder E_A and bundle encoder E_S, respectively. T1w segmentation decoder D_A guides anatomical feature extraction. Bundle decoder D_S reconstructs the output streamlines from the latent space z_s, conditioned on the anatomical features a extracted using E_A. A classifier C_Z is incorporated to ensure the features extracted are specific to bundle c. **b**, ATM inferencing. The probably density function of the streamline embeddings is estimated using KDE. ATM utilizes points sampled based on the probability density function, conditioned on anatomical features a, to generate streamlines of a particular bundle c. **c**, Bundle filtering and trimming. The generated bundle is refined by filtering exccess streamlines with guidance from a T1w-derived tissue mask. The filtered bundle is subsequently trimmed based on the white matter surface.

These generated bundles are then refined through filtering and trimming techniques (Figure 2c).

Our results demonstrate that ATM can successfully generate subject-specific, anatomically accurate white matter bundles directly from T1w MRI, producing reconstructions that closely resemble ground truth generated through traditional dMRI tractography. This breakthrough shows that full white matter pathways can be reconstructed from anatomical MRI alone, without relying on diffusion-based streamline propagation, marking a significant advancement in the field of neuroimaging.

RESULTS

Framework overview. For reproducibility, we used the TractSeg dataset[21], which comprises 72 manually validated bundles from 105 subjects in the WU-Minn Human Connectome Project (HCP) diffusion MRI database[22–24]. We focused primarily on the ATM of three white matter bundle categories: association, commissural, and projection[25]. Each of these categories com-

prises smaller bundles, as detailed in Table S1. To learn the mapping from T1w images to these bundles, tract streamlines were randomly sampled from the pool of 72 TractSeg bundles, with 8,000 streamlines per bundle category. Subsequently, data interpolation, facilitated by the Dipy library[26], was employed to ensure a consistent point count per streamline of 128. The corresponding T1w images from the HCP database were used to generate binary segmentation maps for each target bundle via DSI Studio[3]. The T1w images were downsampled from $260 \times 311 \times 260$ to $80 \times 96 \times 80$ to speed up training and were normalized to an intensity range of $[0, 1]$. The dataset was partitioned into 92 subjects for training, 5 subjects for validation, and 8 subjects for testing. We further assessed bundle variability generated by ATM using the WU-Minn HCP retest data, which included 45 subjects.

ATM (Figure 2) takes N random vectors, each of size 64 drawn from a probability density function estimated using KDE in conjunction with anatomical features extracted from T1w images, to generate N streamlines specific to bundle one-hot encoded with c. To evaluate the performance of the trained ATM model, a set of N = 8,000, 12,000, 16,000, 20,000 streamlines were sampled each for the association, commissural, and projection bundles. To eliminate any implausible or spurious streamlines generated by ATM from the latent space, each individual bundle was filtered based on the corresponding T1w white matter mask. These filtered bundles were trimmed by removing points beyond the white matter surface, and then consolidated into a whole-brain bundle. We further validated ATM performance on 48 smaller bilateral and commissural bundles, listed in Table S2. A set of N = 1,000 streamlines were sampled for these smaller bundles, with all other procedures kept unchanged.

ATM generates complete bundles with characteristic shapes. We demonstrate the effectiveness of ATM in generating major white matter bundles—including association, commissural, and projection bundles—directly from anatomical T1w MRI, without the need for voxel-to-voxel tracking as required in traditional tractography. ATM can generate complete bundles, even those that are thin and difficult to track, such as the fornix in the association bundle. Visualizing the bundles in the context of white matter surfaces and with tract density maps on T1w images shows strong agreement between the generated and ground truth bundles (Figure 3). The bundles align well with the anatomical changes seen in the T1w scans across subjects, as confirmed by the sagittal, coronal, and axial views. The close-up views highlight good bundle-anatomy alignment, confirming that ATM generates bundles that reflect anatomical details. Results for association, commissural, and projection bundles are shown in Figures S2-S4. Results for smaller bundles, including the left arcuate fascicle (AF), anterior thalamic radiation (ATR), cingulum (CG), and thalamo-postcentral (T_POSTC), are provided in Figure S5.

We visualize the latent space of each bundle-specific model using t-distributed stochastic neighbor embedding (t-SNE) implemented with openTSNE[27]. This technique reduces the high-dimensional latent vectors to a two-dimensional space for easier interpretation and visualization. The t-SNE projections of the latent vectors z_s are shown in Figure S6, revealing the structure underlying the data. The clusters of the latent vectors corresponding to the association, commissural, and projection bundles appear more dispersed with a greater mixture between samples. This dispersion likely reflects the higher variability and complexity inherent in these bundles, as they encompass a broader range of anatomical pathways. In contrast, the latent vectors of the smaller bilateral and commissural bundles form more compact clusters in the t-SNE projections,

