LEARNING TO REPRESENT & GENERATE MESHES WITH SPIRAL CONVOLUTIONS

Sergiy Bokhnyak*
Universita Svizzera Italiana
Lugano, Switzerland
bokhns@usi.ch

Giorgos Bouritsas*
Imperial College London
London, UK
g.bouritsas18@imperial.ac.uk

Michael Bronstein
Universita Svizzera Italiana
Imperial College London
Fabula AI
m.bronstein@imperial.ac.uk

Stefanos Zafeiriou
Imperial College London
FaceSoft.io
s.zafeiriou@imperial.ac.uk

ABSTRACT

In this paper, we focus on 3D deformable shapes that share a common topological structure, such as human faces and bodies. Morphable Models were among the first attempts to create compact representations for such shapes; despite their effectiveness and simplicity, such models have limited representation power due to their linear formulation. In this paper, we introduce a mesh autoencoder and a GAN architecture based on the spiral convolutional operator, acting directly on the mesh and leveraging its underlying geometric structure. We provide an analysis of our convolution operator and demonstrate state-of-the-art results on 3D shape datasets compared to the 3D Morphable Model and the recently proposed COMA architecture.

1 INTRODUCTION

In attempting to represent 3D data, the key challenge of geometric deep learning is a meaningful definition of intrinsic operations analogous to convolution and pooling on meshes or point clouds. Among numerous advantages of working directly on mesh data is the fact that it is possible to build invariance to shape transformations (both rigid and nonrigid) into the architecture, as a result allowing to use significantly simpler models and much less training data.

In this paper, we propose a novel representation learning and generative framework for fixed topology 3D shapes. For this purpose, we formulate a spiral convolution operator that acts directly on the mesh, extending the methodology proposed in Lim et al. (2018). In particular, similarly to image convolutions, we construct an intrinsic operator that defines a local neighborhood around each vertex on the mesh, by enforcing an explicit ordering of the neighbors via a spiral scan. This way, we allow for different treatment of each neighbor, yielding anisotropic filters. We use this as a building block for a Spiral Convolutional Autoencoder (SCAE) capable of learning representations and generating 3D shapes from various categories. We evaluate our methods quantitatively on several popular 3D shape datasets and report state-of-the-art reconstruction results, and qualitatively, by showing 'shape arithmetics' in the latent space of the autoencoder. We also present a Spiral Convolutional Wasserstein GAN (SCGAN) that is able to generate novel realistic unseen facial identities.

2 RELATED WORK

Generative models for arbitrary shapes. Perhaps the most common approaches for generating arbitrary shapes are **volumetric CNNs** (Wu et al. (2015)) acting on 3D voxels. In Girdhar et al. (2016), the authors propose an autoencoder on 3D voxel occupancy grids, while Wu et al. (2016)

*Equal Contribution.

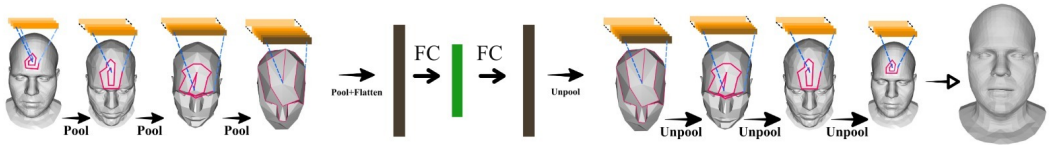

Figure 1: Architecture illustration. Notice the spirals on the mesh.

