Deep orientation-aware functional maps: Tackling symmetry issues in Shape Matching: Supplementary Material

Nicolas Donati
LIX, École Polytechnique
nicolas.donati@polytechnique.edu

Etienne Corman
CNRS, Inria
etienne.corman@cnrs.fr

Maks Ovsjanikov
LIX, École Polytechnique
maks@lix.polytechnique.fr

In this document we provide the proof to the main theorems, more qualitative results, an extensive ablation study and additional insights about the approach described in the main manuscript.

1. Proof of Theorem 1, 2, 3

Theorem 1. *Given a set of shapes $\{S_i\}$ that all contain an* orientation reversing *isometric self-symmetry $\{T_i : S_i \to S_i\}$, s.t. $d_{S_i}(x_j, x_k) = d_{S_i}(T_i(x_j), T_i(x_k))$, then a generic neural network $\mathcal{F}_\Theta$ that is trained by any of the losses introduced in [12, 10, 9, 13, 3] has at least two possible solutions that both lead to the global optimum of the loss.*

Proof. The spectral losses L defined in [12, 10, 9, 13, 3] are fully intrinsic, thus they are invariant under shape isometric changes *i.e.* $L \circ T_i = L$. So, if all shapes admits an isometric self-symmetry, the solution composed with the isometry will have the same loss value. □

Theorem 2. *The complex-linear map Q is a pushforward if and only if there exists an* orientation-preserving *and conformal diffeomorphism $\varphi : M \to N$ satisfying:*

$$\langle X, \nabla C(f) \rangle_{TM} = C\left(\langle Q(X), \nabla f \rangle_{TN}\right), \tag{1}$$

for all $X \in TM, f \in L^2(N)$.

Proof. See Theorem 3.1 in [6], Section 3.5 □

Theorem 3. *Let M, N be two manifolds, and F_M, F_N surface features such that the functional map C estimated from these features is an isometry. Let Q be the complex functional map computed with the feature gradients as described in the main manuscript. Then the maps (C, Q) satisfy Eq. (1), and C is an orientation-preserving isometry.*

Proof. By assumption the functional map $C : L^2(N) \to L^2(M)$ represents the isometric map $\varphi : M \to N$ and exactly transfers the descriptors *i.e.* $C(F_N) = F_M$. Moreover the complex functional map $Q : TM \to TN$ transfers the gradient of the descriptors $Q(\nabla_M F_M) = \nabla_N F_N$ and is complex-linear.

To recover Eq. (1), we take the inner product of the gradient transfer with the gradient of an arbitrary function $f : N \to \mathbb{R}$:

$$\mathbf{g}_p^N\left(Q(\nabla_M F_M), \nabla_N f\right) = \mathbf{g}_p^N\left(\nabla_N F_N, \nabla_N f\right).$$

This equation easily simplifies using the properties of an isometric map: the metric is preserved by the pullback ($\varphi^\star \mathbf{g}^N = \mathbf{g}^M$) and the pushforward commutes with the gradient ($\mathrm{d}\varphi^{-1}(\nabla_N f) = \nabla_M C(f)$), yielding:

$$\begin{aligned}
&\mathbf{g}_p^N\left(Q(\nabla_M F_M), \nabla_N f\right) \\
&= \left((\varphi^\star)^{-1} \mathbf{g}^M\right)_{\varphi^{-1}(p)} \left(\mathrm{d}\varphi^{-1}(\nabla_N F_N), \mathrm{d}\varphi^{-1}(\nabla_N f)\right) \\
&= \mathbf{g}^M_{\varphi^{-1}(p)}\left(\nabla_M C(F_N), \nabla_M C(f)\right)
\end{aligned}$$

So Q and C satisfy Eq. (1) for all complex-linear combination of the gradient descriptors. Therefore, following Thm. 2, Q is the pushforward associated to φ and C must be orientation preserving. □

Meth / Data	SMAL_r
xyz input-3 axis	25.
xyz input-1 axis	5.9
nonOA-FE	34.
no Q-maps (epoch 3)	5.8
no Q-maps (epoch 15)	8.1
Ours (epoch 3)	**4.8**
Ours (epoch 15)	5.1

Table 1. Comparative results ($\times$100) for the different ablations of our method.

