

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Geometric wavelet scattering on graphs and manifolds

Gao, Feng, Hirn, Matthew, Perlmutter, Michael, Wolf, Guy

Feng Gao, Matthew Hirn, Michael Perlmutter, Guy Wolf, "Geometric wavelet scattering on graphs and manifolds," Proc. SPIE 11138, Wavelets and Sparsity XVIII, 111380Q (9 September 2019); doi: 10.1117/12.2529615

SPIE.

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

Geometric wavelet scattering on graphs and manifolds

Feng Gao^{a,b}, Matthew Hirn^{a,c}, Michael Perlmutter^a, and Guy Wolf^d

^aDepartment of Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, Michigan, USA

^bDepartment of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, USA

^cDepartment of Mathematics, Michigan State University, East Lansing, Michigan, USA

^dDepartment of Mathematics and Statistics, Université de Montréal, Montreal, Canada

ABSTRACT

Convolutional neural networks (CNNs) are revolutionizing imaging science for two- and three-dimensional images over Euclidean domains. However, many data sets are intrinsically non-Euclidean and are better modeled through other mathematical structures, such as graphs or manifolds. This state of affairs has led to the development of geometric deep learning, which refers to a body of research that aims to translate the principles of CNNs to these non-Euclidean structures. In the process, various challenges have arisen, including how to define such geometric networks, how to compute and train them efficiently, and what are their mathematical properties.

In this letter we describe the geometric wavelet scattering transform, which is a type of geometric CNN for graphs and manifolds consisting of alternating multiscale geometric wavelet transforms and nonlinear activation functions. As the name suggests, the geometric wavelet scattering transform is an adaptation of the Euclidean wavelet scattering transform, first introduced by S. Mallat, to graph and manifold data. Like its Euclidean counterpart, the geometric wavelet scattering transform has several desirable properties. In the manifold setting these properties include isometric invariance up to a user specified scale and stability to small diffeomorphisms. Numerical results on manifold and graph data sets, including graph and manifold classification tasks as well as others, illustrate the practical utility of the approach.

Keywords: geometric deep learning, scattering transform, convolutional neural network, wavelet, graph, manifold

1. INTRODUCTION

The processing of signal-based data such as auditory recordings, images, and videos for learning tasks is often carried out with convolutional neural networks (CNNs) and related architectures. A primary reason for the success of CNNs is the use of convolution operators, which reduces the number of learned parameters while taking advantage of the underlying Euclidean geometry of the data. However numerous data types of interest, including social networks, molecules in chemistry, two dimensional surfaces as in certain types of medical diagnostic imaging, and computer graphics, amongst others, do not have an underlying Euclidean structure but do have an underlying geometric structure. In many such cases, including those just mentioned, this non-Euclidean structure can be modelled as an abstract graph or Riemannian manifold.

Motivated by the desire to develop learning algorithms for non-Euclidean data and the success of CNNs for Euclidean data, the nascent field of geometric deep learning¹ seeks to develop CNN-type architectures for graph and manifold based data. In order to better understand the theoretical and empirically observable properties of such architectures, several recent papers, including some by the present authors, have generalized the Euclidean

Further author information: (Send correspondence to M.H.)

F.G.: E-mail: gaofeng2@msu.edu

M.H.: E-mail: mhirn@msu.edu, URL: matthewhirn.com

M.P.: E-mail: perlmut6@msu.edu

G.W.: E-mail: guy.wolf@umontreal.ca, URL: guywolf.org

scattering transform to graphs and manifolds, which we generally refer to as geometric scattering transforms.^{2–8} A Euclidean scattering transform⁹ has many similarities to CNNs. It consists of an alternating cascade of linear wavelet convolution operators and nonlinear modulus operations. A geometric scattering transform, analogously, utilizes graph or manifold wavelet transforms and the absolute value nonlinearity in alternating succession to extract invariant information from signals defined on graphs and manifolds.

