

FUSE: A Flow-based Mapping Between Shapes

Lorenzo Olearo¹, Giulio Viganò¹, Filippo Maggioli⁴, Daniele Baieri^{1,2,3},
and Simone Melzi¹

¹ University of Milano-Bicocca, Milan, Italy

² University of Bonn, Bonn, Germany

³ Lamarr Institute, Germany

⁴ Pegaso University, Naples, Italy

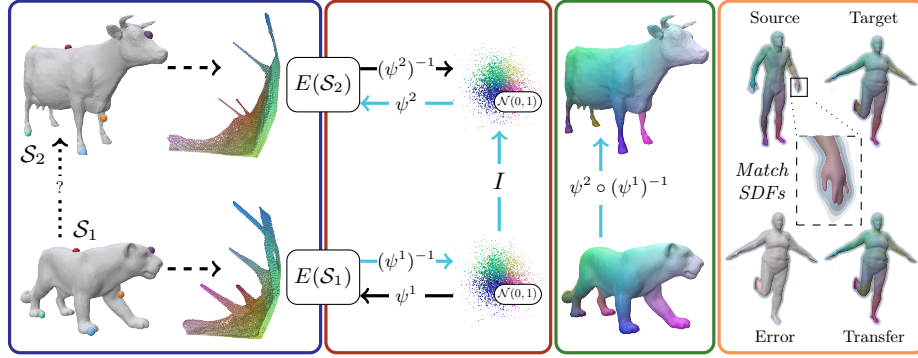


Fig. 1: Our representation establishes a map between shapes S_1, S_2 by optimizing invertible flows (red box) between embedding distributions (blue box) and the unit Gaussian $\mathcal{N}(0, 1)$. The composition of said flows (green box) is a pointwise map between the surfaces, enabling shape matching and supporting various 3D representations, including neural signed distance fields (orange box).

Abstract. We introduce a novel neural representation for maps between 3D shapes based on flow-matching models, which is computationally efficient and supports cross-representation shape matching without large-scale training or data-driven procedures. 3D shapes are represented as the probability distribution induced by a continuous and invertible flow mapping from a fixed anchor distribution. Given a source and a target shape, the composition of the inverse flow (source to anchor) with the forward flow (anchor to target), we map points between the two surfaces. By encoding the shapes with a pointwise task-tailored embedding, this construction provides an invertible and modality-agnostic representation of maps between shapes across point clouds, meshes, signed distance fields (SDFs), and volumetric data. The resulting representation consistently achieves high coverage and accuracy across diverse benchmarks and challenging settings in shape matching. Beyond shape matching, our

framework shows promising results in other tasks, including UV mapping and registration of raw point cloud scans of human bodies.

Keywords: Shape Matching · Flow Matching · Neural Representations

1 Introduction

Shape matching has long been recognized as a “key problem” in geometry processing: indeed, accurate correspondences between two shapes enable simple solutions for several downstream tasks (*e.g.*, pose transfer [55] or interpolation [17]). Most of the research in this field, following the common practices from 3D graphics industry, was built on the triangle mesh representation, which approximates a continuous 2-dimensional manifold embedded in $\mathbb{R}^3$ piecewise linearly with a collection of triangles connected by their common edges and vertices. However, recent progress in the fields of computer vision and deep learning has popularized novel ways of encoding 3D geometry, such as neural representations [43, 49]. These solutions offer several advantages compared to triangle meshes, but lack their large suite of tools and algorithms for analysis and manipulation [68].

In this work, we address this challenge by introducing a novel representation for maps between a source and a target shape that can be applied over various geometric encodings. We namely refer to this representation as FUSE. Constructing a map with this representation requires only two components: i) the ability to sample points from the involved surfaces, and ii) a pointwise feature embedding designed for the target task. Once feature embeddings are computed for the sampled points on both source and target shapes, FUSE estimates an invertible flow from each embedding distribution to a shared anchor distribution, following Flow Matching framework [27]. FUSE then expresses the map between the two shapes as the composition of the inverse flow of the source (source $\rightarrow$ anchor) and the forward flow of the target (anchor $\rightarrow$ target). We note that, given the flows of any two shapes, the map is obtained with no additional optimization, with only landmarks as information shared by the two shapes.

In the context of shape matching, our representation introduces a probabilistic perspective on embedding alignment, where semantically similar features are encouraged to align through a shared probability distribution. Combined with the guarantee of mapping points directly into the target embedding space, our framework produces high-coverage and accurate correspondences that are relevant for different applications. We evaluate the proposed approach through a comprehensive experimental study, comparing it against several shape matching baselines, ranging from well-established axiomatic methods to recent data-driven approaches; FUSE shows high, if not the highest, accuracy on various benchmarks. Our contributions can be summarized as follows:

- We introduce FUSE, the first flow-matching based representation for 3D shape maps that naturally enables correspondence between shapes expressed in different modalities (*e.g.*, meshes, point clouds, and SDFs);

- We quantitatively show that FUSE achieves competitive or superior matching accuracy over state-of-the-art methods, particularly in the context of landmark-based shape matching;
- We showcase the versatility of FUSE, employing it as a backbone for other downstream applications, ranging from shape parametrization to point cloud deformation.

2 Related Work

2.1 Neural Shape Representations

Recent research in geometric deep learning and geometry processing popularized the practice of encoding 3D shapes in the weights of neural networks, which is referred to as neural shape representation. The early contributions in this field are due to [42] and [49], followed by the introduction of NeRF [43], which led to a vast effort to adapt neural shape representations to other tasks besides reconstruction. Examples of these adaptations are to shape deformation [1, 4, 45, 68], text-guided shape editing [30, 65] and generation [29, 51]. Recent contributions also introduced neural representations for 3D simulations [48] and computation of geometric operators over spherical neural surfaces [66].

The work of [23] is closely related to ours, as they propose a method to represent functions continuously over surfaces in intrinsic space. Lastly, FUSE’s architecture is inspired by Geometry Distributions [69], which proposes to represent geometry as a continuous map from a 3D Gaussian distribution to the surface with a denoising diffusion model [56]. However, by targeting the specific goal of representing maps between shapes, we build upon flow matching [27], and model flows directly over an intrinsic feature space, rather than standard Euclidean or extrinsic coordinates. These two choices together enable FUSE to represent mappings, as we prove by targeting the shape correspondence task.

2.2 Shape Matching

Shape correspondence is a fundamental task in geometry processing, with a wide range of approaches proposed over the years, which can largely be distinguished by how they represent the map between shapes and features [3, 8, 40, 57, 59]. Among these, the functional map framework [46] and its follow-up works [11, 14, 32, 39, 41] have emerged as a dominant paradigm, where correspondences are represented as linear operators between functional spaces. These methods offer attractive properties such as smoothness and invariance to deformations, but the reliance on spectral bases limits their generalization to non-mesh representations [39] and in obtaining bijective maps [5, 38, 64].

Another class of approaches leverages optimal transport (OT) [13, 31, 54], in which correspondences are represented by transport plans that minimize the cost of transporting a distribution of features from one shape to another. In the presence of sampling imbalances or ambiguous features, OT plans often exhibit mass-splitting behavior, where individual points distribute probability mass

across multiple targets. This diffusion can degrade correspondence sharpness and stability when recovering discrete maps. Alternatively, registration-based approaches directly optimize non-linear deformation fields in 3D [17, 26, 52, 60]. The fields are often parameterized via deformation networks or complex architectures, achieving accurate alignments at the cost of solving computationally heavy optimization problems and depending on rigid initialization or learned priors.

Hybrid approaches combine the strengths of the above formulations [10, 11, 20, 21, 24, 47, 62], gaining in efficiency and robustness, but still relying on learned priors and struggling to generalize to other representations. Another common approach is template-based matching, representing correspondences as a composition of maps with respect to a common template, selected a priori [19, 36, 37] or directly at test time [34]. On a different perspective, FUSE offers a novel representation for maps. Based on flow matching [27], our method models a correspondence as a composition of continuous invertible flows between intrinsic feature spaces and a shared anchor distribution, inherently enforcing bijectivity and high-coverage in obtained maps, without learned priors or per-pair optimization in a pipeline naturally agnostic of the geometric representation.

3 Background

Shapes and Matching. A 3D shape $\mathcal{S}$ is modeled as a compact two-dimensional manifold embedded in $\mathbb{R}^3$. From a computational standpoint, single shapes may be represented in several ways: polygonal meshes, point clouds, signed distance functions and volumetric meshes are notable examples. In this work, we identify a shape $\mathcal{S}$ as a continuous manifold from which we can sample unordered sets of points $\{x_i\}_{i=0}^n$, encoded by their 3D coordinates (i.e. $x_i \in \mathbb{R}^3$) independently from the adopted representations.

Given 3D shapes $\mathcal{S}_1 = \{x_i\}_{i=1}^{n_1}$ and $\mathcal{S}_2 = \{y_j\}_{j=1}^{n_2}$ with a common semantic, a point-to-point correspondence T_{12} maps each point x_i to its corresponding point y_j . Shape matching is the task of estimating such correspondence. In the non-rigid scenario, estimating correspondences directly from 3D coordinates is challenging; thus, most methods describe the shapes via pointwise *features*, characterizing their semantics and geometry [15, 57, 59]. For each point $x_i \in \mathcal{S}_1$, we compute a vector of features $f_i \in \mathbb{R}^d$ that we can then store as a row in the *embedding matrix* $E_1 \in \mathbb{R}^{n_1 \times d}$ (same goes for $\mathcal{S}_2$). Ideally, one can solve for the correspondence directly by leveraging the similarity between these encodings [3]. In practice, more complex pipelines are required to align the features and retrieve the correspondence from their relations [46].

