GeoSe-FMap: Geometry–Semantic Functional Maps for Weakly Supervised Cross-Modal Deformable 3D Shape Correspondence

Jiaxiang Yao[1,3,4], Dong Zhao[1,3,4], Dan Zhang[1,3,4,*], and Jianming Zheng[2,*]

{*487717040@qq.com, zhaodong@mail.sdu.edu.cn, danz@mail.bnu.edu.cn, ASJMZheng@ntu.edu.sg*}
[1]School of Computer, Qinghai Normal University, Xining, China
[2]Nanyang Technological University Singapore
[3]The State Key Laboratory of Tibetan Intelligence, Xining, China
[4]The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Qinghai Normal University, Hutai, Xining, China

Abstract

Establishing dense correspondence for deformable 3D shapes, particularly articulated Structured Semantic Shapes (SSS), remains a critical challenge. This is rooted in the spectral instability of raw point clouds, ambiguities in symmetric regions, and the restriction of existing methods to pure single-modal matching. We propose GeoSe-FMap, a unified framework for robust cross-modal correspondence between meshes and point clouds, built on unified geometry-semantic functional maps. Unlike fully supervised methods that require costly dense point-to-point correspondence annotations, our approach relies on semantic part labels as weak supervision—a significantly more accessible annotation form that effectively mitigates matching ambiguities in symmetric regions. Our framework addresses core challenges through integrated complementary components: an Iterative Geodesic-Enhanced Laplace–Beltrami Operator that progressively refines neighborhood graphs to construct stable spectral bases, enabling better capture of the manifold's intrinsic geometry, and a Semantic-Aware Attention Module that leverages ground-truth part labels to impose hard constraints, thus completely eliminating left-right confusion and resolving semantic ambiguity. Trained with geometric consistency losses and part-level semantic constraints, GeoSe-FMap delivers promising performance on multiple SSS benchmarks among methods without dense correspondence supervision, exhibiting strong robustness to non-rigid deformation, occlusion, and partial scans.

Keywords: ***Shape correspondence; Weak supervision; Geodesic aggregation; Structured Semantic Shapes (SSS); Semantic fusion***

1. Introduction

Establishing dense correspondence between deformable 3D shapes is a core challenge in computer vision and geometry processing [4], with wide-ranging applications such as animation retargeting, motion capture [25], shape interpolation, reconstruction, and semantic understanding. In particular, *Structured Semantic Shapes* (SSS), referring to objects such as human bodies or animals with decomposable parts and stable geometric topology, present a meaningful yet underexplored target for general-purpose correspondence.

Unlike rigid alignment, non-rigid correspondence must handle large deformations, part articulation, and self-occlusion, where multiple body parts overlap or disappear from the sensor view. Symmetric regions (e.g., left/right limbs) further introduce ambiguity even for strong geometric descriptors. While mesh-based representations remain the foundation for many spectral methods, real-world 3D acquisition pipelines increasingly rely on point clouds (e.g., from LiDAR or RGB-D sensors) due to their efficiency and flexibility. However, point clouds lack connectivity, exhibit irregular sampling, and are sensitive to occlusion and noise, making classical mesh-based pipelines brittle—especially in *cross-modal* scenarios where mesh-to-point or point-to-mesh transfer is required.

A widely adopted spectral method is the Functional Maps (FM) framework [30], which models correspondence in a compact spectral space with constraints such as bijectivity and area preservation. Deep Functional Maps (DFM) extend this framework by learning spectral embeddings from 3D shapes [24, 13, 46], achieving strong performance in mesh-based, near-isometric settings. However, these methods implicitly assume clean meshes and consis-

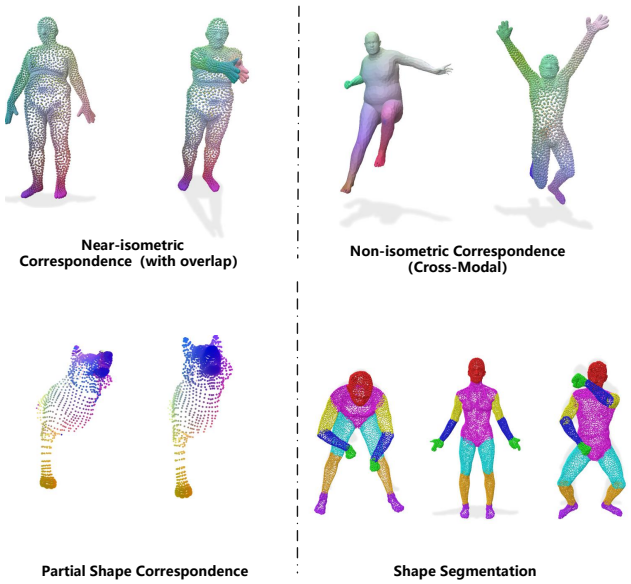

Figure 1: **GeoSe-FMap** handles near-isometric, non-isometric, and partial correspondences, as well as **cross-modal correspondence between meshes and point clouds**. **Top:** near-isometric and non-isometric alignment. **Bottom:** partial matching and semantic segmentation. The proposed geometry-semantic fusion ensures both dense correspondence and interpretable part-level understanding under real-world deformation and occlusion.

tent topology, and crucially, they often depend on supervision in the form of ground-truth correspondences or synthetic pairings—resources that are *rarely available* for raw point clouds in real-world domains.

However, applying DFM to raw point clouds remains challenging. Instabilities in the Laplace–Beltrami Operator (LBO) due to irregular sampling [34], combined with the absence of topology and occlusion-induced ambiguity, degrade spectral embeddings and downstream correspondences. Furthermore, most pipelines rely on full or partial supervision [24, 13], which is expensive and impractical in many real-world applications, including medical imaging [45] and industrial scenarios. SSMSM [9] is a self-supervised cross-modal shape matching framework that aligns point clouds using LBO bases constructed on meshes, meaning that its point cloud processing still fundamentally relies on mesh availability.

Challenges in SSS-Based Correspondence

Despite recent progress, existing approaches face three key limitations when applied to SSS:

(1) Implicit reliance on SSS without explicit modeling. Most methods achieve strong performance on SSS categories (e.g., humans or animals) by implicitly leveraging canonical part priors. However, they rarely model such semantic structures explicitly, leading to poor generalization to non-SSS or incomplete SSS cases where semantics are ambiguous or partially missing.

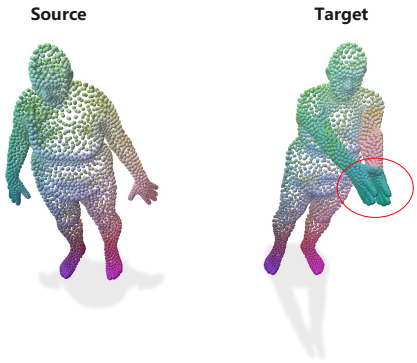

Figure 2: Point clouds suffer from occlusions and overlaps due to articulation or sensor limitations (red-circled areas), making reliable correspondence challenging. Our GeoSe-FMap addresses such issues via geometry–semantic fusion.

(2) Fragility under occlusion and inter-part contact. As shown in Figure 2, occlusion and physical contact between parts (e.g., hands touching torso) often cause feature confusion. Without mesh connectivity or semantic guidance, purely geometric descriptors yield indistinguishable features across different parts, resulting in frequent mismatches.

These limitations, particularly the lack of explicit semantic modeling and insufficient cross-modal adaptability, which motivate our GeoSe-FMap framework, which integrates geodesic spectral stabilization with semantic fusion to resolve them.