proposed a voxel-GAN to generate novel shapes. Among the key drawbacks of volumetric methods are their high computational complexity and coarse representation that can alter the topological structure of the shapes. **Point clouds** are a simple and lightweight alternative to volumetric representations recently gaining popularity. Several methods have been proposed for representation learning of fixed-size point clouds (Achlioptas et al. (2018)) and of arbitrary size (Yang et al. (2018)). Despite their compactness, point clouds are not popular for realistic and high-quality 3D geometry generation due to the absence of neighborhood connectivity and lack of an underlying smooth structure. **Morphable models:** In the case of deformable shapes, such as faces and bodies, where a fixed topology can be obtained by establishing dense correspondences, the most popular methods are still statistical models, mainly based on PCA, as the 3D Morphable Model (3DMM) (Blanz et al. (1999)). Popular follow-up works are the the Large Scale Face Model (LSFM) of Booth et al. (2018) for facial identity, (Cao et al. (2014); Li et al. (2017)) for facial expressions, the SMPL model of Loper et al. (2015) for bodies and that of Romero et al. (2017) for hands. **Geometric Deep Learning** is a set of recent methods trying to generalize neural network architectures to non-Euclidean domains such as graphs and manifolds (Bronstein et al. (2017)). Such methods have achieved very promising results in geometry processing and computer graphics (Masci et al. (2015); Boscaini et al. (2016); Monti et al. (2017)), computational chemistry and drug design (Duvenaud et al. (2015); Gilmer et al. (2017)), and network science. Multiple approaches have been proposed to construct convolution-like operations on graphs and meshes, including spectral methods (Bruna et al. (2014); Defferrard et al. (2016b); Kipf & Welling (2017); Yi et al. (2017)) and local charting based (Masci et al. (2015); Boscaini et al. (2016); Monti et al. (2017); Fey et al. (2018); Verma et al. (2018); Lim et al. (2018)).

3 Spiral Convolutional Networks

For the following discussion, we assume to be given a triangular mesh $M = (V, E, F)$ where $V = \{1, \dots, n\}$, E, and F denote the sets of vertices, edges, and faces, respectively. We furthermore assume to be given a function $f : V \to \mathbb{R}$ representing the vertex-wise features.

One of the key challenges in developing convolution-like operators on graphs or manifolds is the lack of a vector space structure and a global system of coordinates that can be associated with each point. The first intrinsic mesh convolutional architectures such as GCNN (Masci et al. (2015)), ACNN (Boscaini et al. (2016)) or MoNet (Monti et al. (2017)) overcame this problem by constructing a *local* system of coordinates around each vertex of the mesh (patch operators).

One difficulty in the construction of patch operators is their *consistency*, i.e. insensitivity to meshing. Ideally, a system of coordinates constructed on a manifold should be independent on the way the manifold is discretised. A second difficulty is the lack of a canonical reference frame; in particular, a patch on a surface can be oriented arbitrarily. A crucial observation we make in this paper is that these issues are irrelevant for generating shapes with *fixed topology*. Assuming that the mesh structure is fixed, constructing a patch operator amounts to *ordering* of the neighbor points,

$$(f \star g)_i = \sum_{\ell=1}^{L} g_\ell f_{i_\ell}. \tag{1}$$

where $i_1, \dots, i_L$ denote the neighbors of vertex i ordered in some fixed way. In the Euclidean setting, this order is simply a raster scan of pixels in a patch. On meshes, we opt for a simple and intuitive ordering using spiral trajectories inspired by Lim et al. (2018).

Let $i \in V$ be some mesh vertex, and let $R^d(i)$ be the *d-ring*, an ordered set of vertices whose shortest (graph) path to i is exactly d hops long; $R^d_j(i)$ denotes the jth element in the d-ring (trivially, $R^0_1(i) = i$). We define the *spiral patch operator* as the ordered sequence

$spiral(i) = (i, R_1^1(i), R_2^1(i), \dots, R^h_{|R^h|}(i))$, where h denotes the number of hops, similarly to the size of the kernel in classical CNNs.

The uniqueness of the ordering is given by fixing two degrees of freedom: the direction of the rings—whether they are ordered clockwise or counterclockwise—and the first vertex $R_1^1(i)$. These degrees of freedom are set by selecting the first three vertices in the spiral; the rest are ordered inductively. In order to allow for fixed-sized spirals, we truncate their length to a fixed length and do zero-padding for the vertices that have smaller length than the chosen one.

We define *spiral convolution* as

$$(f * g)_i = \sum\nolimits_{\ell \in \text{spiral}(i)} g_\ell \, f_\ell. \tag{2}$$

Lim et al. (2018) choose the starting point of each spiral at random, for every mesh sample and every vertex, which makes the ordering inconsistent, and for every epoch during training, which increases the probability of learning rotation invariant filters. In order for the ordering to be consistent it needs to be defined on a local coordinate system that will be repeatable across meshes. For fixed topologies this can be easily obtained since we can retrieve the same ordering by choosing the same vertex index for every mesh. To make our methods more robust, we fixed a reference vertex i_0 on the mesh and chose the initial point for each spiral to be in the direction of the shortest geodesic path to i_0. Moreover, in Lim et al. (2018), the authors model the vertices on the spiral via a recurrent network, which, besides its higher computational complexity, does not take advantage of the stationary properties of the 3D shape (local statistics are repeated across different patches) which is efficiently treated through weight sharing of our spiral kernel.

