AZIMUTHAL ROTATIONAL EQUIVARIANCE IN SPHERICAL CNNS

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work, we analyze linear operators from $L^2(S^2) \to L^2(S^2)$ which are equivariant to azimuthal rotations, that is, rotations around the z-axis. Several high-performing neural networks defined on the sphere are equivariant to azimuthal rotations, but not to full SO(3) rotations. Our main result is to show that a linear operator acting on band-limited functions on the sphere is equivariant to azimuthal rotations if and only if it can be realized as a block-diagonal matrix acting on the spherical harmonic expansion coefficients of its input. Further, we show that such an operation can be interpreted as a convolution, or equivalently, a correlation in the spatial domain. Our theoretical findings are backed up with experimental results demonstrating that a state-of-the-art pipeline can be improved by making it equivariant to azimuthal rotations.

1 INTRODUCTION

Signals defined on the surface of a sphere arise frequently in applications. For example, omnidirectional cameras are increasingly being utilized in applications that benefit from a large field of view, such as visual odometry and mapping (Matsuki et al., 2018). Spherical signals may also arise from other forms of measurements, such as magnetoencephalography images of brain activity or from planetary data like surface temperature or wind speed measurements (Mudigonda et al., 2017).

Convolutional neural networks (CNNs) (LeCun et al., 1998) have achieved remarkable success in the analysis of images and other data that is naturally arranged in a planar, rectangular grid, but applying them to spherical data has proven more challenging. The reason for this is that it is impossible to map the sphere to a flat surface without introducing distortions. For instance, in the equirectangular projection of spherical images, objects are increasingly stretched out as they approach the poles.

To remedy this, several alternatives to the 2D convolution specifically designed for functions on the sphere have been proposed in the recent literature. Some of these run a 2D CNN on an equirectangular projection of the spherical signal, but modify the kernel (Su & Grauman, 2019), the sampling pattern of the kernel (Coors et al., 2018), or run the CNN on projections to different tangent planes in order to minimize the effect of the distortion (Eder et al., 2020). Other approaches discretize the sphere, where the vertices are intepreted as nodes in a graph, allowing them to apply the machinery of Laplacian-based graph convolutions (Defferrard et al., 2016; 2019) to the problem.

One of the main properties that make regular CNNs so successful is their equivariance to translations. In fact, it is known from abstract algebra that a convolution is deeply tied to the concept of equivariance (Kondor & Trivedi, 2018). This insight was exploited by Esteves et al. (2018) and Cohen et al. (2018), where they define network layers equivariant to $\mathrm{SO}(3)$ rotations. This is accomplished by transforming the input using the spherical Fourier transform (SFT) and then performing convolutions in the spectral domain. These networks have the appealing property of resting firmly on the abstract algebraic underpinnings of convolution.

However, some spherical data comes equipped with a natural orientation. For instance, imagery captured from a self-driving vehicle can be aligned with the gravity direction using information from an inertial measurement unit (IMU). Brain scans and planetary data also come with a preferred orientation, and it has been shown that employing convolutions that utilize this information (and thus are not $\mathrm{SO}(3)$ equivariant), can lead to improved performance in classification and segmentation tasks (Jiang et al., 2019). Instead, what is required in these applications is a convolution which

is equivariant to $\mathrm{SO}(2)$ transformations, i.e., azimuthal rotations, of the input data, but where the requirement of full $\mathrm{SO}(3)$ equivariance may be unnecessarily restrictive.

In this paper, we follow the line of work that examines spherical convolutions in the spectral domain (Kondor et al., 2018; Esteves et al., 2018; Cohen et al., 2018), and consider the problem of how to design $\mathrm{SO}(2)$ equivariant convolutions. Specifically, we make the following contributions:

- Our main contribution is theoretical: We perform a complete characterization of all possible azimuthal-rotation equivariant linear operators $T : L^2(S^2) \to L^2(S^2)$ of band-limited functions. We show that these can be parametrized naturally as block diagonal matrices of complex numbers that act on the spherical harmonic expansion coefficients of their input.
- In addition to the Fourier space characterization, we also show how these operations can be interpreted as correlations between the input signal and filters in the spatial domain.
- We demonstrate that the correlations are natural generalizations of the convolutions proposed by Kondor et al. (2018) and Esteves et al. (2018), and how these, as well as those of Jiang et al. (2019) can be formulated as special cases of our framework by parametrizing the elements of the block-diagonal matrices corresponding to the operator T.
- Through experiments on two datasets, we show that by employing azimuthal-rotation equivariant correlations, we can recreate a state-of-the-art neural network architecture (Jiang et al., 2019) in our framework, allowing it to benefit from improved equivariance.

