ROTATION-INVARIANT GRAPH TRANSFORMERS ENABLE GLOBAL BIOIMAGE REGISTRATION AND CELL MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cell matching is a key step in the joint analysis of different imaging modalities in biology. Its main challenges stem from large differences in sample orientation and geometry, local deformations, missing objects, and variations in image appearance and resolution. Commonly used registration pipelines therefore rely heavily on manual or heuristic pre-alignment. We propose *grreg*, a GNN-based approach for learning context-aware cellular representations that enable direct matching across samples. It combines a rotation- and translation-invariant GCN for local feature extraction and a graph transformer with two kinds of attention: distance-aware self-attention to improve the receptive field of local descriptors, and cross-attention to enable cross-image communication. Our method reaches high matching accuracy, does not require pre-alignment and scales to samples with over 30000 objects through a sparse attention mechanism, as demonstrated on zebrafish embryo development trajectories from Zebrafish Hub. We also introduce a new cross-modal benchmark Platy5+1 to aid further development of bioimage registration.

1 INTRODUCTION

Multimodal cellular atlases, which combine information from different types of microscopy and spatial omics technologies, integrate cellular ultrastructure, molecular measurements, and spatial tissue organization. Such atlases provide a holistic view of biological systems and serve as an invaluable resource to contextualize future studies of the same organism. Recent advances include spatial cell type atlas of mouse brain (Yao et al., 2023), spatio-temporal atlases of zebrafish (Lange et al., 2024) and Drosophila (Wang et al., 2025) development, and cellular resolution nervous system atlases and connectomes (Schlegel et al., 2024; Dorkenwald et al., 2024; Yemini et al., 2021). However, both to construct an atlas and to map new observations, correspondence between the related biological structures needs to be established across samples and modalities.

Practical pipelines rely on image registration (Bischof et al., 2024) and start with finding a coarse rigid transformation followed by multiple steps of finer deformable registration (Klein et al., 2010; Besl & McKay, 1992a; Myronenko & Song, 2010b). Physical pairwise distances between landmarks are then used as a cost matrix for the final matching. Such coarse-to-fine pipelines can be effective, but their dependence on initialization, parameter choices, and heuristics makes them difficult to use and transfer across datasets.

Recent progress in bioimage segmentation allows to reliably identify cells or other objects, therefore finding correspondences can be formulated directly as an object-matching problem. In this approach pairwise costs are typically derived from the similarities or distances between object representations (Wang & Solomon, 2019a; Sarlin et al., 2020), while assignment constraints, spatial regularization, or iterative geometric refinement can be used to resolve ambiguous matches (Myronenko & Song, 2010a; Besl & McKay, 1992b; Torresani et al., 2008).

Designing better representations has long been a central strategy for improving correspondence estimation, from hand-crafted descriptors (Lowe, 2004) to learned features for graph matching. Nevertheless, existing approaches do not fully address the combination of challenges encountered in multimodal biological imaging. Visual appearance is usually modality-specific, while absolute

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

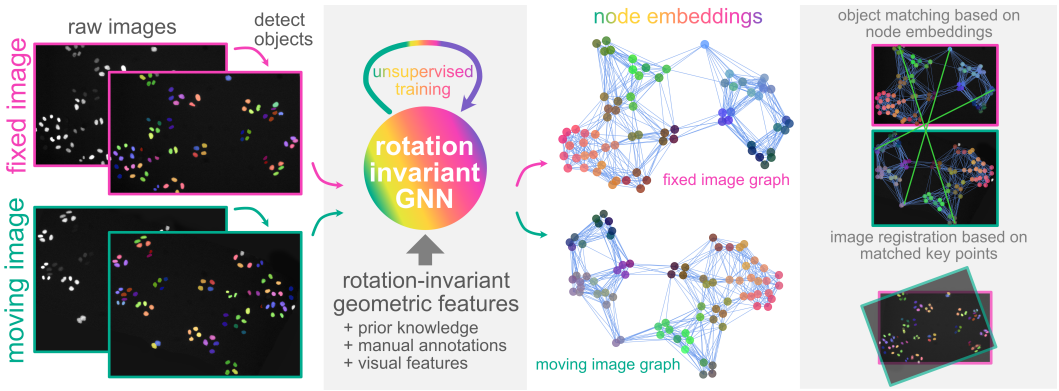


Figure 1: Biological image registration with GNNs.

spatial coordinates are not comparable across samples with arbitrary relative orientation. Cross-sample matching often involves different individuals of the same species or different sections of the same specimen, where sample preparation, elastic deformations, and biological variability introduce changes in tissue geometry. Even local geometric context can be ambiguous, as similar cellular neighborhoods may occur at multiple locations within a tissue or be affected by segmentation errors.

To handle these challenges we propose *grreg* (*gr*aph *reg*istration): a framework for biological image registration and object matching which replaces coarse-to-fine pipelines with one GNN model (Figure 1). First the model extracts local rotation-invariant geometric features specifically adapted for the point clouds without identifiable surfaces and then uses alternating self- and cross-attention to align cell representations between two samples. We introduce distance-aware self-attention for better representation of global context and improve scalability using sparse attention mechanism. We developed a graph augmentation mimicking tissue stiffness to model the difference between samples for self-supervised training with contrastive N-tuples loss. The framework was tested both on 2D and 3D datasets and was demonstrated to find correspondences between samples with up to 30k cells. In addition, we created a new benchmark dataset for cross-modal registration based on 3D electron microscopy and light microscopy of 6 days post fertilization embryo of *Platynereis dumerilii* to enable further development of bioimage registration methods.

2 Related Work

Typical bioimage registration pipelines consist of **pre-alignment** and **deformable registration**. During pre-alignment two images are registered globally using rigid or affine transform. This step is often done manually or based on a heuristic Krentzel et al. (2025); Bogovic et al. (2016). Deformable registration is usually done either with intensity-based methods like Elastix Shamonin et al. (2013); Klein et al. (2010) or point cloud registration like ICP Besl & McKay (1992a); Chen & Medioni (1992) or CPD Myronenko & Song (2010b). Both methods have disadvantages: intensity-based registration can result in implausible deformation fields, or require setting parameters anew for every sample. CPD reduces the sample to a set of points with no easy way to add visual information or other annotations. These iterative methods tend to get stuck in local minima and struggle to find the globally optimal solution.

Global cell matching is closely related to the graph matching: a problem of finding correspondence between nodes of two graphs which minimizes discrepancy in the connectivity of matched nodes. It was used to match neurons in *C. elegans* (Kainmueller et al., 2014; Stricker et al., 2026), however, computational complexity of traditional optimization methods precludes their practical usage for datasets with more than $\sim$ 100 objects (Haller et al., 2022).

This challenge has been addressed in the field of point cloud registration for robotics and CAD model analysis (Zhang et al., 2026). For example, Deep Closest Point Wang & Solomon (2019b) creates point representations with DGCNN Wang et al. (2019), produces approximate correspondences via

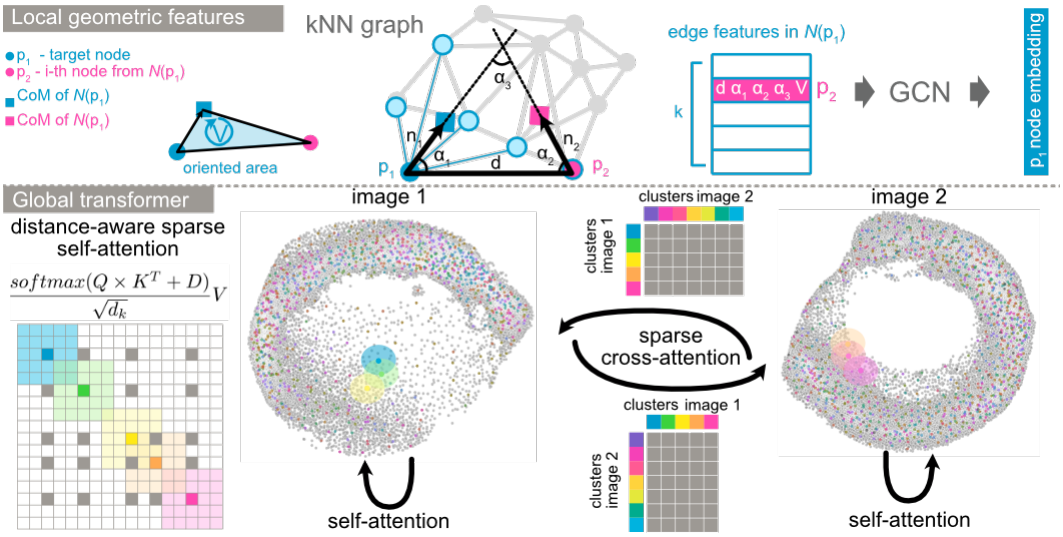


Figure 2: *grreg* architecture.

