Separable Group Equivariant Neural Networks on Lie groups

David M. Knigge

UNIVERSITY OF AMSTERDAM

David W. Romero

VU VRIJE UNIVERSITEIT AMSTERDAM

Erik J. Bekkers

UNIVERSITY OF AMSTERDAM

Short overview

UNIVERSITY OF AMSTERDAM

Broadly, research into Equivariant CNNs can be divided into:

- Regular G-CNNs
  - Use **regular representation** to transform kernel
  - Intuitive, **easy to implement**
  - Applicable to **broad range** of groups
  - **Computational complexity** depends on **group size** and dimensionality ←

- Steerable G-CNNs
  - Constrain kernels to steerable basis
  - Applicable to unimodular groups, e.g. **rotation groups**
  - **Decouple computational complexity** from group size

Summary

UNIVERSITY OF AMSTERDAM

- **Scalability** remains a problem in regular G-CNNs
  - We propose **separating** the **group convolution** operation for affine groups $G = \mathbb{R}^2 \rtimes H$ by **subgroups:**
  - Factorise into a **separate convolution** over H and $\mathbb{R}^2$

Group convolutions

UNIVERSITY OF AMSTERDAM

- Feature maps f are defined over G
- The kernel k now also extends over G (considerably increasing computational complexity)

$$
\begin{aligned}
(f *_{\text{group}} k)(g) &= \int_G f(\tilde{g}) k(g^{-1} \cdot \tilde{g}) \mathrm{d}\tilde{g} \\
&= \int_{\mathbb{R}^2} \int_H f(\tilde{\boldsymbol{x}}, \tilde{h}) k(h^{-1}(\tilde{\boldsymbol{x}} - \boldsymbol{x}), h^{-1} \cdot \tilde{h}) \frac{1}{|h|} \mathrm{d}\tilde{\boldsymbol{x}} \mathrm{d}\tilde{h}
\end{aligned}
$$

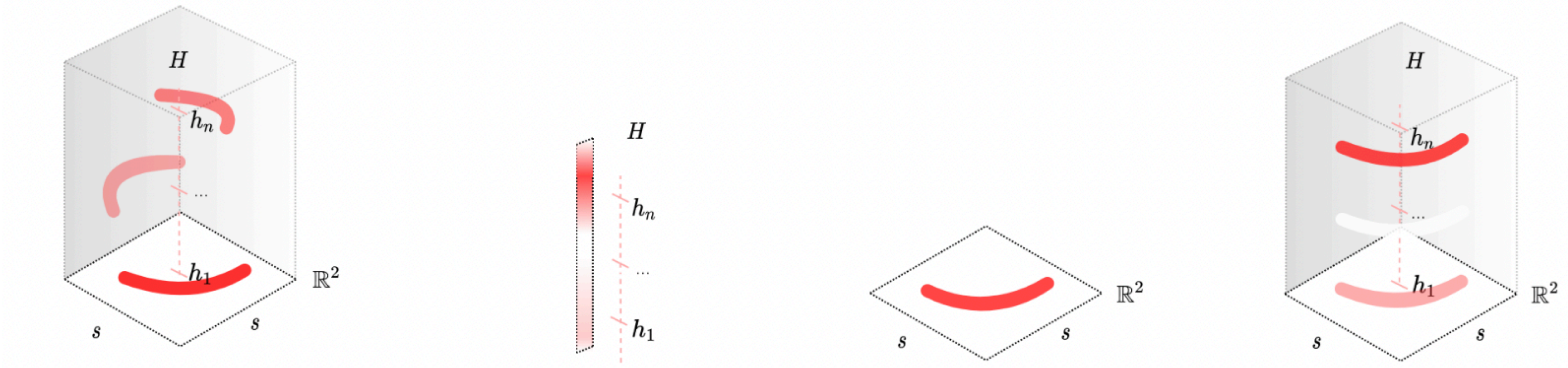
In case of affine groups

where $g = (\boldsymbol{x}, h)$

Separable group convolutions

UNIVERSITY OF AMSTERDAM

- Idea - treat distinct subgroups separately:
  - i.e. separate convolutions over rotation and spatial dimensions
  - Assume: $k(g) = k_{\mathbb{R}^2}(\boldsymbol{x}) \cdot k_H(h)$

