VolterraNet: A higher order convolutional network with group equivariance for homogeneous manifolds

Monami Banerjee†, Rudrasis Chakraborty†, Jose Bouza, and Baba C. Vemuri, *Fellow, IEEE*

Abstract—Convolutional neural networks have been highly successful in image-based learning tasks due to their translation equivariance property. Recent work has generalized the traditional convolutional layer of a convolutional neural network to non-Euclidean spaces and shown group equivariance of the generalized convolution operation. In this paper, we present a novel higher order Volterra convolutional neural network (VolterraNet) for data defined as samples of functions on Riemannian homogeneous spaces. Analagous to the result for traditional convolutions, we prove that the Volterra functional convolutions are equivariant to the action of the isometry group admitted by the Riemannian homogeneous spaces, and under some restrictions, any non-linear equivariant function can be expressed as our homogeneous space Volterra convolution, generalizing the non-linear shift equivariant characterization of Volterra expansions in Euclidean space. We also prove that second order functional convolution operations can be represented as cascaded convolutions which leads to an efficient implementation. Beyond this, we also propose a dilated VolterraNet model. These advances lead to large parameter reductions relative to baseline non-Euclidean CNNs.
To demonstrate the efficacy of the VolterraNet performance, we present several real data experiments involving classification tasks on spherical-MNIST, atomic energy, Shrec17 data sets, and group testing on diffusion MRI data. Performance comparisons to the state-of-the-art are also presented.

Index Terms—Homogeneous spaces, Volterra Series, Convolutions, Geometric Deep Learning, Equivariance

1 Introduction

CNNs were introduced in the 1990s by Lecun [1] and gained enormous popularity in the past decade especially after the demonstration of the significant success on Imagenet data by Krizhevsky et al. [2]. At the heart of CNNs success is its ability to learn a rich class of features from data using a combination of convolutions and nonlinear operations such as ReLU or softmax functions. The success of CNNs however is achieved at the expense of a large number of parameters that need to be learned and a computational burden in the training time. It is well known now that a multi-layer perceptron can approximate any function to the desired level of accuracy with a finite number of neurons in the hidden layer. It is therefore natural to consider parameter efficiency as one of the network design goals to strive for in a deep netowrk. The higher order Volterra series can capture a richer class of features and hence significantly reduce the total number of parameters while maintaining comparable or better classification accuracy relative to the baseline models.

In computer vision and medical imaging, many applications deal with data domains that are non-Euclidean. For instance, the n-sphere ($n \geq 2$), the manifold of symmetric positive definite matrices, the Grassmannian, Stiefel manifold, flag manifolds etc. Most of these manifolds belong to the class of (Riemannian) homogeneous spaces (manifolds). Thus, our goals here are to 1) Introduce a principled framework for defining CNNs on general homogeneous Riemannian manifolds. 2) Introduce a novel higher order convolution layer using Volterra theory [3] on homogeneous Riemannian manifolds which provides significant parameter-efficiency improvements for non-Euclidean CNNs. 3) Establish empirical evidence demonstrating the applicability of our homogeneous Riemannian manifold CNNs and the performance boost provided by the Volterra convolutions.

Much of the recent work in this problem domain has focused on generalizing CNNs to homogeneous spaces by exploiting the weight sharing that the symmetries of the underlying manifold allow. The 2-sphere is a particularly important example. In the recent past, CNNs have been reported in literature [4], [5], [6] which are designed to handle data that are samples of functions defined on a 2-sphere and hence are equivariant to 3D rotations which are members of the $\mathsf{SO}(3)$ group. The spherical convolution[1] network presented in [5], [7] is named Spherical CNN. Recently, Kondor et al. in [8] proposed the Clebsch-Gordan net by replacing the repeated forward and backward Fourier transform operations used in [5]. They showed that by using the Clebsch-Gordan transform as the source of nonlinearity, better performance can be achieved by avoiding the repetitive forward and inverse Fourier transform operations. In [9], authors present polar transformer networks, which are equivariant to rotations and scaling transformations.

- *M. Banerjee is with FaceBook, CA, R. Chakraborty is with University of California, Berkeley, USA. † denotes equal contributions.*
- *J. Bouza and B. C. Vemuri are with University of Florida, Gainesville, FL, USA.*
  E-mail: {monamie.b, rudrasischa, josejbouza}@gmail.com, vemuri@ufl.edu

1. As has been pointed out several times in the literature, the convolution operation in CNNs is actually a correlation and not a convolution. Hence, in this paper, *we will use the term convolution and correlation interchangeably but always imply correlation.*

By combining them with the spatial transformer [10], they achieved the required equivariance to translations as well. Recently, the equivariance of convolutions to more general classes of group actions has been reported in literature [11], and later in [12]. In [7], Esteves et al. used the correlation defined in [13] to propose an $\mathsf{SO}(3)$ equivariant operator and in turn define a spherical convolution. In this paper, we will define correlations and the Volterra series on a homogeneous manifold and show that the equivariance property holds for both.

Volterra kernels were first proposed in image classification literature in [14], [15]. In [15], authors learn the kernels in a data driven fashion and formulate the learning problem as a generalized eigenvalue problem. Volterra theory of nonlinear systems was applied more than two decades ago to a single hidden layer feed-forward neural network with a linear output layer and a fully dynamic recurrent network in [16]. Most recent use of Volterra kernels in deep networks was reported in [17], where, authors presented a single layer of Volterra kernel based convolutions followed by conventional CNN layers. They however did not explore equivariance properties of the network or consider non-Euclidean input domains.

In this paper, we define a Volterra kernel to replace traditional convolution kernels. We present a novel generalization of the convolution group-equivariance property to higher order convolutions expressed using Volterra theory of functional convolution on non-Euclidean domains, specifically, the Riemannian homogenous spaces [18] referred to earlier. Most of these manifolds are commonly encountered in mathematical formulations of various computer vision tasks such as action recognition, covariance tracking etc., and in medical imaging for example, in diffusion magnetic resonance imaging (dMRI), elastography etc. By generalizing traditional CNNs in two possible ways, 1) to cope with data domains that are non-Euclidean and 2) to higher order convolutions expressed using Volterra series, we expect to extend the success of CNNs in yet unexplored ways.

We begin with a significant extension of prior work by the authors of [19], where the authors defined a correlation operation for homogeneous manifolds. Specifically, our extension consists of a proof that not only is the correlation operation group equivariant, but additionally any linear group equivariant function can be written as a correlation on the manifold (Banerjee et al. [19] only showed the first fact). We present experiments to demonstrate better performance of the proposed VolterraNet on spherical-MINST and the Shrec17 data with less number of parameters than previously shown in literature for the Spherical-CNN and the Clebsch-Gordan net. We then present a dilated convolution model based on the VolterraNet and demonstrate its efficacy in group testing on diffusion magnetic resonance data acquired from patients with movement disorders. The domain of this data is another example of a Riemannian homogeneous space.

In summary, our key contributions in this paper are: 1) A principled method for choice of basis in designing a deep network architecture on a Riemannian homogeneous manifold $\mathcal{M}$. 2) A proof of a generalization of the classical linear shift invariance (in our terminology, equivariance) characterization theorem for correlation operations on Riemannian homogeneous manifolds. 3) A novel generalization of convolution operations to higher order Volterra series on non-Euclidean domains specifically, Riemannian homogeneous manifolds which are often encountered both in computer vision and medical imaging applications. 4) A generalization of the classical non-linear shift invariance (in our terminology, equivariance) characterization theorem for Volterra convolution operations on Riemannian homogeneous manifolds. 5) Experiments on real data sets that are publicly available such as the spherical-MNIST, atomic energy and Shrec17. For these real data, we present comparisons to the state-of-the-art methods. 6) An extension of the VolterraNet to Dilated VolterraNet and demonstrate its efficiency via group testing on diffusion MRI brain scans from controls (normal subjects) and movement disorder patients. Further, ablation studies on VolterraNet to demonstrate the usefulness of the higher order convolution operations.

The rest of the paper is organized as follows: In section 3.3 we define the correlation operation on homogenous manifolds and prove a generalization of the Euclidean linear shift invariance (LSI) theorem for this correlation operation. Then, in section 3.4 we present a framework for principled choice of basis in representing functions on a Riemmanian homogeneous manifold. In section 4.1, we define the Volterra higher-order convolution operation and prove a generalization of the non-linear shift invariance theorem for this Volterra operation. Following this, we present a detailed description of the proposed VolterraNet architecture in 5 and a description of the proposed dilated VolterraNet in 6. Finally, section 7 contains the experimental results and section 8 the conclusions.

2 List of Notations

We now summarize the list of notations that will be used throughout this paper.