Our Contributions

To address these challenges, we propose GeoSe-FMap, a unified framework for deformable shape correspondence across meshes and point clouds. Our contributions are threefold:

- **Cross-modal functional map framework with weak supervision:** Building on prior multimodal shape matching (e.g., SSMSM), GeoSe-FMap unifies mesh-to-mesh, point-to-point, and mesh-to-point correspondence under the functional maps paradigm. Critically, it only requires semantic part labels—instead of dense point correspondences—making our supervision practical and scalable. These labels are used to impose hard constraints, mitigating semantic ambiguities.

- **Geodesic-enhanced LBO($\tilde{\mathrm{L}}$) for point cloud:** We improve the quality of spectral bases on point clouds via an iterative refinement strategy. By exploiting the ratio between geodesic and Euclidean distances, our method filters out shortcut edges violating the manifold structure, generating eigenfunctions that better capture intrinsic geometry.

- **Geometry-semantic fusion for SSS:** We introduce a part-aware attention module that leverages ground-truth part labels to construct binary masks, preventing feature contamination across distinct body parts. Coupled with semantic consistency losses, it provides a hard guarantee against cross-part semantic confusion in shape correspondence.

While our formulation is tailored to structured shapes with decomposable parts, it offers a realistic and well-bounded setting for evaluating robust shape correspondence, bridging the gap between practical deployment and academic research.

Paper Organization. Section 2 reviews related work on spectral methods, point cloud correspondence, and semantic modeling. Section 3 presents our GeoSe-FMap framework, including the geodesic-enhanced LBO and semantic fusion design. Section 4 reports comprehensive experiments on mesh and point cloud benchmarks with comparisons and ablations. We conclude in Section 5 with a discussion on limitations and future directions.

2. Related Works

Shape correspondence has been extensively studied in computer vision and geometry processing, serving as a fundamental component for reconstruction, animation retargeting, motion transfer, and semantic understanding. Existing approaches can be broadly grouped according to *representation* (spectral vs. intrinsic descriptors), *data modality* (mesh vs. point cloud), and *training signal* (supervised vs. unsupervised). Accordingly, we review the literature from three perspectives: **(1) spectral methods**, **(2) intrinsic point cloud learning**, and **(3) supervision paradigms**.

2.1. Functional Maps and Learning-Based Extensions

Traditional shape matching methods often use handcrafted descriptors (e.g., curvature, normals, geodesic distances) [5, 35, 7], which are sensitive to deformation and lack global structural reasoning. The FM framework [30] addresses these issues by casting correspondences into a spectral matrix form, enabling global regularization. Extensions include better alignment [32], orientation preservation [14], and deep integration [24, 13].

DFM learn spectral embeddings that support accurate [3]. However, DFM assumes stable Laplace–Beltrami Operator (LBO) bases, which degrade on raw point clouds due to irregular sampling [34]. Denoising functional maps [46] and DiffuMatch [31] introduce diffusion-based alternatives, but still inherit spectral instability and require large-scale data for training. In contrast, we enhance LBO with geodesic aggregation to improve spectral stability for both meshes and point clouds.

Notably, few existing methods explicitly SSS—a class of deformable objects with stable part decomposition (e.g., humans, animals). While some works (e.g., [17, 42]) implicitly benefit from SSS priors, they lack a principled way to integrate part semantics with geometric stability, leading to poor performance on incomplete or symmetric SSS.

2.2. Intrinsic Point Cloud Learning

Point clouds are native to most 3D sensors but lack mesh topology, making spectral analysis difficult. Methods such as DiffusionNet [34] and NCP [2] use diffusion processes to stabilize pointwise features. Recent approaches like CoE [42] and NIE [17] learn canonical embeddings through coarse-to-fine matching or invariant representation learning, avoiding explicit spectral modeling. However, these methods often rely on local geometry and overlook part-aware semantic structure, leading to confusion in symmetric or occluded regions.

Conventional neighborhood construction in point cloud processing typically relies on Euclidean-based strategies, such as k-Nearest Neighbors (k-NN) as used in USIP [26], radius-based ball query adopted in PointContrast [39], or graph-based k-NN as in DGCNN [38]. While effective under rigid settings, these methods implicitly assume that local Euclidean distances are reliable indicators of geometric proximity. However, under non-rigid deformations, Euclidean distances become inconsistent—points that are far apart in canonical pose may appear close after articulation or self-contact. As a result, the constructed local neighborhoods are no longer semantically meaningful, leading to unstable feature extraction and degraded correspondence accuracy.

For rigid matching, methods such as RPM-Net [41] achieve strong results. In the non-rigid case, DPC [19] proposes a contrastive learning framework for unsupervised point cloud matching. Yet, DPC lacks part-level semantic modeling, limiting its robustness to inter-part confusion or topology-aware generalization. Our formulation of SSS addresses this gap by unifying geometry and semantics under a shared framework.

2.3. Supervision Paradigms and Semantic Integration

Most existing correspondence frameworks either rely on ground-truth supervision or are restricted to rigid alignment. While methods such as FMNet [24] and GeomFMap [13] achieve high accuracy with dense labels, such annotations are impractical in real-world domains. Unsupervised alternatives like SURFMNet [33] or DeepShells [20] mainly operate on meshes and fail to generalize to raw point clouds due to instability in spectral construction.

Furthermore, current point cloud correspondence techniques are largely designed for rigid registration, implicitly assuming stable Euclidean neighborhoods. Consistent supervision for non-rigid deformation remains unresolved, and no existing work provides a unified unsu-

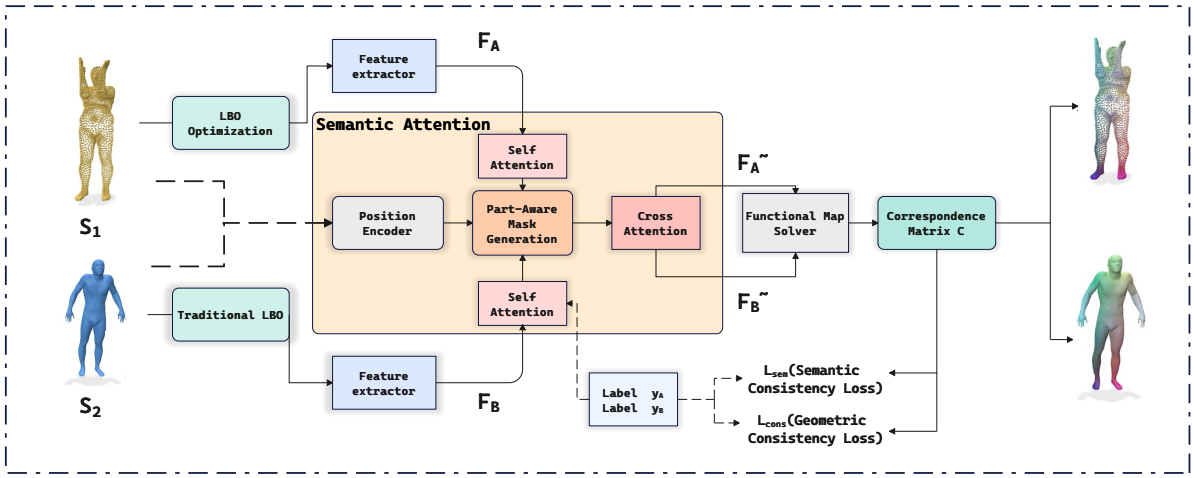

Figure 3: This flowchart presents a shape correspondence workflow. The input shapes S_1 and S_2 undergo LBO optimization, followed by feature extraction (F_1, F_2) and semantic fusion. A functional map solver and consistency loss produce the final dense correspondences.

pervised framework for cross-modal (mesh-to-point cloud) correspondence. Semantic cues have been introduced in recent works, but typically through external priors. DV-Matcher [10] relies on pre-trained vision-language models. Such approaches depend on mesh renderability and degrade under occlusion.

