This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

DiffuMatch: Category-Agnostic Spectral Diffusion Priors for Robust Non-rigid Shape Matching

Emery Pierson[1] Lei Li[2] Angela Dai[2] Maks Ovsjanikov[1]
LIX, Ecole Polytechnique[1] Technical University of Munich[2]

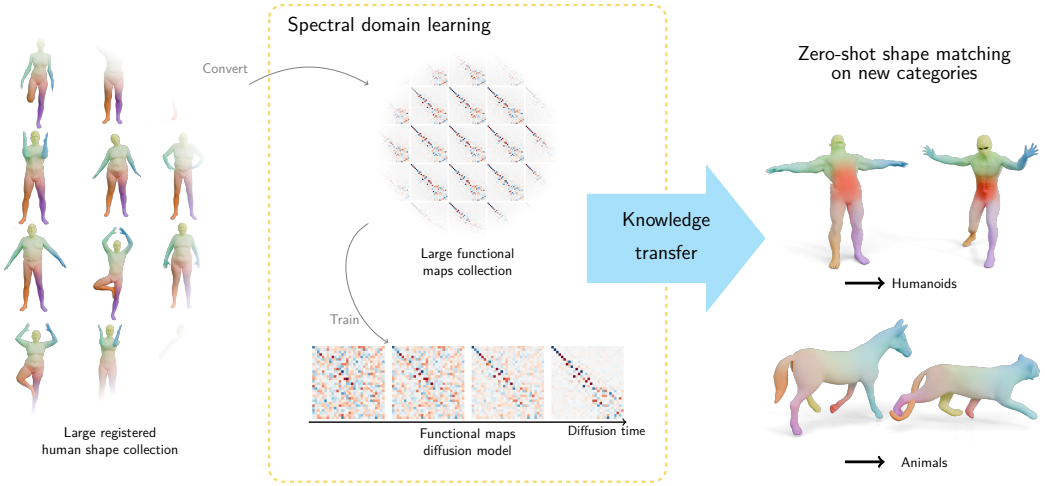


Figure 1. We learn diffusion priors in the spectral domain from a large collection of functional maps computed on registered human shapes. The learned spectral diffusion priors are category-agnostic and generalize robustly to unseen shape categories, enabling accurate zero-shot non-rigid shape matching.

Abstract

Deep functional maps have recently emerged as a powerful tool for solving non-rigid shape correspondence tasks. Methods that use this approach combine the power and flexibility of the functional map framework, with data-driven learning for improved accuracy and generality. However, most existing methods in this area restrict the learning aspect only to the feature functions and still rely on axiomatic modeling for formulating the training loss or for functional map regularization inside the networks. This limits both the accuracy and the applicability of the resulting approaches only to scenarios where assumptions of the axiomatic models hold. In this work, we show, for the first time, that both in-network regularization *and functional map training can be replaced with data-driven methods. For this, we first train a generative model of functional maps* in the spectral domain *using score-based generative modeling, built from a large collection of high-quality maps. We then exploit the resulting model to promote the structural properties of ground truth functional maps on new shape collections. Remarkably, we demonstrate that the learned models are category-agnostic, and can fully replace commonly used strategies such as enforcing Laplacian commutativity or orthogonality of functional maps. Our key technical contribution is a novel distillation strategy from diffusion models in the spectral domain. Experiments demonstrate that our learned regularization leads to better results than axiomatic approaches for zero-shot non-rigid shape matching. Our code is available at:* https://github.com/daidedou/diffumatch/

1. Introduction

Shape matching is a fundamental problem in geometry processing, as it is a necessary step for many applications such as shape interpolation [2], texture transfer [19], and statistical shape analysis [8, 9].

A particularly appealing approach to non-rigid shape matching is the recent deep functional maps framework [20, 42]. It consists of two main blocks: (1) a deep feature extractor that computes descriptor functions approximately preserved across a pair of input shapes, and (2) a differentiable functional map solver that computes the matching in the spectral domain under axiomatic regularizations. Func-

tional maps [51] provide a flexible and compact representation of shape correspondences as small-sized matrices, and have been successfully applied to shape matching, but also other tasks such as shape classification [31] or representation alignment [27]. However, existing methods rely heavily on axiomatic modeling, such as near isometry and local area preservation constraints [15, 67], for formulating the training loss or for functional map regularization inside the networks. This limits their accuracy and generalization to unseen shape pairs where these assumptions may not hold.

This motivates us to explore replacing the axiomatic regularizations in deep functional maps with structural priors learned directly from data. Recently, diffusion models [33, 69] have demonstrated strong capabilities in modeling complex data distributions, achieving impressive results in tasks such as image generation and editing [49, 65]. The rich priors learned by these models have also proven useful in many other tasks, such as 3D generation and reconstruction [57, 75, 77]. Inspired by these successes, we propose to leverage diffusion models to capture structural properties of functional maps directly in the spectral domain. With the spectral diffusion priors as a powerful regularizer, we aim to enhance the robustness and accuracy of functional map estimation for non-rigid shape matching.

In this work, we leverage large-scale datasets of registered non-rigid 3D shapes, primarily human bodies [9], to learn informative structural priors of functional maps. We first construct a large collection of high-quality functional maps from the non-rigid 3D registrations in [9], and then train an unconditional diffusion model in the spectral domain to capture the distribution of these functional maps. Building on this spectral diffusion model, we propose a novel zero-shot deep functional map pipeline, which incorporates a novel data-driven regularization to promote structural properties consistent with those observed in the training data. Specifically, we distill a mask from the trained spectral diffusion model. This mask encodes learned structural priors, replacing conventional axiomatic regularizations, such as Laplacian commutativity or orthogonality, commonly used in deep functional map pipelines. Remarkably, our learned spectral diffusion priors, though trained on human shapes, are category-agnostic and demonstrate strong generalization to unseen shape categories, including humanoids and animals.

In summary, our contributions are:

- We introduce a spectral diffusion model to learn the distribution of functional maps, effectively capturing their structural characteristics in the spectral domain.
- We distill the learned spectral diffusion priors into a mask that serves as a data-driven regularizer, replacing axiomatic regularizations and improving robustness in zero-shot deep functional map pipelines.
- Our spectral diffusion priors, learned from human shapes, demonstrate remarkable adaptability, generalizing effectively to diverse and previously unseen shape categories.

