Extended Abstract Track

On Optimal Lifting to SE(2) in Equivariant Neural Networks

Abstract

Equivariant neural networks, such as PDE-GCNNs, often require data as feature maps over the group. In this paper, we motivate criteria for the optimal lifting of feature maps over $\mathbb{R}^2$ to $SE(2)$. We propose three optimality criteria: fast reconstruction property which ensures that no information is lost during the lifting, spatial locality and orientation locality. The locality conditions make sure that information is organized in a meaningful manner in the lifted space. We formulate corresponding losses which we then numerically minimize, and find that kernels emerge that closely resemble cake wavelets. The results indicate that the cake wavelets are near optimal under presented criteria, and as such provide an excellent starting point for any $SE(2)$ equivariant architecture.

Keywords: Group Convolutions, Geometric Deep Learning, Lifting

1. Introduction

Group-convolutional neural networks (Cohen and Welling, 2016) separate lifting, convolutional and projection layers. Group convolutional networks operate over G-features maps which are signals over a group G, which in this work we take to be the 2D roto-translation group $G = SE(2)$. Inputs, however, are typically feature maps over $\mathbb{R}^2$ (images), and hence a lifting from $\mathbb{R}^2$ to $SE(2)$ is required. While this can be done using a form of group convolution (lifting convolution), some models such as PDE-G-CNNs (Smets et al., 2023) benefit from the inputs to be already defined over the group. As proven in (G.Bellaard et al., 2023b) one does not need to train the lifting layer (as it can be accounted for in the subsequent convection/convolutions layers on the group). Then not training the lifting layer improves geometric interpretability of the PDE-G-CNNs (G.Bellaard et al. (2023a)) in terms of neuro-geometrical association fields that are defined on the group. A natural question then arise of what constitutes an optimal linear lifting of feature maps to the group.

Inspired by wavelet theory, we propose two major criteria that a lifting operation should have: *reconstructability* (invertibility) (Duits et al., 2007; Janssen et al., 2018) and *locality* (minimal variance) (Antoine and Murenzi, 1996; Antoine et al., 1999). *Reconstructability* ensures that no information is lost from lifting. The notion of *locality* enables weight-sharing over localized feature patterns, which is the main factor of success behind the convolution neural network paradigm (LeCun et al., 1998). Since we lift data from $\mathbb{R}^2$ to $SE(2) = \mathbb{R}^2 \rtimes SO(2)$, we want make sure that information not only remains localized in space, but also within the additional rotation axis $SO(2)$. In other words, when a kernel is specific to a certain orientation, its activation can be attributed to that orientation, thus *disentangling* orientation information (Franken and Duits, 2009; Cohen and Welling, 2014).

In this work we define an optimization problem based on the criteria of invertibility and locality and numerically obtain optimal kernels that can be used in lifting convolutions. We show that such kernels closely resemble cake wavelets (Duits, 2005; Janssen et al., 2018).

2. Preliminaries

Much of the success of convolutions/cross-correlations in computer vision is attributed to their translation equivariance. A spatial cross-correlation is given by

$$(f \circledast k)(\boldsymbol{\tau}) := \sum_{\boldsymbol{x} \in \mathbb{Z}^2} f(\mathbf{x})\, \overline{k}(\mathbf{x} - \boldsymbol{\tau}),$$

with k is the convolutional kernel, f is the input image, and τ is a coordinate in the feature map. We note that this operation can also be given in terms of group actions via

$$(f \circledast k)(\boldsymbol{\tau}) := \langle f, \mathcal{L}_\tau k \rangle \,, \tag{1}$$

that is, via the inner product $\langle f, k \rangle := \sum_{\boldsymbol{x} \in \mathbb{Z}^2} f(\boldsymbol{x})\overline{k}(\boldsymbol{x})$, with $\bar{\cdot}$ denoting complex conjugation, and $\mathcal{L}_\tau$ the left regular representation of the translation group on feature maps, i.e., $(\mathcal{L}_\tau k)(\boldsymbol{x}) := k(\boldsymbol{x} - \boldsymbol{\tau})$. Any equivariant bounded linear operator is of the form of (1), and can be interpreted as a form of template matching (Bekkers, 2020).