2. Ablation Study

This section presents an ablation study for the most vital components of our approach, namely: *a)* The input signal fed to the network, *b)* The orientation-aware feature extractor *c)* The orientation-aware loss. We test these ablations on

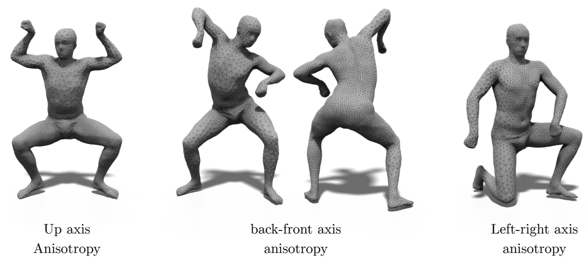

Figure 1. SCAPE [1] dataset remeshed in an anisotropic fashion, used in the first experiment of the main manuscript

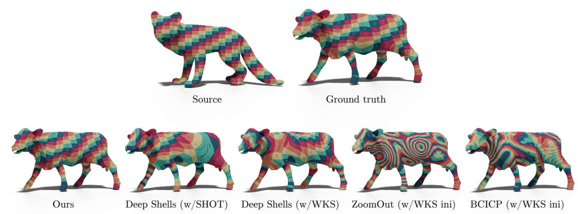

Figure 2. Qualitative comparison to baselines on the SMAL dataset, using texture transfer from source to target shape. Only our method gives accurate correspondence, whereas in this challenging case baselines completely fail to predict the map.

our third experiment, on the SMAL dataset [16] (see main manuscript, Section 5.1 for more details), and to a maximum of 20 epochs. We compare these three ablations to our approach and report the results in Table 1. The ablations are commented in details in the sections below.

2.1. The WKS descriptors as input features

As stated in the main manuscript, many unsupervised deep learning for non-rigid 3D shape matching rely on SHOT descriptors [15] as input signal for the neural network to produce correspondences between shapes [10, 12, 8]. This descriptor is orientation-aware but very sensitive to the input triangulation, resulting in overfitting with the training triangulation as demonstrated in [7], and also in the first experiment of the main manuscript, with anisotropic remeshings.

Therefore we use an input descriptor that is agnostic to the input triangulation so as to not overfit to it: the WKS descriptor [2], which is built using the eigenvectors and eigenvalues of the Laplace-Beltrami operator. These descriptor functions h_k are therefore fully intrinsic, and will display the same intrinsic self-symmetries as the shapes themselves. Namely, with the notations of Theorem 1, $h_k \circ T = h_k$.

Another commonly used option for an input signal is the 3-dimensional extrinsic coordinates of the shape points, as done in [7]. However, this input signal is dependent on the shape orientation in space. Consequently, the input data needs to be centered, and augmented by adding randomly rotated poses. With this input signal, the method is no longer fully intrinsic and therefore potentially unstable to rotations of the input shapes. For this first ablation experiment, we train our method with this input signal (denoted as "xyz input" in Table 1) instead of WKS descriptors. To make the experiment more complete, we report the results with two different data augmentations: *a)* The general case, where there is no prior on the shapes alignment, so the data need to be augmented with all 3D rotations (3 parameters space). We denote this data augmentation as "3 axis" in Table 1. *b)* The special case where the input shapes are all aligned to one axis, but potentially rotated around this axis, so the data needs to be augmented around this axis (1 parameter space). We denote this data augmentation as "1 axis" in Table 1. We stress the fact that this kind of prior on the shapes rigid alignment already makes the method weakly supervised.

We see in Table 1 that even with the prior of shapes aligned to one axis, our method is better (and more general) when trained with WKS descriptors as an input signal.

2.2. The orientation-aware feature extractor

To make our approach unsupervised, it is crucial that the feature extractor should be orientation-aware. Indeed, since we train on shapes exhibiting an isometric self symmetry (the left-right symmetry present in all animal shapes), the only way to disambiguate between left and right is through orientation, since the symmetric map reverses this orientation. DiffusionNet [14] uses gradient features to incorporate this orientation information into potentially symmetric inputs (e.g. WKS descriptors in our case). For this second ablation, we propose to show that without this orientation-aware feature extractor, the method fails to produce informative descriptors, and report the results in Table 1, on row "nonOA-FE" (standing for non orientation-aware feature extractor).