In this letter we review recent results on geometric scattering transforms obtained by the present authors,^{2,4,8} which include formulations for both graphs and manifolds. Its construction is inspired by learning tasks for graphs and manifolds, which can take many forms but have several commonalities. Most such tasks can be placed into, or are closely related to, one of the following categories: (i) multiple signal classification over a single, fixed graph or manifold; (ii) vertex or point classification in a single graph or manifold; (iii) vertex or point classification across several graphs or manifolds; (iv) multiple graph or manifold classification. In all of these cases, though, one is generally required to process a signal defined on a graph or manifold, even for (ii)–(iv), in which case one uses a signal or signals that can universally defined for any graph or manifold as an initial feature set. Thus geometric deep learning is closely related to graph and manifold signal processing.¹⁰ As such, the construction of the geometric scattering transform is based upon the spectral decomposition of the graph Laplacian for graphs and the Laplace-Beltrami operator for manifolds. These spectral theories facilitate a generalized notion of Fourier series that are described in Section 2, and which are the foundation of signal processing on graphs and manifolds. In Section 3 we define geometric wavelet operators for graphs and manifolds, which are then utilized in Section 4 to define the geometric scattering transform. Section 4 also describes some recently proven mathematical properties of the geometric scattering transform. Finally, Section 5 reviews numerical results obtained in the last year, while concluding remarks are presented in Section 6.

2. FOURIER SERIES ON GRAPHS AND MANIFOLDS

Let $G = (V, E, w)$ be a weighted, connected graph and let $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ denote the normalized graph Laplacian of G , where $\mathbf{A}$ is the adjacency matrix of G and $\mathbf{D}$ is the diagonal degree matrix. Additionally, let $\mathcal{M}$ be a compact, smooth, connected d -dimensional Riemannian manifold contained in $\mathbb{R}^n$ with associated Laplace-Beltrami operator Δ . By X we denote either G or $\mathcal{M}$ with $\mathcal{L}_G = \mathbf{L}$ or $\mathcal{L}_{\mathcal{M}} = -\Delta$. In either case, the Laplace operator $\mathcal{L}_X$ is positive semi-definite with a countable (finite, in the case of $X = G$) number of eigenvalues and associated real-valued eigenvectors. Let $\{\lambda_k\}_k$ denote the ordered, non-unique eigenvalues with associated orthonormal eigenvectors $\{\varphi_k\}_k$, i.e.,

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \quad \text{and} \quad \mathcal{L}_X \varphi_k = \lambda_k \varphi_k.$$

We remark that in the case of graphs the normalized graph Laplacian has the additional property that $\lambda_k \leq 2$.

We interpret the eigenvectors of $\mathcal{L}_X$ as the Fourier modes of X , and use them to define Fourier series on X . Let $f \in \mathbf{L}^2(X)$ and define the Fourier series $\hat{f} \in \ell^2$ of f as

$$\hat{f}(k) := \langle f, \varphi_k \rangle = \int_X f(x) \varphi_k(x) d\mu_X(x), \quad k \geq 0,$$

where $\mu_G = \sum_{x \in V} \delta_x$ is the uniform discrete measure and $\mu_{\mathcal{M}}$ is the Riemannian volume. The Fourier series of a signal $f \in \mathbf{L}^2(X)$ decomposes f in terms of the harmonics of X ; Figure 1 visualizes these harmonics for graph and manifold examples. Since $\{\varphi_k\}_k$ is an orthonormal basis for $\mathbf{L}^2(X)$, the following Fourier inversion formula follows immediately:

$$f = \sum_k \hat{f}(k) \varphi_k. \quad (1)$$

Motivated by the Fourier convolution theorem for $\mathbf{L}^2(\mathbb{R}^n)$, we define the convolution operation $*$ between $f, h \in \mathbf{L}^2(X)$ using (1),

$$f * h := \sum_k \hat{f}(k) \hat{h}(k) \varphi_k, \quad (2)$$

which essentially by definition satisfies $\widehat{f * h}(k) = \hat{f}(k) \hat{h}(k)$. We will use this definition of convolution to help define wavelet operators in the next section.