a transformer layer and predicts a rigid transformation that aligns two point clouds. PPFNet (Deng et al., 2018) learns 3D descriptors for a fixed number of patches randomly sampled from the scene with contrastive N-tuples loss. It builds a local coordinate system to calculate rotation-invariant geometric features and combines local and global information via an architecture based on PointNet (Qi et al., 2016; 2017) and max-pooling aggregation. 3D descriptors are then plugged into RANSAC to predict rigid transformation between the scenes. SuperGlue (Sarlin et al., 2020) introduces differential soft assignment layer based on Sinkhorn algorithm to train a network consisting of alternating self- and cross-attention blocks for keypoint matching in 2D images.

These methods use learnable combination of local and global features and pairwise comparison of all points between point clouds to bypass expensive combinatorial optimization, however, scalability beyond a few hundred points remains a common challenge for these architectures. In practice this limitation is circumvented by downsampling or picking only the most meaningful landmarks. This is not directly applicable for point clouds derived from instance segmentation of biological images where each object should have a match in the other sample. In addition, points do not form clear surfaces or other geometric features, therefore relative positions of cells play key role in defining local geometry (Hidalgo-Cenalmor et al., 2026).

3 Problem Formulation

Two matching biological samples are imaged and cell segmentation or other object detection method is applied to the resulting 2D or 3D images (Figure 1). Point clouds consisting of segmentation mask centroids $\boldsymbol{P}_1 = \{\boldsymbol{p}_i \in \mathbb{R}^d : i \in \{1..N_1\}\}$ and $\boldsymbol{P}_2 = \{\boldsymbol{p}_j \in \mathbb{R}^d : j \in \{1..N_2\}\}$ are extracted from each image and converted to kNN graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ using euclidean distances between points. Let $\mathcal{N}(\boldsymbol{p}_i) \subseteq \boldsymbol{P}$ be the k nearest neighbors of $\boldsymbol{p}_i$, $\boldsymbol{X}_1 \in \mathbb{R}^{N_1 \times d_X}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{N_2 \times d_X}$ be d_X-dimensional node feature matrices. For any graph $\boldsymbol{D}_{\mathcal{G}} \in \mathbb{R}^{N \times N}$ is a matrix of pairwise distances between nodes in physical space $d_{ij} = dist(\boldsymbol{p}_i, \boldsymbol{p}_j)$.

Although each cell in one sample should have a corresponding cell in the other sample and vice versa, in practice due to segmentation errors and biological differences, some objects may have multiple plausible matches or not match to anything at all. Furthermore, it is a local quality: the image with less total objects may have more objects locally in some parts of the image and less objects elsewhere. Due to this ambiguity one-to-one matching problem becomes too restrictive, therefore we use a representation learning objective instead. Let $\mathbb{M} = \{(i, j) : i \in \{1..N_1\}, j \in \{1..N_2\}\}$ be a set of index pairs of matching objects. The goal is to find node embeddings $\mathcal{E}_1$ and

Table 1: Comparison of different architectures and CPD for **hela** dataset

	GCN 6 conv		*grreg* all-to-all		*grreg* sparse		CPD pre-aligned	
	acc ↑	Δ_{mean} ↓	acc ↑	Δ_{mean} ↓	acc ↑	Δ_{mean} ↓	acc ↑	Δ_{mean} ↓
asymmetric 01 easy	0.95	0.092	**0.98**	0.037	0.97	0.057	0.98	0.09
asymmetric 01 hard	0.85	0.103	**0.94**	0.059	0.93	0.045	0.97	0.16
asymmetric 02 easy	0.90	0.132	**0.97**	0.060	**0.97**	0.066	0.96	0.15
asymmetric 02 hard	0.65	0.149	0.90	0.027	**0.92**	0.043	0.95	0.15
symmetric 01 easy	0.94	0.185	**0.97**	0.050	0.91	0.069	0.98	0.09
symmetric 01 hard	0.82	0.207	**0.90**	0.040	0.82	0.081	0.97	0.13
symmetric 02 easy	0.86	0.267	0.95	0.043	**0.96**	0.037	0.96	0.14
symmetric 02 hard	0.58	0.264	0.85	0.019	**0.90**	0.029	0.97	0.16

$\mathcal{E}_2$ such that

$$\sum_{i,j\in\mathbb{M}} dist(\mathcal{E}_{1i}, \mathcal{E}_{2j}) \to min; \sum_{i,j\notin\mathbb{M}} dist(\mathcal{E}_{1i}, \mathcal{E}_{2j}) \to max$$

Pairwise distance matrix between graph nodes in the embedding space $\boldsymbol{C} \in \mathbb{R}^{N_1 \times N_2}$, where $c_{ij} = dist(\mathcal{E}_{1i}, \mathcal{E}_{2j})$ can be used as a cost matrix for cell matching as shown in Figure 1.

4 METHODS

4.1 ARCHITECTURE

Point pair features Translation- and rotation-invariant point pair features (PPF) inspired by Deng et al. (2018) were used to encode each node of the graph. Center of mass of the neighborhood was used as a basis for a local coordinate system to eliminate the dependence of local geometric features on the global orientation of the point cloud. As shown in Figure 2, for the target point $\boldsymbol{p}_1$, the center of mass of $\mathcal{N}(\boldsymbol{p}_1)$ is $\boldsymbol{c}_1 = \sum_{j\in\mathcal{N}(\boldsymbol{p}_1)} \frac{\boldsymbol{p}_j}{K}$ and $\boldsymbol{n}_1 = \boldsymbol{c}_1 - \boldsymbol{p}_1$. Similarly for one of the points of the neighborhood $\boldsymbol{p}_2 \in (\boldsymbol{p}_1)$: $\boldsymbol{n}_2 = \boldsymbol{c}_2 - \boldsymbol{p}_2$. Vector $\boldsymbol{d} = \boldsymbol{p}_2 - \boldsymbol{p}_1$ connects the two points.

Since a lot of biological samples exhibit bilateral symmetry, oriented volume that is not reflection-invariant was added to differentiate between the symmetric pairs of cells. In 2D case $V = \frac{1}{2}det([\boldsymbol{n}_1, \boldsymbol{p}_2 - c_1, \boldsymbol{d}])$. In 3D case $V = \frac{1}{6}([\boldsymbol{n}_1, \boldsymbol{c}_2 - \boldsymbol{c}_1, \boldsymbol{n}_2, \boldsymbol{d}])$.

$$PPF(\boldsymbol{p}_1, \boldsymbol{p}_2) = \left(||\boldsymbol{d}||_2^2,\ \angle(\boldsymbol{d}, \boldsymbol{n}_1),\ \angle(\boldsymbol{d}, \boldsymbol{n}_2),\ \angle(\boldsymbol{n}_1, \boldsymbol{n}_2), V\right),$$

where $\angle(\boldsymbol{v}_1, \boldsymbol{v}_2) = atan2(||\boldsymbol{v}_1 \times \boldsymbol{v}_2||, \boldsymbol{v}_1 \cdot \boldsymbol{v}_2)$ is guaranteed to lie in the range $[0, \pi)$.

***grreg* GCN** GCN with L convolutions was used to derive the node embeddings from pairwise geometric features for each sample independently (Figure 2). The update rule is as follows:

$$\boldsymbol{x}_i' = ReLU(\max_{j\in\mathcal{N}(\boldsymbol{p}_i)} h_\Theta([\boldsymbol{x}_j, PPF(\boldsymbol{p}_i, \boldsymbol{p}_j)]))$$

where $\boldsymbol{x}_j$ are node features after previous convolution, $\boldsymbol{x}_i'$ is the new feature and h_Θ is an MLP. Residual connections have been shown to improve the performance of GCNs, therefore PPF features are concatenated to the feature vector at each iteration. The feature dimension doubles with each convolution. PPF-GCN summarizes local context and produces learnt geometric feature vector describing each cell's immediate neighborhood.