2. Related Work

Functional Maps and Regularizations. Since the functional maps seminal work [51], in which orthogonality and Laplacian commutativity penalties are derived to encourage near-isometric maps, many axiomatic approaches have been proposed to improve the computation of functional maps. One of the most common penalties is to encourage bijectivity of the functional maps [52], thereby encouraging bijectivity of the corresponding pointwise correspondence. Ren *et al*. [58] propose to encourage orientation preservation and continuity of maps, along with a new iterative algorithm to improve the quality of maps. Panine *et al*. [54] propose to promote conformality of maps with a new penalty and functional basis for the map computation.

Recently, Zoomout [48] has shown that alternating between functional map and pointwise correspondence representation is a remarkably efficient method to regularize the final map quality automatically. One of the key components is the projection of maps to the proper map space [40, 61], the space of functional maps corresponding to valid point correspondences, from which it is easier to optimize spatial energies like the Dirichlet energy [46], or elastic energies with a separate deformation network [16]. This has lately been used to improve the training of deep functional maps, with different losses encouraging functional maps to be proper [15, 17, 39]. Some approaches propose to use geometric information to refine maps using geometric consistency either in the training of deep functional maps [17], or at test time with precomputed deep features [29, 63, 64], which can be useful in partial shape matching [22, 23]. Another way to mitigate errors is to match multiple shapes at the same time rather than a pair [26, 28], with an increased computational cost. Another direction aims at exploring an alternative to the Laplace-Beltrami operator for constructing the shape basis [7, 10, 29, 76], however, it is not straightforward to incorporate these new approaches in a deep functional maps pipeline.

Only a few works have focused on improving mask regularization. In partial functional maps [62], the authors propose a slanted regularization to encourage the map to follow Weyl's law. Ren *et al*. improved the original Laplacian commutativity by using the Resolvent operator [59], which has better theoretical properties.

To perform well, most of these approaches require a good quality initialization as input, which is given by the mask regularization [4], or by large-scale pre-trained features [79]. In contrast, we propose a data-driven mask computation, which allows for a better initialization of the maps and a distilled loss to optimize the maps.

Knowledge Distillation of Score-based Generative Models. Denoising diffusion models [33] are a class of generative models that learn the mapping between an (unknown) data distribution and a Gaussian distribution. They have shown a great generalization capability, surpassing GANs for image generation [18]. Moreover, the distribution learned by the diffusion models, has numerous desirable properties, as it learns the gradient of the log density, the score, of the (noised) data distribution [70]. The learned score can be distilled in various ways depending on the downstream tasks, such as image inpainting [43], denoising [71], and more recently, text-to-3D generation.

Score distillation sampling (SDS) [57] has indeed quickly become a preferred approach to zero-shot text-to-3D generation using 2D image-based diffusion models. The authors of [57] propose to use the learned score of the diffusion model as the gradient of a desired image given a user text prompt. Coupled with a differentiable scene representation and rendering, the proposed loss allows for the accurate generation of new 3D scenes. However, the approach exhibits undesired properties, such as almost deterministic generation (convergence towards the mean image corresponding to the prompt), low-quality shapes, or color saturation of the generated shapes. To overcome this limitation, HiFA [80] proposes a strategy to mitigate those effects by using negative prompts and forcing realistic images by emphasizing SDS steps with low levels of noise. ProlificDreamer [75] instead proposes to learn a fine-tuned diffusion model for the prompt, to avoid deterministic generation. Those approaches, however, require a conditional diffusion model to work properly. Lukoianov *et al*. [44] proposes DDIM inversion to follow the score to avoid wrong gradient directions. However, the inversion process is approximate and can be slow.

Most methods presented here are designed for image generation. As we will see in Sec. 4, their generalization to the regularization of functional maps is not straightforward, and we therefore propose to adapt the distillation strategy for robust non-rigid shape matching.

3D Generative Modeling for Shape Matching. Generative modeling has proven to be a powerful tool for solving shape-matching tasks. In 3D-CODED [30], the authors train a point-cloud auto-encoder on a large synthetic human dataset and register human scans in a zero-shot approach. The auto-encoder approach has been improved with geometric regularization to avoid degenerated reconstructions. Neural Jacobian Fields [1] improve the overall quality of reconstructions by predicting the Jacobian of deformations instead of directly predicting the deformation field, which implicitly regularizes the final shape. ARAPReg [34] learns a geometrically regularized latent space by penalizing directions that increase the ARAP energy. This strategy has improved to learn correspondences on unregistered datasets [78] but requires a two-step training strategy. Finally, some works directly learn to generate the matching of shapes, whether directly in the spatial domain [25, 50, 73] or in the spectral domain using deep functional maps [32, 42]. In particular, deep functional methods have shown great success for intra-category training of shape matching approaches [14, 15, 40, 66, 72]. A concurrent work [81] proposes to learn conditional distribution of functional maps along with shape descriptors. All the works mentioned above need training on specific categories before being used at test time: a model trained on humans is useful for registering other humans but often fails to generalize well to different categories, such as animals.

We overcome this limitation by training an unconditional diffusion model on the space of functional maps. We propose a novel distillation strategy adapted to the specific problem of functional map regularization. Our approach generalizes to new categories of shapes (*e.g*., animals).

3. Background & Motivation

3.1. Deep Functional Maps

The objective of deep functional maps is to learn shape descriptors to compute high-quality correspondences on pairs of shapes $(\mathcal{S}_1, \mathcal{S}_2)$, represented as triangular meshes. Let n_1, n_2 be their respective number of vertices. The pipeline generally consists of the following steps:

- Compute the first k eigenfunctions of an intrinsic surface operator – usually the Laplace-Beltrami operator – on each shape, serving as a basis of functions on these shapes. The Laplacian is discretized as $S^{-1}W$, where S is the diagonal matrix of mesh vertex areas, and W is the cotangent weight matrix. The eigenfunctions are stored as matrices in the form of $\Phi_1 \in \mathbb{R}^{n_1 \times k}$ and $\Phi_2 \in \mathbb{R}^{n_2 \times k}$.
- A set of d descriptor functions (approximately preserved by the unknown map) $F_1, F_2 = f_\theta(\mathcal{S}_1) \in \mathbb{R}^{n_1 \times d}$, $f_\theta(\mathcal{S}_2) \in \mathbb{R}^{n_2 \times d}$ extracted using a neural network $f_\theta : \mathcal{S} \mapsto \mathbb{R}^d$. After projecting them onto the respective eigenfunctions, the resulting descriptor coefficients are stored as matrices $A_1, A_2 \in \mathbb{R}^{k \times d}$, respectively.
- The functional map matrix C between $\mathcal{S}_1$ and $\mathcal{S}_2$ is computed by solving the following:

$$C = \underset{C}{\operatorname{argmin}} \, ||CA_1 - A_2||^2 + \alpha ||M_{\text{reg}} C||^2, \quad (1)$$

  where the first term is a data preservation term between descriptors, and the second term regularizes the map structure by using a sparsity promoting mask M_{reg}, derived from Laplacian or Resolvent operator commutativity. The whole pipeline $(\mathcal{S}_1, \mathcal{S}_2) \mapsto C$, also called FMReg layer [20] is fully differentiable with respect to θ.
- The weights θ are optimized during training with axiomatic regularization terms, like area preservation,