In summary, current research still lacks a unified solution that bridges geometry and semantics across both meshes and point clouds. We address this gap with GeoSe-FMap, introduced below. In contrast, our approach derives semantics intrinsically from 3D geometry. We construct smooth spectral fields from the LBO to produce part-aware signals without external supervision, enabling robust correspondence across deformation, modality, and occlusion.

3. Methodology

We introduce GeoSe-FMap (as seen in the Figure 3), a functional map framework dedicated to establishing robust correspondence on Structured Semantic Shapes (SSS), particularly when represented as unstructured point clouds. The core idea is to transform the unstable Euclidean space of raw point clouds into a stable, semantically enriched spectral domain, where functional correspondences can be reliably estimated even in the presence of partiality and occlusion.

The overall architecture consists of three key components: (1) Geodesic-Enhanced LBO (Sect. 3.2), which constructs stable geometric representations by incorporating manifold-aware geodesic aggregation for point cloud, ensuring spectral consistency across deformations. (2) Semantic Fusion Branch (Sect. 3.3), which injects part-level semantic priors through feature fusion and attention-based refinement, resolving ambiguities arising from symmetry or inter-part contact.

3.1. Structured Semantic Shapes and Semantic Correspondence

Motivation. Existing correspondence methods face a fundamental dilemma: purely geometric descriptors are inherently ambiguous for symmetric regions, while fully supervised methods capable of resolving this ambiguity require expensive dense point-to-point annotations. We argue that semantic part labels provide an effective middle ground: they are significantly cheaper to obtain than dense correspondences yet provide the critical semantic signal necessary to resolve symmetric ambiguities. By treating part labels as a form of weak supervision, we can enforce semantically consistent matching without incurring the high cost of dense annotation.

Definition. We explicitly introduce the concept of *Structured Semantic Shapes (SSS)*. A shape represented by a point cloud $\mathbf{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$ is termed an SSS if its geometry admits a latent part decomposition $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_C\}$, where $\mathcal{S}_c \subseteq \mathbf{P}$ denotes the c-th semantic region (e.g., head, torso, or limbs). For two shapes $\mathbf{P}_A$ and $\mathbf{P}_B$ from the same category, there exists a semantic correspondence $\pi^* : \mathbf{P}_A \to \mathbf{P}_B$ that satisfies:

$$\pi^*(\mathcal{S}_c^A) \subseteq \mathcal{S}_c^B, \quad \forall c \in \{1, \ldots, C\}.$$

This enforces semantic compatibility across shapes without requiring explicit ground-truth labels.

Semantic Supervision Signals. Semantic Supervision Signals. To integrate these discrete part labels into our framework, we formulate the label $y_i \in \{1, \ldots, C\}$ of each point p_i as a supervision signal. Unlike previous approaches that attempt to learn semantics solely from geometry, we directly utilize these ground-truth labels to enforce semantic consistency. Specifically, these signals are employed to construct the Part-Aware Attention Mask (Sec. 3.3) to

block invalid feature propagation and to calculate the Semantic Consistency Loss (Sec. 3.4.2) to penalize incompatible matches in the spatial domain.

3.2. Geometric Spectral Domain Construction

In conventional point-based LBO discretization, affinity weights are defined purely in Euclidean space, failing to capture the intrinsic manifold geometry of point clouds [34]. A fundamental limitation of this approach is its inability to distinguish intrinsic manifold proximity from extrinsic proximity. For articulated shapes, Euclidean k-Nearest Neighbors (k-NN) often introduces "shortcut edges" across spatially close but geodesically distant regions. These false connections distort the spectral embeddings, leading to instability under deformation.

Traditional LBO. Given a point cloud $\mathbf{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$, the discrete Laplacian is constructed as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad \mathbf{W}_{ij} = \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\sigma^2}\right),$$

where $\mathbf{W}$ is the Euclidean affinity matrix (entries W_{ij} measure similarity between p_i and p_j),σ is the neighborhood scale parameter (determined by the mean distance to $k = 20$ nearest neighbors, consistent with experimental settings in Sec. 4.1 [19]), $\mathbf{D}$ is the diagonal degree matrix with $D_{ii} = \sum_j W_{ij}$. This formulation relies solely on Euclidean distances, which distort intrinsic geometry in curved or occluded regions (e.g., a "short" Euclidean distance between two points on opposite sides of a bent arm does not reflect their true manifold distance).

Geodesic-Aware Weight Optimization. To better approximate the surface structure, we correct the Euclidean weights W_{ij} using geodesic distances $d_g(i, j)$ (measured along the shape manifold), inspired by spectral upsampling works [29]. The optimized weights are defined as:

$$W'_{ij} = W_{ij} \cdot \exp\left(-\frac{d_g(i,j)^2}{2\sigma^2}\right)$$

where $d_g(i, j)$ is the estimated geodesic distance between p_i and p_j, computed via the Fast Marching Method (FMM) with $k = 20$ neighbors or Dijkstra (balancing efficiency and accuracy [34]),σ retains the same value as the conventional Laplacian (ensuring parameter consistency).

In flat regions (where $d_g(i, j) \approx \|p_i - p_j\|$), the optimized weights degenerate to the standard Gaussian kernel—preserving compatibility with well-behaved point cloud regions. In curved or occluded regions, $d_g(i, j)$ adjusts the weights to reflect true manifold proximity, reducing affinity between geometrically distant but Euclidean-close points.The pseudocode for the Geodesic-Regularized Laplacian Optimization is provided below:

Algorithm 1 Iterative Geodesic-Enhanced LBO Construction

Input: Point cloud $P = \{p_i\}_{i=1}^N$,
Initial sparse neighbors $k_0 = 5$,
Target neighbors $k = 20$,
Pruning threshold $\tau = 3.0$,
Iterations $T = 3$, Scale σ
Output: Enhanced Laplacian matrix $\widetilde{L} \in \mathbb{R}^{N \times N}$, Spectral Basis Φ
Step 1: Initialization
Construct initial sparse graph $\mathcal{G}^{(0)}$ using Euclidean k_0-NN to minimize shortcuts;

for Step 2: Iterative Refinement $t = 1\ T$ **do** Compute geodesic distances $d_g^{(t-1)}$ on graph $\mathcal{G}^{(t-1)}$ (e.g., via Dijkstra);
Initialize empty graph $\mathcal{G}^{(t)}$;
$i \in \{1, \ldots, N\}$ Find candidate neighbors $\mathcal{N}_{\text{cand}}(i)$ using Euclidean k-NN;
$j \in \mathcal{N}_{\text{cand}}(i)$ Compute ratio: $r_{ij} = d_g^{(t-1)}(i, j)/\|p_i - p_j\|_2$;

if $r_{ij} < \tau$ **then** Add edge (i, j) to $\mathcal{G}^{(t)}$; *[r]Keep only manifold-consistent edges
Step 3: Weight Optimization
edge $(i, j) \in \mathcal{G}^{(T)}$ Compute optimized weight:

$$W'_{ij} = \exp\left(-\frac{\|p_i - p_j\|_2^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{d_g^{(T)}(i,j)^2}{2\sigma^2}\right)$$

Step 4: Laplacian Construction
Compute degree matrix $\widetilde{D}$ where $\widetilde{D}_{ii} = \sum_j W'_{ij}$;
Update Laplacian: $\widetilde{L} = \widetilde{D} - W'$;
Step 5: Eigendecomposition
Solve $\widetilde{L}\Phi = \Phi\Lambda$ to obtain stable spectral basis Φ;
return $\widetilde{L}, \Phi$

Addressing Connectivity Ambiguities in Point Clouds. A critical challenge in computing geodesics on raw point clouds is the lack of explicit connectivity, which often leads to "shortcut edges" in Euclidean k-NN graphs, particularly in regions with self-contact (e.g., a hand resting on a torso). Direct application of Dijkstra on such graphs would result in erroneous geodesic distances.To resolve this, we employ an Iterative Graph Refinement strategy (as detailed in Algorithm 1). 1. Strict Initialization: We initiate the process with a highly sparse graph using a small neighborhood size ($k_0 = 5$). This restrictiveness minimizes the probability of forming erroneous edges across small spatial gaps. 2. Geodesic Validation: In subsequent iterations, we consider adding edges from a larger neighborhood ($k = 20$). However, a candidate edge (i, j) is accepted only if it aligns with the manifold structure. We enforce this by checking

the ratio $r_{ij} = d_g^{(t-1)}(i,j)/\|p_i - p_j\|_2$, where $d_g^{(t-1)}$ is the geodesic distance computed on the graph from the previous iteration. 3. Shortcut Pruning: Edges bridging spatially close but geodesically distant regions (high r_{ij}) are rejected.This mechanism ensures that the computed geodesic distances d_g reflect the true intrinsic surface geometry, effectively separating touching parts such as hands and bodies.