***grreg* all-to-all** Let $Q = \mathcal{X}\boldsymbol{W}_Q$, $K = \mathcal{X}\boldsymbol{W}_K$ and $\boldsymbol{V} = \mathcal{X}\boldsymbol{W}_V$ be multi-head query, key and value matrices typical for a global graph transformer architecture and $\boldsymbol{D} = \boldsymbol{D}_\mathcal{G}\boldsymbol{W}_d$. We define distance-aware self-attention as

$$\text{D-Attn}(\mathcal{X}) = \frac{softmax(\boldsymbol{Q} \times \boldsymbol{K}^T + \boldsymbol{D})}{\sqrt{d_\mathcal{X}}}\boldsymbol{V}$$

The term $\boldsymbol{Q} \times \boldsymbol{K}^T$ determines the similarity of different nodes in the embedding space while D reintroduces information about physical proximity between nodes, thus allowing to weigh the importance of local geometric features and global physical structure of the dataset in a trainable manner. Cross-attention was used to align representations between the two graphs:

$$\text{X-Attn}(\mathcal{X}^1, \mathcal{X}^2) = \frac{softmax(\mathcal{X}^1 \boldsymbol{W}_Q \times \mathcal{X}^2 \boldsymbol{W}_K^T)}{\sqrt{d_{\mathcal{X}}}} \mathcal{X}^1 \boldsymbol{W}_V$$

In both cases attention block consists of the attention layer, LayerNorm, MLP (linear layer, ReLU, linear layer) and another LayerNorm. Alternating self-attention and cross-attention are applied R times to update the node embeddings as shown in Algorithm 1.

***grreg* sparse** We created sparse versions of D-Attn Block and X-Attn Block inspired by Zaheer et al. (2020) for larger graphs. Each graph is split into overlapping clusters of the same size M. First, K-means clustering with $K = \frac{N}{M}$ random cluster centers is performed to make sure that each node belongs at least to one cluster. If any cluster has more elements than M, the clustering is rerun with $K := K + 1$. Then the closest nodes are added to each cluster until the size reaches M. To reduce the dependency on the specific clustering, during training both graphs are re-clustered during each training step. Centers of clusters are called anchor nodes (shown in different colors in Figure 2). Anchor nodes form graphs $\mathcal{A}_1$ and $\mathcal{A}_2$ which can be viewed as downsampled version of the original graphs. Their node features are initialized with PPF-GCN.

Sparse X-Attn Block is applied only to $\mathcal{A}_1$ and $\mathcal{A}_2$. D-Attn Block is modified in the following way: the attention is calculated only between nodes of $\mathcal{A}$ and nodes within each cluster, as shown in Figure 2. First D-Attn Block is calculated between nodes of $\mathcal{A}$ and then attention is applied within each cluster separately, using matching cluster sizes to parallelize computation. For nodes which belong to more than one cluster, the final embedding is an average of embeddings coming from different clusters, which ensures smooth node embeddings which do not depend on the specific random clustering.

4.2 Self-supervised training

Supervised contrastive N-tuple loss (Deng et al., 2018) was used for training. Each batch consisted of one graph and its augmented version with matching nodes serving as positive pairs.

We developed a random elastic augmentation to model big movements of parts of the graph without disturbing the local neighborhood structure. Empirically denser parts of the tissue should allow for less relative cell movement, therefore another graph is created using ball query with a dataset-specific radius and its edges are added to the kNN graph. A subset of nodes of graph is picked at random. These nodes are displaced randomly with the displacement proportional to the degree of the source node of the edge and their new positions are fixed forming graph $\mathcal{G}'$ with the same connectivity but modified node positions. Then the positions of the remaining nodes in $\mathcal{G}'$ are adjusted using L-BFGS-B to minimize the elastic energy $U = \sum_{(i,j)\in\mathcal{G}'} k_{ij}(l_{ij} - l_{ij0})^2$, where l_{ij} is a length of edge (i, j) in graph $\mathcal{G}'$, l_{ij0} is the length of the same edge in the original graph and k_{ij} is the stiffness coefficient.

After the optimization of positions kNN graph was rebuilt from scratch. To speed up training a dataset of graph pairs was generated and cached in advance. Node drop-out was applied to both graphs during training. After removing the random nodes kNN graph was rebuilt from scratch.

4.3 Cell matching

Linear assignment with pairwise distance in embedding space as a cost matrix is a common way to establish correspondences between nodes. In case of different sizes of the matched sets some elements from larger set are left out to achieve one-to-one matching. We find that even with reasonably good embeddings it leads to unexpected mistakes in the cell assignment due to the locally varying number of objects. To assess how well the embedding space is organized to reflect this, we consider $\mathcal{G}_1$ as a reference and match each point in $\mathcal{G}_2$ to its closest neighbor from $\mathcal{G}_1$ instead of LAP.

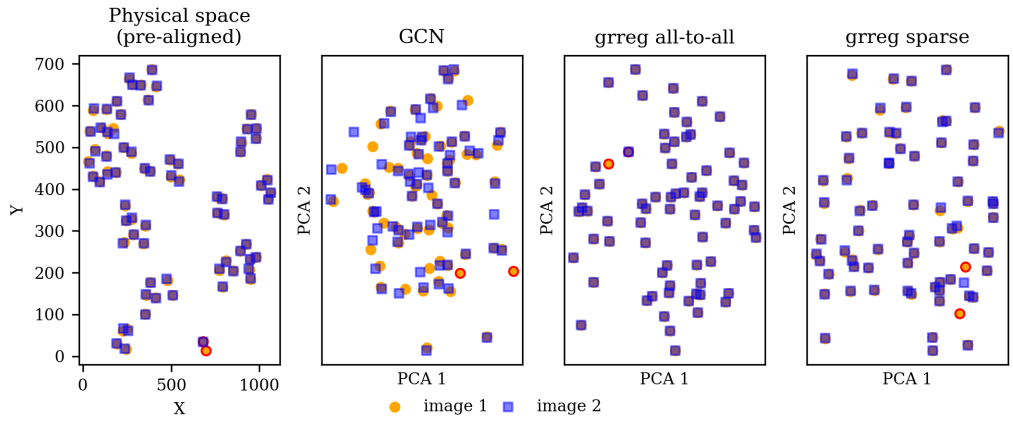


Figure 3: Comparison of embeddings produced by different architectures.

4.4 METRICS

Matching ground truth $\mathbb{M}_{GT} = \{(i,j)|i \in \mathcal{G}_1, j \in \mathcal{G}_2\}$ is a set of matching node pairs. If there is no information about the matches for the particular node, it is excluded from the metric calculation.
Accuracy/K-hop accuracy Often it is not possible to find the exact matching cell based on just the position and for some applications matching to the neighboring cell can be precise enough. In K-hop accuracy the pair is counted as correct if the predicted matching node is at most K graph hops away from its ground-truth matching node. The target graph is treated as undirected for the hop search.

$$Accuracy(\mathbb{M}, \mathbb{M}_{GT}) = \frac{\sum_{(i,j)\in M} \mathbb{1}[(i,j) \in \mathbb{M}_{GT}]}{|\mathbb{M}_{GT} \cup \mathbb{M}|}$$

Mean and maximum mismatch distance The downside of using rotation- and translation-invariant features in combination with linear or greedy assignment is that neighboring nodes from one graph may match similar-looking nodes in the completely different parts of the other image. To analyze matching mistakes, we introduce mean Δ_{mean} and maximum Δ_{max} mismatch distance metrics. Let us define a set of mismatch pairs $\mathbb{M}_{miss} = \mathbb{M} \setminus \mathbb{M}_{GT}$ - all the wrongly matched pairs of nodes. D_{max} - maximum distance between two nodes in a graph, "diameter" of the graph, and D_{min} - the shortest distance between two nodes are used to normalize the metrics:

$$\Delta_{max} = \frac{\max_{(i,j)\in M_{miss}}(dist(i,j)) - D_{min}}{D_{max} - D_{min}}; \; \Delta_{mean} = \frac{\underset{(i,j)\in M_{miss}}{mean}(dist(i,j)) - D_{min}}{D_{max} - D_{min}}$$

If Δ_{max} is close to 1, it means that the matches can be completely wrong - some nodes are matched to nodes that are in a completely opposite part of the graph. Δ_{mean} close to 0 means that all cells are assigned to the correct neighborhood, even if not exactly to the correct matching pair.