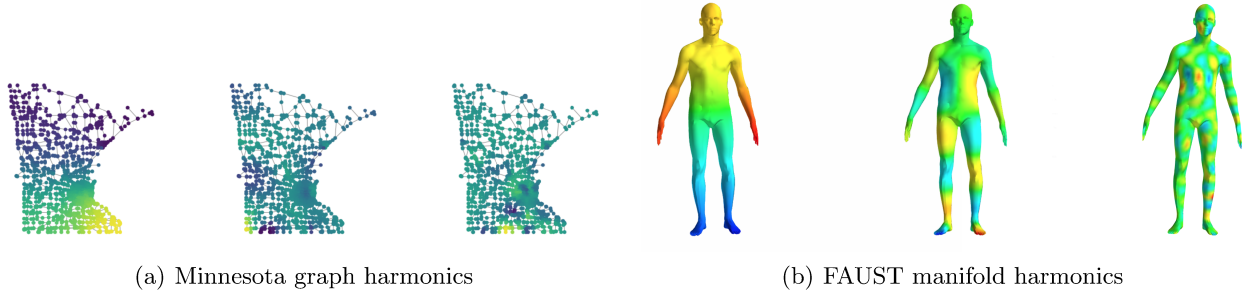


Figure 1. Harmonics corresponding to low, medium, and high frequencies of (a) the Minnesota graph and (b) a manifold from the FAUST dataset.¹¹

3. GEOMETRIC WAVELETS

Graph and manifold filters h may be designed spectrally by specifying their Fourier series $\widehat{h}(k)$ for each $k \geq 0$. Using the definition of $\widehat{f}(k)$ and the convolution definition (2) yields

$$f * h(x) = \int_X \left(\sum_k \widehat{h}(k) \varphi_k(x) \varphi_k(y) \right) f(y) d\mu_X(y),$$

which has the form of an integral operator with kernel $\kappa(x, y) = \sum_k \widehat{h}(k) \varphi_k(x) \varphi_k(y)$. In fact one can use more general kernels to define convolution type operators on X .

We shall be interested in kernels that yield integral operators corresponding to low pass and high pass filter operations on a signal $f \in \mathbf{L}^2(X)$. We use diffusion kernels to specify the low pass kernels. For manifolds we utilize the heat kernel with diffusion time t ,

$$p_{\mathcal{M}}(t, x, y) := \sum_k e^{-t\lambda_k} \varphi_k(x) \varphi_k(y).$$

When $X = G$ is a graph we define $\alpha_k := \mathbf{D}^{1/2} \varphi_k$ and $\beta_k := \mathbf{D}^{-1/2} \varphi_k$ and utilize the following kernel

$$p_G(t, x, y) := \sum_k (1 - \lambda_k/2)^t \alpha_k(x) \beta_k(y),$$

which one can verify is the kernel of the t -step lazy random walk $\mathbf{P}^t$ where $\mathbf{P} := \frac{1}{2}(\mathbf{I} + \mathbf{A}\mathbf{D}^{-1})$. The low pass operator $A_{X,J} : \mathbf{L}^2(X) \rightarrow \mathbf{L}^2(X)$ at the scale 2^J is defined as the integral operator with kernel $p_X(2^J, x, y)$, i.e.,

$$A_{X,J} f(x) := \int_X p_X(2^J, x, y) f(y) d\mu_X(y).$$

Wavelet kernels are obtained by taking the difference of diffusion kernels at consecutive dyadic scales as in diffusion wavelets.¹² On a manifold $X = \mathcal{M}$ the wavelet kernel at the scale 2^j is defined as^{2,8}

$$\psi_{\mathcal{M}}(t, x, y) := \sum_k [e^{-t\lambda_k} - e^{-2t\lambda_k}]^{1/2} \varphi_k(x) \varphi_k(y),$$

whereas on graphs we utilize^{3,4}

$$\psi_G(t, x, y) := p_G(t/2, x, y) - p_G(t, x, y).$$

The corresponding wavelet high pass operator $\Psi_j : \mathbf{L}^2(X) \rightarrow \mathbf{L}^2(X)$ at the scale 2^j is defined as