Flow Matching. Flow Matching (FM) [27, 28] is a framework for learning continuous in time probability paths $(p_t)_{t \in [0,1]}$ that transport a known source distribution p_0 (e.g., a Gaussian) to an unknown target p_1 (e.g., data samples). This is achieved by learning a *velocity field* $u_t^\theta(x)$ describing the instantaneous dynamics of the path, parameterized by a neural network θ . For each $x \in \mathbb{R}^d$, the velocity

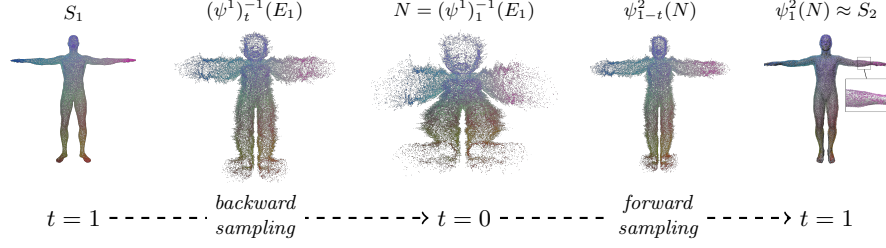


Fig. 2: Visual example of the proposed pipeline. For visualization purpose, we select a pair for which we can solve for the correspondence using the 3D coordinates as features embedding. Estimated correspondence is encoded by the color (matched points have the same color).

field induces a flow $\psi_t(x)$ as the solution of the ODE $\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x))$ with initial condition $\psi_0(x) = x$. Crucially, the flow ψ_t is invertible and diffeomorphic: starting from any point sampled from p_1 , i.e., $x \sim p_1$, integrating the ODE backward yields a point $\psi_0^{-1}(x) \sim p_0$. The bidirectionality enables transport between distributions in both directions. Since u_t is generally intractable, [27, 28] proposed Conditional Flow Matching (CFM), replacing u_t with a tractable conditional velocity field $u_t(x_t | x_1)$, while preserving the same gradients:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, x_1, x_t \sim p_t(\cdot | x_1)} [\|v_t^\theta(x_t) - u_t(x_t | x_1)\|^2]. \quad (1)$$

where $\mathbb{E}_{t, x_1, x_0}$ represents the average over randomly sampled timesteps in $[0, 1]$ and over source and target samples $x_0 \sim p_0, x_1 \sim p_1$. A simple choice for the field is linear interpolation, $x_t = (1-t)x_0 + tx_1$ with $u_t = x_1 - x_0$. CFM training is thus simulation-free, although generating samples still requires integrating the learned ODE. In the remainder of the paper, we will refer to CFM simply as Flow Matching (FM).

4 FUSE

Given the overview presented in the previous section, we can now provide all the details of our novel representation FUSE which we outlined in Fig. 1. The complete source code is available at <https://github.com/Lorenzo01earo/FUSE-Flow-based-Mapping-Between-Shapes>.

4.1 Flow Optimization

Given a pair of shapes $\mathcal{S}_1 = \{x_i\}_{i=1}^{n_1}$, $\mathcal{S}_2 = \{y_j\}_{j=1}^{n_2}$ and an embedding matrix $E_\ell \in \mathbb{R}^{n_\ell \times d}$ (with $\ell = 1$ or 2), we denote the resulting embedding distributions as $p^\ell = E_\ell$ (see blue box of Fig. 1). For each shape, we train an independent flow model that learns a continuous, invertible mapping $\psi^\ell: [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. This flow transports samples from a fixed, shared Gaussian $p_0 = \mathcal{N}(0, 1)^d$ (the

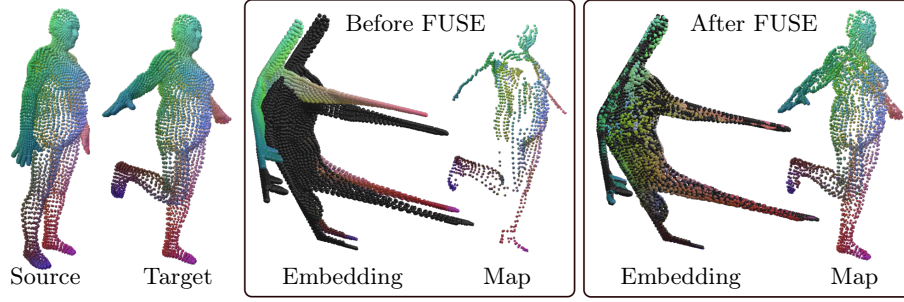


Fig. 3: Effect of FUSE visualized on the first three coordinates of embedding and on the relative obtained map.

anchor distribution) to the shape’s embedding distribution p^ℓ , i.e. $\psi_1^\ell(x) \sim p^\ell$ given $x \sim p_0$. Each flow is parametrized by a neural network and trained via Conditional Flow Matching [27,28] as we described earlier (see red box of Fig. 1). Each flow is trained independently on its shape, requiring only the ability to sample embeddings from the surface.

4.2 Fusing Representations

FUSE determines maps between shapes by composing their flows. Given the two trained flows ψ^1 and ψ^2 , we define the map from $\mathcal{S}_1$ to $\mathcal{S}_2$ as:

$$\Phi^{12} = \psi^2 \circ (\psi^1)^{-1} \quad (2)$$

This composition leverages the Gaussian distribution as a common anchor: we transport the embedding points from $\mathcal{S}_1$ back to the Gaussian via $(\psi^1)^{-1}$, then forward to $\mathcal{S}_2$ via ψ^2 as depicted in Fig. 2. This process offers several advantages:

1. Injection of invariance of the adopted embeddings in the output map;
2. For any pair of shapes represented by the respective flows from a common distribution, no extra training is required to obtain a map between them;
3. Representation invariance: creating a bridge with FUSE only requires sampling from the surface, which is straightforward on most 3D representations.

We can leverage this representation of maps between shapes to perform various tasks on the discrete geometry. For instance, to retrieve the correspondence between points of two discretized shapes, we can perform a nearest neighbor search in the embedding space.

$$T_{12} = \text{NearestSearch}(\Phi^{12}(E(\mathcal{S}_1)), E(\mathcal{S}_2)) \quad (3)$$

The validity of this construction rests on the nature of the FM solutions and the shared anchor distribution. While FM training relies on independent random couplings between Gaussian and data samples, this randomness affects

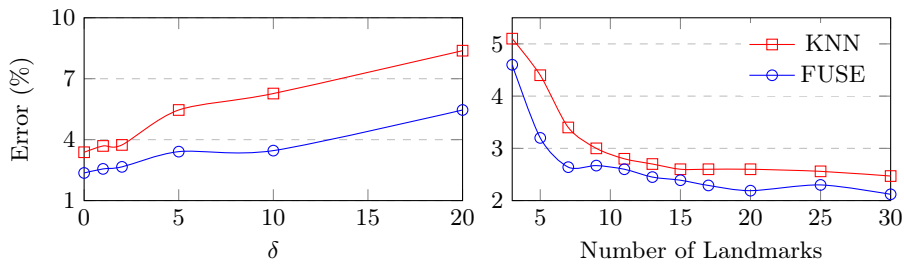


Fig. 4: Stability analysis of FUSE. On the left: robustness under landmark perturbation (δ). On the right: performances varying the number of landmarks.

only the conditional velocity fields, not the learned flow. Theorem 2 in [27] shows that marginalizing the conditional fields $u_t(x | x_1)$ over $x_1 \sim p_1$ recovers the *unique* marginal field u_t generating the probability path between p_1 and p_0 ; since this field is determined by (p_1, p_0) alone, so is the induced flow ψ_t . Consequently, if $E(S_1) = E(S_2) = p_1$, the flows ψ^1 and ψ^2 coincide regardless of the couplings used during training, and when $E(S_1)$ and $E(S_2)$ are similar, their composition $\psi^2 \circ (\psi^1)^{-1}$ aligns $E(S_1)$ to $E(S_2)$ and produces accurate correspondences (Fig. 3). For this reason, points with similar embeddings from different shapes are mapped to nearby regions within the anchor (green box in Fig. 1), provided that both shapes share the same semantic structure.

The same reasoning also highlights when FUSE can be effective. When S_1 and S_2 differ by a transformation to which E is not invariant, the distributions $E(S_1)$ and $E(S_2)$ become dissimilar and not alignable by the flow composition. Instead, when the two embedding distributions share the same semantic structure, by forcing the image of $E(S_1)$ to align with $E(S_2)$ before applying KNN (Fig. 3), FUSE reduces the sensitivity to embedding variations. This is why, as the isometry decreases (Tab. 4) or the landmark placement becomes noisy (Fig. 4), the advantage of FUSE over other methods tends to increase.

4.3 Embedding Selection

The choice of embedding E is critical. First, the generalization of FUSE to any 3D representation depends on E being computable and invariant over 3D representations. Secondly, the embedding determines the properties of the FUSE map (*e.g.*, shape matching needs some E that captures metric and semantic information about the shapes). Every shape matching pipeline relies on assumptions that inject prior knowledge into the process. These assumptions are necessary to resolve fundamental ambiguities, most notably

Table 1: Correspondence error under different prior assumptions. The results are obtained without any test-time refinement. *ULRSSM results are obtained using public checkpoints. Details on metric and datasets in 5.

Method	Prior	Error	
		FAUST	SMAL
Diff3	Vision model	0.089	0.189
ULRSSM*	Training-data	0.043	0.063
FUSE	Landmarks	0.028	0.059

intrinsic symmetries, which cannot be disambiguated from purely geometric information alone. Past methods encode this prior in different ways: [11] learns a feature extractor in an unsupervised manner, where the (data) prior is induced by training over a collection of shapes; [15] leverages a pretrained large-scale vision model to extract semantic information from rendered views of meshes; [3] relies on informative spectral signatures. Each of these strategies introduces representation-specific dependencies that limit their applicability: spectral bases, rendering pipelines, or available training data.

In contrast, 3D coordinates (*i.e.*, “identity” embeddings) are universally available but lack intrinsic geometric or semantic meaning, making them unsuitable for shape correspondence. One effective alternative is the Geodesic Feature Embedding (**GFE**), which represents surface points as vectors of geodesic distances to a small set of landmarks L , yielding a compact embedding in $\mathbb{R}_+^{|L|}$ that encodes the intrinsic geometry. Landmarks provide a minimal yet powerful form of prior: they provably resolve intrinsic symmetries by anchoring the semantic orientation of the shape, and are consistently employed in axiomatic shape matching methods derived from the Functional Maps [46] framework. In our case, landmarks are manually selected on one shape (5 or 6 depending on the semantics of the dataset) and transferred to the rest of the dataset via ground truth correspondences; Fig. 12 in the supplementary material shows the selected landmarks for each dataset. As shown in Tab. 1, despite this simple prior, FUSE with landmark-based GFE achieves the lowest correspondence error on both FAUST and SMAL against state-of-the-art prior-based methods [11, 15].

Importantly, different algorithms have been developed to compute or efficiently approximate geodesic distances on triangle meshes, signed distance fields, and point clouds, such as Dijkstra’s shortest path algorithm, the heat method [12], and fast marching [44]. While computing geodesic distances generally involves estimating the underlying surface, established methods exist (e.g. [12]) that do so implicitly with stable results. As a consequence, FUSE does not require explicit surface estimation over point clouds and can employ them directly as an input representation.

However, geodesic embeddings are not naturally aligned across shapes from a distributional perspective as a non-rigid deformation can significantly alter the geodesic distances between points. With FUSE, the flow composition acts as a probabilistic alignment mechanism: by mapping geodesic features through the Gaussian prior, we obtain distributional alignment across shapes, as shown in Fig. 3 Fig. 3.