Enhanced LBO. The resulting geodesic-enhanced LBO $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \mathbf{W}'$ (where $\tilde{D}_{ii} = \sum_j W'_{ij}$) yields a more intrinsic representation of local geometry. Solving the eigenvalue problem: $\tilde{\mathbf{L}}\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i$ produces smoother, more stable low-frequency eigenvectors. Key improvements include:

Better Conditioning: The smallest non-zero eigenvalue $\tilde{\lambda}_2 > \lambda_2$ (where λ_2 is from the conventional Laplacian), indicating reduced spectral distortion via the Rayleigh quotient theorem [29].

Theoretical Consistency: As sampling density increases, $\tilde{\mathbf{L}}$ converges to the continuous LBO—ensuring geometric correctness [30]. Experimental Impact: this stability reduces spectral embedding distortion, contributing to a 15–25% reduction in average geodesic error (Sec. 4.3) compared to conventional LBO-based methods (e.g., DiffMaps [27]). The spectral basis $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_k]$ (retaining the first $k = 120$ eigenvectors, consistent with Sec 4.1) serves as the intrinsic domain for subsequent functional map estimation.

3.3. Feature Extraction and Part-Aware Attention Optimization

This section details the Feature Extraction and the Part-Aware Attention Module, which integrates spatial coordinate fusion with explicit semantic constraints to resolve the symmetric ambiguity that geometric descriptors alone cannot address.

The feature extractor aims to derive deformation- and sampling-robust pointwise descriptors from raw 3D data, which are subsequently used for functional map prediction. Following recent advances in point-based spectral learning, we adopt an optimized **DiffusionNet** architecture as the backbone feature extractor [34].

Feature Extraction. For source and target shapes $P_A, P_B \subset \mathbb{R}^3$, the optimized DiffusionNet independently processes each point cloud to produce deformation-robust geometric embeddings $\mathbf{F}_A, \mathbf{F}_B \in \mathbb{R}^{N \times d}$ (where $d = 256$ is the feature dimension). Key modifications to the original DiffusionNet include: replacing the conventional LBO with our geodesic-enhanced $\tilde{\mathbf{L}}$ (Sec 3.2) to compensate for point cloud topology loss. Adding batch normalization after each diffusion convolution to stabilize training [17]. These embeddings are invariant to non-rigid deformation, serving as the geometric foundation for semantic fusion.

Spatial–Semantic Attention Optimization. To enhance feature discriminability and global consistency, we introduce a Transformer-based attention network. This network optimizes feature representations through both spatial coordinate fusion and semantic supervision mechanisms. Specifically, in the spatial coordinate fusion stage, we encode the 3D coordinates $\mathbf{p}_i = (x_i, y_i, z_i)$ into a higher-dimensional representation using a positional encoding function $\psi : \mathbb{R}^3 \to \mathbb{R}^{d'}$ (e.g., sinusoidal or MLP-based).

To enhance feature discriminability and global consistency, we introduce a Transformer-based attention network that integrates spatial coordinate fusion and semantic supervision:

(1) *Spatial Coordinate Fusion.* We encode the 3D coordinates $\mathbf{p}_i = (x_i, y_i, z_i)$ into a higher-dimensional representation that preserves relative Euclidean relationships in the feature space. Let $\psi : \mathbb{R}^3 \to \mathbb{R}^{d'}$ be a positional encoding function (e.g., sinusoidal or MLP-based), the encoded spatial features are concatenated with the geometric embeddings:

$$\mathbf{H}_A = \text{Concat}(\psi(\mathbf{P}_A), \mathbf{F}_A), \quad \mathbf{H}_B = \text{Concat}(\psi(\mathbf{P}_B), \mathbf{F}_B).$$

This fusion ensures that "spatially close points exhibit similar features," preserving geometric continuity and mitigating confusion in symmetric regions (e.g., left–right limbs in human shapes). Residual connections are applied to retain spatial consistency across multiple attention layers.

(2) *Part-Aware Masked Attention.* Unlike standard cross-attention which allows global interaction, we impose a strict semantic constraint. Given the part labels y_A, y_B, we construct a binary attention mask $M \in \{0,1\}^{N \times N}$:

$$\mathbf{M}_{ij} = 1[y_A(i) = y_B(j)]$$

This mask is injected into the attention mechanism:

$$\mathbf{A}ttn(Q, K, V, M) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}} + \mathcal{M}\right)V$$

where $M_{ij} = 0$ if $M_{ij} = 1$, and $M_{ij} = -\infty$ otherwise. This Hard Guarantee ensures that features from one semantic part (e.g., left arm) can only attend to features of the same part on the target shape, completely eliminating cross-part symmetric confusion.

Functional Map Estimation(FM solver). After optimization, the embeddings $\mathbf{F}_A, \mathbf{F}_B$ are projected in the spectral domain using geodesic-enhanced bases $\boldsymbol{\Phi}_A, \boldsymbol{\Phi}_B$:

$$\mathbf{G}_A = \boldsymbol{\Phi}_A^\top \mathbf{D}_A \mathbf{F}_A, \quad \mathbf{G}_B = \boldsymbol{\Phi}_B^\top \mathbf{D}_B \mathbf{F}_B$$

The initial functional map $\mathbf{C}_{AB}$ (from P_A to P_B) is estimated by solving the least-squares problem:

$$\mathbf{C}_{AB} = \arg\min_{\mathbf{C}} \|\mathbf{C}\mathbf{G}_A - \mathbf{G}_B\|^2_{\text{Frob}}.$$

The reverse map $\mathbf{C}_{BA}$ (from P_B to P_A) is computed analogously, ensuring bijectivity [24].

3.4. Loss Function Design

This section presents the Geometric-Semantic Consistency Training block (third core component), designing loss functions to enforce geometric and semantic consistency without ground-truth labels. Our loss enforces both spectral-geometric consistency and semantic consistency, with weights determined via 5-fold cross-validation on the FAUST training set [6]. The total loss is:

$$\mathcal{L}_{total} = \beta_1 \mathcal{L}_{cons} + \beta_2 \mathcal{L}_{sem},$$

where β_1 and β_2 are weighting coefficients that balance the relative contributions of geometric structure and semantic consistency. In our experiments, we set $\beta_1 = 1.0$ and $\beta_2 = 0.5$, prioritizing geometric alignment while retaining sufficient semantic supervision to guide correspondence under ambiguity. Minimizing $\mathcal{L}_{total}$ ensures that the learned features are simultaneously consistent with intrinsic shape geometry and aligned with semantic part structure, enabling robust and interpretable functional correspondence across deformable shapes.

