DENSITY TELLS WHERE MATCHING IS AMBIGUOUS: A LINEAR-TIME MAMBA-TRANSFORMER FOR UNORDERED 3D POINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Point cloud registration is a fundamental prerequisite in 3D vision. While existing Transformer-based methods excel at feature matching, their quadratic computational complexity limits scalability. Conversely, linear-time Mamba models inherently struggle to capture non-sequential 3D spatial symmetries and global context. Furthermore, current approaches largely overlook local point density, an intrinsic geometric prior crucial for resolving matching ambiguities in sparse or repetitive regions. To bridge these gaps, we propose the Density-Aware Mamba-Transformer (DAMT). DAMT couples a lightweight attention mechanism, which provides permutation-invariant global initialization via a novel Global Norm Pooling (GNP) module, with a bidirectional Mamba block for linear-scaling deep feature aggregation. To fully exploit density cues without overfitting to sensor noise, we explicitly introduce a Density-Guided Attention Weighting (DGAW) mechanism and dedicated loss constraints, enforcing strict neighborhood density consistency. Extensive experiments across synthetic CAD (ModelNet/ModelLoNet), indoor RGB-D (3DMatch/3DLoMatch), and outdoor LiDAR (KITTI) benchmarks demonstrate that DAMT achieves a highly competitive performance, offering a superior trade-off between registration accuracy and computational efficiency, paving the way for scalable and robust 3D registration.

1 INTRODUCTION

Point cloud registration is a pivotal prerequisite in 3D vision, aiming to align multi-view or multi-sensor spatial data into a unified global frame Guo et al. (2021). While traditional mathematical optimization methods (e.g., ICP, RANSAC) are notoriously sensitive to noise and initialization, deep learning paradigms have significantly enhanced robustness. Specifically, superpoint-based approaches Qin et al. (2022); Yu et al. (2021); Hui et al. (2021) mitigate computational overhead by clustering raw points into local patches for coarse-to-fine matching. However, their efficacy fundamentally relies on the ability to efficiently extract global context and discriminative local features.

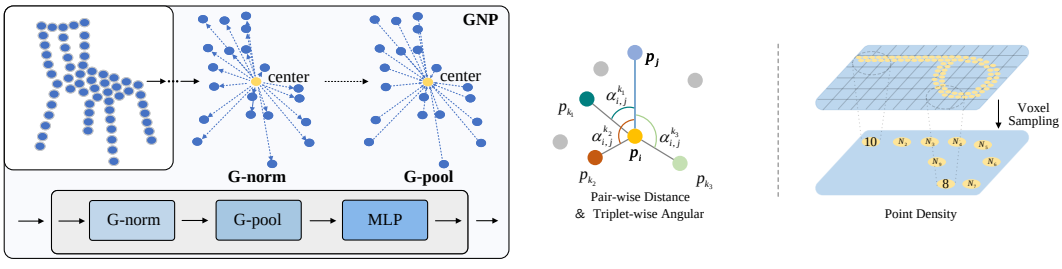


Figure 1: Core components of DAMT. **Left (GNP Module):** A lightweight mechanism that efficiently captures and propagates the global receptive field to individual superpoints, overcoming the computational bottleneck of deep attention layers. **Right (Density Integration):** Explicit modeling of local point cloud density as a structural prior, which seamlessly complements geometric features to resolve matching ambiguities in sparse and challenging scenarios.

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.

To capture such global dependencies, Transformers Vaswani et al. (2017) have been widely adopted Chen et al. (2024a;b). Yet, the self-attention mechanism incurs quadratic computational complexity, severely limiting its scalability for large-scale point clouds. Conversely, while Mamba Gu & Dao (2024) achieves linear complexity and excels in long-sequence modeling, pure State-Space Models (SSMs) inherently struggle to capture non-sequential 3D spatial symmetries. To break this dilemma, we introduce a deliberate division of labor: utilizing a shallow Self-Attention layer exclusively to establish an initial, permutation-invariant global receptive field, followed by a linear-complexity Mamba module for deep feature aggregation.

Beyond global context, local point density is an intrinsic geometric attribute reflecting variations in curvature, sensor distance, and structural articulations He et al. (2022). Despite encoding crucial structural cues that can resolve matching ambiguities in sparse regions, local density patterns are largely overlooked in existing registration pipelines Zhao et al. (2022); Min et al. (2021), leading to sub-optimal performance in challenging environments.

To tackle these challenges, we propose the Density-Aware Mamba-Transformer (DAMT). DAMT circumvents the computation bottleneck of deep attention layers through a hybrid architecture. Specifically, we design a Global Norm Pooling (GNP) module to efficiently capture and propagate long-range global representations. Furthermore, to fully exploit density cues without overfitting to sensor noise, we explicitly introduce a Density-Guided Attention Weighting (DGAW) mechanism. This operationalizes point cloud density as both a structural embedding and a loss constraint, effectively complementing traditional geometric cues for highly discriminative feature learning.

In summary, our key contributions are threefold:

- We propose DAMT, a pioneering Transformer-Mamba hybrid architecture tailored for point cloud registration, offering a scalable and superior trade-off between computational efficiency and matching accuracy.
- We design the GNP module, a lightweight mechanism that efficiently extracts permutation-invariant global context, explicitly overcoming the spatial modeling limitations of pure SSMs.
- We explicitly exploit point cloud density—a crucial yet often overlooked geometric prior—via the novel DGAW mechanism and density-consistency constraints, significantly boosting robustness across diverse sensor distributions.

2 Related work

Keypoint and Superpoint match methods Keypoint-based approaches Yew & Lee (2018); Bai et al. (2020); Lu et al. (2020) detect salient points to establish correspondences, balancing computational efficiency Yew & Lee (2018); Lu et al. (2020) and advanced feature learning Bai et al. (2020); Lu et al. (2019). To further mitigate computational burdens, superpoint-based frameworks Shi et al. (2023); Qin et al. (2022); Huang et al. (2021) group raw points into local patches, facilitating efficient coarse-to-fine matching pipelines. Recent advancements Zhang et al. (2025); Yao et al. (2024) bolster stability by refining local aggregation and enforcing geometric consistency. Ultimately, the success of these paradigms fundamentally hinges on the network's ability to extract discriminative local features coupled with comprehensive global context.

Transformer and State-Space Models Transformers Vaswani et al. (2017) have been widely adopted for point cloud registration by modeling long-range dependencies and integrating geometric cues Qin et al. (2022); Yu et al. (2024); Yew & Lee (2022). Subsequent works have extensively explored attention optimizations to address scalability, efficiency, and multi-modal fusion Liu et al. (2023); She et al. (2024); Huang et al. (2024); Seo et al. (2025); Fung et al. (2025). Despite outperforming early learning-based methods, their quadratic computational complexity imposes prohibitive memory and runtime bottlenecks. Recently, state-space models (SSMs) like Mamba Han et al. (2024); Deng et al. (2025) have been introduced to achieve linear-time complexity. However, due to the forced 1D serialization of unordered 3D data, pure SSMs inherently struggle to capture non-sequential 3D spatial symmetries and global context. To circumvent this, rather than relying on a pure SSM, we propose a hybrid Transformer-Mamba architecture that elegantly combines the

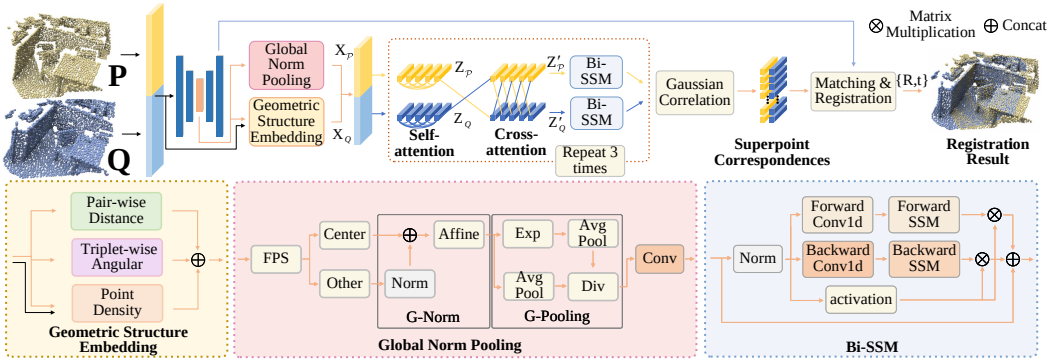


Figure 2: Overview of the proposed DAMT architecture. The method initiates by downsampling the raw point clouds. Structural features are systematically extracted via the Geometric Structure Embedding Module, explicitly integrating density information. Concurrently, the Global Norm Pooling (GNP) module aggregates and propagates global contextual features. Leveraging these fused multi-modal features, the Transformer-Mamba Block performs deep hierarchical feature learning, ultimately outputting robust superpoint correspondences for precise rigid transformation estimation.

permutation-invariant global initialization of attention with the linear-scaling deep feature aggregation of Mamba.

Methods with geometric structure information Explicitly embedding structural priors—such as spatial distances, angular relations, and surface normals—has proven highly effective in enhancing feature discriminability Liu et al. (2021); Yuan et al. (2024); Bai et al. (2021). For instance, GeoTransformer Qin et al. (2022) and its derivatives He et al. (2023); Mu et al. (2024) directly integrate local structural representations into attention mechanisms to refine alignment accuracy. Nevertheless, these methods predominantly focus on distance and angular metrics, largely overlooking point cloud density. As a critical geometric cue that reflects local detail variations and delineates structural boundaries (e.g., edges and corners), explicit density modeling is indispensable for resolving matching ambiguities in sparse regions and achieving robust registration.