5 PLATY5+1: BENCHMARK DATASET FOR CROSS-MODAL REGISTRATION

We created a benchmark dataset for cross-modal registration and object-matching **Platy5+1** that consists of 5 confocal microscopy volumes and an electron microscopy volume of 6 days post fertilization embryos of *Platynereis dumerilii*. All images have instance segmentations, were pre-aligned and resampled to isotropic resolution of $0.129\ \mu m$. 27 cells were manually annotated in each volume to use as ground truth for cell matching. Gene expression for a marker of somatic muscles was measured in the confocal samples. Correlation between the sample's own gene expression and signal transferred from another dataset through matching serves as an additional metric to assess the registration quality. For more information about the dataset see section A.1.

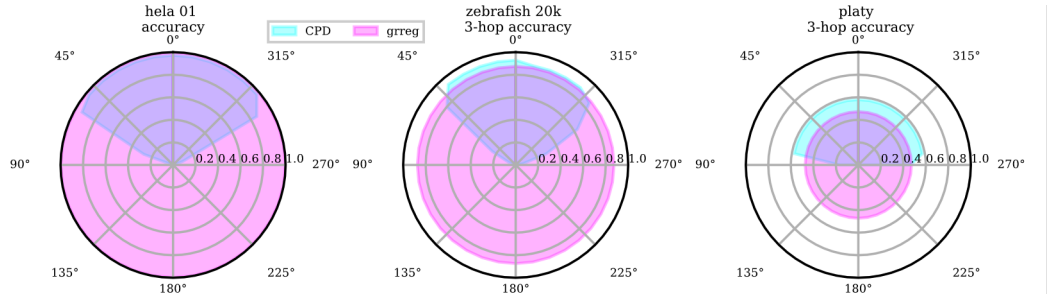


Figure 4: Comparison of the dependence of *grreg* and CPD on the pre-alignment.

Table 2: *grreg* sparse results for **platy** and **zebrafish** datasets

	grreg sparse				CPD pre-aligned		
	acc ↑	3-hop acc ↑	Δ_{mean} ↓	Δ_{max} ↓	acc ↑	3-hop acc ↑	Δ_{mean} ↓
zebrafish 20k easy	0.20	0.87	0.016	0.26	0.85	0.93	0.016
zebrafish 20k hard	0.16	0.87	0.017	0.24	0.56	0.81	0.014
zebrafish 30k easy	0.15	0.63	0.025	0.56	0.78	0.88	0.015
zebrafish 30k hard	0.12	0.63	0.026	0.58	0.68	0.82	0.016
platy	0.09	0.47	0.027	0.15	0.17	0.58	0.020

6 EXPERIMENTS

Cell tracking datasets were used to create registration problems with full ground truth derived from the tracks. The difficulty of the problem can be easily calibrated by changing the time interval between the sampled frames ($dt = 3$ for "easy" and $dt = 6$ for "hard" problems). Cell divisions, movement and segmentation errors mimic the variability between the samples in matching.

hela: 100-400 cells. Ground truth data from Fluo-N2DL-HeLa dataset Neumann et al. (2010) from the Cell tracking challenge Maška et al. (2023) has two tracks named 01 and 02. Since datasets with repeating neighborhood structures and specifically symmetric datasets are especially challenging for *grreg* framework, we created symmetric version of each track which consisted of the original images and their mirrored versions put side by side. 10 augmented samples per frame were generated with 10% of nodes moved by a maximum 5% of the graph diameter. 0.1 dropout ratio was used.

zebrafish: 20000-30000 cells. Time series of light sheet microscopy volumes depicting the development of zebrafish embryo and corresponding cell lineages were taken from Zebrahub Lange et al. (2024). Cell tracking was performed using an automated method Bragantini et al. (2025) without further manual correction with 50% of tracks getting an error within 60 frames. Track ZSNS001 was used to derive two registration datasets: **zebrafish 20k** contains frames 0-50 from the beginning of the track and **zebrafish 30k** contains frames 500-550 from the end of the track. 5 augmented samples per frame were generated with 5% of nodes moved by at maximum 2% of the graph diameter. 0.1 dropout ratio was used.

platy: 11000 cells. Only confocal volumes from **Platy5+1** were used. 20 augmented samples per frame were generated with 10% of nodes moved by at maximum 4% of the graph diameter. 0.15 dropout ratio was used.

For **hela** dataset we compared 3 architectures: *grreg* GCN, *grreg all-to-all* and *grreg sparse*. Due to GPU memory limitations only *grreg sparse* was trained for **zebrafish** and **platy**. All *grreg all-to-all* and *sparse* models were based on the PPF-GCN with 2 convolutions and had 2 transformer layers. All kNN graphs were built with $k = 6$. Gradient clipping and cosine learning rate scheduler, commonly used for training large transformer models Rae et al. (2021); Hoffmann et al. (2022), were used to improve the convergence. Further details on experimental setup can be found in section A.2. Coherent Point Drift was used as a baseline (Section A.3).

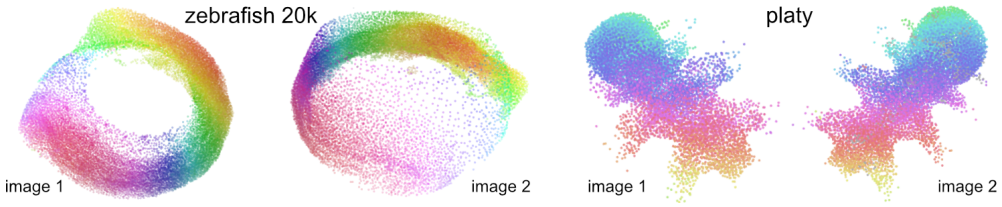


Figure 5: Matching examples for zebrafish and platy

7 RESULTS

7.1 *grreg* ENABLES ACCURATE GLOBAL REGISTRATION

Grreg enables accurate cell matching for **hela** dataset. It reaches near-perfect result for **easy** version of registration problem where similarity between registered images is high and over 0.9 accuracy on most **hard** datasets (Table 1). Low mean and maximum mismatch distance suggests that *grreg* was able to successfully use global context to resolve ambiguity between similar cells in different parts of the sample and identify the correct neighborhood.

Sparse models achieved similarly high matching accuracy, although only a portion of nodes was used for self- and cross-attention. Furthermore, we tested different number of anchor nodes and found that it is possible to use as little as 5% of nodes without loss of performance (Figure 8). At 5% of nodes for some images it meant that as little as 4 anchor nodes were used.

CPD paired with hungarian matching - a commonly used method for deformable registration - reaches higher accuracy when the samples are perfectly pre-aligned (Table 1) but if the initial orientation is outside of the tolerance zone, registration process gets stuck in a local minimum and the accuracy falls to 0. Using rotation- and translation-invariant geometric features ensures that the registration result does not depend on the initial alignment (Figure 4).

Let us inspect the embedding space produced by *grreg*. The leftmost panel of Figure 3 shows the overlay of two point clouds from **hela 01 asymmetric easy** in physical space. Some cells drifted a little from their initial positions, and one cell (outlined in red) in the lower right corner does not have a clear match likely due to a segmentation error, but overall the point clouds are very similar. *Grreg all-to-all* (third panel) places matching points in the exact same locations in the embedding space, successfully reproducing the structure of the graph in the rotation- and translation-invariant feature space. As expected, highlighted cell does not have an exact correspondence but is placed near to its potential matching point.

7.2 *grreg* SCALES TO SAMPLES WITH THOUSANDS OF CELLS

Grreg sparse was used to register volumetric images of zebrafish embryos extracted from an early development time series. Figure 5 shows the matching: points in the sample on the left are colored according to their XYZ coordinates, points in the sample on the right are colored same as their matches. It was possible to achieve 12 to 20% exact accuracy for different datasets and 87% 3-hop accuracy for **zebrafish 20k**. Due to errors in the object detection and tracking, as well as movement of the cells, it was expected that it might not be possible to find the exact match, but it should be possible to place the cell in the close neighborhood of the correct match (Table 2). Low Δ_{mean} confirms that overall the matching is consistent and allows to register two point clouds no matter the initial orientation (Figure 4). Sparse transformer architecture allowed to handle volumes with up to 32 000 cells. Network inference and matching take $\sim$ 18 seconds for the samples from **zebrafish 30k**.

