Möbius Convolutions for Spherical CNNs

Thomas W. Mitchel
tmitchel@jhu.edu
Johns Hopkins University
USA

Noam Aigerman
Vladimir G. Kim
aigerman@adobe.com
vokim@adobe.com
Adobe Research
USA

Michael Kazhdan
misha@cs.jhu.edu
Johns Hopkins University
USA

Figure 1: Equivariant convolutions require two ingredients: a frame operator and a density operator. Filters assign weights based on the relative positions of points and the frame operator $T$ corrects for the deformation of the local “tangent space” under a Möbius transformation $g$. Similarly, the density operator $\rho$ adjusts for the change in the area measure used for integration, proportional to the conformal scale factor $\lambda_g^2$.

ABSTRACT

Möbius transformations play an important role in both geometry and spherical image processing – they are the group of conformal automorphisms of 2D surfaces and the spherical equivalent of homographies. Here we present a novel, Möbius-equivariant spherical convolution operator which we call Möbius convolution; with it, we develop the foundations for Möbius-equivariant spherical CNNs. Our approach is based on the following observation: to achieve equivariance, we only need to consider the lower-dimensional subgroup which transforms the positions of points as seen in the frames of their neighbors. To efficiently compute Möbius convolutions at scale we derive an approximation of the action of the transformations on spherical filters, allowing us to compute our convolutions in the spectral domain with the fast Spherical Harmonic Transform. The resulting framework is flexible and descriptive, and we demonstrate its utility by achieving promising results in both shape classification and image segmentation tasks.

CCS CONCEPTS

• Computing methodologies → Neural networks; Shape analysis; Image processing.

KEYWORDS

Neural networks, Group equivariance, Möbius transformations, Conformal transformations, Convolution

1 INTRODUCTION

Convolutional neural networks (CNNs) are effective because convolution responds to a contextualized window, forcing the learning to be translation-equivariant. However, vanilla CNNs assume a fixed coordinate frame and lose effectiveness in the presence of deformations that change the frame. This has lead to the development of more general notions of convolution equivariant to transformation...
groups including rotations [Cohen and Welling 2016; Worrall et al. 2017] and dilations [Finzi et al. 2020; Sosnovik et al. 2019; Worrall and Welling 2019]. Critically, the notion of rotation-equivariance has facilitated the generalization of CNNs to domains without a canonical orientation at each point such as the sphere [Cohen et al. 2019b, 2018; Esteves et al. 2020] and arbitrary surfaces [de Haan et al. 2020; Mitchel et al. 2021; Wiersma et al. 2020]. The resulting networks are isometry-equivariant – able to repeatably characterize local features in the presence of distance-preserving transformations – and have excelled in fundamental geometry processing tasks such as shape classification, segmentation, and correspondence.

Despite their success, rotation- and isometry-equivariant CNNs can fail to achieve adequate performance in the presence of the kinds of complex deformations commonly found in real-world image and shape data [Mitchel et al. 2021]. Such deformations may potentially be better modeled by higher-dimensional transformation groups. For example, homographies (projective transformations) better approximate changes in camera viewpoints than similarities (rotations and dilations) [Hartley and Zisserman 2003] and, for spherical images, can be represented using conformal transformations [El Mir et al. 2014; Schlemier and Segerman 2016]. For geometry processing, conformal (angle-preserving) transformations encompass a broader class of deformations than isometries that still preserve the sense of "shape". [Crane et al. 2011; Gu et al. 2004; Lévy et al. 2002].

We present a novel spherical convolution operator, equivariant to Möbius transformations, which we call Möbius convolution (MC). Using this operator, we develop the foundations for Möbius-equivariant spherical CNNs. Our convolutions are flexible, and we demonstrate the utility of our Möbius-equivariant CNNs by achieving promising results on standard benchmarks in both genus-zero shape classification and spherical image segmentation.

Our approach is based on a key observation: while the Möbius group is six-dimensional, its action on the characterization of the position of a point relative to its neighbor can be described by a four-dimensional subgroup. By defining an equivariant frame operator (Figure 1, left) at each point with which we align the filter, we correct for the change in the relative positions induced by this subgroup. To compute convolutions, input features are mapped to a density distribution (Figure 1, right), controlling for the change in area measure, and integrated against the aligned filters over the sphere, rather than the group itself. To facilitate efficient evaluations, we parameterize filters using log-polar basis functions from which we derive an approximation of the action of the frames, allowing us to compute our convolutions via the Fast Spherical Harmonic Transform [Driscoll and Healy 1994; Kostelec and Rockmore 2008]. Our implementation is publicly available at github.com/twmitchel/MobiusConv.

2 RELATED WORK

Group-equivariant CNNs were first introduced by [Cohen and Welling 2016; Cohen et al. 2019a], wherein kernels are parameterized in terms of equivariant basis functions on the group itself. Convolution is performed by lifting features from the domain and searching over all possible transformations of the kernels. This becomes intractable for non-compact groups where the domains of integration are unbounded and the representations are infinite-dimensional, though recent work by [Finzi et al. 2020] mitigates these problems by considering only the origin-preserving subgroups and integrating with respect to an equivariant Monte Carlo estimator to facilitate evaluations. This approach is flexible, and has since been extended to handle general affine transformations and homographies [MacDonald et al. 2021]. In tandem, several fully-connected networks have been developed that achieve equivariance to non-compact transformation groups including the Lorentz group [Bogatskiy et al. 2020], the Poincaré group [Villar et al. 2021], and the symplectic group [Finzi et al. 2021] – all closely related to Möbius transformations.