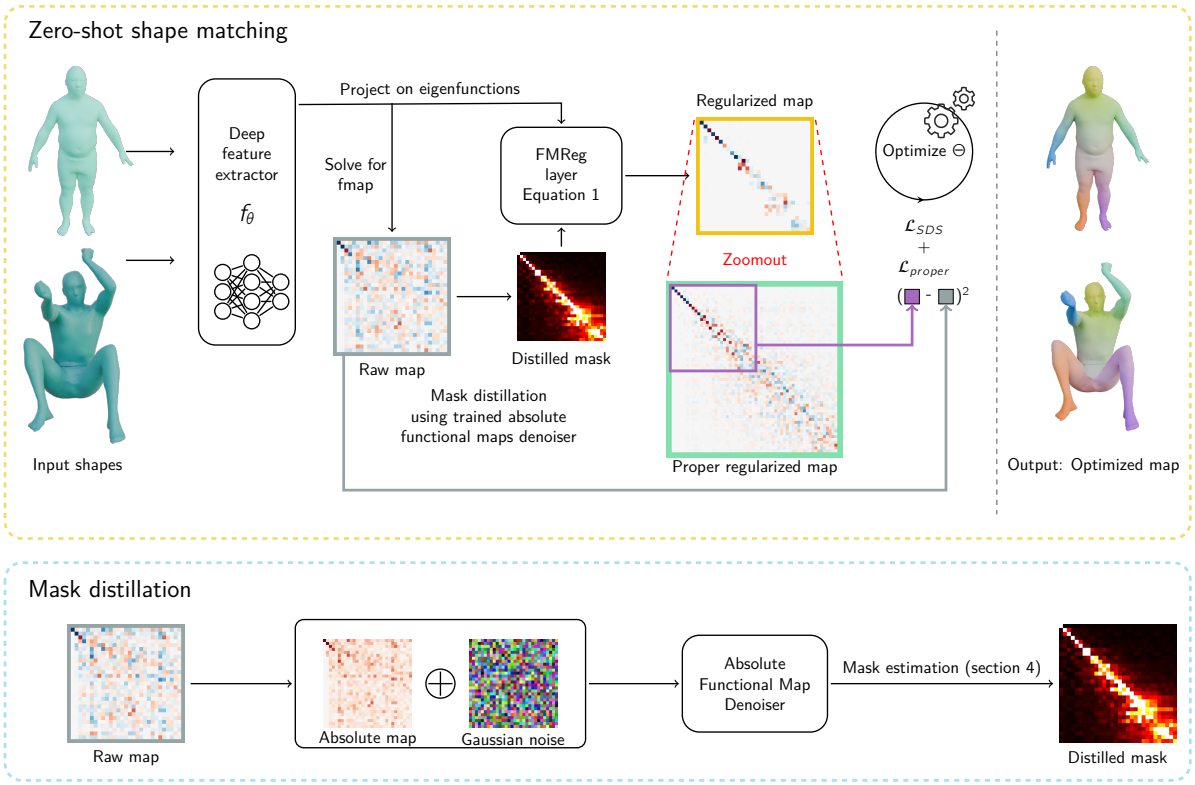


Figure 2. Our diffusion-based zero-shot shape matching pipeline. Given two shapes, we distill a mask using the spectral diffusion denoiser applied to the estimated "raw" functional map from features. We then use the distilled mask in a regularized functional map solver (FMReg) and apply Zoomout [48] to obtain a proper, regularized map. We minimize both score distillation and L2 distance to the proper map.

which is formulated as an orthogonality penalty:

$$\mathcal{L}_{\text{ortho}}(C) = ||CC^T - I||^2, \tag{2}$$

or other regularizations, such as orientation preservation.

- A test-time, we use the learned pipeline to extract C. Post-processing algorithms such as Zoomout can be used to increase the accuracy of the map, before extracting the point-to-point-map using the aligned eigenfunctions $\Phi_1 C^T$ and Φ_2 and nearest neighbor search or more accurate techniques [53].

Orthogonality and commutativity penalties are essential to ensure that the correspondences are plausible. We propose to replace those axiomatic penalties in deep functional maps with data-driven penalties by distilling priors from trained spectral diffusion models in a zero-shot manner.

3.2. Score-based Generative Modeling

In this section, we follow the formalism of [37] to present denoising score models. The general objective of generative modeling is to learn a distribution $p_\psi(x)$ (where ψ are the learned parameters) corresponding to an (unknown) data distribution using the available samples of this distribution. Denoising score matching [35], and in particular, denoising diffusion models, are a specific class of models that learn to model the score function, defined as:

$$s(x) = \nabla_x \log p(x), \tag{3}$$

instead of the density p. This formulation overcomes the problem of normalizing constants of the data density. Knowing the score function is equivalent to knowing the data distribution, as one can sample from it using Langevin dynamics [55]. Since the score is unknown in parts of the sample space without data, denoising score matching learns the score functions $s_\theta(x_\sigma, \sigma) = \nabla_{x_\sigma} \log q(x_\sigma, \sigma)$ at different noise scales [74], where $x_\sigma = x + n_\sigma$, with $n_\sigma \sim \mathcal{N}(o, \sigma^2 I)$. This is done by learning a denoiser network $D_\psi(x + n_\sigma, \sigma)$ by minimizing the following loss:

$$\mathbb{E}_{x\sim p_{\text{data}}} \mathbb{E}_{n_\sigma \sim \mathcal{N}(0,\sigma^2 I)} ||D_\psi(x + n_\sigma, \sigma) - x||^2, \tag{4}$$

where the optimized parameters are the parameters ψ of the denoiser. We drop the denoiser parameters sign ψ to avoid confusion with other learnable parameters, since for the rest of the paper, the denoiser is considered as trained. After training, new samples are generated by progressively denoising random samples, following a probability ordinary differential equation [70]. Moreover, the score at noise level

σ can be estimated using:

$$\nabla_{x_\sigma} \log p(x_\sigma; \sigma) = (D(x_\sigma; \sigma) - x)/\sigma^2 \tag{5}$$

Score Distillation Sampling. Score distillation sampling (SDS) [57] is a generic way of transferring knowledge from a diffusion model learned on a *source domain* Ω, to regularize or generate samples y in a *target domain*. It can be summarized as follows: (1) obtain a trained denoiser $D(x_\sigma, \sigma)$, with $x \in \Omega$, the *source domain*, on which it is easy to train the denoiser; (2) differentiably extract $x = g(y_\theta)$ from the *target* to the *source* domain, where the representation y_θ of samples in the target domain is optimizable, and g is a differentiable mapping from the *target* to the *source* domain. In the original work [57], the *source domain* is images, and the *target domain* is 3D scenes. 3D scenes are parameterized using Neural Radiance Fields y_θ, and the function $g(y_\theta)$ is simply the differentiable rendering of a novel view.