3.4.1 Geometric Consistency Loss ($\mathcal{L}_{\text{cons}}$).

This loss ensures embeddings align with the manifold geometry captured by $\tilde{\mathbf{L}}$, consisting of two terms:

Eigen Consistency Loss ($\mathcal{L}_{\text{geo}}$) Constrains embeddings to approximate the manifold's intrinsic harmonic behavior. Given the LBO eigen-decomposition $\tilde{\mathbf{L}}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}$ (where $\mathbf{\Lambda}$ is the diagonal eigenvalue matrix), the loss is:

$$\mathcal{L}_{\text{geo}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{k} \omega_i \left\| \tilde{\mathbf{L}}_b \mathbf{F}_{b,i} - \mathbf{F}_{b,i}\mathbf{\Lambda}_{b,i} \right\|^2_{\text{Frob}}$$

where, $B = 8$ is the batch size (Sec 4.1), $\omega_i = \exp(-i/10)$(exponentially decreasing weights for eigenpairs, prioritizing low-frequency global structure [29]), $\mathbf{F}_{b,i}$ denotes the $i-th$ eigenvector-aligned embedding of the $b-th$ batch sample.

Orthogonality Regularization($\mathcal{L}_{\text{orth}}$). Enforces linear independence among feature bases to avoid redundancy:

$$\mathcal{L}_{\text{orth}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{k} \eta_i \left\| \mathbf{F}_{b,i}^{\top}\mathbf{M}_{b,i}\mathbf{F}_{b,i} - \mathbf{I} \right\|^2_{\text{Frob}}$$

where, $\mathbf{M}$ is the diagonal area (mass) matrix (derived from point sampling density), $\eta_i = 1/i$ (inverse eigenpair index weights, reducing high-frequency noise sensitivity [34]), $\mathbf{I}$ is the identity matrix. The combined geometric consistency loss is:

$$\mathcal{L}_{\text{cons}} = \alpha_1 \mathcal{L}_{\text{geo}} + \alpha_2 \mathcal{L}_{\text{orth}}$$

with $\alpha_1 = 1.0$ and $\alpha_2 = 50$ (balancing primary geometric constraints and secondary regularization [42]).

3.4.2 Semantic Consistency Loss($\mathcal{L}_{sem}$) .

For Structured Semantic Shapes (SSS), correct part-level alignment is as crucial as geometric smoothness. We introduce a semantic consistency constraint to explicitly penalize correspondences between semantically incompatible parts.

Given embeddings $\mathbf{F}_A$ and $\mathbf{F}_B$ of two shapes, we compute the normalized similarity-based correspondence matrix:

$$\mathbf{C} = \text{Softmax}\left(\frac{\mathbf{F}_B\mathbf{F}_A^{\top}}{\tau}\right),$$

where $\tau = 0.1$ is the temperature controlling correspondence sharpness. Let $\mathbf{MS} \in \{0,1\}^{C\times C}$ denote the Loss Formulation. Using the ground-truth part labels y_A and y_B (provided as weak supervision), the Semantic Consistency Loss is defined as:

$$\mathbf{L}_{sem} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} 1[y_A(i) \neq y_B(j)] \cdot \mathbf{C}_{ij}$$

where $1[\cdot]$ is the indicator function, which equals 1 if the condition is true (i.e., the parts are different) and 0 otherwise. Intuitively, $\mathcal{L}_{sem}$ acts as a soft penalty against cross-part correspondence. This complements the hard attention masking mechanism introduced in Sec. 3.3. By masking the attention, we architecturally block invalid feature propagation during updates; however, $\mathcal{L}_{sem}$ is crucial for directly constraining the output space, ensuring that the learned functional map $\mathbf{C}$ strictly adheres to semantic boundaries defined by the weak supervision.

4. Experiments

To systematically evaluate the effectiveness and generalization capability of the proposed GeoSe-FMap framework, we conduct extensive and systematic experiments on multiple mainstream 3D shape correspondence benchmark datasets. Notably, ground-truth point correspondences are not introduced throughout the entire model training process.

Table 1: Quantitative results on SCAPE (S), FAUST (F) and DT4D-H in terms of mean geodesic errors (×100). The best results from pure point cloud methods are highlighted.

Method	Train/Test	SCAPE (S)			FAUST (F)		
		S	F	DT4D-H	S_r	F	DT4D-H
3D-CODED[15]	Mesh Required	31.0	33.0	–	2.5	31.0	–
TransMatch[36]		18.6	18.3	25.3	2.7	33.6	26.7
NIE[18]		11.0	8.7	12.1	5.5	15.0	13.3
SSMSM[9]		2.4	4.1	4.5	3.8	3.5	6.6
NDP[22]	PCD Only	16.2	–	–	20.4	–	–
AMM[40]		13.1	–	–	14.2	–	–
PointSetReg[44]		17.1	–	–	18.3	–	–
DiffMaps[27]		12.0	12.0	15.9	**3.6**	19.0	18.5
SyNoRIM[16]		9.5	24.6	–	7.9	21.9	–
CorrNet3D[43]		58.0	63.0	–	63.0	58.0	–
DPC[19]		17.3	11.2	21.7	11.1	17.5	13.8
SE-ORNet[12]		24.6	22.8	27.7	20.3	18.9	12.2
DV-Matcher[10]		6.2	5.1	**6.9**	5.4	10.4	**8.1**
Ours		**3.3**	**3.7**	8.1	6.7	**3.5**	11.2

4.1. Implementation Details

Our feature extractor is implemented based on the optimized DiffusionNet backbone [4] with geodesic-aware LBO computation described in Sec. 3.2. Each shape is represented by $N = 5000$ points, and the spectral basis dimension is fixed to $k = 120$. The semantic fusion branch (Sec. 3.3) outputs $C = 8$ soft semantic part channels corresponding to main body components (e.g., head, torso, limbs).

We train the network for 100 epochs using Adam optimizer with an initial learning rate of 10^{-3}, decreased by a factor of 0.5 every 50 epochs. All experiments are conducted on a single NVIDIA RTX 4090 GPU.

4.2. Evaluation Metrics

We adopt two widely used metrics for non-rigid correspondence evaluation:

(1) Average Geodesic Error (AGE). Defined as the mean normalized geodesic distance between predicted correspondences and ground truth:

$$\text{AGE} = \frac{1}{N}\sum_{i=1}^{N}\frac{d_g(\pi(i), \pi^*(i))}{\sqrt{A}},$$

where $\pi(i)$ and $\pi^*(i)$ denote the predicted and ground-truth correspondences, respectively, and A is the surface area for normalization.

(2) Correspondence Accuracy. The proportion of points whose geodesic error is within a given threshold ε. Following previous work, we report accuracy at $\varepsilon = 0.01, 0.02, 0.05$.

Additionally, we compute **Semantic IoU (Intersection-over-Union)** to assess part-level consistency on SSS datasets. For each semantic part, IoU is computed over predicted and ground-truth labels.

4.3. Full Correspondence

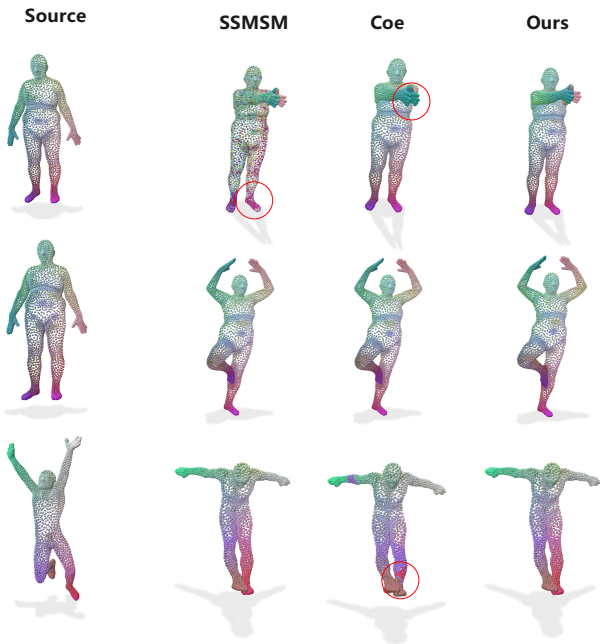

Figure 4: To showcase the performance of different methods in the overlapping regions of point clouds, red circles are used for marking. By observing the marked overlapping areas (to be highlighted with red circles), it can be seen that our method achieves more accurate and consistent correspondence in these regions.