Generality and robustness. While GFE is our primary choice, FUSE can be applied to any pointwise embedding. To understand when it is effective, we vary both the embedding type (Tab. 2) and the landmark configuration (Fig. 4). In our experiments, FUSE does not appear to suffer from the increasing dimensionality: adding landmarks from 3 to 30 yields progressively better results that consistently outperform KNN (Fig. 4, right); we observe the same behaviour when perturbing landmark positions by a random fraction of the shape diameter (Fig. 4, left). Instead, Tab. 2 reveals that a critical factor is *smoothness*: GFE

Table 2: Euclidean error on FAUST and SMAL datasets with different embeddings. WKS* and WKS-L* are smoothed versions, to show the improvement of FUSE over KNN in the presence of low-energy embeddings.

Method	FAUST							SMAL				
	x, y, z	GFE	WKS	WKS*	WKS-L	WKS-L*	ULRSSM	x, y, z	GFE	WKS	WKS-L	ULRSSM
Dim	3	5	20	20	20	20	256	3	5	20	20	256
KNN	0.1572	0.0454	0.1007	0.1125	0.0428	0.0522	0.0723	0.2903	0.0879	0.2736	0.0697	0.1061
FUSE	0.1488	0.0289	0.1015	0.0901	0.0556	0.0416	0.0829	0.3226	0.0595	0.3039	0.0668	0.1117
Dirichlet	0.0002	0.0001	0.0056	0.0008	0.0022	0.0004	0.1847	0.0001	0.0001	0.0257	0.0131	0.637

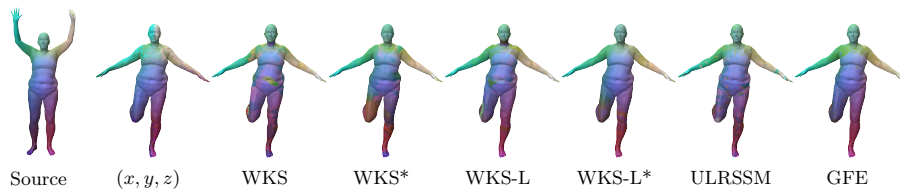


Fig. 5: Correspondences obtained by FUSE with different embeddings. We note that (x, y, z) does not encode the semantic correspondence and cannot capture the non-rigid deformation. WKS does not solve the intrinsic symmetries of the shape. With learned embeddings (ULRSSM), our methods introduce high-frequency artifacts.

has very low Dirichlet energy and benefit greatly from the flow mapping, while WKS [3] and learned embeddings [11] have substantially higher Dirichlet energy, resulting in poor performance. We refer to this as *curse of frequency*: high-frequency embeddings induce complex distributions that are harder for the flow to align, a challenge mirroring known difficulties in Functional Maps optimization [41, 46]. GFE exhibits lower spectral complexity, making them particularly well-suited for FUSE. We note that smoothing the embeddings via projection on a small set of Laplace-Beltrami eigenbasis [25, 58, 61], (WKS* and WKS-L* in Tab. 2) consistently improves FUSE’s accuracy over raw features, even in non-landmark-based scenarios. This suggests that integrating smoothly learned features with flow-based models is a viable direction and fosters future work.

We note that the invertibility of the flows does not by itself induce a bijection between the shapes: the final KNN step (Sec. 4.2) assigns each source point its nearest target independently, so the matching is not strictly bijective, as is standard in point-based correspondence. Since both the GFE and the flow are defined continuously over the surface, the construction naturally extends toward a continuous, fully bijective map, which we leave to future work.

4.4 Flow matching over Diffusion

Building on the recent work of [69], which represents 3D shapes as probability distributions generated via a Denoising Diffusion Implicit Model (DDIM) [56], we adopt the same probabilistic representation while constructing maps between

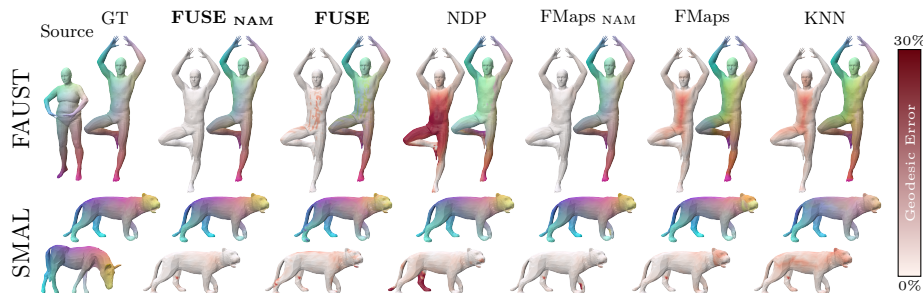


Fig. 6: Qualitative evaluation: We report a pair of shapes with visualization of the p2p as transferred colormap and in terms of the geodesic error visualized via a hot-colormap. We note the lower but less smooth error of Ours compared to FMaps.

shapes using flow matching rather than DDIM. Although FM and DDIM can be similar, the former offers distinct advantages for our purposes. FM establishes an explicit relationship between the Gaussian distribution and the data distribution. This ensures that the inverse flow maps any shape back to the exact Gaussian distribution in a deterministic manner. This property guarantees that all shapes lie in an ideally shared latent space. In contrast, methods such as DDIM only asymptotically approach the Gaussian through large noise schedules, resulting in deviations from the anchor space.

We quantify this effect in Tab. 3, evaluating shape pairs with GFE using two criteria: (1) the accuracy of nearest-neighbor correspondences both in the anchor space and in the embedding space, and (2) the JS and KL divergence between embedding distributions. We compute these metrics before and after inverting the flows using DDIM (FUSE_{DDIM}) and FM (FUSE). The results show that FUSE achieves substantially lower divergence and higher correspondence accuracy, confirming that accurate anchoring to the Gaussian directly correlates with improved matching performance. DDIM’s inversion leads to larger deviations from the anchor and thus inferior map quality. In the supplementary materials, we provide a mathematical derivation of this crucial difference, along with additional examples.

Table 3: FUSE_{DDIM} vs FUSE in terms of shape matching error (first four rows) and probability divergence (last two rows) on the FAUST dataset. Lower is better for all the metrics.

Metric	Raw	FUSE _{DDIM}	FUSE
Euclidean Error (anchor)	0.045	0.099	0.042
Geodesic Error (anchor)	0.043	0.095	0.039
Euclidean Error	-	0.037	0.028
Geodesic Error	-	0.035	0.027
Mean KL Divergence	0.2528	0.3519	0.0375
Mean JS Divergence	0.1928	0.2128	0.0829

5 Results

We evaluate our representation, considering shape matching as the main task, but also presenting possible extensions of FUSE targeting other applications.

5.1 Non-Rigid Shape Matching

We first analyze the performance of FUSE against more standardized benchmarks, considering pairs of meshes and point clouds. We then explore novel and more challenging scenarios by matching SDFs and inter-representation pairs. We restrict our analysis to landmarks-based methods, as discussed in Sec. 4.3.

Baselines. We evaluate FUSE against baselines from different paradigms. Methods using our same geodesic feature encoding include nearest neighbor search in the embedding space (KNN) and in the anchor distribution space (FUSE_{ANCHOR}), OT (Optimal Transport with Sinkhorn relaxation [13]), Functional Maps (FMaps) with geodesic descriptors [46], and refined variants FMaps_{ZO} (ZoomOut [41]) and FMaps_{NAM} (Neural Adjoint Maps [62]). Further, we also consider the original Functional Maps formulation with WKS [3], the neural extrinsic Neural Deformation Pyramid (NDP) approach [26], the vision-foundation Diff3F [15], and the data-driven ULRSSM [11]. Finally, we use FUSE to initialize ZoomOut and NAM (FUSE_{ZO} and FUSE_{NAM}).

Datasets. We evaluate our method on well-established datasets for non-rigid shape matching: **FAUST** [7] (near-isometric human meshes), **SMAL** [11, 70] (non-isometric quadruped meshes), **SHREC20** [16] (strongly non-isometric animal meshes), and **KINECT** [6, 62] (human point clouds). We manually pick 5 landmarks for humans and 6 for animals, Fig. 12 in the supplementary shows the selected landmarks for each dataset.

Metrics. We evaluate the quality of correspondence using four metrics: two measures of accuracy, Euclidean (Eucl.) and geodesic (Geod.) distances [22], as well as Dirichlet energy [35] and coverage of the computed maps. Lower values are better for the first three metrics, while a higher value is better for Coverage. Quantitative results are reported in Tab. 4, qualitative in Fig. 6, and plots with the full distribution of the metrics in Fig. 13 of the supplementary material.

Mesh correspondences. While FMaps-based methods tend to produce smoother maps with smaller Dirichlet energy, our method standouts in terms of both Euclidean and geodesics accuracy. Due to the nature of the flows, invertible by design, FUSE achieves better coverage of the target surface overall. The introduction of refinement strategies (*i.e.*, ZO and NAM) adds pairwise optimization,

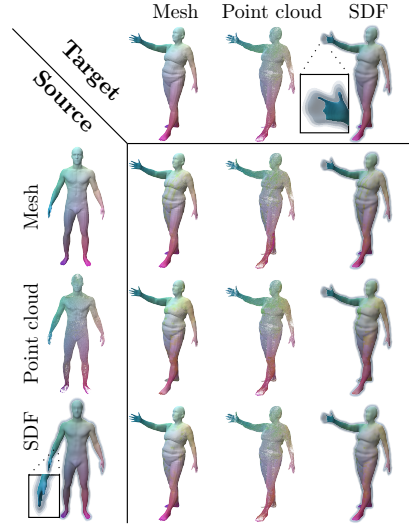


Fig. 7: Coordinate transfer among different 3D representations enabled by FUSE.

Table 4: Averaged results from all the considered datasets for our method and the other baseline solutions. For each column, the best result is highlighted in bold.

