Supplementary Material
Learning to Convolve: A Generalized Weight-Tying Approach

Anonymous Authors[1]

Here we provide proofs of the equivariance properties of the of the group convolution and unitary group convolution. We also show example of filters and activations.

1. Equivariance Of The Group Convolution

Here we provide a copy of Cohen & Welling (2016)'s equivariance proof of the discrete group convolution. For a signal $f : X \to \mathbb{R}$, filter $\psi : X \to \mathbb{R}$, domain X, group G, and group action $\mathcal{L}_g$ where $\mathcal{L}_g[f](x) = f(\mathcal{L}_g^{-1}[x])$, we have

$$
\begin{aligned}
[\mathcal{L}_t[f] \star_G \psi](g) &= \sum_{x \in X} \mathcal{L}_t[f](x)\psi(\mathcal{L}_g^{-1}[x]) && (1)\\
&= \sum_{x \in G} f(\mathcal{L}_t^{-1}[x])\psi(\mathcal{L}_g^{-1}[x]) && (2)\\
&= \sum_{x' \in G} f(x')\psi(\mathcal{L}_g^{-1}[\mathcal{L}_t[x']]) && (3)\\
&= \sum_{x' \in G} f(x')\psi(\mathcal{L}_{g^{-1}t}[x']]) && (4)\\
&= \sum_{x' \in G} f(x')\psi(\mathcal{L}_{(t^{-1}g)^{-1}}[x']]) && (5)\\
&= [f \star_G \psi](t^{-1}g) && (6)\\
&= \mathcal{L}_t[f \star_G \psi](g) && (7)
\end{aligned}
$$

From line 1 to 2 we used the definition $\mathcal{L}_g[f](x) = f(\mathcal{L}_g^{-1}[x])$; from line 2 to 3 we performed as change of variables $x' = \mathcal{L}_t^{-1}[x]$ or equally $x = \mathcal{L}_t[x']$; from line 3 to 4 we applied the composition rule for actions; from line 4 to 5 we used the rule $(ab)^{-1} = b^{-1}a^{-1}$ and in the remaining lines we used the definitions of the group convolution and actions.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2. The Equivariance Loss

In the equivariance loss we make use of the following statement

$$\mathcal{R}_S[f] \star_{\mathbb{Z}^d} \mathcal{R}_R[\psi] = \mathcal{R}_S[f \star_{\mathbb{Z}^d} \mathcal{R}_{S^{-1}R}[\psi]]. \quad (8)$$

The derivation is as follows. We begin by noting that the roto-translation operator can be written $\mathcal{L}_{R,z} = \mathcal{T}_z\mathcal{R}_R$, where $\mathcal{T}_z$ is the translation operator and $\mathcal{R}_R$ is the rotation operator. Then we consider the convolution of an S-rotated image $\mathcal{R}_S[f]$ and filters ψ

$$
\begin{aligned}
[\mathcal{R}_S[f] \star_G \psi](R, z) &= \sum_{x \in G} \mathcal{R}_S[f](x)\mathcal{T}_z[\mathcal{R}_R[\psi]](x) && (9)\\
&= [\mathcal{R}_S[f](x) \star_G \mathcal{R}_R[\psi]](x) && (10)
\end{aligned}
$$

which constitutes the LHS of the expression. Now for the RHS.

$$
\begin{aligned}
&\mathcal{R}_S[f] \star_{\mathbb{Z}^d} \mathcal{R}_R[\psi] && (11)\\
&= [\mathcal{L}_{S,0}[f] \star_G \psi](R, z) && (12)\\
&= \sum_{x \in G} \mathcal{L}_{S,0}[f](x)\mathcal{L}_{R,z}[\psi](x) && (13)\\
&= \sum_{x \in G} f(x)\mathcal{L}_{(S,0)^{-1}}[\mathcal{L}_{R,z}[\psi]](x) && (14)\\
&= \sum_{x \in G} f(x)\mathcal{L}_{(S^{-1},0)}[\mathcal{L}_{R,z}[\psi]](x) && (15)\\
&= \sum_{x \in G} f(x)\mathcal{L}_{(S^{-1}R,S^{-1}z)}[\psi](x) && (16)\\
&= \sum_{x \in G} f(x)\mathcal{L}_{(S^{-1}R,S^{-1}z)}[\psi](x) && (17)\\
&= \sum_{x \in G} f(x)\mathcal{T}_{S^{-1}z}[\mathcal{R}_{S^{-1}R}[\psi]](x) && (18)\\
&= \sum_{x \in G} f(x)\mathcal{R}_{S^{-1}R}[\psi](x - S^{-1}z) && (19)\\
&= [f \star_{\mathbb{Z}^d} \mathcal{R}_{S^{-1}R}[\psi]](S^{-1}z) && (20)\\
&= \mathcal{R}_S[f \star_{\mathbb{Z}^d} \mathcal{R}_{S^{-1}R}[\psi]](z) && (21)
\end{aligned}
$$

which constitutes the RHS of the expression.