$\mathcal{M}$	Riemannian homogeneous space (manifold)
G	a group
$\mathsf{SO}(n)$	n-dimensional special orthogonal group of matrices
$I(\mathcal{M})$	Isometry group admitted by $\mathcal{M}$
$L^2(\mathcal{M}, \mathbf{R})$	Space of real-valued square integrable functions on $\mathcal{M}$
$L^2(G, \mathbf{R})$	Space of real-valued square integrable functions on G
$\omega^{\mathcal{M}}$	Volume form of $\mathcal{M}$
μ_G	Haar measure of G
$g \cdot x / L_g(x)$	Action of $g \in G$ on $x \in \mathcal{M}$
$gh / L_g(h)$	Action of $g \in G$ on $h \in G$
$g \cdot f / L^*_{g^{-1}}(f)$	Action of $g \in G$ on $f : \mathcal{M} \to \mathbf{R}$ and is given by $x \mapsto f(g^{-1} \cdot x)$
$\mathbf{S}^n$	n-sphere
$\mathbf{R}^+$	Space of positive reals
P_3	Space of 3×3 symmetric positive-definite matrices
$\mathsf{GL}(n)$	General linear group of $n \times n$ matrices
$\mathsf{O}(n)$	Space of $n \times n$ orthogonal matrices
$\mathbf{R} \setminus \{0\}$	Space of reals without the origin
$\mathsf{Stab}(x)$	Stabilizer of an element $x \in \mathcal{M}$
$g^{\mathcal{M}}$	Riemannian metric on the manifold $\mathcal{M}$
$logm$	Matrix log operation

3 Correlation on Riemannian homogeneous spaces

In this section, we define a correlation operation which generalizes the Euclidean convolution layer to arbitrary Riemannian homogeneous spaces. Further, we prove a generalization to Riemannian homogeneous spaces of the linear shift-invariant system (LSI) characterization of Euclidean convolutions. Similar theorems were first proved in [11] and later in [12]. This result is not meant to be novel but to motivate the analogous result for higher order convolutions on Riemannian homogeneous spaces that we prove subsequently.

3.1 Background

We will briefly review the differential geometry of Riemannian homogeneous spaces from an informal perspective. Formal definitions will be deferred to the appendix for conceptual clarity.

As mentioned earlier, Riemannian homogeneous spaces are Riemannian manifolds which 'look' the same locally at each point with respect to some symmetry group G, meaning that the action of G on $\mathcal{M}$ is transitive. We will specifically consider Riemannian manifolds with a transitive action of the isometry group $I(\mathcal{M})$. For the rest of the paper, we use $G = I(\mathcal{M})$ unless mentioned otherwise. For example, the 2-sphere is a homogeneous space with $G = \mathsf{SO}(3)$. An important fact about homogeneous spaces is that they can be identified as a quotient space. In general, if $\mathcal{M}$ is a homogeneous space with group G acting on it, and H_x is some stabilizer (see definition in Appendix A) of of a point $x \in \mathcal{M}$ then $\mathcal{M} \simeq G/H_x$. Returning to the 2-sphere example, if we take H to be the stabilizer of the north pole, a subgroup of $\mathsf{SO}(3)$ isomorphic to $\mathsf{SO}(2)$, then $G/H \simeq \mathsf{SO}(3)/\mathsf{SO}(2) \simeq \mathbf{S}^2$. For a detailed exposition on these concepts, we refer the reader to [18].

3.1.1 Assumptions

For the remainder of this paper we assume $\mathcal{M}$ to be a Riemannian homogeneous space admitting a transitive action of the group G, which we call the symmetries of $\mathcal{M}$. We also assume that G is a locally compact topological group. Further we assume that any function $f : \mathcal{M} \to \mathbf{R}$ is square integrable, i.e. $\int_{\mathcal{M}} |f(x)|^2 \omega^{\mathcal{M}}(x) < \infty$, where $\omega^{\mathcal{M}}$ is a suitable volume form on $\mathcal{M}$. As mentioned before, we denote the space of square integrable functions on $\mathcal{M}$ by $L^2(\mathcal{M}, \mathbf{R})$.

3.2 Euclidean LSI Theorem

We begin this section by recalling the Euclidean Linear Shift Invariant (LSI) theorem.

Definition 1. *Let $F : U \to V$ be a bounded linear operator between spaces U and V consisting of functions $\mathbf{R}^n \to \mathbf{R}$. For $f \in U \cup V$ and $x \in \mathbf{R}^n$ we define $\tau_x(f)(z) = f(z - x)$. The set $\{\tau_x\}_{x \in \mathbf{R}^n}$ forms a group under composition. We say F is translation equivariant (i.e. shift invariant in the traditional literature) if*

$$\tau_x(F(g)) = F(\tau_x(g))$$

for all $g \in U$, $x \in \mathbf{R}^n$.

Theorem 1. *Let $w : \mathbf{R}^n \to \mathbf{R}$ be a weight kernel, then the operator given as $G_w : U \to V$ is defined by $G_w(f) = f \star w$, where $\star$ is the Euclidean convolution operation which is a bounded, linear, and translation equivariant operator. Further, if F is any bounded linear translation equivariant operator, then there exists $w : \mathbf{R}^n \to \mathbf{R}$ such that $F = G_w$, i.e. $F(f) = f \star w$, for all $f \in U$.*

Thus, Euclidean convolutions with a weight kernel have an interesting and powerful characterization as linear shift invariant operators. Next we show that the correlation operation on any Riemannian homogeneous manifolds satisfy a generalization of the aforementioned LSI theorem.

3.3 Generalizing Convolutions and the LSI Theorem to Riemannian homogeneous spaces

We begin by defining the correlation operation for arbitrary homogeneous Riemannian manifolds. Some equivalent definitions have been made several times in the literature, first for specific manifolds as in [5], [7], then in more generality such as in [11] and later in [12]. We then state a generalization of the LSI theorem for this correlation operation, which we call the Linear Group Equivariant (LGE) theorem. Note that similar theorems were first proved in [11] and later in [12]. We present this theorem not as a novel result, but as motivation for a non-linear version of the theorem which we will prove in section 4. Regardless, we present a much simpler proof (compared to [11], [12]) of the result in the appendix.

Definition 2 (Correlation). *The correlation between $f : \mathcal{M} \to \mathbf{R}$ and $w : \mathcal{M} \to \mathbf{R}$ is given by, $(f \star w) : G \to \mathbf{R}$ defined as follows:*

$$(f \star w)(g) := \int_{\mathcal{M}} f(x)\,(g \cdot w)(x)\,\omega^{\mathcal{M}}(x) \tag{1}$$

The correlation between $f : G \to \mathbf{R}$ and $w : G \to \mathbf{R}$ is given by, $(f \star w) : G \to \mathbf{R}$ defined as follows:

$$(f \star w)(g) := \int_{G} f(h)\,(g \cdot w)(h)\,\mu_G(h) \tag{2}$$

where μ_G is the Haar measure on G (which is guaranteed to exist based on our assumption that G is a locally compact topological group). Please see the discussion at the end of this subsection for details.

Equation 1 is described in words as follows: the weight kernel w is "shifted" using the action of the symmetry group, and the point-wise product of the shifted weight kernel and the function f is integrated over the manifold. A similar interpretation can be given to the correlation on groups in Eq. 2. This generalizes the work on the 2-sphere presented in [5], [7] for an arbitrary Riemannian homogeneous space $\mathcal{M}$.

We show that this correlation operation is equivariant to the isometry group G of the underlying homogeneous space $\mathcal{M}$. In order to state this theorem, we first formally define equivariance.

Definition 3 (Equivariance). *Let X and Y be sets and G be a group acting on X and Y (in literature these sets are termed as G sets [20]). Then, $F : X \to Y$ is said to be equivariant to the action of G if*

$$F(g \cdot x) = g \cdot F(x) \tag{3}$$

for all $g \in G$ and all $x \in X$.

We are now ready to state the following theorems:

Theorem 2. *Let S and U be $L^2(\mathcal{M}, \mathbf{R})$ or $L^2(G, \mathbf{R})$ and $F : S \to U$ be a function given by $f \mapsto (f \star w)$. Then F is equivariant with respect to the pullback action of G, i.e.*

$$(\phi \cdot f) \star w = \phi \cdot (f \star w)$$

for $\phi \in G$ a symmetry of $\mathcal{M}$ where

$$\phi \cdot h \coloneqq h \circ \phi^{-1}$$

for any $h : \mathcal{M} \to \mathbf{R}$ square-integrable.

Proof. See appendix B. ■

This constitutes the forward direction of the LGE theorem. Now we show the converse statement, namely, every linear group equivariant function is a correlation.