Method	FAUST (quasi-iso)				SMAL (non-iso)				SHREC20 (strong non-iso)				Kinect (point cloud)			
	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑
KNN	0.0454	0.0434	0.0022	0.2452	0.0879	0.0836	0.0016	0.1364	0.1083	0.1036	0.0063	0.1446	0.1119	0.1095	0.0025	0.0995
OT [13]	0.0926	0.0856	0.0016	0.0890	0.0779	0.0741	0.0012	0.1129	0.1252	0.1168	0.0060	0.1001	0.1327	0.1273	0.0047	0.0748
FMap [46]	0.0401	0.0400	0.0016	0.3236	0.0756	0.0744	0.0018	0.2162	0.1381	0.1411	0.0077	0.2037	0.0810	0.0779	0.0031	0.1683
FMapwks [3]	0.0331	0.0334	0.0017	0.3759	0.0647	0.0623	0.0023	0.2585	0.1197	0.1131	0.0203	0.2380	0.0723	0.0718	0.0045	0.2211
FMapzo [41]	0.0207	0.0195	0.0009	0.6969	0.0625	0.0611	0.0065	0.4706	0.1454	0.1480	0.0152	0.3495	0.0706	0.0710	0.0062	0.4043
FMapNAM [62]	0.0245	0.0240	0.0009	0.6323	0.0617	0.0589	0.0015	0.4694	0.1041	0.0962	0.0082	0.2860	0.0689	0.1118	0.0026	0.3494
NDP [26]	0.0702	0.0886	0.0020	0.4893	0.0555	0.0675	0.0010	0.4818	0.0698	0.0649	0.0013	0.5016	0.0457	0.0438	0.0012	0.4190
Diff3f [15]	0.0697	0.1109	0.0497	0.3075	0.1594	0.2278	0.0560	0.2146	0.0876	0.1036	0.0443	0.1579	0.1394	0.1732	0.0755	0.1448
ULRSSM [11]	0.0553	0.0886	0.0036	0.4427	0.1283	0.1943	0.0114	0.2789	0.1173	0.1357	0.0133	0.2376	0.1880	0.2087	0.0362	0.0565
	0.0092	0.0109	0.0011	0.7302	0.0370	0.0374	0.0030	0.4936	0.3286	0.3496	0.2299	0.1337	N/A	N/A	N/A	N/A
FUSEANCHOR	0.0425	0.0390	0.0241	0.3483	0.0579	0.0542	0.0185	0.3708	0.0912	0.0806	0.0538	0.3619	0.0983	0.0925	0.0328	0.2459
FUSEDDIM	0.0402	0.0407	0.0023	0.2973	0.0921	0.0879	0.0016	0.1500	0.0856	0.0797	0.0030	0.2214	0.1106	0.1101	0.0025	0.1326
FUSE	0.0289	0.0274	0.0066	0.5320	0.0595	0.0576	0.0075	0.5156	0.0975	0.0896	0.0339	0.5278	0.0969	0.0914	0.0137	0.4931
FUSE zo	0.0200	0.0189	0.0009	0.7127	0.0437	0.0413	0.0023	0.5361	0.1275	0.1312	0.0093	0.3763	0.0690	0.0662	0.0046	0.4173
FUSE NAM	0.0179	0.0168	0.0009	0.7027	0.0422	0.0388	0.0010	0.5520	0.0658	0.0603	0.0027	0.5342	0.0770	0.0702	0.0023	0.3904

further improves the base results of the FUSE pipeline and while considerably increasing the smoothness of the maps. Notably, initializing these refinements methods with FUSE yields better results than using the standard FMaps initialization, particularly evident in the case SHREC20 with strongly non-isometric pairs. Geodesic distances are particularly sensitive to deviations from isometry, and even though corresponding points can be associated with structurally similar distance distributions, actual values may wildly vary. We also observe that, in the more challenging scenario of non-isometric pairs, FUSEANCHOR achieves marginally better accuracy than our approach, at the cost of a considerably higher Dirichlet energy and lower coverage. This suggests that the lack of smoothness in the maps produced by FUSE depends on the flow inversion from the surface to the anchor distribution, which is a non-smooth operation. ULRSSM [11] achieves the best results on FAUST and SMAL, but under performs substantially on SHREC20. As a data-driven method, it requires extensive training on a large dataset, it is restricted to meshes, whereas our flows can be efficiently trained from different representations and easily composed.

Point cloud correspondences. Point cloud matching presents a unique challenge: FUSE requires reliable geodesic distances and sufficient surface coverage, both of which are difficult to guarantee on unstructured data, since the surface information outside of the vertices is unknown. This degrades FUSE’s performance relative to extrinsic methods like NDP [26], which bypasses geodesics entirely at the cost of failing under strong non-rigid deformations (e.g., FAUST). However, FUSE still outperforms non-refined geodesic methods (KNN, OT, FMaps), proving that the flow mechanism itself adds value.

5.2 Alternative Representations

Since FUSE only requires the ability to sample points from a surface and to compute single-to-all geodesic distances, it enables shape matching in a broader class of representations, such as Signed Distance Functions (SDFs), as well as inter-representation matching (see Fig. 7).

Inter-representation matching. We conduct experiments on shape pairs from different FAUST-like collections: from the original FAUST meshes, we extract the vertices to use as point clouds, and we compute SDFs as discussed in the previous paragraph. For each shape, geodesics are computed sticking to the constraints of the representation (*i.e.*, heat geodesics [12] on point clouds, Dijkstra on mesh or voxel grid). As mentioned in Sec. 5.1, FUSE inevitably struggles on matching tasks involving only point clouds, but achieves results comparable to mesh-to-mesh matching when at least one shape is represented as mesh or SDF. We provide an example of matching for all possible combinations of meshes, point clouds and SDFs in Fig. 7. A deeper analysis is presented in the supplementary material.

Table 5: Quantitative evaluation of the SDF-based FAUST dataset without considering the topological error.

Method	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑
KNN	0.0562	0.0569	0.0023	0.1744
NDP	0.0525	0.0643	0.0016	0.5139
OT	0.0921	0.0881	0.0018	0.0936
FUSE ANCHOR	0.0472	0.0469	0.0175	0.2784
FUSE	0.0375	0.0385	0.0069	0.4653

SDFs shape matching. To our knowledge, no dataset exists to evaluate quantitatively matching between SDFs, so we introduce a new benchmark derived from FAUST. For each shape, we train a neural implicit model following [4], and for evaluation, we project the original mesh vertices onto the corresponding SDF surface, ensuring accuracy is always measured on the original mesh geom-

etry. We note that this implicit reconstruction may alter the topology of some shapes. To apply FUSE, we uniformly sample $\sim 10^5$ points from the zero level set and voxelize each SDF surface to compute geodesic distances via Dijkstra. As in Sec. 5.1, these geodesic embeddings are then used to train the flow models. Tab. 5 reports the results for all shapes whose SDFs preserve valid topology.

FUSE provides a matching estimation without pair-specific supervision, additional training, or fine-tuning and with results comparable or better than other methods. While ZO and NAM refinement methods can be applied between meshes and point clouds to greatly improve the matching quality, these methods cannot be used on SDFs where FUSE added value is most significant, as it provides the most accurate maps compared to other baselines.

5.3 Applications

FUSE allows to model 3D surfaces as probability distributions via surface sampling and to compute maps without explicit optimization, effectively making it a “bridge” between distinct representations. We exploit this peculiar property to use FUSE as a backbone for various tasks other than pure shape matching.

Human scans fitting and skinning. Given the scan of a human body, represented as a 3D point cloud $\mathcal{S}$, we use FUSE to match it to a triangle mesh representing the neutral pose $\mathcal{T}$ of a templated body model, obtained with SMPL-X [50]. By exploiting the correspondence, we can easily fit the body model’s parameters to approximate $\mathcal{S}$ via minimization of the MSE between corresponding points. This procedure, whose results we show in the left column of Fig. 8, yields style

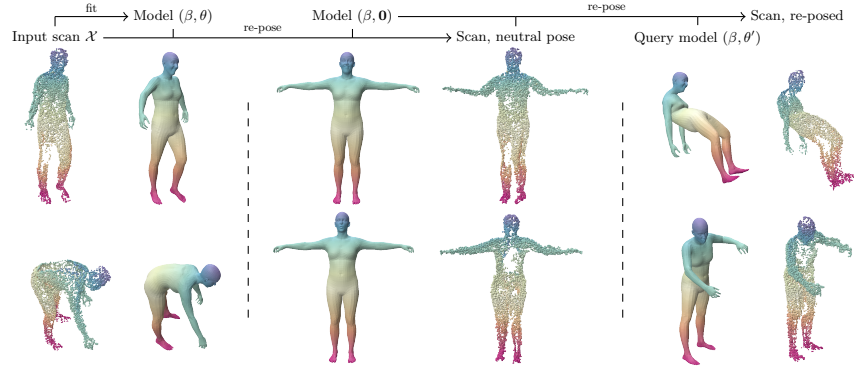


Fig. 8: Results for human-scan fitting and skinning. Input scans identities are preserved in the edited poses (see preserved scan defects on right leg, back and chest). We first optimize β and θ using FUSE correspondences, then re-pose the scan by applying point-wise offsets from the neutral pose.

and pose vectors (β, θ) . If we compute the vertex offsets from the the SMPL-X output for the optimized θ to the neutral pose $\mathcal{T}$, we can apply them “as-is” to the corresponding points in $\mathcal{S}$ to obtain the point cloud in neutral pose. Finally, given a novel pose vector θ' , we can iterate this process to obtain the scan in the required pose. Two example applications of our pipeline are depicted in Fig. 8.

Mesh parametrization. Given a genus-zero surface $\mathcal{S}$ and a reference sphere Φ , we train flows from a Gaussian to each shape using 3D coordinates as embeddings. Composing these flows yields a bijection between $\mathcal{S}$ and Φ , allowing us to transfer the sphere’s UV parametrization to the shape. As FUSE maps may lack smoothness (see Sec. 5.1), transferred UVs are used to supervise an MLP that predicts continuous UV coordinates, enforcing a smooth prior. Since the MLP operates in the embedding space, the procedure applies to any representation supported by FUSE. In this setting, 3D coordinates are sufficient, as no semantic consistency is required. In Fig. 9 we show an example.

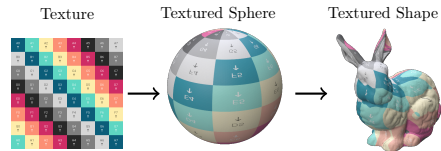


Fig. 9: Illustration of our parametrization via texture mapping.

Volume matching. Our representation-agnostic formulation naturally extends to volumetric shapes, three-dimensional manifolds with well-defined geodesic distances, enabling targeting of volume-matching tasks. We apply FUSE to tetrahedral meshes, comparing it to Volumetric FMaps [33]. Results from Tab. 6 and Fig. 10 mirror our analysis on the surface scenario: comparable or improved accu-

racy, particularly in case of strong non-isometry, where the functional approach struggles. Details on the metrics are in the supplementary materials.

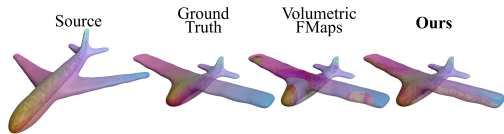


Fig. 10: Qualitative comparison of Ours vs Volumetric Functional Maps on a pair of non isometric tetrahedral meshes.

Table 6: Volumetric correspondence by accuracy and isometricity.

Mesh	Isometry ↓	Accuracy ↓	
		VolFmaps	Ours
Octopus	0.0449	0.0151	0.0215
Cat	0.0561	0.0237	0.0422
Dino_skel	0.0851	0.0652	0.0426
Airplane	0.2207	0.1273	0.0942
Average	0.1017	0.0578	0.0501

6 Conclusions and Limitations

We presented FUSE, a representation for maps between shapes based on Flow Matching. By composing per-shape invertible flows through a shared anchor distribution, FUSE produces accurate, high-coverage correspondences that are agnostic to the underlying geometric representation, supporting meshes, point clouds, SDFs, and volumetric data. Its probabilistic formulation offers an alternative to classical correspondence-based pipelines and provides a concrete starting point for works integrating learned features with flow-based alignment.