In **platy** dataset each sample is a different individual animal and the matching is enabled by the stereotypical development. Out of 27 annotated cells, on average 47% 3-hop accuracy was reached with *grreg sparse* (Table 2). Low Δ_{mean} and Δ_{max} of 0.15 of the graph diameter and muscle gene expression correlation of 0.59 (compared to 0.67 with CPD) show that despite individual differences between the organisms and varying poses, *grreg sparse* accomplishes a reasonable registration consistent throughout the volume (Figure 5).

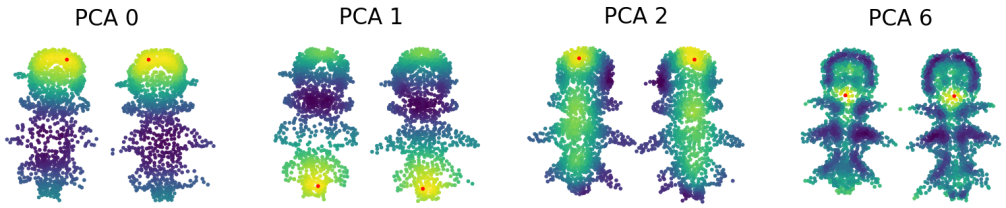


Figure 6: *grreg* embeddings for **platy**

Similarly to **hela**, CPD achieves higher accuracy when applied to perfectly pre-aligned images (Table 2) but fails without pre-alignment. CPD enforces that cells located next to each other should map to the objects located near to each other. This coincides with the bias introduced by deriving ground truth matches from tracking data. For **platy**, a more challenging dataset, the gap between *grreg* and pre-alignment + CPD is smaller.

7.3 Interpretation of *grreg* embeddings

Embeddings for two matched samples were combined and PCA was applied to the joint point cloud. Points in the sample were then colored according to the PCA component values to visualize different features. PCA components with higher variance encode the distance from a cell to a certain prominent anatomical feature. For example, for *Platynereis* samples (Figure 6) PCA1 encodes the proximity to the head, PCA2 - proximity to the tail and PCA3 determines the side of the body. For zebrafish (Figure 7 the first few PCA components determine the distance from some anchor locations on the circle formed by the body of the embryo. PCA3 determines the side of the body (not shown). Components with lower variance are less correlated with global coordinates and help to distinguish cells locally. We observed that *grreg* embeddings are easily interpretable and determine a position of each cell as a combination of learnt basis functions specific for a given set of samples.

7.4 Ablation

PPF-GCNs with different number of convolutions were trained separately, without the transformer part, and have shown significantly worse performance than transformers, especially for more challenging datasets Table 1. High Δ_{max} suggests that, despite having most of the graph within the receptive field, GCN struggles to create a global coordinate system and instead relies on local features for matching. This is illustrated in the second panel of Figure 3: small differences in the cell coordinates significantly change the embeddings which makes the final discrete assignment problem much less straightforward.

Increasing the number of convolutions allows GCN to "see" wider neighborhood. For **hela** dataset increasing number of convolutions from 2 to 6 improves the accuracy but further increase does not help and can even worsen the performance (Figure 9), consistent with oversmoothing (Arnaiz-Rodriguez & Errica, 2025) observed in various GNNs. PPFs capture local geometric structure but the information about the relative position of long-range nodes is lost without the distance-aware self-attention. This observation highlights the global nature of the cell matching problem and confirms that the transformer architecture is essential for correct matching.

8 Conclusions

This paper demonstrates the power of attention-based rotation-invariant GNNs for global cell matching in biomedical images. *Grreg*'s architecture consists of rotation- and translation-invariant GCN for local feature extraction and a graph transformer with two kinds of attention: distance-aware self-attention, which improves the receptive field of local descriptors, and cross-attention, which enables cross-image communication. Our method handles inexact and incomplete matching by using a contrastive objective. Our experiments show that *grreg* achieves high registration accuracy, does not depend on good pre-alignment and can work for very large volumes with up to 32 000 cells. We believe that *grreg* is an important step towards automated integration and analysis of large-scale cross-modal imaging datasets.

AI USE STATEMENT

In this work, we used generative AI tools to aid or polish writing and create and edit code for result visualization. We have not used generative AI tools for any tasks with required disclosure. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by 2 authors, text was proof-read. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Adrian Arnaiz-Rodriguez and Federico Errica. Oversmoothing, oversquashing, heterophily, long-range, and more: Demystifying common beliefs in graph machine learning. *arXiv [cs.LG]*, May 2025.

Albina Asadulina, Aurora Panzera, Csaba Verasztó, Christian Liebig, and Gáspár Jékely. Whole-body gene expression pattern registration in platynereis larvae. *Evodevo*, 3(1):27, December 2012.

P J Besl and Neil D McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, February 1992a.

Paul J. Besl and Neil D. McKay. lt;titlegt;method for registration of 3-d shapeslt;/titlegt;. In Paul S. Schenker (ed.), *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pp. 586–606. SPIE, April 1992b. doi: 10.1117/12.57955. URL http://dx.doi.org/10.1117/12.57955.

Johanna Bischof, Georgina Fletcher, Paul Verkade, Claudia Kuntner, Julia Fernandez-Rodriguez, Linda Chaabane, Leor Ariel Rose, Andreas Walter, Michiel Vandenbosch, Marc A M J van Zandvoort, Assaf Zaritsky, Antje Keppler, and Maddy Parsons. Multimodal bioimaging across disciplines and scales: challenges, opportunities and breaking down barriers. *Npj Imaging*, 2(1):5, March 2024.

John A Bogovic, Philipp Hanslovsky, Allan Wong, and Stephan Saalfeld. Robust registration of calcium images by learned contrast synthesis. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1123–1126, April 2016.

Jordão Bragantini, Ilan Theodoro, Xiang Zhao, Teun A P M Huijben, Eduardo Hirata-Miyasaki, Shruthi VijayKumar, Akilandeswari Balasubramanian, Tiger Lao, Richa Agrawal, Sheng Xiao, Jan Lammerding, Shalin Mehta, Alexandre X Falcão, Adrian Jacobo, Merlin Lange, and Loïc A Royer. Ultrack: pushing the limits of cell tracking across biological scales. *Nat. Methods*, 22(11): 2423–2436, November 2025.

Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image Vis. Comput.*, 10(3):145–155, April 1992.

Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-Chieh Yu, Claire E McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, Will Silversmith, Casey Schneider-Mizell, Chris S Jordan, Derrick Brittain, Akhilesh Halageri, Kai Kuehner, Oluwaseun Ogedengbe, Ryan Morey, Jay Gager, Krzysztof Kruk, Eric Perlman, Runzhe Yang, David Deutsch, Doug Bland, Marissa Sorek, Ran Lu, Thomas Macrina, Kisuk Lee, J Alexander Bae, Shang Mu, Barak Nehoran, Eric Mitchell, Sergiy Popovych, Jingpeng Wu, Zhen Jia, Manuel A Castro, Nico Kemnitz, Dodam Ih, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Forrest Collman, Davi D Bock, Gregory S X E Jefferis, H Sebastian Seung, Mala Murthy, and FlyWire Consortium. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, October 2024.

Antje Hl Fischer, Thorsten Henrich, and Detlev Arendt. The normal development of platynereis dumerilii (nereididae, annelida). *Front. Zool.*, 7:31, December 2010.

Stefan Haller, Lorenz Feineis, Lisa Hutschenreiter, Florian Bernard, Carsten Rother, Dagmar Kainmüller, Paul Swoboda, and Bogdan Savchynskyy. A comparative study of graph matching algorithms in computer vision. *arXiv [cs.CV]*, July 2022.