The paper is organized as follows. Sec. 2 gives some preliminaries and Sec. 3 introduces the notion of equivariance. Our main theoretical results on azimuthal equivariance and azimuthal correlations are given in Sec. 4 and Sec. 5, respectively. Sec. 6 provides an in-depth summary of related work and finally, we present experimental results and comparisons in Section 7.

2 Preliminaries

A point ω on the unit sphere S^2 can be parametrized by spherical coordinates $\omega(\theta, \phi) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$, where θ is the polar angle measured down from the z-axis and ϕ the azimuthal angle measured from the x-axis (see Fig. 2). The angle θ varies between 0 and π, while ϕ varies between 0 and 2π.

The Hilbert space $L^2(S^2)$ consists of all functions $f : S^2 \to \mathbb{C}$ which are square integrable on the unit sphere, and comes equipped with the inner product

$$\langle f, h \rangle = \int_{S^2} f \bar{h} d\omega = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} f(\theta, \phi) \overline{h(\theta, \phi)} \sin\theta d\theta d\phi.$$

The spherical harmonics are a set of functions on the sphere that form an orthogonal basis of $L^2(S^2)$. This means that a function $f \in L^2(S^2)$ can be represented in the orthogonal basis via

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{f}_{lm} Y_l^m(\theta, \phi), \tag{1}$$

where Y_l^m is the spherical harmonic of degree l and order m, and the $\hat{f}_{lm} \in \mathbb{C}$ are the (spherical) Fourier coefficients. Note that we always have $-l \leqslant m \leqslant l$. Here,

$$Y_l^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi}, \tag{2}$$

where P_l^m is the associated Legendre function of degree l and order m. For the exact definition of Legendre functions, see Driscoll & Healy (1994). For our purposes, it will be important to know that for fixed m, $\int_{\theta=0}^{\pi} P_k^m(\cos\theta) P_l^m(\cos\theta) \sin\theta d\theta = 0$ whenever $k \neq l$. Similarly, as the spherical harmonics form a complete orthogonal basis, we have $\langle Y_l^m, Y_{l'}^{m'} \rangle = \delta_{l,l'} \delta_{m,m'}$. Given f, the Fourier coefficients $\hat{f}_{lm}$ can be computed as $\hat{f}_{lm} = \langle f, Y_l^m \rangle = \int_{S^2} f \overline{Y_l^m} d\omega$.

3 Equivariance and Linear Operators

With the notation in place, we are now in a position to turn our attention towards equivariance. We will start by looking at the group of translations over reals, which will be insightful for further analysis. Then, we will turn to functions on S^2 and the group of rotations $\mathrm{SO}(3)$.

3.1 TRANSLATIONS

For $f \in L^2(\mathbb{R}^n)$ and $t \in \mathbb{R}^n$, the *translation* of f by t is the function $\Lambda_t f(x)$ given by $\Lambda_t f(x) = f(x-t)$. Clearly, $||\Lambda_t f||_2 = ||f||_2$ for all $f \in L^2(\mathbb{R}^n)$. So, Λ_t is a bounded and linear operator from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$.

We say that a linear operator T is *equivariant* with respect to translations if, for all $t \in \mathbb{R}^n$, we have

$$\Lambda_t T = T \Lambda_t, \tag{3}$$

which means that the linear operator T commutes with translation Λ_t.

For $h \in L^2(\mathbb{R}^n)$, let T_h be the operator performing *convolution* of $f \in L^2(\mathbb{R}^n)$ with h, defined by the function

$$T_h f(x) = (h * f)(x) = \int_{\mathbb{R}^n} h(y) f(x-y) dy, \;\; \forall x \in \mathbb{R}^n. \tag{4}$$

This operator is equivariant with respect to translations. Indeed, for $t \in \mathbb{R}^n$,

$$T_h \Lambda_t f(x) = \int_{\mathbb{R}^n} h(y) f(x-y-t) dy = (h * f)(x-t) = T_h f(x-t) = \Lambda_t T_h f(x). \tag{5}$$

It turns out that Eq. (4) is the most general form of a translation equivariant operator for functions on $\mathbb{R}^n$. If T is a translation equivariant operator, then there exists a function h such that $Tf = h * f$ for all f, provided one is allowed to pick h from the (larger) function space that also includes distributions. This is a classical result in functional analysis, see Hörmander (1960) for a proof.

3.2 ROTATIONS

Now, let us consider a real-valued function f defined on S^2. Let $R \in \mathrm{SO}(3)$ be a rotation. Then we define Λ_R as the operator that rotates the graph of the function $f : S^2 \to \mathbb{R}$ by R. Specifically, $\Lambda_R f(\omega) = f(R^{-1}\omega)$.

