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

Supplementary Material

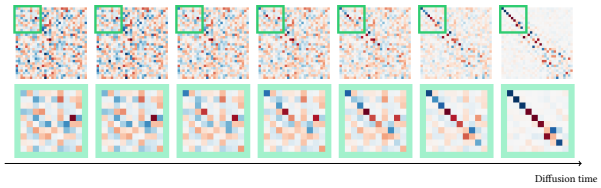

Figure 7. Generation process of a functional map using a diffusion model. For low-frequency elements (green square), the sign of diagonal elements at the Gaussian noise step never changes during the denoising process. This explains why spectral regularization with SDS fails to correct misalignments effectively.

In this supplementary material, we first provide insights on the sign ambiguity problem of functional maps in Sec. 8. We provide more experimentation details about datasets (Sec. 9), baselines (Sec. 10), and implementation (Sec. 11) for the experiments in Sec. 5 of the paper. Next, we show the behavior of our method with a plot of the loss during zero-shot optimization in Sec. 12, a visualization of denoising trajectories in Sec. 13, and finally an analysis of descriptors in Sec. 14. We also show that the matching provided by our method allows competitive reconstruction of input shapes by combining it with the ARAP energy in Sec. 15. Finally, in Sec. 16, we provide a simple experiment providing insights on sparsity-promoting mask efficiency.

8. Sign Ambiguity of Functional Maps

This phenomenon occurs because functional maps are nearly discrete at low frequencies. Indeed, it has been observed that the ground truth maps at low frequency follow a diagonal structure [19], where the values of the diagonal elements are ± 1 (modulo volume changes). This affects the overall trajectory of generation - where signs of the diagonal elements remain unchanged (Fig. 7) - and thus the capacity of diffusion models to provide efficient spectral regularization. Thus, to better capture the underlying structure of the functional maps from data, we chose to adopt a **sign-agnostic** approach.

9. Datasets

Near-isometric Shape Matching. The FAUST remeshed [6, 20] version contains 10 individuals in 10 different poses. SCAPE [2] contains 50 challenging poses of one individual. SHREC [17] contains 50 humans from different datasets, with 407 annotated pairs using an automatic human registration algorithm (partial shape matching pairs are excluded).

Non-isometric Shape Matching. The matching version of the DT4D dataset [16] contains more than 400 shapes, with more than 1000 annotated pairs remeshed using the LRVD algorithm [24], from which we use the intra-category and inter-category test sets from [14]. The SMAL remeshed dataset [9], which contains around 400 animal pairs extracted from real images using the SMAL deformation model [25]. The animal shape pairs from the TOSCA are from *cat, dog, horse* and *wolf* categories.

10. Baselines

We compare our method against several baselines for shape matching. 3D-CODED [11] is an autoencoder trained specifically for shape matching. The shape latent vectors are computed and refined by optimizing the obtained registrations. Neural Jacobian Fields [1] is a model that predicts the Jacobian of deformation instead of vertex positions and generalizes to unregistered meshes. Smooth shells [10] is an axiomatic approach that refines functional maps in a coarse-to-fine approach to obtain plausible final correspondences. Shape-Non-Rigid-Kinematics (SNK) [4] is a state-of-the-art zero-shot algorithm to train deep feature extractors on pairs of shapes. We also compare to a state-of-the-art deep functional maps approach, Simplified Fmaps [15]. All trainable models are trained on the D-FAUST dataset. Finally, we also show the results of using a feature extractor with random weights combined with different masks.

11. Experimental Details

Feature Extractor. We follow the zero-shot experimental settings from SNK [4]. The feature extractor consists of four DiffusionNet blocks of dimension 256, and we use 128 eigenvectors for the heat diffusion. The input features of the feature extractor are XYZ features on the oriented versions of each dataset [22]. We set $\lambda = 0.1$ for humans and $\lambda = 1e-3$ for the other datasets, respectively. For the Ini+Zoomout scenario with our mask, we set $\lambda = 1$.

Diffusion Model Training. We train our spectral diffusion model for 1000 steps. The training setting is the same as in [13], with optimal reweighting of the losses and using

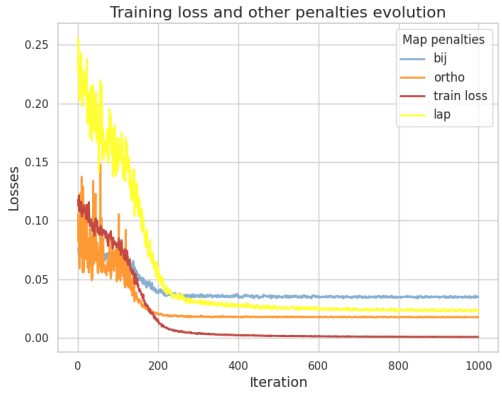

Figure 8. Loss and other penalties during optimization of the matching.