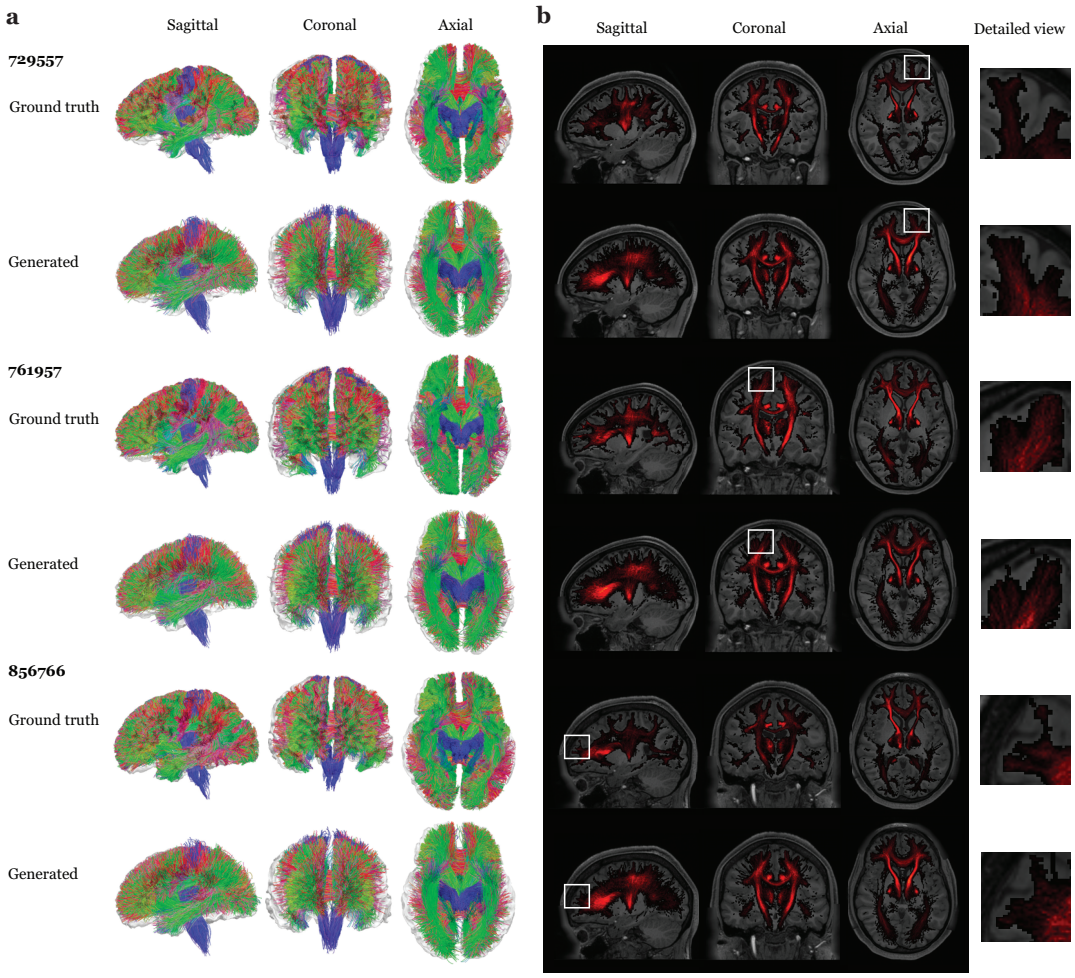

Fig. 3 | Bundle similarity with ground truth. **a**, Whole-brain bundles in the context of white matter surfaces, **b**, Tract density maps shown in subject-specific T1w spaces. For each subject, the top row represents the ground truth and the bottom row represents the bundle generated from the T1w image.

indicating that ATM effectively captures and preserves the local features of the streamlines, resulting in robust latent representations. This clustering highlights ATM's capability to differentiate between various white matter structures based on their unique characteristics, demonstrating its potential for accurate bundle generation.

High geometric agreement with ground truth. Using the bundle adjacency (BA) metric from Bundle Analytics (BUAN)[28], we assessed the quality of the generated bundles in terms of shape similarity with respect to the ground truth. Mean BA scores across subjects range from 82% to 97% (Table S3), indicating that ATM generated bundles highly resemble the ground truth.

We compared the generated and ground truth bundles in terms of spatial coverage using Tractometer[29,30], focusing on volumetric bundle overlap, overreach, and F1 scores (Table S3). ATM consistently produces bundles with competitive overlap, overreach, and F1 scores, ranging

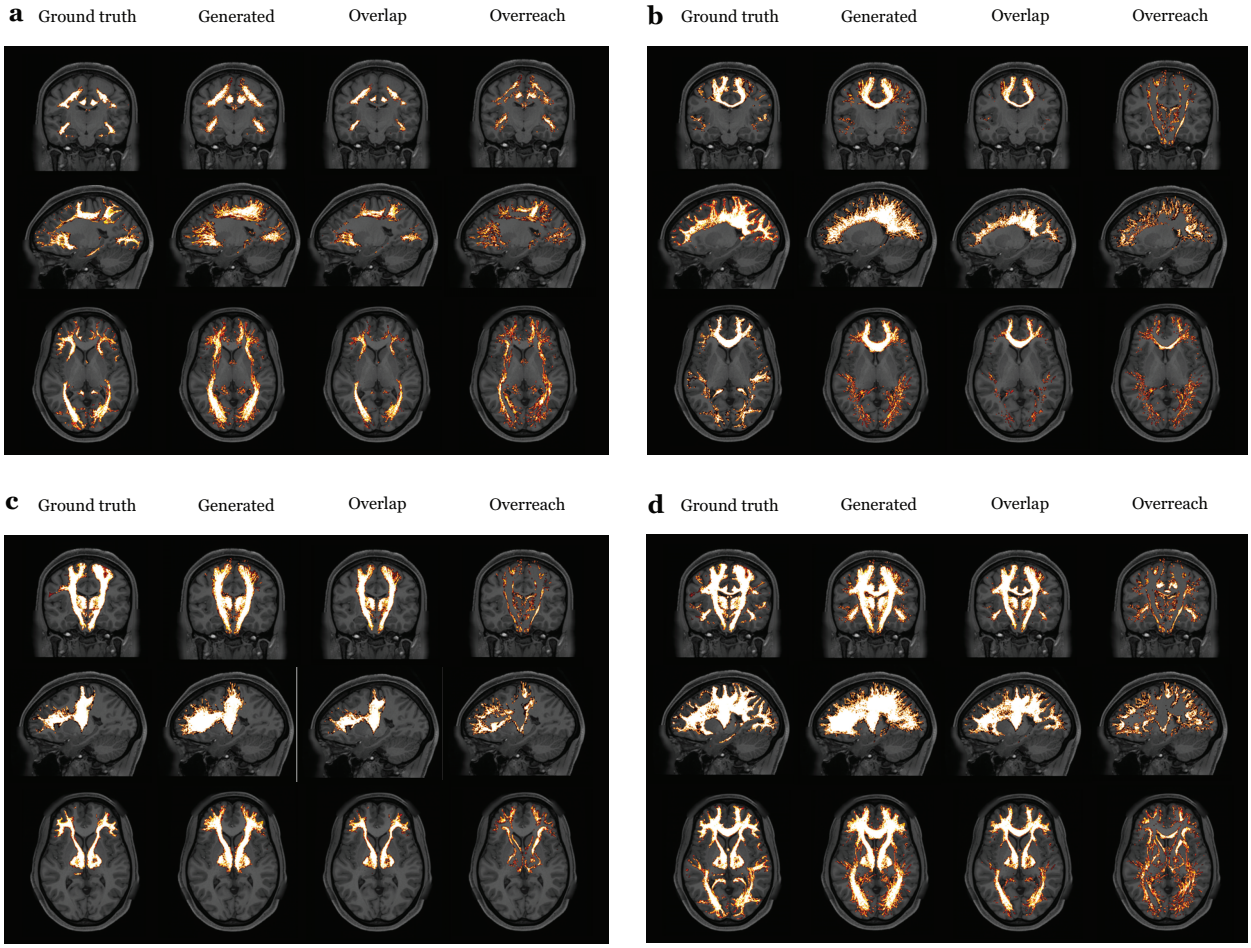

Fig. 4 | Bundle coverage. Spatial overlap and overreach between the ground truth and generated **a**, association, **b**, commissural, **c**, projection, and **d**, whole-brain combined bundles of a single representative subject shown in the T1w space.