Equivariant CNNs that integrate over the domain on which the group acts generally parameterize kernels in terms of basis functions that rotate or dilate with the local coordinate system [Sosnovik et al. 2019; Weiler and Cesa 2019; Worrall et al. 2017; Worrall and Welling 2019]. This approach has been extended to volumetric domains [Weiler and Cesa 2019], point clouds [Qi et al. 2017], and canonical domains such as the sphere [Cohen et al. 2018; Esteves et al. 2020]. However, finite-dimensional equivariant bases often don’t exist for non-commutative and non-compact transformation groups of interest, precluding generalizations to groups that act projectively.

Equivariance is a necessary condition for the transposition of convolutional frameworks to domains without canonical coordinate systems such as arbitrary 2D surfaces. Existing surface networks can be broadly categorized into several emerging paradigms: Representational methods [Hanocka et al. 2019; Lahav and Tal 2020] exploit the ubiquitous representation of surfaces as triangle meshes to form operators equivariant to local similarities; Diffusive approaches [Sharp et al. 2020; Smirnov and Solomon 2021; Yi et al. 2017] formulate convolution in terms of spectral kernels and accelerate computations in a low-frequency eigenbasis; and Transporting networks [Mitchel et al. 2021; Wiersma et al. 2020] propagate tangent vector features that transform with local coordinate systems. Despite their success in a variety of tasks, existing surface networks are only repeatable up to isometries, and we believe our approach to be the first surface network equivariant to conformal transformations.

3 METHOD OVERVIEW

Our method is based on a powerful observation: instead of dealing with the six-dimensional group of Möbius transformations, we only need to consider a four-dimensional subgroup. We first review the action of Möbius transformations on the sphere, and define a notion of relative positions between points analogous to the logarithm map on surfaces, and show how the latter transforms under the former.

We then introduce Möbius convolutions, which provide a method for Möbius-equivariant spatial aggregations on the sphere. Möbius convolutions are part of the extended convolution framework [Mitchel et al. 2020a] which allows filters to adaptively transform as they shift over a manifold by defining an equivariant frame at each point. This framework forms the basis for recently proposed state-of-the-art surface descriptors [Mitchel et al. 2020b] and CNNs [Mitchel...
et al. 2021]. We facilitate an efficient discretization by parameterizing filters using log-polar basis functions from which we derive a linearized approximation of the action of the frames, allowing us to compute our convolutions via the fast spherical harmonic transform [Driscol and Healy 1994; Kostelec and Rockmore 2008].

We complete the foundations for Möbius-equivariant CNNs by introducing a conformally-equivariant normalization layer based on filter response normalization [Singh and Krishnan 2020] and we validate equivariance by direct experimental evaluation. The principle module in applications is a simple Möbius convolution ResNet (MCRResNet) block [He et al. 2016], which is self-contained and flexible. We demonstrate the utility of our framework by achieving promising results on standard benchmarks in both genus-zero shape classification and spherical image segmentation.

4 Möbius Transformations

Möbius transformations can be understood by associating the twosphere with the Riemann sphere, \( \mathbb{C} = \mathbb{C} \cup \{ \infty \} \), via the stereographic projection taking the north pole to the origin. Möbius transformations are described by the action of SL(2, C), the group of matrices in \( \mathbb{C}^{2 \times 2} \) with unit determinant, on \( \mathbb{C} \) by fractional linear transformations. That is, for any \( g = \left[ \begin{array}{cc} a & b \\ c & d \end{array} \right] \in \text{SL}(2, \mathbb{C}) \) and \( z \in \mathbb{C} \):

\[
gz = \frac{az + b}{cz + d} \tag{1}
\]

4.1 Generalizing the logarithm and exponential maps

To parameterize the sphere about a point \( z \in \mathbb{C} \), we borrow the notion of the exponential and logarithm maps from Riemannian geometry, defining the generalized logarithm of \( z \) as a rotation \( \log_z \in \text{SU}(2) \) taking \( z \) to the origin,

\[
\log_z = \frac{1}{|c| \sqrt{1 + |z|^2}} \left( \begin{array}{c} c \\ cz \end{array} \right)
\]

with \( c = \sqrt{z} \). (Though any choice of \( c \) gives a rotation taking \( z \) to the origin, the above choice ensures that the great circle going through the origin and \( z \) is mapped to the real line, enabling the use of the fast Spherical Harmonic Transform in the implementation of §6.)

Then, for any point \( y \in \mathbb{C} \), we can express the “position” of \( y \) in the frame of \( z \) as \( \log_y \in \mathbb{C} \). By analogy to Riemannian geometry, the generalized logarithm maps \( \mathbb{C} \) to the “tangent space” at \( z \). We make this explicit by denoting the image of the logarithm map at \( z \) as \( \mathbb{C}^2 \):

\[
\log_z : \mathbb{C} \rightarrow \mathbb{C}^2
\]

though formally \( \mathbb{C} \) and \( \mathbb{C}^2 \) are the same space – the Riemann sphere. Similarly, we define the generalized exponential of \( z \) as the inverse of the generalized logarithm, \( \exp_z = \log^{-1}_z : \mathbb{C}^2 \rightarrow \mathbb{C} \).

4.2 Action of the origin-preserving subgroup

We will think of convolution filters as functions defined on a canonical “tangent space” describing the weight with which a point contributes to its neighbor in terms of the position of the neighbor in the frame of that point. To design our network, we need to define a Möbius-equivariant convolution operator. To this end we need to understand how the position of a point in the frame of its neighbor changes under the action of Möbius transformations.

Describing the transformation from one coordinate frame to the other is straightforward: Beginning in \( \mathbb{C} \), we 1). Map to \( \hat{\mathbb{C}} \) by applying \( \exp_z \). Transform \( \hat{\mathbb{C}} \) by \( g \) and 3). Map back to \( \hat{\mathbb{C}}_{gz} \) using \( \log_{gz} \).

4.3 Möbius Convolutions for Spherical CNNs

As in [Mitchel et al. 2020a] we implement convolution by shifting a filter over the domain, aligning the shifted filter using a frame field, and distributing the values of a density function to neighboring points, with distribution weights given by the aligned filter.