3 Method

3.1 Overview

Given a source point cloud $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, ..., N\}$ and a target point cloud $\mathcal{Q} = \{\mathbf{q}_i \in \mathbb{R}^3 \mid i = 1, ..., N\}$, registration estimates the rigid transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ minimizing the spatial discrepancy over a set of corresponding point pairs $\mathcal{C}$:

$$\min_{\mathbf{R},\mathbf{t}} \frac{1}{|\mathcal{C}|} \sum_{(\mathbf{p}_i,\mathbf{q}_i)\in\mathcal{C}} \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2 \tag{1}$$

where $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$; the normalization by $|\mathcal{C}|$ ensures scale-invariance and numerical stability. Figure 2 illustrates our Density-Aware Mamba-Transformer (DAMT), which follows a coarse-to-fine matching paradigm: the Global Norm Pooling (GNP) module (Sec. 3.2) propagates a permutation-invariant global context to superpoints, the Transformer-Mamba block (Sec. 3.3) fuses density cues and geometric priors into discriminative descriptors for robust coarse matching, and fine matching then recovers dense point-to-point correspondences, with training objectives and losses detailed in Sec. 3.4.

3.2 Global Norm Pooling

To efficiently capture long-range spatial context without incurring the quadratic complexity of deep Transformers, we propose the lightweight Global Norm Pooling (GNP) module, inspired by the Local Norm Pooling in Mamba3D Han et al. (2024). Initially, local features from raw point clouds

$\mathcal{P}$ and $\mathcal{Q}$ are extracted and downsampled via a KPConv-FPN backbone, yielding superpoints $\hat{\mathcal{P}}$ and $\hat{\mathcal{Q}}$. Rather than relying on computationally heavy hierarchical stacks, the GNP module extracts and propagates a permutation-invariant global context to each superpoint through two streamlined operations: feature propagation (G-Norm) and feature aggregation (G-Pooling), as illustrated in Fig. 1 (Left).

G-Norm: Propagation Taking the source superpoints $\hat{\mathcal{P}}$ as an example, we apply Farthest Point Sampling (FPS) to select $K_c = 32$ anchor center points, explicitly denoted as $\hat{\mathcal{P}}_c$, capturing a multi-centric structural core. To efficiently propagate features without incurring global pairwise computation, each remaining point $\hat{p}_o \in \hat{\mathcal{P}} \setminus \hat{\mathcal{P}}_c$ is deterministically assigned to its nearest anchor center via $\arg\min_{c \in \hat{\mathcal{P}}_c} \|\hat{p}_o - c\|_2$. This establishes a many-to-one nearest-center mapping scheme, restricting the spatial distance computation complexity to a highly scalable $\mathcal{O}(|\hat{\mathcal{P}}| \cdot K_c)$.

Let $\mathbf{F}_O$ denote the features of the remaining discrete points, and $\mathbf{F}_C^{\rightarrow O}$ denote the broadcasted features of their correspondingly assigned centers. We normalize $\mathbf{F}_O$ relative to these assigned centers and apply a learnable affine transformation to derive the propagated feature $\widehat{\mathbf{F}}_O$:

$$\begin{cases} \mathbf{F}'_O = \dfrac{\mathbf{F}_O - \mathbf{F}_C^{\rightarrow O}}{\sqrt{\mathrm{Var}(\mathbf{F}_O - \mathbf{F}_C^{\rightarrow O}) + \epsilon}} \\ \widehat{\mathbf{F}}_O = (\mathbf{F}'_O \oplus \mathbf{F}_C^{\rightarrow O})\mathbf{W}_\gamma + \beta \end{cases} \tag{2}$$

where $\mathbf{W}_\gamma$ and β are learnable projection weights and biases, $\oplus$ denotes channel-wise concatenation, $\epsilon = 10^{-5}$ is a stability constant, and $\mathrm{Var}(\cdot)$ computes the spatial variance across the assigned point clusters.

G-Pooling: Aggregation Standard maximum pooling across unordered point sets often discards fine-grained structural details. To execute global aggregation while mitigating feature collapse in partial-overlap scenarios—where structural centers may shift—we introduce a channel-independent spatial softmax attention mechanism:

$$\widehat{\mathbf{F}}_C = \sum_{i=1}^{|\hat{\mathcal{P}}|} \Big(\mathrm{Softmax}(\widehat{\mathbf{F}}_O \mathbf{W}_s)\Big)_i \odot \widehat{\mathbf{F}}_O^i \tag{3}$$

where $\mathbf{W}_s$ is a learnable projection matrix and $\odot$ denotes element-wise multiplication. Crucially, the $\mathrm{Softmax}(\cdot)$ operator is applied spatially but independently across each feature channel. This effectively acts as a soft-gating mechanism that dynamically reweighs the structural prior without entangling disparate feature channels.

The identical computational procedure is applied symmetrically to the target set $\hat{\mathcal{Q}}$. By establishing this lightweight global graph, the GNP module efficiently equips every superpoint with a robust global receptive field, optimally preparing them for the subsequent geometry-aware matching stages.

3.3 Transformer-Mamba Block

Self-attention Following the GNP computation, we concatenate the global feature $\widehat{\mathbf{F}}_C$ with the local features $\hat{\mathbf{F}}_{\mathcal{P}}$ and $\hat{\mathbf{F}}_{\mathcal{Q}}$ extracted by the KPConv-FPN, yielding augmented representations $\mathbf{X}_{\mathcal{P}}$ and $\mathbf{X}_{\mathcal{Q}}$. To establish geometry-aware correlations, we directly inject a decoupled geometric structural embedding $\mathbf{r}_{i,j}^{geo}$ and a novel Density-Guided Attention Weighting (DGAW) factor $\Phi_{i,j}^{De}$ into the core attention mechanism. Taking $\hat{\mathcal{P}}$ as an example, the output feature matrix $\mathbf{Z}_{\mathcal{P}} = \{\mathbf{z}_{p1}, \mathbf{z}_{p2}, ..., \mathbf{z}_{pn}\}$ is obtained via:

$$\mathbf{z}_{pi} = \sum_{j=1}^{|\hat{\mathcal{P}}|} \sigma\left(\frac{\left(\mathbf{x}_{pi}\mathbf{W}^Q\right)\left(\mathbf{x}_{pj}\mathbf{W}^K + \mathbf{r}_{i,j}^{geo}\mathbf{W}^R\right)^T}{\sqrt{d_t}} + \Phi_{i,j}^{De} \right)(\mathbf{x}_{pj}\mathbf{W}^V) \tag{4}$$

where $\sigma(\cdot)$ is the softmax function, d_t indicates the channel dimension, and $\mathbf{W}^Q$, $\mathbf{W}^K$, $\mathbf{W}^V$, $\mathbf{W}^R$ are projection matrices. By formulating density as an explicit attention-gating constraint rather than a naive feature addition, the network dynamically filters out geometrically inconsistent outliers in sparse scenarios.

Point Density Embedding To account for varying scanning mechanisms (e.g., LiDAR density decaying with distance), we normalize the raw point count by the expected density profile at its respective radial distance. This yields a range-normalized relative density feature:

$$\mathbf{r}_i^{De} = \frac{|\{p_k \in \mathcal{P} \mid \|p_k - \hat{p}_i\|_2 < r\}|}{\rho_{expected}(d_i)} \tag{5}$$

where r is the local search radius (detailed in Appendix E), d_i denotes the radial distance to the sensor origin, and $\rho_{expected}(d_i)$ is the theoretical density expectation. For LiDAR scans, we model this as $\rho_{expected}(d_i) = \alpha/d_i^2$. Crucially, α serves merely as a base scaling constant (e.g., empirically initialized to the average point count at unit distance). Because $\mathbf{r}_i^{De}$ is subsequently processed by the non-linear DGAW module, precise sensor-specific calibration of α is unnecessary; any global scalar miscalibration is inherently absorbed by the learnable MLP weights. This guarantees that the network focuses on relative structural density variations rather than absolute scalar values, ensuring robust generalization. For uniformly sampled CAD models, we compute a global expectation: $\rho_{expected} = (N/V) \cdot \frac{4}{3}\pi r^3$.

For indoor RGB-D scans (e.g., 3DMatch) featuring severe occlusions, this uniform expectation acts purely as a base normalization scalar. The non-linear DGAW module mitigates inherent noise by learning relative density consistencies, effectively extracting robust structural cues even from non-ideal distributions (theoretical details in Appendix F).

Density-Guided Attention Weighting (DGAW) Following Qin et al. (2022), we compute pairwise distance and triplet-wise angular embeddings using high-frequency sinusoidal functions (Appendix B) to formulate the pure geometric embedding $\mathbf{r}_{i,j}^{geo}$:

$$\mathbf{r}_{i,j}^{geo} = \mathbf{r}_{i,j}^{Di}\mathbf{W}^{Di} + \max_x \left\{\mathbf{r}_{i,j,x}^{A}\mathbf{W}^{A}\right\} \tag{6}$$

While existing methods capture distance and angular cues, they often overlook local density. Instead of a linear summation, we explicitly enforce neighborhood density consistency to enhance discriminability in geometrically ambiguous regions. The density consistency factor $\Phi_{i,j}^{De}$ is computed as:

$$\Phi_{i,j}^{De} = \mathrm{MLP}\left(\mathbf{r}_i^{De} \oplus \mathbf{r}_j^{De} \oplus \left|\mathbf{r}_i^{De} - \mathbf{r}_j^{De}\right|\right) \tag{7}$$

This design explicitly models the structural density variation, heavily penalizing matches across drastically different sampling densities.

Cross-attention Cross-attention exchanges reciprocal features between the two point clouds. Given self-attention features $\mathbf{Z}_\mathcal{P}$ and $\mathbf{Z}_\mathcal{Q}$, the cross-attention matrix $\mathbf{Z}'_\mathcal{P} = \{\mathbf{z}'_{p1}, \mathbf{z}'_{p2}, ..., \mathbf{z}'_{pn}\}$ is computed as:

$$\mathbf{z}'_{pi} = \sum_{j=1}^{|\hat{\mathcal{Q}}|} \sigma\left(\frac{(\mathbf{z}_{pi}\mathbf{W}^Q)(\mathbf{z}_{qj}\mathbf{W}^K)^T}{\sqrt{d_t}}\right)(\mathbf{z}_{qj}\mathbf{W}^V) \tag{8}$$

Mamba Block The refined reciprocal features are then fed into the Mamba block, which adopts a Transformer-like design with dual state-space models (SSMs), LayerNorm, and residual connections. Detailed continuous-to-discrete formulations of the SSM are provided in Appendix A.

