Eurographics Symposium on Geometry Processing 2022
M. Campen and M. Spagnuolo
(Guest Editors)

Volume 41 (2022), Number 5

Supplementary Material of Localized Shape Modelling with Global Coherence: An Inverse Spectral Approach

M. Pegoraro[1] and S. Melzi[1,2] and U. Castellani[3] and R. Marin[†1] and E. Rodolà[†1]

[1]Sapienza University of Rome, Italy
[2]Bicocca University of Milan, Italy
[3]University of Verona, Italy

Abstract

As mentioned in the main manuscript, we report further details on our experiments and additional results in this document. We organize the supplementary material as follows: in Section 1 we report details about the data used in our experiments and architectures; in Section 2 we show further results of our analysis on the reconstruction quality with different combinations of operators, spectrum sizes and localizations; in Section 4 we perform further experiments on the semantic control by local and global switching and interpolations; in Section 2.8 we show a static Figure of the interactive application shown in the attached video; in Section 5 we report more results on unorganized point clouds training.

CCS Concepts

• ***Computing methodologies*** → ***Shape analysis;*** *Shape representations;*

1. Experimental Setup

1.1. Dataset

Here we report additional details about the datasets involved in our experiments.

CUBE. In the CUBE dataset, the local variations are extrusions of simple geometric patterns (circle, ellipsis, square, and rectangle) applied on a selected face same for all cuboids. We vary dimensions and rotations of these patterns avoiding isometric shapes (that are indistinguishable by their eigenvalues). For the global variations, we scaled the cube along the dimension orthogonal to the face with local variation by a factor in the range $[0.6, 2]$, obtaining cuboids with different depths.

SURREAL. The shapes in SURREAL are generated by SMPL [LMR*15], a standard generative template with 6890 vertices and two sets of parameters: one for the subject identity and one for its pose. Since pose changes generate near-isometric shapes, we set all the individuals in the same T-pose. The shape parameters are sampled from the ones available from SURREAL dataset [VRM*17].

AIRPLANES. In the AIRPLANES dataset we chose the segment of the tail as the local region because we think that the tail is a semantically significant region of the airplane: it is related to the airplane type (e.g., Boeing, Jet, Fighter) and its size.

1.2. Architecture and training details

Our architecture is a simple decoder composed of 4 fully connected layers. All the hidden layers use batch normalization followed by a selu activation, while the last layer has a linear activation. We report the number of nodes for each layer in Tab. 1. For the SURREAL dataset we add a dropout layer with a 0.1 drop rate to all hidden layers. We trained our network on 90% of the dataset and used the remaining 10% for testing. During training we used Adam optimizer with a learning rate of $2 * 10^{-3}$ for the first 1000 epochs and then we reduce it to $1.8 * 10^{-3}$ for the rest of the training. We fixed the maximum number of epoch in each dataset making sure each method reached convergence. The output of Π is a matrix $X \in \mathbb{R}^{n \times 3}$ encoding the vertex coordinates. In the second part of Table 1 we show the training parameters.

1.3. Computation time

We trained all the models on a NVIDIA GeForce GTX 1050 Ti. On the CUBE, SURREAL and SMAL dataset, the training time is about 2.2 hours; while on the AIRPLANES dataset is about 12 hours.

† Equal contribution

	CUBE	SURREAL	SMAL	AIRPLANES
Number of Nodes				
Layer 1	258	258	258	258
Layer 2	1024	512	512	1024
Layer 3	2048	1536	1536	2048
Layer 4	22050	20670	20670	1500
Output size	7350 x 3	6890 x 3	3889 x 3	500 x 3
Number of epochs	2000	1000	1000	4500
Batch size	64	32	32	8

Table 1: *Networks parameters for the different datasets involved in our experiments.*

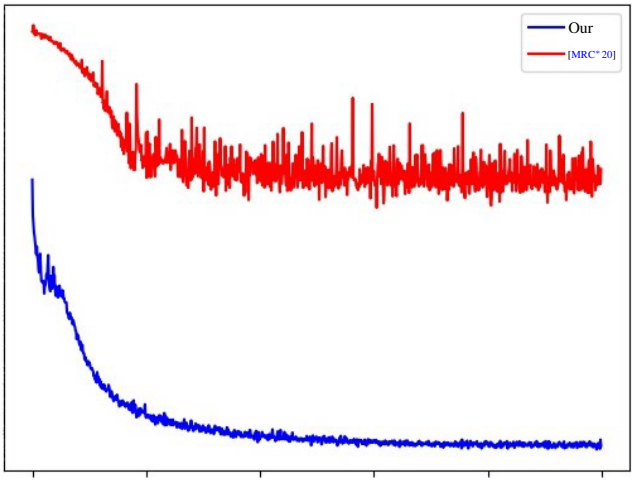

Figure 1: *Training loss comparison: our method in blue; in red the method proposed by [MRC*20].*

1.4. Comparison to the autoencoder of [MRC*20]

One of the main advantage of our method is the simplicity of the model: a single decoder composed of fully connected layers. This allows us to perform a more direct analysis of the linkage between spectral geometry processing and semantic modeling. On the contrary the model proposed by [MRC*20] is composed of an autoencoder enhanced with an invertible module blurring a similar analysis. In fact, the correspondence between the spectrum and the object geometry established by [MRC*20] passes through a latent space impacted by other components. Moreover, our specialized architecture performs better than the architecture proposed in [MRC*20] in synthesizing shapes from the spectrum. An other advantage of our model choice is the training. Fig. 1 shows the training loss curves of our model (in blue) and of [MRC*20] (in red). It emerges that our model not only reaches lower errors but has also a more stable training.

2. Shape from spectrum

This section provides further results on our analysis of the reconstruction of a 3D shape from its spectrum. If not differently stated, the shapes we adopt in all our experiments and figures have never been seen during the training and belong to the test set or a completely different dataset.

2.1. Different evaluation metrics and Nearest-Neighbor comparison

Evaluation metrics. In the main manuscript, we considered two extrinsic measures (global and localized MSE) and an intrinsic measure (Area error). Here we report a more complete analysis, including additional metrics.