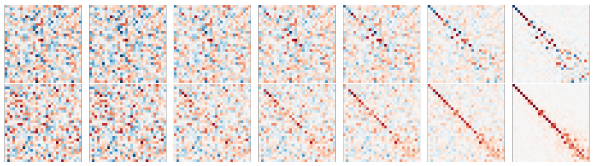

Figure 9. Example generation trajectories using spectral diffusion models on functional maps (top) and absolute functional maps (bottom)

the variance-preserving SDE, which reproduces the trajectory of DDPM [12]. No normalization of functional maps is applied, as the values inside the matrices range from -1 to 1 already.

Zero-Shot Training. We train our deep functional map approach for 1000 gradient steps using Adam optimizer. The overall training on a single pair takes approximately 180 seconds on a NVIDIA L40S GPU.

Evaluation. For the evaluation, we refined our optimized maps using Zoomout to obtain a final map dimension of 150x150, as commonly done in the deep functional maps approach [4, 15].

12. Loss Behavior

We plot the loss behavior during optimization in Figure 8. The loss is smoothly optimized and converges rapidly.

13. Generating Functional Maps and Absolute Functional Maps

We show two example denoising trajectories, from the original and absolute spectral diffusion models in Figure 9.

Figure 10. After applying DiffuMatch, we select a point on the source shape and compute the distance of this point to all points on both target and source shapes, in the descriptor space. We plot the obtained distances on both shapes. The closest points are points that are geodesically close to the select point.

14. Quality of Learned Descriptors

Learned descriptors using our approach are meaningful thanks to our proper loss. Indeed, it has been shown that when properness is encouraged, the extracted correspondence is approximately the same whether it is extracted from the functional map or by nearest neighbor search [5]. We visually verify this in Figure 10, where we show the nearest points to a selected point using nearest neighbor in the feature space (after projection on the space spanned by the first 30 eigenfunctions – the only ones used in the map computation), showing that our method enables meaningful descriptor learning in addition to the quality of the shape matching.

15. Comparison of Reconstruction of Deformation Models

As stated in the paper, deformation models are not suitable for generalization to new type of categories. In this section, we provide reconstructions from 3D-CODED and NJF of the source shape in section 4.2. Moreover, as SNK provides a shape reconstruction as output, we also show the reconstruction provided by SNK. Finally, we extract shape correspondence Π from DiffuMatch and reconstruct the vertex position of the shape in the target mesh topology, by solving for the closest possible solution minimizing the As-Rigid-As-Possible (ARAP) [23]. Let X be the vertex of the source mesh, the reconstruction Y_{rec} in the target mesh topology is given by:

$$Y_{rec} = \underset{Y}{\text{argmin}} ||Y - \Pi X|| + E_{arap}(Y).$$

As our matching is nearly perfect, the provided reconstruction, shown in Figure 11 is visually better than the one

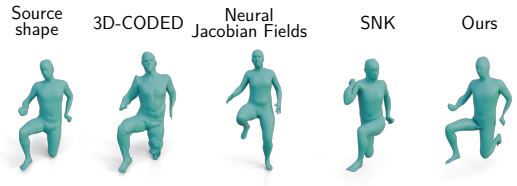

Figure 11. Reconstruction of source shape using different approaches. For our reconstruction, we solve for the closest vertex positions to the matched shape minimizing the ARAP energy from the target shape.

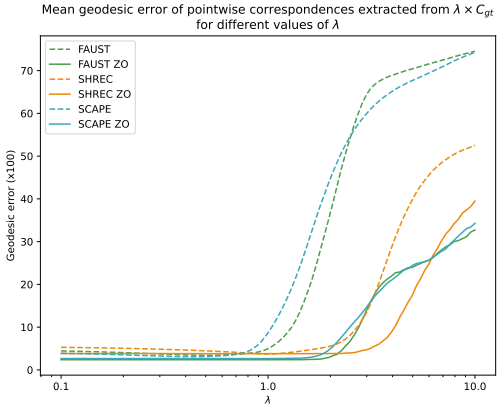

Figure 12. Given a ground truth functional map $C \in \mathcal{M}_n(\mathbb{R})$, and a scalar, $0 < \lambda < 1$, the matrix λC represents approximately the same pointwise correspondence as C. By applying Zoomout to both C and λC, we obtain the same map. The observation does not always hold when $lambda > 1$. We plot the geodesic errors of λC for different values of λ.

given by other approaches, up to some artifacts due to our matching being computing on the first 30 eigenfunctions only. The capabilities of our model can also be extended to reconstruction of input meshes in a new topologies.