$$\Psi_{X,j} f(x) := \int_X \psi_X(2^j, x, y) f(y) d\mu_X(y).$$

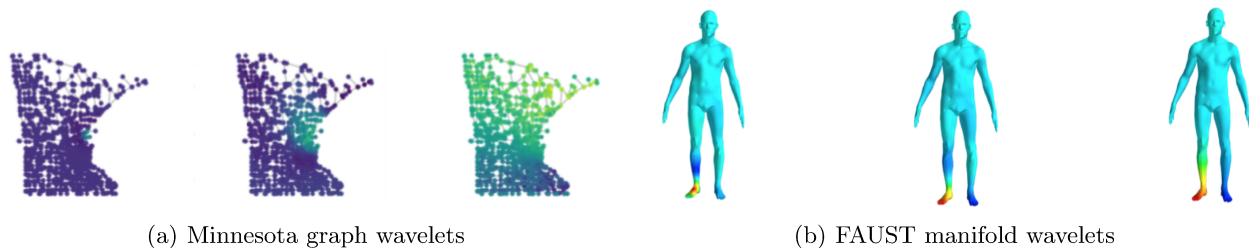


Figure 2. Geometric wavelets in order of increasing scale for (a) the Minnesota graph and (b) a manifold from the FAUST dataset.¹¹ Figure 2(a) adapted from [4, Figure 1(a)] and Figure 2(b) adapted from [8, Figure 1].

Finally, the geometric wavelet transform $W_{X,J} : \mathbf{L}^2(X) \rightarrow \ell^2(\mathbf{L}^2(X))$ of a signal $f \in \mathbf{L}^2(X)$ is given by

$$W_{X,J}f := \{A_{X,J}f, \Psi_{X,j}f : j \in I_{X,J}\},$$

where the index sets $I_{X,J}$ are given by

$$I_{G,J} := \{j \in \mathbb{Z} : 1 \leq j \leq J\} \quad \text{and} \quad I_{\mathcal{M},J} := \{j \in \mathbb{Z} : j \leq J\}. \quad (3)$$

The geometric wavelet transform $W_{X,J}f$ averages the signal f at the scale 2^J via $A_{X,J}f$, which captures the low frequencies of f (as measured by the eigenvalues of $\mathcal{L}_X$). The high frequencies of f are recovered with $\Psi_{X,j}f$ by testing the signal f against the difference of diffusions at consecutive dyadic scales, which are made progressively smaller to capture higher and higher frequencies. In the case of graphs, which are discrete objects with a finite number of vertices, the number of wavelets is finite and the smallest wavelet is defined as the difference between the 1-step diffusion and the 2-step diffusion. Alternatively, on manifolds the diffusion time can be made infinitesimally small which leads to a countably infinite number of wavelets, which is reflected in the definitions of the index sets in (3). Images of geometric wavelets are given in Figure 2.

4. GEOMETRIC SCATTERING

The construction of the geometric wavelet scattering transform is motivated by the desire to have a representation of $f \in \mathbf{L}^2(X)$ that is (locally) invariant to graph/manifold isometries while at the same time preserving as much information in f as possible. It consists of an alternating cascade of geometric wavelet transforms and absolute value nonlinearities. After each wavelet transform, the low frequency component of the previous layer is extracted, the collection of which forms the geometric wavelet scattering transform. Remaining high frequency information is passed on to the next layer, part of which is mapped to the low frequencies via the absolute value nonlinearity. Formally it is defined as:

$$S_{X,J}f := \{A_{X,J}f, A_{X,J}|\Psi_{X,j_m}|\Psi_{X,j_{m-1}} \cdots |\Psi_{X,j_1}f|| : m \geq 1, j_l \in I_{X,J}\}.$$

Figure 3 illustrates the geometric wavelet transform as an alternating sequence of geometric wavelet transforms and absolute values.