At each iteration, SDS consists in sampling $x = g(y_\theta) \in \Omega$ and perturbing the x with noise $n_\sigma \sim \mathcal{N}(0, \sigma)$. Then, SDS guides the target representation ψ by applying the following gradient to the parameters:

$$\nabla_\theta \mathcal{L}_{\text{SDS}} = \mathbb{E}_{\sigma, x_\sigma \sim \mathcal{N}(x,\sigma)}[(x_\sigma - D(x_\sigma, \sigma))/\sigma] \frac{\partial g}{\partial \theta}. \tag{6}$$

The gradient is not backpropagated through the denoiser as it is costly and unstable due to the noising step. In practice, only a single denoising step is applied for better performance.

3.3. Motivation

As we can generate training functional maps easily (Sec. 4.1), our goal is to leverage the learned functional maps distribution from score models, for the matching of unseen shapes.

A first solution could be to learn a conditional distribution by conditioning the functional map diffusion model on point descriptors to generate the target map, as in recent diffusion-based rigid shape-matching approaches [36]. However, the learned model would be category-specific, as with classic deep functional maps methods.

A second solution could be to use the learned probability likelihood of the score model [70] as a proxy for measuring map quality. A similar idea, based on axiomatic constraints instead of learned penalties, has been explored in MapTree [60]. However, such an approach outputs a set of *candidate* maps, including symmetry-swapping maps, as purely intrinsic approaches do not differentiate between them. Moreover, evaluating the likelihood of a diffusion model is costly due to the integration of the generation trajectory.

A third potential solution is to adapt Score Distillation Sampling to the deep functional maps setting. Indeed, the recent Shape-Non-Rigid Kinematics (SNK) [5] work has shown that deep functional maps networks can be used in a zero-shot setting. In this paper, the authors exploit Neural Correspondence Priors [4] to obtain good initializations and optimize the map using spectral and spatial axiomatic constraints.

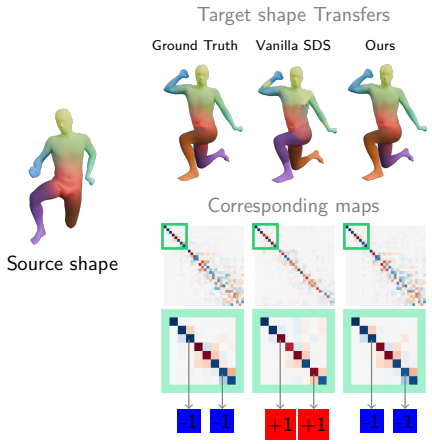


Figure 3. Comparison between vanilla SDS and our approach. The same diagonal structure is encouraged between the two approaches. However, using our proposed mask allows for fixing misalignments (sign flips in the functional maps, highlighted at the bottom) of the initialization. In contrast, the vanilla SDS converges towards the closest map with the same diagonal.

We now discuss the adaptation of SDS to our problem. Given two shapes $\mathcal{S}_1, \mathcal{S}_2$, the *source domain* is shape correspondences, represented as functional maps (Sec. 3.1). The *target domain* is descriptor functions. The pointwise descriptors of shapes F_i are parameterized by a deep neural network $F_i = f_\theta(\mathcal{S}_i)$, from which we estimate the functional map C_{12} with the FMReg layer [20]. In the SDS notation, x is the functional map C, θ corresponds to the weights of the feature extractor, and g is the FMReg layer.

In Fig. 3, we perform a preliminary experiment with SDS. Notably, the recovered map shows significant mismatches (left and right legs are reversed) compared to the ground truth. We also observe that the structure promoted by the approach is nearly diagonal, as commonly observed in functional maps. The wrong signs along the diagonal cause the observed inconsistencies. We argue that the sign ambiguity of functional maps affects the performance (more discussions in the first section of supp. material). To better capture the underlying structure of the functional maps from data, we adopt a **sign-agnostic** approach in the following section.

4. Method

We first train a spectral diffusion model of functional maps (Sec. 4.1). Second, we devise a zero-shot training approach by distilling the knowledge learned by the trained model

(Sec. 4.2), and the pipeline is illustrated in Fig. 2.

4.1. Training

Given a dataset of registered shapes, our objective is to train a spectral functional maps diffusion model $D(C_\sigma, \sigma)$. Using the registered shapes, we first build a dataset of ground-truth functional maps C_{gt}. Given $\mathcal{S}_1, \mathcal{S}_2$ two registered shapes, and Φ_1, Φ_2 their respective eigenfunctions, the ground-truth map between the two shapes is given by:

$$C_{12-\text{gt}} = \Phi_2^\dagger \Phi_1,$$

where $(\cdot)^\dagger$ is the Moore-Penrose pseudoinverse. We extract functional maps of fixed size $n \times n$ of template to shape correspondences.

The extracted functional maps are thus matrices $C \in \mathcal{M}_n(\mathbb{R})$, which are analogous to images. We thus build upon the available image-based architectures and use the Diffusion Transformer architecture [56], which has shown great capabilities for image generation.

Our spectral denoiser $D_\psi(C_\sigma, \sigma)$ takes as input a matrix $C_\sigma \in \mathcal{M}_n(\mathbb{R})$ and noise level σ. To be sign-agnostic, we train a diffusion model on **absolute functional maps**, with input training data as the set of $|C_{\text{gt}}|$ (which is C_{gt} with $x \to |x|$ applied on each element).

4.2. Zero-shot shape matching

In this section, we are now given as input two *unseen shapes* $\mathcal{S}_1, \mathcal{S}_2$. We aim to estimate the functional map C_{12} between the two shapes. We use the deep functional maps framework [20] to differentiably estimate the functional map C_{12}, from pointwise descriptors $F_i = f_\theta(\mathcal{S}_i)$, parameterized by neural network weights θ.

We optimize the parameters θ by applying a distillation loss to the functional maps. The remaining section discusses the construction of our distillation loss.