Theorem 3. *Let S and U be $L^2(\mathcal{M}, \mathbf{R})$ or $L^2(G, \mathbf{R})$ and $F : S \to U$ be a linear equivariant function with respect to the pullback action of $I(\mathcal{M})$. Then, $\exists w \in S$ such that, $(F(f))(g) = (f \star w)(g)$, for all $f \in S$ and $g \in G$.*

Proof. See appendix B. ■

Together with these two theorems, we can generalize the LSI theorem to homogeneous spaces using the correlation defined in Def. 2 .

A NOTE ON VOLUME FORMS / MEASURES

In Def. 2, we specify the Haar measure for integration of a function on G. If f and w are functions on G, then the Haar measure μ_G has several desirable properties. For example, the Haar measure is invariant to "translations", i.e. if $S \subset G$ is measurable then $\mu_G(S) = \mu_G(gS)$ for any $g \in G$. Further, using the Haar measure provides a convolution theorem which makes correlation a simple multiplication under the generalized Fourier transform for groups. This particular property is vital for efficient implementations.
Note that on the other hand, we do not specify a specific volume form $\omega^{\mathcal{M}}$ for integration of function on $\mathcal{M}$ in definition 2. In many cases, the Haar measure on G will induce a G-invariant volume form on $\mathcal{M} \simeq G/H$, but stating the exact conditions for this to be possible requires some work. Instead, we define the correlation using an arbitrary volume form. In the next section we will give a construction which induces such a G-invariant volume form on $\mathcal{M}$.

3.4 Basis functions for L^2-functions on homogeneous spaces

Our goal in this section is to induce a natural basis on $L^2(\mathcal{M}, \mathbf{R})$ from the canonical basis on $L^2(G, \mathbf{R})$ where G is the group acting on the homogeneous manifold $\mathcal{M}$. The basis on G consists of matrix elements of irreducible unitary representations, which provides a Fourier transform on G (for more details reader is referred to [21]). We show that this construction matches the commonly used basis for specific manifolds, e.g. the spherical harmonics and Wigner-D functions in [5]. This construction can be used to induce basis on arbitrary Riemannian homogeneous spaces.

3.4.1 Basis on $L^2(\mathcal{M}, \mathbf{R})$ induced from $L^2(G, \mathbf{R})$

To induce a basis on $L^2(\mathcal{M}, \mathbf{R})$, we use the principal fiber bundle structure of the homogeneous manifold $\mathcal{M}$. A fiber bundle is a space that locally looks like a product space. It is expressed as a base space B with the fibers making up a fiber space F, and their union being the total space denoted by E. There is a projection map $\pi : E \to B$ mapping fibers to their "base point" on B. A principal fiber G-bundle is a fiber bundle with a continuous (right) action of a group G, such that the action of G is free, transitive and preserves the fibers. For more details on fiber bundle theory see [22].

As mentioned in 3.1, $\mathcal{M}$ can be identified with G/H for G the group action on $\mathcal{M}$ and H the stabilizer of a point $x \in \mathcal{M}$, usually called the "origin". It is well known that this identification induces a principal fiber G-bundle structure on $\mathcal{M}$ via the projection map.

Proposition 1. *[18] The homogeneous space, $\mathcal{M}$ identified as G/H together with the projection map $\pi : G \to G/H$ is a principal bundle with H as the fiber. Furthermore there exists a diffeomorphism $\psi : G/H \to \mathcal{M}$ given by $gH \mapsto g \cdot o$, where o is the "origin" of $\mathcal{M}$.*

Moreover, a section is a continuous right inverse of π, which is denoted by $\sigma : B \to E$. In literature [18], a zero section (denoted by $S \subset G$) is the section containing the identity element of H. Let $\sigma_0 : S \to \mathcal{M}$ be a diffeomorphism. Given $\{v_\alpha : G \to \mathbf{R}\}$ be the set of basis of $L^2(G, \mathbf{R})$. Then, we can get the induced basis on $L^2(\mathcal{M}, \mathbf{R})$ as $\{\widetilde{v}_\alpha = v_\alpha \circ \sigma_0^{-1}\}$. A schematic of an example fiber bundle is shown in Fig. 1.

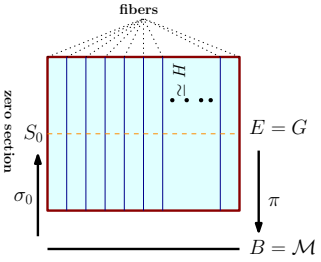


Fig. 1. Fiber bundle (B, E, π).

Example: Consider the example of $\mathcal{M} = \mathbf{S}^2$, where $G = \mathsf{SO}(3)$, $H = \mathsf{SO}(2)$. A choice of basis on $L^2(G, \mathbf{R})$ is Wigner D-functions denoted by $\left\{D^j_{l,m} | j \in \{0, 1, \cdots, \infty\}, -j \le l, m \le j\right\}$. Let (α, β, γ) be the parametrization of $\mathsf{SO}(3)$ and the zero section (S) be denoted by $\{\alpha, \beta, 0\}$. Then, $\left\{D^j_{l,0}\right\}$ are the choice of basis on $S \subset G$, which gives the induced basis on $\mathbf{S}^2$ as $\widetilde{D}^j_l(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}} D^j_{l,0}(\phi, \theta, 0)$. Further, observe that $\left\{\widetilde{D}^j_l\right\}$ are the spherical harmonics basis.

4 Higher Order Correlation on Riemannian Homogeneous Spaces

In this section, we define a higher order correlation operator on Riemannian homogeneous spaces using the Volterra series, state a theorem demonstrating it's symmetry equivariance and show how to compute it efficiently using first order correlation operations. Further, we prove that the set of functions which can be written as sums of products of linear operators and are G-equivariant can be expressed as a Volterra series. This partially generalizes the non-linear shift equivariance characterization of Volterra expansions in Euclidean space.

4.1 Volterra Series on Homogenous Spaces

We now generalize the Volterra Series to Riemannian homogeneous spaces.

Definition 4 (Volterra series expansion)**.** *We define the Volterra expansion of a function $f : \mathcal{M} \to \mathbf{R}$ or $f : G \to \mathbf{R}$ by $F(f) = \sum_{n=1}^{\infty} (f \star_n w_n)$. If $f : \mathcal{M} \to \mathbf{R}$ and $w_n : (\mathcal{M})^{\oplus n} \to \mathbf{R}$ then $(f \star_n w_n) : G \to \mathbf{R}$ is defined as,*

$$(f \star_n w_n)(g) := \int_{\mathcal{M}} \cdots \int_{\mathcal{M}} f(x_1) \cdots f(x_n)\,(g \cdot w_n)(x_1, \cdots, x_n)\omega^{\mathcal{M}}(x_1) \cdots \omega^{\mathcal{M}}(x_n)$$

If instead $f : G \to \mathbf{R}$ and $w_n : (G)^{\oplus n} \to \mathbf{R}$ then $(f \star_n w_n) : G \to \mathbf{R}$ is defined as,

$$(f \star_n w_n)(g) := \int_{G} \cdots \int_{G} f(h_1) \cdots f(h_n)\,(g \cdot w_n)(h_1, \cdots, h_n)\mu_G(h_1) \cdots \mu_G(h_n)$$

where μ_G is the Haar measure on G (which again, is guaranteed to exist based on our assumption that G is a locally compact topological group).

One can easily see that, Definition 2 is a special case of definition 4 when $n = 1$. When $n > 1$, we will call it the n^{th} order Volterra expansion. Higher order terms of the Volterra expansion express polynomial relationships between function values. An illustration of the second order Volterra kernel is provided in Figure 2. As we can see, the second order Volterra kernel has a regular correlation weight kernel at each location on the manifold $\mathcal{M}$. The results of applying these weight kernels get multiplied together to get the output of $f \star_2 w_2$. A biological motivation is provided in [17] for the (Euclidean) Volterra series. Now, we prove that F as defined in Definition 4 is equivariant to the symmetry group actions admitted by a homogeneous space.