To that end, we deactivate the gradient-based blocks of DiffusionNet, which results in a new orientation-agnostic feature extractor which can still produce excellent results [14]. We then train our method using this feature extractor and WKS as input signal. We see in Table 1 that this ablation greatly impairs the method.

2.3. The complex functional maps block and the orientation-aware loss

We remove the complex functional map block from the loss by setting $w_{\text{Q-ortho}} = 0$. As discussed in Theorem 1, the resulting network is not guaranteed or encouraged to produce orientation-preserving correspondence. We report the result of this ablation in Table 1, on rows "no Q-maps".

We observe that this ablation still seems to converge to

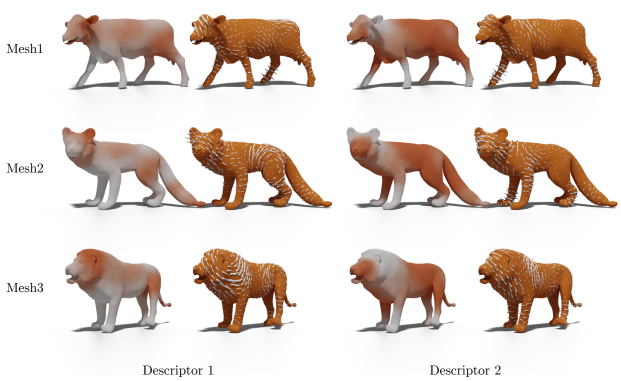

Figure 3. Visualization of two different scalar descriptors learned by our method, along with their vector valued counterparts, on 3 mesh from the SMAL dataset. Contrary to the descriptors produced by [7], these descriptors are fully intrinsic and generally not localised. However, we see that both our scalar and vector valued descriptors are robust to strong distortions.

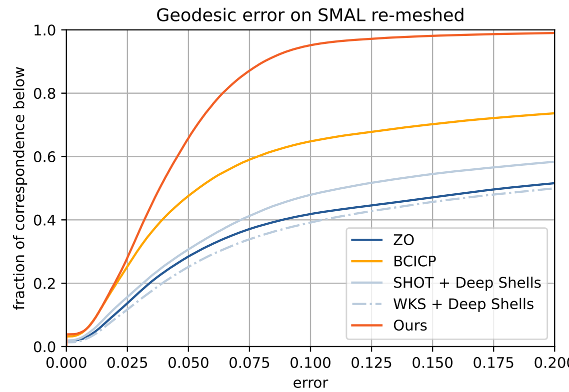

Figure 4. Quantitative results of the different methods using the protocol introduced in [11], on the SMAL remeshed test set (third experiment of the main manuscript).

well oriented maps in this case. This may be explained by the fact that DiffusionNet can produce non-symmetric descriptors from symmetric input like WKS, using shape orientation through gradients. Therefore, if two input shapes are consistently oriented, the symmetric input signal will be broken “in the same direction” by DiffusionNet gradient-based blocks. Conversely, if two shapes are non-consistently oriented (e.g. one with inward normals, one with outwards normals), the symmetry will be broken “in opposite directions”. In fact, using this remark one can retrieve symmetrized output descriptor functions (by symmetrized, here we mean composited with the intrinsic symmetric map T_i of the shape S_i) generated by DiffusionNet from symmetric descriptors such as WKS, by simulating a change in shape orientation (which corresponds to a conjugation operation on the tangent bundle structure, or more practically to setting `gradY = -gradY` in DiffusionNet gradient operator entries).

In practice, a second beneficial effect of our complex functional map loss is the reduction of overfitting. Indeed, in the experiment reported in Table 3 of the main manuscript, the train set is made of 32 SMAL re-meshed shapes and the test set is made of 17 shapes other SMAL remeshed shapes. Learning methods are thus liable to overfit to their training set. We see in Table 1 (where we report the geodesic error at epoch 3 and epoch 15 both with and without the complex functional map layer/loss) that without the complex functional map loss, the method is more prone to overfitting, as it looses generality if trained for too many epochs. To summarize, our complex functional map block and loss theoretically guarantee our approach to be orientation-preserving, and in practice also improve the pipeline stability with respect to overfitting.