between 54 − 70%, 43 − 68%, and 49 − 66%, respectively. Figure 4 shows higher overreach scores for the association bundle than the commissural and projection bundles, likely due to greater shape variability in thinner tracts like the fornix[21,31]. ATM yields competitive performance compared with several state-of-the-art methods for streamline generation, including Generative Sampling in Bundle Tractography using Autoencoders (GESTA)[31], Track-to-Learn[15], DeepTract[16] and CtTrack[17]. Based on the ISMRM 2015 Tractography Challenge dataset[32], GESTA yields overlap and overreach scores between 65 − 84% and 64 − 104%, Track-to-Learn achieves overlap scores between 55 − 62%, and DeepTract[16] reports overlap, overreach, and F1 scores of 69%, 22.7%, and 70.1%. Additionally, CtTrack[17], evaluated on two HCP subjects, reports overlap, overreach, and F1 scores of 50%, 16%, and 60%, respectively.

ATM generated bundles consistently showed high performance in streamline validity metrics assessed with Tractometer[29,30], measuring the validity of individual streamlines within each

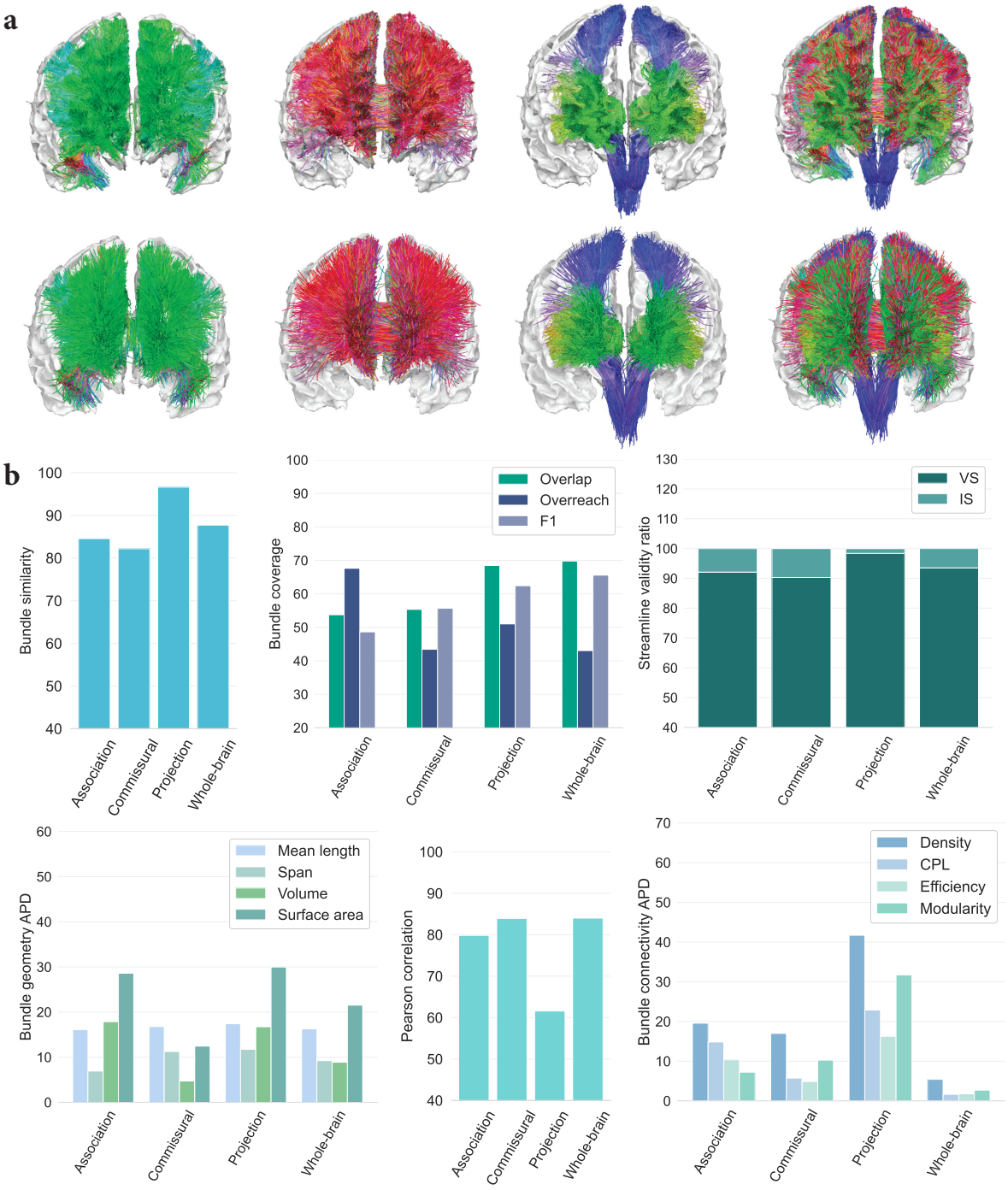

Fig. 5 | Bundle geometry and connectivity. **a**, Comparison between (top) ground truth and (bottom) generated association, commissural, projection, and whole-brain bundles. **b**, Numerical results in terms of bundle-specific similarity, coverage, validity, geometry, and connectivity.

bundle. Results indicate that the generated bundles achieve valid streamline (VS) ratios between approximately 90% and 98%, with invavlid streamline (IS) ratios generally below 10% (Table S3). Based on a test set of two HCP subjects, CtTrack[17] achieved a VS ratio of 57%. Based on the ISMRM 2015 Tractography Challenge dataset[32], Track-to-Learn[15] and DeepTract[16] respectively reported VS ratios of 68% and 70%.