As extrinsic measures, we report the same error optimized by the loss, referred to as **MSE**. With **MSE-$\mathcal{R}$** and **MSE-$\mathcal{R}^C$** we denote the same measure computed inside or outside $\mathcal{R}$. As intrinsic measures, we consider:

Area: as in the main manuscript, it is the average difference of the area elements of each vertex, which relates to surface stretch.

Metric: the vertex-wise metric distortion, computed as the difference in the geodesic distances from a fixed set of 100 uniformly sampled points to all the points in the mesh.

Align: the MSE reconstruction error of the local region after the best rigid alignment, obtained by solving the Procrustes problem between the local patches.

The **Area-$\mathcal{R}$** and **Metric-$\mathcal{R}$** are the same as above, computed for the local region.

Nearest-Neighbor comparison. We directly compare each method against the nearest-neighbor baseline; as done in [MRC*20], given the spectrum of a new test shape, we look in the training set for the spectrum which is the nearest in the L2 sense. Then, we consider as baseline output the training shape associated with this spectrum. To compare with this baseline we include **ENN** which measures the MSE of the baseline, and **EM** $<$ **ENN**, which indicates the percentage in which the method outperforms the baseline. Every method uses a different dataset, so each one has its baseline; this is why we considered it a column rather than multiple rows.

2.2. Further Results

All the results are summarized in Tables 2, 3 and 4. The columns represent the evaluation metrics presented in the previous Section, while the rows are the different combinations of local and global spectrum.

2.3. CUBE

With the CUBE dataset, we want to test how our model associates the orthogonal set of global and local variations with respect to the spectra.

The first three rows of Table 2 shows the results of the global spectrum with a different number of eigenvalues: LBO k. All the methods have a similar MSE with LBO 50 slightly better. The slightly higher error of LBO 80 may be due to the increasing uncertainty that is generated when we compute a higher frequency. In fact, in the computation of the LBO, the higher is the eigenvalue and the higher is the error that can be added to the computation. This factor encourages us to keep a lower number of global eigenvalues, while adding eigenvalues from a local spectrum.

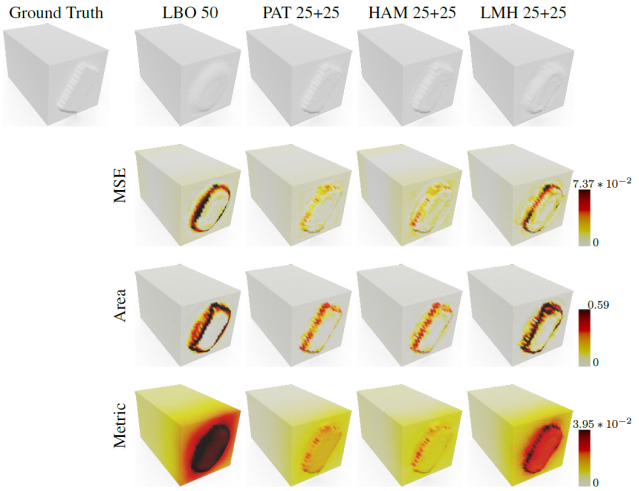

Figure 2: *Qualitative results on the CUBE. In the top left we display the ground truth shape. In the first row there are the reconstructions of different methods. In the following rows we plot on the reconstruction the extrinsic (MSE) and intrinsic (Area,Metric) measures. Errors are color-coded, growing from white to dark red.*

In Figure 2, we can see a qualitative example. In each row, we plot a different error on 4 different methods reported in Table 2: LBO 50, PAT 25+25, HAM 25+25, LMH 25+25. Overall the error accumulates on the border of the extrusion. This is due to the difficulty of the decoder to generate steeper details. LBO50 produces the smoothest result: we believe that this is linked to the absence of high frequency on the input spectrum. The MSE metric mildly highlights the back face of the cuboids, while the Area metric only concentrates on the pattern face. Even if, in Table 2, the MSE and Area mean error correlates, these qualitative differences allow us to distinguish the cause of the errors. In fact, we believe that the error on the back face is an accumulation error due to the difficulty of our model to stretch the triangles of the mesh.

This example supports our ideas on the limits of MSE. It represents a global measure that mixes up the structure and position of the shape. Therefore it is a good measure to validate the global accuracy of the reconstruction, but it hides where locally the reconstruction is better.

In Figure 5, we show the complete experiments of Figure 7 in the main paper. We trained our model with the same parameters on a second version of the CUBE dataset, where all the cube's faces manifest the same pattern. With this modification, a correlation between the selected region and the rest of the cube exists. Our method (second row) can effectively learn this correlation by modifying at the same time the shape of the local pattern on all the faces and the depth of the cube. The global interpolation (third row) only

Method	MSE ($*10^{-6}$)	MSE-$\mathcal{R}$ ($*10^{-6}$)	MSE-$\mathcal{R}^C$ ($*10^{-6}$)	ENN ($*10^{-6}$)	EM <ENN	Area ($*10^{-2}$)	Area-$\mathcal{R}$ ($*10^{-2}$)	Metric ($*10^{-3}$)	Metric-$\mathcal{R}$ ($*10^{-3}$)
LBO 30	11	62.5	0.65	15	60%	1.96	6.80	6.63	33.8
LBO 50	10.7	62	**0.45**	15	65%	1.80	6.76	6.41	3.33
LBO 80	11.1	63.5	0.63	14	61%	1.73	6.59	6.54	33.7
PAT 20+10	6.65	37.8	**0.42**	309	88%	1.78	6.25	4.81	24.2
PAT 15+15	5.66	27.1	1.36	1090	95%	2.31	6.01	3.96	17.5
PAT 10+20	4.91	26.3	0.63	1720	100%	**1.43**	5.37	3.59	17.9
PAT 40+10	6.76	38.5	**0.42**	79	86%	1.59	6.13	4.89	24.9
PAT 25+25	**3.59**	**19.1**	**0.5**	2060	100%	**1.33**	**4.52**	**2.76**	**13.9**
PAT 10+40	6.70	23.4	3.36	1860	100%	1.51	4.62	3.63	**13.2**
PAT 40+40	**3.77**	**18.3**	0.85	2310	100%	**1.35**	**4.44**	**2.84**	**12.8**
HAM 25+25	**4.07**	**20.1**	0.86	2050	99%	**1.33**	**4.56**	**3.05**	14.1
LMH 25+25	14.7	69.2	3.81	137	65%	1.96	6.85	6.08	27.7