An operator T is *equivariant* with respect to rotations, if for all $R \in \mathrm{SO}(3)$, we have

$$T \Lambda_R = \Lambda_R T,$$

i.e., if it commutes with the rotation operator.

The characterization of all rotation equivariant operators is well-known in the literature. In Dai & Xu (2013), it is shown that a linear, bounded operator $T : L^2(S^2) \to L^2(S^2)$ is equivariant to rotations if and only if there exists a sequence of numbers $\{\hat{\mu}_l\}$, $l = 0, 1, \ldots$, such that

$$(\widehat{Tf})_{lm} = \hat{\mu}_l \hat{f}_{lm} \tag{6}$$

is fulfilled for all Fourier coefficients $\hat{f}_{lm}$ indexed by l and m of an arbitrary function $f \in L^2(S^2)$. Such linear operators T can also be interpretated as convolutions. In Driscoll & Healy (1994), spherical convolution is defined as

$$(h * f)(\omega) = \int_{R \in \mathrm{SO}(3)} h(R\eta) f(R^{-1}\omega) dR. \tag{7}$$

Here, ω is any point on the sphere and η the north pole. The measure is $dR = \sin\theta \, d\theta \, d\phi \, d\psi$ in terms of Euler angle coordinates of a rotation. It is also proven that the Fourier transform of the convolution is given by

$$(\widehat{h * f})_{lm} = 2\pi \sqrt{\frac{4\pi}{(2l+1)}} \sum_{l=0}^{\infty} \hat{h}_{l0} \hat{f}_{lm}. \tag{8}$$

Note that only the coefficients with order $m = 0$ for h are present in the formula. Filters, which only have terms of order $m = 0$ in their Fourier expansion, are known as *zonal filters* and consequently they are constant for fixed θ when varying ϕ. Also note that the convolution fulfills (6), so it is indeed a rotation equivariant operation. Conversely, it is also clear from (6) that every bounded linear operator T can be associated with such a filter h.

4 AZIMUTHAL-ROTATION EQUIVARIANT LINEAR OPERATORS

At last, we are ready to perform our analysis of azimuthal-rotation equivariant linear operators. Let $T : L^2(S^2) \to L^2(S^2)$ be a linear operator acting on functions on the unit sphere. What conditions must be imposed on T in order for it to have equivariance to rotations around the z-axis? Formally we should have, for all such azimuthal rotations $R_\phi \in \mathrm{SO}(3)$, that $T\Lambda_{R_\phi} = \Lambda_{R_\phi} T$.

Consider the action of T on an arbitrary function $f \in L^2(S^2)$. Let $\hat{f}_{lm}$ be its expansion coefficients according to (1). Applying T to this function gives, due to the linearity of T,

$$Tf(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{n} \hat{f}_{lm} T Y_l^m(\theta, \phi). \tag{9}$$

The action of T is thus fully determined by its action on the spherical harmonic basis functions. In particular, it will be equivariant if and only if the spherical harmonic basis functions are equivariant.

Let $g_{lm}(\theta, \phi) = TY_l^m(\theta, \phi)$ be the action of T on Y_l^m. Since $g_{lm} \in L^2(S^2)$, we can expand it,

$$g_{lm}(\theta, \phi) = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \hat{g}_{l'm'}^{lm} Y_{l'}^{m'}(\theta, \phi). \tag{10}$$

The requirement that T be equivariant then gives, for all azimuthal angles ϕ_0,

$$T\left[Y_l^m(\theta, \phi - \phi_0)\right] = g_{lm}(\theta, \phi - \phi_0). \tag{11}$$

The left hand side becomes, after using the definition of the spherical harmonic Y_l^m in (2) and moving the constant phase $e^{-im\phi_0}$ outside T and using the definition of g_{lm} and its expansion (10):

$$T\left[Y_l^m(\theta, \phi - \phi_0)\right] = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \hat{g}_{l'm'}^{lm} e^{-im\phi_0} Y_{l'}^{m'}(\theta, \phi). \tag{12}$$

The right hand side of (11) is simply $g_{lm}(\theta, \phi - \phi_0) = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \hat{g}_{l'm'}^{lm} e^{-im'\phi_0} Y_{l'}^{m'}(\theta, \phi)$.