Theorem 4. *Let S and U be $L^2(\mathcal{M}, \mathbf{R})$ or $L^2(G, \mathbf{R})$ and $F : S \to U$ be a function given by $f \mapsto \sum_{n=1}^{\infty} (f \star_n w_n)$. Then, F is equivariant.*

Proof. Observe that the sum of equivariant operators is equivariant. Hence, we only need to check that $f \star_n w_n$ is equivariant for all n. Let $g, h \in G$, let $n \in \mathbf{N}$. Then,

$$\begin{aligned}
(g.f \star_n w_n)(h) &= \left(L^*_{g^{-1}} f \star_n w_n\right)(h) \\
&= \int_{\mathcal{M}} \cdots \int_{\mathcal{M}} L^*_{g^{-1}} f(x_1) \cdots L^*_{g^{-1}} f(x_n) \\
&\quad (L^*_{h^{-1}} w_n)(x_1, \cdots, x_n)\omega^{\mathcal{M}}(x_1) \cdots \omega^{\mathcal{M}}(x_n) \\
&= \int_{\mathcal{M}} \cdots \int_{\mathcal{M}} f(y_1) \cdots f(y_n) w_n\left((h^{-1}g) \cdot y_1\right. \\
&\quad \left., \cdots, (h^{-1}g) \cdot y_n\right) \omega^{\mathcal{M}}(g \cdot y_1) \cdots \omega^{\mathcal{M}}(g \cdot y_n) \\
&= \int_{\mathcal{M}} \cdots \int_{\mathcal{M}} f(y_1) \cdots f(y_n) w_n\left((h^{-1}g) \cdot y_1\right. \\
&\quad \left., \cdots, (h^{-1}g) \cdot y_n\right) \omega^{\mathcal{M}}(y_1) \cdots \omega^{\mathcal{M}}(y_n) \\
&= (f \star_n w_n)(g^{-1}h) \\
&= L^*_{g^{-1}} (f \star_n w_n)(h) \\
&= (g \cdot (f \star_n w_n))(h)
\end{aligned}$$

Here, $(L^*_g f)(h) = f(g^{-1}h)$ (see appendix for details), since, $g, h \in G$ and n are arbitrary F is equivariant. ■

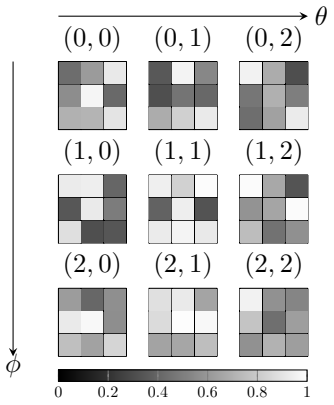


Fig. 2. Visualization (in the spirit of [17]) of second-order term $w_2 : \mathcal{M}^2 \to \mathbf{R}$ of a Volterra kernel (here on a 2-manifold parametrized by (θ, ϕ)). The coordinates above each grid represent the first entry $w_2(\mathbf{x}, \cdot)$, and within each grid the gray-scale value represents the weight of the associated kernel $w_2(\mathbf{x}, \mathbf{y})$.

In the other direction, we also show that for the aforementioned set of functions, every G-equivariant function can be written as a Volterra series.

Theorem 5. *Let S and U be $L^2(\mathcal{M}, \mathbf{R})$ or $L^2(G, \mathbf{R})$ and $F : S \to U$ be a non-linear G-equivariant function which can be written as $F = \sum_{i \in I} F_i$, where each F_i is a product of two linear functions, i.e., $F_i = F_{i,1} F_{i,2}$. Then, $\exists \{w_i\}_{i \in I} \subset S$ such that, $(F(f))(g) = \sum_{i \in I} (f \star_2 w_i)(g)$, for all $f \in S$ and $g \in G$.*

Proof. It suffices to show that for each term $F_i = F_{i,1} F_{i,2}$ (for $F_{i,k}$ a linear G-equivariant function) there exists w_i such that $F_i = f \star_2 w_i$. If $w_i(x, y) = w_{i,1}(x) w_{i,2}(y)$ (i.e. w_i is separable), then $f \star_2 w_i = (f \star w_{i,1})(f \star w_{i,2})$. But by the previous theorem, there exists $w_{i,k}$ such that $F_{i,k} = f \star w_{i,k}$, completing the proof. ■

These results partially generalize the well know non-linear shift equivariance characterization of Volterra expansions in Euclidean space and justifies the use of the Volterra series as a higher-order generalization of the correlation operation Definition 2.

4.2 Efficient Computation of the Second-Order Volterra Kernel

The Volterra series presented in the previous definition is significantly more expressive than the correlation operation defined in Definition 2 since it captures higher order relationships between inputs, but it requires the computation of iterated integrals and does not have an efficient GPU implementation. Note that for separable second order kernel w_2, $(f \star_2 w_2)(g)$ can be factored as $((f \star \tilde{w}_2)(g))((f \star \bar{w}_2)(g))$. Thus, we can compute the second order Volterra series with separable kernel as a product of traditional correlation operations. In general, we can use a convex combination of first order and second order terms of the Volterra series to define second order Volterra network.

A schematic diagram for the second order Volterra correlation operator is shown in Fig. 3. This representation of a second order kernel using product of two separable kernels is analogous to tensor product approximation of a function and can be shown to achieve approximation error

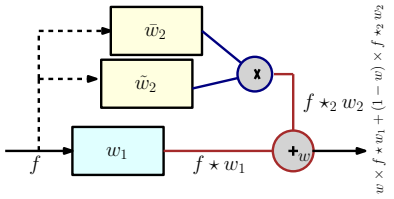


Fig. 3. Second order Volterra correlation operator with first order kernel w_1 and separable second order kernel w_2.

of an arbitrary precision [23]. The separability assumption on the kernels leads to efficient computation which is especially valuable in the network setting where these operations are performed numerous times.

5 Architecture

We now present the basic modules for implementing our correlation and higher-order Volterra operations as layers in a deep network.

5.1 Correlation on homogeneous spaces

Using definition 2 we can define:

Correlation on $\mathcal{M}$ - Corr$^{\mathcal{M}}$ (f, w): Let $f \in L^2(\mathcal{M}, \mathbf{R})$ be the input function and $w \in L^2(\mathcal{M}, \mathbf{R})$ be the mask. Then, using definition 2, Corr$^{\mathcal{M}}(f, w)$ is defined as $(f \star w) : G \to \mathbf{R}$. We have shown in Theorem 2, that Corr$^{\mathcal{M}}(f, w)$ is equivariant to the action of G. Hence, we can use Corr^{G} layer as the next layer.

Correlation on G - Corr^{G}(f, w): Let $\widetilde{f} \in L^2(G, \mathbf{R})$ be the input function and $w \in L^2(G, \mathbf{R})$ be the mask. Then analogous to Corr$^{\mathcal{M}}$, we can define Corr$^{G}(f, w)$ as $\left(\widetilde{f} \star w\right) : G \to \mathbf{R}$ using definition 2. We have used Theorem 2 to show that Corr$^{G}(f, w)$ is equivariant to the action of G. *Since this is an operation equivariant to G, we can cascade Corr^{G}.*

5.2 Volterra on homogeneous spaces

We can see that because the basic architecture of second order Volterra series consists of the following modules:

Second order Volterra on $\mathcal{M}$ - Corr$_2^{\mathcal{M}}(f, w_1, w_2)$: Let $f \in L^2(\mathcal{M}, \mathbf{R})$ be the input function and $w_1 : \mathcal{M} \to \mathbf{R}$ and $w_2 : (\mathcal{M})^{\oplus 2} \to \mathbf{R}$ be the kernels. Then, Corr$_2^{\mathcal{M}}(f, w_1, w_2) := \sum_{j=1}^{2} (f \star_j w_j) : G \to \mathbf{R}$. We have shown in Theorem 4, that Corr$_2^{\mathcal{M}}(f, w_1, w_2)$ is equivariant to the action of G. Hence, we can use Corr$_2^{G}(f, w_1, w_2)$ layer as the next layer.

Second order Volterra on G - Corr$_2^{G}(f, w_1, w_2)$: Let $f \in L^2(G, \mathbf{R})$ be the input function and $w_1 : G \to \mathbf{R}$ and $w_2 : (G)^{\oplus 2} \to \mathbf{R}$ be the kernels. Then, Corr$_2^{G}(f, w_1, w_2) := \sum_{j=1}^{2} (f \star_j w_j) : G \to \mathbf{R}$. We have used Theorem 4 to show that Corr$_2^{G}(f, w_1, w_2)$ is equivariant to the action of G. *Since this is an operation equivariant to G, we can cascade Corr$_2^{G}(f, w_1, w_2)$.*

5.3 Other Layers

Activation function: Since the outputs of all the above layers are functions from G to $\mathbf{R}$, we will use the standard activation operation on $\mathbf{R}$.