The (regular) group convolution paradigm is implemented by the appropriate inner product, that is, if f and k are feature maps over the group than integration/summation takes place over the group ($\langle f, k \rangle := \sum_{g \in G} f(g)\overline{k}(g)$), and the appropriate left regular representation $\mathcal{L}_g$. For example, when the feature maps are over $\mathbb{R}^2$ and one applies an $SE(2)$ group convolution, and $(\mathcal{L}_g k)(\boldsymbol{x}) := k(\mathbf{R}_\theta^T(\boldsymbol{x} - \boldsymbol{\tau}))$, with $g = (\boldsymbol{\tau}, \theta)$. Thus, the $SE(2)$ lifting convolution is given by

$$(f \circledast_G k)(\boldsymbol{\tau}, \theta) = \sum_{\boldsymbol{x} \in \mathbb{Z}^2} f(\boldsymbol{x})\overline{k}_\theta(\boldsymbol{x} - \boldsymbol{\tau}) \,, \tag{2}$$

in which with $k_\theta(\boldsymbol{x}) := k(\mathbf{R}_\theta^T \boldsymbol{x})$ we denote the base kernel rotated over an angle θ. We will limit our discussion to the lifting convolution (Eq. 2) and refer to (Bekkers, 2020) for the group convolution and projection cases.

We further note that we will use the term orientation and rotation interchangingly, as $SO(2)$ is homomorphic to the ring of orientations S^1 with the group product $\theta \cdot \theta' = \theta + \theta'$ mod 2π. We follow the wavelet theoretic framework by Duits et al. (2007) and further refer to optimally lifted feature fields over $SE(2)$ as functions over *position-orientation space* $\mathbb{R}^2 \times S^1$ using the term *orientation scores*. We finally note, that we will discretize the rotation axis into N rotations, and as such, are effectively considering the cyclic permutation group C_N instead of the continuous $SO(2)$.

3. Methodology

We derive optimal kernels by parametrizing them through their Fourier spectrum and minimizing three losses corresponding to reconstructibility and locality in space and orientation.

The Fast Reconstruction Property It is important that during lifting no information is lost. That is, one should be able to reconstruct the original image using the lifted function. In this work, we aim for the *fast reconstruction property* Duits et al. (2007); Janssen et al. (2018), which requires reconstruction by simply summing over the orientation axis. I.e., for all f we want $f(\boldsymbol{x}) = \sum_{\theta \in S^1} (f \circledast_G k)(\boldsymbol{x}, \theta)$.

Due to linearity of both the lifting and the fast reconstruction, we obtain a constraint on the kernel itself. Namely,

$$f(\boldsymbol{x}) = \sum_{\theta \in S^1} (f \circledast_G k)(\boldsymbol{x}, \theta) \;\; \Leftrightarrow \;\; f(\boldsymbol{x}) = (f \circledast \sum_{\theta \in S^1} k_\theta)(\boldsymbol{x}) \;\; \Leftrightarrow \;\; f(\boldsymbol{x}) = (f \circledast \tilde{k})(\boldsymbol{x}) \,,$$

where we define $\tilde{k}[\boldsymbol{x}] := \sum_{\theta \in S^1} k_\theta[\boldsymbol{x}]$ as the sum of the rotated kernels. It is then clear that this equation only holds when $\tilde{k}$ equals the unit impulse (δ-function). Alternatively, invoking the Fourier Convolution Theorem, the Fourier transform $\mathcal{F}(\tilde{k})$ must be 1 within the bandlimit of f. Thus, the Fourier spectrum of $\tilde{k}$ should match a disk with a radius given by the Nyquist frequency $\mathcal{N}$, and this gives the key constraint on the kernel