So the azimuthal equivariance condition (11) requires that the above two expressions are equal, i.e.,

$$\sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \hat{g}_{l'm'}^{lm} e^{-im\phi_0} Y_{l'}^{m'}(\theta, \phi) = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \hat{g}_{l'm'}^{lm} e^{-im'\phi_0} Y_{l'}^{m'}(\theta, \phi). \tag{13}$$

Now, two functions in $L^2(S^2)$ are identical if and only if their coefficients in the basis are identical. This gives us the constraint $\hat{g}_{l'm'}^{lm} e^{-im\phi_0} = \hat{g}_{l'm'}^{lm} e^{-im'\phi_0}$. Thus, for any non-zero coefficient $\hat{g}_{l'm'}^{lm}$, we must have that $m = m'$. The expansion (10) of g_{lm} must therefore have the form

$$g_{lm}(\theta, \phi) = \sum_{l'=|m|}^{\infty} \hat{g}_{l'm}^{lm} Y_{l'}^{m}(\theta, \phi). \tag{14}$$

The inner sum has vanished, since the terms are only non-zero for $m = m'$, and the sum starts at $l' = |m|$ since no terms exist in (10) for $l' < |m|$. The coefficients $\hat{g}_{l'm}^{lm}$ thus parametrize the azimuthal rotationally equivariant transform T. To summarize, we have proven the following.

Proposition 1. *A linear, bounded operator $T : L^2(S^2) \to L^2(S^2)$ is equivariant to azimuthal rotations if and only if there exists a sequence of numbers $\{\hat{\mu}_{l'}^{lm}\}$ such that*

$$(\widehat{Tf})_{lm} = \sum_{l'=|m|}^{\infty} \hat{\mu}_{l'}^{lm} \hat{f}_{l'm} \tag{15}$$

is fulfilled for all Fourier coefficients $\hat{f}_{lm}$ indexed by l, m of an arbitrary function $f \in L^2(S^2)$.

Note that T maps a spherical harmonic Y_l^m to a linear combination of spherical harmonics with the same order m. Thus, another way of stating the proposition is that a bounded linear operator

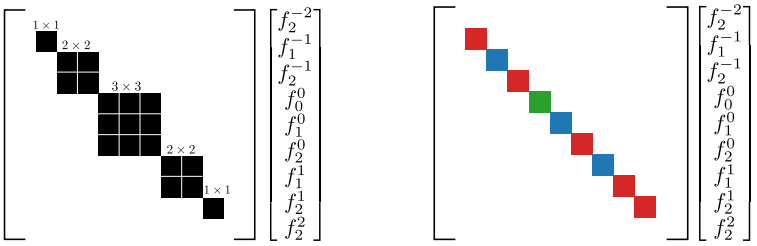


Figure 1: *Left*: The matrix representation of a general linear $\mathrm{SO}(2)$ equivariant operator T acting on the spherical harmonic expansion coefficients of a function $f \in L^2(S^2)$. Note that the coefficients have been grouped by their order (superscript). *Right:* The matrix representation of an $\mathrm{SO}(3)$ equivariant linear operator. The matrix elements shown in the same color are identical, since all coefficients of the same degree are multiplied by the same number.

$T : L^2(S^2) \to L^2(S^2)$ is equivariant to azimuthal rotations if and only if the subspaces spanned by the spherical harmonics of a fixed degree form invariant subspaces under the action of T.

This result has a natural interpretation if we restrict ourselves to spherical harmonics of some maximum degree L, i.e., when we are working with band-limited signals. This occurs commonly in practice, e.g., with a signal obtained through an equiangular sampling on the sphere. If the expansion coefficients are listed in a vector and grouped by m-values, then the matrix representation of T will be a block diagonal matrix (cf. Fig. 1).

The $\mathrm{SO}(3)$ equivariant convolution as defined by Driscoll and Healy, cf. (7), with the transform formula given in (8) is a special case of the above proposition where the matrix T is diagonal and where the coefficients for all spherical harmonics of the same degree are identical, cf. Fig. 1.

5 Azimuthal convolutions and correlations on S^2

Translation and rotation equivariant operations can be realized as convolutions, or equivalently, correlations. Previously we characterized azimuthal-rotation equivariant linear operators in the Fourier domain (Proposition 1). A natural question arises: Can we interpret these operators as convolutions or correlations? Below, we show that these operators can be interpreted in terms of a correlation.

Let $h, f : S^2 \to \mathbb{R}$. Then, a first attempt at a definition of correlation is $(h \star f)(\phi) = \int_{\omega \in S^2} h(R_\phi^{-1}\omega) f(\omega) d\omega$. However, such a correlation is only defined on S^1, since it is a function of only the azimuthal angle ϕ. Instead, we extend the domain to S^2 by considering a filter parametrized by the polar angle θ, i.e., $h = h_\theta$. So, by varying θ, we get a different filter and consequently a different correlation response.