Table 2: *Reconstruction error on the CUBE test set. For each column, we highlighted the top three results in red with decreasing intensity.*

Method	MSE ($*10^{-6}$)	MSE-$\mathcal{R}$ ($*10^{-6}$)	MSE-$\mathcal{R}^C$ ($*10^{-6}$)	ENN ($*10^{-6}$)	EM <ENN	Area ($*10^{-3}$)	Area-$\mathcal{R}$ ($*10^{-3}$)	Metric ($*10^{-3}$)	Metric-$\mathcal{R}$ ($*10^{-3}$)
LBO 30	1.7	2.12	1.61	15.4	99.57%	8.19	10.1	2.62	5.23
PAT 25+5	1.1	1.42	1.03	19.6	100%	15	23.9	2.21	3.32
PAT 20+10	**0.77**	0.75	**0.77**	28.8	100%	5.24	**5.09**	1.87	**2.25**
PAT 15+15	**0.71**	**0.5**	**0.76**	39.1	100%	**4.58**	**4.94**	**1.81**	**2.27**
HAM 25+5	0.98	1.13	0.95	19	100%	5.33	5.65	**1.93**	2.47
HAM 20+10	**0.73**	**0.67**	**0.75**	27.4	100%	**5.11**	**5.09**	**1.86**	**2.29**
HAM 15+15	0.86	**0.57**	0.92	38.9	100%	**5.10**	**5.21**	2.11	2.37
LMH 25+5	2.7	2.49	2.75	29.9	100%	10.6	11.7	3.37	5.71
LMH 20+10	2.4	2.06	2.47	32.5	100%	9.64	8.62	3.26	3.97
LMH 15+15	1.5	0.98	1.61	31.2	100%	7.15	6.16	2.62	2.67
[MRC*20]$_{big}$ LBO 30	2.51	2.26	2.56	15.09	99.96%	15.74	22.87	4.35	7.24
[MRC*20]$_{big}$ PAT 15+15	2.51	1.8	2.67	51.3	100%	16.9	22.25	4.52	9.28
[MRC*20] PAT 15+15	3.1	3.89	2.92	51.3	100%	19.81	32.93	4.46	8.44

Table 3: *Reconstruction error on SURREAL test set. For each column, we highlighted the top three results in red with decreasing intensity.*

changes the cube's length and leaves the pattern on the faces unchanged. Viceversa, the local interpolation (fourth row) changes as aspected the pattern in the selected region and the ones in the other faces but preserves the same cube's depth. These results confirm that our encoding can control the factors of variation both when they are correlated and when they are not, generating a shape that maintain a global stylistic coherence.

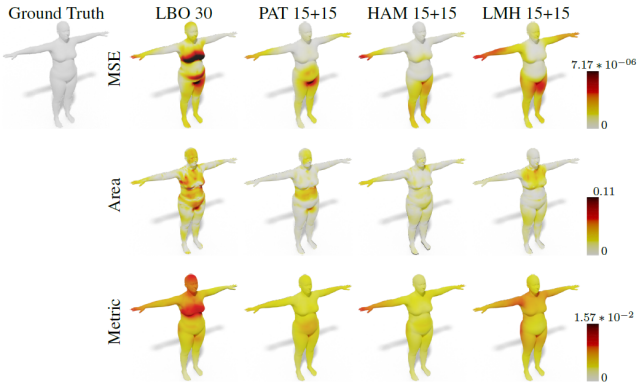

Figure 3: *Qualitative result on the SURREAL. The errors are shown on the reconstructed surfaces of a female with encoded color growing from white to dark red.*

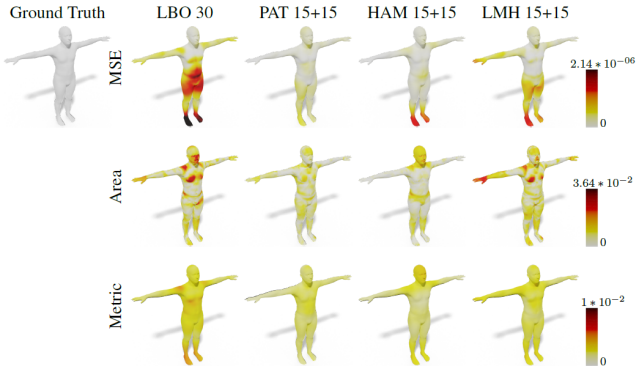

Figure 4: *Qualitative result on the SURREAL. The errors are shown on the reconstructed surfaces of a male with encoded color growing from white to dark red.*

2.4. SURREAL

The SURREAL dataset is our first case of realistic data. We perform an analysis similar to before. Firstly, we test our model with a global spectrum only, first row in Table 3. Then we consider different combinations of global and local spectra computed from different localized operators. In this analysis, we concentrate more on the performance of different localized operators.

As in the CUBE dataset, the local addition improves both the MSE and the intrinsic metrics. In particular, the head region reaches an error lower than the rest of the body. PAT 15+15 is the best combination, just followed by HAM 20+10. These results confirm once again the Laplace-Beltrami operator computed on a patch and the hamiltonian operator as the best-localized operators. Moreover, a combination of global and local representations in which they have similar proportions seems to hold the best results.

In this case, the ENN errors don't change drastically between global and localized operators, but it is still significantly higher than the MSE. The EM < ENN accuracy is 100% in all tests, except in LBO30. The similar values of ENN in this dataset allows us to make a clearer interpretation with respect to the one done with the CUBE dataset. In fact, while in the CUBE dataset the shapes are simple cuboids with details only on one face, the shapes in this dataset are more complex and have details that can vary all over the surface. Then, since in SURREAL the LBO can encode more variations at a global level than in the CUBE dataset, the ENN error is already higher in the global case and slightly increase in the local ones. This observation highlights the quantity of information encoded in a spectrum when the shape has greater or lesser details spread across the surface.