Applying 1D SSMs directly to 3D point clouds is sub-optimal due to the lack of canonical ordering Deng et al. (2025). Since the preceding Attention layers effectively encode spatial structures, we smoothly transition to the SSM by serializing the geometry-aware 3D superpoints into a 1D continuous sequence using a Z-order space-filling curve. This preserves spatial locality, ensuring proximal 3D points remain adjacent in 1D. We then apply a bi-directional SSM (bi-SSM) configuration to capture the global structure from both forward and reverse traversals:

$$\text{bi-SSM}(\mathbf{Z}'_\mathcal{P}) = \mathbf{Z}'_\mathcal{P} + \phi(\mathbf{Z}'_\mathcal{P}) + \psi(\mathbf{Z}'_{\mathcal{P}_r}) \tag{9}$$

where $\phi(\cdot)$ and $\psi(\cdot)$ symbolize forward and reverse sequence modeling, respectively, and $\mathbf{Z}'_{\mathcal{P}_r}$ is the reversed 1D sequence. The overall feature computation process is summarized as:

$$\begin{cases} \mathbf{Z} & = LN(\text{Self-attention}(\mathbf{X}) + \mathbf{X}) \\ \mathbf{Z}' & = LN(\text{Cross-attention}(\mathbf{Z}) + \mathbf{Z}) \\ \mathbf{Z}'' & = \text{bi-SSM}(\mathbf{Z}') + \mathbf{Z}' \end{cases} \tag{10}$$

where LN represents layer normalization.

3.4 Loss Function

The overall loss function comprises an overlap-aware circle loss $\mathcal{L}_{oc}$, a point matching loss $\mathcal{L}_p$, and a point density loss $\mathcal{L}_{den}$:

$$\mathcal{L} = \mathcal{L}_{oc} + \mathcal{L}_p + \mathcal{L}_{den} \tag{11}$$

The complete mathematical formulations and parameter settings for $\mathcal{L}_{oc}$ and $\mathcal{L}_p$ are detailed in Appendix C. To enforce geometric consistency, $\mathcal{L}_{den}$ supervises the predicted density embeddings. Instead of using a non-differentiable set cardinality metric, we apply a Mean Squared Error (MSE) constraint against the ground-truth local density distribution $D(\hat{p})$:

$$\mathcal{L}_{den} = \frac{1}{|\hat{\mathcal{P}}|} \sum_{\hat{p} \in \hat{\mathcal{P}}} \| f_{den}(\hat{p}) - D(\hat{p}) \|_2^2 \tag{12}$$

where $f_{den}(\hat{p})$ is the network's predicted continuous density feature, and $D(\hat{p})$ is the normalized pre-computed density at point $\hat{p}$.

3.5 Feature Matching and Registration

Given the density-aware, geometry-encoded superpoint features $\widehat{\mathbf{F}}_{\mathcal{P}}$ and $\widehat{\mathbf{F}}_{\mathcal{Q}}$, we derive $(\mathbf{R}, \mathbf{t})$ by coarse-to-fine matching: in the coarse stage, dual-softmax over pairwise feature distances yields superpoint correspondences; in the fine stage, optimal transport expands them to point level via dense similarities in matched patches. The Local Global Registration (LGR) module Qin et al. (2022) with SVD then solves for $\mathbf{R}$ and $\mathbf{t}$ analytically. This pipeline turns structural and density priors into accurate pose estimates; details are in Appendix E.

4 Experiments

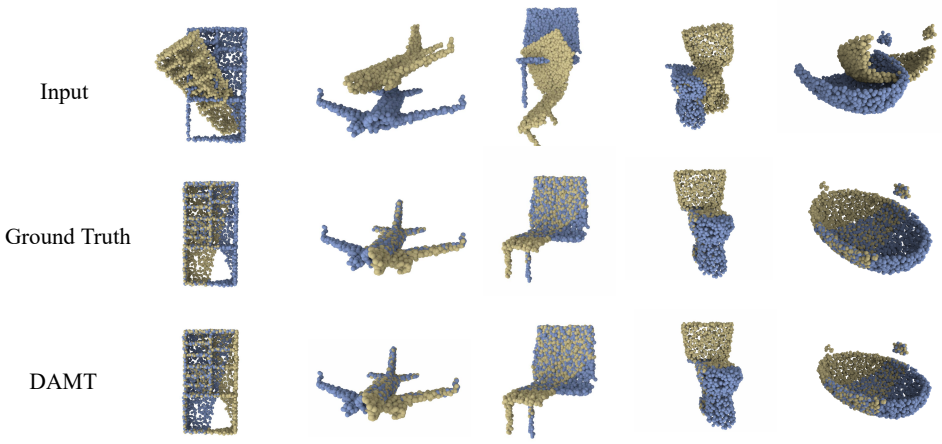


Figure 3: Qualitative visualization results on ModelNet. From top to bottom: the input point cloud pairs, the ground truth registration, and the registration results generated by DAMT.

To ensure standardized comparisons with existing methods, we evaluate DAMT across three distinct domains: outdoor LiDAR (KITTI Geiger et al. (2012)), synthetic CAD (ModelNet and ModelLoNet

Wu et al. (2015)), and indoor RGB-D (3DMatch and 3DLoMatch Huang et al. (2021)). The inclusion of low-overlap variants (ModelLoNet and 3DLoMatch) explicitly tests the model's robustness under challenging, sparse correspondence scenarios. Finally, we conduct detailed ablation studies to isolate the individual contributions of the GNP module, the Mamba mechanism, and the density-aware embeddings.

4.1 Implementation Details

Our model is implemented in PyTorch. The Mamba block state dimension is set to 16. For the geometric structure embedding, we set hyperparameters $\sigma_d = 0.1$, $\sigma_a = 15^\circ$, and $k = 10$ nearest neighbors. The loss weight ω is set to 0.5. Note that the runtime reported in our tables measures the pure network forward pass and excludes the time taken by the post-processing 50k RANSAC iterations.

4.2 Experimental Results

KITTI We evaluate our method on the KITTI odometry benchmark, a widely adopted standard for large-scale outdoor LiDAR registration, which ensures direct and fair comparisons with existing state-of-the-art methods. We adopt four standard evaluation metrics: Root Mean Square Error of translation (RTE, in cm) and rotation (RRE, in °), Registration Recall (RR, in %), the percentage of successful registrations, and inference Time (s) measured on a single RTX 3090 GPU. Following previous works Yao et al. (2024); Huang et al. (2024); Wang et al. (2025), we sample 30k points per point cloud and train using the Adam optimizer for 160 epochs; during testing, we execute 50k RANSAC iterations to robustly estimate the rigid transformations. Table 1 presents the quantitative results: DAMT improves on the recent leading method CAST in both RTE and RRE, attains a perfect 100.0% RR, and runs significantly faster than recent estimators such as PARE-Net, demonstrating its robustness in sparse outdoor environments.

Table 1: Outdoor KITTI evaluation results. DAMT attains 2.3 cm RTE, 0.22° RRE and 100.0% RR at 0.15 s per pair, mitigating the accuracy–speed trade-off of recent estimators.

Model	RTE(cm)↓	RRE(°)↓	RR(%)↑	Time(s)↓
3DFeat-Net	25.9	0.25	96.0	38.07
Predator	6.8	0.27	99.8	0.77
GeoTrans	6.8	0.24	99.8	0.26
BUFFER	7.1	0.26	99.8	0.27
Rethinking	6.6	0.24	99.8	-
DiffusionPCR	6.3	0.23	99.8	-
MATR	6.6	0.25	99.8	-
END	6.6	0.25	99.8	-
PARE-Net	4.9	0.23	99.8	0.21
CAST	2.5	0.27	**100.0**	0.23
PTT	6.3	0.25	99.8	-
BUFFER-X	7.7	0.27	99.8	-
DAMT(Ours)	**2.3**	**0.22**	**100.0**	**0.15**

Table 2: ModelNet and ModelLoNet evaluation results. On ModelNet, GNP propagates global context efficiently; on the low-overlap ModelLoNet, our density embedding resolves feature ambiguity.

Methods	ModelNet			ModelLoNet		
	RRE↓	RTE↓	CD↓	RRE↓	RTE↓	CD↓
RPM-Net	1.712	0.018	8.5e-4	7.342	0.124	5.0e-3
Predator	1.856	0.019	8.9e-4	5.24	0.132	8.3e-3
RegTR	1.331	0.011	7.8e-4	3.93	0.087	3.7e-3
GeoTrans	1.856	0.019	8.8e-4	5.462	0.133	7.9e-3
UDPReg	1.712	0.018	3.1e-2	3.58	0.069	4.2e-2
MTAR	1.277	0.016	6.0e-4	1.618	0.122	3.5e-3
END	1.292	0.014	5.0e-4	1.632	0.108	2.9e-3
PTT	1.30	0.010	7.5e-4	3.65	0.068	3.1e-3
IAPReg	1.138	0.010	-	2.842	0.031	-
DAMT(Ours)	**1.127**	**0.009**	**4.7e-4**	**1.613**	**0.029**	**2.2e-3**

ModelNet & ModelLoNet We evaluate on ModelNet, a synthetic 3D CAD benchmark for registration under ideal conditions, and on ModelLoNet, its low-overlap variant testing generalization under sparse correspondences. Metrics follow KITTI (RTE in m, RRE in °), plus Chamfer Distance (CD) for the global geometric consistency of the registered clouds. Following previous works Qin et al. (2022); Wang et al. (2024; 2025), we sample 30k points per point cloud, train for 40 epochs with Adam on an RTX 3090 GPU, and run 50k RANSAC iterations at test time. Table 2 reports the results: DAMT attains the lowest RTE (0.009 m) and CD (4.7e-4) on ModelNet and the lowest CD (2.2e-3) and RTE (0.029 m) on the challenging ModelLoNet, with competitive RRE (1.127° / 1.613°), outperforming recent methods such as END and MTAR. Qualitative results are in Figure 3.