Definition 1. *Let $h_\theta, f : S^2 \to \mathbb{R}$. Then, we define the* **azimuthal correlation** *as*

$$(h_\theta \star f)(\theta, \phi) = \int_{\omega \in S^2} h_\theta(R_\phi^{-1}\omega) f(\omega) d\omega.$$

Let $\hat{f}_{lm}$ and $\hat{h}^\theta_{lm}$ be the Fourier expansion coefficients of f and h_θ, Further, we will expand each coefficient $\hat{h}^\theta_{lm}$ in the associated Legendre basis,

$$\hat{h}^\theta_{lm} = \sum_{k=|m|}^{\infty} \hat{h}_{klm} P_k^m(\cos\theta). \tag{16}$$

Now we can show that every azimuthal-rotation equivariant linear operator can be expressed as a correlation. See Fig. 2 for an illustration.

Proposition 2. *For functions h_θ, f in $L^2(S^2)$, the transform of the correlation is given by*

$$\widehat{(h_\theta \star f)}_{lm} = (-1)^m \sqrt{\frac{4\pi(l+m)!}{(2l+1)(l-m)!}} \sum_{l'=|m|}^{\infty} \hat{h}_{ll'm} \hat{f}_{l'm}, \tag{17}$$

where f and h_θ are expanded in Fourier and Legendre series according to (16).

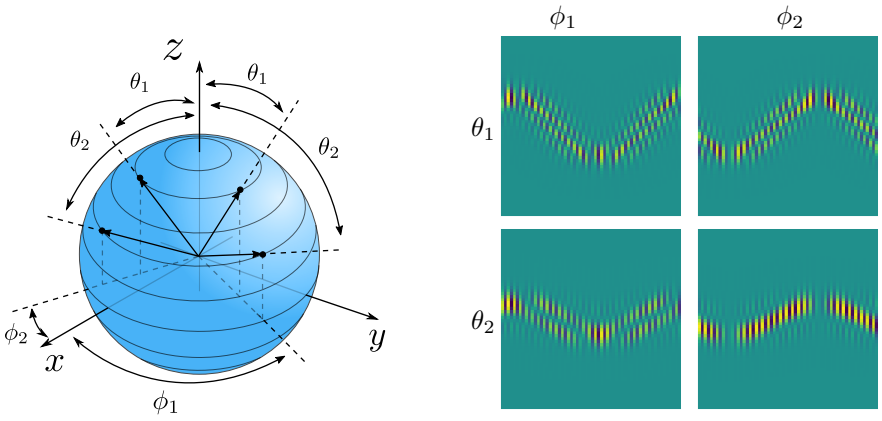


Figure 2: Correlation interpretation of an $\mathrm{SO}(2)$ equivariant transformation. To the left, four points are shown, positioned at two azimuths ϕ_1 and ϕ_2, and polar angles θ_1 and θ_2. To compute the response at a point (θ, ϕ), the polar-dependent filter h_θ is rotated around the z-axis by ϕ, as shown in (b), before being multiplied with the input signal f and integrated over the sphere. This is similar to regular correlation in $\mathbb{R}^2$, except that changing θ does not translate the filter, but rather morphs the filter into a new one. The operator shown is the partial derivative $\partial/\partial\phi$.

The proof is straight-forward by substituting the expansions and simplifying, see Appendix.

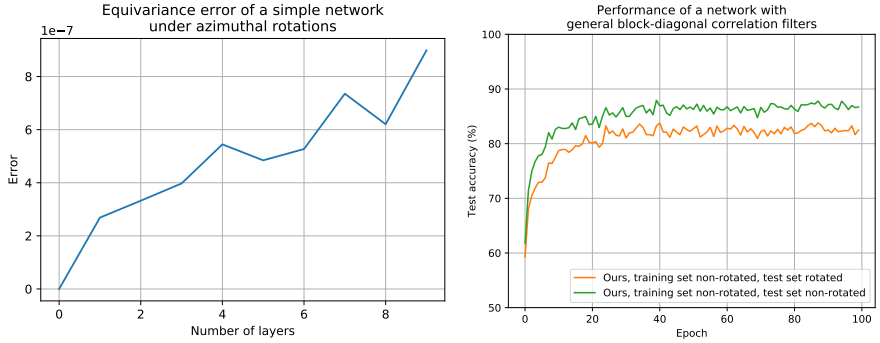
6 Comparison of Our Results With the Literature

Our work is closely related to, and inspired by, recent theoretical developments on spherical CNNs and $\mathrm{SO}(3)$ equivariance, including Cohen et al. (2018); Esteves et al. (2018); Kondor & Trivedi (2018); Defferrard et al. (2019). See also Esteves (2020) for a more comprehensive review. However, these works do not consider the case of purely azimuthal equivariance, and the characterization we present in Proposition 1 is, to our knowledge, new. Our second result, Proposition 2 which shows that azimuthal equivariant operators can be expressed as correlations is by no means a surprise, but it does not follow from, for instance, the general framework of Kondor & Trivedi (2018).