Invariant last layer: As both layers, Corr$_2^{\mathcal{M}}$ and Corr$_2^{G}$ are equivariant to the action of G, so are the cascaded layers. Since, if the input signal is transformed by a group element $g \in G$, so is the output of Corr$_2^{\mathcal{M}}$ as this layer is equivariant. Thus the output of Corr$_2^{\mathcal{M}}$ is transformed by the same group element g. Hence, the input of Corr$_2^{G}$ is transformed by g and due to the equivariance so is the output of Corr$_2^{G}$. This justifies that the cascaded layers are equivariant to the action of G. Hence, after the cascaded correlation layers, the output $\widetilde{f} \in L^2(G, \mathbf{R})$ lies on a G set. Similar to the Euclidean CNN, we want the last layer to be G invariant. Hence, we will integrate $\widetilde{f}$ on the domain G and return a scalar. Note that in the experiment, we learn multiple channels analogous to the Euclidean CNN, where in each channel, we learn a G equivariant $\widetilde{f} \in L^2(G, \mathbf{R})$. Thus, after the integration, we have c scalars, where c is the number of channels, which will be input to the softmax fully connected layer similar to the Euclidean CNN. We abbreviate this last layer as **iL** (invariant layer).

A schematic diagram of our proposed VolterraNet is shown in Fig. 4.

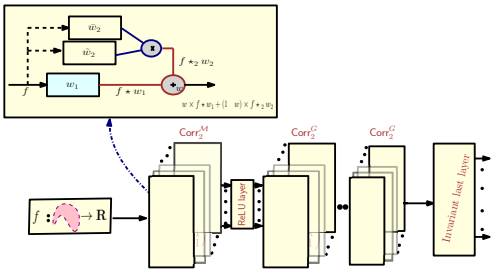


Fig. 4. Schematic diagram of a second order Volterranet

6 Dilated VolterraNet

In this section, we propose a dilated VolterraNet framework which is suitable for sequential data. Sequential data here refers to a sequence of data points (signal measurements at voxels) along the neuronal fiber tracts that are extracted from diffusion MRI data sets. Neuronal fiber tracts in certain regions of the brain are disrupted by movement disorders such as Parkinsons disease. The sensory motor area tract (pathway) in the brain is one such neuronal pathway where the disease caused changes are expected to be observed. By treating this pathway as a sequence of points where diffusion sensitized MR signal is acquired, we propose to apply the dilated VolterraNet (described below) to analyze this data.

It is known that the sequential data should involve recurrent structure [24], but as pointed out in [25], convolutional architectures often outperform recurrent models in sequential data analysis. Furthermore, recurrent models are computationally more expensive than the convolutional models. But, note that in order to mimic the infinite memory capabilities of a recurrent model, one needs to increase the receptive field by using the dilated convolutions. We will first recap the definition of Euclidean dilated convolution [25] and then describe the proposed dilated VolterraNet.

6.1 Euclidean Dilated Convolution:

Given a one-dimensional input sequence $\mathbf{x} : \mathbf{N} \to \mathbf{R}^n$ and a kernel $w : \{0, \cdots, k-1\} \to \mathbf{R}$, the dilated convolution function $(\mathbf{x} \star_d w) : \mathbf{N} \to \mathbf{R}^n$ is defined as,

$(\mathbf{x} \star_d w)(s) = \sum_{i=0}^{k-1} w(i)\mathbf{x}(s - d \times i)$, where $\mathbf{N}$ is the set of natural numbers and k and d are the kernel size and the dilation factor respectively. Note that with $d = 1$, we get the normal convolution operator. In a dilated CNN, the receptive field size will depend on the depth of the network as well as on the choice of k and d.

6.2 Dilated VolterraNet

Now we present a dilated VolterraNet model by combining the VolterraNet with the dilated CNN model. Given a one-dimensional input sequence $\{f_i : \mathcal{M} \to \mathbf{R}\}$, we will first apply $\mathsf{Corr}_2^{\mathcal{M}}$ and cascaded Corr_2^G layers to each point in the sequence independently. The output of a Corr_2^G layer is a function $G \to \mathbf{R}$. Let the output of the last Corr_2^G layer be $\{g_i : G \to \mathbf{R}\}$. Then, we discretize the group G, to represent each g_i by a vector $\mathbf{x}_i$ (as shown in Fig. 5). The steps of discretization, i.e., length of $\mathbf{x}_i$, are chosen via grid search in the experimental section. This is analogous to the standard practice in literature [5], [7]. Polar coordinates on G are used to discretize G and then we use the dilated CNN by treating each sample as a vector. This essentially amounts to choosing a uniform grid in the parameter space using Rodrigues vectors [26], although more sophisticated techniques can be employed in this context [27]. Now, we input $\{\mathbf{g}_i\}$ to the Euclidean dilated CNN (since the components of $\mathbf{g}_i$ are real) to construct a dilated VolterraNet framework. In Fig. 5, we present a schematic of dilated VolterraNet with input $\{f_i\}$ followed by $\mathbf{Corr}_2^{\mathcal{M}} \to \mathbf{ReLU} \to \mathbf{Corr}_2^G \to \mathbf{ReLU} \to \mathbf{Corr}_2^G$.

A self explanatory schematic diagram of the dilated VolterraNet architecture is shown in Fig. 5.

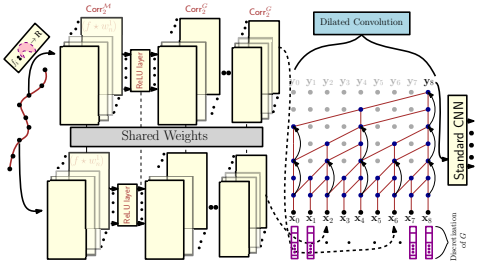


Fig. 5. Schematic diagram of dilated VolterraNet

7 Experiments

In this section, we present experiments on spherical MNIST, atomic energy and Shrec17 data sets respectively. We present comparisons of performance of our VolterraNet to Spherical CNN by Cohen et al. [5] and Clebsch-Gordan net by Kondor et al. [8]. Further, we also present a separate comparison with the spherical CNN presented most recently in [7] on the shrec17 data set. The separate comparison was necessary due to the fact that the loss function used in [7] was distinct from the one used in [5], [8]. Finally, we extend the VolterraNet to a dilated version, a higher order analogue of the dilated CNN and use it to demonstrate its efficacy in group testing on diffusion MRI (dMRI) data acquired from movement disorder patients. The data in this example reside in a product space, $\mathbf{S}^2 \times \mathbf{R}^+$, which is a Riemannian homogeneous space distinct from $\mathbf{S}^2$. This experiment serves as an example demonstrating the ability of VolterraNet to cope with manifolds other than the sphere.

Choice of basis: In our experiments, we have three examples of manifolds, $\mathbf{S}^2$, $\mathbf{S}^2 \times \mathbf{R}^+$ and P_3. For $\mathsf{SO}(3)$, we use the Weigner basis and for $\mathbf{S}^2 \simeq \mathsf{SO}(3)/\mathsf{SO}(2)$, we use the induced basis, i.e., the Spherical Harmonics basis. For $\mathbf{S}^2 \times \mathbf{R}^+$, we use the product basis of each of the spaces, i.e., Spherical Harmonics for $\mathbf{S}^2$ and the canonical basis for $\mathbf{R}^+$. Since P_3 can be written as $\mathsf{GL}(3)/\mathsf{O}(3)$, we use the induced basis on P_3 which are induced from the canonical basis on $\mathsf{GL}(3)$.

In all the experiments we compare the VolterraNet architecture to the state of the art models, and additionally compare it to an architecture which we call Homogeneous CNN (HCNN) which replaces the Volterra non-linear convolutions/correlations with the correlation operations from definition 2. We have released an implementation of the VolterraNet architecture with the Spherical MNIST experiment which can be found at, https://github.com/cvgmi/volterra-net.

7.1 Synthetic Data Experiment: Classification of data on P_3

In this section, we first describe the process of synthesizing functions $f : P_3 \to [0, 1]$. In this experiment, we generated data samples drawn from distinct Gaussian distributions defined on P_3 [28]. Let $\mathcal{X}$ be a P_3 valued random variable that follows $\mathcal{N}(M, \sigma)$, then, the p.d.f. of X is given by [28]:

$$f_{\mathcal{X}}(X; M, \sigma) = \frac{1}{C(\sigma)} \exp\left(-\frac{d^2(M, X)}{2\sigma^2}\right), \qquad (4)$$

where, $d(.,.)$ is the affine invariant geodesic distance on P_3 as given by $d(M, X) = \sqrt{\text{trace}\left((\text{logm}\,(M^{-1}X))^2\right)}$.

We first chose two sufficiently spaced apart location parameters M_1 and M_2 and then for the i^{th} class we generate Gaussian distributions with location parameters that are perturbations of M_i and with variance 1. This gives us two clusters in the space of Gaussian densities on P_3, which we will classify using HCNN and VolterraNet. In this case, the HCNN network architecture is given by: $\mathbf{Corr}^{P_3} \to \mathbf{ReLU} \to \mathbf{Corr}^{\mathsf{GL}(3)} \to \mathbf{ReLU} \to \mathbf{Corr}^{\mathsf{GL}(3)} \to \mathbf{ReLU} \to \mathbf{iL} \to \mathbf{FC}$. and for VolterraNet the correlation operations are replaced with the corresponding Volterra convolutions.

