Edge-Convolution Latent Inference for Brain Tractography Embedding

Samuel Kostadinov[1,2], Pierfrancesco Ardino[1], Ludovico Coletta[1,3], Luca Zigiotto[4], Silvio Sarubbo[4,5], Laurent Petit[6,7], and Paolo Avesani[1,3]

[1] Neuroinformatics Laboratory, Fondazione Bruno Kessler, Italy
[2] Department of Engineering and Information Sciences, University of Trento, Italy
[3] Centre of Mind/Brain Sciences, University of Trento, Italy
[4] ASUIT, S. Chiara Hospital, Trento, Italy
[5] CISMed, University of Trento, Italy
[6] Groupe d'Imagerie Neurofonctionnelle, Institut des Maladies Neurodégénératives (GIN-IMN),UMR5293, CNRS, CEA, Université de Bordeaux, Bordeaux, France
[7] IRP/LIA OpTeam, CNRS Biologie et Université de Bordeaux, France - Université de Sherbrooke, Canada
{skostadinov, pardino, lcoletta, avesani}@fbk.eu,
luca.zigiotto@asuit.tn.it, silvio.sarubbo@unitn.it,
laurent.petit@u-bordeaux.fr

Abstract. Brain connectivity studies increasingly use diffusion MRI tractography to build whole-brain structural networks as tractograms with millions of streamlines. But streamlines are variable-length polylines with irregular point spacing, which makes geometric processing and distance computation challenging, and complicates downstream tasks such as artifact detection, bundle identification, streamline manipulation, and tractogram comparison. To address this issue, common fixes either resample streamlines to a fixed length (distorting geometry) or learn distance-based embeddings (e.g., dissimilarity representations, autoencoders) that require aggressive resampling and often generalize poorly across subjects. Deep models also treat streamlines as unordered point clouds, discarding the edge relationships that define anatomical trajectories. We present ECLIPSE (Edge-Convolution Latent Inference for Path/Stream Embedding), a geometry-aware framework that embeds streamlines by explicitly modeling point-to-point edges along the path. Using edge convolution to propagate information between neighboring vertices, ECLIPSE preserves sequential structure while capturing smoothness, curvature, and local anatomical context. Across datasets spanning different step sizes and tractography algorithms, ECLIPSE generalizes robustly to geometric and tractography variability, yielding more faithful embeddings for downstream tasks – including bundle identification and clustering – than existing streamline embedding methods.

Keywords: Tractography · Geometric Deep Learning · Autoencoder · Unsupervised Learning

1 Introduction

Brain connectivity offers a unifying framework to study how cortical and subcortical regions interact through large-scale communication pathways, rather than as isolated modules. Diffusion MRI provides a non-invasive window onto this architecture by measuring the directionality of water diffusion in white matter, enabling tractography algorithms to reconstruct putative fiber pathways linking distant regions. These techniques have driven a shift in neuroimaging: from analyzing scalar signals on a three-dimensional voxel grid to representing the brain as a graph, in which nodes correspond to regions of interest and edges encode structural or functional connections. A tractogram is a whole-brain reconstruction of macroscale white matter connections obtained by using tracking algorithms on diffusion-weighted MRI data (DWI). It consists of millions of streamlines, each representing a structural pathway throughout the brain. Streamlines are encoded as polylines, that is, sequences of connected 3D points, without intrinsic directional information. Different streamlines can contain varying numbers of points, and the spacing between consecutive points along a streamline can be non-uniform across the tractogram, with important implications for both geometric processing and quantitative analysis. Tractogram processing encompasses several core tasks, including artifact detection, bundle identification, and tractogram-to-tractogram similarity. These operations rely on the notion of distance between streamlines, with Hausdorff distance being a popular choice, as it can compare vector representations of different lengths. However, many scalable off-the-shelf algorithms that assume Euclidean geometry cannot be used, limiting the direct adoption of conventional machine-learning pipelines. A common workaround is to resample each streamline to a fixed number of points with uniform spacing. This makes downstream computation easier, but it is inherently lossy and can distort geometry. Short streamlines are often oversampled, amplifying interpolation artifacts (e.g., spline-induced deviations), whereas long streamlines are undersampled, erasing fine-scale anatomical detail. More principled approaches replace ad hoc heuristics with geometry-informed embeddings that map each streamline into a fixed-length representational space, enabling the straightforward application of standard machine learning and statistical tools. Within this direction, dissimilarity representations [15] offer a general strategy in which each streamline is encoded by its distances to a set of prototypes, as adopted in Tractome [16] for whole-brain clustering. Deep learning-based autoencoders, such as FINTA [11], have recently been proposed to learn task-oriented embeddings for bundle identification, compressing streamline geometry into low-dimensional latent spaces. However, both strategies have drawbacks: dissimilarity representations are typically tied to a specific subject or tractogram and therefore do not generalize naturally across individuals, whereas approaches like FINTA require resampling to a fixed number of points, reintroducing the interpolation biases and geometric distortions that embedding methods were meant to avoid.