By retaining only the low frequency part of the functions at each layer, the geometric wavelet scattering transform encodes a locally isometric invariant description of the signal f . For manifolds, it is proven in [8, Theorem 2] that if one computes $M < \infty$ layers of the geometric wavelet scattering transform, then

$$\|S_{X,J}f - S_{X,J}V_{\zeta}f\| \leq C(\mathcal{M})(M+1)^{1/2}2^{-dJ}\|\zeta\|_{\infty}\|f\|_2, \quad \forall f \in \mathbf{L}^2(\mathcal{M}), \zeta \in \text{Isom}(\mathcal{M}),$$

where $\text{Isom}(\mathcal{M})$ is the isometry group of $\mathcal{M}$, $V_{\zeta}f(x) := f(\zeta^{-1}(x))$, and $\|\zeta\|_{\infty} := \sup_{x \in \mathcal{M}} r(x, \zeta(x))$ with $r(x, x')$ being the geodesic distance between $x, x' \in \mathcal{M}$. An analogous result for graphs is given in [5, Theorem 4.2]. On manifolds, [8, Theorem 3] proves additionally the geometric scattering transform is stable to the action of diffeomorphisms on bandlimited signals $f \in \mathbf{L}^2(\mathcal{M})$. The stability of the transform with respect to graph perturbations is also of interest and has been studied in several recent works.^{3,5,7} For brevity we omit the precise statements of these results in this letter, and instead refer the reader to the original papers.

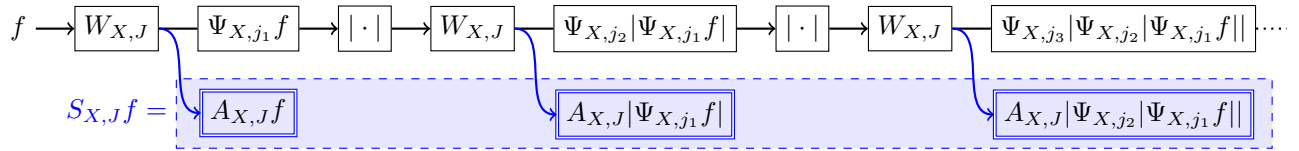


Figure 3. The geometric wavelet scattering transform as an alternating cascade of geometric wavelet transforms and absolute value nonlinearities. At the $(m+1)^{\text{st}}$ layer the low frequencies of the functions from the m^{th} layer are extracted (blue), while the remaining high frequency information is passed on to the next layer (black). Figure adapted from [8, Figure 2].

In practice one must decide, based on prior knowledge or cross-validation, the scale of isometric invariance, which is indexed by J . In the case of manifold classification, shape retrieval, and graph classification, each data point is new a graph or manifold with an associated label and one would like to learn a model for the labeling function. In most cases this labeling function is invariant to isometries, meaning that, for example, the extrinsic coordinates of the manifold do not matter or the particular labeling of the vertices of a graph does not affect its label. One must therefore make global comparisons between graphs or manifolds, modulo isometric transformations. In this case we take $J \rightarrow \infty$, which leads to the simplified expression

$$\lim_{J \rightarrow \infty} A_J |\Psi_{X,j_m} |\Psi_{X,j_{m-1}} \cdots |\Psi_{X,j_1} f(x)||| \propto \|\Psi_{X,j_m} |\Psi_{X,j_{m-1}} \cdots |\Psi_{X,j_1} f|\|_1, \quad \forall x \in X. \quad (4)$$

5. EMPIRICAL PERFORMANCE IN LEARNING TASKS

We consider both graph and manifold supervised learning tasks. Section 5.1 describes results on graph classification first reported in [4, Section 4.1], while Section 5.2 describes the manifold classification results contained in [8, Section 4.2]. Finally, Section 5.3 considers signal classification on the 2-sphere $\mathbb{S}^2$ using the spherical MNIST dataset; these results are originally from [8, Section 4.1].

5.1 Graph Classification

In [4, Section 4.1] we applied the geometric wavelet scattering transform to graph classification of social network data,¹³ which included six social network data sets extracted from scientific collaborations (COLLAB), movie collaborations (IMDB-B & IMDB-M), and Reddit discussion threads (REDDIT-B, REDDIT-5K, REDDIT-12K). In each data set, every graph is given a label and the goal is to predict the label of unseen graphs.