Rather than have the user provide the transformation field and density directly, our framework derives these from an input signal in such a way so as to ensure the ensuant convolution is Möbius-equivariant. Specifically, we construct a frame operator \( \mathcal{I} \) and density operator \( \rho \) that take in a real-valued function on \( \mathbb{C} \) and return a lower-triangular frame field and a real-valued density field, respectively,

\[
\mathcal{I} : L^2(\mathbb{C}, \mathbb{R}) \rightarrow L^2(\mathbb{C}, \mathbb{R}) \quad \text{and} \quad \rho : L^2(\mathbb{C}, \mathbb{R}) \rightarrow L^2(\mathbb{C}, \mathbb{R})
\]

Given a Möbius transformation \( g \in \text{SL}(2, \mathbb{C}) \) and a point on the Riemann sphere, \( z \in \mathbb{C} \), the frame operator corrects for the deformation of the tangent space resulting from \( g \), as characterized by the origin-preserving transformation \( D_g^2 \) from Equation (3) (Figure 1, left). Similarly, the density operator adjusts for the change in the area measure used for integration, given by the conformal scale factor \( \lambda_g^2(z) \) (Figure 1, right).

Given operators \( \mathcal{I} \) and \( \rho \), the Möbius convolution of a function \( \psi \) with with a filter \( f \), both in \( L^2(\mathbb{C}, \mathbb{R}) \), can formally be expressed as the function in \( L^2(\mathbb{C}, \mathbb{R}) \) with

\[
(\psi * f)(y) = \int_{\mathbb{C}} \rho_\psi(z) [\mathcal{I}_\psi(z) f] (\log_y z) \, dz,
\]

where \( \mathcal{I}_\psi(z) f \equiv f \circ [\mathcal{I}_\psi(z)]^{-1} \) denotes the standard action of Möbius transformations on \( f \) by left shifts.

That is, to get the value at a point \( y \in \mathbb{C} \), we 1). Iterate over all neighbors \( z \); 2). Compute the position of \( y \) in the frame at \( z \); 3). Evaluate the filter at that point; and 4). Accumulate the density at \( z \) weighted by the filter value. Following [Mitchel et al. 2020a], instead of aggregating features with respect to a single frame at each point, the influence of each frame is spread across the neighbors,
making the construction robust to noise and other nuisance factors affecting the stability of the frame field.

5.1 Equivariance

Following the above discussion, $\mathcal{Z}$ and $\rho$ must satisfy certain conditions to ensure that Möbius convolutions are Möbius-equivariant; i.e. that for any function $\psi$, filter $f$, and Möbius transformation $g$, Möbius convolution commutes with the action of $g$ by left shifts,

$$g(\psi * f) = (g \psi) * f.$$  
(6)

A sufficient condition for the equation to hold is if for all $\psi \in L^2(\mathbb{C}, \mathbb{R})$ and $g \in \text{SL}(2, \mathbb{C})$, the operators $\mathcal{Z}$ and $\rho$ satisfy

$$D^g_z \mathcal{Z}_\psi(z) = \mathcal{Z}_g \psi(gz) \quad \text{and} \quad \lambda^2 y(z) \rho_{gy}(z) = \rho_g \psi(gz),$$  
(7)

for all $z \in \mathbb{C}$. The condition for $\mathcal{Z}$ follows from Equation (3), and that on $\rho$ comes from the change of variables in the integral of Equation (5). In particular these conditions are satisfied when the frame and density operators are defined as

$$\mathcal{Z}_\psi(x) \equiv \begin{pmatrix} \frac{1}{2} \nabla d \log_x |\psi|_0 & 0 \\ d \log_x |\psi|_0 & \frac{1}{2} \left( d \log_x |\psi|_0 \right)^2 \end{pmatrix},$$  
(8)

and

$$\rho_{\psi}(x) \equiv \left| d \log_x |\psi|_0 \right|^2,$$  
(9)

where $d \log_x |\psi|_0$, $\nabla d \log_x |\psi|_0 \in \mathbb{C}$ are the differential and Hessian of $\log_x \psi$ evaluated at the origin. We note that at a point where $\mathcal{Z}_\psi(x)$ is ill-defined – those at which $d \log_x |\psi|_0$ vanishes – $\rho_{\psi}(x)$ also vanishes and the point contributes nothing to the convolution.

Proof that the conditions in Equation (7) imply Möbius-equivariance, and that the operators defined in Equations (8–9) satisfy the conditions of Equation (7) is provided in the supplement.

6 DISCRETIZATION

To efficiently compute Möbius convolutions at the scale necessary to build CNNs, we develop an implementation based on the fast Spherical Harmonic Transform [Driscol and Healy 1994; Kostelec and Rockmore 2008]. We give an outline of this process below and leave the details to the supplement.

Identity Convolution with the Spherical Harmonic Transform. To simplify the calculation, we first consider a simpler non-equivariant convolution, we call an identity convolution, where we replace the frame and density operators from Equation (5) with the trivial frame field $\mathcal{Z}_\psi(z) = e$ (with $e$ the identity) and the density $\rho_{\psi}(x) = \psi(x)$,

$$\psi * e f(y) = \int_C \psi(z) f(\log z y) \ dx,$$  
(10)

using $* e$ to distinguish it from the equivariant convolution.