Another class of deep learning approaches represents streamlines as unordered point clouds, which discard the edge structure between consecutive points that encodes the path's trajectory. This over-simplification limits the

ability to capture key anatomical properties of white matter trajectories, such as smoothness, curvature changes, and local interactions with neighboring structures, which are inherently defined along the streamline geometry. To address this limitation, a new family of embedding methods able to incorporate the path structure – rather than relying solely on global point aggregation – is needed.

To fill this gap, we propose ECLIPSE (*Edge-Convolution Latent Inference for Path/Stream Embedding*), a geometry-aware framework that leverages edge convolution to capture anatomically meaningful structure in tractography.

By explicitly modeling local geometric relationships between neighboring points, ECLIPSE progressively propagates information along streamlines and encodes how trajectories evolve in space. In contrast to architectures such as FINTA, which primarily depend on global pooling of point-wise features, ECLIPSE preserves the sequential structure of streamlines and aggregates edge-informed features into a latent representation that is more faithful to the underlying anatomy. In doing so, it provides a more principled embedding for downstream tasks such as bundle identification and clustering. Through extensive experimental evaluation, we report results from a set of experiments designed to validate the properties of the proposed model and assess its effectiveness across a range of challenging scenarios. In particular, we evaluate its robustness and generalization in datasets with different step sizes, in multiple tractography algorithms, and in a clinical population, thus demonstrating the model's ability to handle geometric variability and tractography heterogeneity.

2 Related works

2.1 Representation learning and autoencoders

Learning compact, fixed-dimensional representations is a core requirement in many brain tractography applications, and autoencoders have become the dominant approach for learning streamline embeddings [3]. Numerous variants have been explored, such as variational [8] and denoising autoencoders [24], and applied to downstream tasks such as tractogram post-processing [10], bundle segmentation [5], or streamline clustering [9].

However, most existing models rely on convolutional or recurrent architectures that treat streamlines as fixed-length sequences. For instance, FINTA [11] applies one-dimensional convolutions to resampled streamlines, enforcing a fixed point count that depends on the sampling resolution and cannot generalize across step sizes. Furthermore, a fixed point count results in high preprocessing overhead and suffers from over-smoothing in reconstruction. Recurrent alternatives [25] avoid explicit resampling via padding, but still rely on a strong directional sequence prior – at odds with the inherently undirected nature of tractography streamlines. Other approaches, such as Siamese networks [4] or supervised multiscale embeddings [7], rely on annotated ground-truth data and thus offer limited flexibility. Taken together, these limitations highlight the need for an autoencoder that operates directly on variable-length streamlines, preserves local geometric structure, and avoids global sequence assumptions.

2.2 Geometric learning

Geometric deep learning has demonstrated strong potential for learning from irregular, non-Euclidean data. Graph neural networks and point-based models (such as PointNet [18] and Graph U-Net [6], respectively) have recently achieved notable success in shape analysis and molecular modeling [2]. However, these architectures are not tailored to tractography streamlines: PointNet-style global pooling discards ordered local geometry, while graph pooling operations modify graph topology and resolution, which is suboptimal for thin, path-like structures.

Even though edge convolution provides a locality-preserving alternative that enables progressive information propagation while maintaining the original graph structure, geometric learning approaches remain largely underexplored in tractography. A notable exception is Verifyber [1], which leverages edge convolution for anatomical plausibility classification, but does not address latent embedding learning or reconstruction.

To address these limitations, we introduce ECLIPSE (*Edge-Convolution Latent Inference for Path/Stream Embedding*), a graph-based autoencoder inspired by Verifyber that learns geometry-aware embeddings directly from variable-length, native-resolution streamlines without global pooling or graph coarsening.

3 Material and methods

3.1 Datasets

BIL&GIN In this work, we used four datasets. The main dataset, used for training and testing, is BIL&GIN [13]. This dataset comprises 453 healthy subjects acquired with a 3T MRI scanner. The detailed acquisition protocol is described in the accompanying publication [13]. Whole-brain tractograms were generated via diffusion tensor imaging (DTI) and probabilistic tracking. We used the tractograms of 206 subjects, whose tractograms contain, on average, 5.35 million streamlines (min-max= 4.16-7.01M).

HCP As one of the validation datasets, we used data from the Human Connectome Project (HCP, 3T multishell, 41 subjects of the test-retest young adult release) [23]. Whole-brain tractograms were generated via spherical constrained deconvolution (CSD) and anatomically constrained probabilistic tractography [22]. Each tractogram has, on average, 2.99 million streamlines (min-max= 2.96-3.0M).