Mask regularization is a sign-agnostic regularization that has proven essential in deep functional maps as it provides reliable initialization towards the final solution [4]. In this section, we seek to provide a masked regularization in the form of Eq. (1). We search for sparsity-promoting masks M_σ, such that given a ground truth map C_{gt}, we have:

$$||M_\sigma \cdot C_{gt}|| \simeq 0 \tag{7}$$

It is equivalent to say that C_{gt} maximizes the likelihood

$$p(C_\sigma; \sigma) \propto \exp(-||M_\sigma \cdot C_\sigma||^2). \tag{8}$$

Under this hypothesis, the score function derives as:

$$s(C_\sigma; \sigma) = \nabla_x \log p(x : \sigma) = -2M_\sigma^2 \cdot C_\sigma. \tag{9}$$

By using Equation (5), we obtain:

$$M_\sigma^2 \cdot C_\sigma = (C_\sigma - D(C_\sigma; \sigma))/2\sigma^2, \tag{10}$$

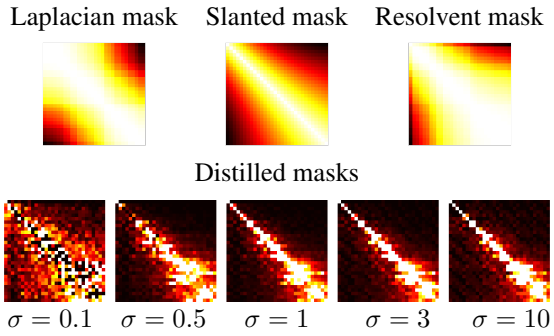


Figure 4. Top: usual masks for functional map regularization. Bottom: estimated distillation masks at different noise levels

which reduces as the following formula for computing M_σ (by taking the mean over the noise distribution):

$$M_\sigma^2 = \mathbb{E}_{n_\sigma \sim \mathcal{N}(0, \sigma^2 I)} \left[(C_\sigma - D(C_\sigma; \sigma))/(2\sigma^2 C_\sigma) \right]. \tag{11}$$

Applying the formula directly would cause numerical instabilities when dividing by C_σ that can contain 0 values, if arbitrary values of noise are sampled. We avoid this by sampling only $n_\sigma > 0$, which ensures only positive values when working with absolute functional maps $|C|$. The formula for the mask is finally:

$$M_\sigma^2 = \mathbb{E}_{n_\sigma \sim \mathcal{N}(0, \sigma^2 I), n_\sigma > 0} \left[(|C|_\sigma - D(|C|_\sigma; \sigma))/(2\sigma^2 |C|_\sigma) \right]. \tag{12}$$

We can distill the learned structure from the spectral diffusion model into a mask M_σ for different noise levels, given any functional map matrix C_{init}. We show in Fig. 4 the estimated masks for different noise levels. We can incorporate this mask into the functional map computation: (1) we estimate a "raw" functional map C_{raw} based on the input descriptors using Eq. (1) with $\alpha = 0$; (2) we then estimate a mask M_σ from C_{raw} and solve Eq. (1) with M_σ to obtain a mask-regularized map C_{reg}.

Proper SDS Similarly to SDS, we do not backpropagate through the mask optimization during optimization. Moreover, it has been shown that projecting the functional map on the "proper" map space [6, 61] (space of maps computed from a pointwise map) is necessary for convergence. We apply Zoomout [48] to the regularized functional map C_{reg} and obtain a proper regularized map C_{proper}. We minimize the L_2 distance between the raw map and the proper map:

$$\mathcal{L}_{\text{proper}}(C_{\text{raw}}) = ||C_{\text{raw}} - C_{\text{proper}}||^2, \tag{13}$$

where we only backpropagate to the feature extractor weights through C_{raw}. Finally, we also apply SDS to the absolute raw map. Thus, our total loss is :

$$\mathcal{L}_{\text{total}}(C_{\text{raw}}) = \mathcal{L}_{\text{proper}}(C_{\text{raw}}) + \mathcal{L}_{SDS}(|C_{\text{raw}}|) \tag{14}$$

	Methods	Humans			Humanoids		Animals	
		FAUST	SCAPE	SHREC19	DT4D-Intra	DT4D-Inter	SMAL	TOSCA
Axiom.	Ini + Zoomout (Laplacian)	3.8	7.5	13.1	1.8	16.5	18.3	8.1
	Ini + Zoomout (Resolvent)	3.2	5.7	12.4	**1.6**	13.4	19.1	5.4
	Smooth shells [24]	2.5	4.7	12.2	/	/	16.3	/
Learned	3D-CODED [30]	7.5	17.2	13.4	45.0	61.4	54.6	32.8
	Neural Jacobian Fields [1]	5.9	11.7	9.6	43.4	32.8	49.2	50.2
	Simplified Fmaps [45]	**1.7**	**2.3**	**3.4**	2.0	8.9	42.1	5.1
Zero-shot	SNK [5]	1.8	4.7	5.8	2.0	9.0	**9.1**	3.6
	Ini + Zoomout (our mask)	2.4	6.6	8.3	2.1	11.7	12.9	8.3
	Ours	1.9	4.4	3.9	1.8	**8.6**	10.1	**2.9**

Table 1. Comparison of matching accuracy of axiomatic, learned, and zero-shot shape matching methods. The learning-based methods are trained on human shapes from Dynamic FAUST. The lower the better.

Summary Given a pair of shapes, and a trained spectral diffusion model, at test time we optimize a shape pointwise feature extractor, $F_i = f_\theta(\mathcal{S}_i)$. (1) Given features on the two shapes, we first estimate a functional map C^{12}_{raw} between the shapes by solving Eq. (1) with $\alpha = 0$. (2) We use this map to distill a mask M_σ from the diffusion model, and solve Eq. (1) a second time to obtain a regularized map C^{12}_{reg}, on which we apply Zoomout to obtain C^{12}_{proper}. (3) This map, along with the diffusion model, is used to compute $\mathcal{L}_{\text{total}}$ (Eq. (14)). (4) The parameters f_θ of the feature extractor are optimized through back-propagation. (5) We convert the optimized functional map C^{12} to a point-to-point map with the standard approach [51]. Note that our pipeline differs from previous deep functional maps approaches in that we *avoid axiomatic priors*, such as Laplacian commutativity or orthogonality, both at mask estimation and training. Instead, all of our regularization and objective terms, except for the basic properness term, are derived *solely from available training data*.

5. Experiments

5.1. Experimental Details

Functional Map Diffusion Model. We train our diffusion model on 30×30 maps, using template-to-shape maps on the D-FAUST dataset, for a total of $\sim 40,000$ maps. The architecture is the DiT-S [56] Diffusion Transformer with a patch size of 5. We train our model with EDM [37] for 1000 epochs with the variance preserving loss.