$$\sum_{\theta \in S^1} \hat{k}_\theta(\boldsymbol{\omega}) = 1, \qquad \forall_{\boldsymbol{\omega} \in \mathbb{R}^2} \; ||\boldsymbol{\omega}|| < \mathcal{N} \,, \tag{3}$$

where $\hat{k}_\theta := \mathcal{F}(k_\theta)$ is the Fourier transform of k_θ. We translate this constraint into the loss

$$\boxed{L^{\text{recon}} = \sum_{||\boldsymbol{\omega}|| < \mathcal{N}} |1 - \sum_{\theta \in C_n} \hat{k}_\theta(\boldsymbol{\omega})|^2 \,.} \tag{4}$$

The Orientation Localization Property The fast reconstruction property restricts the wavelet to sum to a $\delta-$function. However, this does not impose a unique solution. The trivial solution to the fast reconstruction property would be the kernel that is a δ-function weighted by $\frac{1}{|C_n|}$. Since then each rotated kernel is still a δ-function, the lifting merely creates copies of the input for each θ, and it results into what we call a *trivial lifting*. In the trivial solution, the information is *maximally entangled*, as any oriented information is spread evenly over the orientation axis and it is a result of lack of orientation sensitivity of the δ-function. We address this issue by maximizing sensitivity to oriented patterns.

In order to turn the notion of orientation sensitivity into a measurable quantity associated with the kernel k, we again fall back to Fourier theory. We now consider $\boldsymbol{\omega}$ as a random variable under a distribution defined by $\hat{k}$. Namely we define $p_{\hat{k}}(\boldsymbol{\omega}) := \frac{|\hat{k}(\boldsymbol{\omega})|^2}{\int_{\mathbb{R}^2} |\hat{k}(\boldsymbol{\omega})|^2 \mathrm{d}\omega}$ and define $\hat{k}$ directly in Fourier space.

The kernel $p_{\hat{k}}$ represents the probability density that a frequency in an input signal f is picked up by the kernel k when performing the convolution. Note that any frequency $\boldsymbol{\omega} = (\omega_x, \omega_y)$ is associated with an orientation through $\theta(\boldsymbol{\omega}) = \arg(\omega_x + i\omega_y) \in [-\pi, \pi)$. Then—following the standard definition of variance—the variance of the orientation $\theta(\boldsymbol{\omega})$ associated with a frequency $\boldsymbol{\omega}$ is given by the expected squared distance of $\theta(\boldsymbol{\omega})$ relative a mean orientation θ_0, namely,

$$\boxed{L^{\theta\text{-variance}} = \int_{\mathbb{R}^2} |\theta(\boldsymbol{\omega}) - \theta_0|^2 p_{\hat{k}}(\boldsymbol{\omega}) \mathrm{d}\omega \,.} \tag{5}$$

In summary, a kernel k is maximally sensitive to a specific orientation axis θ_0 if its variance (Eq. 5) is minimized, that is, if it only responds to features close to that reference orientation θ_0. Since in the group convolution (Eq. 2) the rotations shift this specific orientation θ_0 by rotating the kernel by an amount of θ, we conclude that $(f \circledast_G k)(\boldsymbol{x}, \theta)$ at each location $\boldsymbol{x}$ indicates the presence of frequency content centered around an orientation of $\theta_0 - \theta$. Further, note that the trivial solution produces maximum variance as the frequency distribution associated with the (band-limited) δ-function is the uniform disk.

