Spherical CNNS

Liangjie Cao

July 6, 2018

Abstract

Convolutional Neural Networks (CNNs) have become the method of choice for learning problems involving 2D planar images. However, a number of problems of recent interest have created a demand for models that can analyze spherical images. Examples include omnidirectional vision for drones, robots, and autonomous cars, molecular regression problems, and global weather and climate modelling. A naive application of convolutional networks to a planar projection of the spherical signal is destined to fail, because the space-varying distortions introduced by such a projection will make translational weight sharing ineffective. In this paper the authors introduce the building blocks for constructing spherical CNNs. They propose a definition for the spherical cross-correlation that is both expres- sive and rotation-equivariant. The spherical correlation satisfies a generalized Fourier theorem, which allows us to compute it efficiently using a generalized (non-commutative) Fast Fourier Transform (FFT) algorithm. They demonstrate the computational efficiency, numerical accuracy, and effectiveness of spherical CNNs applied to 3D model recognition and atomization energy regression.

1. Introduction

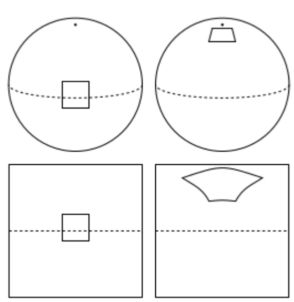

Figure 1. Any planar projection of a spherical signal will result in distortions. Rotation of a spherical signal cannot be emulated by translation of its planar projection.

Convolutional networks are able to detect local patterns regardless of their position in the image. Like patterns in a planar image, patterns on the sphere can move around, but in this case the "move" is a 3D rotation instead of a translation. In analogy to the planar CNN, we would like to build a network that can detect patterns regardless of how they are rotated over the sphere.

As shown in Figure 1, there is no good way to use translational convolution or cross-correlation[1] to analyze spherical signals. The most obvious approach, then, is to change the definition of cross- correlation by replacing filter translations by rotations. Doing so, they run into a subtle but important difference between the plane and the sphere: whereas the space of moves for the plane (2D translations) is itself isomorphic to the plane, the space of moves for the sphere (3D rotations) is a different, three-dimensional manifold called SO(3)[2]. It follows that the result of a spherical correlation (the output feature map) is to be considered a signal on $SO(3)$, not a signal on the sphere, S^2. For this reason, they deploy $SO(3)$ group correlation in the higher layers of a spherical CNN [2]. Then they address both of these problems using techniques from non-commutative harmonic analysis [1, 3].

2. Correlation On The Sphere And Rotation Group

They will explain the S^2 and $SO(3)$ correlation by analogy to the classical planar $\mathbb{Z}^2$ correlation. The planar correlation means the value of the output feature map at translation $x \in \mathbb{Z}^2$ is computed as an inner product between the input feature map and a filter, shifted by x. And the spherical correlation means the value of the output feature map evaluated at rotation $R \in SO(3)$ is computed as an inner product between the input feature map and a filter, rotated by R.

[1] Despite the name, CNNs typically use cross-correlation instead of convolution in the forward pass. In this paper the authors generally use the term cross-correlation, or correlation for short.

[2] To be more precise: although the symmetry group of the plane contains more than just translations, the translations form a subgroup that acts on the plane. In the case of the sphere there is no coherent way to define a composition for points on the sphere, and so the sphere cannot act on itself (it is not a group).

*In order to define the spherical correlation, the authors need to know not only how to rotate points $x \in S^2$ but also how to rotate filters (*i.e. *functions) on the sphere. To this end, they introduce the rotation operator L_R that takes a function f and produces a rotated function $L_R f$ by composing f with the rotation R^{-1}:*

$$[L_R f](x) = f(R^{-1}x) \tag{1}$$

Due to the inverse on R, they have $L_{RR'} = L_R L_{R'}$.

The inner product on the vector space of spherical signals is defined as:

$$< \psi, f >= \int_{S^2} \sum_{k=1}^{K} \psi_k(x) f_k(x) dx \tag{2}$$

The integration measure dx denotes the standard rotation invariant integration measure on the sphere, which can be expressed as $d\alpha\ R\ sin(\beta)d\beta/4\pi$ in R spherical coordinates. The invariance of the measure ensures that $\int_{S^2} f(Rx)dx = \int_{S^2} f(x)dx$, for any rotation $R \in SO(3)$. That is, the volume under a spherical height map does not change when rotated. Using this fact, they can show that $L_{R^{-1}}$ is adjoint to L_R, which implies that L_R is unitary:

$$< L_R\psi, f >= \int_{S^2} \sum_{k=1}^{K} \psi_k(R^{-1}x) f_k(x) dx \tag{3}$$

Then we know that this two function is equal.

References

[1] *G. S. Chirikjian and A. B. Kyatkin.* Engineering Applications of Noncommutative Harmonic Analysis. *Crc Press, 2001. 1*

[2] *T. S. Cohen and M. Welling. Group equivariant convolutional networks. In* ICML*, 2016. 1*

[3] *G. B. Folland.* A Course in Abstract Harmonic Analysis. *Crc Press, 1995. 1*