3. Architecture

Here we detail the architecture used in our experiments.

Table 1. The architectures of the translational and roto-translational equivariant models. After every convolution we place a batch normalization layer and a ReLU nonlinearity. Across the two models we have fixed the number of channels, such that the number of parameters is roughly the same. 'conv$N$' stands for a standard translational convolution of size $N \times N$ and 'Gconv$N$' stands for a roto-translational group convolution. Horizontal lines correspond to max pooling of kernel size 2 and stride 2. The global max pool corresponds to a max pool over the spatial dimensions and the orientation dimensions of the activation tensor.

TRANSLATIONAL	ROTO-TRANSLATIONAL
conv3-96	Gconv-33
conv3-96	Gconv-33
conv3-96	Gconv-33
conv3-192	Gconv-67
conv3-192	Gconv-67
conv3-192	Gconv-67
conv3-192	Gconv-67
conv1-192	Gconv-67
conv1-192	Gconv-67
global max pool	global max pool
softmax-layer	softmax-layer

3.1. Visualization of bases and reconstructions

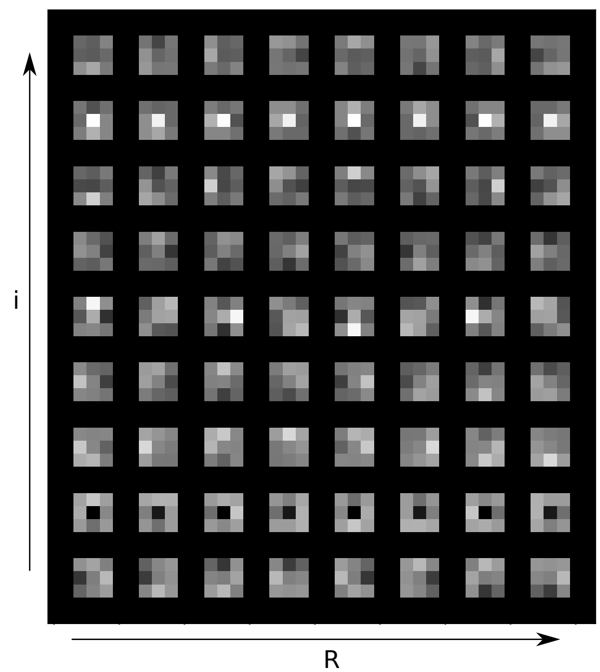

Figure 1. A basis with 9 elements at 8 orientations from an PARTIAL model. $\{e^i_R\}_{i,R}$

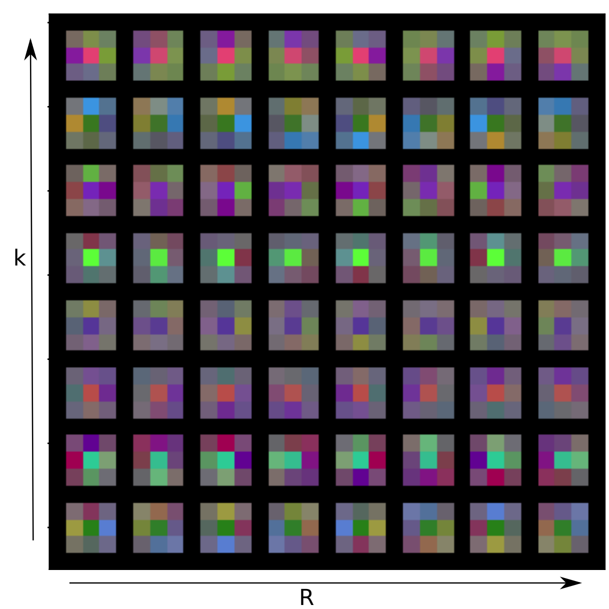

Figure 2. A set of filters from the first layer of an PARTIAL model. $\mathcal{R}_R[\psi_k]$

References

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pp. 2990–2999, 2016.

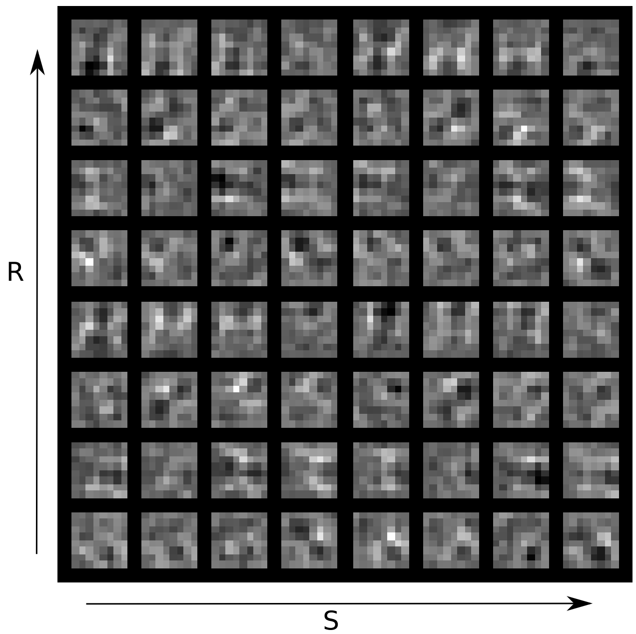

Figure 3. A set activations from an PARTIAL model's layer 6. $\mathcal{R}_R[f] \star_{\mathbb{Z}^d} \mathcal{R}_S[\psi]$