Spectral convolutional operators developed in Defferrard et al. (2016a); Kipf & Welling (2017) for graphs and used in Ranjan et al. (2018)—COMA—for mesh autoencoders, suffer from the fact that they are inherently *isotropic*. The reason is that on a general graph, there is no canonical ordering of the neighbouring vertices and one has to resort to permutation-invariant operators. While a necessary evil in general graphs, spectral filters on meshes are rather weak: they are locally rotationally-invariant, i.e., have a fixed value in each ring. On the other hand, spiral convolutional filters leverage the fact that on a mesh one can canonically order the neighbours (up to selection of the orientation and reference direction); consequently, our filters are inherently anisotropic and can be expressive with just one hop $h = 1$.

4 Evaluation

We compare the following methods: **PCA**: 3D Morphable Model by Blanz et al. (1999), **COMA**: ChebNet-based Convolutional Mesh Autoencoder (COMA) by Ranjan et al. (2018), **SCAE (small)**: Ours Spiral Convolution Autoencoder, where we used the same architecture as in COMA replacing the spectral filter with the spiral one, **SCAE (ours)**: Our proposed SCAE framework, where we enhanced our model with a larger parameter space, in terms of filter widths. For all the cases, we choose as signal on the mesh the normalized deformations from the mean shape. The **datasets** we evaluate on are COMA from Ranjan et al. (2018) (facial expressions), MeIn3D from Booth et al. (2016) (facial identities), and DFAUST from Bogo et al. (2017) (body poses).

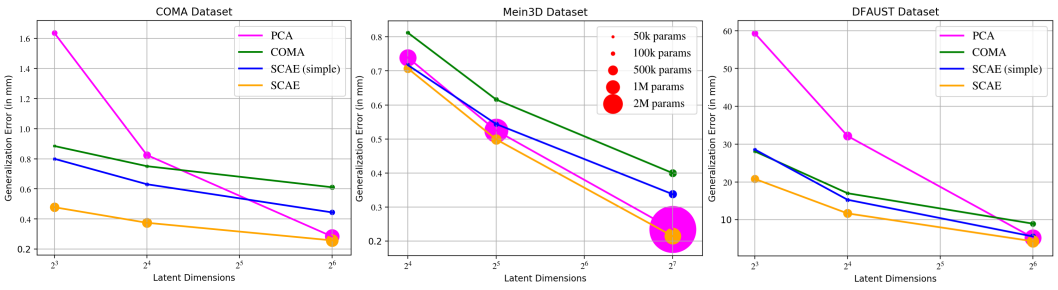

Figure 2: Quantitative evaluation of our SCAE against the baselines, in terms of generalization error

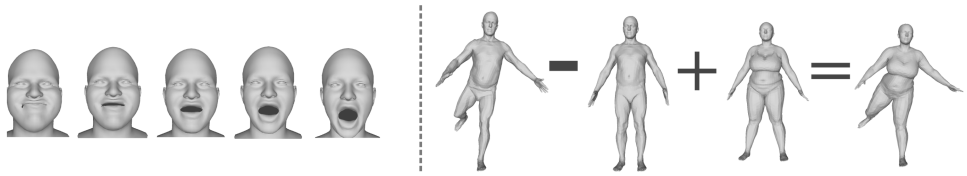

Figure 3: Interpolation on COMA (left) and shape analogy on DFAUST (right)