3DMatch & 3DLoMatch We evaluate on 3DMatch, an RGB-D indoor benchmark with noisy low-resolution scans, and on 3DLoMatch, its 10%–30% overlap subset for extremely sparse correspondences. Metrics are Feature Matching Recall (FMR, %), Inlier Ratio (IR, %), Registration Recall (RR, %) and inference time (s). Following Qin et al. (2022), we sample 30k points per cloud,

Table 3: Indoor 3DMatch and 3DLoMatch evaluation results.

Dataset	3DMatch					3DLoMatch					Time (s)↓
# Samples	5000	2500	1000	500	250	5000	2500	1000	500	250	Average
Feature Matching Recall(%)↑											
D3Feat	95.6	95.4	94.5	94.1	93.1	67.3	66.7	67.0	66.7	66.5	-
Predator	96.6	96.6	96.5	96.3	96.5	78.6	77.4	76.3	75.7	75.3	-
YOHO	98.2	97.6	97.5	97.7	96.0	79.4	78.1	76.3	73.8	69.1	-
CoFiNet	98.1	98.3	98.1	98.2	**98.3**	83.1	83.5	83.3	83.1	82.6	-
GeoTrans	97.9	97.9	97.9	97.9	97.6	88.3	88.6	88.8	88.6	88.3	-
SemReg	**98.6**	-	-	-	-	89.8	-	-	-	-	-
PARE-Net	98.5	-	-	-	-	89.2	-	-	-	-	-
IAPReg	97.6	-	-	-	-	89.6	-	-	-	-	-
DAMT(Ours)	**98.6**	**98.4**	**98.3**	**98.3**	**98.3**	**90.0**	**89.4**	**89.4**	**89.2**	**88.9**	-
Inlier Ratio (%)↑											
D3Feat	39.0	38.8	40.4	41.5	41.8	13.2	13.1	14.0	14.6	15.0	-
Predator	58.0	58.4	57.1	54.1	49.3	26.7	28.1	28.3	27.5	25.8	-
YOHO	64.4	60.7	55.7	46.4	41.2	25.9	23.3	22.6	18.2	15.0	-
CoFiNet	49.8	51.2	51.9	52.2	52.2	24.4	25.9	26.7	26.8	26.9	-
GeoTrans	71.9	75.2	76.0	82.2	85.1	43.5	45.3	46.2	52.9	57.7	-
SemReg	56.2	-	-	-	-	29.4	-	-	-	-	-
PARE-Net	76.9	-	-	-	-	47.5	-	-	-	-	-
IAPReg	72.7	-	-	-	-	45.2	-	-	-	-	-
DAMT(Ours)	**77.6**	**79.1**	**83.7**	**85.5**	**86.6**	**48.3**	**50.1**	**57.4**	**59.0**	**60.9**	-
Registration Recall (%)↑											
D3Feat	81.6	84.5	83.4	82.4	77.9	37.2	42.7	46.9	43.8	39.1	0.289
Predator	89.0	88.9	90.6	88.5	87.8	59.8	61.2	62.4	60.8	58.1	0.759
YOHO	90.8	90.3	89.1	88.6	85.5	65.2	65.5	63.2	56.5	48.0	13.529
CoFiNet	89.3	88.9	88.4	87.4	87.0	67.5	66.2	64.2	63.1	61.0	0.306
GeoTrans	92.0	91.8	91.8	91.4	91.2	75.0	74.8	74.2	74.1	73.5	0.192
SemReg	94.7	-	-	-	-	75.9	-	-	-	-	-
PARE-Net	95.0	-	-	-	-	80.5	-	-	-	-	0.183
CAST	95.2	95.1	94.5	94.0	93.5	75.1	74.8	74.5	73.9	73.2	0.182
IAPReg	93.8	-	-	-	-	75.9	-	-	-	-	-
DAMT(Ours)	**95.4**	**95.3**	**94.9**	**94.8**	**94.6**	**81.2**	**76.3**	**76.0**	**75.8**	**75.4**	**0.125**

train 40 epochs with Adam, and test at sample sizes from 5000 down to 250 with 50k RANSAC iterations. Table 3 reports the results: on 3DMatch, DAMT attains FMR and IR comparable to recent state of the art at 0.125s per pair; on the challenging 3DLoMatch, it reaches an RR of 81.2% at 5000 samples, outperforming PARE-Net, and it attains the highest RR at 250 samples, surpassing CAST. This resilience to sparse, low-overlap input indicates that GNP stabilizes global semantics while the density embedding supplies structural cues under fragmented overlap. Qualitative results are in Figure 4.

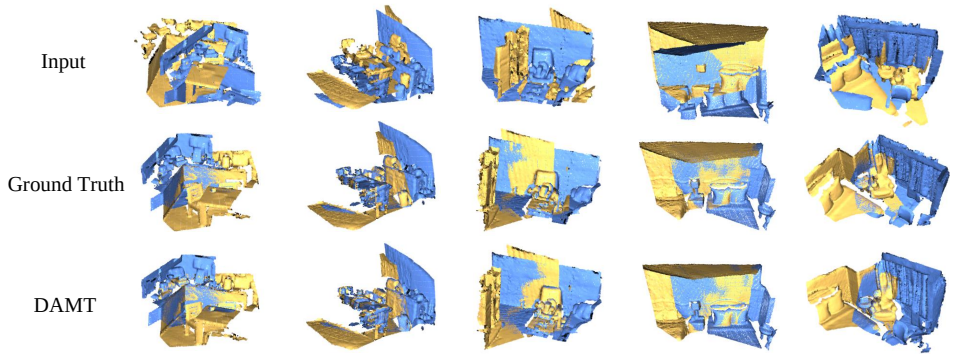


Figure 4: Qualitative registration results on 3DMatch. DAMT accurately aligns point cloud pairs across varying overlap ratios, demonstrating robust feature matching in challenging indoor scenarios.

Table 5: Ablation experiments on Mamba vs. Self-Attention.

Model	PIR↑	IR↑	RRE↓	RTE↓	RMSE↓	RR↑
DAMT	0.955	**0.817**	**1.127**	**0.009**	**0.010**	**63.3**
A	0.910	0.794	1.359	0.016	0.015	60.9
B	**0.957**	0.816	1.217	0.011	0.011	62.6

Table 6: Ablation on 1D Serialization Strategies for the Mamba Block on ModelNet.

Serialization Strategy	RRE↓	RTE↓	RR↑	Time↓
Random Sequence	1.632	0.017	55.4	**0.1**
Graph Traversal (BFS)	1.138	0.010	62.6	12.4
Hilbert Curve	1.135	0.010	63.1	4.5
Z-order Curve (Ours)	**1.127**	**0.009**	**63.3**	0.8

Table 4: Ablation experiments on Module Contribution Analysis. GNP supplies global semantic anchors, the bi-SSM (Mamba) block deep sequential features, and DGAW the density prior that raises RR.

GNP	Den	Mam	PIR↑	IR↑	RRE↓	RTE↓	RMSE↓	RR↑
✓			0.948	0.811	1.237	0.012	0.011	61.8
	✓		0.949	0.809	1.275	0.013	0.012	61.4
		✓	0.939	0.809	1.238	0.017	0.015	61.9
✓	✓		0.953	0.812	1.261	0.012	0.011	62.1
	✓	✓	0.950	0.810	1.259	0.013	0.012	62.2
✓		✓	0.948	0.814	1.192	0.012	0.011	62.7
✓	✓	✓	**0.955**	**0.817**	**1.127**	**0.009**	**0.010**	**63.3**

Takeaway: the three modules are complementary—DAMT attains the best value on all six metrics.

Modeling density as a relative expectation prior lets DAMT extract structural features adaptively, without dataset-specific tuning. It performs consistently on synthetic CAD (ModelNet), indoor RGB-D (3DMatch) and outdoor LiDAR (KITTI), showing that its invariant geometric-density cues mitigate the cross-modality generalization bottleneck.

4.3 Ablation Study

To validate the effectiveness of each component within DAMT, we conducted comprehensive ablation experiments on ModelNet.

Module Contribution Analysis Table 4 ablates our three core modules individually and jointly. No pair compensates for the missing third: relative to the best partial variant on RRE and RR (GNP + Mamba, 1.192° / 62.7%), adding DGAW further lowers RRE to 1.127° and raises RR to 63.3%.

Mamba vs. Self-Attention. Table 5 compares DAMT with pure-Mamba (Model A) and pure self-attention (Model B) baselines. Model A degrades sharply (RR 60.9%): vanilla Mamba cannot capture spatial symmetries under forced 1D serialization, which our lightweight self-attention initialization restores. Appendix D.1 adds the scaling evidence: 9.2 GB at 100k points, versus Out-of-Memory (OOM) for pure attention.

Impact of 1D Serialization Strategy Table 6 compares serialization strategies on ModelNet. A Random Sequence destroys spatial locality (RR 55.4%); BFS preserves geometry but resists parallelization (12.4 ms); Hilbert and Z-order keep high RR (63.1%/63.3%). We adopt Z-order, whose bitwise interleaving costs only 0.8 ms.

Comprehensive Supplementary Evaluations. Due to space limits, extensive analyses validating DAMT's completeness and robustness are detailed in the Appendix. These include: (1) theoretical proofs decoupling our density formulation from sensor noise (Appendix F); (2) structural validations and GNP module decompositions (Appendix D.2, D.3, D.4); (3) hyperparameter sensitivity evaluations (Appendix E); (4) comparisons against alternative density heuristics (Appendix E.3); and (5) mathematical clarifications (Appendix G) alongside extended qualitative visualizations across all benchmarks (Appendix I).