FUSE has three main limitations. First, symmetry disambiguation depends entirely on the embedding: geodesic landmarks handle this well, but embeddings lacking symmetry-breaking information (e.g., WKS) lead to degraded results. Second, the reliance on consistent geodesic distances limits its effectiveness on point clouds and makes it unsuitable for partial matching in its current form. Third, the final KNN retrieval step (Sec. 4.2) makes the matching not strictly bijective, as is standard in point-based methods, though the continuity of both the GFE and the flow already supports a continuous solution. Extending FUSE to partial settings [9, 67], enforcing full bijectivity, and integrating learned smooth features are natural directions for future work.

Acknowledgements

Lorenzo Olearo’s PhD is funded by MUR under the grant “Dipartimenti di Eccellenza 2023-2027” of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy. Daniele Baieri’s research is supported by a Humboldt PostDoc Fellowship from the Alexander von Humboldt Foundation and previously by the PRIN project GEOPRIDE, 2022-NAZ-0115, CUP H53D23003400001. This work was supported by the NVIDIA Corporation with the RTX A5000 GPUs granted through the Academic Hardware Grant Program to the University of Milano-Bicocca for the project “Learned representations for implicit binary operations on real-world 2D-3D data.”

References

1. Aigerman, N., Gupta, K., Kim, V.G., Chaudhuri, S., Saito, J., Groueix, T.: Neural jacobian fields: learning intrinsic mappings of arbitrary meshes. *ACM Trans. Graph.* **41**(4) (Jul 2022). <https://doi.org/10.1145/3528223.3530141>, <https://doi.org/10.1145/3528223.3530141>
2. Attaiki, S., Ovsjanikov, M.: Understanding and improving features learned in deep functional maps. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2023)
3. Aubry, M., Schlickewei, U., Cremers, D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. pp. 1626–1633. IEEE (2011)
4. Baieri, D., Maggioli, F., Rodolà, E., Melzi, S., Löhner, Z.: Implicit-ARAP: Efficient handle-guided neural field deformation via local patch meshing. In: *The Thirty-ninth Annual Conference on Neural Information Processing Systems* (2026), <https://openreview.net/forum?id=zp7W2QmxHS>
5. Bastian, L., Xie, Y., Navab, N., Löhner, Z.: Hybrid functional maps for crease-aware non-isometric shape matching. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3313–3323 (June 2024)
6. Bhatnagar, B.L., Xie, X., Petrov, I., Sminchisescu, C., Theobalt, C., Pons-Moll, G.: Behave: Dataset and method for tracking human object interactions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (jun 2022)
7. Bogo, F., Romero, J., Loper, M., Black, M.J.: Faust: Dataset and evaluation for 3d mesh registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3794–3801 (2014)
8. Boscaini, D., Masci, J., Melzi, S., Bronstein, M.M., Castellani, U., Vandergheynst, P.: Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. *Computer Graphics Forum* **34**(5), 13–23 (2015)
9. Bracha, A., Dagès, T., Kimmel, R.: On unsupervised partial shape correspondence. In: Cho, M., Laptev, I., Tran, D., Yao, A., Zha, H. (eds.) *Computer Vision – ACCV 2024*. pp. 316–332. Springer Nature Singapore, Singapore (2025)
10. Cao, D., Eisenberger, M., El Amrani, N., Cremers, D., Bernard, F.: Spectral meets spatial: Harmonising 3d shape matching and interpolation. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3658–3668 (2024). <https://doi.org/10.1109/CVPR52733.2024.00351>
11. Cao, D., Roetzer, P., Bernard, F.: Unsupervised learning of robust spectral shape matching. *ACM Trans. Graph.* **42**(4) (jul 2023). <https://doi.org/10.1145/3592107>, <https://doi.org/10.1145/3592107>
12. Crane, K., Weischedel, C., Wardetzky, M.: The heat method for distance computation. *Commun. ACM* **60**(11), 90–99 (Oct 2017). <https://doi.org/10.1145/3131280>, <http://doi.acm.org/10.1145/3131280>
13. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. p. 2292–2300. NIPS’13, Curran Associates Inc., Red Hook, NY, USA (2013)
14. Donati, N., Corman, E., Ovsjanikov, M.: Deep orientation-aware functional maps: Tackling symmetry issues in shape matching. *CVPR* (2022)
15. Dutt, N.S., Muralikrishnan, S., Mitra, N.J.: Diffusion 3d features (diff3f): Decorating untextured shapes with distilled semantic features. In: *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4494–4504 (June 2024)
16. Dyke, R.M., Lai, Y.K., Rosin, P.L., Zappalà, S., Dykes, S., Guo, D., Li, K., Marin, R., Melzi, S., Yang, J.: Shrec20: Shape correspondence with non-isometric deformations. *Computers & Graphics* **92**, 28–43 (2020). <https://doi.org/https://doi.org/10.1016/j.cag.2020.08.008>, <https://www.sciencedirect.com/science/article/pii/S0097849320301266>
 17. Eisenberger, M., Novotny, D., Kerchenbaum, G., Labatut, P., Neverova, N., Cremers, D., Vedaldi, A.: Neuromorph: Unsupervised shape interpolation and correspondence in one go. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7473–7483 (2021)
 18. Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boissunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N.T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D.J., Tavenard, R., Tong, A., Vayer, T.: Pot: Python optimal transport. *Journal of Machine Learning Research* **22**(78), 1–8 (2021), <http://jmlr.org/papers/v22/20-451.html>
 19. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: 3d-coded: 3d correspondences by deep deformation. In: *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part II*. p. 235–251. Springer-Verlag, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-030-01216-8_15, https://doi.org/10.1007/978-3-030-01216-8_15
 20. Hartman, E., Pierson, E., Bauer, M., Charon, N., Daoudi, M.: Bare-esa: A riemannian framework for unregistered human body shapes. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 14135–14145 (2023). <https://doi.org/10.1109/ICCV51070.2023.01304>
 21. Jiang, P., Sun, M., Huang, R.: Non-rigid shape registration via deep functional maps prior. In: *Thirty-seventh Conference on Neural Information Processing Systems (2023)*, <https://openreview.net/forum?id=XvfEYqEbIb>
 22. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. In: *ACM Transactions on Graphics (TOG)*. vol. 30, p. 79. ACM (2011)
 23. Koestler, L., Grittner, D., Moeller, M., Cremers, D., Löhner, Z.: Intrinsic neural fields: Learning functions on manifolds. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) *Computer Vision – ECCV 2022*. pp. 622–639. Springer Nature Switzerland, Cham (2022)
 24. Le, T., Nguyen, K., Sun, S., Ho, N., Xie, X.: Integrating efficient optimal transport and functional maps for unsupervised shape correspondence learning. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 23188–23198 (2024). <https://doi.org/10.1109/CVPR52733.2024.02188>
 25. Lévy, B.: Laplace-Beltrami eigenfunctions towards an algorithm that understands geometry. In: *Proc. SMI*. pp. 13–25. IEEE, Washington, DC (2006)
 26. Li, Y., Harada, T.: Non-rigid point cloud registration with neural deformation pyramid. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems. NIPS ’22*, Curran Associates Inc., Red Hook, NY, USA (2022)
 27. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747* (2022)
 28. Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R.T., Lopez-Paz, D., Ben-Hamu, H., Gat, I.: Flow matching guide and code. *arXiv preprint arXiv:2412.06264* (2024)

29. Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9298–9309 (October 2023)
30. Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.Y., Russell, B.: Editing conditional radiance fields. In: Proceedings of the International Conference on Computer Vision (ICCV) (2021)
31. Liu, Y., Zhu, L., Yamada, M., Yang, Y.: Semantic correspondence as an optimal transport problem. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4463–4472 (2020)
32. Maggioli, F., Baieri, D., Rodolà, E., Melzi, S.: Rematching: Low-resolution representations for scalable shape correspondence. In: European Conference on Computer Vision. pp. 183–200. Springer (2025)
33. Maggioli, F., Livesu, M., Melzi, S.: Volumetric functional maps (2026)
34. Magnet, R., Bloch, K., Taverne, M., Melzi, S., Geoffroy, M., Khonsari, R.H., Ovsjanikov, M.: Assessing craniofacial growth and form without landmarks: A new automatic approach based on spectral methods. *Journal of Morphology* **284**(8), e21609 (2023). <https://doi.org/https://doi.org/10.1002/jmor.21609>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmor.21609>
35. Magnet, R., Ren, J., Sorkine-Hornung, O., Ovsjanikov, M.: Smooth non-rigid shape matching via effective dirichlet energy optimization. In: 2022 International Conference on 3D Vision (3DV). IEEE (sep 2022). <https://doi.org/10.1109/3dv57658.2022.00061>, <https://doi.org/10.1109/3dv57658.2022.00061>
36. Marin, R., Melzi, S., Rodolà, E., Castellani, U.: Farm: Functional automatic registration method for 3d human bodies. *Computer Graphics Forum* **39**(1), 160–173 (2020). <https://doi.org/https://doi.org/10.1111/cgf.13751>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13751>
37. Marin, R., Corona, E., Pons-Moll, G.: Nicp: Neural icp for 3d human registration at scale. In: European Conference on Computer Vision (2024)
38. Marin, R., Melzi, S., Musoni, P., Bardon, F., Tarini, M., Castellani, U.: CMH: Coordinates Manifold Harmonics for Functional Remeshing. In: Biasotti, S., Lavoué, G., Veltkamp, R. (eds.) *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association (2019). <https://doi.org/10.2312/3dor.20191063>
39. Marin, R., Rakotosaona, M.J., Melzi, S., Ovsjanikov, M.: Correspondence learning via linearly-invariant embedding. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS '20, Curran Associates Inc., Red Hook, NY, USA (2020)
40. Melzi, S., Ovsjanikov, M., Roffo, G., Cristani, M., Castellani, U.: Discrete time evolution process descriptor for shape analysis and matching. *ACM Transactions on Graphics (TOG)* **37**(1), 4:1–4:18 (Jan 2018)
41. Melzi, S., Ren, J., Rodolà, E., Sharma, A., Wonka, P., Ovsjanikov, M.: Zoomout: Spectral upsampling for efficient shape correspondence. *ACM Transactions on Graphics (TOG)* **38**(6), 155:1–155:14 (Nov 2019)
42. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2019)
43. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
44. Mémoli, F., Sapiro, G.: Fast computation of weighted distance functions and geodesics on implicit hyper-surfaces. *Journal of Computational Physics* **173**(2),