TractoInferno Furthermore, we used a subset of participants from TractoInferno [17]. This dataset originally comprises 354 healthy subjects with DWIs acquired on different scanners and under different imaging protocols. As mentioned in [17], DTI-based tractographies were generated using: (i) deterministic (DET); (ii) probabilistic (prob);(iii) particle-filtered (PFT);(iv) and surface-enhanced (SET) approaches. We obtained 20 subjects from the original authors. For each

subject, the four tractographies were evaluated independently. The number of streamlines varies substantially, ranging from a minimum of 367,290 in PFT to a maximum of 5,070,308 in SET.

Clinical Data To validate our results in a clinical setting, we used a dataset from the Santa Chiara hospital in Trento [20]. The dataset comprises 31 subjects with glioma, whose scans were acquired with a 1.5T scanner. Whole-brain tractograms were generated via CSD using probabilistic tractography. We extracted streamlines with an estimated displacement of at least 5mm and analyzed only those streamlines.

Data preprocessing The BIL&GIN tractograms were aligned in the Johns Hopkins University (JHU) space [14]. To ensure comparability across datasets, we aligned the tractograms from HCP and TractoInferno to the same reference space using a non-linear transformation. For the clinical data, we used a rigid transformation to preserve tumor-induced geometry alterations. Streamlines exiting the image's bounding box due to registration were removed (less than 1% across datasets). For the BIL&GIN and HCP datasets, we leveraged the availability of binary plausibility labels to divide streamlines into anatomically plausible and non-plausible subsets, as previously described [1]. In the BIL&GIN dataset, only plausible streamlines were used for training, while we tested all the models on both subsets (separately). In the HCP dataset, all models were evaluated only on the plausible subset. BIL&GIN and HCP have tractograms sampled with a constant 1mm step size, while TractoInferno had compressed streamlines. To assess the generalization capabilities of our model, we resampled the BIL&GIN data to 2mm, 3mm, and also with compressed streamlines using a non–constant step size. Lastly, we prepared the data sampled at 256 points per streamline, as required for training FINTA.

3.2 Models

The proposed model, ECLIPSE, is inspired by Verifyber, an edge-convolution-based network introduced for detecting tractography artifacts. In Verifyber, each streamline is represented as a graph, $G = (V, E)$, where nodes correspond to 3D points, and edges connect consecutive points along the trajectory. With ECLIPSE, we exploit this intuition and local geometric message passing to learn a latent representation of the streamlines without resampling using an autoencoder. Let $\mathbf{x}_i \in \mathbb{R}^d$ denote the feature vector associated with node $\mathbf{v}_i$. Edge convolution (EC) updates node features by aggregating information from neighboring nodes $\mathbf{v}_j \in \mathcal{N}(i)$. Specifically, edge features are computed as:

$$\mathbf{e}_{ij} = h_{\Theta}(x_i \oplus (x_j - x_i)) \tag{1}$$

where, h_{Θ} is a learnable function and $\oplus$ denotes concatenation. Updated node features are obtained via a symmetric aggregation operator $\square$ (e.g., max pooling),

$$\mathbf{x}'_i = \square_{j \in \mathcal{N}(i)} \mathbf{e}_{ij} \tag{2}$$

Edge Convolution (EC) has two implementation variants: EC and Dynamic Edge Convolution (DEC). While mathematically identical, they differ in how neighborhoods are computed. EC requires the neighborhood as input (fixed), whereas DEC dynamically computes it with k-Nearest-Neighbor search in feature space. This dynamic computation allows information to flow between similar nodes, but introduces the notion of distance into the network. Building on these operators, ECLIPSE consists of two geometric encoding layers, followed by a fully connected bottleneck that produces a compact latent representation Figure 1. Finally, a decoder mirrors the encoder to reconstruct the input streamline. The model is trained end-to-end using a mean squared error (MSE) reconstruction loss, thereby encouraging the bottleneck to capture informative, geometry-aware embeddings. We evaluate two variants of ECLIPSE, both with a fixed embedding dimensionality of 32: one employing two EC layers, and one combining an EC layer with a DEC layer. In both cases, node-level latent features are aggregated to form a single embedding per streamline. All models have been trained using the AdamW optimizer [12] with a learning rate of 0.1, weight decay of 5×10^{-4}, and a step-based learning rate scheduler (step size 10, decay factor 0.7, minimum learning rate 6.68×10^{-4}). During training, using the whole brain tractogram proved to be computationally unfeasible. To address this issue, at each epoch, the model uses 10000 randomly sampled streamlines from each subject.

We release the source code and the checkpoints of our model and the FINTA implementation through `http://doi.org/10.5281/zenodo.17981408`.