Zero-shot Optimization. We use DiffusionNet [68] as our feature extractor. The estimated functional map size is 30×30. We follow the zero-shot experimental settings from SNK [5]. We set $\sigma = 1$ for our distilled mask as we found the best results from this specific value, with $N = 100$ noisy samples, which can be done in a single batch denoising. Iterating the process did not provide significant improvement. We apply Zoomout to increase the map size from 30×30 to 40×40 to compute $\mathcal{L}_{proper}$.

5.2. Datasets and Comparison

Near-isometric Shape Matching. We first test the generalization of our approach on human data. Notably, we test on the oriented versions of the remeshed FAUST [8], SCAPE [3], and SHREC [47] datasets, on the usual test sets from commonly used train/test splits.

Non-isometric Shape Matching. We then test our approach on unseen data types, namely humanoids and animals. For humanoids, we used the remeshed split of the DynamicThings4D dataset (DT4D) [41], from which we use the intra-category and inter-category test sets from [40]. For animals, we tested our approach on the SMAL remeshed dataset [21] and animal shape pairs from the TOSCA dataset [13], as done in [72].

Baselines. We compare our method with axiomatic [24], learned [1, 30, 45] and zero-shot [5] baselines. A detailed description is provided in the supplementary material.

5.3. Results

We follow the Princeton benchmark evaluation protocol [38] and evaluate the accuracy of the maps using the geodesic error of the computed correspondence. We present the results on near isometric data on the left of Tab. 1. We outperform both axiomatic and other zero-shot approaches on this task. Notably, we are close to the state-of-the-art deep functional map Simplified Fmaps approach [45]. Also, our distilled mask provides, in general, a good quality initialization, competitive with other approaches, and outperforms the Laplacian and Resolvent masks (Ini + Zoomout).

The results on non-isometric data are in the right of Tab. 1. Notably, our approach outperforms other approaches on the DT4D-Inter challenge and the TOSCA

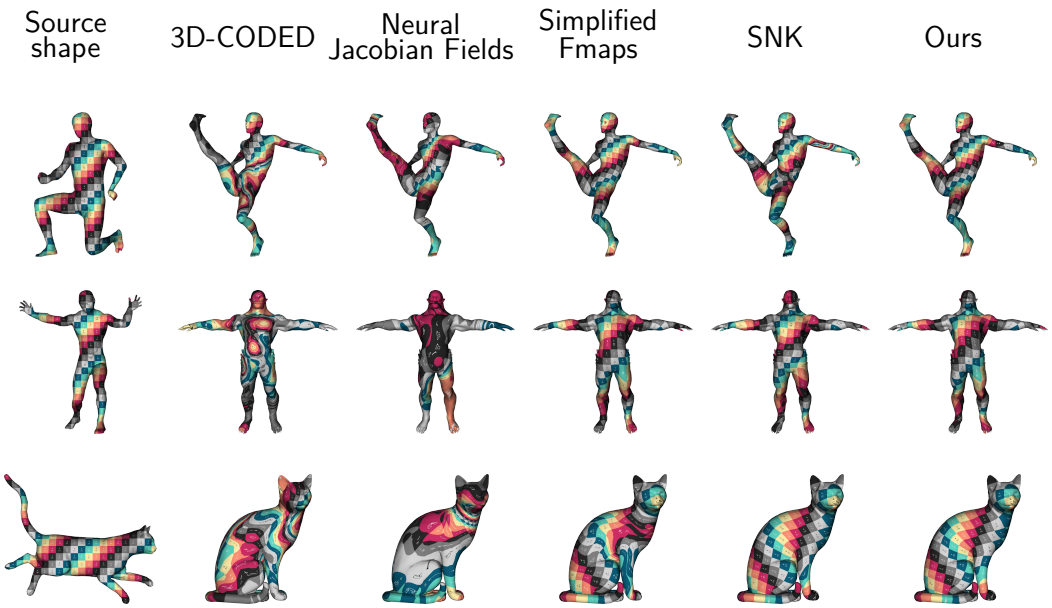


Figure 5. Texture transfer on different examples. We observe that deformation-based models such as 3D-CODED or Neural Jacobian Fields struggle with challenging poses and fail to generalize to unseen meshes. In the meantime, the learned pointwise descriptors of Simplified Fmaps generalize well to humanoids but struggle when confronted with new categories like animals. The SNK zero-shot approach can match shapes from different categories but struggles with challenging poses. Our approach can infer qualitative shape correspondences on each of those challenging examples.

dataset, including Simplified Fmaps. 3D-CODED and Neural Jacobian Fields, based on learned deformation models, perform poorly on humanoids and animals, as the learned deformations are category-specific. Moreover, Simplified Fmaps fails on the SMAL dataset, suggesting that learned descriptors do not generalize well to animals. We also outperform SNK on most datasets, and our distilled mask provides a better initialization than the traditional Laplacian and resolvent masks (Ini + Zoomout).

5.4. Ablation study

We ablate the different components of our approach in Table 2. As stated in Sec. 3.3, vanilla SDS fails to correct misalignments and shows poor results. Using $\mathcal{L}_{proper}$ alone is efficient but far from state-of-the-art performance. The best is to combine $\mathcal{L}_{proper}$ and $\mathcal{L}_{SDS}$. Finally, we added axiomatic penalties (orthogonal, bijectivity, and Laplacian losses) to DiffuMatch. We find that the final accuracy is nearly the same, indicating that our formulation already encompasses these axiomatic regularizations.

Approach	Geod Error (SHREC)
Vanilla SDS	57.3
Mask + Zoomout	8.3
$\mathcal{L}_{proper}$	7.7
Mask + $\mathcal{L}_{SDS}$	7.1
Mask + $\mathcal{L}_{proper}$	6.7
Mask + $\mathcal{L}_{proper}$ + $\mathcal{L}_{SDS}$ (ours)	**4.4**
Ours + Axiomatic	4.3

Table 2. Ablation study of the different components of our approach

6. Limitations and future work

Despite the efficiency of our method, our method might not handle well highly non-isometric shapes or partial shapes [62], which is a known issue of functional map-based methods [11, 12]. A potential direction to mitigate this problem is to jointly learn the basis along with spectral regularization. Moreover, our diffusion model is trained only on human shapes with limited diversity. Using or generating more registered training data will be crucial towards a unified model for functional maps.