Model	mean acc.	std. acc.
VolterraNet	**91.50**	**0.08**
HCNN	86.50	0.02

TABLE 1
Comparative mean and stdev. on the synthetic data

The data consists of 500 samples from each class, where each sample is drawn from a Gaussian distribution on P_3. The classification accuracies in a ten-fold partition of the data are shown in Table 1. In most deep learning applications, one is used to seeing a high classification accuracy, but we believe that this can be achieved here as well by increasing the number of layers and possibly overfitting the data. The purpose of this synthetic experiment was not to seek an "optimal" classification accuracy but to provide a flexible framework which if "optimally" tuned can yield a good testing accuracy for data whose domain is a non-compact Riemannian homogeneous space.

7.2 Spherical MNIST Data Experiment

The spherical MNIST data are generated using the scheme described in [5]. There are two instances of this data, one in which we project MNIST digits on the northern hemisphere (denoted by 'NR') and the other where we apply random rotation afterwards (denoted by 'R'). The spherical signal is discretized using a bandwidth of 60.

We selected the same baseline model as was chosen in [5], which is a Euclidean CNN with 5×5 filters and $32, 64, 10$ channels with a stride of 3 in each layer. This CNN is trained by mapping the digits from the northern hemisphere onto the plane. The Spherical CNN model [5] we used has the following architecture (as was reported in [5]), $\textbf{Corr}^{\mathbf{S}^2} \rightarrow \textbf{ReLU} \rightarrow \textbf{Corr}^{\mathbf{SO}(3)} \rightarrow \textbf{ReLU} \rightarrow \textbf{FC}$ with bandwidths 20, 12 and the number of channels 20, 40 respectively. We used the same architecture for Clebsch-Gordan net as was reported in [8].

For our method, we used a second order Volterra network with the following architecture: $\textbf{Corr}_2^{\mathbf{S}^2} \rightarrow \textbf{ReLU} \rightarrow \textbf{iL} \rightarrow \textbf{FC}$ with bandwidth 30, 20 respectively and number of features 25, 10 respectively. We chose a batchsize of 32 and learning rate of 5×10^{-3} with ADAM optimization [29].

We performed three sets of experiments: non rotated training and test sets (denoted by 'NR/NR'), non rotated training and randomly rotated test sets (denoted by 'NR/R') and randomly rotated both training and test sets (denoted by 'R/R'). The comparative results in terms of classification accuracy are shown in Table 2.

Method	NR/NR	NR/R	R/R	# params.
Baseline CNN	97.67	22.18	12.00	68000
Spherical CNN [5]	95.59	94.62	93.40	58550
Clebsch-Gordan net [8]	96.00	95.86	95.80	342086
VolterraNet	**96.72**	**96.10**	**96.71**	**46010**

TABLE 2
Comparison of classification accuracy on Spherical MNIST data

We can see that the VolterraNet performed better than all the three competing networks for both the 'R/R' and 'NR/R' cases. Note that in terms of number of parameters, VolterraNet used **46010**, while Spherical CNN used **58550** and Clebsch-Gordan net used **342086**. The baseline CNN used **68000** parameters. Thus in comparison, we have approximately an 86% reduction in parameters over the Clebsch-Gordan net with almost equal or better classification accuracy. In comparison to the Spherical CNN, we have approximately a 21% reduction in the parameters over the Spherical CNN while achieving significantly better performance. This clearly depicts the usefulness of our proposed VolterraNet in comparison to existing networks used in processing this type of data in a non-Euclidean domain.

7.3 3D Shape Recognition Experiment

We now report results for shape classification using the Shrec17 dataset [30] which consists of 51300 3D models spread over 55 classes. This dataset is divided into a $70/10/20$ split for train/validation/test. Following the method in [5], we perturbed the dataset using random rotations. We processed the dataset as in [5]. Basically, we represented each 3D model by a spherical signal using a ray casting scheme. For each point on the sphere, a ray towards the origin is sent which collects the ray length, cosine and sine of the surface angle. Additionally, the convex hull of the 3D shape gives 3 more channels, which results in 6 input channels. The spherical signal is discretized using Driscoll-Healy grid [13] grid with a bandwidth of 128.

The Spherical CNN model [5] we used has the following architecture (as was reported in [5]): $\textbf{Corr}^{\mathbf{S}^2} \rightarrow \textbf{BN} \rightarrow \textbf{ReLU} \rightarrow \textbf{Corr}^{\mathbf{SO}(3)} \rightarrow \textbf{BN} \rightarrow \textbf{ReLU} \rightarrow \textbf{Corr}^{\mathbf{SO}(3)} \rightarrow \textbf{BN} \rightarrow \textbf{ReLU} \rightarrow \textbf{FC}$ with bandwidths 32, 22 and 7 and the number of channels 50, 70 and 350 respectively. We used the same architecture for Clebsch-Gordan net as was reported in [8].

In our method, we used a second order Volterra network with the following architecture: $\textbf{Corr}^{\mathbf{S}^2} \rightarrow \textbf{BN} \rightarrow \textbf{ReLU} \rightarrow \textbf{Corr}_2^{\mathbf{SO}(3)} \rightarrow \textbf{BN} \rightarrow \textbf{ReLU} \rightarrow \textbf{iL} \rightarrow \textbf{FC}$ with bandwidths 10, 8, 8 respectively and number of features 60, 80, 100 respectively. We chose a batch size of 100 and a learning rate of 5×10^{-3} with ADAM optimization [29]. Table 3 summarizes comparison of VolterraNet with other existing deep network architectures that reported results on this data in literature. From this table, it is evident that VolterraNet almost always yields classification accuracy results within the top three methods, while having the best parameter efficiency.

Comparison with Esteves et al. [7]: We also compared our VolterraNet with recent work of Esteves et al. [7] using an extra in-batch triplet loss [34] (as used in Esteves et al. [7]). We show the comparison results in Table 3 (last two rows), which clearly shows that, **(a)** The VolterraNet outperforms the network in [7] (which is the state-of-the-art algorithm in terms of parameter efficiency). **(b)** The triplet loss boosts the performance of VolterraNet relative to the baseline loss of cross entropy.

7.4 Regression Experiment: Prediction of Atomic Energy

Here, we report the application of our VolterraNet to the QM7 dataset [35], [36], where the goal is to regress over atomization energies of molecules given atomic positions $(\mathbf{p}_i)$ and charges $(\mathbf{z}_i)$. Each molecule consists of at most 23 atoms and the molecules are of 5 types (C, N, O, S, H). We use the Coulomb Matrix (CM) representation proposed by [36], which is rotation and translation invariant but not permutation invariant. We used a similar experimental setup to that described in [5] for this regression problem. We define a sphere $\mathbf{S}_i$ around $\mathbf{p}_i$ for each i^{th} atom. We define the potential functions

$$U_{\mathbf{z}}(\mathbf{x}) = \sum_{i \neq j, \mathbf{z}_j = \mathbf{z}} \frac{\mathbf{z}_i^t \mathbf{z}}{\|\mathbf{x} - \mathbf{p}_i\|}, \tag{5}$$

for every $\mathbf{z}$ and for every $\mathbf{x}$ on the sphere $\mathbf{S}_i$. This yields a spherical signal consisting of 5 features which were discretized using the Driscoll-Healy grid [13] with a bandwidth of 20. For the VolterraNet, we used one $\mathbf{S}^2$ and $\mathsf{SO}(3)$ second order Volterra block with bandwidths $12, 8, 8, 4$ and number of features $8, 10, 20, 50$ respectively.