- 730–764 (2001). <https://doi.org/https://doi.org/10.1006/jcph.2001.6910>, <https://www.sciencedirect.com/science/article/pii/S0021999101969104>
45. Novello, T., da Silva, V., Schardong, G., Schirmer, L., Lopes, H., Velho, L.: Neural implicit surface evolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14279–14289 (October 2023), https://openaccess.thecvf.com/content/ICCV2023/html/Novello_Neural_Implicit_Surface_Evolution_ICCV_2023_paper.html
 46. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)* **31**(4), 30:1–30:11 (2012)
 47. Pai, G., Ren, J., Melzi, S., Wonka, P., Ovsjanikov, M.: Fast sinkhorn filters: Using matrix scaling for non-rigid shape correspondence with functional maps. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 384–393. IEEE Computer Society, Los Alamitos, CA, USA (jun 2021)
 48. Pan, S., Brunton, S.L., Kutz, J.N.: Neural implicit flow: a mesh-agnostic dimensionality reduction paradigm of spatio-temporal data. *J. Mach. Learn. Res.* **24**(1) (Jan 2023)
 49. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
 50. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3D hands, face, and body from a single image. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 10975–10985 (2019)
 51. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. In: The Eleventh International Conference on Learning Representations (2023), <https://openreview.net/forum?id=FjNys5c7VyY>
 52. Raganato, A., Pasi, G., Melzi, S.: Attention And Positional Encoding Are (Almost) All You Need For Shape Matching. *Computer Graphics Forum* (2023). <https://doi.org/10.1111/cgf.14912>
 53. Ramachandran, P., Zoph, B., Le, Q.V.: Searching for activation functions. arXiv preprint arXiv:1710.05941 (2017)
 54. Saleh, M., Wu, S.C., Cosmo, L., Navab, N., Busam, B., Tombari, F.: Bending graphs: Hierarchical shape matching using gated optimal transport. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11747–11757 (2022). <https://doi.org/10.1109/CVPR52688.2022.01146>
 55. Song, C., Wei, J., Li, R., Liu, F., Lin, G.: 3d pose transfer with correspondence learning and mesh refinement. In: Thirty-Fifth Conference on Neural Information Processing Systems (2021)
 56. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021), <https://openreview.net/forum?id=St1giarCHLP>
 57. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. *Computer graphics forum* **28**(5), 1383–1392 (2009)
 58. Taubin, G.: A signal processing approach to fair surface design. In: Proc. CGIT. pp. 351–358. ACM, New York, NY (1995)
 59. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Proc. ECCV. pp. 356–369. Springer (2010)

60. Trappolini, G., Cosmo, L., Moschella, L., Marin, R., Melzi, S., Rodolà, E.: Shape registration in the time of transformers. *Advances in Neural Information Processing Systems* **34**, 5731–5744 (2021)
61. Vallet, B., Lévy, B.: Spectral geometry processing with manifold harmonics. *Computer Graphics Forum* **27**(2), 251–260 (2008)
62. Viganò, G., Ovsjanikov, M., Melzi, S.: Nam: Neural adjoint maps for refining shape correspondences. *ACM Trans. Graph.* **44**(4) (Jul 2025). <https://doi.org/10.1145/3730943>, <https://doi.org/10.1145/3730943>
63. Viganò, G., Longari, G., Pereira, L., Miolane, N., Melzi, S.: Geomfum: A python package for machine learning with functional maps (09 2025). <https://doi.org/10.13140/RG.2.2.31903.01441>
64. Viganò, G., Melzi, S.: Bijective upsampling and learned embedding for point clouds correspondences. *Computers and Graphics* **122**, 103985 (2024). <https://doi.org/https://doi.org/10.1016/j.cag.2024.103985>, <https://www.sciencedirect.com/science/article/pii/S0097849324001201>
65. Wang, C., Chai, M., He, M., Chen, D., Liao, J.: Clip-nerf: Text-and-image driven manipulation of neural radiance fields. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 3825–3834 (2021), <https://api.semanticscholar.org/CorpusID:245006265>
66. Williamson, R., Mitra, N.J.: Neural geometry processing via spherical neural surfaces. *Computer Graphics Forum* **44**(2), e70021 (2025). <https://doi.org/https://doi.org/10.1111/cgf.70021>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.70021>
67. Xie, Y., Ehm, V., Roetzer, P., El Amrani, N., Gao, M., Bernard, F., Cremers, D.: Echomatch: Partial-to-partial shape matching via correspondence reflection. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 11665–11675 (2025)
68. Yang, G., Belongie, S., Hariharan, B., Koltun, V.: Geometry processing with neural fields. In: *Thirty-Fifth Conference on Neural Information Processing Systems* (2021)
69. Zhang, B., Ren, J., Wonka, P.: Geometry distributions. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 1495–1505 (October 2025)
70. Zuffi, S., Kanazawa, A., Jacobs, D., Black, M.J.: 3D menagerie: Modeling the 3D shape and pose of animals. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017*. pp. 5524–5532. IEEE, Piscataway, NJ, USA (Jul 2017)

Table 7: Full quantitative results for all 9 cases of inter-representation shape matching. Best results per source-target pair are highlighted in **bold** and second best are underlined.

Source Representation Method		Target Representation: Mesh				Target Representation: PointCloud				Target Representation: SDF			
		Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑
Mesh	KNN	0.0454	0.0434	0.0022	0.2452	0.1480	0.1404	<u>0.0016</u>	0.0254	0.0543	0.0549	<u>0.0022</u>	0.1842
	OT [13]	0.0926	0.0856	0.0016	0.0890	<u>0.0747</u>	<u>0.0722</u>	0.0028	0.0472	0.0965	0.0910	0.0017	0.0945
	NDP [26]	0.0702	0.0886	<u>0.0020</u>	0.4893	0.0535	0.0643	0.0016	0.4828	0.0572	0.0712	0.0017	0.5114
	KNN-IN-GAUSS	<u>0.0425</u>	<u>0.0390</u>	0.0241	0.3483	0.1121	0.1047	0.0395	0.2262	<u>0.0505</u>	<u>0.0494</u>	0.0247	0.3107
	Ours	0.0289	0.0274	0.0066	0.5320	0.0957	0.0890	0.0184	<u>0.3761</u>	0.0373	0.0375	0.0081	<u>0.4816</u>
Point Cloud	KNN	0.1580	0.1508	0.0162	0.0129	0.0695	0.0685	0.0034	0.1540	0.1533	0.1465	0.0153	0.0117
	OT [13]	0.1662	0.1556	0.0145	0.0339	0.0836	0.0802	<u>0.0025</u>	0.1156	0.1629	0.1533	0.0142	0.0378
	NDP [26]	0.0547	0.0650	0.0017	0.4813	0.0530	0.0638	0.0018	0.4648	0.0578	<u>0.0715</u>	0.0018	0.4753
	KNN-IN-GAUSS	0.0805	0.0757	0.0300	0.1823	0.0650	<u>0.0639</u>	0.0197	0.2333	<u>0.0645</u>	0.0643	0.0123	<u>0.3995</u>
	Ours	<u>0.0695</u>	<u>0.0680</u>	<u>0.0129</u>	<u>0.3725</u>	<u>0.0645</u>	0.0643	0.0123	<u>0.3995</u>	0.0730	0.0716	<u>0.0107</u>	0.3045
SDF	KNN	0.0554	0.0554	0.0020	0.1697	0.1451	0.1387	0.0016	0.0254	0.0562	0.0569	0.0023	0.1744
	OT [13]	0.0916	0.0873	<u>0.0019</u>	0.0851	<u>0.0751</u>	<u>0.0735</u>	<u>0.0033</u>	0.0489	0.0921	0.0881	<u>0.0018</u>	0.0936
	NDP [26]	0.0530	0.0627	0.0015	0.5085	0.0535	0.0630	0.0016	0.4691	0.0525	0.0643	0.0016	0.5139
	KNN-IN-GAUSS	<u>0.0475</u>	<u>0.0465</u>	0.0194	0.2706	0.1051	0.0979	0.0204	0.1612	<u>0.0472</u>	<u>0.0469</u>	0.0175	0.2784
	Ours	0.0358	0.0361	0.0061	<u>0.4773</u>	0.1004	0.0932	0.0102	<u>0.2680</u>	0.0375	0.0385	0.0069	<u>0.4653</u>

1 Additional FUSE Details

1.1 Implementation Details

To implement the flows, we use the open-source `flow_matching` library [28] and parametrize the vector field using a simple neural architecture. Specifically, we employ a 256-dimensional, 6-layer MLP with Swish activations [53], conditioned on both the input features $x \in \mathbb{R}^d$ and a scalar time variable t . Before input to the MLP, the features x and t are concatenated and processed through a Fourier feature encoder with six frequencies, resulting in $(d+1)(6 \times 2) + (d+1)$ features. A final linear layer projects the hidden representation back to the original input dimension d . Optimization is carried out similarly as in [69]. We trained on a single NVIDIA A5000 GPU for 10^4 epochs, sampling $5 \cdot 10^4$ points per epoch uniformly from the surface of each input shape. This setup requires less than 2 GB of GPU memory and takes approximately 2 minutes per shape. Our implementation will be made publicly available upon publication.

Comparing our setup to the one proposed in [69], our choices drastically reduce computation time for training a single flow. While some design decisions in GeomDist can be advantageous for learning high-quality representations of highly detailed input shapes, our approach optimizes representations that are sufficiently accurate for our needs within minutes instead of the several hours reported in the original GeomDist implementation.

1.2 Further Details on the Comparison Between FM and DDIM

In this section, we show additional details highlighting the differences between our method, built upon the FM framework, and the equivalent implementation with DDIM, which is the basis of GeomDist [69].