Assuming that $\psi$ and $f$ are $B$ band-limited functions, they can be expressed in terms of their spherical harmonic decompositions as

$$\psi = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} \Psi_{lm} Y^m_l \quad \text{and} \quad f = \sum_{l'=0}^{B-1} \sum_{m'=l'}^{l'} f_{l'm'} Y^m_{l'}.$$  
(11)

Recalling that $\log z$ is a rotation, we know that it preserves the frequency content and expand

$$f \circ \log z = \sum_{l''=0}^{B-1} \sum_{m''} \sum_{l''=0}^{B-1} f_{l''m''} D^{l''-m''}_{-l''m''} (\log z) Y^{m''}_{l''},$$

where $D^{l''-m''}_{-l''m''}$, is the Wigner-D function giving the $(l'', m'')$-th spherical harmonic coefficient of the rotation of $Y_{l''}^{m''}$. Furthermore, using the fact that in the $z = y = z$ Euler angle notation our definition of $\log z$ corresponds to a rotation described by an Euler triplet whose first entry is zero, it follows that the integral vanishes

$$0 = \int_C Y^m_l(z \cdot D^{l''-m''}_{-l''m''} (\log z)) \ dx,$$

whenever $m'' \neq m$. Thus, expanding Equation (10) we get

$$\psi * f(y) = \sum_{l=0}^{B-1} \sum_{m=\text{even}}^{l} Y^m_l \left( \sum_{m''=\text{odd}}^{l''} f_{l''m''} \Delta m''_{l''} \right) Y^m_{l''}$$  
(12)

where the value of $\Delta m''_{l''}$ is independent of $\psi$ and $f$,

$$\Delta m''_{l''} = \int_C Y^{m''}_l(z \cdot D^{l''-m''}_{-l''m''} (\log z)) \ dz.$$  
(13)

From this, we can compute identity convolutions efficiently by: 1). Computing the spherical harmonic coefficients of $\psi$ and $f$ using the fast SHT; 2). Summing the two sets of coefficients according to Equation (12) to get the coefficients of the convolution; and 3). Applying the fast inverse SHT to reconstruct the convolution.

The complexity of steps 1 and 3 are proportional to those of the fast SHT, which is $O(B^3 \log^2 B)$. However, in our implementation we compute the discrete Legendre transform via sparse matrix multiplication for a total complexity of proportional to $O(B^3 \log B)$, which we find to be more efficient on the GPU. Step 2 has complexity $O(B^4)$, and we show that this computation can be re-used in computing the full (equivariant) convolution.
Spherical Log-Polar Bases. To efficiently compute arbitrary M"obius convolutions as in Equation (5), we approximate them as sums of identity convolutions. To do so, we choose a basis for our filters that enables an approximation of the action of $\Xi_\psi$.

To this end we use linear combinations of log-polar (Fourier-Mellin) basis functions [Vilenkin 1978; Vilenkin and Klimyk 1991],

$$B_{ms}(z) \equiv \frac{|z|^{is}}{|z|^t} \left( \begin{array}{c} z \\ |z| \end{array} \right)^m$$

with $m \in \mathbb{Z}$ and $s, t \in \mathbb{R}$. These functions are localized about $z = 0$ (resp. $z = \infty$) when $t > 0$ (resp. $t < 0$), are discontinuous at $z \in \{0, \infty\}$ when $m \neq 0$ (since the argument of $z$ is not defined), and singular at $z = 0$ (resp. $z = \infty$) when $t > 0$ (resp. $t < 0$). We note that, for the purposes of integration, the singularity can be ignored when $t < 1$. Loosely, this follows from the fact that $\int_0^1 \frac{1}{x^n} x^{1-t} + c$ which is bounded at $x = 0$ whenever $t \in (0, 1)$. Noting also that $B_{ms}(z) = B_{-m,-s}(1/z)$, it follows that the functions are continuous away from $\{0, \infty\}$ and, for the purposes of integration, singularities at $\{0, \infty\}$ can be ignored when $|t| < 1$.

The first several basis functions at integer frequencies $|m| \leq 2$ and $|s| \leq 3$, with $t = 0.15$ are shown in Figure 2. The complex-valued functions are visualized using the HSV scale: hue and value are determined by the arguments and magnitudes of the function values, and saturation is fixed at one.

Using these as basis functions, we consider filters in the span

$$f = \sum_{m=-M}^M \sum_{s=-N}^N b_{ms} B_{ms}$$

where $s$ is constrained to be an integer and $t$ is positive so as to localize the filter about the origin. As our filters are real-valued we have $B_{-m-n} = b_{mn}$. giving $(2M+1)(2N+1)$ real parameters.

Approximating the Transformation of Filters. Unfortunately, it is not the case that the space of band-limited filters spanned by the $B_{ms}$ is fixed under the action of the lower-triangular subgroup. This is because, in general, for non-compact, non-commutative groups, a space of functions fixed under the action of the group (i.e. a representation) will be infinite-dimensional.

However, we show in the supplement that, given a filter as in Equation (15) and a lower-triangular matrix $L \in \mathbb{L}$, the transformation of the filter $f$ by $L$ can be expanded as

$$Lf = \sum_{m=-M}^M \int_{-\infty}^{\infty} \sum_{j=1}^3 \hat{j}^{+\sigma_j}(L, b) B_{ms}^\sigma_j ds,$$

where $b$ is the $(2M+1) \times (2N+1)$-dimensional vector of coefficients of $f$, $m$ is now summed over all integers, $s$ is continuous and integrated over the real line, $\sigma_1$, $\sigma_2$, and $\sigma_3$ (the localization values) are any real values satisfying $t < \sigma_1 < 2, t - 1 < \sigma_2 < 0,$ and $\sigma_3 = t$, and $\hat{j}^{+\sigma_j}$ are functions taking a lower-triangular matrix and a set of filter coefficients, and returning the coefficient of $B_{ms}^\sigma_j$ in the expansion of the transformed filter. Here, the integral is equivalent to an inverse Mellin transform with frequency variable $s$, and the bounds on $\sigma_1$ and $\sigma_2$ are necessary to ensure invertibility [Vilenkin 1978; Vilenkin and Klimyk 1991].