5 Conclusion

We propose the DAMT for point cloud registration, a hybrid architecture coupling Transformer's global context modeling with Mamba's linear-time efficiency. Our lightweight GNP module efficiently extracts and propagates permutation-invariant global representations. Furthermore, we embed local point cloud density as a structural prior via a novel attention-weighting mechanism and consistency loss, mitigating feature ambiguities in sparse regions. Extensive evaluations across synthetic CAD, indoor RGB-D, and outdoor LiDAR benchmarks demonstrate DAMT's superior trade-off between registration accuracy and computational efficiency.

AI USE STATEMENT

Generative AI tools were used only to assist with LaTeX format typesetting and language-level proofreading. The authors remain responsible for all scientific claims, experimental results, citations, and final manuscript content, and verified the generated formatting and compilation output.

REFERENCES

Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6358–6366, 2020. doi: 10.1109/CVPR42600.2020.00639.

Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15854–15864, 2021. doi: 10.1109/CVPR46437.2021.01560.

Guangyan Chen, Meiling Wang, Yi Yang, Li Yuan, and Yufeng Yue. Fast and robust point cloud registration with tree-based transformer. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 773–780, 2024a. doi: 10.1109/ICRA57147.2024.10610004.

Guangyan Chen, Meiling Wang, Qingxiang Zhang, Li Yuan, and Yufeng Yue. Full transformer framework for robust point cloud registration with deep information interaction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13368–13382, 2024b. doi: 10.1109/TNNLS.2023.3267333.

Liyuan Deng, Yunpeng Bai, Yongkang Dai, Xiaoshui Huang, Hongping Gan, Dongshuo Huang, Hao Jiacheng, and Yilei Shi. Mamtiff-cad: Multi-scale latent diffusion with mamba+ for complex parametric sequence. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10517–10526, 2025. doi: 10.1109/ICCV51701.2025.00979.

Sheldon Fung, Xuequan Lu, Dasith de Silva Edirimuni, Wei Pan, Xiao Liu, and Hongdong Li. Semreg: Semantics constrained point cloud registration. In *Computer Vision – ECCV 2024*, pp. 293–310. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-72940-9_17.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012. doi: 10.1109/CVPR.2012.6248074.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12):4338–4364, 2021. doi: 10.1109/TPAMI.2020.3005434.

Xu Han, Yuan Tang, Zhaoxuan Wang, and Xianzhi Li. Mamba3d: Enhancing local features for 3d point cloud analysis via state space model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4995–5004, 2024. ISBN 9798400706868. doi: 10.1145/3664647.3681173.

Leping He, Shuaiqing Wang, Qijun Hu, Qijie Cai, Muyao Li, Yu Bai, Kai Wu, and Bo Xiang. Gfoicp: Geometric feature optimized iterative closest point for 3-d point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2023. doi: 10.1109/TGRS.2023.3317822.

Yun He, Xinlin Ren, Danhang Tang, Yinda Zhang, Xiangyang Xue, and Yanwei Fu. Density-preserving deep point cloud compression. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2323–2332, 2022. doi: 10.1109/CVPR52688.2022.00237.

Renlang Huang, Yufan Tang, Jiming Chen, and Liang Li. A consistency-aware spot-guided transformer for versatile and hierarchical point cloud registration. In *Advances in Neural Information Processing Systems*, volume 37, pp. 70230–70258. Curran Associates, Inc., 2024. doi: 10.52202/079017-2245.

Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4265–4274, 2021. doi: 10.1109/CVPR46437.2021.00425.

Le Hui, Jia Yuan, Mingmei Cheng, Jin Xie, Xiaoya Zhang, and Jian Yang. Superpoint network for point cloud oversegmentation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5490–5499, 2021. doi: 10.1109/ICCV48922.2021.00546.

Jiuming Liu, Guangming Wang, Zhe Liu, Chaokang Jiang, Marc Pollefeys, and Hesheng Wang. Regformer: An efficient projection-aware transformer network for large-scale point cloud registration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8417–8426, 2023. doi: 10.1109/ICCV51070.2023.00776.

Weixiao Liu, Hongtao Wu, and Gregory S. Chirikjian. Lsg-cpd: Coherent point drift with local surface geometry for point cloud registration. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15273–15282, 2021. doi: 10.1109/ICCV48922.2021.01501.

Fan Lu, Guang Chen, Yinlong Liu, Zhongnan Qu, and Alois Knoll. Rskdd-net: Random sample-based keypoint detector and descriptor. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 21297–21308. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/f40723ed94042ea9ea36bfb5ad4157b2-Paper.pdf.

Weixin Lu, Guowei Wan, Yao Zhou, Xiangyu Fu, Pengfei Yuan, and Shiyu Song. Deepvcp: An end-to-end deep neural network for point cloud registration. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12–21, 2019. doi: 10.1109/ICCV.2019.00010.

Taewon Min, Eunseok Kim, and Inwook Shim. Geometry guided network for point cloud registration. *IEEE Robotics and Automation Letters*, 6(4):7270–7277, 2021. doi: 10.1109/LRA.2021.3097268.

Juncheng Mu, Lin Bie, Shaoyi Du, and Yue Gao. Colorpcr: Color point cloud registration with multi-stage geometric-color fusion. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21061–21070, 2024. doi: 10.1109/CVPR52733.2024.01990.

Zheng Qin, Hao Yu, Changiian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11133–11142, 2022. doi: 10.1109/CVPR52688.2022.01086.

Minkyun Seo, Hyungtae Lim, Kanghee Lee, Luca Carlone, and Jaesik Park. Buffer-x: Towards zero-shot point cloud registration in diverse scenes. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3851–3862, 2025. doi: 10.1109/ICCV51701.2025.00367.

Rui She, Qiyu Kang, Sijie Wang, Wee Peng Tay, Kai Zhao, Yang Song, Tianyu Geng, Yi Xu, Diego Navarro Navarro, and Andreas Hartmannsgruber. Pointdifformer: Robust point cloud registration with neural diffusion and transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3351286.

Chenghao Shi, Xieyuanli Chen, Huimin Lu, Wenbang Deng, Junhao Xiao, and Bin Dai. Rdmnet: Reliable dense matching based point cloud registration for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 24(10):11372–11383, 2023. doi: 10.1109/TITS.2023.3286464.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Meiling Wang, Guangyan Chen, Yi Yang, Li Yuan, and Yufeng Yue. Point tree transformer for point cloud registration. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(7): 6756–6772, 2025. doi: 10.1109/TCSVT.2025.3526680.

Yong Wang, Pengbo Zhou, Guohua Geng, Li An, Mingquan Zhou, and Kang Li. Matr: Multi-compound adaptive transformer for point cloud registration. *IEEE Transactions on Industrial Informatics*, 20(11):12903–12912, 2024. doi: 10.1109/TII.2024.3430499.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1912–1920, 2015. doi: 10.1109/CVPR.2015.7298801.

Runzhao Yao, Shaoyi Du, Wenting Cui, Canhui Tang, and Chengwu Yang. Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In *ECCV*, pp. 287–303, 2024.

Zi Jian Yew and Gim Hee Lee. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

Zi Jian Yew and Gim Hee Lee. Regtr: End-to-end point cloud correspondences with transformers. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6667–6676, 2022. doi: 10.1109/CVPR52688.2022.00656.

Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. In *Advances in Neural Information Processing Systems*, volume 34, pp. 23872–23884. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/c85b2ea9a678e74fdc8bafe5d0707c31-Paper.pdf.

Zhiyuan Yu, Zheng Qin, Lintao Zheng, and Kai Xu. Learning instance-aware correspondences for robust multi-instance point cloud registration in cluttered scenes. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19605–19614, 2024. doi: 10.1109/CVPR52733.2024.01854.

Yongzhe Yuan, Yue Wu, Xiaolong Fan, Maoguo Gong, Wenping Ma, and Qiguang Miao. Egst: Enhanced geometric structure transformer for point cloud registration. *IEEE Transactions on Visualization and Computer Graphics*, 30(9):6222–6234, 2024. doi: 10.1109/TVCG.2023.3329578.

Yue Zhang, Yue Wu, Wenping Ma, Maoguo Gong, Hao Li, and Biao Hou. Partial point cloud registration with multi-view 2d image learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10329–10337, 2025. doi: 10.1609/aaai.v39i10.33121.

Mingyang Zhao, Lei Ma, Xiaohong Jia, Dong-Ming Yan, and Tiejun Huang. Graphreg: Dynamical point cloud registration with geometry-aware graph signal processing. *IEEE Transactions on Image Processing*, 31:7449–7464, 2022. doi: 10.1109/TIP.2022.3223793.

SUPPLEMENTARY MATERIAL

The attachment zip file contains the complete source code implementation (***DAMT_code.zip***) and supporting materials (***Technical_Appendix.pdf***).

DAMT_code.zip The code is developed based on Python, relying on the PyTorch framework and CUDA. The experimental environment can be easily replicated through the provided environment files, and the detailed configuration of the entire code, dataset preparation, and training/evaluation commands can be found in the README.md file in the code directory. All training scripts, evaluation codes, and configurations to reproduce the results on KITTI, ModelNet, and 3DMatch/3DLoMatch benchmarks are included.