Bundle geometry was further assessed by measuring mean length, span, volume, and surface area (Table S4). The absolute percent differences (APDs) of these geometric measurements with respect to the ground truth ranged from 16-17%, 7-12%, 5-18%, and 12-29%, showing strong morphological agreement between the generated and ground truth bundles. For reference, a recent study using a convolutional-recurrent neural network (CoRNN)[11] on 10 HCP subjects reported mean APD of approximately 5%, 5%, 20%, and 15% for the bundle mean length, span, volume, and surface area, respectively. Visual and numerical comparisons in Figure 5 reinforce our conclusion that ATM generates bundles that are faithful to subject-specific anatomy.

Additional validation with smaller bilateral and commissural bundles (Table S2) shows high performance based on BA and Tractometer metrics, as highlighted in Figure 6, Table S5, and Table S6. These findings suggest that ATM effectively captures bundle-specific shapes across all bundle categories through subject-specific T1w images.

High-fidelity connectomics. We assessed the connectomes derived from the bundles based on the Desikan–Killiany-Tourville (DKT) atlas[33,34], with edges defined by streamline counts between pairs of the 62 DKT cortical regions. Numetical results based on based on Pearson correlation and the graph-theoretic measures, such as network density, characteristic path length (CPL), global efficiency, and modularity, are reported in Figure 5 and Table S4. We found high Pearson correlations for association, commissural, and projection bundles as well as their combination, with population averages of about 80%, 84%, 62%, and 84%, respectively. Discrepancy between the generated and ground truth bundles is low, with density, CPL, efficiency, and modularity APDs ranging respectively from $5-41\%$, $1-23\%$, and $1-16\%$, and $2-32\%$ (Table S4). This suggests a high level of network-level agreement with the ground truth, except for the projection bundle, which exhibits a slightly higher difference, ranging from $23-41\%$ across metrics. Evaluation based on the smaller white matter bundles also demonstrate strong connectomic agreement (Figure 6, Table S6). Evaluated with a different set of 97 cortical regions, CoRNN[11] reported approximately 80% Pearson correlation, and 5% APD for both CPL and modularity.

Streamline sample size. As the number of streamlines increases, bundle coverage overlap, overreach, and F1 scores also show a corresponding rise, as demonstrated in Table S3. This indicates that with more streamlines, there is improved spatial matching between the reconstructed and ground truth bundles. Figure S7 illustrates representative qualitative results that highlight these trends visually. Despite these changes in coverage-related metrics, the overall bundle geometric and connectomic fidelity remain relatively stable, showing no significant change, as indicated in Table S4.

Test-retest reliability. ATM shows high test-retest consistency, with inter-session BA scores of 93%–99%, overlap of 56%–70%, and overreach of 29%–43% (Table S7). The high BA scores suggest good bundle reproducibility, though coverage measures show some variability. We also assessed

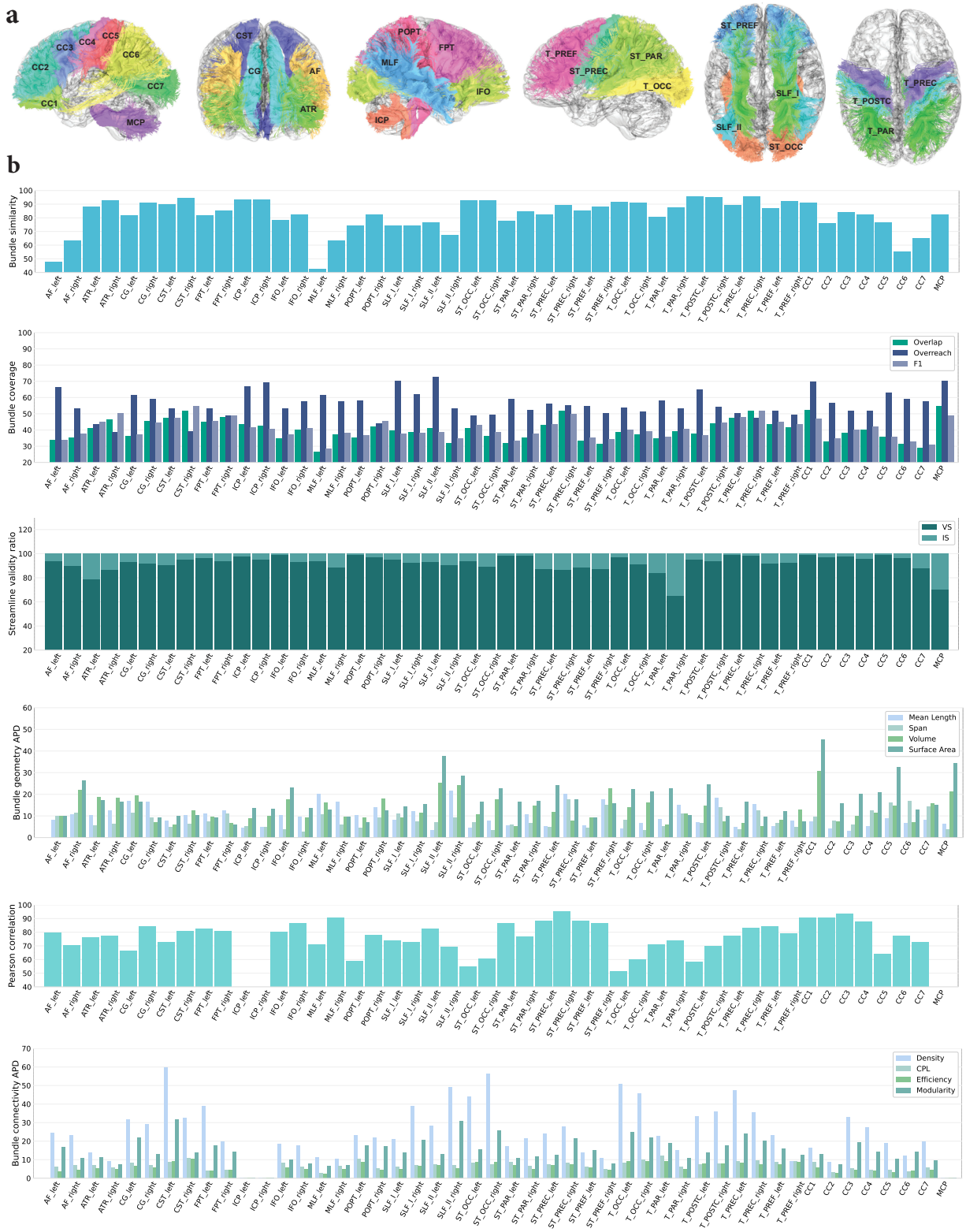

Fig. 6 | Assessment of smaller bilateral and commissural bundles. a, Bilateral and commissural bundles inclucded for evaluation. **b**, Numerical results in terms of bundle-specific similarity, coverage, and connectivity.