Qualitative examples of the results in Table 3 can be found in Figure 3 and in Figure 4. The former is a more robust woman, while the latter is a thin male. All the combinations are able to create shapes qualitatively similar to the ground truth. The addition of a local spectrum computed on the head greatly improves the reconstruction not only of the head, but also on the torso. We believe that our model learns to associate the information encoded on the head spectrum with other important features of a subject such as his robustness. Similar to the example in Figure 2, the Area errors accumulate in different parts than the MSE, allowing us to separate reconstruction errors due to the position in the space from the ones due to the "topology" of the shape. For instance, the errors in the hands are high in the MSE metric but low in Area one. This suggests that locally the hands are well reconstructed but globally, they are not in the right position because of an accumulation error on the vertices of the arm.

In the last three rows of Table 3 we report the errors obtained both with the best spectra combinations (PAT 15+15) and the global spectrum only. Results show not only that our decoder approach is better than the full architecture even in the LBO setting, but that [MRC*20] is not equally capable of combining local information with its latent space.

2.5. SMAL

Since animals have several regions which may encode some shape semantic (e.g., tail, head, paws, ...), we used SMAL to focus our analysis on the contribution of different local regions. In particular, we localize the operators on both the head and tail to see their impact on the generation capacity. For this reason, we modify the taxonomy of the Table 4: we consider two distinct regions error $\mathcal{H}$ and $\mathcal{T}$ that correspond respectively to the head and tail of the shape; we differentiate our model trained on a different region with the subscript $_H$ for the head region and $_T$ for the tail. We also change the columns of the errors computed on a portion of the surface. The MSE-$\mathcal{R}$ column is replaced with MSE-$\mathcal{H}$ that represents the error computed on the head region; while the MSE-$\mathcal{R}^C$ column is replaced with MSE-$\mathcal{T}$ that represents the error computed on the tail region. The same applies to the Area and Metric columns.

A difference with respect to the previous results is the ENN. Its values are higher than MSE, but with a lower gap. Moreover, the EM<ENN accuracy does not reach the 100% in any test. The lower gap could suggest that the spectra combination produces less distinctive encoding in this dataset. Therefore our model has more difficulty generating a more accurate shape. We also remember that SMAL is composed of different animal species and therefore has a higher variation. This characteristic allows the decoder to give more

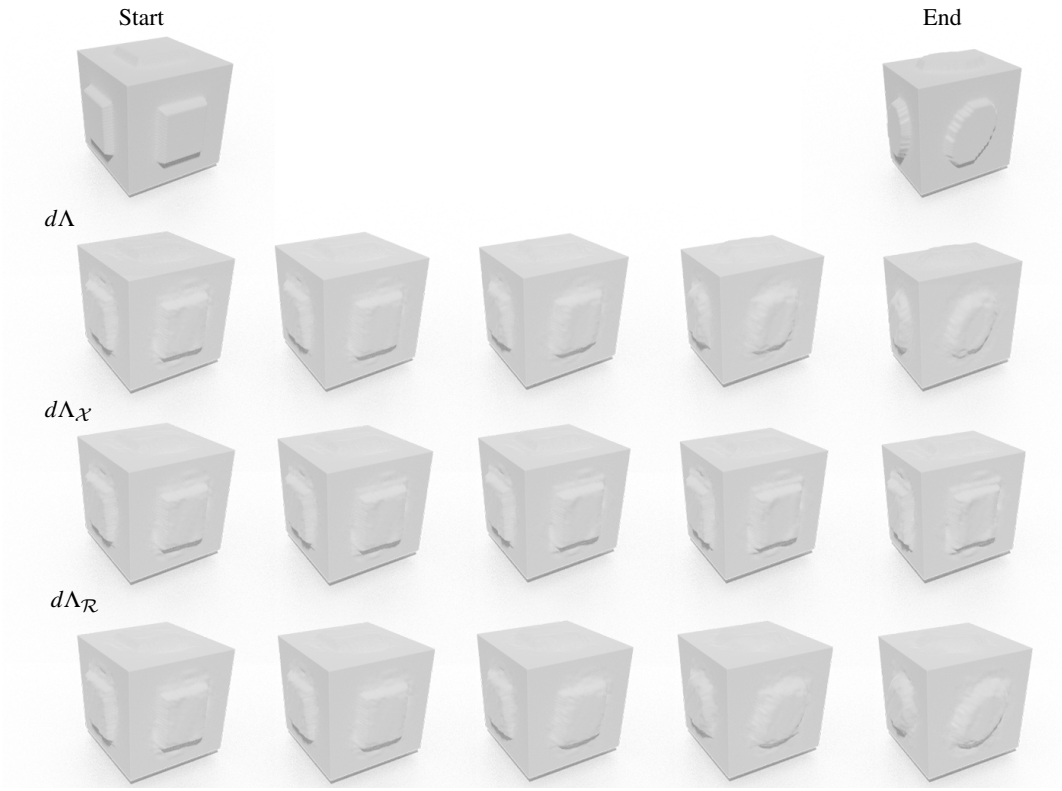

Figure 5: *Interpolation results for a pair of cubes with same patterns applied to the all the faces.* First row*: Two input cubes.* Second row*: Interpolation of the entire encoding.* Third row*: Interpolation of the global part only; observe how the patterns on the faces do not change, while the volume of the entire cube is correctly interpolated.* Last row*: Interpolation of the local part of the encoding (red in the bar plots), inducing a change in the patterns only.*

Method	MSE ($*10^{-6}$)	MSE-$\mathcal{H}$ ($*10^{-6}$)	MSE-$\mathcal{T}$ ($*10^{-6}$)	ENN ($*10^{-6}$)	EM <ENN	Area ($*10^{-2}$)	Area-$\mathcal{H}$ ($*10^{-2}$)	Area-$\mathcal{T}$ ($*10^{-2}$)	Metric ($*10^{-3}$)	Metric-$\mathcal{H}$ ($*10^{-3}$)	Metric-$\mathcal{T}$ ($*10^{-3}$)
LBO 30	**1.39**	**1.48**	**4.30**	3.84	80.2%	**1.90**	2.58	**4.25**	**3.58**	**7.40**	**23.6**
PAT_H 15+15	**1.08**	**1.07**	**3.07**	5.47	86.6%	**1.63**	**2.12**	**3.83**	**3.05**	**5.96**	**20.3**
PAT_T 15+15	3.93	4.20	11.5	11	64.6%	3.05	4.03	6.26	6.13	14.1	36.9
HAM_H 15+15	**1.11**	**1.13**	**3.50**	5.22	86.6%	**1.54**	**2.01**	**3.59**	**3.01**	**5.77**	**19.6**
HAM_T 15+15	3.13	3.35	8.81	8.17	68.2%	2.77	3.68	5.75	7.5	13	32.9
LMH_H 15+15	1.78	1.9	5.11	4.47	82%	1.93	**2.52**	4.47	3.62	7.6	24.5
LMH_T 15+15	3.57	3.8	9.76	13	67.4%	2.94	3.93	5.9	5.89	12.6	35.8

Table 4: *Reconstruction error on the SMAL test set. For each column, we highlighted the top three results in red with decreasing intensity.*

importance to the data encoded in the global spectrum since there are shapes that differ also in their global structure, i.e. an hippo compared to a tiger is shorter and bigger. This deduction is also enforced by the short gap between LBO 30 and PAT_H $15+15$ where the addition of the local spectrum brings less information with respect to the humans of SURREAL.