GeomDist (GD) is based on a Denoising Diffusion Implicit Model (DDIM) framework. In this setting, the noisy data X_t evolves according to the discretized

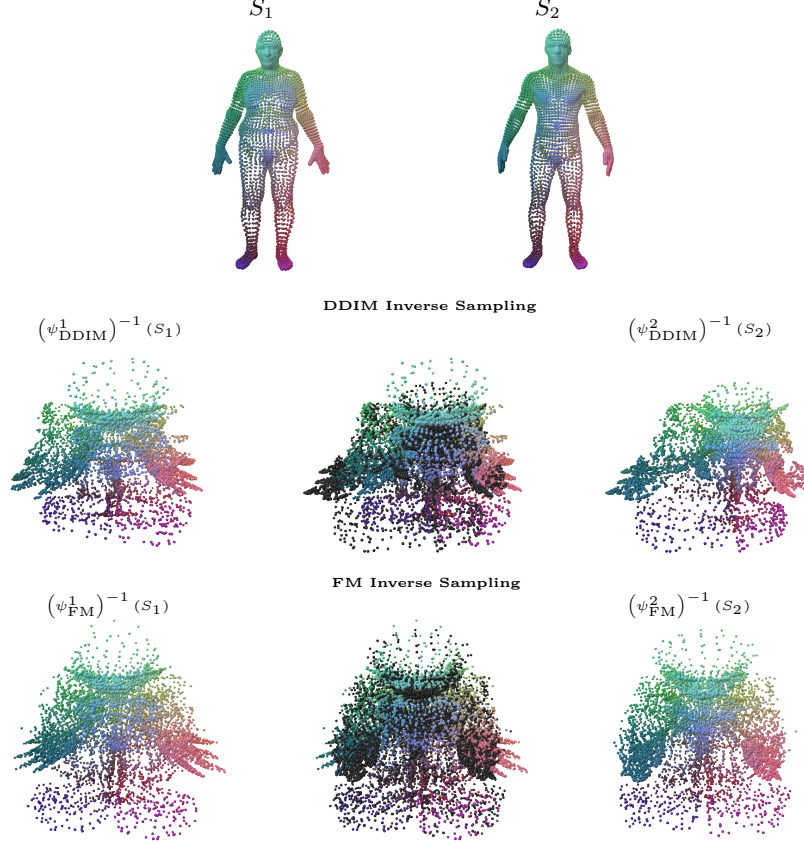


Fig. 11: Visualization of differences during inverse sampling between Flow Matching (FM) and DDIM. The first row shows two shapes S_1 and S_2 . The second row shows the result of the inverse sampling via DDIM for the two shapes. The column in the middle shows the aligned visualization to highlight the differences. The last row shows the corresponding inverse sampling results using Flow Matching. We observe a notable difference in the alignment of the two inverted shapes in the case of DDIM, which is quantitatively analyzed in Table 3. This difference highlights the improved accuracy and stable anchoring properties of FM compared to DDIM.

update:

$$X_t = X_0 + t\epsilon, \quad (4)$$

where X_0 is the clean shape, $\epsilon \sim \mathcal{N}(0, I)$ is Gaussian noise. We note that the Gaussian noise only dominates the data distribution asymptotically, as t grows large. This implies that the latent isotropic Gaussian distribution is reached only for $t \rightarrow \infty$, as opposed to the finite-time interpolation guaranteed by Flow Matching. Thus, GD's inversion map asymptotically approximates the Gaussian

Table 8: Average inference computational time (in seconds) and euclidean error for different methods on FAUST dataset.

Method	Ours	Ours-ZO	Ours-NAM	fmap-wks	fmap	fmap-ZO	fmap-NZO	KNN-IN-GAUSS	ndp	ot	knn
time (s) ↓	2.50	4.27	26.59	25.80	0.57	7.95	39.26	2.37	3.36	1.47	0.05
euclidean error ↓	0.0266	0.0093	0.0096	0.0270	0.0436	0.0097	0.0143	0.0393	0.0616	0.0862	0.0608

prior, leading to deviations from the anchor distribution and hindering anchor space similarities between different shapes.

This difference underpins the improved accuracy and tighter distributional alignment we observe in our flow-matching-based method. Indeed, FUSE’s interpolation can be written as

$$X_t = (1 - t)X_0 + tX_1, \quad (5)$$

where X_1 is the clean shape and $X_0 \sim \mathcal{N}(0, I)$ is Gaussian noise. We note that in this case, we have a clear notion of the time at which we sample one of the two distributions, maintaining consistency between different flows. In Fig. 11, we show a qualitative comparison between two three-dimensional flows. Inverting the source and target flows to their Gaussian-like base distributions reveals a subtle discrepancy in the DDIM case. Under a Flow-Matching noise schedule, however, this discrepancy largely disappears. For visualization purposes, the inverted target is shown in black while the inverted source remains colored.

2 Evaluation

Datasets. We evaluate our method on well-established datasets for non-rigid shape matching:

- **FAUST** (meshes): a dataset of near-isometric human subjects in different poses [7] in a 1:1 correspondence.
- **SMAL** (meshes): composed of different classes of quadrupeds in varying poses with strongly non-isometric deformations [70]; we use the remeshed version of the dataset [11], where shapes have different discretizations.
- **SHREC20** (meshes): 14 animal shapes including highly non-isometric pairs and shapes with missing parts [16]. No dense registration is provided, each pair is annotated with sparse set of landmarks; we consider all the pairs of the 5 test set and evaluate on the common landmarks between each shape pair.
- **KINECT** (point clouds): Kinect acquisitions of 15 human shapes interacting with objects from the BEHAVE dataset [6] with SMPL registrations, introduced in [62].

Landmarks are placed manually at shape protrusions: 5 for humans and 6 for animals. For SHREC20 specifically, 6 points are selected from the pool of common points shared across all shapes, with evaluation performed on the remaining common point pairs. Fig. 12 illustrates the landmarks on a representative shape from each dataset.

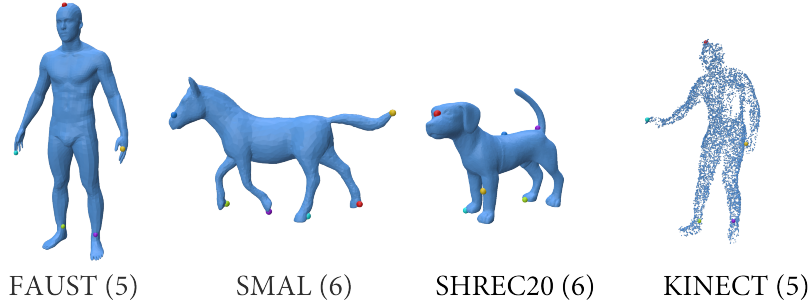


Fig. 12: Landmarks used to compute geodesic embeddings (number per dataset in parentheses), hand-picked on one shape of the dataset and transferred to others via ground truth correspondences. Note that SHREC20 landmarks are selected from the small set of points common to all shapes in the datasets.

Baselines. For Functional Maps-based baselines, we used implementations from the GeomFuM library [63]. For NDP [26], we adopted the original source code. For volumetric FMaps [33], we considered the original MATLAB source code. For FMaps and FMaps-WKS, maps were computed using 20 basis functions. For the WKS variant, we leveraged 200 basis functions to compute descriptors, using both landmark-based and global descriptors as standard. ZoomOut and NAM [62] were implemented by performing spectral upsampling from 20 to 120 basis functions with steps of 5 basis functions. Optimal Transport (OT) computations were performed using the Sinkhorn relaxation method [13] available via the PythonOptimalTransport library [18].

2.1 Evaluation Metrics

In this Section, we provide details on how we computed all evaluation metrics employed throughout the paper. The Euclidean error is computed as:

$$E_{\text{eucl}} = \frac{1}{|S_1|} \sum_{x \in S_1} \frac{\|T_{12}(x) - T_{\text{gt}}(x)\|_2}{D_{\text{eucl}}} \quad (6)$$

where $D_{\text{eucl}} = \max_{u,v \in S_2} \|u - v\|_2$. Similarly, the geodesic error is defined as:

$$E_{\text{geod}} = \frac{1}{|S_1|} \sum_{x \in S_1} \frac{d(T_{12}(x), T_{\text{gt}}(x))}{D} \quad (7)$$

where d is the geodesic distance function on S_2 , T_{12} is the estimated correspondence map, T_{gt} is the ground truth map, and $D = \max_{u,v \in S_1} d(u, v)$ is the shape diameter of S_2 .

The Dirichlet energy of the correspondence map is computed as:

$$E_{\text{Dir}} = \text{trace}(T_{12}(X_1)^\top \mathbf{A} T_{12}(X_1)) \quad (8)$$

where $\mathbf{A}$ and X_1 are respectively the Laplacian matrix and the vertex coordinates matrix associated with the source shape S_1 .

Coverage measures the fraction of the target shape *covered* by the mapping. Importantly, we normalize this measure by the cardinality ratio $|S_2|/|S_1|$ to ensure comparability across shapes with varying numbers of vertices:

$$\text{Cov} = \frac{|\{y \in S_2 : \exists x \in S_1, T_{12}(x) = y\}|}{|S_2|} \cdot \frac{|S_2|}{|S_1|} \quad (9)$$

The isometry metric evaluates the relative difference between pairwise geodesic distances on tetrahedra that are in ground truth correspondence as:

$$E_{\text{iso}} = \sum_{x,y} \frac{|d_1(x,y) - d_2(T_{\text{gt}}(x), T_{\text{gt}}(y))|}{d_1(x,y)} \quad (10)$$

where d_1 and d_2 are geodesic distances on the source and target shapes respectively.

2.2 Evaluation Metrics Distribution

For a better visualization of the numerical results reported in Tab. 4 of the main paper, we provide in Fig. 13 the distribution of per-pair Euclidean error, Geodesic error, Dirichlet energy, and Coverage for each method across all four benchmarks.

2.3 Inter-Representation Quantitative Results

We follow the setting described in Sec. 5.2 to evaluate our method in the context of inter-representation matching. Out of all the baselines considered in Sec. 5.1, we only take into account those suitable for inter-representation matching, namely KNN, Optimal Transport (OT) [13] and NDP [26], together with FUSE and KNN in the Gaussian anchor space (KNN-IN-GAUSS).

Tab. 7 reports the full quantitative results for all the 9 inter-representation matching scenarios considered in Fig. 7 of the main paper. Our method stands out in terms of Euclidean and Geodesic errors when considering matching between well-defined surfaces, such as meshes and SDFs. Consistent with the pointcloud-to-pointcloud matching results on the KINECT dataset reported in Tab. 4, FUSE struggles when matching to or from point clouds, where NDP [26] instead achieves the best results across all metrics, likely due to the shared information at inference time. Furthermore, our method relies on accurate geodesic distance computations and uniform sampling of the surface, both of which are challenging to achieve on point clouds. Nonetheless, FUSE still achieves the second-best performance across most metrics in these scenarios.

3 Smoothness

In this section, we provide additional details and analysis on why smooth features are necessary for the functioning of the FUSE. Firstly, we demonstrate the effect that varying the smoothness of a feature has on the performance of FUSE by providing an example. Secondly, we demonstrate that smoothing embeddings enables FUSE to achieve better results even with high dimensional pretrained embeddings. We define the $*$ operation as the spectral filtering operation. Given an embedding E , the spectral filter of $E*$ can be obtained as follows:

$$E* = \Phi_k \Phi_k^T M E \quad (11)$$

where Φ_k is the k truncated basis matrix obtained by the decomposition of the Laplace-Beltrami Operator of the input shape [61]. This operation is the same as that used in the main paper in Table 2, with $k = 10$.

Controlled experiment on smoothness. To visualize the effect of smoothness on the effectiveness of FUSE, we consider a pair of shapes from the FAUST dataset and compute their Wave Kernel Signature embeddings. We then perform spectral filtering while varying the number of eigenfunctions used. Increasing the number of eigenfunctions decreases the smoothness of the embedding and increases its Dirichlet energy.