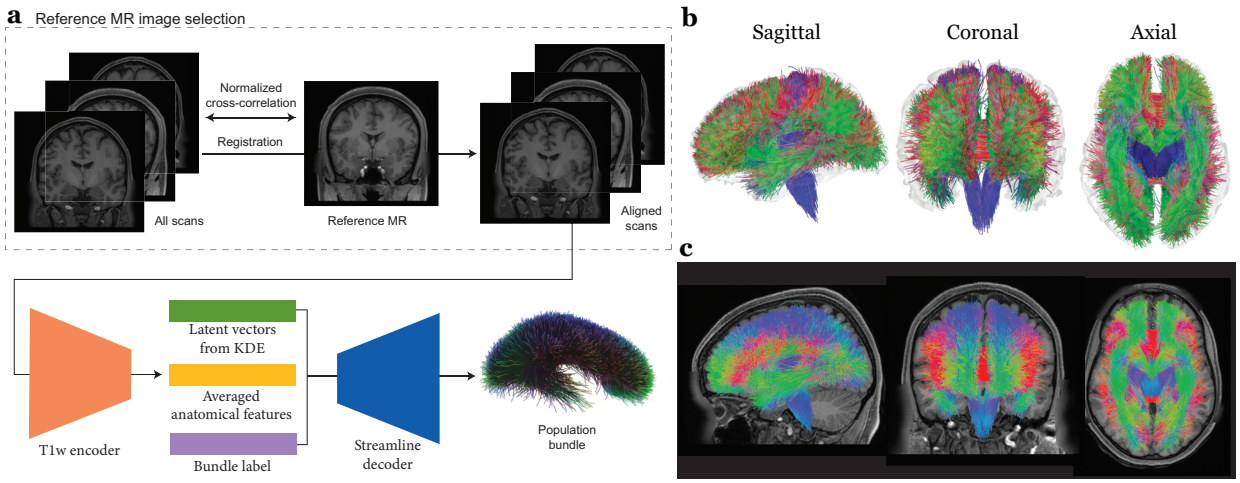

Fig. 7 | Population-level bundle generation. **a**, Steps involved in of population bundle inference. Generated bundle shown in the context of **b**, the white matter surface and **c**, the T1w space.

inter-session geometric and connectomic APDs for generated bundles (Table S8). Geometric APDs show high agreement, with only 1–2% differences. Connectomes weighted by streamline count exhibit high Pearson correlations and low APD values for key metrics, i.e., density, CPL, efficiency, and modularity, with variability from 3% to 46%.

Population-level inference. We tested whether ATM could infer population-level bundles from T1w images. First, we calculated the normalized cross-correlation between the T1w images and selected the image with the lowest total distance as the reference. Using Advanced Normalization Tools (ANTs)[35], all images were aligned to this reference image. ATM anatomical features were then extracted from the aligned images and averaged to generate bundles. Figure 7 illustrates this process, showing the generated whole-brain bundle within the reference white matter surface and T1w space. The resulting bundles align well in the reference space, reflecting the averaged anatomical features from multiple images.

DISCUSSION

We introduced ATM, a deep learning model that extracts anatomical features from T1-weighted images to infer individual-specific white matter bundles. ATM overcomes challenges commonly faced by traditional tractography methods, such as navigating complex configurations like crossing, kissing, and bending fibers, as well as constricting bottlenecks. By generating streamlines directly from anatomical images, ATM reduces the need for high-quality dMRI data and enables anatomy-based tractograms to guide diffusion tractography. Our results show a strong resemblance between ATM predictions and actual bundles, with competitive performance in similarity, coverage, validity, geometry, and connectivity, and inter-session consistency of 93–99%. ATM also simplifies population-level bundle inference using averaged anatomical features, making it easy to create bundle atlases from population data.

Recent advances in deep learning (DL)-based tractography have primarily focused on predicting local fiber orientations (e.g., Learn-to-Track[13]) or modeling fiber orientation distributions (e.g., DeepTract[16] and CTtrack[17]). However, these methods often struggle with tracking errors, especially in regions with complex fiber configurations. A new approach, Generative Sampling in Bundle Tractography using Autoencoders (GESTA)[31], uses an autoencoder-based framework to generate complete streamlines from the latent space of a trained model. While GESTA improves spatial coverage of white matter bundles without relying on local streamline propagation, its sampling process depends on the quality of seed streamlines and does not account for individual anatomical differences. In contrast, ATM incorporates T1-weighted (T1w) images to guide subject-specific streamline generation, allowing a direct mapping between anatomy and tracts. Another recent method, the Convolutional-Recurrent Neural Network (CoRNN)[11], adopts a teacher-student framework. In this approach, the student model samples streamlines from T1w MRI by mimicking the teacher model, which propagates streamlines from diffusion MRI (dMRI). However, CoRNN's point-to-point tracking remains prone to the same errors as traditional dMRI-based tractography, particularly in complex fiber regions, potentially leading to false-positive streamlines. In contrast, ATM demonstrates competitive performance in both coverage and accuracy without relying on diffusion-based streamline propagation.

ATM has certain limitations that necessitate further improvements. While it is capable of generating streamlines that accurately represent anatomical pathways, it is not immune to errors. These errors can undermine the model's reliability in clinical and research applications. These errors may arise from the imperfection of ATM in fully encoding anatomical information relevant for precise streamline generation. For instance, the model might overlook subtle variations in the white matter anatomy or fail to account for complex anatomical relationships, leading to inaccuracies in the generated streamlines. Although the current model has been successful in creating subject-specific bundles, more research is needed to identify and better understand the specific anatomical features that ATM captures. This deeper understanding could inform future enhancements to the model, allowing it to produce more accurate and representative streamlines.