Another closely related work is that of Jiang et al. (2019). They present an efficient and powerful framework for analyzing spherical images with state-of-the-art performance. It is based on learning filters that are linear combinations differential operators, or more specifically, $\frac{\partial}{\partial\theta}$, $\frac{\partial}{\partial\phi}$ and the Laplacian ∇^2. The resulting filters are not $\mathrm{SO}(3)$ equivariant, but they are indeed linear and azimuthal equivariant operators (and thus azimuthal correlations, cf. Proposition 2) and consequently a special case of our framework. However, the discretization of the sphere into an icosahedral mesh at lower resolutions breaks the equivariance in Jiang et al. (2019). This hurts the generalization performance as confirmed by our experiments. This limitation can be alleviated by applying fully azimuthal-rotation equivariant operations.

7 Experiments

In this section we implement neural networks based on the presented framework. First, we verify that a correlation layer $L^2(S^2) \to L^2(S^2)$ designed according to Proposition 1 is indeed equivariant to azimuthal rotations. Then, as in Cohen et al. (2018); Jiang et al. (2019), we perform experiments on the Omni-MNIST and ModelNet40 (Wu et al., 2015) datasets, where the task is to perform classification of spherical images and 3D shape models, respectively. We demonstrate that the state-of-the-art architecture of Jiang et al. (2019) can be recreated in our framework, and that the classification results of both approaches on unrotated data are virtually identical, but with our approach showing increased generalization performance to unseen $\mathrm{SO}(2)$ rotations on the test set of ModelNet40.



(a) (b)

Figure 3: (a) The equivariance error Δ as a function of the number of consecutive correlation and ReLU layers in the network Φ. (b) Accuracy during training of the network with general block-diagonal correlation filters. The accuracy is plotted both for the original test set, as well as for a rotated version of the test set, where each image has been randomly rotated about the z-axis.

7.1 Equivariance error

To evaluate the equivariance error in our network layers, we follow the approach of Cohen et al. (2018). We create prototype networks Φ by composing randomly initialized differential type correlation layers and ReLU. We sample $n = 500$ azimuthal rotations R_i and n signals f_i with 12 channels each. Next we compute the equivariance error $\Delta = \frac{1}{n}\sum_{i=1}^{n} \mathrm{std}(\Lambda_{R_i}\Phi(f_i) - \Phi(\Lambda_{R_i} f_i))/\mathrm{std}(\Phi(f_i))$ which should be zero for a perfectly equivariant network. The obtained equivariance errors are presented in Fig. 3a. The low order of the error and that the error does not increase much with the number of layers indicates that the correlations are azimuthally equivariant.

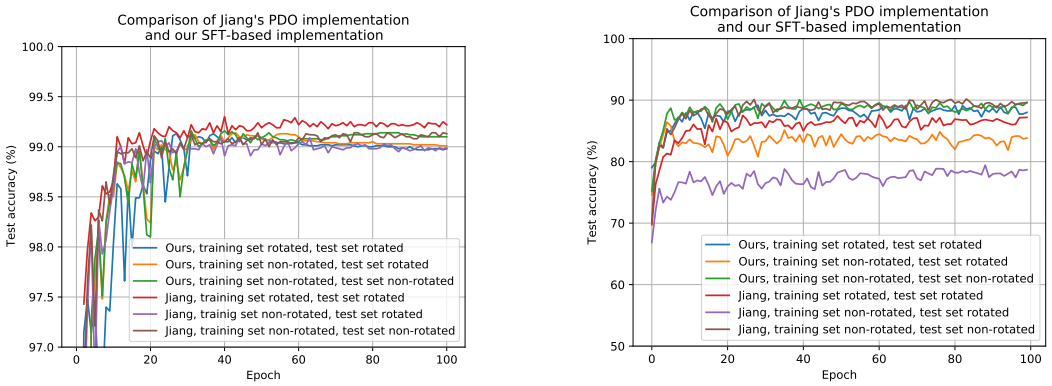
7.2 Digit classification on Omni-MNIST

The Omni-MNIST dataset is obtained by projecting the images of the MNIST handwritten digit dataset onto the surface of a sphere. The resulting dataset consists of 60,000 spherical images for training, and 10,000 for testing. As in Jiang et al. (2019), the images are projected onto the north pole, and then rotated to the equator.