Figure 6 shows the different errors plotted on a tiger. Overall, all the methods generate shapes that are qualitative very similar to

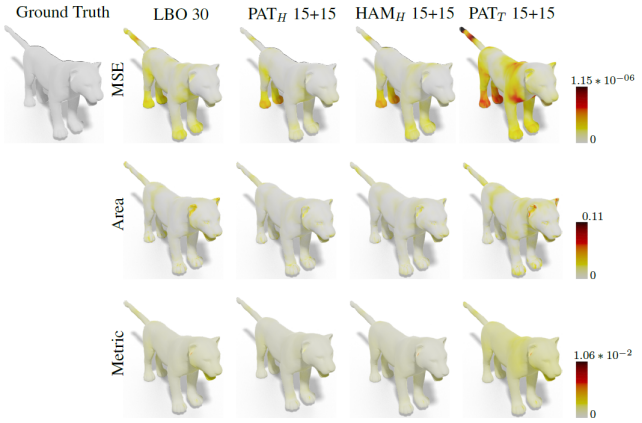

Figure 6: *Qualitative results on the SMAL with extrinsic and intrinsic measures plotted on the reconstructed surface of a tiger with a white to dark red colormap.*

the ground truth. PAT $_H$ $15+15$ has lower results not only on the head, but also on the tail. This suggests a correlation between the two regions. On the contrary PAT$_T$ $15+15$ has worse results. In particular, even though the local spectrum is computed on the tail, the error on that region is higher. This may be due to the inability of the tail spectrum to encode enough information relevant to the whole shape like the head does. As a consequence, the decoder has fewer informative features to generate the whole shape causing a sparse error that in the MSE highly accumulates on the tails.

Other qualitative examples from the SMAL dataset are reported in Figure 7. We show the MSE and Area measures of LBO 30, PAT$_H$ $15+15$ and PAT$_T$ $15+15$ on all the remaining classes. The error accumulates mainly on the characteristic region of the different animals: the tail and ears in the wolf, the ears and crest in the zebra, the ears in the cow and the snout in the hippo.

2.6. Main Insights of our analysis

We summarize the main insights of our analysis:

Global versus Local. The standard eigenvalue representation of [MRC*20] is outperformed by mixed PAT encoding, regardless of different $k+h$ values. Nevertheless, even in the presence of orthogonal transformations, introducing a local representation helps the global reconstruction as well. We believe this is possible only if the network can relate each operator spectrum with a shape variation.

Different localized operators. We see that maximizing the locality of the considered representation is beneficial. PAT emerges as the best representation, tightly followed by HAM which is almost equivalent. LMH performs the worst. The discussion on this point is further detailed in Section 5.4.

Locality proportion. Our experiments suggest that a balanced mix of global and local information provides the best representation for the inverse spectral problem. Also, increasing the number of eigenvalues is more beneficial in a mixed setup than in the standard LBO

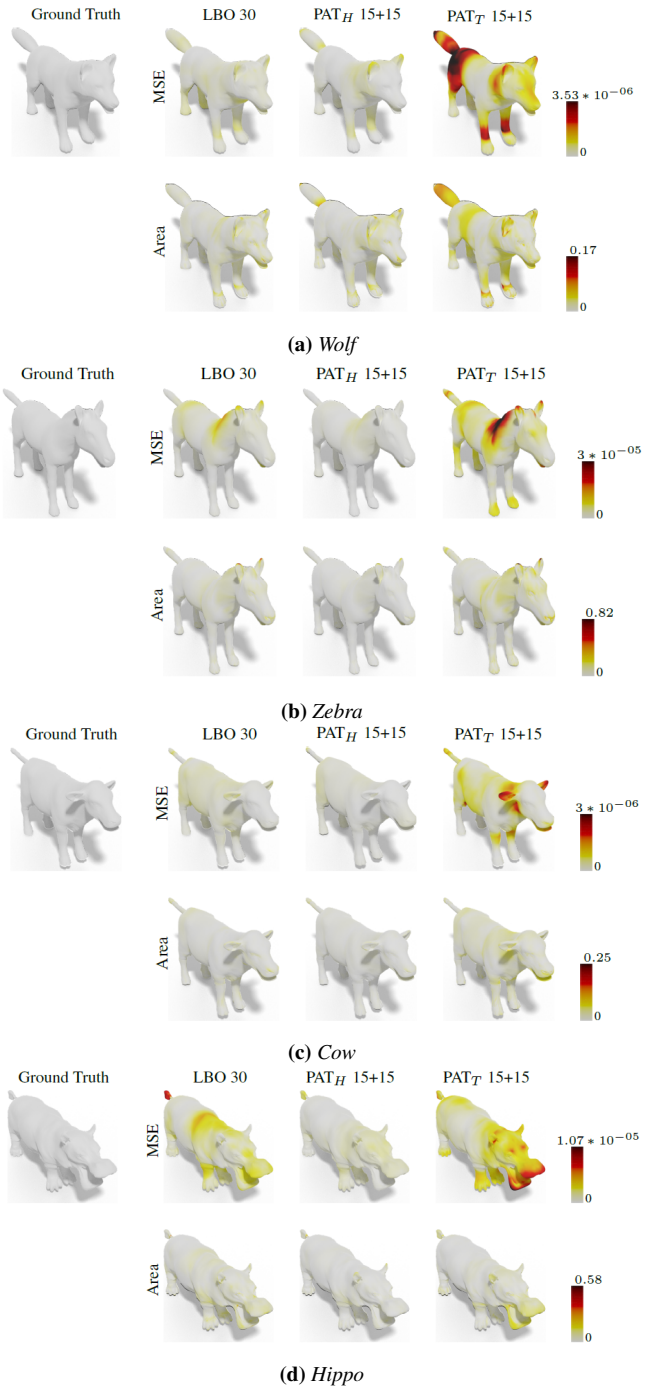

Figure 7: *Qualitative results on the SMAL with extrinsic and intrinsic measures plotted on the reconstructed surface of different classes with a white to dark red colormap.*