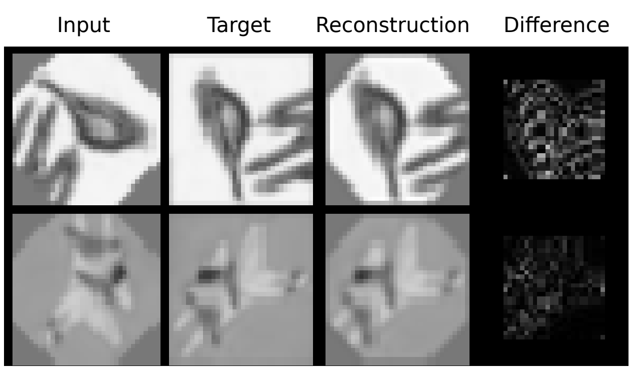

Figure 4. A set of 2 pairs from the reconstruction task, when training the basis. The loss is normalized to the scale of the loss, otherwise it would be too small to distinguish anything.

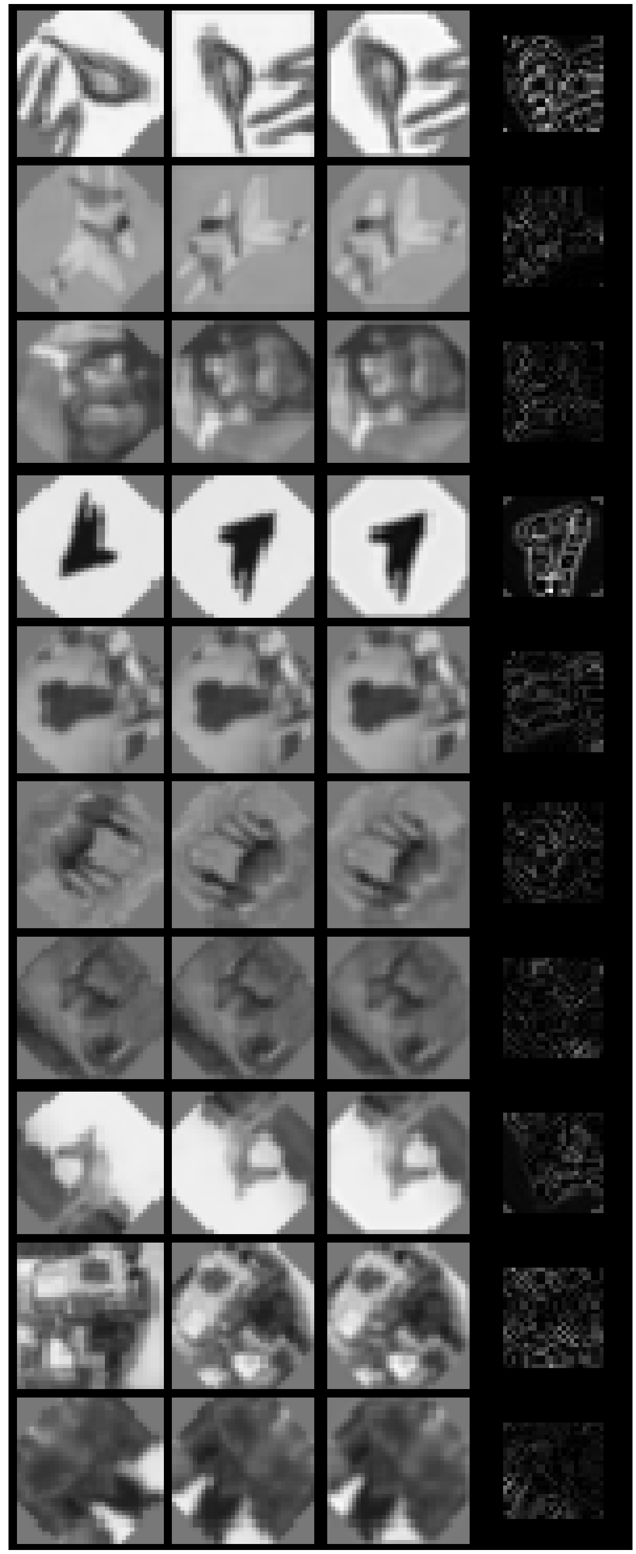

Figure 5. A set of 10 pairs from the reconstruction task, when training the basis. The columns represent in order: the input, the target, the reconstruction, the loss.