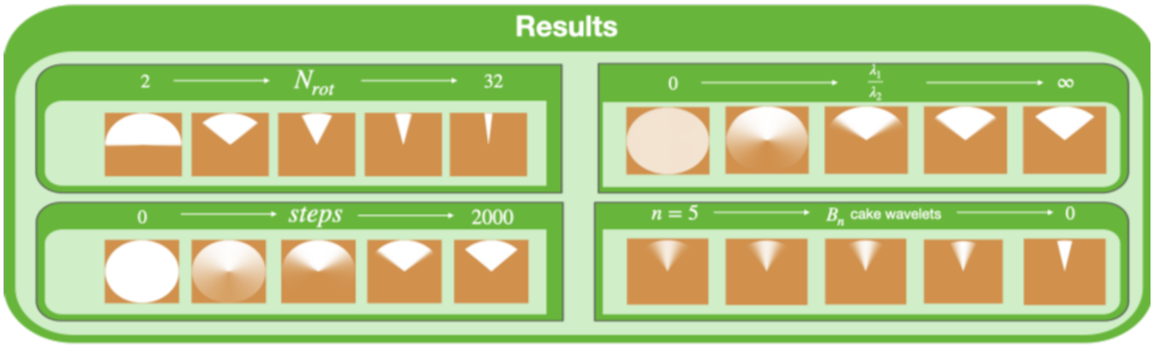

Figure 1: Optimization of the kernel in the Fourier domain.

The Spatial localization Property Similar to (5), the spatial localization can be measured with the variance of the kernel with respect to position, which, assuming $\boldsymbol{x}_0 = \mathbf{0}$, is given by $\int_{\mathbb{R}^2} \|\boldsymbol{x}\|^2 p_k(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$. We note, however, that locality of a spatial signal corresponds to smoothness of its frequency spectrum, and as such, we can equivalently quantify spatial locality through the Fourier parametrization of k by mimizing the total gradient norm using

$$\boxed{L^{\boldsymbol{x}\text{-variance}} = \int\limits_{\mathbb{R}^2} ||\nabla_{\omega_x} \; \hat{k}\left(\boldsymbol{\omega}\right)||^2 \mathrm{d}\omega \approx \sum_{\boldsymbol{\omega} \in \mathbb{Z}^2} ||\nabla_{\omega_x} \; \hat{k}\left(\boldsymbol{\omega}\right)||^2 .} \tag{6}$$

4. Results and Conclusion

Results The result of minimizing the three losses given in Eqs. 4-6 can be seen in Fig. 1. The orientation and spatial localization losses are respectively balanced by λ_1 and λ_2 respectively, and we note that the ratio λ_1/λ_2 balances the two types of locality as by Heisenberg's uncertainty principle spatial certainty (locality) induces uncertainty in momentum/frequency and thus in orientation. If the optimization is executed without encouraging spatial localization, the numerical optimization converges to sharp slices in the Fourier domain. This result intuitively makes sense: reconstructibilty requires that all rotated copies sum to one, and thus frequency content need to be distributed, then the orientation variance is minimized by centering all mass around a preferred orientation. The result is that the (band-limited) frequency spectrum (the cake) is sliced into equal pieces.

As the spatial localization weight increases, orientation variance has to increase due to Heisenberg's uncertainty principle, and this is recognized as a smoothing of the slices. This gives them the appearance of cake wavelets as proposed in Duits (2005). The $\lambda_2 = 0$ case gives precisely the B_0 cake wavelet (see Appendix A). As $\lambda_2 \gg \lambda_1$, the kernel approaches the trivial lifting kernel, i.e. the δ-function.

Conclusion If one wishes to lift signals onto the discrete $SE(2)$ group using a predefined (non-trainable) kernel, then a kernel that is a slice in the Fourier domain provides the optimal localization in orientation. If one wants to balance this with spatial localization, cake wavelets appear to be the near-optimal choice under the fast reconstruction property.

References

J-P Antoine and Romain Murenzi. Two-dimensional directional wavelets and the scale-angle representation. *Signal processing*, 52(3):259–281, 1996.

J-P Antoine, Romain Murenzi, and Pierre Vandergheynst. Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns. *Applied and Computational Harmonic Analysis*, 6(3):314–345, 1999.

Erik J Bekkers. B-spline cnns on lie groups. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=H1gBhkBFDH.

Taco Cohen and Max Welling. Learning the irreducible representations of commutative lie groups. In *International Conference on Machine Learning*, pages 1755–1763. PMLR, 2014.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

Remco Duits. Perceptual organization in image analysis: a mathematical approach based on scale, orientation and curvature. 2005.