For the comparison we chose a variety of different dimensionalities of the latent space, based on the variance explained by PCA. The chosen values are variance ratios of roughly 85%, 95%, and 99%. As can be seen from the graphs in Fig 2, our SCAE achieves smaller generalization errors in every case it was tested on. For the COMA and DFAUST datasets all hierarchical intrinsic architectures outperform PCA for small latent sizes. That should probably be attributed to the fact that the localized filters used allow for effective reconstruction of smaller patches of the shape, such as arms and legs (for the DFAUST case), whilst PCA attempts a more global reconstruction, thus its error is distributed equally across the entire shape (please refer to Figure 2 of the supplementary material, where we visually compare all three methods by color coding the per vertex error on exemplar shapes). Regarding the comparison between the hierarchical architectures, it is again apparent here that our spiral based autoencoder has increased capacity, both thanks to the convolutional kernel, which makes our simple SCAE surpass COMA, and thanks to the increased parameter space, which makes our larger SCAE outperform the other methods by a considerably large margin. Despite the fact that for higher dimensions PCA can explain more than 99% of the total variance, SCAE still manages to outperform it, while keeping the parameter size substantially smaller. Especially in the MeIn3D dataset, the large vertex count ($\sim$ 28K vertices) makes PCA impractical.

Moreover, we qualitatively assess (Fig 3) the representational power of our models by: **Interpolations**, where we decode intermediate latent vectors between encodings of two sufficiently different samples; **Analogies**, where we use the decoder to produce a mesh D such that it satisfies $A{:}B{::}C{:}D$, by solving $e(B) - e(A) = e(D) - e(C)$ given meshes A, B, C and encoder e.

4.1 SCGAN EVALUATION

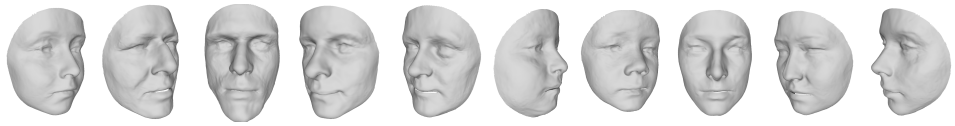

Figure 4: SCGAN synthesized identities

Our SCGAN, follows the SCAE architecture and is trained by the wasserstein divergence loss of Arjovsky et al. (2017) and Gulrajani et al. (2017). In Fig 4, we sampled several faces from the latent distribution of the SCGAN. Notice the diversity in terms of identity – gender, ethnicity, age – as well as the realistically looking details of the facial geometry. Compared to the most popular approach for synthesizing faces, i.e. the 3DMM, our model learns to produce fine details on the facial structure, making them hard to distinguish from real 3D scans, whereas the highly smoothed faces produced by the 3DMM frequently look artificial.

5 CONCLUSION

In this paper we introduced a representation learning and generative framework, based on spiral convolutions, for fixed topology 3D deformable shapes. We show that our mesh autoencoders achieve state-of-the-art results in mesh reconstruction and present the generation capabilities of our models through vector arithmetics in the autoencoder's latent space, as well as by using our GAN to synthesize realistic, novel facial identities. Regarding future work, we plan to extend our framework to generative models for 3D shapes of arbitrary topology, as well as to other domains that have capacity for an implicit ordering of their primitives, such as point clouds.

REFERENCES

Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. *International Conference on Machine Learning (ICML)*, 2018.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

Volker Blanz, Thomas Vetter, et al. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, 1999.

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 (CVPR)*, 2017.

James Booth, Anastasios Roussos, Stefanos Zafeiriou, Allan Ponniah, and David Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5543–5552, 2016.

James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. Large scale 3d morphable models. *International Journal of Computer Vision (IJCV)*, 2018.

Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR)*, 2014.

Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 2014.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems (NIPS)*, 2016a.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NIPS)*, 2016b.

David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing systems (NIPS)*, 2015.

Matthias Fey, Jan Eric Lenssen, Frank Weichert, and Heinrich Müller. Splinecnn: Fast geometric deep learning with continuous b-spline kernels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

Rohit Girdhar, David F Fouhey, Mikel Rodriguez, and Abhinav Gupta. Learning a predictable and generative vector representation for objects. In *European Conference on Computer Vision (ECCV)*. Springer, 2016.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 2017.

Isaak Lim, Alexander Dielen, Marcel Campen, and Leif Kobbelt. A simple approach to intrinsic correspondence learning on unstructured 3d meshes. *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*, 2018.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 37–45, 2015.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3d faces using convolutional mesh autoencoders. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 2017.

Nitika Verma, Edmond Boyer, and Jakob Verbeek. Feastnet: Feature-steered graph convolutions for 3d shape analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 2598–2606, 2018.

Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Li Yi, Hao Su, Xingwen Guo, and Leonidas J Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.