Another significant challenge is the presence of spurious and noisy streamlines in the generated bundles. This issue often arises from the inherent randomness associated with the latent space sampling process. Such spurious bundles can compromise the overall quality of the tractography results, leading to misleading interpretations in downstream analyses. To minimize the occurrence of these spurious bundles, future iterations of ATM should consider optimizing model architectures and refining loss functions. By careful adjustments, the model can be enhanced to generate cleaner and more reliable streamlines. Moreover, the current filtering algorithm used in ATM could benefit from enhancements through the incorporation of advanced deep learning models, such as Filtering in Tractography using Autoencoders (FINTA)[36]. FINTA employs an autoencoder framework to efficiently filter streamlines by comparing the distance between reference streamlines and those that need to be filtered. Since ATM is also based on an autoencoder architecture, integrating FINTA's latent space filtering technique could be a seamless process. This integration would enable ATM to leverage sophisticated filtering capabilities, improving the overall quality of the generated bundles.

ATM is currently trained on bundles and T1-weighted (T1w) MRI data scanned from healthy individuals. While this training provides a solid foundation, it may limit the model's

effectiveness when applied to subjects who possess different neuroanatomical features or present with various pathologies. For instance, individuals with neurological disorders or developmental abnormalities often exhibit atypical white matter configurations that are not well-represented in the training data. As a result, ATM may struggle to generate accurate white matter bundles in these populations, potentially compromising its clinical utility and reliability in understanding complex brain structures. Fortunately, ATM has a conditioning mechanism that allows it to be retrained using data reflecting anatomical variations associated with abnormal conditions. By incorporating data from individuals with different pathologies, ATM can learn how streamline trajectories adapt to these unique anatomical characteristics. This adaptability is crucial for improving the model's performance across diverse populations, making it more versatile and clinically relevant.

Additionally, expanding ATM to incorporate other imaging modalities, such as T2-weighted (T2w) MRI, could significantly enhance its analytical capabilities. Integrating data from both T1w and T2w images would provide a more comprehensive view of the brain's anatomy, allowing ATM to leverage different types of information. For instance, T2w images can provide insights into fluid-filled structures and changes in tissue properties that T1w images alone may not capture. This broader approach could not only improve the accuracy of generated white matter bundles but also help in refining the model to reduce tracking errors, thereby enhancing its reliability.

Overall, ATM has proven effective in generating accurate white matter bundles from anatomical images, even in the absence of diffusion MRI (dMRI) data. This capability is particularly important because dMRI, while a standard method for mapping white matter pathways, is not always available in clinical settings or for certain patient populations. ATM's strength lies in its ability to capture precise bundle shapes and subject-specific anatomical details. By leveraging the rich anatomical information present in T1-weighted MRI images, ATM generates white matter bundles that closely resemble the ground truth geometrically. This level of accuracy ensures that the generated bundles can effectively represent the underlying white matter architecture. The inferred bundles produced by ATM can be valuable in in-vivo white matter pathway analysis in understanding of neural connectivity, supporting clinical decision-making, and contributing to the advancement of neuroimaging research.

METHODS

ATM architecture. The ATM network (Figure 2) consists of two main components:

1. A segmentation module (Figure 2a) with a volumetric image encoder E_{A} to extract anatomical features a from a T1w image x and a decoder D_{A} to predict the binary segmentation map $\tilde{x}$ of the bundle of interest, using a conditioning mechanism based on feature-wise linear modulation[37] (FiLM) layers to target bundle c.

2. A streamline variational autoencoder (VAE) module (Figure 2b) that employs an encoder E_{S} to embed streamline information in a low-dimensional space conditioned on the anatomical information a, and the bundle of interest c, using FiLM layers. The latent sample $\mathbf{z}_s$, together with the bundle of interest c, and anatomical features a conditioned using FiLM layers, are inputs to the decoder D_{S} to reconstruct the target streamline $\tilde{s}$. A classifier C_{Z} is used to ensure the extracted streamline embedding $\mathbf{z}_s$ is of bundle c.

See Figure S1 for the architectural details of ATM.

As depicted in Figure 2a, for a given input T1w image $x \in \mathcal{X}$ and streamlines $s \in \mathcal{S}$, the image encoder E_{A} extracts anatomical features a conditioned on the bundle type c. With the high-level anatomical features a, the anatomical MRI image decoder D_{A} constructs its corresponding binary segmentation $\tilde{x} = D_{\mathrm{A}}(a)$. Conditioned on the extracted anatomical features a and the type of bundle c, the streamline encoder E_{S} generates mean vector $\boldsymbol{\mu}_s$ and standard deviation vector $\boldsymbol{\sigma}_s$. The latent vector $\mathbf{z}_s$ is obtained through reparameterization, following $\mathbf{z}_s = \boldsymbol{\mu}_s + \boldsymbol{\epsilon} \times \boldsymbol{\sigma}_s$, where $\boldsymbol{\epsilon}$ is sampled from $\mathcal{N}(0,1)$. The latent classifier C_{Z} is used to ensure the embeddings are of the bundle of interest c. Finally, the streamline decoder D_{S} learns the mapping between latent vector $\mathbf{z}_s$ and the reconstructed streamlines $\tilde{s} = D_{\mathrm{S}}(\mathbf{z}_s, a, c)$.

The segmentation module of ATM undergoes training with the primary objective of minimizing the segmentation loss, $\mathcal{L}_{\mathrm{SEG}}$, which measures the dissimilarity between the ground truth $\hat{x}$ and the reconstructed binary segmentation $\tilde{x}$ of input image x, specifically for a designated bundle c. This optimization is achieved through the utilization of the Dice loss function:

$$\mathcal{L}_{\mathrm{SEG}} = 1 - 2\left(\frac{\sum_m^M (\hat{x}_m \cdot \tilde{x}_m) + \mathrm{e}}{\sum_m^M \hat{x}_m^2 + \sum_m^M \tilde{x}_m^2 + \mathrm{e}}\right), \tag{1}$$

where e denotes the smoothing effect and m indicates the m-th voxel of the image. In the context of streamline reconstruction and generation, ATM incorporates the VAE loss function[38,39]:

$$\mathcal{L}_{\mathrm{VAE}} = \mathbb{E}[(s - \tilde{s})^2] + \beta D_{\mathrm{KL}}(q(\mathbf{z}_s|s,c)||p(\mathbf{z})), \tag{2}$$

where β is a regularization coefficient that constraints the latent bottleneck, $p(\mathbf{z})$ is the Gaussian prior distribution $\mathcal{N}(0,1)$, and $D_{\mathrm{KL}}(\cdot||\cdot)$ is the Kullback-Leibler divergence. The first term in the equation is the mean square error loss, evaluating how effectively ATM can reconstruct the input streamlines. The second term regularizes the latent space $q(\mathbf{z}_s|s,c)$ to follow the prior distribution $p(\mathbf{z})$. ATM uses cross-entropy loss between the actual bundle label c and the predicted label c' to