Datasets. We evaluate the proposed method on four widely used datasets: FAUST[6], SCAPE[1], SHREC19[29], and DT4D-H[23]. FAUST contains 10 human categories (10 poses per category, for a total of 100 meshes); SCAPE contains 71 meshes of the same person in different poses, with the last 20 used for testing; SHREC'19 contains 44 shapes of diverse body shapes and poses, with no training set and used only for testing; and DT4D-H is a re-meshed subset of DeformingThings4D, containing 10 human-like shape categories (with significant differences in pose and style, making it a challenging benchmark). During training, 80% of the shapes are used for model learning and 20% for testing.

Result. As shown in Table 1, our method consistently outperforms most pure point cloud baselines on both SCAPE and FAUST. Compared to DiffMaps and DPC, our approach reduces the mean geodesic error by 30–50% on SCAPE-S (from 12.0/17.3 to 3.3) and by 40–70% on FAUST-F (from 19.0/17.5 to 3.5). Even against the recent DV-Matcher, which already delivers competitive performance, our method achieves a further $\sim$45% improvement on SCAPE-F (from 5.1 to 3.7) and $\sim$56% on FAUST-F (from 10.4 to 3.5). Beyond numerical gains, our approach also demonstrates better robustness to partiality and structural variations, producing smoother and more semantically consistent correspondences across diverse poses and identities. Figure 4 highlights a key advantage of our approach: robust correspondence in challenging overlapping and self-contact regions. While existing methods (e.g., SSMSM and CoE) often produce inconsistent or locally flipped matches in articulated areas such as arms or legs, GeoSe-FMap maintains smooth and semantically aligned mappings even under heavy occlusion. Notably, our framework operates in a setting without ground-truth correspondences, templates, or mesh connectivity, yet remains competitive with and in challenging regions even outperforms several supervised counterparts. This demonstrates the effectiveness of combining geodesic spectral stabilization with intrinsic semantic fusion.

We observe that on the DT4D-H benchmark, DV-Matcher slightly outperforms our method (8.1 vs. 11.2). We attribute this result to the significant variations in body style and the non-isometric deformation characteristics inherent to the DT4D-H dataset. Our GeoSe-FMap is constructed upon the intrinsic geometric consistency of surfaces; while this characteristic enables superior matching accuracy on standard deformable benchmarks like FAUST and SCAPE, it is naturally more sensitive to the extreme topological noise present in certain DT4D-H samples. Nevertheless, our method achieves competitive results without relying on large-scale external foundation models or mesh rendering processes.

4.4. Partial Shape Correspondence.

Datasets. We conducted a systematic evaluation of the proposed method on the SHREC'16 dataset [11], a widely used benchmark for the partial shape matching task in shape analysis. SHREC'16 contains 120 partial animal shape samples spanning various categories such as cats, dogs, horses, and lions. Each sample preserves only local key regions of the animal (e.g., head, torso, or limb segments), rather than the full body, making it one of the more challenging datasets for this task.

Result. Figure 5 shows our results on the SHREC'16 partial animal shape matching benchmark. Despite severe incompleteness, varying sampling densities, and disconnected components, GeoSe-FMap produces highly consistent feature distributions between source and target shapes, as indicated by the aligned color patterns. This confirms that our method generalizes well across diverse animal categories and local morphological variations, demonstrating strong robustness in partial shape correspondence scenarios.

4.5. Cross-Dataset and Cross-Modality Generalization

Dataset. To assess both dataset-level and modality-level generalization, we train our model exclusively on the synthetic mesh-based SURREAL dataset [37]and evaluate it directly on three real-world point cloud datasets: FAUST, SCAPE, and SHREC'19. This setup enforces a challenging Mesh $\rightarrow$ Point Cloud transfer, without any fine-tuning or domain adaptation.

Results. As shown in Table 2, GeoSe-FMap achieves consistently strong performance despite being trained on a different dataset and a different geometric modality. It outperforms most unsupervised baselines and remains competitive with several supervised counterparts. Notably, our method requires only a small amount of synthetic training data, yet generalizes to unseen real scans with large pose and topology variations. Qualitative results in the Figure 6 further confirm that the predicted correspondences remain spatially coherent even under severe cross-domain shifts.

4.6. Application Scenarios of Animal Shapes

Datasets. To verify the applicability of the proposed method to shape categories other than human bodies, we conducted tests on the SMAL remeshed dataset[47]. This dataset contains 49 animal shapes (covering 8 species); we divided the data according to the settings in [21], using 29 shapes (5 species) for training and 20 shapes (3 species) for testing. This division method ensures that all animal categories are distributed in both the training set and the test set.

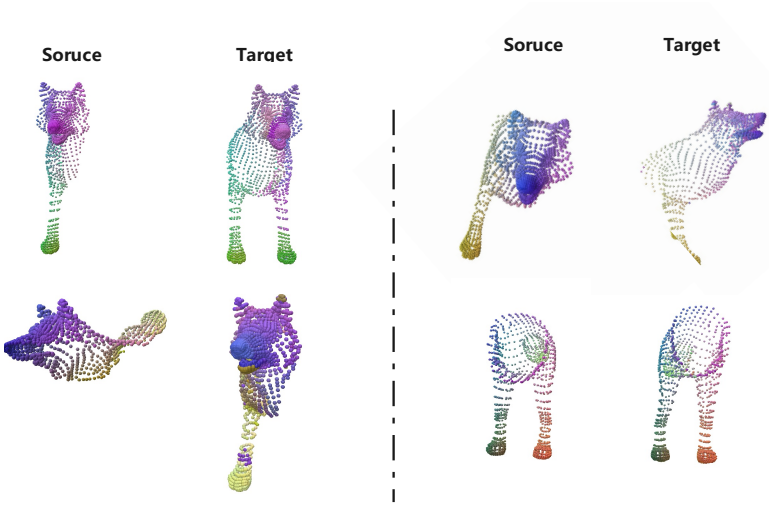

Figure 5: Results of our method on the SHREC16 dataset demonstrate that, even for partial point clouds with varying sampling rates and disconnected components, our approach achieves favorable correspondence.

Figure 6: Cross-dataset generalisation evaluated on the FAUST, SCAPE and SHREC'19 datasets and trained on the SURREAL dataset.

Result. The experimental results are shown in the Figure 7. From the results presented in the figures, the proposed method appears to achieve a certain degree of correspondence and transformation between the source and the target, regardless of the animal shapes (e.g., horses, hippopotamuses) or other different shapes. This initially indicates that the method has a certain adaptability to various animal shape categories and can capture the features of different animal shapes and perform corresponding processing.

4.7. Shape Segmentation

Datasets. The experiment adopts the composite human shape dataset from [28]. This dataset contains human meshes from multiple sources, such as [1], covering shapes of different body types and poses. Moreover, it provides implicit semantic segmentation references (e.g. joints, limb boundaries), which facilitates qualitative verification of segmentation rationality.

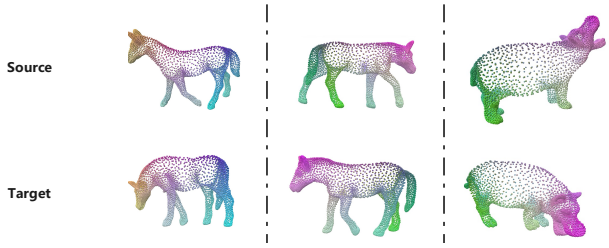

Figure 7: The figure illustrates shape correspondence results for animals. Across diverse animal forms (horse and hippopotamus), the results visually demonstrate effective shape matching via point cloud distribution and color coding.