Figure 8: *Qualitative comparisons between different regions on the SURREAL. On the top left we display the ground truth with the different regions highlighted: in red the head $\mathcal{H}$; in green the forearm $\mathcal{F}$; in blue the torso $\mathcal{T}$. The rows are different methods, while the columns different measures. Errors are color-coded, growing from white to dark red.*

setup. We think the information rapidly faints in subsequent eigenvalues, while new operators provide a clearer pattern for the network to harvest.

Autoencoder vs decoder-only. Results show not only that our decoder approach is better than the full architecture even in the LBO setting, but that [MRC*20] is not equally capable of combining local information with its latent space.

Evaluation metrics. We emphasize that extrinsic metrics are not always reflected in the intrinsic ones. While the extrinsic measures directly test the network on the purpose of its training, our measures also reflect the model's intrinsic properties. We find them complementary, and encourage follow-up works to rely on similar measurements both for training and test.

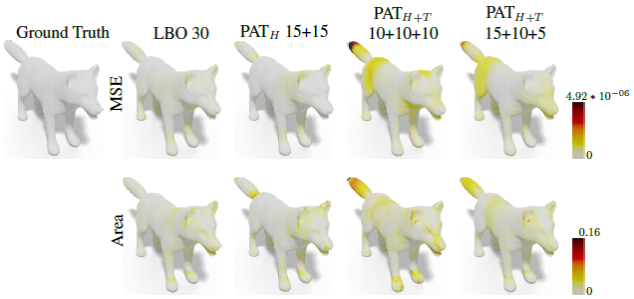

Figure 9: *Qualitative results of different regions on the SMAL. We plot on the reconstructed shape extrinsic (first row) and instrinsic (second row) measures with a colormap from white to dark red.*

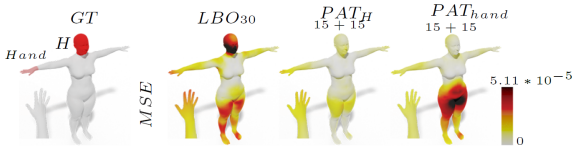

Figure 10: *An example of reconstruction comparing head or hand selection.*

2.7. Region selection results

In this section, we want to investigate more deeply the importance of $\mathcal{R}$

In Fig. 8, we show the mean squared and the area error for an example of shape reconstruction comparing PAT$_F$ $15+15$ and PAT$_T$ $15+15$ with LBO 30 and PAT $15+15$. Concerning LBO 30, we see that PAT$_F$ $15+15$ and PAT$_T$ $15+15$ presents a similar intrinsic error on the head. PAT$_F$ $15+15$ has a low Area error on the arms but high on the torso which produces a higher MSE also on the arm. On the contrary, PAT$_T$ $15+15$ is able to improve the torso eliminating the LBO error on the chest. In all cases, *PAT* $15+15$ performs significantly better.

2.8. Multi-region

Figure 9 shows a qualitative comparison of multi-region selection over SMAL. PAT$_{H+T}$ $15+10+5$ has a lower error than PAT$_{H+T}$ $10+10+10$ especially on the back of the shape. Since PAT$_{H+T}$ $15+10+5$ has a slightly lower MSE than PAT$_{H+T}$ $10+10+10$, we think that the performance may be correlated to the number

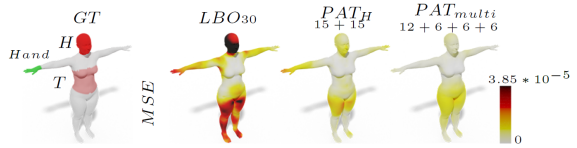

Figure 11: *An experiment of different representation for the input: on the right, it considers* 12 *eigenvalue from the global spectrum,* 6 *from the head,* 6 *from the torso, and* 6 *from the hand.*

of eigenvalues assign to each spectrum. In fact, the combination $15+10+5$ has fewer eigenvalues from the tail which is the region with the higher errors. As consequence, the decoder has a more informative encoding from which generates the correct shape.

The PAT_{H+T} $15+15+15$ combination seems to add enough information from all the spectra to improve the performance. In the last two columns of Figure 8 can be seen a significant improvement on the lower part of the body.

This last series of experiments highlight the importance of $\mathcal{R}$. Not all portions of $\mathcal{X}$ are good candidates as local regions since they don't have enough high-frequency information. This suggests the necessity of research of meaningful area on which to compute the local spectra. Moreover, the proportion of eigenvalues assigned to the different spectra affect just as much the overall performance.

Multiple-Local areas. Splitting between regions and global representation requires a careful design. From a general perspective, our experiments show that substituting part of the global information with some local one also provides better reconstruction in the global areas for many different scenarios. This is strengthened by the correlation between the local and global parts. However, pushing further the number of the local regions maybe not be trivial. As a complement of already seen results, we consider here (Figures 10 and 11) the hand region: in the first row as the only local region, in the second one in conjunction of head and torso. These experiments convey all the same message: each part should be represented with enough information. A study on the perfect balance between regions would be domain-dependent and exciting for future works.

3. Interactive manipulation

Our method can be efficiently used in real-time to synthetize shapes and control their deformations by acting on the different parts of our representation. An example can be seen in the *video* attached to this document. In that short sequence, we have two sliders that modify the different components of the proposed spectral encoding. On the left side of the video, we show the spectrum that we input to the network, highlighting with brighter color the part subject to the current modification. In gray, we kept the original spectrum as a reference. On the right side of the video, we visualize the shape produced by our model. The color depicted on the surface encodes the difference between two subsequent modification frames; this visualization helps to identify where the modification represented by the sliders is acting on the 3D geometry. Moreover, in Fig. 12 we report an illustrative image of free manipulation provided by our model. We chose a reference shape and modified its global and local encoding separately. Similarly to the video, for each shape, we highlight on the surface the area variations encoded by the colors. In the first row, we decrease the global part of the encoding generating alterations scattered on the body, but with minimal interaction with the head. In the second row, we increase the values of the local part of our encoding obtaining a more feminine physiognomy and variations localized on the head and thorax, while the legs are almost left unchanged.