Using these embeddings, we compute correspondences via KNN and FUSE, which highlights a distinctive behavior of our method. In Figure 14, we visualize this experiment. We observe that FUSE improves the results when the embeddings are smooth, whereas for high-frequency embeddings, it degrades performance with respect to KNN. Interestingly, the combination of smoothing and FUSE yields the overall best results.

Smoothed pretrained embeddings. We then consider pretrained embeddings obtained with the method of [11]. We show that applying the smoothing operation with $k = 10$ improves the overall performance.

In Table 9, we can clearly see that smoothing combined with FUSE yields better results than the raw pretrained embeddings. This observation motivates future work on integrating FUSE directly into the pretraining phase of embedding learning, and it is consistent with previously documented benefits of smoothness in learned representations [2].

Table 9: Euclidean error on FAUST with original and modified embeddings computed using a pretrained feature extractor.

	Original	*
KNN	0.072	0.074
FUSE	0.082	0.068
Dirichlet	0.1847	0.0016

3.1 Computational Cost Analysis

In Flow Matching, the vector field is parameterized using a neural network. Consequently, beyond the computational cost of evaluating the network itself, one

Table 10: Results on the FAUST dataset when using different multidimensional anchor distributions for FUSE. We report the average euclidean error (Eucl.), geodesic error (Geod.), Dirichlet energy (Dirich.) errors, and coverage (Cov.). We highlight in **bold** choice of distribution used in the main paper.

Anchor distribution	Eucl. ↓	Geod. ↓	Dirich. ↓	Cov. ↑
<i>Standard Gaussian</i>	0.0289	0.0274	0.0066	0.5320
<i>Unit Sphere Volume</i>	0.0292	0.0290	0.0094	0.5877
<i>Unit Volume</i>	0.0319	0.0318	0.0146	0.5840
<i>Unit Sphere Surface</i>	0.2654	0.2702	0.2477	0.3137

must also take into account the numerical integration of the field to obtain the actual flow. The computational cost of this process depends on both the choice of solver and the number of integration steps employed. In Tab. 11, we present a comparison between Euler and Midpoint solvers with varying numbers of integration steps, applied to both the forward and inverse flow processes on the FAUST dataset. Notably, the accuracy of the resulting maps is predominantly influenced by the precision of the forward flow integration, whereas lower integration precision in the inverse flow has a less pronounced impact. As expected, the Midpoint solver with a larger number of integration steps achieves better accuracy than naive Euler integration, albeit at the expense of increased computational time. Conversely, employing fewer steps with Midpoint or using a simpler solver such as Euler accelerates inference but results in a measurable reduction in mapping accuracy.

3.2 FUSE with Different Anchors

A shape can be represented as a flow starting from any distribution from which samples can be drawn. Naturally, the choice of this starting distribution can influence the quality of the final maps obtained via the flow composition employed in FUSE. In our experiments we used a standard Gaussian distribution as the anchor, however, alternative choices are not only possible but may also be better suited for other applications beyond the scope of this work.

The fundamental requirement for the anchor distribution in FUSE is robustness with respect to flow inversion from the surface of the shape to itself. In Fig. 15, we provide a qualitative comparison of different starting distributions in $\mathbb{R}^3$ for visualization purposes, while Tab. 10 reports corresponding quantitative results on the FAUST dataset with different multidimensional distributions. As illustrated in Fig. 15, sampling from the initial distribution X_0 and subsequently applying the flow $\psi(X_0)$ produces a shape distribution that deviates slightly from the target shape surface on which the flow was trained. By performing a backward integration of the flow, $\psi(\psi(X_0))^{-1}$, we recover an approximation of the identity map, with discrepancies coming from the non-smoothness of the mapping from volume to surface of the inverted flow.

Table 11: Comparison between Euler and Midpoint solvers with different numbers of integration steps for the backward (bwd-steps) and forward (fwd-steps) flows on the FAUST dataset. We report average Euclidean error (Eucl.), geodesic error (Geod.), Dirichlet energy (Dir.), coverage (Cov.), and elapsed time in seconds (Time). The configuration used in the main paper is highlighted in **bold**.

Solver	bwd-steps	fwd-steps	Eucl. ↓	Geod. ↓	Dir. ↓	Cov. ↑	Time (s) ↓
Midpoint	10	10	0.0419	0.0428	0.0072	0.2933	0.41
Midpoint	10	64	0.0344	0.0349	0.0070	0.4259	1.43
Midpoint	64	10	0.0438	0.0450	0.0084	0.3527	1.46
Midpoint	64	64	0.0289	0.0274	0.0066	0.5320	2.50
Midpoint	32	32	0.0337	0.0340	0.0066	0.4692	1.26
Midpoint	16	16	0.0368	0.0377	0.0070	0.3669	0.64
Euler	64	64	0.0364	0.0360	0.0075	0.3945	1.30
Euler	10	64	0.0756	0.0700	0.0133	0.2479	0.74
Euler	64	10	0.0827	0.0778	0.0065	0.1954	0.75
Euler	32	32	0.0514	0.0496	0.0092	0.2791	0.64

In contrast, when sampling directly from the surface of the shape and applying the inverse flow to project these points back to the anchor distribution, we observe a more pronounced deviation from the expected anchor distribution. This behavior is particularly evident in Fig. 2 and Fig. 11, where the distributions obtained via the inverted flow do not exhibit Gaussian-like characteristics. We attribute this phenomenon to the fact that the sampled points occupy only a subset of the shape’s distribution parametrized by the flow; consequently, the resulting inverted distribution corresponds to a subset of the anchor distribution.

To avoid smoothness issues during the inversion of a surface into a volume, one could consider using an initial distribution that is naturally defined on a lower-dimensional manifold, such as the surface of the unit sphere, thus having true diffeomorphic mappings between the shape and the anchor. However, as shown in Tab. 10 and Fig. 15, this choice leads to significantly worse results compared to volumetric anchors. This is likely due to the challenges associated with accurately learning flows on lower-dimensional manifolds embedded in higher-dimensional spaces, which can result in instabilities during both training and inference.



Fig. 13: Distribution of Euclidean and Geodesic error, Dirichlet energy, and Coverage for each method on FAUST, SMAL, SHREC20, and KINECT. Each box spans the inter quartile range; the central line is the median while individual points are outliers. Methods left of the dashed line are ours.

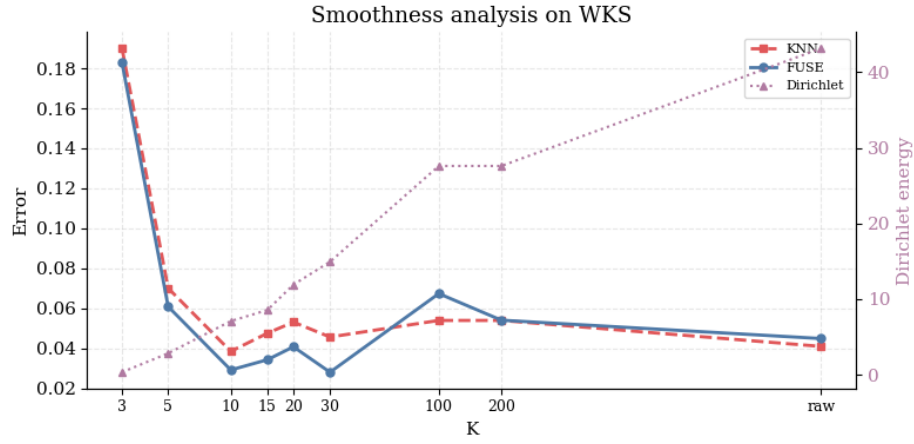


Fig. 14: Influence of the number of eigenfunctions used to perform low-pass filtering on a pair of shapes from FAUST using the WKS descriptors. Comparing KNN and FUSE across different smoothing levels, we observe that FUSE outperforms the baseline for smooth embeddings but underperforms for high-frequency embeddings. We also report the unnormalized Dirichlet energy for each feature used.

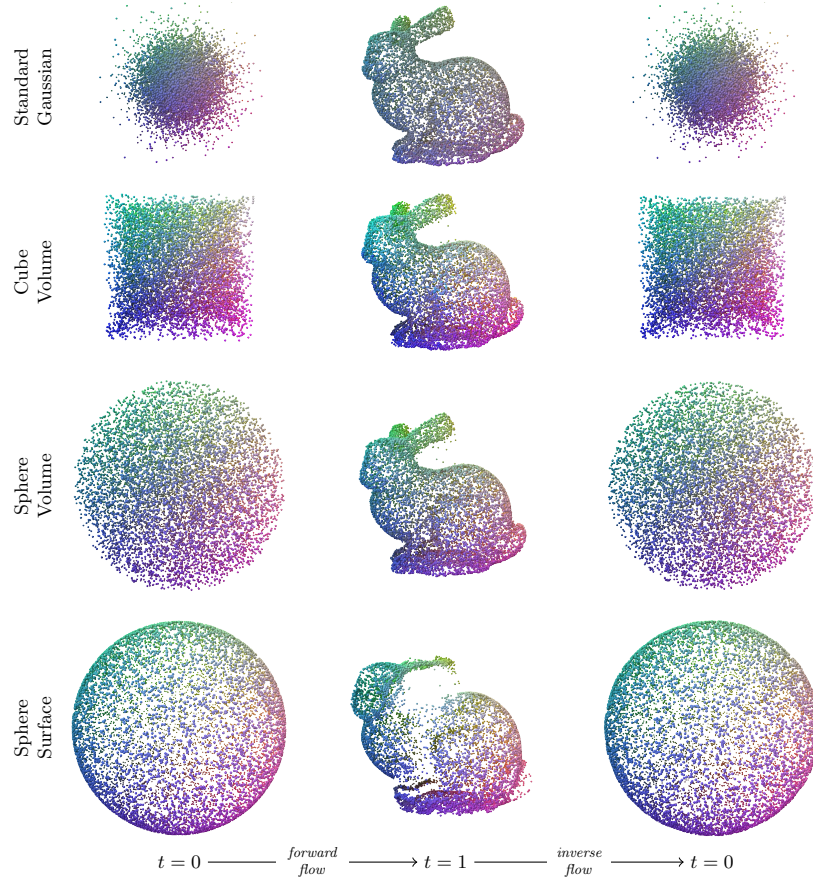


Fig. 15: FUSE with different anchor distributions. Each row shows a flow starting from a different initial distribution at $t=0$, reaching the surface of the Stanford bunny at $t=1$, and then back to the initial distribution at $t=0$ by inverting the flow. For visual purposes, we show samples in $\mathbb{R}^3$; however, in practice, FUSE uses multidimensional anchors as described in the main paper.