Technical_Appendix.pdf The "Technical_Appendix.pdf" of DAMT includes:

- Section A Preliminaries: Mamba
- Section B Pair-wise Distance and Triplet-wise Angular Embeddings
- Section C Loss Function Details
  - Sec. C.1 Overlap-aware circle loss
  - Sec. C.2 Point matching loss
- Section D Further Analysis on the Architecture
  - Sec. D.1 Efficiency-Accuracy Trade-off of Hybrid Mamba-Transformer
  - Sec. D.2 Structural Design Validation
  - Sec. D.3 Ablation on GNP Center Points (K_c)
  - Sec. D.4 Global Context Strategy Comparison
  - Sec. D.5 GNP Block Composition
- Section E Hyperparameter Sensitivity and Robustness Analysis
  - Sec. E.1 Sensitivity to Neighborhood Radius (r)
  - Sec. E.2 Embedding Dimensionality and Loss Weights
  - Sec. E.3 Comparison with Alternative Density Estimation Methods
- Section F Theoretical Rationale of Point Density Embedding and Noise Robustness
  - Sec. F.1 Why Local Density Aids Matching
  - Sec. F.2 Decoupling Sensor Noise from True Structural Boundaries
- Section G Clarifications on Mathematical Formulations
  - Sec. G.1 Channel-independent Spatial Softmax in G-Pooling
  - Sec. G.2 Logical Bridge for Density Embedding (Eqs. (6) and (7))
- Section H Additional Discussion on Module Robustness
- Section I Additional Visualizations
- Section J Future Work

A PRELIMINARIES: MAMBA

Mamba is a state-space model that maps an input sequence $x(t)$ to an output sequence $y(t)$ via a hidden state $h(t)$. The State-Space Model (SSM) is a linear time-invariant (LTI) system defined by the following continuous-time ordinary differential equations:

$$\begin{cases} \dot{h}(t) = Ah(t) + Bx(t) \\ y(t) = Ch(t) + Dx(t) \end{cases} \tag{13}$$

Here, A, B, C, and D represent the state, control, output, and feedthrough matrices, respectively. To adapt this continuous system to discrete sequence data, the model needs to be discretised. The

parameters A and B are defined using the Zero Order Hold (ZOH) discretisation rule, which introduces a sampling step Δ and a Taylor expansion. In this case, the parameters can be approximated as:

$$\begin{cases} \bar{A} = e^{\Delta A}, \\ \bar{B} = (e^{\Delta A} - I)A^{-1}B, \\ \bar{C} = C, \end{cases} \tag{14}$$

where I is the identity matrix and e denotes the matrix exponential. After the transformation, the system discretises to:

$$h_k = \bar{A}h_{k-1} + \bar{B}x_k, \quad y_k = \bar{C}h_k \tag{15}$$

Finally, the model output y can be represented as the convolution of input x and a convolution kernel, which endows Mamba with the potential for parallel computation as: $y = \bar{K} * x$, where $\bar{K}$ is the convolution kernel represented by:

$$\bar{K} = (C\bar{B}, C\bar{A}\bar{B}, ..., C\bar{A}^{M-1}\bar{B}) \tag{16}$$

B Pair-wise Distance and Triplet-wise Angular Embeddings

Following the geometric insights from Qin et al. (2022), the pair-wise distance embedding $\mathbf{r}_{i,j}^{Di}$ is computed by applying a high-frequency sinusoidal function to the normalized distance $d_{i,j} = \|\hat{p}_i - \hat{p}_j\|_2/\sigma_d$. For the angular embedding $\mathbf{r}_{i,j,x}^{A}$, we select the k-nearest neighbors $\hat{P}_k$ of the superpoint $\hat{p}_i$. For each neighboring point $\hat{p}_x \in \hat{P}_k$, we compute the relative angle $\alpha_{i,j}^{x} = \angle(\Delta_{x,i}, \Delta_{j,i})/\sigma_a$ and apply the sinusoidal function:

$$\begin{cases} r_{i,j,2k}^{Di} = \sin\left(\dfrac{d_{i,j}/\sigma_d}{10000^{2k/d_t}}\right) \\ r_{i,j,2k+1}^{Di} = \cos\left(\dfrac{d_{i,j}/\sigma_d}{10000^{2k/d_t}}\right) \\ r_{i,j,x,2l}^{A} = \sin\left(\dfrac{\alpha_{i,j}^{x}/\sigma_a}{10000^{2l/d_t}}\right) \\ r_{i,j,x,2l+1}^{A} = \cos\left(\dfrac{\alpha_{i,j}^{x}/\sigma_a}{10000^{2l/d_t}}\right) \end{cases} \tag{17}$$

where σ_d and σ_a act as hyper-parameters tuning the network's sensitivity, and $\Delta_{i,j}$ signifies the vector difference.

C Loss Function Details

C.1 Overlap-aware circle loss

Following formulations in recent geometric processing works Qin et al. (2022), the overlap-aware circle loss on the source point cloud $\mathcal{P}$ is defined as:

$$\mathcal{L}_{oc}^{\mathcal{P}} = \frac{1}{|\mathcal{A}|} \sum_{\mathcal{G}_i^{\mathcal{P}} \in \mathcal{A}} \log \Bigg[1 + \sum_{\mathcal{G}_j^{\mathcal{Q}} \in \varepsilon_p^i} e^{\lambda_i^j \beta_p^{i,j} (d_i^j - \Delta_p)} \cdot \sum_{\mathcal{G}_k^{\mathcal{Q}} \in \varepsilon_n^i} e^{\beta_n^{i,k} (\Delta_n - d_i^k)} \Bigg] \tag{18}$$

where $\mathcal{A}$ is the set of patches in $\mathcal{P}$ that contain at least one positively matched sample (overlap ratio $\geq 10\%$). $\mathcal{G}_i^{\mathcal{P}}$ is a patch in $\mathcal{A}$. ε_p^i and ε_n^i are the sets of its positive and negative patches in the target $\mathcal{Q}$, respectively. d_i^j denotes the feature distance between $\mathcal{G}_i^{\mathcal{P}}$ and $\mathcal{G}_j^{\mathcal{Q}}$. $\Delta_p = 0.1$ defines the feature distance threshold for positive sample matching, and $\Delta_n = 1.4$ sets the margin between the boundaries of positive and negative samples. $\beta_p^{i,j} = \gamma(d_i^j - \Delta_p)$ and $\beta_n^{i,k} = \gamma(\Delta_n - d_i^k)$ are the weighting factors. Finally, the overlap-aware circle loss is formulated as $\mathcal{L}_{oc} = (\mathcal{L}_{oc}^{\mathcal{P}} + \mathcal{L}_{oc}^{\mathcal{Q}})/2$. This loss prevents high-confidence positive classes from being overly suppressed by positive gradients, leading to more reliable superpoint matching.

C.2 POINT MATCHING LOSS

The point matching loss is calculated by averaging the individual losses of all sampled superpoint matches: $\mathcal{L}_p = \frac{1}{N_g}\sum_{i=1}^{N_g}\mathcal{L}_{p,i}$. The single patch matching loss $\mathcal{L}_{p,i}$ is calculated as:

$$\begin{aligned}\mathcal{L}_{p,i} = &- \sum_{(x,y)\in\mathcal{M}_i} \log \bar{z}^i_{x,y} \\ &- \sum_{x\in\mathcal{I}_i} \log \bar{z}^i_{x,n_i+1} - \sum_{y\in\mathcal{J}_i} \log \bar{z}^i_{m_i+1,y}\end{aligned} \tag{19}$$

where $\bar{z}$ is the soft assignment matrix of each superpoint correspondence equipped with a dustbin mechanism. m_i and n_i explicitly denote the number of points in the source patch $\mathcal{P}_i$ and the target patch $\mathcal{Q}_i$, respectively. Therefore, indices $m_i + 1$ and $n_i + 1$ correspond to the added dustbin column/row for unmatched points. $\mathcal{M}_i$ is the set of matched points within a certain radius, while $\mathcal{I}_i$ and $\mathcal{J}_i$ denote the sets of unmatched points in the two respective patches.

D FURTHER ANALYSIS ON THE ARCHITECTURE

D.1 EFFICIENCY-ACCURACY TRADE-OFF OF HYBRID MAMBA-TRANSFORMER

While a pure Self-Attention model (Model B) achieves a comparable Registration Recall (RR) of 0.626 on the ModelNet dataset, its quadratic complexity inherently limits its scalability for dense point clouds. As demonstrated in Table 7 on the KITTI outdoor benchmark, our hybrid DAMT architecture intrinsically resolves this bottleneck. It achieves a significant 31.8% reduction in inference latency alongside noticeably lower FLOPs compared to Model B. This confirms that integrating Mamba is not merely a computational trade-off but provides a substantial efficiency gain while maintaining state-of-the-art accuracy.

Table 7: Efficiency comparison on KITTI. Metrics: RRE↓ / Latency↓ / FLOPs↓ / Params↓. Model B denotes the pure Self-Attention baseline.

Method	RRE (°)	Time (s)	FLOPs (G)	Params (M)
GeoTrans	0.24	0.26	28.4	14.2
Model B	0.23	0.22	24.1	12.8
DAMT (Ours)	**0.22**	**0.15**	**16.5**	**11.6**

Scalability across Varying Point Cloud Sizes To fully substantiate our claimed efficiency gains and explicitly demonstrate the near-linear scalability of the Mamba block against the quadratic explosion of standard Transformers, we conduct a scalability stress test. We evaluate the Inference Time and Peak GPU Memory usage across scaling point cloud sizes ($N = 30\text{k}, 60\text{k}, 100\text{k}$) on an NVIDIA RTX 3090 GPU.

As detailed in Table 8, as the number of points increases, the purely attention-based architecture (Model B) exhibits a quadratic surge in both Peak Memory and latency, rapidly exceeding standard GPU memory limits and inevitably leading to an Out-of-Memory (OOM) error at 100k points. In stark contrast, by deliberately delegating the deep feature aggregation to the bi-SSM (Mamba) module, our DAMT scales much more gracefully. The memory footprint and inference latency of DAMT grow at a near-linear rate, successfully processing 100k points with only 9.2 GB of memory.

Table 8: Scalability test: Peak GPU Memory (GB) and Inference Time (s) across varying point cloud sizes (N). Model B denotes the Attention-Only baseline.