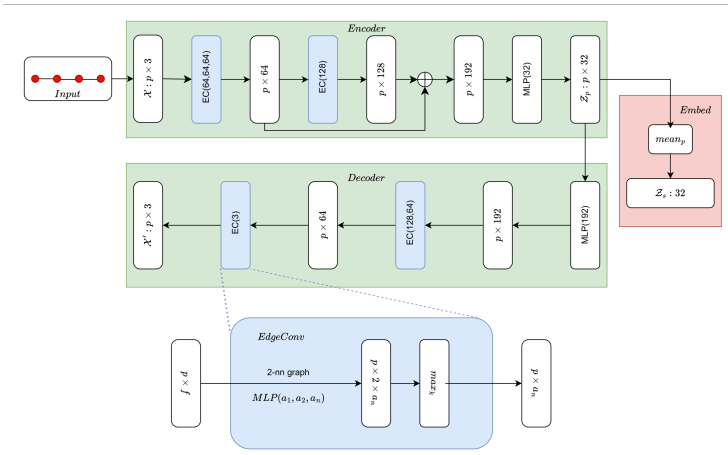

Fig. 1: ECLIPSE Architecture. The encoder-decoder network model with the embedding space and the details of the edge convolution step.

4 Experiments and Results

This section presents the experimental evaluation and results obtained for all considered models, including the different variants of ECLIPSE and the baseline FINTA. Specifically, Section 4.1 reports results for streamline reconstruction, while Section 4.2 evaluates the robustness of the models under varying experimental conditions. Section 4.3 analyzes computational efficiency and Section 4.4 provides a detailed characterization of reconstruction errors across all models. Finally, Section 4.5 compares the learned representations in downstream tasks to highlight differences in model behavior and applicability.

4.1 Reconstruction error

FINTA We first trained FINTA under the same variable-length setting adopted for ECLIPSE. Since FINTA requires fixed-size inputs, variable-length streamlines were padded to 256 points by replicating the final point. Padding positions were masked during both input processing and loss computation. This configuration enables evaluation on variable-length streamlines without resampling and serves as the primary baseline for robustness experiments. Under this setting, FINTA achieved a test mean squared error (MSE) of 0.475, as denoted in Table 1. For a fair comparison under identical input conditions, we additionally trained and evaluated FINTA using streamlines resampled to a fixed length of 256 points, following the original formulation of the model. In this setting, FINTA achieved a lower test MSE of 0.284, reflecting the benefits of operating on uniformly resampled data without the drawback of padding but at the cost of increased preprocessing and reduced flexibility.

ECLIPSE We then trained the proposed ECLIPSE autoencoder. In its initial configuration, the encoder consists of one edge convolution (EC) layer followed by one dynamic edge convolution (DEC) layer. As described in Section 3.2, EC operates on a fixed graph defined by the streamline topology, while DEC dynamically recomputes neighborhoods using k-nearest neighbors in feature space. A fully connected bottleneck projects the encoded features into a latent space, and a symmetric decoder reconstructs the input streamline. When evaluated on resampled streamlines with 256 points — matching the fixed-length setting used for FINTA – this configuration, referred to as ECLIPSE_DEC, achieved a test MSE of 0.147 in the variable-length setting and an MSE of 0.09 in the resampled setting, outperforming FINTA under identical input conditions. Since DEC explicitly incorporates Euclidean distance into neighborhood construction, we further investigated whether this operation could introduce sensitivity to sampling resolution and step-size variability. To this end, we trained a second variant of ECLIPSE with the same architecture, replacing the DEC layer with a second EC layer. This modification removes distance-based neighborhood recomputation and constrains message passing to the native streamline topology. This variant, referred to as ECLIPSE, further reduced the reconstruction error,

Table 1: Reconstruction comparison in two settings. On the left, the step size is fixed. On the right, reconstruction error under variable step-size conditions, including constant resampling (256 points per streamline) and mixed step sizes, as encountered in compressed tractograms.

Model	**fixed step size**			**flexible step size**	
	1 *mm*	2 *mm*	3 *mm*	*const*	*mixed*
FINTA (padding)	0.47	4.11	13.02	45.77	22.45
FINTA (resampled)				0.28	
ECLIPSE (DEC)	0.14	2.36	15.73	0.09	81.59
ECLIPSE	**0.06**	**0.24**	**1.58**	**0.07**	**5.98**

achieving a test MSE of 0.06 in the first setting, and 0.08 in the second setting. In the remainder of this work, ECLIPSE denotes the model composed exclusively of EC layers, while ECLIPSE_DEC refers to the variant incorporating a DEC layer.

Qualitative evaluation After training both FINTA and ECLIPSE, we analyzed and compared the reconstructed streamlines produced by the two models. The empirical results indicate that FINTA tends to oversmooth the geometric structure of the streamlines, whereas ECLIPSE better preserves local variations and sharper changes in trajectory. A qualitative comparison of representative reconstructions is illustrated in Figure 2.