We compute the loss and report it in Table 4. We can see that VolterraNet performs better than the competing methods. For Spherical CNN [5] and Clebsch-Gordan net [8], we used

Method	P@N	R@N	F1@N	mAP	NDCG	# params.
Tasuma_ReVGG [31]	0.70	0.76	0.72	0.69	0.78	3M
Furuya_DLAN [32]	0.81	0.68	0.71	0.65	0.75	8.4M
SHREC16-bai_GIFT [33]	0.68	0.66	0.66	0.60	0.73	36M
Deng_CM-VGG5-6DB	0.41	0.70	0.47	0.52	0.62	-
Spherical CNN [5]	0.70	0.71	0.69	0.67	0.76	1.4Mil
Clebsch-Gordan net [8]	0.70	0.72	0.70	0.68	0.76	-
Ours (VolterraNet)	**0.71** (2nd)	**0.70** (3rd)	**0.70** (3rd)	**0.67** (3rd)	**0.75** (4th)	**396297**
[7] (w triplet loss)	0.72	0.74	0.69	-	-	0.5M
Ours (w triplet loss)	**0.73**	**0.74**	**0.70**	**0.68**	**0.76**	**396297**

TABLE 3
Comparison results in terms of classification accuracy on the shrec17 data

Method	MSE
MLP/ Random CM [37]	5.96
LGIKA (RF) [38]	10.82
RBF Kernels/ Random CM [37]	11.42
RBF Kernels/ Sorted CM [37]	12. 59
MLP/ Sorted CM [37]	16.06
Spherical CNN [5]	8.47
Clebsch-Gordan net [8]	7.97
Ours (VolterraNet)	**5.92** (1st)

TABLE 4
Comparison results on atomic energy prediction

similar architectures as described in the respective papers. Spherical CNN [5] and Clebsch-Gordan net [8] use $1.4M$ and $1.1M$ parameters respectively, while the VolterraNet used 128460, nearly an order of magnitude reduction of parameters, while achieving the best classification accuracy. This illustrates the parameter efficiency gains that we get from using a higher order correlation, a richer feature, in the VolterraNet.

7.5 Network architecture for dMRI data Using Dilated VolterraNet

Diffusion MRI (dMRI) is an imaging modality that non-invasively measures the diffusion of water molecules in tissue samples being imaged. It serves as an interesting example of our framework since dMRI data can naturally be described by functions on a Riemannian homogeneous space. In this section we describe the dMRI data and its processing using the framework presented in this paper, which will help the reader understand the results of the following subsections.

In each voxel of a dMRI data set, the signal magnitude is represented by a real number along each gradient magnetic field over a hemi-sphere of directions in 3D. Hence, in each voxel, we have a function $f : \mathbf{S}^2 \times \mathbf{R}^+ \to \mathbf{R}$. The proposed network architecture has two components: *intra-voxel layers* and *inter-voxel layers*. The *intra-voxel layers* extract features from each voxels, while the *inter-voxel layers* use dilated convolution to capture the interaction between extracted features. In our application in the next section we extract a sequence of voxels lying along a nerve fiber bundle in the brain known to be affected in Parkinson disease. Hence we have a sequence of functions along the fiber bundle $\{f_i : \mathbf{S}^2 \times \mathbf{R}^+ \to \mathbf{R}\}$, making the application of the dilated VolterraNet in section 6.2 appropriate.

7.5.1 Extracting intra-voxel features

We extract intra-voxel features (independently) from each voxel. As mentioned before, in each voxel we have a function $f : \mathbf{S}^2 \times \mathbf{R}^+ \to \mathbf{R}$. Since $\mathbf{S}^2 \times \mathbf{R}^+$ is a Riemannian homogeneous space (endowed with the product metric), we will use a cascade of the Volterra correlation layers defined earlier (with standard non-linearity between layers) to extract features which are *equivariant* to the action of $\mathsf{SO}(3) \times (\mathbf{R} \setminus \{0\})$. These features are extracted independently within each voxel. Observe that this equivariance property is natural in the context of dMRI data. Since in each voxel of the dMRI data, the signal is acquired in different directions (in 3D), we want the features to be equivariant to the 3D rotations and scaling.

7.5.2 Extracting inter-voxel features

After the extraction of the intra-voxel features (which are equivariant to the action of G), we seek to derive features based on the interactions *between* the voxels. Here we use the standard dilated convolution (as described in 6.1) layers to capture the interaction between features extracted from voxels.

Now, we are ready to give the details of the data used for the experiment of our proposed Dilated-VolterraNet. For this experiment, we used a second order Dilated-VolterraNet with 3 dilated layers of kernel size (5×5) and dilation factors of 1, 2 and 4 respectively.

7.6 Dilated VolterraNet Experiment: Group testing on movement disorder patients

This dMRI data was collected from 50 PD patients and 44 controls at the University of Florida and are accessible via request from the NIH-NINDS Parkinson's Disease Biomarker Program portal https://pdbp.ninds.nih.gov/. All images were collected using a 3.0 T MR scanner (Philips Achieva) and 32-channel quadrature volume head coil. The parameters of the diffusion imaging acquisition sequence were as follows: gradient directions = 64, b-values = 0/1000 s/mm2, repetition time =7748 ms, echo time = 86 ms, flip angle = 90°, field of view = 224 × 224 mm, matrix size = 112 × 112, number of contiguous axial slices = 60, slice thickness = 2 mm. Eddy current correction was applied to each data set by using standard motion correction techniques.

We first extracted the sensory motor area tracts called M1 fiber tracts (as shown in Fig.6) using the FSL software [40] from both the left ('LM1') and right hemispheres ('RM1'). We applied the Dilated-Volterra to the raw signal measurements along the fiber tracts. Our

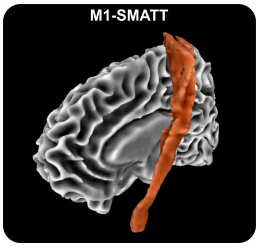


Fig. 6. M1 Template [39]

$\mathbf{Corr}^{\mathbf{S}^2}$	$\mathbf{Corr}^{\mathbf{SO}(3)}$	p-val.	
		'LM1'	'RM1'
N	N	**0.01**	**0.02**
Y	N	**0.04**	**0.03**
N	Y	0.13	0.24
Y	Y	0.15	0.26

TABLE 5
Ablation studies for Dilated-VolterraNet

model was trained on both the control (normal subjects) group and the PD group data sets, i.e., we learned two Dilated VolterraNet models, one each for the control and the PD groups respectively. Using the method from [41] we compute the distance between these two models, denoted by d. Now we permute the class labels between the classes, retrain two models and compute the network distance d_j. If there are significant differences between the classes we should expect that $d > d_j$. We repeat this experiment for $j = 1, \ldots, 1000$ and let p be the proportion of experiments for which $d \leq d_j$. This is a permutation test of the null hypothesis: there is no significant difference between the tract models learned from the two different classes. We also performed ablation studies with regards to the order of the model in the Dilated-VolterraNet to study the effect of higher order convolutions. We used the following architecture $\mathbf{Corr}_2^{\mathbf{S}^2} \rightarrow \mathbf{BN} \rightarrow \mathbf{ReLU} \rightarrow \mathbf{Corr}_2^{\mathbf{SO}(3)} \rightarrow \mathbf{BN} \rightarrow \mathbf{ReLU}$ as our baseline model and then replaced $\mathbf{Corr}_2^{\mathbf{SO}(3)}$ and $\mathbf{Corr}_2^{\mathbf{S}^2}$ to $\mathbf{Corr}^{\mathbf{SO}(3)}$ and $\mathbf{Corr}^{\mathbf{S}^2}$ respectively in an alternating fashion. The ablation study result is presented in Table 5. The 'N' in $\mathbf{Corr}^{\mathbf{SO}(3)}$ ($\mathbf{Corr}^{\mathbf{S}^2}$) indicates that we used second order for the respective convolution operator. The table shows that a second order representation in later layers is very useful and hence a model with $\mathbf{Corr}^{\mathbf{SO}(3)}$ performs poorly but a model with $\mathbf{Corr}^{\mathbf{S}^2}$ and $\mathbf{Corr}_2^{\mathbf{SO}(3)}$ performs as good as the model with both second order kernels. Both models reject the null hypothesis with 95% confidence.

We compared our dilated VolterraNet with the standard (no dilation) VoletrraNet and as expected we needed $\approx 1.5\times$ parameters in case of standard VolterraNet to achieve p-values of 0.03 and 0.04 for LM1 and RM1 respectively, which is similar in performance to its dilated counterpart. Additionally, we compared our network's performance to the performance of a similar dMRI architecture (recurrent model) namely, the SPD-SRU [39] and the baseline model used for comparison in [39] (see section 5.2 of [39] for details on the baseline model). We found that the baseline method yielded a p-value of 0.17 and 0.34 respectively for 'LM1' and 'RM1'. Whereas, the SPD-SRU architecture yielded a p-values of 0.01 and 0.032 respectively. We can conclude that both using standard and Dialted VolterraNet we can reject the null hypothesis with 95% confidence whereas Dilated VoletrraNet can achieve the statistically significant result with $\approx 33\%$ reduction in number of parameters compared to its standard counterpart.