Obviously, the infinite summation and the integration in Equation (16) make evaluation unfeasible. We propose a practical implementation by truncating the summation over the angular frequency $m$, and replacing the integration over the real line with a discrete approximation using quadrature. The summation is a result of the addition theorem for Bessel functions [Watson 1958] which appear in the derivation of $\hat{j}^{+\sigma_j}$, it converges rapidly at low frequencies and can be well-approximated with only several terms [Chirikjian and Kyatkin 2016]. The use of quadrature is motivated by the observation that for a fixed transformation $L$ and filter coefficients $b$ the function $\hat{j}^{+\sigma_j}$ tends to be smooth and falls off quickly away from $s = 0$. Using the approximation, we get

$$Lf \approx \sum_{m=-M}^{M'} \sum_{q=1}^Q \sum_{j=1}^3 w_q \hat{j}^{+\sigma_j}(L, b) B_{ms}^\sigma_j$$

where $\{q\} \subset \mathbb{R}$ are the quadrature points and $\{w_q\}$ are the weights.

We remark that the principle idea behind the expansion in Equation (16) involves exploiting the symmetry of the spherical log-polar basis functions under M"obius transformations taking $z$ to $-z^{-1}$. This allows us to replace the projective action of $L$ with the affine action of the upper-triangular matrices – the group of rotations, translations, and dilations – whose representations are better understood [Vilenkin 1978; Vilenkin and Klimyk 1991].

Efficient M"obius Convolutions. Plugging the approximation in Equation (17) into the definition of M"obius convolution in Equation (5) and moving the sums outside the integral gives

$$\psi \ast f \approx \sum_{m=-M}^{M'} \sum_{q=1}^Q \left( \sum_{j=1}^3 \left( \hat{j}^{+\sigma_j}(L, b) \right) \right).$$

Thus, by approximating the pointwise action of the frame operator $\Xi_\psi(z)$, we can approximate an arbitrary M"obius convolution as a sum of identity convolutions. The components of $\hat{j}^{+\sigma_j}$, depending only on $s$ can be pre-computed for a fixed set of quadrature points so that, in practice, the complexity of evaluating the function at run time is linear in the coefficients of $b$.

6.1 Complexity

In the approximation of M"obius convolution in Equation (18), the right side of the identity convolutions are independent of the filter coefficients, so the innermost bracketed term in Equation (12) can be pre-computed for a given band-limit $B$ (for every angular frequency $m$, quadrature point $s_q$, and localization index $\sigma_j$). Thus, the $O(B^2)$ computational bottleneck in computing the identity convolution need only be performed once and the total complexity of computing the M"obius convolution is $O(M'QB^3 \log B)$. In applications, we find that setting $M' = M + 1$ and using a $Q = 30$ point trapezoidal quadrature rule in Equation (17) allows us to both suitably approximate the transformation of the filter and scale up to $B = 64$.

7 M"OBUS-EQUIVARIANT SPHERICAL CNNS

Constructing M"obius-equivariant spherical CNNs with M"obius convolutions is straightforward and requires no specialized architecture. The atomic units are the same as those found in regular
CNNs – a convolutional layer, followed by normalization and a non-linearity.

7.1 Convolutional layers

For a Möbius convolution layer mapping $C$-channel input features $\psi \in L^2(\mathbb{C}, \mathbb{R}^C)$ to $C^\prime$-channel output features $\psi' \in L^2(\mathbb{C}, \mathbb{R}^{C^\prime})$, the $c^\prime$-th output feature $\psi_{c^\prime}'$ is computed in the usual manner by summing the convolutions of the input features with the filters in the $c^\prime$-th row of the bank. However, the structure of Equation (18) allows us to preform the reduction over the input channels before computing the convolutions in the sum, such that

$$
\psi'_{c^\prime} = \sum_{-M \leq m \leq M} \sum_{1 \leq j \leq 3} \sum_{c=1}^{C} \left( \sum_{q=1}^{Q} \rho_j \psi_j \mathcal{Z}_{c^\prime} \cdot b_{c^\prime} \mathcal{Z}_{c^\prime} \cdot b_{c^\prime} \mathcal{Z}_{c^\prime} \right),
$$

where $b_{c^\prime}$ denotes the $(2M+1)(2N+1)$ parameters for the $(c, c^\prime)$-th filter in the bank. Thus, for each convolutional layer mapping $C$ input features to $C^\prime$ output features, we only need to compute $C^\prime$ Möbius convolutions instead of $C \times C^\prime$.

This advantage is not without caveat. As discussed in the supplement, a naive implementation of the inner sum over the input channels produces large intermediate tensors at high resolutions ($B \geq 64$), which can quickly fill GPU memory. Our layers are implemented in PyTorch [Paszke et al. 2019], where we fuse this operation to reduce its overhead.

7.2 Normalization and Non-linearities

Standard normalization techniques don’t commute with Möbius transformations, since the mean and standard deviation of spherical signals are not invariant under dilations. Instead, we introduce a conformally-equivariant normalization layer based on filter response normalization [Singh and Krishnan 2020], replacing the square mean with the Dirichlet energy which is invariant under Möbius transformations. The normalization is applied on a per-channel basis independent of the batch size via the mapping

$$
\psi_c \mapsto \frac{\alpha_c \psi_c + \beta_c}{\sqrt{\int_{\mathbb{C}} \rho_c(z) dz + \epsilon_c}},
$$

where $\rho_c$ is defined as in Equation (9) and $\alpha_c, \beta_c \in \mathbb{R}$ and $\epsilon_c \in \mathbb{R}_{>0}$ are learnable per-channel parameters.