4. Different representations from training time

In Fig. 13 we report an additional result on semantic control with different representations. We start from a sparse point cloud (3445 vertices), depicted on the left, from which we compute the global spectrum with the robust Laplacian [SC20] and combine it with the local spectrum from a mesh representing a different subject in a different pose, visualized in the middle. On the right we show the resulting shape, which maintains the identity of the second one, but with a thinner body like the first shape. We remark that the network is trained only on meshes; thus we appreciate the robustness of our model also to unseen and noisy data.

5. Unorganized pointclouds

Here we present other examples of our airplane experiments (Fig. 8 in the main paper).

In Fig. 14 we perform a spectrum switch. The two input planes have a similar tail but a different structure. Even in this subtle case, when we change only the local encoding, our method interpolates the two tails without modifying the airplane length, the presence of the turbines, and keeping the wings loyal to the starting plane. On the contrary, the global switch affects the whole plane like in the others interpolations experiments.

In Fig. 16 we test our model by looking at the shape generated from the spectra obtained by interpolating the input spectral encoding of two shapes (depicted on the left). In the first row, we report the results from the interpolation of the whole spectral encoding. We can see that the deformation is smooth both in size (i.e., length of the structure) and features (i.e., turbines appearing, tail morphing). In the second row, we fix the local part of the encoding, and interpolate the global. Coherently, changing the whole structure also requires changing the tail structure (different kinds of airplanes have different tails). Finally, in the third row, we only manipulate the local part maintaining the global one. The local interpolation mainly impacts the tail region (a close-up is depicted in Fig. 15), which follows the interpolation pattern of other rows. Remarkably, other global aspects of the airplanes are only slightly modified (i.e., the turbines and the shapes of the wings are almost left unchanged). We consider this result significant, since the spectrum of the tail seems representative enough to relate with different airplanes. Moreover, our global plus local spectral encoding provides nice interpolation results. In Fig. 17 we report the same example of Fig. 16, but using a 20+10 network instead of a 15+15 one. The results are consistent, showing a certain resilience to different settings.

References

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graph. 34*, 6 (2015), 248:1–248:16. 1

[MRC*20] MARIN R., RAMPINI A., CASTELLANI U., RODOLÀ E., OVSJANIKOV M., MELZI S.: Instant recovery of shape from spectrum via latent space connections. In *International Conference on 3D Vision (3DV)* (2020). 2, 3, 4, 5, 7, 8

[SC20] SHARP N., CRANE K.: A Laplacian for Nonmanifold Triangle Meshes. *Computer Graphics Forum (SGP) 39*, 5 (2020). 10

[VRM*17] VAROL G., ROMERO J., MARTIN X., MAHMOOD N., BLACK M. J., LAPTEV I., SCHMID C.: Learning from synthetic humans. In *CVPR* (2017). 1

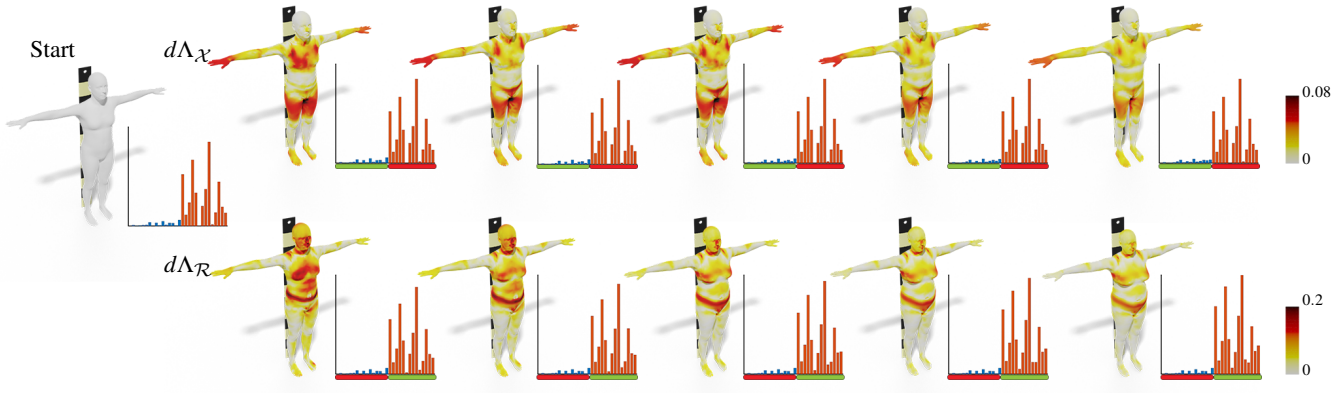

Figure 12: *Free manipulation of a SURREAL shape. Given a shape (on the left), we decrease the global values ($\Lambda_{\mathcal{X}}$-first row) and increase the local values ($\Lambda_{\mathcal{R}}$-second row) separately. For each shape, we plot the area variations for each vertex and show the correspondent encoding as barplot (blue:global, red:local). We highlight in green the values that changes and in red the values that we keep constant.*

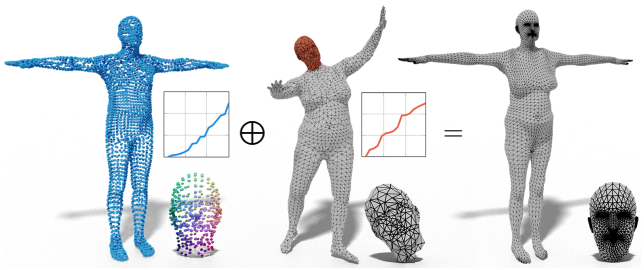

Figure 13: *Combination of global spectrum (in blue) of a point cloud (left) with a local spectrum (in red) from a mesh (center) with different discretization and pose.*