ensure that the streamline embeddings $\mathbf{z}_s$ match the correct bundle type c:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{n} c_i \log(c_i'), \tag{3}$$

where N is the number of streamlines and $c_i' = C_{\text{Z}}(\mathbf{z}_s^{(i)})$ denotes the predicted output for the i-th streamline embedding. To ensure the quality of the generated streamlines, ATM explicitly integrates a regularization term $\mathcal{L}_{\text{ADJ}}$ that keeps adjacent points on a streamline closely spaced, preventing them from becoming dispersed or disorganized. This is achieved by calculating the average and maximum of the average Euclidean distances between adjacent points across all streamlines:

$$\mathcal{L}_{\text{ADJ}} = \frac{1}{N}\sum_{i=1}^{N} B(i) + \max_{i\in\{1,\ldots,N\}} B(i), \tag{4}$$

where $\mathbf{p}$ represents a point on the streamline, and $B(i) = \frac{1}{K}\sum_{j=1}^{K} ||\mathbf{p}_j^{(i)} - \mathbf{p}_{j-1}^{(i)}||$ is the average distance between consecutive points across K points on the i-th streamline. The first and second terms then calculate the mean and maximum of these average distances across all N streamlines.

Finally, ATM is optimized with the following total loss function:

$$\mathcal{L}_{\text{ATM}} = \lambda_{\text{SEG}}\mathcal{L}_{\text{SEG}} + \lambda_{\text{VAE}}\mathcal{L}_{\text{VAE}} + \lambda_{\text{CE}}\mathcal{L}_{\text{CE}} + \lambda_{\text{ADJ}}\mathcal{L}_{\text{ADJ}}. \tag{5}$$

Streamline inference. Prior to streamline generation, the probability density function of streamline embeddings $\mathbf{z}_s$ extracted via the trained streamline encoder E_{S} is estimated using KDE. For bundle-specific generation, encoder E_{A} extracts the anatomical features a of the input T1w image x given the bundle type c. Conditioned on a and c, the streamline decoder D_{S} maps a sampled latent vector $\mathbf{z}_r$ from the probability density function to reconstruct streamline $\tilde{s}$.

Streamline filtering and trimming. A streamline within a bundle is represented as a sequence of points. To remove unwanted and noisy streamlines, we use tissue segmentation maps from T1-weighted scans, which include white matter and gray matter masks generated with MRtrix3[40]. Any streamline extending beyond the white matter and gray matter is discarded. Additionally, streamlines shorter than the threshold length t_1 are excluded from the bundle.

We further refine the streamlines by trimming them based on the white matter surface. Starting from one end point, p_s, of each streamline s, we find the nearest vertex on the white matter surface mesh. We then compute two vectors: u, which points from p_s to the vertex, and n, the vertex's normal vector. If the angle between u and n is greater than 180 degrees, indicating that p_s is outside the surface, we discard the point. This process is repeated iteratively until p_s is inside the surface. The same method is applied to the other end of the streamline s.

ATM implementation details. ATM was implemented using Tensorflow[41] on a machine with a CPU (Intel-I9-12900K) and a GPU (NVIDIA GeForce RTX3090 24GB). ATM training was guided by T1w images and binary segmentations indicating tract occupancy on the voxel grid. The segmentation module is trained using the Dice loss function[42]. The training of the streamline VAE aligns with beta-VAE[38], incorporating an adjustable hyperparameter, β, to control the emphasis

on learning statistically independent latent factors. Mean squared error (MSE) is employed to minimize the streamline reconstruction loss. Cross-entropy loss is used to ensure the streamline embeddings are of the targeted bundle of interest. Regularization of streamline adjacent points is incorporated to prevent point dispersion. ATM is trained using the Adam[43] optimizer with a learning rate of 1×10^{-4} for a total of 2,000 epochs. Throughout the study, we set the weights of the loss functions as follows: $\lambda_{\text{SEG}} = 1$, $\lambda_{\text{VAE}} = 1$, $\lambda_{\text{CE}} = 1$, $\lambda_{\text{ADJ}} = 1$, and $\beta = 1$.

Bundle metrics. ATM generated bundles were evaluated based on similarity, coverage, validity, geometry, and connectivity:

1. *Similarity*: The quality of the generated bundles were evaluated using the bundle adjacency (BA) metric from Bundle Analytics (BUAN)[28], measuring the shape similarity between generated and ground truth bundles.

2. *Coverage*: Bundle spatial coverage was evaluated using Tractometer[29,30], focusing on volumetric overlap, overreach, and F1 scores. The overlap score represents the proportion of voxels within the ground truth bundle that are traversed by at least one valid streamline from the corresponding ATM-generated bundle. The overreach score indicates the proportion of voxels outside the ground truth bundle that are traversed by these streamlines. The F1 score is calculated as the harmonic mean of voxel precision, which represents the proportion of true positive results, and sensitivity (also known as overlap), which quantifies the proportion of actual positive voxels correctly identified.

3. *Validity*: Streamline validity was assessed using valid streamline (VS) and invalid streamline (IS) ratios as defined in Tractometer[29,30]. The VS ratio indicates the proportion of streamlines that are part of valid bundles. The IS ratio indicates the proportion of streamlines that are not part of valid bundles.

4. *Geometry*: Bundle geometry was asssessed with DSI Studio[3,44] using the following measurements:
   - Mean length: Average streamline length within a bundle.
   - Span: Average distance between streamline endpoints.
   - Volume: Total volume occupied by a bundle.
   - Surface area: Total surface area of a bundle.

5. *Connectivity*: Connectomes derived from bundles using the Desikan–Killiany-Tourville (DKT) atlas[33,34] were compared with the ground truth via Pearson correlation. Key graph-theoretic properties were also evaluated using the Brain Connectivity Toolbox[45]:
   - Density: Proportion of actual connections relative to all possible connections.
   - Characteristic path length (CPL): Average shortest path between brain regions.
   - Global efficiency: Average inverse shortest path length across the network.
   - Modularity (Mod): Degree to which the network divides into distinct modules.

Supplementary Information The manuscript contains supplementary material.

Acknowledgements This work was supported in part by the United States National Institutes of Health (NIH) under grants R01 MH125479, R01 EB008374, R01 MH133836, and R01 EB035160.