Classification with the geometric wavelet scattering representation is carried out by letting $J \rightarrow \infty$ and using (4) along with other p -norms. For the input signal f we used both the eccentricity and clustering coefficient of each vertex. The resulting “scattering feature vector” is passed into a support vector machine (SVM) with Gaussian (RBF) kernel; we call the resulting classifier GS-SVM. Using ten-fold cross validation we compare the GS-SVM classifier to ten recent graph classification methods, the results of which are given in Table 1. The comparison methods include four graph kernel approaches, Weisfeiler-Lehman graph kernels [14, WL], Graphlet kernels [15, Graphlet], deep graph kernels [13, DGK], and Weisfeiler-Lehman optimal assignment kernels [16, WL-OA], in addition to six geometric deep learning algorithms: deep graph convolutional neural network [17, DGCNN], 2D convolutional neural networks [18, 2DCNN], Patchy-san [19, PSCN], graph capsule convolutional neural networks [20, GCAPS-CNN], recurrent neural network autoencoders [21, S2S-N2N-PP], and the graph isomorphism network [22, GIN]. Additional details are given in [4, Section 4.1].

5.2 Manifold Classification

Manifold classification is analogous to graph classification. In [8, Section 4.2] we considered manifold classification using the FAUST database,¹¹ which consists of 100 manifold meshes spanning ten people in ten poses. Using this data one can consider person classification or pose classification. In either case one lets $J \rightarrow \infty$ and uses (4); unlike in the graph case, we restrict our scattering feature vector to contain only the $p = 1$ norms as in (4). This initial signals f consisted of 352 SHOT features.²³ Table 2 reports the results, which in both classification tasks show an improvement over using the SHOT features alone. Classification results are reported for five-fold cross validation using an SVM classifier with RBF kernel; all hyper-parameters, including the number of layers in the wavelet scattering transform, were estimated using nested cross-validation.

Table 1. Comparison of the proposed GS-SVM classifier with leading graph kernel and deep learning methods on social network graph datasets. Results reported as average classification accuracy plus/minus the standard deviation over ten folds. Table and results taken from [4, Table 1].

	COLLAB	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	REDDIT-12K	
WL	77.82 $\pm$ 1.45	71.60 $\pm$ 5.16	N/A	78.52 $\pm$ 2.01	50.77 $\pm$ 2.02	34.57 $\pm$ 1.32	Graph kernel
Graphlet	73.42 $\pm$ 2.43	65.40 $\pm$ 5.95	N/A	77.26 $\pm$ 2.34	39.75 $\pm$ 1.36	25.98 $\pm$ 1.29	
WL-OA	80.70 $\pm$ 0.10	N/A	N/A	89.30 $\pm$ 0.30	N/A	N/A	
DGK	73.00 $\pm$ 0.20	66.90 $\pm$ 0.50	44.50 $\pm$ 0.50	78.00 $\pm$ 0.30	41.20 $\pm$ 0.10	32.20 $\pm$ 0.10	
DGCNN	73.76 $\pm$ 0.49	70.03 $\pm$ 0.86	47.83 $\pm$ 0.85	N/A	48.70 $\pm$ 4.54	N/A	Deep learning
2D CNN	71.33 $\pm$ 1.96	70.40 $\pm$ 3.85	N/A	89.12 $\pm$ 1.70	52.21 $\pm$ 2.44	48.13 $\pm$ 1.47	
PSCN ($k = 10$)	72.60 $\pm$ 2.15	71.00 $\pm$ 2.29	45.23 $\pm$ 2.84	86.30 $\pm$ 1.58	49.10 $\pm$ 0.70	41.32 $\pm$ 0.42	
GCAPS-CNN	77.71 $\pm$ 2.51	71.69 $\pm$ 3.40	48.50 $\pm$ 4.10	87.61 $\pm$ 2.51	50.10 $\pm$ 1.72	N/A	
S2S-P2P-NN	81.75 $\pm$ 0.80	73.80 $\pm$ 0.70	51.19 $\pm$ 0.50	86.50 $\pm$ 0.80	52.28 $\pm$ 0.50	42.47 $\pm$ 0.10	
GIN-0 (MLP-SUM)	80.20 $\pm$ 1.90	75.10 $\pm$ 5.10	52.30 $\pm$ 2.80	92.40 $\pm$ 2.50	57.50 $\pm$ 1.50	N/A	
GS-SVM	79.94 $\pm$ 1.61	71.20 $\pm$ 3.25	48.73 $\pm$ 2.32	89.65 $\pm$ 1.94	53.33 $\pm$ 1.37	45.23 $\pm$ 1.25	