16. Importance of Mask Regularization.

Mask regularization plays a key role in most (deep) functional map pipelines [3]. We run a simple experiment to show that the functional map space is particularly well-suited for this type of penalty. Multiplying a ground truth functional map matrix $C \in \mathcal{M}_n(\mathbb{R})$ by any scalar $0 < \lambda < 1$ raises approximately the same pointwise correspondence as the original one from C. We also observed the same phenomena after applying Zoomout [18], where the obtained correspondences are the same. This phenomenon is illustrated in Figure 12.

As most masks are sparsity-promoting masks, their mask penalty minimizers have multiple solutions, which are $\lambda \times X$ where X is any solution. As we observed, optimized maps can be proportional to the ground truth solution and still output a correct pointwise correspondence. Based on this insight and the efficiency of mask regularization in functional map computation, we proposed to distill the knowledge of our trained diffusion model by extracting a sign-agnostic mask that will promote structures seen in the training set.

Method	Computation time
3D-Coded	160s
Neural Jacobian Fields	3.26s
SimplifiedFmaps	1.08s
SNK	130s
Ini + Zoomout (our mask)	0.75s
Ours full	150s

Table 3. Computation costs for different methods.

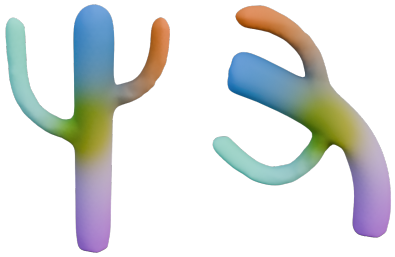

Figure 13. DiffuMatch result on a cactus pair.

Figure 14. Partial matching results on SHREC16

17. Computation Time

A single run of DiffuMatch takes approximately 150 seconds on an NVIDIA L40S GPU. In the case where computation time is a bottleneck, the scenario Ini (feature extractor with random weights) + Zoomout *with our distilled mask* is competitive as it requires little computation time. We provide a comparison of computation time with some other competing methods in Tab. 3

18. Generalization

Non articulated shapes We showcase that DiffuMatch can perform well on a pair of two cactus meshes in Fig. 13.
Partial shape matching We show in Fig 14 some partial matching results. DiffuMatch can work on pairs where the partiality is moderate. However, when the partiality becomes significant, DiffuMatch is prone to failure, with an

error of 19.8 and 23.4 on SHREC16 cuts and holes partial shape matching challenges [21]. This is to be expected, as functional maps have a different structure between full and partial correspondence [21], and methods applied to partial shape matching often rely on modified losses [7, 8] or require feature pre-training [8].

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. 1

[2] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: Shape completion and animation of people. 2023. 1

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

[4] 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. 1, 2

[5] 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. 2

[6] 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] Dongliang Cao and Florian Bernard. Unsupervised deep multi-shape matching. In *European Conference on Computer Vision*, pages 55–71. Springer, 2022. 4

[8] Dongliang Cao, Paul Roetzer, and Florian Bernard. Unsupervised learning of robust spectral shape matching. *ACM Transactions on Graphics (TOG)*, 42:1 – 15, 2023. 4

[9] 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. 1

[10] 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. 1

[11] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. 3d-coded : 3d correspondences by deep deformation. In *ECCV*, 2018. 1

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

[13] 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. 1

[14] 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. 1

[15] 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. 1, 2

[16] 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. 1

[17] 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. 1

[18] 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. 3

[19] 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. 1

[20] 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. 1

[21] 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. 4

[22] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional maps for shape matching. *Advances in Neural Information Processing Systems*, 33:19264–19275, 2020. 1

[23] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, pages 109–116. Citeseer, 2007. 2

[24] Dong-Ming Yan, Guanbo Bao, Xiaopeng Zhang, and Peter Wonka. Low-resolution remeshing using the localized restricted voronoi diagram. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2014. 1

[25] Silvia Zuffi, Angjoo Kanazawa, David W Jacobs, and Michael J Black. 3d menagerie: Modeling the 3d shape and pose of animals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6365–6373, 2017. 1