Following normalization we apply thresholded activations as non-linearities, which have been shown to better compliment filter response normalization than other activation layers [Singh and Krishnan 2020]. Here, we replace the ReLU with the Mish activation [Misra 2019] which we find improves training speed and performance. Specifically, non-linearities are applied pointwise as,

$$
\psi_c \mapsto \text{Mish}(\psi_c - Y_c) + Y_c,
$$

where $Y_c \in \mathbb{R}$ is a learnable per-channel threshold value. We note that the thresholded activation is not fundamental to our framework, and can be replaced with other activation layers if desired.

Figure 3: The equivariance error plotted as a function of the maximum conformal scale factor. Notably, moving from $U(1)$ (rotations) to $C_{\mathbb{R}\pi}$ (rotations and dilations) does not provide a benefit – one must consider the full lower-triangular subgroup $L \subset \text{SL}(2, \mathbb{C})$.

8 EVALUATION

We validate our claim of Möbius-equivariance empirically and demonstrate the utility of Möbius-equivariant CNNs by achieving strong results in both geometry and spherical-image processing tasks. In the former paradigm, we perform genus-zero shape classification by conformally mapping surfaces to the sphere; in the latter, we consider the task of omnidirectional image segmentation.

Our principal module in applications is an MCResNet block [He et al. 2016], consisting of two Möbius convolutions, each followed by the normalization layer and non-linearity described in Equations (20-21), with a residual connection between the input and output streams. We use $M = N = 1$ band-limited filters and set $t = 0.15$, optimizing for $\sigma_1$ and $\sigma_2$ and the quadrature points $\{s_q\}$ as described in the supplement. Our framework is implemented in PyTorch [Paszke et al. 2019], and we fit our networks using SGD with Nesterov momentum [Sutskever et al. 2013], training for 60 epochs with an initial learning rate of $10^{-5}$, decaying to $10^{-4}$ on a cosine annealing schedule [Loshchilov and Hutter 2017].

8.1 Equivariance

We empirically validate the equivariance of our framework by quantifying the degree to which our layers commute with Möbius transformations of increasing area distortion. We consider a 32-channel, $B = 64$ band-limited MCResNet block with the equivariant residual connection removed to avoid bias. We control the area distortion of a Möbius transformation $g$ by composing a series of random rotations and inversions so that the maximal scale factor over $\mathbb{C}$ equals a fixed value [Baden et al. 2018]. Denoting $R$ as the mapping induced by passing features through the MCResNet layer, we follow [de Haan et al. 2020; Sosnovik et al. 2019; Worrall and Welling 2019] and define the equivariance error for a fixed maximum scale factor $\alpha \in \mathbb{R}_{>0}$ as

$$
\text{Error} = \frac{\mathbb{E} \left( R(g \psi) - g R(\psi) \right)^2}{\mathbb{Var} g R(\psi)} \text{ with } \max_{z \in \mathbb{C}} \lambda^2_0(z) = \alpha,
$$

where $\mathbb{E}$ and $\mathbb{Var}$ denote the mean and variance computed over 100 randomly initialized models, Möbius transformations, and features.
Table 1: Genus-zero shape classification. Several conformally-deformed meshes from the SHREC ’11 dataset [2011] are shown below.

| Method          | Orig. Accuracy | Conf. Accuracy |
|-----------------|----------------|----------------|
| MC (ours)       | 99.1%          | 86.5%          |
| DiffusionNet [2020] | 99.5%          | 64.9%          |
| FC [2021]       | 99.2%          | 40.7%          |
| CubeNet [2021]  | 47.1%          | 21.2%          |

As a baseline, we compare our proposed approach against three other paradigms. In the first, we replace Möbius convolution with a standard 5 × 5 convolution layer taking ρ as input; in the second, we restrict the transformation field to rotations so that Ξ(z) ∈ U(1); in the third, we loosen the restriction to include dilations with Ξ(z) ∈ C2≥0. We note that the second and third paradigms are isometry-equivariant, and that the latter is also equivariant to the conformal transformations of the (non-compactified) plane.

The results are shown in Figure 3, where the equivariance error in Equation (22) is plotted as a function of the maximum scale factor. The green curve is our proposed method with Ξ(z) ∈ L. Using our method, the error stays very low, indicating that Möbius convolution is approximately equivariant even in the presence of significant changes in scale (λ ≥ 12). Notably, we see no improvement moving from Ξ(z) ∈ U(1) to Ξ(z) ∈ C2≥0, suggesting that rotations and dilations alone fail to well-characterize the local deformations induced by Möbius transformations.

8.2 Conformal Shape Classification

Next, we use Möbius convolutions to classify genus-zero shapes. Over the last half-decade, the SHREC ’11 dataset [Lian et al. 2011] has become a popular choice for evaluating network performance in shape classification tasks and several recent approaches have achieved near-perfect accuracy on the dataset [Milano et al. 2020; Mitchell et al. 2021; Sharp et al. 2020; Wiersma et al. 2021]. However, shapes within each of the 30 categories in the dataset differ only by (approximate) isometric deformations. To better highlight the strengths of our approach, we extend the dataset to include deformations given by random conformal transformations with several examples shown above Table 1.

To apply our framework, we conformally map each mesh to the sphere via mean curvature flow [Kazhdan et al. 2012] and use a simple network consisting of a single 16-channel, B = 64 band-limited MCResNet block followed by a global mean pool and a fully-connected layer to give predictions for the 30 shape categories.

We fit our network on both the original SHREC ’11 dataset and our conformally-augmented version using 10 samples per class. For comparisons, we report the results of Field Convolutions (FC) and DiffusionNet [Mitchel et al. 2021] and DiffusionNet [Sharp et al. 2020] - two state-of-the-art surface networks - on the original dataset and train both networks on the conformally-augmented version. We also train CubeNet [Shakerinava et al. 2021], a top performing rotation-equivariant spherical network, on both datasets. As inputs, each network takes the Heat Kernel Signature (HKS) [Sun et al. 2009] computed at 16 different timescales; since the HKS isn’t conformally-invariant, we use the values computed on the original meshes when training and testing on their conformally-augmented counterparts.