7. Conclusion

In this work, we presented a functional map score generative model, trained on registered human shapes, to learn the structure induced by the distribution of functional maps. Based on our model, we proposed a novel functional map penalty and a zero-shot training pipeline to match shape pairs at test time. The results demonstrate state-of-the-art results for zero-shot shape matching on diverse benchmarks, including categories unseen during training. We believe our approach will serve as a first step towards foundation models for shape matching.

Acknowledgements This work was performed using HPC resources from GENCI-IDRIS (Grant 2025-AD010613760R2). Parts of this work were supported by the ERC Consolidator Grant 101087347 (VEGA), as well as gifts from Ansys and Adobe Research, and the ERC Starting Grant SpatialSem (101076253).

References

[1] Noam Aigerman, Kunal Gupta, Vladimir G. Kim, Siddhartha Chaudhuri, Jun Saito, and Thibault Groueix. Neural jacobian fields: learning intrinsic mappings of arbitrary meshes. *ACM Trans. Graph.*, 41(4), 2022. 3, 7

[2] Marc Alexa, Daniel Cohen-Or, and David Levin. As-rigid-as-possible shape interpolation. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, page 157–164, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 1

[3] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: Shape completion and animation of people. 2023. 7

[4] Souhaib Attaiki and Maks Ovsjanikov. Ncp: Neural correspondence prior for effective unsupervised shape matching. *Advances in Neural Information Processing Systems*, 35:28842–28857, 2022. 2, 5, 6

[5] Souhaib Attaiki and Maks Ovsjanikov. Shape non-rigid kinematics (snk): A zero-shot method for non-rigid shape matching via unsupervised functional map regularized reconstruction. *Advances in Neural Information Processing Systems*, 36:70012–70032, 2023. 5, 7

[6] Souhaib Attaiki and Maks Ovsjanikov. Understanding and improving features learned in deep functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1326, 2023. 6

[7] David Bensaïd, Amit Bracha, and Ron Kimmel. Partial shape similarity by multi-metric hamiltonian spectra matching. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 717–729. Springer, 2023. 2

[8] Federica Bogo, Javier Romero, Matthew Loper, and Michael 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. 1, 7

[9] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J Black. Dynamic faust: Registering human bodies in motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6233–6242, 2017. 1, 2

[10] Amit Bracha, Oshri Halim, and Ron Kimmel. Shape Correspondence by Aligning Scale-invariant LBO Eigenfunctions. In *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2020. 2

[11] Amit Bracha, Thomas Dagès, and Ron Kimmel. On unsupervised partial shape correspondence. In *Proceedings of the Asian Conference on Computer Vision*, pages 4488–4504, 2024. 8

[12] Amit Bracha, Thomas Dagès, and Ron Kimmel. Wormhole loss for partial shape matching. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2025. 8

[13] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008. 7

[14] Dongliang Cao and Florian Bernard. Unsupervised deep multi-shape matching. In *European Conference on Computer Vision*, pages 55–71. Springer, 2022. 3

[15] Dongliang Cao, Paul Roetzer, and Florian Bernard. Unsupervised learning of robust spectral shape matching. *ACM Transactions on Graphics (TOG)*, 42:1 – 15, 2023. 2, 3

[16] Dongliang Cao, Marvin Eisenberger, Nafie El Amrani, Daniel Cremers, and Florian Bernard. Spectral meets spatial: Harmonising 3d shape matching and interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3658–3668, 2024. 2

[17] Dongliang Cao, Zorah Lähner, and Florian Bernard. Synchronous diffusion for unsupervised smooth non-rigid 3d shape matching. In *European Conference on Computer Vision*, pages 262–281. Springer, 2024. 2

[18] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 3

[19] Huong Quynh Dinh, Anthony Yezzi, and Greg Turk. Texture transfer during shape transformation. *ACM Transactions on Graphics (ToG)*, 24(2):289–310, 2005. 1

[20] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8592–8601, 2020. 1, 3, 5, 6

[21] Nicolas Donati, Etienne Corman, Simone Melzi, and Maks Ovsjanikov. Complex functional maps: A conformal link between tangent bundles. In *Computer Graphics Forum*, pages 317–334. Wiley Online Library, 2022. 7

[22] Viktoria Ehm, Maolin Gao, Paul Roetzer, Marvin Eisenberger, Daniel Cremers, and Florian Bernard. Partial-to-partial shape matching with geometric consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27488–27497, 2024. 2

[23] Viktoria Ehm, Paul Roetzer, Marvin Eisenberger, Maolin Gao, Florian Bernard, and Daniel Cremers. Geometrically consistent partial shape matching. In *2024 International Conference on 3D Vision (3DV)*, pages 914–922. IEEE, 2024. 2

[24] Marvin Eisenberger, Zorah Lahner, and Daniel Cremers. Smooth shells: Multi-scale shape registration with functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12265–12274, 2020. 7

[25] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. Deep shells: Unsupervised shape correspondence with optimal transport. *Advances in Neural information processing systems*, 33:10491–10502, 2020. 3

[26] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. G-msm: Unsupervised multi-shape matching with graph-based affinity priors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[27] Marco Fumero, Marco Pegoraro, Valentino Maiorca, Francesco Locatello, and Emanuele Rodolà. Latent functional maps: a spectral framework for representation alignment. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 2

[28] Maolin Gao, Zorah Lahner, Johan Thunberg, Daniel Cremers, and Florian Bernard. Isometric multi-shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14183–14193, 2021. 2

[29] Maolin Gao, Paul Roetzer, Marvin Eisenberger, Zorah Lähner, Michael Moeller, Daniel Cremers, and Florian Bernard. Sigma: Scale-invariant global sparse shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 645–654, 2023. 2

[30] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. 3d-coded : 3d correspondences by deep deformation. In *ECCV*, 2018. 3, 7

[31] Oshri Halimi and Ron Kimmel. Self functional maps. In *2018 International Conference on 3D Vision (3DV)*, pages 710–718. IEEE, 2018. 2

[32] Oshri Halimi, Or Litany, Emanuele Rodola, Alex M Bronstein, and Ron Kimmel. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2019. 3

[33] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3

[34] Q. Huang, X. Huang, B. Sun, Z. Zhang, J. Jiang, and C. Bajaj. Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5795–5805, Los Alamitos, CA, USA, 2021. IEEE Computer Society. 3

[35] Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005. 4

[36] Haobo Jiang, Mathieu Salzmann, Zheng Dang, Jin Xie, and Jian Yang. Se (3) diffusion model-based point cloud registration for robust 6d object pose estimation. *Advances in Neural Information Processing Systems*, 36:21285–21297, 2023. 5

[37] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022. 4, 7

[38] Vladimir G Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. *ACM transactions on graphics (TOG)*, 30(4):1–12, 2011. 7