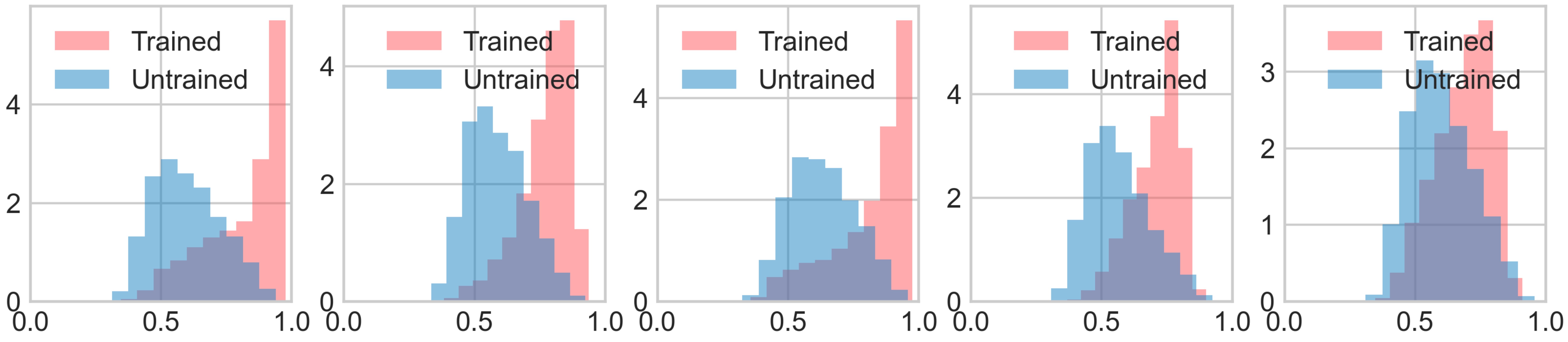


(a) $k : \mathbb{R}^2 \rtimes H \to \mathbb{R}$ (b) $k_H : H \to \mathbb{R}$ (c) $k_{\mathbb{R}^2} : \mathbb{R}^2 \to \mathbb{R}$ (d) $k_H \cdot k_{\mathbb{R}^2}$

Separable group convolutions

UNIVERSITY OF AMSTERDAM

- It turns out there is empirical motivation for this approach, learned G-conv kernels exhibit redundancy along the H axis!



Visualisation of redundancy in group convolution kernels in subsequent layers

Separable group convolutions

UNIVERSITY OF AMSTERDAM

- Idea: treat distinct subgroups separately:
    - i.e. separate convolutions over rotation and spatial dimensions
    - Assume: $k(g) = k_{\mathbb{R}^2}(\boldsymbol{x}) \cdot k_H(h)$

$$(f *_{group} k)(g) \approx \sum_{\tilde{\boldsymbol{x}} \in \mathbb{Z}^2} \sum_{\tilde{h} \in \mathscr{H}_\epsilon} f(\tilde{\boldsymbol{x}}, \tilde{h}) k(h^{-1}(\tilde{\boldsymbol{x}} - \boldsymbol{x}), h^{-1} \cdot \tilde{h}) \frac{1}{|h|}$$

$$= \sum_{\tilde{\boldsymbol{x}} \in \mathbb{Z}^2} \Big[\sum_{\tilde{h} \in \mathscr{H}_\epsilon} f(\tilde{\boldsymbol{x}}, \tilde{h}) k_H(h^{-1} \cdot \tilde{h}) \Big] \frac{1}{|h|} k_{\mathbb{R}^2}(h^{-1}(\tilde{\boldsymbol{x}} - \boldsymbol{x}))$$

Does not depend on $\boldsymbol{x}$, precompute this!