8 Conclusions

In this paper, we presented a novel generalization of CNNs to non-Euclidean domains specifically, Riemannian homogeneous spaces. More precisely, we introduced higher order convolutions – represented using a Volterra series – on Riemannian homogeneous spaces. We call our network a Volterra homogeneous CNN abbreviated as VolterraNet. The salient contributions of our work are: (i) A proof of equivariance of higher order convolutions to group actions on homogeneous Riemannian manifolds. Proofs of generalized Linear Shift Invariant (equivariant) and Nonlinear Shift Invariant (eqivariant) theorems for correlations and Volterra series defined on Riemannian homogeneous spaces. (ii) We prove that second order Volterra convolutions can be expressed as a cascade of convolutions. This allows for efficient implementation of second-order Volterra representation used in the VolterraNet. (iii) In support of our conjecture on the reduced number of parameters, real data experiments empirically demonstrate that VolterraNet requires less number of parameters to achieve the baseline accuracy of classification in comparison to both Spherical-CNN and Clebsch-Gordan net. (iv) We also presented a dilated VolterraNet that was shown to be effective on a group testing experiment on movement disorder patients. Our future work will be focused on performing more real data experiments to demonstrate the power of VolterraNet for a variety of data domains that are Riemannian homogeneous spaces.

Acknowledgements

This research was in part funded by the NSF grant IIS-1724174 to BCV. We thank Professor David Vaillancourt of the University of Florida, Department of Applied Physiology and Kinesiology for providing us with the diffusion MRI scans used in this work.

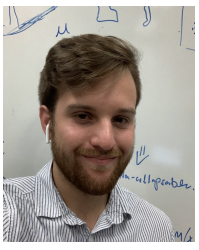
References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[2] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Tech. Report, University of Toronto, Tech. Rep., 2009.
[3] V. Volterra, *Theory of functionals and of integral and integro-differential equations*. Courier Corporation, 2005.
[4] E. Worrall, J. Garbin, D. Turmukhambetov, and J. Brostow, "Harmonic networks: Deep translation and rotation equivariance," in *Proceedings of the IEEE CVPR*. IEEE, 2017, pp. 5026–5037.
[5] T. Cohen, M. Geiger, J. Koehler, and M. Welling, "Spherical CNNs," in *Proceedings of ICLR*. JMLR, 2018.
[6] ——, "Convolutional networks for spherical signals," in *Proceedings of ICML*. JMLR, 2017.
[7] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, "Learning so (3) equivariant representations with spherical cnns," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 52–68.
[8] R. Kondor, Z. Lin, and S. Trivedi, "Clebsch-gordan nets: a fully fourier space spherical convolutional neural network," *arXiv preprint arXiv:1806.09231*, 2018.
[9] C. Esteves, C. Allen-Blanchette, X. Zhou, and K. Daniilidis, "Polar transformer networks," *arXiv preprint arXiv:1709.01889*, 2017.
[10] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *Advances in neural information processing systems (NIPS)*, 2015, pp. 2017–2025.
[11] R. Kondor and S. Trivedi, "On the generalization of equivariance and convolution in neural networks to the action of compact groups," *arXiv preprint arXiv:1802.03690*, 2018.
[12] T. Cohen, M. Geiger, and M. Weiler, "A general theory of equivariant cnns on homogeneous spaces," *arXiv preprint arXiv:1811.02017*, 2018.
[13] J. R. Driscoll and D. M. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in applied mathematics*, vol. 15, no. 2, pp. 202–250, 1994.

[14] R. Kumar, A. Banerjee, and B. C. Vemuri, "Volterrafaces: Discriminant analysis using volterra kernels," 2009.
[15] R. Kumar, A. Banerjee, B. C. Vemuri, and H. Pfister, "Trainable convolution filters and their application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 7, pp. 1423–1436, 2012.
[16] N. Hakim, J. Kaufman, G. Cerf, and H. Meadows, "Volterra characterization of neural networks," in *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*. IEEE, 1991, pp. 1128–1132.
[17] G. Zoumpourlis, A. Doumanoglou, N. Vretos, and P. Daras, "Non-linear convolution filters for cnn-based learning," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4771–4779.
[18] S. Helgason, *Differential geometry and symmetric spaces*. Academic press, 1962, vol. 12.
[19] M. Banerjee, R. Chakraborty, D. Archer, D. Vaillancourt, and B. C. Vemuri, "Dmr-cnn: A cnn tailored for dmr scans with applications to pd classification," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 388–391.
[20] D. S. Dummit and R. M. Foote, *Abstract algebra*. Wiley Hoboken, 2004, vol. 3.
[21] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis: Volume I Structure of Topological Groups Integration Theory Group Representations*. Springer Science & Business Media, 2012, vol. 115.
[22] P. W. Michor, *Topics in differential geometry*. American Mathematical Soc., 2008, vol. 93.
[23] W. Hackbusch and B. N. Khoromskij, "Tensor-product approximation to operators and functions in high dimensions," *Journal of Complexity*, vol. 23, no. 4-6, pp. 697–714, 2007.
[24] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
[25] S. Bai, J. Z. Kolter, and V. Koltun, "Convolutional sequence modeling revisited," 2018.
[26] W. R. Hamilton, *Elements of quaternions*. Longmans, Green, & Company, 1866.
[27] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Discretization of so (3) using recursive tesseract subdivision," in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2017, pp. 49–55.
[28] G. Cheng and B. C. Vemuri, "A novel dynamic system in the space of spd matrices with applications to appearance tracking," *SIAM journal on imaging sciences*, vol. 6, no. 1, pp. 592–615, 2013.
[29] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, vol. 5, 2015.
[30] L. Yi, L. Shao, M. Savva, H. Huang, Y. Zhou, Q. Wang, B. Graham, M. Engelcke, R. Klokov, V. Lempitsky *et al.*, "Large-scale 3d shape reconstruction and segmentation from shapenet core55," *arXiv preprint arXiv:1710.06104*, 2017.
[31] A. Tatsuma and M. Aono, "Multi-fourier spectra descriptor and augmentation with spectral clustering for 3d shape retrieval," *The Visual Computer*, vol. 25, no. 8, pp. 785–804, 2009.
[32] T. Furuya and R. Ohbuchi, "Deep aggregation of local 3d geometric features for 3d model retrieval." in *BMVC*, 2016, pp. 121–1.
[33] C. R. Qi, H. Su, M. Nießner, A. Dai, M. Yan, and L. J. Guibas, "Volumetric and multi-view cnns for object classification on 3d data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5648–5656.
[34] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
[35] L. C. Blum and J.-L. Reymond, "970 million druglike small molecules for virtual screening in the chemical universe database gdb-13," *Journal of the American Chemical Society*, vol. 131, no. 25, pp. 8732–8733, 2009.
[36] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical review letters*, vol. 108, no. 5, p. 058301, 2012.
[37] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, A. V. Lilienfeld, and K.-R. Müller, "Learning invariant representations of molecules for atomization energy prediction," in *Advances in Neural Information Processing Systems*, 2012, pp. 440–448.
[38] A. Raj, A. Kumar, Y. Mroueh, P. T. Fletcher, and B. Schölkopf, "Local group invariant representations via orbit embeddings," *arXiv preprint arXiv:1612.01988*, 2016.
[39] R. Chakraborty, C.-H. Yang, X. Zhen, M. Banerjee, D. Archer, D. Vaillancourt, V. Singh, and B. C. Vemuri, "Statistical recurrent models on manifold valued data," *ArXiv e-prints*, 2018.
[40] D. B. Archer, D. E. Vaillancourt, and S. A. Coombes, "A template and probabilistic atlas of the human sensorimotor tracts using diffusion mri," *Cerebral Cortex*, vol. 28, no. 5, pp. 1685–1699, 2017.
[41] U. Triacca, "Measuring the distance between sets of ARMA models," *Econometrics*, vol. 4, no. 3, p. 32, 2016.

Monami Banerjee received her Ph.D. in computer science from the Univ. of Florida in 2018. She is currently a research staff member at Facebook Oculus, Menlo Park. Her research interests lie at the intersection of Geometry, Computer Vision and Medical Image Anlaysis.

Rudrasis Chakraborty received his Ph.D. in computer science from the Univ. of Florida in 2018. He is currently a post doctoral researcher at UC Berkeley. His research interests lie in the intersection of Geometry, ML and Computer Vision.



Jose Bouza is a fourth year Mathematics and Computer Science undergraduate at the University of Florida. His primary interests encompass computer vision and applied topology.

Baba C. Vemuri received his PhD in Electrical and Computer Engineering from the University of Texas at Austin. Currently, he holds the Wilson and Marie Collins professorship in Engineering at the University of Florida and is a full professor in the Department of Computer and Information Sciences and Engineering. His research interests include Statistical Analysis of Manifold-valued Data, Medical Image Computing, Computer Vision and Machine Learning. He is a recipient of the IEEE Technical Achievement Award (2017) and is a Fellow of the IEEE (2001) and the ACM (2009).