[39] Lei Li, Souhaib Attaiki, and Maks Ovsjanikov. Srfeat: Learning locally accurate and globally consistent non-rigid shape correspondence. In *2022 International Conference on 3D Vision (3DV)*, pages 144–154. IEEE, 2022. 2

[40] Lei Li, Nicolas Donati, and Maks Ovsjanikov. Learning multi-resolution functional maps with spectral attention for robust shape matching. *Advances in Neural Information Processing Systems*, 35:29336–29349, 2022. 2, 3, 7

[41] Yang Li, Hikari Takehara, Takafumi Taketomi, Bo Zheng, and Matthias 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. 7

[42] Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE international conference on computer vision*, pages 5659–5667, 2017. 1, 3

[43] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022. 3

[44] Artem Lukoianov, Haitz Sáez de Ocáriz Borde, Kristjan Greenewald, Vitor Campagnolo Guizilini, Timur Bagautdinov, Vincent Sitzmann, and Justin Solomon. Score distillation via reparametrized DDIM. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3

[45] Robin Magnet and Maks Ovsjanikov. Memory-scalable and simplified functional map learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4041–4050, 2024. 7

[46] Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. Smooth non-rigid shape matching via effective dirichlet energy optimization. In *International Conference on 3D Vision (3DV)*, 2022. 2

[47] Simone Melzi, Riccardo Marin, Emanuele Rodolà, Umberto Castellani, Jing Ren, Adrien Poulenard, P Ovsjanikov, et al. Shrec'19: matching humans with different connectivity. In *Eurographics Workshop on 3D Object Retrieval*, pages 1–8. The Eurographics Association, 2019. 7

[48] Simone Melzi, Jing Ren, Emanuele Rodolà, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. Zoomout: spectral upsampling for efficient shape correspondence. *ACM Trans. Graph.*, 38(6), 2019. 2, 4, 6

[49] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. 2022. 2

[50] Aymen Merrouche, Joao Pedro Cova Regateiro, Stefanie Wuhrer, and Edmond Boyer. Deformation-guided unsupervised non-rigid shape matching. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 3

[51] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graph.*, 31(4), 2012. 2, 7

[52] Maks Ovsjanikov, Etienne Corman, Michael Bronstein, Emanuele Rodolà, Mirela Ben-Chen, Leonidas Guibas,

Frederic Chazal, and Alex Bronstein. Computing and processing correspondences with functional maps. In *SIGGRAPH ASIA 2016 Courses*, New York, NY, USA, 2016. Association for Computing Machinery. 2

[53] Gautam Pai, Jing Ren, Simone Melzi, Peter Wonka, and Maks Ovsjanikov. Fast sinkhorn filters: Using matrix scaling for non-rigid shape correspondence with functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 384–393, 2021. 4

[54] Mikhail Panine, Maxime Kirgo, and Maks Ovsjanikov. Non-isometric shape matching via functional maps on landmark-adapted bases. In *Computer graphics forum*, pages 394–417. Wiley Online Library, 2022. 2

[55] Giorgio Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3):378–384, 1981. 4

[56] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023. 6, 7

[57] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 5

[58] Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics (ToG)*, 37(6):1–16, 2018. 2

[59] Jing Ren, Mikhail Panine, Peter Wonka, and Maks Ovsjanikov. Structured regularization of functional map computations. In *Computer Graphics Forum*, pages 39–53. Wiley Online Library, 2019. 2

[60] Jing Ren, Simone Melzi, Maks Ovsjanikov, and Peter Wonka. Maptree: Recovering multiple solutions in the space of maps. *ACM Transactions on Graphics*, 39(6):1–17, 2020. 5

[61] Jing Ren, Simone Melzi, Peter Wonka, and Maks Ovsjanikov. Discrete optimization for shape matching. In *Computer Graphics Forum*, pages 81–96. Wiley Online Library, 2021. 2, 6

[62] Emanuele Rodolà, Luca Cosmo, Michael M Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. In *Computer graphics forum*, pages 222–236. Wiley Online Library, 2017. 2, 8

[63] Paul Roetzer and Florian Bernard. Spidermatch: 3d shape matching with global optimality and geometric consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14543–14553, 2024. 2

[64] Paul Roetzer, Ahmed Abbas, Dongliang Cao, Florian Bernard, and Paul Swoboda. Discomatch: Fast discrete optimisation for geometrically consistent 3d shape matching. In *European Conference on Computer Vision*, pages 443–460. Springer, 2024. 2

[65] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[66] Jean-Michel Roufosse, Abhishek Sharma, and Maks Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1617–1627, 2019. 3

[67] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional maps for shape matching. *Advances in Neural Information Processing Systems*, 33:19264–19275, 2020. 2

[68] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG)*, 41(3): 1–16, 2022. 7

[69] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 2

[70] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 3, 4, 5

[71] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. 3

[72] Mingze Sun, Shiwei Mao, Puhua Jiang, Maks Ovsjanikov, and Ruqi Huang. Spatially and spectrally consistent deep functional maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14497–14507, 2023. 3, 7

[73] Giovanni Trappolini, Luca Cosmo, Luca Moschella, Riccardo Marin, Simone Melzi, and Emanuele Rodolà. Shape registration in the time of transformers. *Advances in Neural Information Processing Systems*, 34:5731–5744, 2021. 3

[74] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 4

[75] Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 3

[76] Simon Weber, Thomas Dages, Maolin Gao, and Daniel Cremers. Finsler-laplace-beltrami operators with application to shape analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3131–3140, 2024. 2

[77] Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, and Aleksander Holynski. Reconfusion: 3d reconstruction with diffusion priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21551–21561, 2024. 2

[78] Haitao Yang, Xiangru Huang, Bo Sun, Chandrajit L. Bajaj, and Qixing Huang. Gencorres: Consistent shape matching via coupled implicit-explicit shape generative models. In *The Twelfth International Conference on Learning Representations*, 2024. 3

[79] Gal Yona, Roy Velich, Ehud Rivlin, and Ron Kimmel. Neural descriptors: Self-supervised learning of robust local surface descriptors using polynomial patches. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 218–230. Springer, 2025. 2

[80] Junzhe Zhu, Peiye Zhuang, and Sanmi Koyejo. HIFA: High-fidelity text-to-3d generation with advanced diffusion guidance. In *The Twelfth International Conference on Learning Representations*, 2024. 3

[81] Aleksei Zhuravlev, Zorah Lähner, and Vladislav Golyanik. Denoising functional maps: Diffusion models for shape correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26899–26909, 2025. 3