Results are shown in Table 1 in the form of the mean classification accuracy over three randomly sampled test-train splits. Our simple Möbius convolution network matches the state-of-art performance of FC and DiffusionNet on the original dataset and significantly outperforms both on the conformally augmented version, despite the fact that the transformations between the spherically-parameterized meshes aren’t perfect Möbius transformations. Like FC and DiffusionNet, our method is equivariant to isometric deformations of the meshes as they manifest as Möbius transformations after parameterization to the sphere, serving to explain our strong performance on the original dataset. However, Möbius-equivariance allows our rudimentary network to better account for conformal deformations between similar shapes and suggests that a new class of conformally-equivariant surface networks may outperform existing isometry-equivariant networks in challenging shape analysis and recognition tasks. CubeNet performs poorly on both datasets, indicating that Möbius-equivariance isn’t easily learned in a rotation-equivariant framework.

8.3 Omni-directional Image Segmentation

Last, we demonstrate the utility of Möbius convolutions by moving from geometry to image processing, where we apply them to semantically segment omni-directional images from the Stanford 2D3DS dataset [Armeni et al. 2017]. Here, we use MCResNet blocks to construct a U-Net [Ronneberger et al. 2015] architecture with 32, 64, 128, 256, 128, 64, 32 channels per layer, applying max pooling or nearest neighboring upsampling before each increase or decrease in channel width. As with other state-of-the-art equivariant spherical networks [Esteves et al. 2020; Shakerinava et al. 2021], we find our method performs best as a feature extractor for a small network of standard convolutional layers due to the consistent latitudinal orientation of the images; we append six 3 × 3 2D convolutions to the end of our network to predict labels. To measure performance, we report the mean per-class segmentation accuracy and intersection over union (IoU) averaged over the three official folds.

Results are shown in Table 2, and we attain performance comparable to the state-of-the-art. Existing spherical networks compute...
convolutions either in the spatial domain [Jiang et al. 2018; Shakernia et al. 2021; Zhang et al. 2019] or, like our method, in the spectral domain via expansions in spherical basis functions. In the latter case, efficiency and scalable filter support come at the cost of fidelity, as some high-frequency information is lost when computing the forward SHT due to the fixed band-limit assumption. This puts spectral methods at a disadvantage in precision labeling tasks with a large class imbalance, where spectral aliasing can blur sharp boundaries and over-smooth localized features necessary to make accurate predictions. Compounded by the devaluation of equivariance due to the consistent orientation of the images, this is a challenging task for our framework. However, we outperform existing rotation-equivariant spectral approaches, demonstrating that we are able to achieve Möbius-equivariance without sacrificing descriptive power. A visualization of our results and further discussion are provided in the supplement.

9 CONCLUSION

We present a novel, Möbius-equivariant spherical convolution operator which we call Möbius convolution. With it, we develop the foundations for Möbius-equivariant spherical CNNs and demonstrate the utility of this framework by achieving strong results in both geometry and spherical-image processing tasks.

More generally, this work represents an effort to move both image and surface convolutional neural networks beyond standard rotation- and isometry-equivariance and into the realm of conformal-equivariance. In particular, our experiments suggest that the latter transition may be especially relevant in the context of shape analysis and recognition and we hope this work serves to catalyze the development of a new generation of surface networks better able to handle the kinds of complex deformations found in real-world shape data.

REFERENCES

Iro Armeni, Sasha Sax, Amir R Zamar, and Silvio Savarese. 2017. Joint 2d-3d-semantic data for indoor scene understanding. arXiv preprint arXiv:1702.01165 (2017).

Alex Baden, Keenan Crane, and Michah Kazhdan. 2018. Möbius registration. In Computer Graphics Forum, Vol. 37. Wiley Online Library. 211–220.

Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi, David Miller, and Ravi Konda. 2020. Lorentz group equivariant neural networks for particle physics. In International Conference on Machine Learning. PMLR, 992–1002.

Gregory S Chirikjian and Alexander B Kyatkin. 2016. Harmonic Analysis for Engineers: and Applied Scientists: Updated and Expanded Edition. Courier Dover Publications.

Carlos Esteves, Christine Allen-Blanche, Ameesh Makadia, and Kostas Daniilidis. 2018. Learning so(3) equivariant representations with spherical cnns. In European Conference on Computer Vision. 52–68.

Carlos Esteves, Ameesh Makadia, and Kostas Daniilidis. 2020. Spin-weighted spherical cnns. Advances in Neural Information Processing Systems (2020).

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. 2020. Generalized convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In International Conference on Machine learning. PMLR, 3165–3176.

Marc Finzi, Max Metzler, and Andrew Gordon Wilson. 2021. A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups. arXiv preprint arXiv:2104.09459 (2021).

Xiaofeng Gu, Yalin Wang, T.F. Chan, P.M. Thompson, and Shing-Tung Yau. 2004. Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Transactions on Medical Imaging 23, 8 (2004), 949–958.

Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. 2019. MeshCNN: A network with an Edge. Transactions on Graphics 38, 4 (2019), 90.

R. Hartley and A. Zisserman. 2003. Multiple View Geometry in Computer Vision. Cambridge University Press. https://books.google.com/books?id=szIXRPh69YQ.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Computer Vision and Pattern Recognition. 770–778.

Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Phillip Marcus, Matthias Niessner, et al. 2018. Spherical CNNs on Unstructured Grids. In International Conference on Learning Representations.

Michael Kazhdan, Jake Solomon, and Mirela Ben-Chen. 2012. Can mean-curvature flow be modified to be non-singular?. In Computer Graphics Forum, Vol. 31. Wiley Online Library. 1745–1754.

Peter J Kostelec and Daniel N Rockmore. 2008. FFTs on the rotation group. Journal of Fourier analysis and applications 14, 2 (2008), 145–179.