We run the method of Jiang et al. (2019) as a reference, using their implementation, and also recreate their architecture in our SFT-based framework. The network consists of a stack of residual blocks. Each block contains 1×1 convolutions that mix feature maps, as well as spherical convolutional layers that process each feature map in parallel by forming a linear combination of the image, its horizontal and vertical derivatives, and its Laplacian. Note that since all of these operations are $\mathrm{SO}(2)$ equivariant, they can each be represented by a block-diagonal matrix acting on the spherical harmonic expansion coefficients of the feature maps. Batch-normalization and ReLU are also performed within the residual blocks. Please refer to the paper of Jiang et al. (2019) for more details about the network architecture.

Downsampling is performed in the spectral domain, where we simply reduce the maximum degree L of included SFT coefficients by a factor of two, rounding up.

Both our implementation and the original implementation by Jiang et al. (2019) are trained with stochastic gradient descent, using the Adam optimizer and a batch size of 16. The initial learning rate is 0.01, and is decreased by a factor of two every ten epochs. For each epoch during training, we compute the accuracy on the test set, using both the original version, and a randomly azimuthally rotated version of the test set. The results are shown in Fig. 4a. The training is also run on an unrotated and a randomly azimuthally rotated version of the test set.



(a) Omni-MNIST

(b) Modelnet40

Figure 4: Accuracy during training of the discretized sphere implementation of Jiang et al. (2019) and our SFT-based implementation. The accuracy is plotted both for the original test set, as well as for a rotated version of the test set, where each image has been randomly rotated about the z-axis.

7.3 3D SHAPE CLASSIFICATION ON MODELNET40

ModelNet40 (Wu et al., 2015) is a dataset containing CAD models of 40 different object classes. We run our experiments on the aligned version of the dataset provided by Sedaghat et al. (2017) and follow Cohen et al. (2018) by creating spherical images by ray casting the CAD models onto an enclosing sphere. Each model is represented by six channels – the ray length, the sine of the surface angle and the cosine of the surface angle, as well as the same three properties obtained from ray casting the convex hull of the model.

We compare to the implementation by Jiang et al. (2019). We recreate their network architecture (c.f. Sec. 7.2) and train using stochastic gradient descent with Adam and a batch size of 16. The learning rate is set to $5 \cdot 10^{-3}$ and decreased by a factor of 0.7 every 25 epochs. Again we evaluate and train the model both on the aligned ModelNet40 dataset as well as a version of the dataset with each CAD model rotated by a random azimuthal rotation. We train for 100 epochs and show the results in Fig. 4b. Of special note is the case with a non-rotated training set and a rotated test set, where we see that our network manages to generalize better than the one by Jiang et al. (2019). The reason for the low degree of generalization from non-rotated to rotated data in the baseline model might be the hierarchical nature of the icosahedral grid and down-sampling used by Jiang et al. (2019), where a grid point on the sphere will influence the output more if it belongs to a lower level in the hierarchy. If an important point in the input is rotated to a grid point in a higher hierarchy level, the output might change drastically. This problem is not present in our design since we downsample in the frequency domain and therefore the information from every grid point is treated equally.

Additionally, we created another network using the same architecture as before, but with general block-diagonal correlation filters (instead of learning a linear combination of differential filters). The performance results of this network are shown in Fig. 3b. The accuracy of the general network is comparable to the accuracy of the networks with differential feature maps in Fig. 4b, but it should be noted that the number of trainable parameters is much larger.

8 CONCLUSIONS

In this paper, we have examined how to design convolutions which operate on spherical data and are equivariant to $\mathrm{SO}(2)$ rotations. Specifically, we have performed a complete characterization of bounded, linear operators from $L^2(S^2) \to L^2(S^2)$ which exhibit $\mathrm{SO}(2)$ equivariance. We showed that, for band-limited signals, these can be realized as block-diagonal matrices in the spectral domain, and also demonstrated how these operators may be interpreted as correlations in the spatial domain. Using this framework, we implemented an existing state-of-the-art pipeline, which in our framework showed better generalization performance to $\mathrm{SO}(2)$ rotations not seen during training.

REFERENCES

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *Int. Conf. on Learning Representations*, 2018.

Benjamin Coors, Alexandru Paul Condurache, and Andreas Geiger. SphereNet: Learning spherical representations for detection and classification in omnidirectional images. In *European Conf. on Computer Vision*, 2018.

Feng Dai and Yuan Xu. *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer, 2013.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, 2016.