Author Contributions Y.-F.T.: Methodology, investigation, visualization, data curation, writing – original draft, writing — review and editing. S.L.: Methodology. K.M.H.: Methodology. R.C.-W.P.: Reources. C.-M.T.: Funding acquisition, supervision, writing — review and editing. P.-T.Y.: Conceptualization, supervision, funding acquisition, resources, investigation, validation, writing — review and editing.

Data Availability The TractSeg data used in this study is available at `https://zenodo.org/records/1477956`.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to P.-T.Y.

1. Maier-Hein, K. H. *et al.* The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications* **8**, 1349 (2017).

2. Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. & Aldroubi, A. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* **44**, 625–632 (2000).

3. Yeh, F.-C., Verstynen, T. D., Wang, Y., Fernández-Miranda, J. C. & Tseng, W.-Y. I. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLOS One* **8**, e80713 (2013).

4. Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34**, 144–155 (2007).

5. Tournier, J.-D., Calamante, F. & Connelly, A. MRtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* **22**, 53–66 (2012).

6. Reisert, M. *et al.* Global fiber reconstruction becomes practical. *NeuroImage* **54**, 955–962 (2011).

7. Alexander, D. C. & Seunarine, K. K. Mathematics of crossing fibers. *Diffusion MRI: Theory, Methods, and Application* 451–464 (2010).

8. Tournier, J.-D. The biophysics of crossing fibers. *Diffusion MRI: theory, methods, and application* 465–81 (2010).

9. Rheault, F., Poulin, P., Caron, A. V., St-Onge, E. & Descoteaux, M. Common misconceptions, hidden biases and modern challenges of dMRI tractography. *Journal of Neural Engineering* **17**, 011001 (2020).

10. Schilling, K. G. *et al.* Prevalence of white matter pathways coming into a single white matter voxel orientation: The bottleneck issue in tractography. *Human Brain Mapping* **43**, 1196–1213 (2022).

11. Cai, L. Y. *et al.* Convolutional-recurrent neural networks approximate diffusion tractography from T1-weighted MRI and associated anatomical context. *bioRxiv* (2023).

12. Neher, P. F., Götz, M., Norajitra, T., Weber, C. & Maier-Hein, K. H. A machine learning based approach to fiber tractography using classifier voting. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. LNCS 9349, 45–52 (Springer, 2015).

13. Poulin, P. *et al.* Learn to track: deep learning for tractography. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. LNCS 10433, 540–547 (Springer, 2017).

14. Wegmayr, V., Giuliari, G., Holdener, S. & Buhmann, J. Data-driven fiber tractography with neural networks. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 1030–1033 (IEEE, 2018).

15. Théberge, A., Desrosiers, C., Descoteaux, M. & Jodoin, P.-M. Track-to-Learn: A general framework for tractography with deep reinforcement learning. *Medical Image Analysis* **72**, 102093 (2021).

16. Benou, I. & Riklin Raviv, T. Deeptract: A probabilistic deep learning framework for white matter fiber tractography. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. LNCS 11766, 626–635 (Springer, 2019).

17. Hosseini, S. *et al.* CTtrack: A CNN+ transformer-based framework for fiber orientation estimation & tractography. *Neuroscience Informatics* **2**, 100099 (2022).

18. Lin, Z. *et al.* Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. *Medical Physics* **46**, 3101–3116 (2019).

19. Wegmayr, V. & Buhmann, J. M. Entrack: Probabilistic spherical regression with entropy regularization for fiber tractography. *International Journal of Computer Vision* **129**, 656–680 (2021).

20. Feng, Y., Chandio, B. Q., Thomopoulos, S. I., Chattopadhyay, T. & Thompson, P. M. Variational autoencoders for generating synthetic tractography-based bundle templates in a low-data setting. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (2023).

21. Wasserthal, J., Neher, P. & Maier-Hein, K. H. TractSeg – fast and accurate white matter tract segmentation. *NeuroImage* (2018).

22. Sotiropoulos, S. N. *et al.* Advances in diffusion MRI acquisition and processing in the human connectome project. *NeuroImage* (2013).

23. Van Essen, D. C. *et al.* The WU-Minn human connectome project: an overview. *NeuroImage* (2013).

24. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the human connectome project. *NeuroImage* (2013).

25. Mori, S., Wakana, S., Van Zijl, P. C. & Nagae-Poetscher, L. *MRI Atlas of Human White Matter* (Elsevier, 2005).

26. Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics* **8**, 8 (2014).

27. Poličar, P. G., Stražar, M. & Zupan, B. openTSNE: A modular Python library for t-SNE dimensionality reduction and embedding. *Journal of Statistical Software* **109**, 1–30 (2024).

28. Chandio, B. Q. *et al.* Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. *Scientific Reports* **10**, 17149 (2020).

29. Côté, M.-A. *et al.* Tractometer: towards validation of tractography pipelines. *Medical Image Analysis* **17**, 844–857 (2013).

30. Renauld, E., Théberge, A., Petit, L., Houde, J.-C. & Descoteaux, M. Validate your white matter tractography algorithms with a reappraised ISMRM 2015 tractography challenge scoring system. *Scientific Reports* **13**, 2347 (2023).

31. Legarreta, J. H., Petit, L., Jodoin, P.-M. & Descoteaux, M. Generative sampling in bundle tractography using autoencoders (gesta). *Medical Image Analysis* **85**, 102761 (2023).

32. Maier-Hein, K. *et al.* Tractography challenge ISMRM 2015 data (2015). URL `http://dx.doi.org/10.5281/zenodo.572345`.

33. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

34. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience* **6**, 171 (2012).

35. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* **12**, 26–41 (2008).

36. Legarreta, J. H. *et al.* Filtering in tractography using autoencoders (FINTA). *Medical Image Analysis* **72**, 102126 (2021).

37. Perez, E., Strub, F., De Vries, H., Dumoulin, V. & Courville, A. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32 (2018).

38. Higgins, I. *et al.* Beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations* (2016).

39. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

40. Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137 (2019).

41. Abadi, M. *et al.* Tensorflow: Large-scale machine learning on heterogeneous systems. Mountain View, CA: Tensorflow (2015).

42. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18**, 203–211 (2021).

43. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

44. Yeh, F.-C. Shape analysis of the human association pathways. *NeuroImage* **223**, 117329 (2020).

45. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1060 (2010).