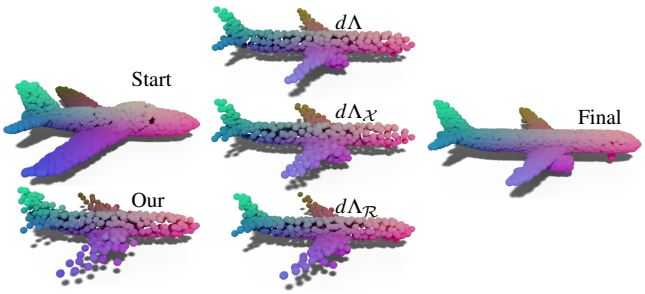

Figure 14: *A spectrum switch, similar to the interpolation shown in the main manuscript. On the left: the starting airplane and the corresponding output generated by our network. On the right, the second airplane. In the middle, from the top there are the reconstruction generated: using the whole second spectral enconding ($d\lambda$); concatenating the global spectral encoding of the second with the local one of the first($\Lambda_{\mathcal{X}}$); concatenating the global spectral encoding of the first with the local spectral encoding of the second($\Lambda_{\mathcal{R}}$). Notice how the first two impact the whole plane, while the third changes the tail and preserves the global structure (e.g., wings and turbines). Both the starting and final airplanes are taken from the test set.*

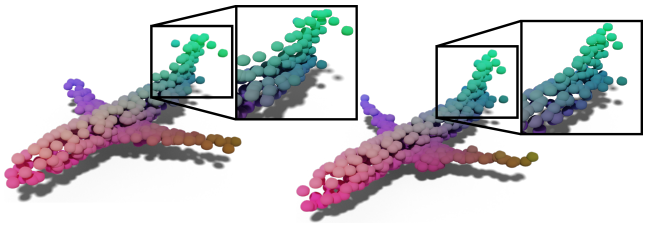

Figure 15: *The first (left) and the last (right) steps for the $d\Lambda_{\mathcal{R}}$ interpolation depicted in Figure 17, with a close-up on the tails.*

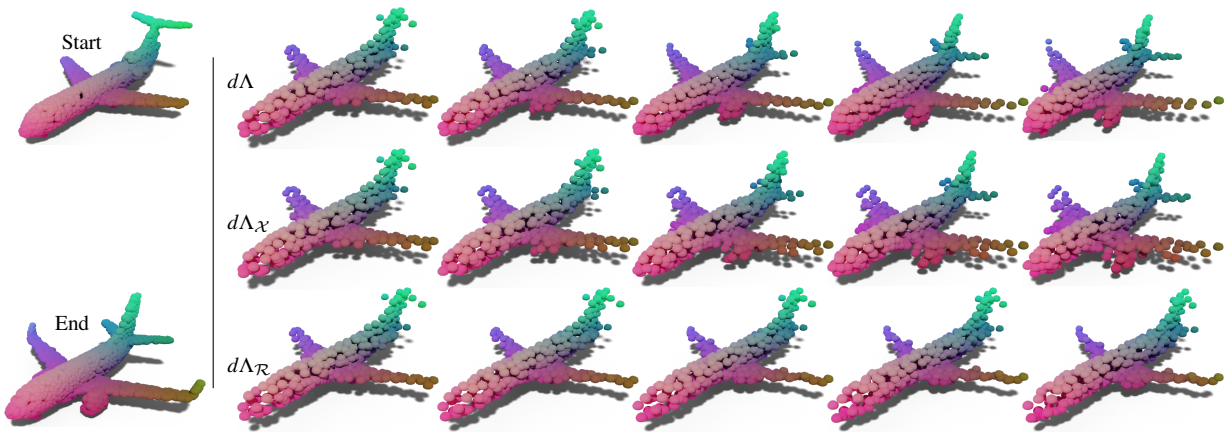

Figure 16: *Casting different kinds of spectra interpolation into our network gives us different degrees of control. On the left, the models used as initial and final steps; on the right, we interpolated the entire spectral encoding ($d\Lambda$-first row), only the global frequencies ($d\Lambda_{\mathcal{X}}$-second row), and only the local ones ($d\Lambda_{\mathcal{R}}$-third row).*

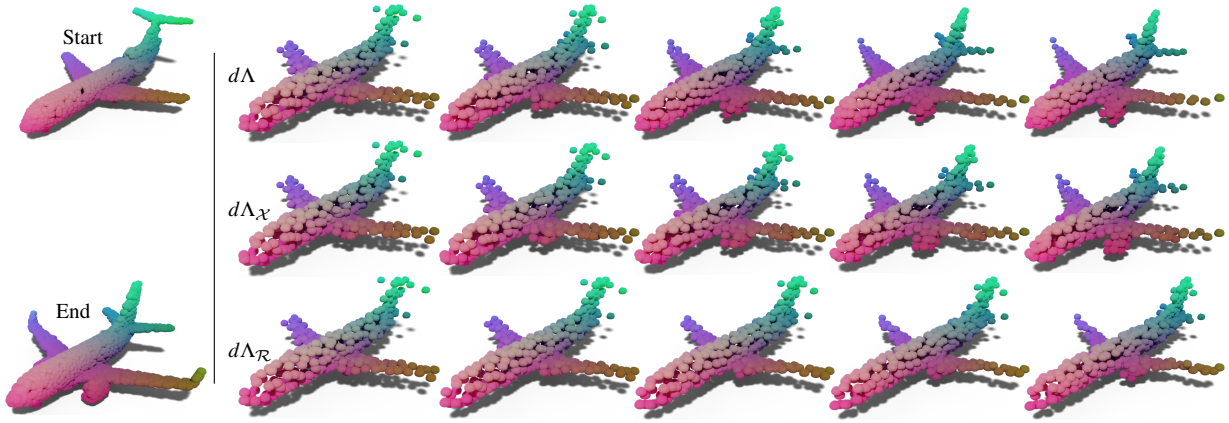

Figure 17: *Casting different kinds of spectra interpolation into our 20+10 network. On the left, the models used as initial and final steps; on the right, we interpolated the entire spectral encoding ($d\Lambda$-first row), only the global frequencies ($d\Lambda_{\mathcal{X}}$-second row), and only the local ones ($d\Lambda_{\mathcal{R}}$-third row).*