Result. The primary objective of semantic shape segmentation is to partition a 3D object into structurally meaningful regions (e.g., head, torso, limbs for human bodies). As illustrated in Figure 8, our method produces highly consistent

Figure 8: The figure shows point cloud segmentation results of 3D human models. Across various poses (standing, walking, kicking, arm-raising), the segmentation is consistent and accurate, benefiting subsequent tasks like human pose analysis and motion capture.

Table 2: Results on the SURREAL dataset. The best results of methods in each column are highlighted. Our method not only outperforms other methods on most metrics but also demonstrates stable overall performance, fully verifying its competitiveness on this benchmark dataset.

Geo. (×100)	**FAUST**	**SCAPE**	**Shrec19**
FMNet [3]	12.2	15.3	22.7
DiffFMaps [27]	26.5	34.8	42.2
GeomFMaps[13]	10.4	8.7	14.1
SURFMNet [33]	16.0	14.7	27.8
Deep Shells[20]	12.5	14.1	15.9
ConsistFMaps[8]	19.3	17.3	24.2
CorrNet3D [43]	18.1	18.3	18.8
DPC [19]	13.4	15.8	17.4
SSMSM[9]	6.8	6.4	9.8
Ours	5.8	6.7	9.5

part-level segmentation across diverse poses such as standing, walking, kicking, and arm-raising. This stability under articulation suggests strong potential for downstream tasks including human pose understanding and motion capture.

4.8. Ablation Study

In this section, we analyze the contribution of the proposed semantic segmentation module and its interactions with other core components such as LBO optimization and semantic attention. We adopt the improved DiffusionNet as the backbone and conduct controlled comparisons by embedding or removing individual modules. To ensure fair comparison, all experiments follow the same configuration as the cross-dataset setup in Sec. 4.5.

Table 3: Ablation study on the contribution of loss functions and architectural components, reported as geodesic error (×100) on FAUST and SCAPE."w/o" is the abbreviation of "without".

Geo. error (×100)	**FAUST**	**SCAPE**
Ablation study on loss terms		
w/o L_{cons}	20.1	25.8
w/o L_{sem}	3.8	4.5
Ablation study on network components		
w/o LBO Optimization	3.9	4.3
w/o Semantic Attention	4.0	3.4
Ours	**3.7**	**3.3**

Effect of Loss Terms. As seen in Table 3, removing the consistency loss (L_{cons}) leads to a dramatic increase in error, reaching 20.1 (FAUST) and 25.8 (SCAPE). This confirms that L_{cons} is essential for enforcing globally coherent correspondence across the entire shape. Removing the semantic loss (L_{sem}) also degrades performance (3.8 / 4.5), though less severely, indicating that L_{sem} effectively guides the network to learn semantically structured features. This effect is further visualized in Figure 9, where removing semantic attention results in mismatches in symmetric or interacting regions (highlighted in red), whereas the full model preserves consistent alignment.

Effect of Network Components. Excluding LBO optimization raises the error to 3.9 / 4.3, demonstrating that

Figure 9: Ablation study on the impact of different loss terms. Removing semantic attention leads to mismatched regions (red circles), while the full model maintains consistent correspondence.

geodesic-regularized spectral stabilization improves feature alignment with underlying manifold geometry. Removing the semantic attention module yields 4.0 / 3.4, implying that attention helps the network focus on discriminative semantic regions when resolving local ambiguities.

Conclusion. The complete model achieves the best performance (3.7 / 3.3), showing that geometric stabilization and semantic guidance are complementary and jointly contribute to robust correspondence. Moreover, Table 4 shows that the semantic segmentation branch achieves a mIoU of 0.7371 and over 96% training accuracy, confirming that the learned semantic signals are both discriminative and reliable.

Table 4: Semantic segmentation performance of GeoSe-FMap, reported in terms of mIoU and classification accuracy.

Method	mIoU	Test Acc.	Train Acc.
Ours (GeoSe-FMap)	0.7371	76.59	96.89

4.9. Computational Complexity and Runtime Analysis

A contribution of our method is the Geodesic-Enhanced LBO. We analyze its overhead compared to the standard Euclidean LBO. The standard construction relies on k-NN search, with a time complexity of $O(N \log N)$ for N points. In contrast, our approach (Algorithm 1) incorporates an iterative refinement strategy requiring geodesic distance estimation. Assuming a sparse graph construction with E edges (where $E \approx kN$), executing Dijkstra's algorithm or the Fast Marching Method (FMM) T times yields a complexity of roughly $O(T \cdot (E + N \log N))$.

While our approach theoretically increases the computational cost during the graph construction phase, it is crucial to note that this is a one-time offline pre-computation step. The resulting LBO and its spectral bases are computed only once per shape and cached. Consequently, the online training and inference phases remain completely unaffected and share the same efficiency as standard spectral-based pipelines (e.g., DiffusionNet [34]).

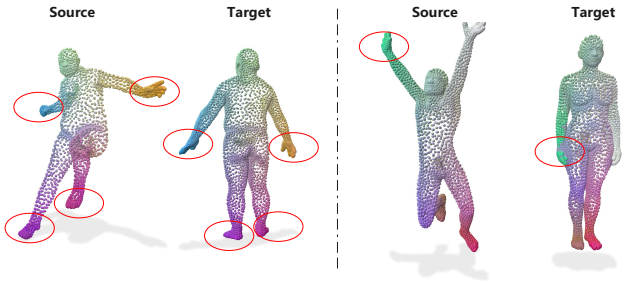

Figure 10: Point clouds often exhibit occluded or overlapping regions.

5. Conclusion

We introduced GeoSe-FMap, a functional map framework grounded in the concept of Structured Semantic Shapes (SSS), a class of deformable objects whose geometry follows an underlying part-based organization. By jointly leveraging geodesic-enhanced spectral operators and semantic consistency constraints, the proposed method enables robust correspondence across deformation, modality, and partiality without requiring mesh connectivity, templates, or expensive dense point-to-point supervision.

Comprehensive experiments demonstrate that GeoSe-FMap achieves competitive or superior performance compared to both supervised and mesh-based baselines, while exhibiting strong cross-modal generalization (mesh-to-point cloud) and resilience to occlusion. These results indicate that unifying intrinsic geometry with latent semantic structure provides a powerful prior for non-rigid correspondence. Despite these advantages, we explicitly define the applicability boundary of our method:

Theoretical Boundary on Non-SSS Shapes: Our Geometry-Semantic Fusion mechanism relies on the existence of decomposable semantic parts to construct the Part-Aware Attention Mask. For non-SSS objects (e.g., amorphous blobs or loose clothing) that lack stable semantic topology, this mask becomes degenerate, causing the

method to revert to purely geometric spectral matching. This limitation is empirically reflected in our results on the DT4D-H dataset (Table 1), where the presence of topological noise and ambiguous semantics leads to a performance drop compared to methods that do not rely on explicit SSS priors.

Topology and Noise Sensitivity: As shown in Fig. 10, severe topology-breaking overlaps can still lead to correspondence ambiguity. Furthermore, the current pipeline relies on a pre-constructed Laplace–Beltrami Operator, which remains sensitive to noise in raw point clouds.

Future work may explore learning spectral operators directly from point sets or extending SSS priors to category-agnostic scenarios. In summary, embedding semantic structure into geometric correspondence opens a promising direction toward generic, weakly supervised 3D perception across modalities and shape categories.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (62462055) and the College Students' Innovation and Entrepreneurship Training Program of Qinghai Normal University (qhnucxcy2026039, qhnucxcy2026040).