Iván Hidalgo-Cenalmor, Adán Olguín-Olguín, Carolina Prieto, Johannes Kumra Ahnlide, Pontus Nordenfelt, Ricardo Henriques, Mario Del Rosario, and Guillaume Jacquemet. NucleiSky enables cross-scale multimodal registration of microscopy data using nuclei constellations. *J. Cell Sci.*, (jcs.265152), August 2026.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W Rae, Oriol Vinyals, and Laurent Sifre. Training compute-optimal large language models. *arXiv [cs.CL]*, March 2022.

Dagmar Kainmueller, Florian Jug, Carsten Rother, and Gene Myers. Active graph matching for automatic joint segmentation and annotation of C. elegans. *Med. Image Comput. Comput. Assist. Interv.*, 17(Pt 1):81–88, 2014.

Kenta-Tanaka et al. probreg. URL `https://probreg.readthedocs.io/en/latest/`.

Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien P W Pluim. Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging*, 29(1): 196–205, January 2010.

Daniel Krentzel, Matouš Elphick, Marie-Charlotte Domart, Christopher J. Peddie, Romain F. Laine, Cameron Shand, Ricardo Henriques, Lucy M. Collinson, and Martin L. Jones. Clem-reg: an automated point cloud-based registration algorithm for volume correlative light and electron microscopy. *Nature Methods*, 22(9):1923–1934, September 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02794-0. URL `http://dx.doi.org/10.1038/s41592-025-02794-0`.

Merlin Lange, Alejandro Granados, Shruthi VijayKumar, Jordão Bragantini, Sarah Ancheta, Yang-Joon Kim, Sreejith Santhosh, Michael Borja, Hirofumi Kobayashi, Erin McGeever, Ahmet Can Solak, Bin Yang, Xiang Zhao, Yang Liu, Angela M Detweiler, Sheryl Paul, Ilan Theodoro, Honey Mekonen, Chris Charlton, Tiger Lao, Rachel Banks, Sheng Xiao, Adrian Jacobo, Keir Balla, Kyle Awayan, Samuel D'Souza, Robert Haase, Alexandre Dizeux, Olivier Pourquie, Rafael Gómez-Sjöberg, Greg Huber, Mattia Serra, Norma Neff, Angela Oliveira Pisco, and Loïc A Royer. A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. *Cell*, 187(23):6742–6759.e17, November 2024.

David G Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60 (2):91–110, November 2004.

Martin Maška, Vladimír Ulman, Pablo Delgado-Rodriguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A Guerrero Peña, Tsang Ing Ren, Elliot M Meyerowitz, Tim Scherr, Katharina Löffler, Ralf Mikut, Tianqi Guo, Yin Wang, Jan P Allebach, Rina Bao, Noor M Al-Shakarji, Gani Rahmon, Imad Eddine Toubal, Kannappan Palaniappan, Filip Lux, Petr Matula, Ko Sugawara, Klas E G Magnusson, Layton Aho, Andrew R Cohen, Assaf Arbelle, Tal Ben-Haim, Tammy Riklin Raviv, Fabian Isensee, Paul F Jäger, Klaus H Maier-Hein, Yanming Zhu, Cristina Ederra, Ainhoa Urbiola, Erik Meijering, Alexandre Cunha, Arrate Muñoz-Barrutia, Michal Kozubek, and Carlos Ortiz-de Solórzano. The cell tracking challenge: 10 years of objective benchmarking. *Nat. Methods*, 20(7):1010–1020, July 2023.

Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, December 2010a. ISSN 0162-8828. doi: 10.1109/tpami.2010.46. URL `http://dx.doi.org/10.1109/TPAMI.2010.46`.

Andriy Myronenko and Xubo Song. Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2262–2275, December 2010b.

Beate Neumann, Thomas Walter, Jean-Karim Hériché, Jutta Bulkescher, Holger Erfle, Christian Conrad, Phill Rogers, Ina Poser, Michael Held, Urban Liebel, Cihan Cetin, Frank Sieckmann, Gregoire Pau, Rolf Kabbe, Annelie Wünsche, Venkata Satagopam, Michael H A Schmitz, Catherine Chapuis, Daniel W Gerlich, Reinhard Schneider, Roland Eils, Wolfgang Huber, Jan-Michael Peters, Anthony A Hyman, Richard Durbin, Rainer Pepperkok, and Jan Ellenberg. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464 (7289):721–727, April 2010.

Oel Adam Phillip, Cros Cyril Christophe Daniel Robert, Pan Leslie, Buglakova Elena, Haury Parra Samuel, Santangeli Luca, Gerber Tobias, Papalopoulos Nikolaos, Vergara Hernando Martinez, Puga David, Disela Vanessa, Niederhaus Lara, Bertucci Paola Yanina, Witte Victoria, Bilgesu Kilic Asli, Quintana Urzainqui Idoia, Acevedo Julieta Maria, Musser Jacob, Kreshuk Anna, and Arendt Detlev. A whole-body cell type atlas mapped into an electron microscopy volume of an annelid worm. *bioRxiv*, 2025.

C Qi, Hao Su, Kaichun Mo, and L Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 77–85, December 2016.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.*, 30, 2017.

Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, Eliza Rutherford, Tom Hennigan, Jacob Menick, Albin Cassirer, Richard Powell, George van den Driessche, Lisa Anne Hendricks, Maribeth Rauh, Po-Sen Huang, Amelia Glaese, Johannes Welbl, Sumanth Dathathri, Saffron Huang, Jonathan Uesato, John Mellor, Irina Higgins, Antonia Creswell, Nat McAleese, Amy Wu, Erich Elsen, Siddhant Jayakumar, Elena Buchatskaya, David Budden, Esme Sutherland, Karen Simonyan, Michela Paganini, Laurent Sifre, Lena Martens, Xiang Lorraine Li, Adhiguna Kuncoro, Aida Nematzadeh, Elena Gribovskaya, Domenic Donato, Angeliki Lazaridou, Arthur Mensch, Jean-Baptiste Lespiau, Maria Tsimpoukelli, Nikolai Grigorev, Doug Fritz, Thibault Sottiaux, Mantas Pajarskas, Toby Pohlen, Zhitao Gong, Daniel Toyama, Cyprien de Masson d'Autume, Yujia Li, Tayfun Terzi, Vladimir Mikulik, Igor Babuschkin, Aidan Clark, Diego de Las Casas, Aurelia Guy, Chris Jones, James Bradbury, Matthew Johnson, Blake Hechtman, Laura Weidinger, Iason Gabriel, William Isaac, Ed Lockhart, Simon Osindero, Laura Rimell, Chris Dyer, Oriol Vinyals, Kareem Ayoub, Jeff Stanway, Lorrayne Bennett, Demis Hassabis, Koray Kavukcuoglu, and Geoffrey Irving. Scaling language models: Methods, analysis & insights from training gopher. *arXiv [cs.CL]*, December 2021.

Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4938–4947, 2020.

Philipp Schlegel, Yijie Yin, Alexander S Bates, Sven Dorkenwald, Katharina Eichler, Paul Brooks, Daniel S Han, Marina Gkantia, Marcia Dos Santos, Eva J Munnelly, Griffin Badalamente, Laia Serratosa Capdevila, Varun A Sane, Alexandra M C Fragniere, Ladann Kiassat, Markus W Pleijzier, Tomke Stürner, Imaan F M Tamimi, Christopher R Dunne, Irene Salgarella, Alexandre Javier, Siqi Fang, Eric Perlman, Tom Kazimiers, Sridhar R Jagannathan, Arie Matsliah, Amy R Sterling, Szi-Chieh Yu, Claire E McKellar, FlyWire Consortium, Marta Costa, H Sebastian Seung, Mala Murthy, Volker Hartenstein, Davi D Bock, and Gregory S X E Jefferis. Whole-brain annotation and multi-connectome cell typing of drosophila. *Nature*, 634(8032):139–152, October 2024.

Denis P Shamonin, Esther E Bron, Boudewijn P F Lelieveldt, Marion Smits, Stefan Klein, Marius Staring, and Alzheimer's Disease Neuroimaging Initiative. Fast parallel image registration on CPU and GPU for diagnostic classification of alzheimer's disease. *Front. Neuroinform.*, 7:50, 2013.

Sebastian Stricker, Christoph Karg, Lisa Hutschenreiter, Bogdan Savchynskyy, and Dagmar Kainmueller. Cycle-consistent multi-graph matching for self-supervised annotation of C. elegans. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 7208–7217. IEEE, March 2026.