Remco Duits, Michael Felsberg, Gösta Granlund, and Bart ter Haar Romeny. Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the euclidean motion group. *International Journal of Computer Vision*, 72(1):79–102, 2007.

Erik Franken and Remco Duits. Crossing-preserving coherence-enhancing diffusion on invertible orientation scores. *International Journal of Computer Vision*, 85:253–278, 2009.

G.Bellaard, D.Bon, G.Pai, B.M.N.Smets, and R.Duits. Analysis of (sub-)Riemannian PDE-G-CNNs. *JMIV*, 65:819–843, 2023a.

G.Bellaard, G.Pai, J.O. Bescos, and R.Duits. Geometric adaptations of PDE-G-CNNs. *LNCS*, 14009:538–550, 2023b.

Michiel HJ Janssen, Augustus JEM Janssen, Erik J Bekkers, J Oliván Bescós, and Remco Duits. Design and processing of invertible orientation scores of 3d images. *Journal of mathematical imaging and vision*, 60:1427–1458, 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.

Bart MN Smets, Jim Portegies, Erik J Bekkers, and Remco Duits. PDE-based group equivariant convolutional neural networks. *Journal of Mathematical Imaging and Vision*, 65(1):209–239, 2023.

Appendix A. Cake Wavelets

Cake wavelets, introduced by Duits (2005), provide an efficient method for orientation score transforms, enabling the separation of position-orientation information in signals. They are defined by a smoothed slice in the Fourier domain. The slices is constructed with a radial envelop,

$$g(\rho) := \mathcal{M}_N(\rho^2/t) = e^{-\rho^2/t} \sum_{k=0}^{N} \frac{(\rho^2/t)^k}{k!} \tag{7}$$

where ρ represents the radial coordinate, where $t > 0$ is a scaling parameter, and where N is the order of the power series, and an angular smoothing is performed using B-spline interpolation:

$$\mathcal{B}^k(\theta) = (\mathcal{B}^{k-1} * \mathcal{B}^0)(\theta) \tag{8}$$

where $*$ denotes convolution and k is the order of the B-spline. The Fourier transform of the cake wavelet is given by:

$$\hat{\psi}_{\text{cake}}(\omega) = B^k\left(\frac{(\phi \mod 2\pi) - \pi/2}{s_\theta}\right)(\mathcal{M}_N(\rho))^{\frac{1}{2}}, \qquad \omega = (\rho\cos\varphi, \rho\sin\varphi). \tag{9}$$

Here, B^k denotes the B-spline of order k, ϕ is the angular coordinate, s_θ represents the angular scale, and ρ is the radial coordinate. The use of radial B-splines in this formulation ensures that the sum over rotations satisfies the fast reconstruction property.

Typically, the scaling parameter $t > 0$ is chosen such that the inflection point of radial function g (where the second order derivative vanishes) is close to the Nyquist frequency $\mathcal{N}$ due to sampling, i.e. for $g''(\gamma \cdot \mathcal{N}) = 0$ we choose $t = \frac{2\gamma^2}{1+2N}$ with $0 \ll \gamma < 1$. In practice $\gamma = 0.7$ is a reasonable choice.

Finally, a radial envelope is applied in the spatial domain:

$$\psi_{\text{cake}}(\boldsymbol{x}) = \mathcal{F}^{-1}\left[\hat{\psi}_{\text{cake}}\right](\boldsymbol{x})\; G_{\sigma_s}(\boldsymbol{x}) \tag{10}$$

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform and G_{σ_s} is a Gaussian with standard deviation σ_s. This windowing will kill possibly long tails of the cakewavelet if one has a highy Nyquist frequency (i.e. many samples) and hardly affect the stability of the orientation score transform Duits (2005).

However, in this paper, we focus on the Fourier transform of the cake-wavelet $\hat{\psi}_{\text{cake}}$. For analytic descriptions in the spatial domain one has to work with Zernike polynomials as in Janssen et al. (2018) but here we only need analytic descriptions in the Fourier domain.