3. More quantitative results

For completeness, we also report in Figure 4 the accuracy of our method and some baselines on the third experiment we conducted in the main manuscript (train on 32 SMAL remeshed shapes, test on 17 other SMAL remeshed shapes), using the evaluation protocol introduced in [11]. We see that our method gives the best correspondence quality by far, as in this case it always finds the well-oriented map (which results in the tail of the error curve being very close to 1).

4. More qualitative results

4.1. Anisotropic remeshing

In Figure 1, we show the anisotropic remeshing of SCAPE dataset [1], generated with Mmg [5, 4]. We use this anosotropic remeshing in the first experiment of the main manuscript to show that SHOT [15] based learning methods do not generalize to unseen triangulation. Indeed, we see in Figure 1 that the triangle scale is a function of the element coordinate. For the first seven shapes of SCAPE test set, we constrain the triangle size to be dependent on the position on the up axis. For the next seven shapes, we constraint the triangle size to be dependent on the position on the back-front axis. For the remaining six shapes, we constraint the triangle size to be dependent on the position on the left-right axis. With this remeshing, a network overfitting the triangulation combinatorics will most likely fail to predict the desired map. Our method, which is triangulation agnostic, remains almost unaltered, as shown in the first experiment of the main manuscript.

4.2. Another qualitative comparison on SMAL

We report in Figure 2 a second texture transfer performed by baselines compared with our method on the SMAL test shapes. For this example the distortion between the two shapes is stronger than in the one displayed in the main manuscript. However, our method still manages to predict accurate correspondences, while baselines fail to produce even a reasonable mapping in this case.

4.3. Visualization of the scalar & vector valued descriptors learned by our method

Lastly, we propose to visualize descriptors learned by our network, also on the SMAL dataset, in Figure 3. Since our method also exploits the gradients of the scalar descriptors learned by DiffusionNet, we also visualize these gradients (here rotated by $\pi/2$ to better make singularities stand out). Our method enforces learned descriptors *and their gradients* to correspond between source and target shapes, which was not done in any previous work to the best of our knowledge. Consequently, the features obtained with our method are all the more robust, since their gradients are also well preserved under shape non-rigid deformation.

References

[1] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics*, 24(3):408–416, July 2005. 2, 3

[2] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 1626–1633. IEEE, 2011. 2

[3] Mehmet Aygün, Zorah Lähner, and Daniel Cremers. Unsupervised dense shape correspondence using heat kernels. In *2020 International Conference on 3D Vision (3DV)*, pages 573–582. IEEE, 2020. 1

[4] Charles Dapogny, Cécile Dobrzynski, and Pascal Frey. Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems. *Journal of computational physics*, 262:358–378, 2014. 3

[5] Cécile Dobrzynski and Pascal Frey. Anisotropic delaunay mesh adaptation for unsteady simulations. In *Proceedings of the 17th international Meshing Roundtable*, pages 177–194. Springer, 2008. 3

[6] Nicolas Donati, Etienne Corman, Simone Melzi, and Maks Ovsjanikov. Complex functional maps: A conformal link between tangent bundles. *Computer Graphics Forum*, 41(1):317–334, 2022. 1

[7] 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. 2, 3

[8] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. Deep shells: Unsupervised shape correspondence with optimal transport. *Advances in Neural Information Processing Systems*, 34, 2020. 2

[9] Dvir Ginzburg and Dan Raviv. Cyclic functional mapping: Self-supervised correspondence between non-isometric deformable shapes. In *European Conference on Computer Vision*, pages 36–52. Springer, 2020. 1

[10] 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. 1, 2

[11] Vladimir G Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. In *ACM Transactions on Graphics (TOG)*, volume 30, page 79. ACM, 2011. 3

[12] 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. 1, 2

[13] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional maps for shape matching. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19264–19275. Curran Associates, Inc., 2020. 1

[14] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces, 2020. 2

[15] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *Proc. ECCV*, pages 356–369. Springer, 2010. 2, 3

[16] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017*, pages 5524–5532, Piscataway, NJ, USA, July 2017. IEEE. 2