Table 2. Manifold classification accuracy on the FAUST dataset for pose classification and person classification. “SHOT only” refers to integrals of SHOT features.²³ Table and results taken from [8, Table 1].

Task/Model	SHOT only	Geometric scattering
Pose classification	0.92	0.95
Person classification	0.61	0.81

5.3 Signal Classification on a Fixed Non-Euclidean Manifold

Finally we consider classification of the spherical MNIST database, which consists of the the handwritten MNIST digits projected onto the 2-sphere S^2 . Using $J = -2$, two geometric wavelet scattering layers, and RBF kernel SVM, [8, Section 4.1] reported competitive results with fully learned geometric deep networks designed specifically for the sphere. A subset of these results are displayed in Table 3.

6. CONCLUSION

In this letter we described the geometric wavelet scattering transform, which we use as a general name for the generalization of the wavelet scattering transform to graphs and manifolds.²⁻⁸ The geometric wavelet scattering transform is a type of geometric convolutional network in which the standard notion of convolution is replaced with graph and manifold analogues, and standard wavelet filters are replaced with geometric versions. Here we considered diffusion wavelets but more general graph and manifold wavelets may be used. In addition to defining the geometric wavelet scattering transform, we described its local isometric invariance property and recalled some of its recent empirical results, but numerous other recent developments, notably manifold and graph stability properties, were omitted. As the field of geometric deep learning continues to grow in popularity, mathematical understanding of graph and manifold neural networks will be a key component to advancing the field; we expect the geometric scattering transform to continue to play a role in this effort.

ACKNOWLEDGMENTS

This research was partially funded by: grant P42 ES004911, through the National Institute of Environmental Health Sciences of the NIH, supporting *F.G.*; IVADO (l’institut de valorisation des données) [*G.W.*]; the Alfred P. Sloan Fellowship (grant FG-2016-6607), the DARPA Young Faculty Award (grant D16AP00117), NSF grant 1620216 and NSF CAREER award 1845856 [*M.H.*].

Table 3. Spherical MNIST classification accuracy. Table and results taken from [8, Table 3b].

Method	Accuracy
S2CNN ²⁴	0.96
FFS2CNN ²⁵	0.96
Method from ²⁶	0.99
Harr wavelet scattering ²⁷	0.90
Geometric wavelet scattering	0.95