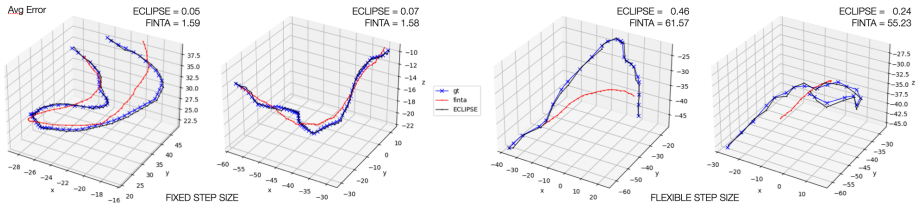

Fig. 2: Qualitative comparison of streamline reconstruction. The two left panels show the reconstruction at a 1mm step size with ground truth reference (black curve). FINTA (red curve) tends to smooth the streamline, while ECLIPSE (blue curve) preserves the original geometry. The two right panels show examples of reconstruction in a compressed-streamline setting. FINTA shows significant oversmoothing or complete failure in reconstructing the streamline, while ECLIPSE better preserves the geometry. On top, the reconstruction error of the two models.

4.2 Error characterization

To gain deeper insight into the behavior of ECLIPSE, we analyze its reconstruction errors and potential biases relative to FINTA. Specifically, we examine the relationship between reconstruction error and streamline length, as well as the correspondence between distances in the learned embedding space and geometric distances between streamlines. Finally, we demonstrate that simpler heuristics, such as uniform resampling, yield lossy representations and fail to preserve meaningful geometric information.

Correlation between error and streamline length We first analyzed potential biases of the models with respect to the streamline length by evaluating both FINTA and ECLIPSE on all subjects in the test set and computing the average reconstruction error per streamline. The results indicate that the reconstruction error of FINTA increases with streamline length relative to that of ECLIPSE, suggesting a length-dependent bias. A representative visualization of this behavior is shown in Figure 3.

Correlation between embedding distance and real distance We further investigated the relationship between distances in the learned embedding space and distances in physical space. For each subject in the test set, a random subset of streamline pairs was sampled, and embeddings were computed using both FINTA and ECLIPSE. For each pair, the MDF Hausdorff distance between streamlines and the cosine similarity between the corresponding embeddings were computed. Both models show a correlation between embedding and geometric distance. However, the embeddings produced by FINTA appear sparser and less stable for geometrically close streamlines, whereas ECLIPSE yields more compact and consistent representations. This difference may stem from the underlying architectural choices: the global embedding strategy employed by FINTA may emphasize absolute streamline position, while the geometry-aware design of ECLIPSE prioritizes local point relationships, resulting in embeddings that reflect similar pathways rather than merely similar spatial locations. A representative visualization of this behavior is shown in Figure 3.

Resampling approximation loss As discussed in Section 1, a common heuristic for handling variable-length streamlines is to resample them to a fixed number of points. Although computationally efficient, this strategy introduces lossy distortions. When the target number of points exceeds the original point count, interpolation alters the geometric representation; conversely, when the target number is lower, fine-grained geometric details may be discarded. To quantify this effect, we computed the distance between randomly sampled streamlines and their resampled counterparts with a fixed number of points. Distances were measured using dynamic time warping [21] (DTW), a widely used metric for sequence comparison that estimates the optimal alignment between two sequences. The results show that the greater the discrepancy between the native point count and

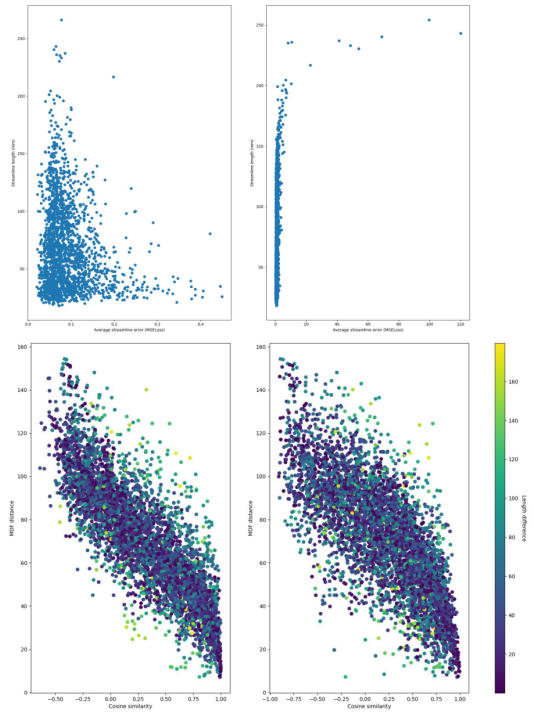

Fig. 3: Analysis of reconstruction and embedding behavior for FINTA and ECLIPSE. (Top) Relationship between average reconstruction error and streamline length, showing a length-dependent increase in error for FINTA (top right), while ECLIPSE (top left) remains largely invariant to streamline length. (Bottom) Relationship between embedding cosine similarity and geometric distance (MDF Hausdorff), illustrating a clearer and more stable correspondence for ECLIPSE (bottom left), particularly for geometrically close streamlines, compared to sparser and less stable embeddings produced by FINTA (bottom right).

the resampled point count, the larger the deviation from the original streamline geometry.