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM Siggraph 2005 Papers*, pages 408–416. 2005. 9, 10

[2] S. Attaiki and M. Ovsjanikov. Ncp: Neural correspondence prior for effective unsupervised shape matching. *Advances in Neural Information Processing Systems*, 35:28842–28857, 2022. 3

[3] S. Attaiki, G. Pai, and M. Ovsjanikov. Dpfm: Deep partial functional maps. In *2021 International Conference on 3D Vision (3DV)*, pages 175–185. IEEE, 2021. 3, 11

[4] S. Attaiki, N. Sharp, and J. Solomon. Understanding functional maps: A data-driven perspective. *Transactions on Graphics (TOG)*, 42(4):1–14, 2023. 1, 8

[5] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 1626–1633. IEEE, 2011. 3

[6] F. Bogo, J. Romero, M. Loper, and M. J. Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3794–3801, 2014. 7, 9

[7] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1704–1711. IEEE, 2010. 3

[8] D. Cao and F. Bernard. Unsupervised deep multi-shape matching. In *European conference on computer vision*, pages 55–71. Springer, 2022. 11

[9] D. Cao and F. Bernard. Self-supervised learning for multimodal non-rigid 3d shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17735–17744, 2023. 2, 8, 11

[10] Z. Chen, P. Jiang, and R. Huang. Dv-matcher: Deformation-based non-rigid point cloud matching guided by pre-trained visual features. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27264–27274, 2025. 4, 8

[11] L. Cosmo, E. Rodola, M. M. Bronstein, A. Torsello, D. Cremers, Y. Sahillioğlu, et al. Shrec'16: Partial matching of deformable shapes. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 61–67. Eurographics Association, 2016. 9

[12] J. Deng, C. Wang, J. Lu, J. He, T. Zhang, J. Yu, and Z. Zhang. Se-ornet: Self-ensembling orientation-aware network for unsupervised point cloud shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5364–5373, 2023. 8

[13] L. Donati, A. Sharma, and M. Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8592–8601, 2020. 1, 2, 3, 11

[14] N. Donati, E. Corman, S. Melzi, and M. Ovsjanikov. Complex functional maps: A conformal link between tangent bundles. In *Computer Graphics Forum*, volume 41, pages 317–334. Wiley Online Library, 2022. 3

[15] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. 3d-coded: 3d correspondences by deep deformation. In *Proceedings of the european conference on computer vision (ECCV)*, pages 230–246, 2018. 8

[16] J. Huang, T. Birdal, Z. Gojcic, L. J. Guibas, and S.-M. Hu. Multiway non-rigid point cloud registration via learned functional map synchronization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2038–2053, 2022. 8

[17] P. Jiang, M. Sun, and R. Huang. Neural intrinsic embedding for non-rigid point cloud matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21835–21845, 2023. 3, 6

[18] P. Jiang, M. Sun, and R. Huang. Non-rigid shape registration via deep functional maps prior. *Advances in Neural Information Processing Systems*, 36:58409–58427, 2023. 8

[19] I. Lang, D. Ginzburg, S. Avidan, and D. Raviv. Dpc: Unsupervised deep point correspondence via cross and self construction. In *2021 International Conference on 3D Vision (3DV)*, pages 1442–1451. IEEE, 2021. 3, 5, 8, 11

[20] H. Li and et al. Correspondence via deep shells: Unsupervised non-rigid shape matching with surface convolutions. *NeurIPS*, 2021. 3, 11

[21] L. Li, N. Donati, and M. Ovsjanikov. Learning multi-resolution functional maps with spectral attention for robust shape matching. *Advances in Neural Information Processing Systems*, 35:29336–29349, 2022. 9

[22] Y. Li and T. Harada. Non-rigid point cloud registration with neural deformation pyramid. *Advances in Neural Information Processing Systems*, 35:27757–27768, 2022. 8

[23] Y. Li, H. Takehara, T. Taketomi, B. Zheng, and M. Nießner. 4dcomplete: Non-rigid motion estimation beyond the observable surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12706–12716, 2021. 9

[24] O. Litany, T. Remez, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 5659–5667, 2017. 1, 2, 3, 7

[25] S. Liu, S. Gai, F. Da, and F. Waris. Geometry-aware 3d pose transfer using transformer autoencoder. *Computational Visual Media*, 10(6):1063–1078, 2024. 1

[26] X. Luo, T. Wang, et al. Usip: Unsupervised stable interest point detection from 3d point clouds. In *CVPR*, 2019. 3

[27] R. Marin, M.-J. Rakotosaona, S. Melzi, and M. Ovsjanikov. Correspondence learning via linearly-invariant embedding. *Advances in Neural Information Processing Systems*, 33:1608–1620, 2020. 6, 8, 11

[28] H. Maron, M. Galun, N. Aigerman, M. Trope, N. Dym, E. Yumer, V. G. Kim, and Y. Lipman. Convolutional neural networks on surfaces via seamless toric covers. *ACM Trans. Graph.*, 36(4):71–1, 2017. 10

[29] S. Melzi, J. Ren, E. Rodola, A. Sharma, P. Wonka, and M. Ovsjanikov. Zoomout: Spectral upsampling for efficient shape correspondence. *arXiv preprint arXiv:1904.07865*, 2019. 5, 6, 7, 9

[30] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (ToG)*, 31(4):1–11, 2012. 1, 3, 6

[31] E. Pierson, L. Li, A. Dai, and M. Ovsjanikov. Diffumatch: Category-agnostic spectral diffusion priors for robust non-rigid shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5745–5756, 2025. 3

[32] J. Ren, A. Poulenard, P. Wonka, and M. Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics (ToG)*, 37(6):1–16, 2018. 3

[33] J.-M. Roufosse, A. Sharma, and M. Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1617–1627, 2019. 3, 11

[34] N. Sharp, S. Attaiki, K. Crane, and J. Solomon. Diffusionnet: Discretization agnostic learning on surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1810, 2020. 2, 3, 5, 6, 7, 12

[35] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11*, pages 356–369. Springer, 2010. 3

[36] G. Trappolini, L. Cosmo, L. Moschella, R. Marin, S. Melzi, and E. Rodolà. Shape registration in the time of transformers. *Advances in Neural Information Processing Systems*, 34:5731–5744, 2021. 8

[37] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid. Learning from synthetic humans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 109–117, 2017. 9

[38] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5):1–12, 2019. 3

[39] Y. Xie et al. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *ECCV*, 2020. 3

[40] Y. Yao, B. Deng, W. Xu, and J. Zhang. Fast and robust non-rigid registration using accelerated majorization-minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9681–9698, 2023. 8

[41] Z. Y. Yew and G. H. Lee. Rpm-net: Robust point matching using learned features. In *European Conference on Computer Vision (ECCV)*, pages 48–65. Springer, 2020. 3

[42] H. Zeng, M. Gao, and D. Cremers. Coe: Deep coupled embedding for non-rigid point cloud correspondences. In *2025 International Conference on 3D Vision (3DV)*, pages 286–295. IEEE, 2025. 3, 7

[43] Y. Zeng, Y. Qian, Z. Zhu, J. Hou, H. Yuan, and Y. He. Corrnet3d: Unsupervised end-to-end learning of dense correspondence for 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6052–6061, 2021. 8, 11

[44] M. Zhao, J. Jiang, L. Ma, S. Xin, G. Meng, and D.-M. Yan. Correspondence-free non-rigid point set registration using unsupervised clustering analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21199–21208, 2024. 8

[45] J. Zhou, T. Liu, and L. Wang. Cost analysis of 3d medical image annotation for shape matching. *Medical Image Analysis*, 78:102376, 2022. 2

[46] A. Zhuravlev, Z. Lähner, and V. Golyanik. Denoising functional maps: Diffusion models for shape correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26899–26909, 2025. 1, 3

[47] S. Zuffi, A. Kanazawa, D. W. Jacobs, and M. J. Black. 3d menagerie: Modeling the 3d shape and pose of animals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6365–6373, 2017. 9