Method	$N = 30\text{k}$		$N = 60\text{k}$		$N = 100\text{k}$	
	Mem	Time	Mem	Time	Mem	Time
Model B	4.2	0.22	14.8	0.85	OOM	OOM
DAMT (Ours)	**2.8**	**0.15**	**5.3**	**0.32**	**9.2**	**0.55**

Ultimately, the Transformer module excels at high-precision initial correspondence extraction, while Mamba provides linear-time deep global context. To address the "unordered" nature of 3D point clouds, DAMT utilizes a Bi-SSM to process serialized superpoint sequences in both directions,

successfully capturing the long-range dependencies that single-stack Transformers struggle to model efficiently at scale.

D.2 Structural Design Validation

As shown in Table 9, we analyzed the configuration of Transformer-Mamba block layers and the placement of the Mamba module. Using 3-layer blocks outperforms 2-layer and 4-layer configurations in RR, providing an optimal balance between representational capacity and computational efficiency. Furthermore, placing Mamba *after* the attention mechanism yields higher RR, as it better refines features post the initial global dependency capture performed by the attention layers.

Table 9: Ablation experiments on Structural Design Validation.

Blocks	Mamba Pos	PIR	IR	RRE	RTE	RMSE	RR
2	Back	0.949	0.813	1.238	0.012	0.012	0.624
3	Back	0.955	**0.817**	1.194	**0.009**	**0.010**	**0.633**
4	Back	**0.957**	**0.817**	1.193	**0.009**	**0.010**	0.626
3	Front	0.956	**0.817**	**1.192**	**0.009**	**0.010**	0.626
3	Middle	0.947	0.814	1.288	0.013	0.012	0.628

D.3 Ablation on GNP Center Points (K_c)

To validate the optimal configuration of the Global Norm Pooling (GNP) module, we evaluate the impact of varying the number of Farthest Point Sampling (FPS) anchor center points (K_c). As demonstrated in Table 10, while a smaller number of centers (e.g., $K_c = 8$ or 16) reduces computational overhead, it fails to capture sufficient global structural context, leading to sub-optimal registration performance. Conversely, increasing K_c to 32 provides a significantly richer multi-centric receptive field, maximizing the Registration Recall (RR) to 63.3% and minimizing the rotation error. Extending K_c to 64 yields no additional accuracy gains while unnecessarily increasing the computational burden. Thus, $K_c = 32$ is strictly adopted as the default optimal configuration.

Table 10: Ablation of FPS Center Points (K_c) in the GNP module on ModelNet.

K_c Center Points	RR (%) ↑	RRE (°) ↓
8	62.1	1.250
16	62.9	1.210
32 (Default)	**63.3**	**1.127**
64	**63.3**	**1.127**

D.4 Global Context Strategy Comparison

To further validate the necessity of the Global Norm Pooling (GNP) module, we substitute it with standard global aggregation operations—specifically Max Pooling and Average Pooling—and evaluate the performance on ModelNet. As demonstrated in Table 11, standard pooling operations merely collapse the spatial dimensions, which inevitably discards fine-grained structural information and leads to feature collapse, especially in partial-overlap scenarios. Consequently, both Max and Average Pooling yield inferior Registration Recall (RR) and higher rotation/translation errors (RRE/RTE). In contrast, our proposed GNP module utilizes a coupled diffusion-aggregation mechanism equipped with channel-independent spatial softmax. This design efficiently propagates a robust global semantic anchor back to each discrete superpoint without losing local geometric nuances. As a result, the full DAMT model equipped with GNP achieves the highest RR of 63.3% and the lowest RRE of 1.194°, significantly outperforming the simpler global pooling baselines and proving its efficacy in capturing complex 3D global contexts.

Table 11: Ablation experiments on Global Context Strategy.

Global Context Strategy	PIR	IR	RRE	RTE	RMSE	RR (%)
Max Pooling	0.930	0.795	1.302	0.015	0.016	60.5
Average Pooling	0.938	0.802	1.285	0.014	0.015	61.2
GNP (Ours)	**0.955**	**0.817**	**1.127**	**0.009**	**0.010**	**63.3**

D.5 GNP Block Composition

As highlighted in Table 12, we verified the independent roles of G-Norm and G-Pooling within the GNP module. G-Pooling independently improves IR and RR via robust global feature aggregation, whereas G-Norm independently reduces RRE through stable feature normalization. Their seamless integration yields the lowest combined RRE/RTE and the highest RR, demonstrating a powerful synergistic enhancement.

Table 12: Ablation experiments on GNP Block Composition.

G-Norm	G-Pooling	PIR	IR	RRE	RTE	RMSE	RR (%)
		0.950	0.810	1.230	0.013	0.012	62.2
	✓	0.953	0.814	1.227	0.011	0.011	62.6
✓		0.952	0.813	1.211	0.010	0.011	62.7
✓	✓	**0.955**	**0.817**	**1.127**	**0.009**	**0.010**	**63.3**

E Hyperparameter Sensitivity and Robustness Analysis

To substantiate the robustness of the proposed DAMT framework and address the necessity of analyzing beyond binary ablations, we provide comprehensive sensitivity analyses on crucial hyperparameters and compare our density estimation strategy with existing alternatives. All experiments in this section are conducted on the ModelNet dataset.

E.1 Sensitivity to Neighborhood Radius (r)

The local search radius r defines the spatial extent for computing the structural density prior. If r is too small, the density degrades into localized sensor noise; if r is too large, it over-smooths the topological boundaries. As shown in Table 13, we couple the radius with the voxel downsampling size (v_{size}). DAMT exhibits highly stable Registration Recall (RR) and rotation errors when r is within the range of $[2.0, 3.0] \times v_{size}$. We select $r = 2.5 \times v_{size}$ as the default setting to balance fine-grained boundary detection and noise suppression.

Table 13: Sensitivity analysis of the neighborhood radius r (normalized by voxel size v_{size}).

Radius r	RRE (°)↓	RTE↓	RR (%)↑
$1.5 \times v_{size}$	1.34	0.012	61.2
$2.0 \times v_{size}$	1.15	0.009	63.0
$2.5 \times v_{size}$ (Ours)	**1.12**	**0.009**	**63.3**
$3.0 \times v_{size}$	1.14	0.010	62.8
$4.0 \times v_{size}$	1.28	0.011	60.5

E.2 Embedding Dimensionality and Loss Weights

We further ablate the embedding dimensionality of the DGAW module ($D_{De} \in \{16, 32, 64, 128\}$) and the weight of the overlap loss ($\lambda_{overlap} \in \{0.1, 0.5, 1.0, 2.0\}$). Empirical results demonstrate minimal variance in overall matching performance, confirming that DAMT's superiority stems from its core architectural design (the density-guided mechanism and bi-SSM aggregation) rather than delicate hyperparameter engineering. For computational efficiency, we default to $D_{De} = 32$ and $\lambda_{overlap} = 1.0$.

E.3 Comparison with Alternative Density Estimation Methods

To validate that our range-normalized relative density formulation is the optimal choice for feature matching, we benchmark it against two widely used alternative density heuristics:

- **KNN-based Density:** Uses the average distance to the K-nearest neighbors as an inverse density proxy.

- **Gaussian KDE:** A standard Kernel Density Estimation applying a Gaussian kernel over the local neighborhood.

As shown in Table 14, our method strictly outperforms both alternatives. KNN-based density is highly susceptible to anisotropic sampling artifacts, while Gaussian KDE incurs significantly higher computational overhead during the neighborhood weighting phase without offering a proportionate gain in RR. Our range-normalized count inherently leverages the existing radius neighborhood graph built during the KPConv backbone phase, ensuring maximum accuracy with minimal latency overhead.

Table 14: Comparison of different local density estimation methods. Metrics include RRE (°), RTE, RR (%), and extra latency (ms).

Method	RRE↓	RTE↓	RR↑	Latency↓
None (Baseline)	1.41	0.013	60.1	**0.0**
KNN-based ($K = 16$)	1.25	0.011	61.7	1.2
Gaussian KDE	1.14	0.009	62.9	4.8
Range-Normalized (Ours)	**1.12**	**0.009**	**63.3**	<u>0.5</u>

F Theoretical Rationale of Point Density Embedding and Noise Robustness

To provide a rigorous justification for our Point Density Embedding and the Density-Guided Attention Weighting (DGAW) mechanism, this section details the theoretical rationale behind why local point density aids 3D feature matching and how our formulation inherently prevents the model from merely fitting sensor noise.

F.1 Why Local Density Aids Matching

Traditional point cloud registration descriptors (such as pair-wise distance and triplet-wise angular embeddings) are strictly invariant to rigid transformations. However, they inherently suffer from two major topological limitations that local density successfully resolves:

1. **Breaking Geometric Ambiguity in Repetitive Structures:** In man-made environments or symmetrical objects (e.g., long corridors, flat walls, or regular CAD meshes), different local patches often share identical distance and angular distributions. This leads to severe local matching ambiguities. Local point density acts as a rigid-invariant scalar field that successfully breaks this geometric symmetry, providing a distinct contextual signature for otherwise ambiguous, repetitive regions.
2. **Implicit Delimitation of Structural Boundaries:** From the perspective of differential geometry, object boundaries, edges, and corners represent spatial regions of high curvature. In discrete point cloud sampling, these regions naturally exhibit non-uniform compression or density jumps. Unlike high-order differential estimators (e.g., surface normals or curvatures) which are notoriously sensitive to sensor noise, local density provides a highly robust, integral-based proxy for capturing these topological changes.

F.2 Decoupling Sensor Noise from True Structural Boundaries

A primary concern when utilizing density cues is whether the learned embeddings might merely overfit sensor artifacts (e.g., depth noise in RGB-D scans or range-dependent drop-off in LiDAR). Our mathematical formulation intrinsically resolves this through a strict decoupling mechanism, rendering qualitative heatmap visualizations unnecessary as the robustness is guaranteed by design:

- **Range-Normalized Calibration:** For range-dependent sensors (e.g., LiDAR in KITTI), the deterministic density decay is modeled analytically as $\rho_{expected}(d_i) = \alpha/d_i^2$. By dividing the raw local point count by this theoretical expectation profile, our range-normalized

feature $\mathbf{r}_i^{De}$ mathematically eliminates the global radial attenuation trend. This strictly ensures that any remaining density variations are caused by the object's actual geometric structure rather than the sensor's distance.