4.3 Robustness to data changes

We evaluated model robustness along four key factors: non-plausible streamlines, dataset, step size, and tractography type. To assess generalization across step sizes, streamlines were resampled at varying step sizes. For these experiments, the original FINTA model could not be directly applied due to its requirement for resampled tractograms. Therefore, we used the padded-input variant of FINTA, and unless otherwise specified, all references to FINTA correspond to this version.

Non-plausible streamlines The first experiment assessed performance on anatomically non-plausible streamlines, using a step size consistent with training data. All models generalized to these inputs, with only a modest increase in reconstruction error (Table 2).

Dataset generalization In the second experiment, models were evaluated on a previously unseen dataset (HCP), sampled at 1 mm to match the training data resolution. Similar to the non-plausible streamlines test, all models exhibited generalization, with a slight increase in reconstruction error (Table 2).

Step size The third experiment evaluated robustness to different step sizes. Multiple versions of the BIL&GIN dataset were prepared, including constant step sizes of 2mm and 3mm, as well as a compressed version with non-uniform step sizes. While this setup reduces storage requirements and I/O overhead, it tests the ability of the models to generalize to new sampling resolutions. Results indicate that FINTA and ECLIPSE_DEC fail to generalize, exhibiting a sharp increase in reconstruction errors. In contrast, ECLIPSE maintains substantially lower errors, although performance degrades slightly under extreme step-size changes (Table 1).

Tractography type The fourth test evaluated robustness to different tractography algorithms using the TractoInferno dataset. This includes four distinct tractography types, as explained in Section 3.1. Each model was independently evaluated on all four types. As these tractograms consist of compressed streamlines, higher reconstruction errors were expected. Consistent with previous results, FINTA and ECLIPSE_DEC exhibited limited generalization across tracking algorithms, whereas ECLIPSE consistently achieved lower reconstruction errors (Table 2).

Clinical data Lastly, we evaluated the robustness of the model on clinical data. As explained in Section 3.1, we used a dataset considering only streamlines deformed by the presence of a glioma. The results of this test indicate that FINTA and ECLIPSE_DEC both maintain relatively low error, but ECLIPSE achieves better results, as shown in Table 2.

4.4 Model efficiency

As mentioned in Section 3.1, we prepared the dataset in different ways to test our models. The direct consequence is that, depending on sampling size, the data occupy different amounts of disk space. As an example, considering both plausible and non-plausible streamlines, the original BIL&GIN dataset occupies 909 GB of space. The version resampled for FINTA to 256 points 3.1 TB, while the version with compressed streamlines occupies 261 GB. The versions resampled at constant step sizes of 2mm and 3mm occupy 459 GB and 310 GB, respectively. We

Table 2: Generalization Error. The first column shows the reconstruction error on non-plausible streamlines while the autoencoder was trained on plausible ones. The second left column reports the error on a dataset different from the training set (HCP vs B&G), with a fixed step size of 1 mm. Columns 3–6 report the reconstruction error for specific types of compressed tractography on TractoInferno: deterministic, probabilistic, particle filtering, and superficial enhanced. Column 7 reports the reconstruction error for the clinical data.

Model	**B&G**	**HCP**	**TractoInferno**				**Clinical data**
	PROB	*PROB*	*DET*	*PROB*	*PFT*	*SET*	*PROB*
FINTA (padding)	0.82	0.83	49.23	15.29	16.64	20.28	1.182
ECLIPSE (DEC)	0.28	0.29	238.45	78.31	58.07	107.50	0.467
ECLIPSE	**0.08**	**0.07**	**27.85**	**3.83**	**5.57**	**13.68**	**0.201**

also compared the models in terms of network size and number of parameters. FINTA contains approximately 4.7 million parameters, whereas ECLIPSE has only 97 thousand parameters. These results indicate that ECLIPSE achieves substantially lower model complexity and reduced data footprint while maintaining superior reconstruction performance relative to FINTA.

4.5 Streamline clustering

A further evaluation investigates the impact of different embeddings on streamline clustering, a common downstream task. The analysis is conducted using an atlas of white matter bundles [19]. We hypothesize that clusters should group streamlines according to the canonical pathways of known bundles, thereby allowing assessment of cluster quality by their coherence with established neuroanatomy. After aligning the atlas to the JHU reference space, which is also used to train the autoencoder models, we map the streamlines to the embedding spaces of FINTA and ECLIPSE, respectively. We then cluster the embedded streamlines using a mini-batch k-means algorithm, setting the number of clusters equal to the number of bundles in the atlas (N = 77). We manually select a subset of clusters by inspecting their medoids, whose trajectories resemble the skeletons of known bundles. Figure 4 illustrates how ECLIPSE, but not FINTA, can reconstruct the expected single-bundle anatomy.