Alon Lahav and Ayellet Tal. 2020. Meshwalker: Deep mesh understanding by random walks. Transactions on Graphics 39, 6 (2020), 1–13.

Bruno Lévy, Sylvain Petitjean, Nicolas Ray, and Jerome Maillot. 2002. Least Squares Conformal Maps for Automatic Texture Atlas Generation. ACM Trans. Graph, 21, 3 (2002), 362–371.

Zhoubnin Lian, Afral Godil, Benjamin Bustos, Mohamed Daoudi, Jeroen Hermans, Shun Kawamura, Yukinori Kurita, Guillaume Lavoué, Hien Nguyen, Ryutaro Ohihuchi, Yuki Ohkita, Yuya Ohishi, Fathi Porikli, Martin Reuter, Ivan Sipiran, Dirk Smeets, Paul Sutents, Heri Tabia, and Dirk Vanderveen. 2011. SIGBEC 11: Track Shape Retrieval on Non-rigid 3D Watertight Meshes. In Eurographics Workshop on 3D Object Retrieval. 79–88.

Ilya Loushchov and Frank Hutter. 2017. SGD: Stochastic Gradient Descent with Warm Restarts. arXiv:1608.03983 [cs.LG].

Lachlan Ewen MacDonald, Sameera Ramasinge, and Simon Lucey. 2021. Enabling equivariance for arbitrary Lie groups. https://doi.org/10.48550/ARXIV.2111.08251.

Francesco Milano, Antonio Loquercio, Antonio Rosinol, Davide Scaramuzza, and Luca Carlone. 2020. Primal-Dual Mesh Convolutional Neural Networks. arXiv:2010.12455 [cs.CV].

Diganta Misra. 2019. Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.06861 (2019).

Thomas W Mitchell, Benedict Brown, David Koller, Tim Weyrich, Szymon Rusinkewicz, and Michael Kazhdan. 2020. Efficient spatially adaptive convolution and correlation. arXiv preprint arXiv:2006.13188 (2020).

Thomas W Mitchell, Vladimir G Kim, and Michael Kazhdan. 2021. Field Convolutions for Surface CNNs. In International Conference on Computer Vision. 10001–10011.

Thomas W Mitchell, Szymon Rusinkewicz, Gregory S Chirikjian, and Michael Kazhdan. 2020b. ECHO: Extended Convolution Hologram of Orientations for Local Surface Description. In Computer Graphics Forum. Wiley Online Library.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32 (2019), 8026–8037.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 79–88.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV].

Saul Schleimer and Henry Segerman. 2016. 3D printing. In Surfaces and Interfaces: An Introduction to Surface Science. 79–88.

Carlos Esteves, Christine Allen-Blanche, Ameesh Makadia, and Kostas Daniilidis. 2018. Learning so(3) equivariant representations with spherical cnns. In European Conference on Computer Vision. 52–68.

Saurabh Singh and Shankar Krishnan. 2020. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks. In Computer Vision and Pattern Recognition. 11237–11246.

Dmitry Smirnov and Justin Solomon. 2021. HodgeNet: learning spectral geometry on graph signals. arXiv preprint arXiv:2102.05425 (2020).

James R Driscoll and Dennis M Healy. 1994. Computing Fourier transforms and convolutions on the 2-sphere. Advances in Applied Clifford Algebras 15, 2 (1994), 205–250.

Saul Schleimer and Henry Segerman. 2016. 3D printing. In Surfaces and Interfaces: An Introduction to Surface Science. 79–88.
Ivan Sosnovik, Michał Szmaja, and Arnold Smeulders. 2019. Scale-Equivariant Steerable Networks. In *International Conference on Learning Representations*.

Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. 2009. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, Vol. 28. Wiley Online Library, 1383–1392.

Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*. PMLR, 1139–1147.

N. Ja. Vilenkin. 1978. *Special functions and the theory of group representations*. Vol. 22. American Mathematical Soc.

N. Ja. Vilenkin and A. U. Klimyk. 1991. Representation of Lie Groups and Special Functions: Volume 1: Simplest Lie Groups, Special Functions and Integral Transforms (Mathematics and its Applications).

Soledad Villar, David Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. 2021. Scalars are universal: Equivariant machine learning, structured like classical physics. *Advances in Neural Information Processing Systems* 34 (2021).

George Neville Watson. 1995. *A treatise on the theory of Bessel functions*. Cambridge university press.

Maurice Weiler and Gabriele Cesa. 2019. General E (2)-Equivariant Steerable CNNs. *Advances in Neural Information Processing Systems* 32 (2019), 14334–14345.

Ruben Wiersma, Elmar Eisemann, and Klaus Hildebrandt. 2020. CNNs on surfaces using rotation-equivariant features. *Transactions on Graphics* 39, 4 (2020), 92–1.

Ruben Wiersma, Ahmad Nasikun, Elmar Eisemann, and Klaus Hildebrandt. 2021. DeltaConv: Anisotropic Point Cloud Learning with Exterior Calculus. *arXiv preprint arXiv:2111.08799* (2021).

Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. 2017. Harmonic networks: Deep translation and rotation equivariance. In *Computer Vision and Pattern Recognition*. 5028–5037.

Daniel E Worrall and Max Welling. 2019. Deep scale-spaces: Equivariance over scale. *arXiv preprint arXiv:1905.11697* (2019).

Li Yi, Hao Su, Xingwen Guo, and Leonidas J. Guibas. 2017. SyncSpecCNN: Synchronous Spectral CNN for 3D Shape Segmentation. In *Computer Vision and Pattern Recognition*.

Chao Zhang, Stephan Liwicki, William Smith, and Roberto Cipolla. 2019. Orientation-aware semantic segmentation on icosahedron spheres. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3533–3541.