REFERENCES

- [1] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P., “Geometric deep learning: Going beyond Euclidean data,” *IEEE Signal Processing Magazine* **34**(4), 18–42 (2017).
- [2] Perlmutter, M., Wolf, G., and Hirn, M., “Geometric scattering on manifolds,” in [*NeurIPS Workshop on Integration of Deep Learning Theories*], (2018). arXiv:1812.06968.
- [3] Gama, F., Ribeiro, A., and Bruna, J., “Diffusion scattering transforms on graphs,” in [*International Conference on Learning Representations*], (2019).
- [4] Gao, F., Wolf, G., and Hirn, M., “Geometric scattering for graph data analysis,” in [*Proceedings of the 36th International Conference on Machine Learning, PMLR*], **97**, 2122–2131 (2019).
- [5] Zou, D. and Lerman, G., “Graph convolutional neural networks via scattering,” *Applied and Computational Harmonic Analysis* (2019). In press.
- [6] Zou, D. and Lerman, G., “Encoding robust representation for graph generation,” arXiv:1809.10851 (2019).
- [7] Gama, F., Bruna, J., and Ribeiro, A., “Stability of graph scattering transforms,” arXiv:1906.04784 (2019).
- [8] Perlmutter, M., Gao, F., Wolf, G., and Hirn, M., “Geometric scattering networks on compact Riemannian manifolds,” arXiv:1905.10448 (2019).
- [9] Mallat, S., “Group invariant scattering,” *Communications on Pure and Applied Mathematics* **65**, 1331–1398 (October 2012).
- [10] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P., “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine* **30**(3), 83–98 (2013).
- [11] Bogo, F., Romero, J., Loper, M., and Black, M. J., “FAUST: Dataset and evaluation for 3D mesh registration,” in [*Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*], (2014).
- [12] Coifman, R. R. and Maggioni, M., “Diffusion wavelets,” *Applied and Computational Harmonic Analysis* **21**(1), 53–94 (2006).
- [13] Yanardag, P. and Vishwanathan, S., “Deep graph kernels,” in [*Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*], 1365–1374, ACM (2015).
- [14] Shervashidze, N., Schweitzer, P., Leeuwen, E. J. v., Mehlhorn, K., and Borgwardt, K. M., “Weisfeiler-Lehman graph kernels,” *Journal of Machine Learning Research* **12**(Sep), 2539–2561 (2011).
- [15] Shervashidze, N., Vishwanathan, S., Petri, T., Mehlhorn, K., and Borgwardt, K., “Efficient graphlet kernels for large graph comparison,” in [*Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*], van Dyk, D. and Welling, M., eds., *Proceedings of Machine Learning Research* **5**, 488–495, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA (2009).
- [16] Kriege, N. M., Giscard, P.-L., and Wilson, R., “On valid optimal assignment kernels and applications to graph classification,” in [*Advances in Neural Information Processing Systems 29*], Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., eds., 1623–1631, Curran Associates, Inc. (2016).
- [17] Zhang, M., Cui, Z., Neumann, M., and Chen, Y., “An end-to-end deep learning architecture for graph classification,” in [*AAAI Conference on Artificial Intelligence*], 4438–4445 (2018).
- [18] Tixier, A. J.-P., Nikolentzos, G., Meladianos, P., and Vazirgiannis, M., “Classifying graphs as images with convolutional neural networks,” *arXiv preprint*, arXiv:1708.02218 (2017).
- [19] Niepert, M., Ahmed, M., and Kutzkov, K., “Learning convolutional neural networks for graphs,” in [*International conference on machine learning*], 2014–2023 (2016).
- [20] Verma, S. and Zhang, Z.-L., “Graph capsule convolutional neural networks,” in [*Joint ICML and IJCAI Workshop on Computational Biology*], (2018). arXiv:1805.08090.
- [21] Taheri, A., Gimpel, K., and Berger-Wolf, T., “Learning graph representations with recurrent neural network autoencoders,” in [*KDD Deep Learning Day*], (2018).
- [22] Xu, K., Hu, W., Leskovec, J., and Jegelka, S., “How powerful are graph neural networks?,” in [*International Conference on Learning Representations*], (2019).
- [23] Tombari, F., Salti, S., and Di Stefano, L., “Unique signatures of histograms for local surface description,” in [*European conference on computer vision*], 356–369 (2010).
- [24] Cohen, T. S., Geiger, M., Koehler, J., and Welling, M., “Spherical cnns,” in [*Proceedings of the 6th International Conference on Learning Representations*], (2018).

- [25] Kondor, R., Lin, Z., and Trivedi, S., “Clebsch-Gordan nets: a fully Fourier space spherical convolutional neural network,” in [*Advances in Neural Information Processing Systems 31*], 10117–10126 (2018).
- [26] Jiang, C. M., Huang, J., Kashinath, K., Prabhat, Marcus, P., and Niessner, M., “Spherical CNNs on unstructured grids,” in [*International Conference on Learning Representations*], (2019).
- [27] Chen, X., Cheng, X., and Mallat, S., “Unsupervised deep Haar scattering on graphs,” in [*Conference on Neural Information Processing Systems 27*], 1709–1717 (2014).