5 Discussion and conclusions

ECLIPSE demonstrates effective and robust behavior as an edge-convolution-based autoencoder for brain tractography. Through extensive empirical evaluation, we show that the model learns fixed-dimensional embeddings from variable-length streamlines, while preserving latent information about the geometry and anatomy of white matter pathways. Its markedly lower reconstruction error compared

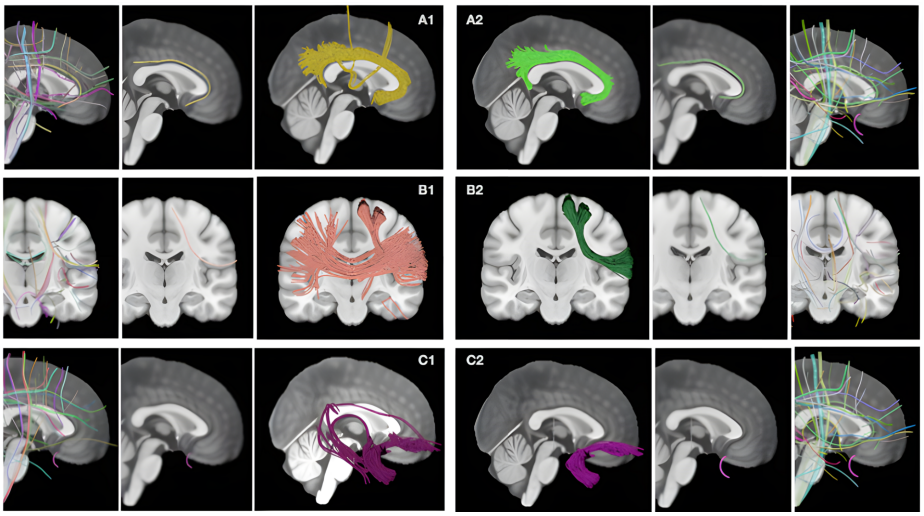

Fig. 4: Streamline clustering analysis. Comparison of the impact of two streamline embeddings in a downstream task of clustering: FINTA on the left side, ECLIPSE on the right side. We focus the assessment on three main bundles: (A) right cingulate, (B) left frontal aslant tract, (C) right uncinate fasciculus. For each use case, we report the visualization of the 77 clusters' medoid, the selected cluster, and the associated streamlines.

with FINTA indicates that ECLIPSE provides a substantially improved representation learning framework for tractography data, highlighting a meaningful advancement in streamline embedding for connectomics studies. The pairwise evaluation of the two models investigates the relation between the real-world streamline distances with distances in the learned embedding space. Each showed a correlation, but FINTA produced a sparser, less stable distribution of latent distances for geometrically similar streamlines (Figure 3). Moreover, the reconstruction error of FINTA increased with streamline length, whereas ECLIPSE maintained more uniform reconstruction accuracy across short and long trajectories, indicating a more balanced and stable geometric representation (Figure 3).

Both ECLIPSE and FINTA generalize to non-plausible streamlines and to data acquired and processed with different pipelines, indicating that neither model overfits to a narrow distribution of plausible training examples. However, when assessing generalization across different step sizes and tractography types, FINTA fails to generalize, whereas ECLIPSE maintains substantially lower reconstruction error (Table 2). Moreover, ECLIPSE shows better generalization capabilities to a clinical setting, where streamlines may be distored and noisier, increasing the reliability of the method in clinical scenarios. This behavior highlights the importance of local geometric operators compared to global ones. From a computational standpoint, ECLIPSE is also more efficient than FINTA. By avoiding streamline resampling and operating directly on the original stream-

lines, the model reduces both network complexity and data footprint. This leads to lower memory consumption and diminished input/output overhead, which are critical advantages when processing whole-brain tractograms. The preliminary assessment on a downstream task, such as clustering, is promising. The embeddings produced by ECLIPSE yield more anatomically coherent and better separated clusters than those obtained with FINTA (Figure 4), indicating that the proposed representation captures more discriminative streamline features related to anatomical pathways. ECLIPSE also has several limitations that point to important directions for future work. First, the model embeds each streamline in isolation, without explicitly considering the context provided by neighboring pathways. A context-aware embedding, which integrates local tract neighborhood information, could help mitigate noise and reconstruction artifacts that are intrinsic to tractography. Second, although the extensive evaluation supports the validity of the learned representation space, the current experiments are restricted to glioma, making the study of different conditions, such as lesions or malformations, necessary.