Separable group convolutions

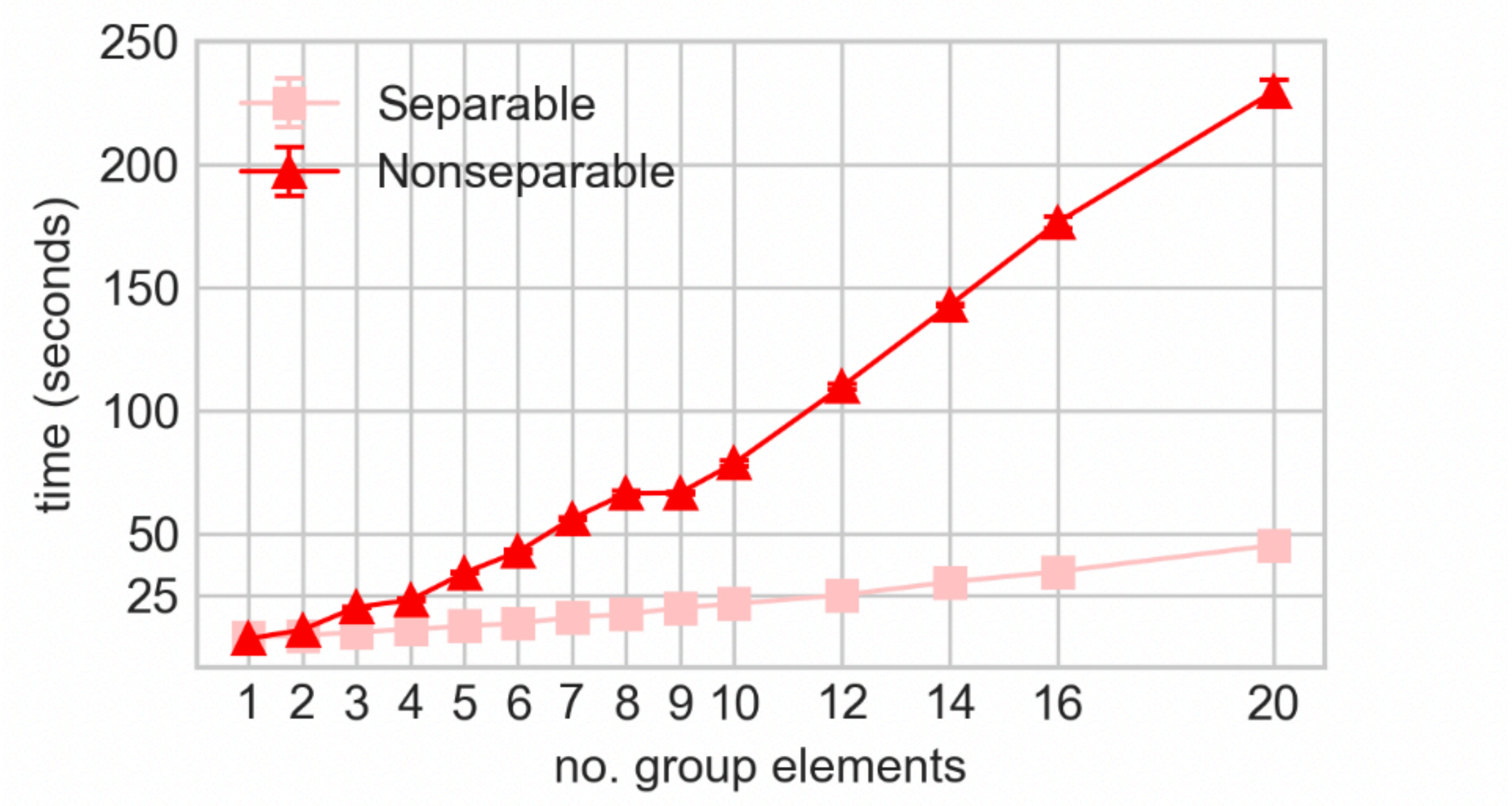
UNIVERSITY OF AMSTERDAM

Computational efficiency

- Drastically **reduces computational complexity:**

- In turn allowing for larger groups

Process time per epoch on MNIST-rot



Separable group convolutions

Sim(2)-equivariance

Table 2. Test error (%) on rotated MNIST for separable G-CNNs in comparison to other equivariant baselines: G-CNN (Cohen & Welling, 2016a), H-Net (Worrall et al., 2017), RED-NN (Salas et al., 2019), LieConv (Finzi et al., 2020), SFCNN (Weiler et al., 2018b), E(2)-NN (Weiler & Cesa, 2019). † Separable along dilation and rotations dimensions. + Train-time augmentation by continuous rotations.

Baseline Methods							Separable G-CNNs (Ours)			
G-CNN	H-Net	RED-NN	LieConv	SFCNN	SFCNN_+	E(2)-NN_+	SE(2)	$\text{SE}(2)_+$	$\text{Sim}(2)^\dagger$	$\text{Sim}(2)^\dagger_+$
2.28	1.69	1.39	1.24	0.88	0.714	0.68	$0.89 \pm .008$	$0.66 \pm .023$	$0.66 \pm .009$	**0.59**$\pm .008$

Table 3. Test error (%) on CIFAR10 with All-CNN-C architecture, for our separable group convolutions in comparison to other baselines: All-CNN-C (Springenberg et al., 2014), $p4$- and $p4m$-G-CNN (Cohen & Welling, 2016a). † Separable along dilation and rotations dimensions. + Train-time augmentation by random horizontal flips and random cropping. n-Sim(2)-CNNs where n is the SIREN hidden size in units. 6-Sim(2)-CNNs have approximately equal numbers of parameters to the original All-CNN-C.

Baseline Methods					Separable G-CNNs (Ours)					
1.4m param.	1.37m param.		1.27m param.		1.14m param.		1.33m param.		3.22m param.	
All-CNN-C	$p4$-G-CNN	$p4$-G-CNN_+	$p4m$-G-CNN	$p4m$-G-CNN_+	5-$\text{Sim}(2)^\dagger$	5-$\text{Sim}(2)^\dagger_+$	6-$\text{Sim}(2)^\dagger$	6-$\text{Sim}(2)^\dagger_+$	16-$\text{Sim}(2)^\dagger$	16-$\text{Sim}(2)^\dagger_+$
9.08	8.84	7.67	7.59	7.04	8.50	7.41	8.22	6.47	7.27	**5.50**

Separable group convolutions

Separable vs. nonseparable