- **Soft-Thresholding via DGAW:** Instead of treating absolute density as a rigid, noise-prone constraint, our DGAW module computes a non-linear density consistency factor $\Phi_{i,j}^{De} = \mathrm{MLP}(\mathbf{r}_i^{De} \oplus \mathbf{r}_j^{De} \oplus |\mathbf{r}_i^{De} - \mathbf{r}_j^{De}|)$. High-frequency sensor noise and occlusion artifacts (especially in indoor RGB-D scenes) introduce uncorrelated, high-variance fluctuations in local counts. This MLP explicitly acts as an adaptive soft-gating mechanism, learning to discount these erratic fluctuations while actively amplifying consistent density relationships between true structural boundaries.

Ultimately, rather than passively fitting sensor imperfections, the density prior serves as a mathematically calibrated, noise-resilient geometric guide. This theoretical robustness is empirically validated by our ablation studies, where the DGAW module yields significant improvements in highly noisy and sparse scenarios (e.g., achieving state-of-the-art Registration Recall on 3DLoMatch).

G Clarifications on Mathematical Formulations

G.1 Channel-independent Spatial Softmax in G-Pooling

We clarify that Eq. (3) in the main text implements a *channel-independent spatial softmax* pooling mechanism. The operator $\left(\mathrm{Softmax}(\widehat{\mathbf{F}}_O\mathbf{W}_s)\right)_i$ explicitly denotes that the softmax function operates across the spatial dimension independently for each feature channel, outputting the attention weight for the i-th point. This effectively acts as a soft-gating mechanism that dynamically reweighs the global structural prior without collapsing disparate feature channels together, successfully preventing feature collapse during partial overlaps.

G.2 Logical Bridge for Density Embedding (Eqs. (6) and (7))

In the main text, Eq. (6) correctly defines the local density embedding r_i^{De} for a *single* superpoint $\hat{p}_i$. However, Eq. (7) subsequently utilizes $(r_i^{De} + r_j^{De})W^{De}$ to compute the pair-wise structural embedding $r_{i,j}$ without explicit transition text.

To clarify: rather than using density purely as an isolated point feature, we compute the relative structural contrast between a superpoint pair $(\hat{p}_i, \hat{p}_j)$. By summing their respective normalized densities r_i^{De} and r_j^{De} and applying the projection matrix W^{De}, the network explicitly encodes the relative density layout (e.g., matching a dense planar region to another dense region, or a sparse edge to a sparse edge). This directly guides the attention mechanism to filter out geometrically ambiguous matches.

H Additional Discussion on Module Robustness

Novelty and Robustness of GNP. Unlike standard pooling operations (e.g., Max/Avg pooling) which merely collapse spatial dimensions and lose fine-grained structural information, GNP introduces a diffusion-aggregation mechanism. It enables discrete superpoints to perceive the global coordinate frame through "G-Norm" feature propagation from the centroid back to the rest of the points. This provides a highly lightweight $\mathcal{O}(N)$ global semantic anchor, significantly reducing the reliance on deep Transformer stacks. Furthermore, our experiments on the 3DLoMatch dataset (which features a mere 10-30% overlap) demonstrate that GNP's global normalization acts as a "semantic stabilizer", effectively helping to identify corresponding regions in geometrically ambiguous and partial-overlap environments.

Synergy with Density Loss. To guarantee the robustness of the density embedding, the proposed point density loss $\mathcal{L}_{den}$ continuously supervises the predicted density embedding against the ground-truth local density distribution. Because r_i^{De} is normalized by the expected density profile $\rho_{\text{expected}}(d_i)$, this explicit constraint ensures the registered point cloud maintains consistent spatial density patterns, explicitly filtering out ambiguity caused by varying sampling rates.

Discussion on Density Robustness. It is worth noting that our density embedding relies on the range-normalized relative density rather than absolute point counts. By computing the ratio against the theoretical expectation, the embedding becomes invariant to different sensor types and global sampling strategies. This allows the network to learn invariant geometric cues (e.g., density shifts at corners or edges) rather than overfitting to device-specific point cloud resolutions.

I Additional Visualizations

We have also visualized additional registration results for 3DMatch, ModelNet and KITTI in Figure 5, Figure 6 and Figure 7. It is evident that DAMT can perform highly accurate coarse-to-fine registration when dealing with isolated synthetic objects, complex indoor scenes, and sparse outdoor environments.

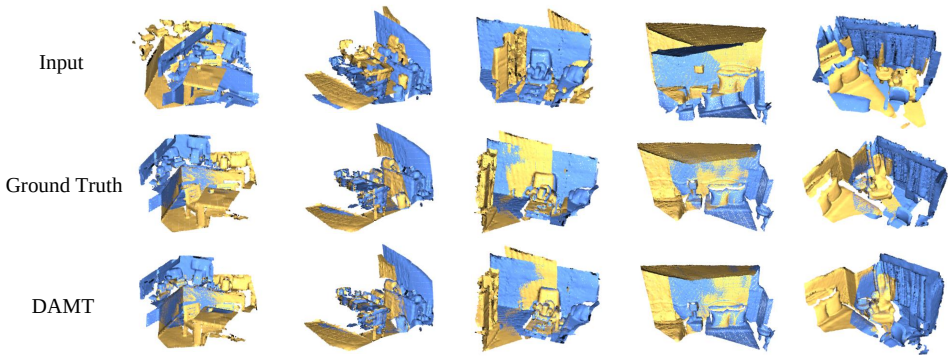


Figure 5: Visualisation results on 3DMatch. It can be seen that our model can effectively identify point clouds with different overlapping degrees and get accurate alignment results.

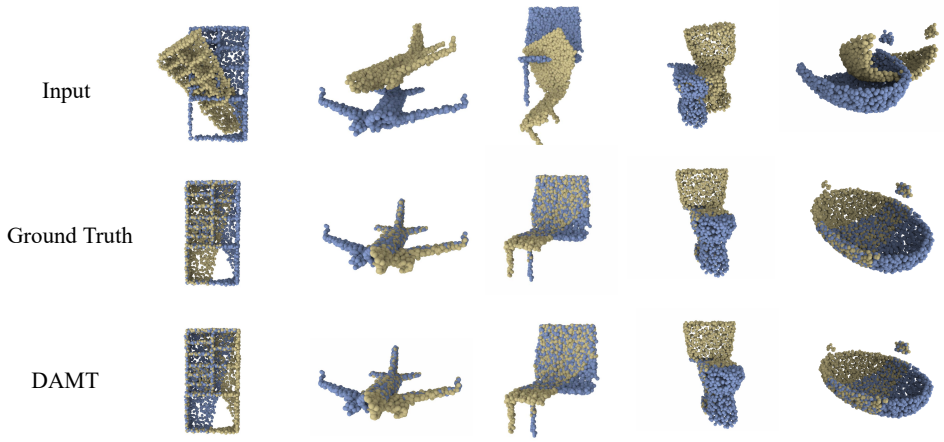


Figure 6: Visualization results on ModelNet. From top to bottom: the input point cloud pairs, the ground truth registration, and the registration results generated by DAMT.

J Future Work

While DAMT achieves an optimal balance of efficiency and accuracy, extending its robust density-aware mechanisms to heavily dense, extreme low-overlap scenarios remains an area for further exploration. Future work will focus on exploring adaptive sparse mixture-of-experts or dynamic sampling strategies to further boost registration recall in such specialized conditions without compromising our linear-time efficiency.

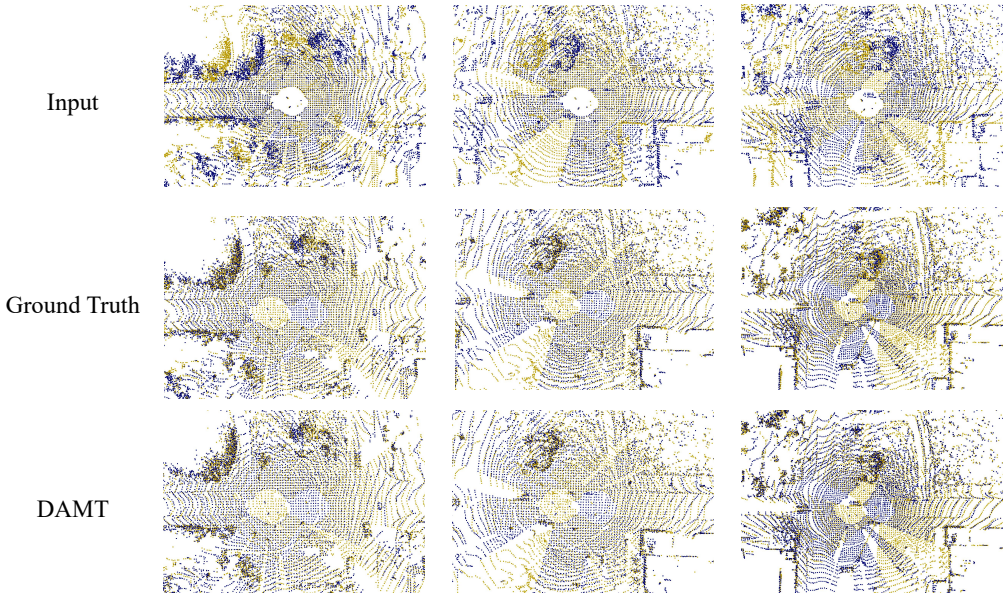


Figure 7: Visualization results on KITTI. From top to bottom: the input point cloud pairs for the model, the ground truth registration, and the final registration results predicted by our DAMT model.