References

1. Astolfi, P., Verhagen, R., Petit, L., Olivetti, E., Sarubbo, S., Masci, J., Boscaini, D., Avesani, P.: Supervised tractogram filtering using geometric deep learning. Medical Image Analysis **90**, 102893 (2023)
2. Atz, K., Grisoni, F., Schneider, G.: Geometric deep learning on molecular representations. Nature Machine Intelligence **3**(12), 1023–1032 (2021)
3. Berahmand, K., Daneshfar, F., Salehi, E.S., Li, Y., Xu, Y.: Autoencoders and their applications in machine learning: a survey. Artificial intelligence review **57**(2), 28 (2024)
4. Chen, Y., Zhang, C., Song, Y., Makris, N., Rathi, Y., Cai, W., Zhang, F., O'Donnell, L.J.: Deep fiber clustering: anatomically informed unsupervised deep learning for fast and effective white matter parcellation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 497–507. Springer (2021)
5. Dumais, F., Legarreta, J.H., Lemaire, C., Poulin, P., Rheault, F., Petit, L., Barakovic, M., Magon, S., Descoteaux, M., Jodoin, P.M., et al.: Fiesta: Autoencoders for accurate fiber segmentation in tractography. NeuroImage **279**, 120288 (2023)
6. Gao, H., Ji, S.: Graph u-nets. CoRR **abs/1905.05178** (2019), `http://arxiv.org/abs/1905.05178`
7. Goel, A., Singh, B., Joshi, A., Jha, R.R., Ahuja, C., Nigam, A., Bhavsar, A.: Tractoembed: Modular multi-level embedding framework for white matter tract segmentation. In: International Conference on Pattern Recognition. pp. 240–255. Springer (2024)
8. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
9. Legarreta, J.H., Petit, L., Jodoin, P.M., Descoteaux, M.: Clustering in tractography using autoencoders (cinta). In: International Workshop on Computational Diffusion MRI. pp. 125–136. Springer (2022)

10. Legarreta, J.H., Petit, L., Jodoin, P.M., Descoteaux, M.: Generative sampling in bundle tractography using autoencoders (gesta). Medical Image Analysis **85**, 102761 (2023)
11. Legarreta, J.H., Petit, L., Rheault, F., Theaud, G., Lemaire, C., Descoteaux, M., Jodoin, P.M.: Tractography filtering using autoencoders. (2020)
12. Loshchilov, I., Hutter, F.: Fixing weight decay regularization in adam. CoRR **abs/1711.05101** (2017), http://arxiv.org/abs/1711.05101
13. Mazoyer, B., Mellet, E., Perchey, G., Zago, L., Crivello, F., Jobard, G., Delcroix, N., Vigneau, M., Leroux, G., Petit, L., et al.: Bil&gin: A neuroimaging, cognitive, behavioral, and genetic database for the study of human brain lateralization. Neuroimage **124**, 1225–1231 (2016)
14. Oishi, K., Faria, A., Jiang, H., Li, X., Akhter, K., Zhang, J., Hsu, J.T., Miller, M.I., van Zijl, P.C., Albert, M., et al.: Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and alzheimer's disease participants. Neuroimage **46**(2), 486–499 (2009)
15. Olivetti, E., Nguyen, T.B., Garyfallidis, E.: The approximation of the dissimilarity projection. In: 2012 Second International Workshop on Pattern Recognition in NeuroImaging. pp. 85–88 (2012). https://doi.org/10.1109/PRNI.2012.13
16. Porro-Muñoz, D., Olivetti, E., Sharmin, N., Nguyen, T.B., Garyfallidis, E., Avesani, P.: Tractome: a visual data mining tool for brain connectivity analysis. Data mining and knowledge discovery **29**(5), 1258–1279 (2015)
17. Poulin, P., Theaud, G., Rheault, F., St-Onge, E., Bore, A., Renauld, E., de Beaumont, L., Guay, S., Jodoin, P.M., Descoteaux, M.: Tractoinferno-a large-scale, open-source, multi-site database for machine learning dmri tractography. Scientific Data **9**(1), 725 (2022)
18. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
19. Radwan, A.M., Sunaert, S., Schilling, K., Descoteaux, M., Landman, B.A., Vandenbulcke, M., Theys, T., Dupont, P., Emsell, L.: An atlas of white matter anatomy, its variability, and reproducibility based on constrained spherical deconvolution of diffusion mri. NeuroImage **254**, 119029 (2022)
20. Riccardi, C., Coletta, L., Ghezzi, S., Amorosino, G., Zigiotto, L., Jovicich, J., Sarubbo, S., Avesani, P.: Supervised white matter bundle segmentation in glioma patients with transfer learning. Medical Image Analysis p. 103766 (2025)
21. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing **26**(1), 43–49 (2003)
22. TO, N.I.C.C.G.: Support neuroscience research: Par-02-059. Journal of Investigative Medicine **50**(3) (2002)
23. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)
24. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., Bottou, L.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research **11**(12) (2010)
25. Zhong, S., Chen, Z., Egan, G.: Auto-encoded latent representations of white matter streamlines for quantitative distance analysis. Neuroinformatics **20**(4), 1105–1120 (2022)