Raju Tomer, Alexandru S Denes, Kristin Tessmar-Raible, and Detlev Arendt. Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell*, 142 (5):800–809, September 2010.

Lorenzo Torresani, Vladimir Kolmogorov, and Carsten Rother. Feature correspondence via graph matching: Models and global optimization. In *Computer Vision – ECCV 2008*, volume 5303 of *Lecture Notes in Computer Science*, pp. 596–609. Springer, 2008. doi: 10.1007/978-3-540-88688-4_44.

Hernando M Vergara, Constantin Pape, Kimberly I Meechan, Valentyna Zinchenko, Christel Genoud, Adrian A Wanner, Kevin Nzumbi Mutemi, Benjamin Titze, Rachel M Templin, Paola Y Bertucci, Oleg Simakov, Wiebke Dürichen, Pedro Machado, Emily L Savage, Lothar Schermelleh, Yannick Schwab, Rainer W Friedrich, Anna Kreshuk, Christian Tischer, and Detlev Arendt. Whole-body integration of gene expression and single-cell morphology. *Cell*, 184(18): 4819–4837.e22, September 2021.

Hernando Martínez Vergara, Paola Yanina Bertucci, Peter Hantz, Maria Antonietta Tosches, Kaia Achim, Pavel Vopalensky, and Detlev Arendt. Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of platynereis dumerilii. *Proc. Natl. Acad. Sci. U. S. A.*, 114(23):5878–5885, June 2017.

Mingyue Wang, Qinan Hu, Zhencheng Tu, Lingshi Kong, Tengxiang Yu, Zihan Jia, Yuetian Wang, Jiajun Yao, Rong Xiang, Zhan Chen, Yan Zhao, Yanfei Zhou, Qing Ye, Kang Ouyang, Xianzhe Wang, Yinqi Bai, Zhenyu Yang, Hanxiang Wang, Yanru Wang, Hanxiang Jiang, Tao Yang, Jing Chen, Yunting Huang, Ni Yin, Wenyuan Mo, Wenfu Liang, Chang Liu, Xiumei Lin, Chuanyu Liu, Ying Gu, Wei Chen, Longqi Liu, Xun Xu, and Yuhui Hu. A drosophila single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation. *Cell*, 188(17):4734–4753.e31, August 2025.

Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. *arXiv [cs.CV]*, pp. 3523–3532, May 2019a.

Yue Wang and Justin M. Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019b.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.*, 38(5):1–12, October 2019.

Zizhen Yao, Cindy T J van Velthoven, Michael Kunst, Meng Zhang, Delissa McMillen, Changkyu Lee, Won Jung, Jeff Goldy, Aliya Abdelhak, Matthew Aitken, Katherine Baker, Pamela Baker, Eliza Barkan, Darren Bertagnolli, Ashwin Bhandiwad, Cameron Bielstein, Prajal Bishwakarma, Jazmin Campos, Daniel Carey, Tamara Casper, Anish Bhaswanth Chakka, Rushil Chakrabarty, Sakshi Chavan, Min Chen, Michael Clark, Jennie Close, Kirsten Crichton, Scott Daniel, Peter DiValentin, Tim Dolbeare, Lauren Ellingwood, Elysha Fiabane, Timothy Fliss, James Gee, James Gerstenberger, Alexandra Glandon, Jessica Gloe, Joshua Gould, James Gray, Nathan Guilford, Junitta Guzman, Daniel Hirschstein, Windy Ho, Marcus Hooper, Mike Huang, Madie Hupp, Kelly Jin, Matthew Kroll, Kanan Lathia, Arielle Leon, Su Li, Brian Long, Zach Madigan, Jessica Malloy, Jocelin Malone, Zoe Maltzer, Naomi Martin, Rachel McCue, Ryan McGinty, Nicholas Mei, Jose Melchor, Emma Meyerdierks, Tyler Mollenkopf, Skyler Moonsman, Thuc Nghi Nguyen, Sven Otto, Trangthanh Pham, Christine Rimorin, Augustin Ruiz, Raymond Sanchez, Lane Sawyer, Nadiya Shapovalova, Noah Shepard, Cliff Slaughterbeck, Josef Sulc, Michael Tieu, Amy Torkelson, Herman Tung, Nasmil Valera Cuevas, Shane Vance, Katherine Wadhwani, Katelyn Ward, Boaz Levi, Colin Farrell, Rob Young, Brian Staats, Ming-Qiang Michael Wang, Carol L Thompson, Shoaib Mufti, Chelsea M Pagan, Lauren Kruse, Nick Dee, Susan M Sunkin, Luke Esposito, Michael J Hawrylycz, Jack Waters, Lydia Ng, Kimberly Smith, Bosiljka Tasic, Xiaowei Zhuang, and Hongkui Zeng. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, 624(7991):317–332, December 2023.

Eviatar Yemini, Albert Lin, Amin Nejatbakhsh, Erdem Varol, Ruoxi Sun, Gonzalo E Mena, Aravinthan D T Samuel, Liam Paninski, Vivek Venkatachalam, and Oliver Hobert. NeuroPAL: A

multicolor atlas for whole-brain neuronal identification in C. elegans. *Cell*, 184(1):272–288.e11, January 2021.

Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, and Amr Ahmed. Big bird: Transformers for longer sequences. *arXiv [cs.LG]*, July 2020.

Yu-Xin Zhang, Jie Gui, Baosheng Yu, Xiaofeng Cong, Xin Gong, Wenbing Tao, and Dacheng Tao. Deep learning-based point cloud registration: A comprehensive survey and taxonomy. *Int. J. Comput. Vis.*, 134(4), April 2026.

B Duygu Özpolat, Nadine Randel, Elizabeth A Williams, Luis Alberto Bezares-Calderón, Gabriele Andreatta, Guillaume Balavoine, Paola Y Bertucci, David E K Ferrier, Maria Cristina Gambi, Eve Gazave, Mette Handberg-Thorsager, Jörg Hardege, Cameron Hird, Yu-Wen Hsieh, Jerome Hui, Kevin Nzumbi Mutemi, Stephan Q Schneider, Oleg Simakov, Hernando M Vergara, Michel Vervoort, Gáspár Jékely, Kristin Tessmar-Raible, Florian Raible, and Detlev Arendt. The nereid on the rise: Platynereis as a model system. *Evodevo*, 12(1):10, September 2021.

A Appendix

Algorithm 1 *grreg* all-to-all

$\mathcal{X}_0^1 \leftarrow$ PPF-GCN$(\mathcal{G}_1)$
$\mathcal{X}_0^2 \leftarrow$ PPF-GCN$(\mathcal{G}_2)$
for $i \leftarrow 1, R$ **do**
  $\mathcal{X}_i^{1\prime} \leftarrow$ D-Attn Block$(\mathcal{X}_{i-1}^1)$
  $\mathcal{X}_i^{2\prime} \leftarrow$ D-Attn Block$(\mathcal{X}_{i-1}^2)$
  $\mathcal{X}_i^1 \leftarrow$ X-Attn Block$(\mathcal{X}_i^{1\prime}, \mathcal{X}_i^{2\prime})$
  $\mathcal{X}_i^2 \leftarrow$ X-Attn Block$(\mathcal{X}_i^{2\prime}, \mathcal{X}_i^{1\prime})$
end for

A.1 Platy5+1

Aiming to build a dataset representing a real-world large-scale registration problem, we turn to the data acquired in the efforts to build a comprehensive atlas of the *Platynereis dumerilii* worm. *P. dumerilii* is a sea annelid which serves as a convenient model organism for evolutionary and developmental biology Özpolat et al. (2021). Larval development of *P. dumerilii* is synchronized between individual animals until the late nectochaete stage (5-7 days post fertilization) Fischer et al. (2010). At this stage the larva already has about 11000 cells as well as many differentiated tissues and cell types while remaining relatively small ($\sim$ 300 μm by 100 μm) and amenable to various imaging techniques. Stereotypical developmental enables researches to image different individual animals with different techniques and correlate the results Tomer et al. (2010); Asadulina et al. (2012); Vergara et al. (2017; 2021).