Michaël Defferrard, Martino Milani, Frédérick Gusset, and Nathanaël Perraudin. DeepSphere: a graph-based spherical CNN. In *Int. Conf. on Learning Representations*, 2019.

James R. Driscoll and Dennis M. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.

Marc Eder, Mykhailo Shvets, John Lim, and Jan-Michael Frahm. Tangent images for mitigating spherical distortion. In *Conf. Computer Vision and Pattern Recognition*, 2020.

Carlos Esteves. Theoretical aspects of group equivariant neural networks. *arXiv preprint arXiv:2004.05154*, 2020.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning SO(3) equivariant representations with spherical CNNs. In *European Conf. on Computer Vision*, 2018.

Lars Hörmander. Estimates for translation invariant operators in L^p spaces. *Acta Math.*, 104(1-2): 93–140, 1960.

Chiyu Jiang, Jingwei Huang, Karthik Kashinath, Philip Marcus, Matthias Niessner, et al. Spherical CNNs on unstructured grids. In *Int. Conf. on Learning Representations*, 2019.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Int. Conf. on Machine Learning*, 2018.

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–Gordan nets: A fully fourier space spherical convolutional neural network. In *Advances in Neural Information Processing Systems*, 2018.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Hidenobu Matsuki, Lukas von Stumberg, Vladyslav Usenko, Jörg Stückler, and Daniel Cremers. Omnidirectional dso: Direct sparse odometry with fisheye cameras. *IEEE Robotics and Automation Letters*, 3(4):3693–3700, 2018.

Mayur Mudigonda, Sookyung Kim, Ankur Mahesh, Samira Kahou, Karthik Kashinath, Dean Williams, Vincen Michalski, Travis O'Brien, and Mr Prabhat. Segmenting and tracking extreme climate events using neural networks. In *Deep Learning for Physical Sciences (DLPS) Workshop, held with NIPS Conference*, 2017.

Nima Sedaghat, Mohammadreza Zolfaghari, Ehsan Amiri, and Thomas Brox. Orientation-boosted voxel nets for 3D object recognition. In *British Machine Vision Conf.*, 2017.

Yu-Chuan Su and Kristen Grauman. Kernel transformer networks for compact spherical convolution. In *Conf. Computer Vision and Pattern Recognition*, 2019.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A deep representation for volumetric shapes. In *Conf. Computer Vision and Pattern Recognition*, 2015.

APPENDIX A: PROOF OF PROPOSITION 2

Proof. From the definition of correlation we obtain

$$(h_\theta \star f)(\theta,\phi) = \int_{\omega\in S^2} h_\theta(R_\phi^{-1}\omega) f(\omega) d\omega = \int_{\theta'=0}^{\pi}\int_{\phi'=0}^{2\pi} h_\theta(\theta',\phi') f(\theta',\phi+\phi')\sin\theta' d\theta' d\phi'. \tag{18}$$

Substituting in the Fourier expansions of f and h, and employing the orthogonality of the spherical harmonics, we get

$$(h_\theta \star f)(\theta,\phi) = \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \hat{h}^{\theta}_{l'm'}\hat{f}_{l'm'}e^{im'\phi}. \tag{19}$$

Now, employing the expansion (16) of $\hat{h}^{\theta}_{l'm'}$ into the basis of the associated Legendre functions, we find

$$(h_\theta \star f)(\theta,\phi) = \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\sum_{k=|m'|}^{\infty} \hat{h}_{kl'm'}P_k^{m'}(\cos\theta)\hat{f}_{l'm'}e^{im'\phi}. \tag{20}$$

By reindexing this last sum, we can instead write it in the form

$$(h_\theta \star f)(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=|m|}^{\infty} \hat{h}_{ll'm}\hat{f}_{l'm}P_l^m(\cos\theta)e^{im\phi} = \tag{21}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left[(-1)^m\sqrt{\frac{4\pi(l+m)!}{(2l+1)(l-m)!}}\sum_{l'=|m|}^{\infty}\hat{h}_{ll'm}\hat{f}_{l'm}\right]Y_l^m(\theta,\phi), \tag{22}$$

where, in the last equality, we have used the definition (2) of the spherical harmonics. Note that the last expression is a spherical harmonic expansion, and the expression in the brackets is the corresponding expansion coefficient, given by

$$\widehat{(h_\theta \star f)}_{lm} = (-1)^m\sqrt{\frac{4\pi(l+m)!}{(2l+1)(l-m)!}}\sum_{l'=|m|}^{\infty}\hat{h}_{ll'm}\hat{f}_{l'm}. \tag{23}$$

This proves the proposition. □