Our benchmark is based on the data made available with Phillip et al. (2025). Each confocal volume consists of two channels: DAPI staining highlighting nuclei and a reporter for gene expression for a selected gene. We chose the volumes stained for StMhc gene which is a marker of somatic muscles. StMhc is highly expressed throughout the body of the animal, making it a perfect marker for registration quality assessment without manual landmark detection. In Phillip et al. (2025), confocal volumes, originally acquired at $0.5\ \mu m \times 0.13\ \mu m \times 0.13\ \mu m$ resolution, were preprocessed to enhance the signal and nuclei were segmented based on DAPI channel. We extend the segmentation to cells, taking the part of the volume closest to each nucleus mask as the cell mask. Average intensity of StMhc staining signal within the cell area is then assigned to each nucleus.

In Vergara et al. (2021), the *P. dumerilii* atlas was enriched with an electron microscopy volume, originally acquired at $25\ nm \times 10\ nm \times 10\ nm$ resolution, with a corresponding nuclear and cellular segmentation. In addition, muscle cells were segmented based on cellular morphology.

Anisotropy, large difference in the appearance of the modalities and differences between individual organisms make this benchmark a realistic representation of the challenges of biological image registration.

The matching ground truth consists of cell pairs, annotated by *P. dumerilii* experts across the 5 confocal and 1 EM volumes. In addition, we use a dataset-level registration quality metric measured as the correlation of StMhc gene signal between the registered volumes. The cells for ground truth were chosen to be distributed throughout the body of the animal and represent different types of local neighborhoods. If the cell had a bilaterally symmetric pair, both cells were annotated. To account for bilateral symmetry of the organism, we provide two versions of ground truth annotations: direct and mirrored.

Based on the ground truth matches, we recommend using the following metrics to evaluate the matching quality: accuracy, 1-, 2- and 3-hop accuracy, Δ_{mean} and Δ_{max}. These metrics allow to assess the single-cell precision of the registration but can be stochastic due to small number of the annotated correspondences. In contrast, muscle signal correlation can be used to assess the average quality of registration.

Let S_1 and S_2 be muscle signals in nodes of $\mathcal{G}_1$ and $\mathcal{G}_2$. Registration produces matching between nodes $M = \{(i, j), i \in \mathcal{G}_1, j \in \mathcal{G}_2\}$. To each node $j \in \mathcal{G}_2$ we assign S_{1i} - muscle signal of the corresponding node in $\mathcal{G}_1$. If there are multiple corresponding nodes, we assign an average value. It results in feature vector S'_2.

$$muscle\ corr = corr(S_2, S'_2)$$

This metric works best to assess the registration quality between confocal volumes, although it is possible to apply it to registration with EM volume, using binary muscle segmentation as muscle signal.

A.2 Experimental setup

For **hela** dataset we compared 3 architectures: *grreg* GCN, *grreg all-to-all* and *grreg sparse*. Due to memory limitations only *grreg sparse* was trained for **zebrafish** and **platy**. All *grreg all-to-all* and *sparse* models were based on the rotation-invariant GCN with 2 convolutions and had 2 transformer layers. All kNN graphs were built with $k = 6$. Gradient clipping and cosine learning rate scheduler, commonly used for training large transformer models Rae et al. (2021); Hoffmann et al. (2022), were used to improve the convergence.

Networks for **hela** were trained on a single NVIDIA GeForce RTX 3090 for 500 epochs (up to 5 hours for GCN and *all-to-all* and up to 11 hours for *sparse*). The final embedding size was set to 128. We performed learning rate warmup over 46000 iterations to the maximum of 0.0001, with learning rate decreasing until the total number of iterations. 2 attention layers with 16 heads were used in transformer models. For *grreg sparse* number of points in each cluster was set to 15, 30 and 60 points for anchor node ratios 0.2, 0.1 and 0.05 correspondingly.

Models for **zebrafish** were trained for 200 epochs using a single Nvidia H100 GPU, 13 min per epoch for **zebrafish 20k** and 50 min per epoch for **zebrafish 30k**. The final embedding size was set to 128. We performed learning rate warmup over 37500 iterations to the maximum of 0.00001, with learning rate decreasing until the total number of iterations. 2 attention layers with 16 heads were used in *grreg sparse*. Number of points per cluster was set to 60 for the anchor node ratio 0.05.

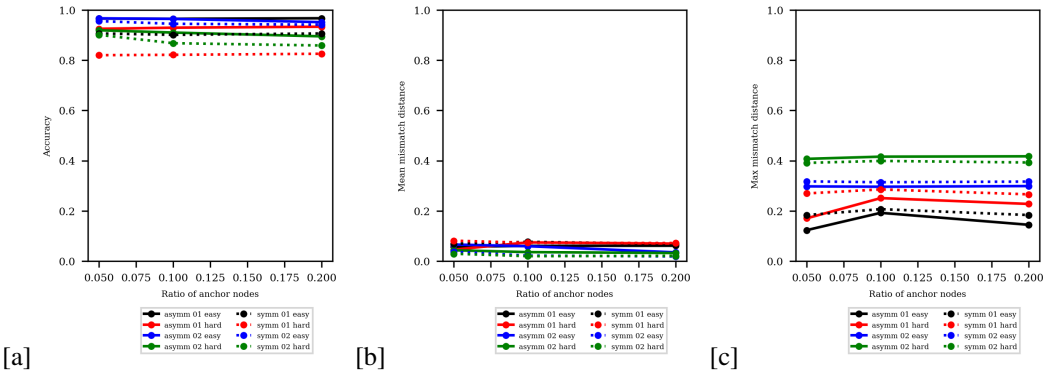
Model for **platy** was trained for 800 epochs with 8 min per epoch using a single Nvidia H100 GPU. The final embedding size was set to 4. We performed learning rate warmup over 37500 iterations to the maximum of 0.0001, with learning rate decreasing until the total number of iterations. 2 attention layers with 4 heads were used. Number of points per cluster was set to 60 for the anchor node ratio 0.2.

A.3 Coherent point drift baseline

Coherent point drift (CPD) Myronenko & Song (2010a) was used as a baseline as it is a method commonly used for deformable registration. Implementation from probreg library Kenta-Tanaka et al. was used for all experiments. For all point cloud pairs moving point cloud was rotated around



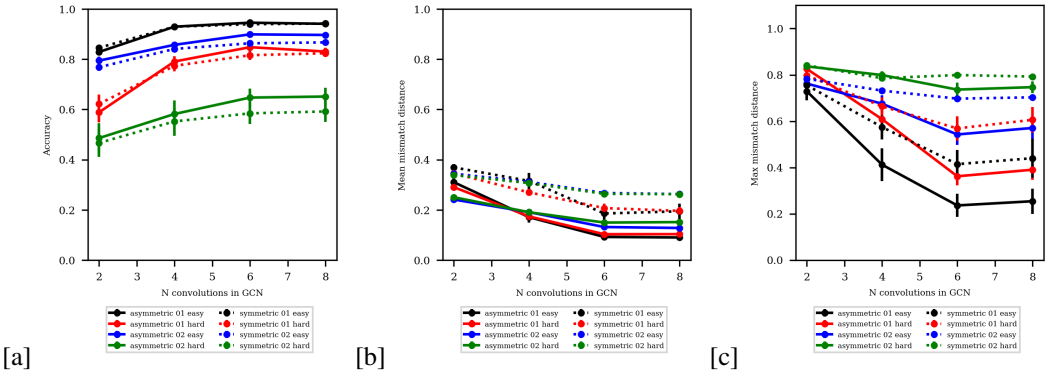
Figure 7: *grreg* embeddings for **zebrafish 20k**



[a] [b] [c]

Figure 8: Influence of number of anchor nodes on the registration accuracy.

center of mass to demonstrate the dependence of registration result on the initial position. At first we applied rigid CPD and then deformable, with the parameters established by a grid search. For **hela** we set $w = 0$, $\lambda = 1$ and $\beta = 2$, $max\ iter = 50$. For **zebrafish** $w = 0$, $\lambda = 1$ and $\beta = 2$, $max\ iter = 100$. For platy $w = 0.00001$, $\lambda = 0.1$ and $\beta = 100$, $max\ iter = 100$.As the matching accuracy falls quickly with the rotation, we report the accuracy at rotation angle $\alpha = 0°$ - pre-aligned samples - and $\alpha = 180°$ - completely misaligned samples.



[a] [b] [c]

Figure 9: Influence of number of convolutions on the registration accuracy.