# Summary of scale equivariance in recent models

Park Chan Ho

## 1 Introduction

This document aims to summarize and compare the construction of scale equivariance in different models. By exploring through the mathematical foundation behind each ideas, I will attempt to propose a direction in which will improve the performance of models or scale convolution.

The papers that will be explored are the following:

1. A general Theory of equivariant CNNs on homogeneous spaces[1]
2. B-spline CNNs on Lie Groups[2]
3. Scale-equivariant steerable network[3]
4. Deep Scale Spaces: Equivariance over scale[4]
5. Scale Equivariant Neural Networks with decomposed convolution filters[5]

First we will look at the generalized group convolution analysis by [1] and then relate it and show the equivalence with the other definitions stated in different works.

## 2 Group Equivariant Convolution

### 2.1 Generalized Group Equivariant convolution by [1]

This perspective looks as the feature maps and associated group structure as a fiber bundle; implying that the feature map as fields over a continuous space.

$p : E \to B$ satisfying for all a in E there exists an open neighborhood U, and a homeomorphism $\psi$ such that $\psi : p^{-1}(U) \to U\times$ F

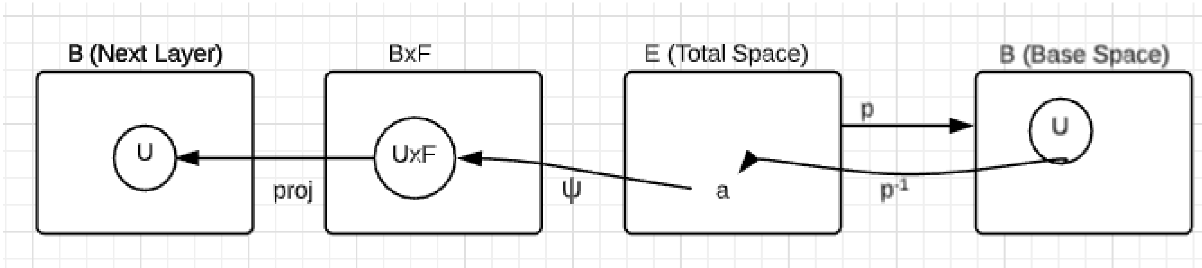


Figure 1: Image general structure of equivariant mapping (bundle)

$\psi \circ proj = p$ where $proj(u, f) = u$

Association with group will convert each of the components to the following meaning.

E(Total Space) = G x V where V represents a vector space

B(Base Space) = G/H

F(fiber) = H group twisted copy of G/H

p = projection of element of E to its coset representation in G/H

proj = group pooling

Induced representation $\pi = Ind_H^G \rho$ describes the action of G on fields (feature map or input image in the context of group convolution). This group representation will satisfy group action on fields as such: $[\pi_G(g)f](k) = f(g^{-1}k)$

The association of group has underlying assumptions as such: 1.Groups acts linearly on input and hidden layers. 2. Stabalizer at all points are isomorphic and equivalent to stabalizer group H

With such construction, the equivariant kernel is defined to be a equivariant maps between induced representations ($[k \cdot [\pi_1(u)f]](g) = [\pi_2(u)[k \cdot f]](g)$). This leads to the equivariant kernel function space being isomorphic to the space of left-equivariant kernels on $G/H_1$. This allows us to express the kernel in terms of a linear combination of basis functions of such isomorphic spaces. This idea is utilized in [3],[5]

With this constraint, we are now able to define a group convolution as such:

$$[k \cdot f](g) = \int_G k(g, g')f(g')dg' = \int_G k(g^{-1}g')f(g')dg' = [k \star f](g)$$

Depending on the representation of the stabalizer group H, the non linearity functions that will not affect the equivariance differs. For example for regular representation, all the non linearities are applicable due to its commutativity with convolution

## 2.2 Group Equivariant convolution by [2]

Before proceeding to the definitions and constraints described by the paper, these are some notations to be noted.

$L^N(X)$ = Vector space defined on elements of group X

$K \in L(X_2 \times X_1)$ = Operator mapping $L^N(X_1) \rightarrow L^N(X_2)$

$\widetilde{\star}$ denotes the general convolution symbol while $\star_{R^d}$ denotes the $R^d$ convolution.

$B^{R^d,n}$ is the set of basis functions in $R^d$ of order $n$

The association of lie group has underlying assumptions as such: 1.Groups acts transitively on X,Y implying X,Y are homogenous spaces. 2. Kernel is linear and bounded

The kernel constraints are as such:

1. $K(x', x) = \frac{d\mu_x(g_{x'}^{-1} \circ x)}{d\mu_x(x)} k(g_{x'}^{-1} \circ x)$ here $g_{x'} \in G$ is an element such that $x' = g_{x'} \circ x'_0$ where $x'_0$ is the origin of space $X_2$ (Proof in appendix)

2. If $X_2 = G/H$ where H is stabalizer of G, then kernel is constrained by

$\forall_{h \in H}, \forall_{x \in X} : k(x) = \frac{d\mu_x(g_{x'}^{-1} \circ x)}{d\mu_x(x)} k(h^{-1} \circ x)$ (Proof in appendix)

The model divides group convolution in the 3 stages:

1. Lifting layer

$(Kf)(g) = (k \widetilde{\star} f)(g) = (k_h \star_{R^d} f)(x)$ (Proof in appendix)

2. Group Convolution Layer

$(KF)(g) = (K \star F)(g) = (K_h \star_{R^d} F))(x)$ (Proof in appendix)

3. Projection Layer

$(KF)(x) = \int_H F(x, \widetilde{h}) d\mu(\widetilde{h})$

Based on these kernel constraints and convolution, this paper expresses the kernel in terms of linear combination of the B-spline basis with respect to the lie group of subject. B-spline basis on lie group can be written as:

$$K(g) = K(x,h) := \Sigma_{i=1}^{N} c_i B^{R^d,n}(\frac{x-x_i}{s_x}) B^{R^d,n}(\frac{Log(h_i^{-1}h)}{s_h})$$

A more concrete example (scale group) is stated in part 3.

## 2.3 Relation between the construction of [1] and [2]

In this section, I will show how the ideas from [1] and [2] are related and specifically how [2] is a special case of [1].

It is clear that the lifting, group convolution and projection stage from [2] corresponds to $p^{-1}, \Phi$ and $proj$ of [1] respectively.

To compare the kernel constraints, we can prove that the kernel satisfying constraints of [2] will satisfy the conditions stated in [1]; that is convolution have the property of Mackey Functions

E(Total Space) = G x V = $L^N(X)$

Remark. V represents a vector space, generally $\mathbb{R}^d$ in this case d = N

B(Base Space) = G/H

F(fiber) = f and F

p = $(K)(x) = \int_H F(x,h) d\mu(h)$ representation in G/H

proj = group pooling

That is $f(gh) = \rho(h^{-1}) f(g)$ where $\rho$ are group representations of stabalizer group H

In terms of the convolution we aim to prove:

$$[k \star f](gh) = \rho(h^{-1})[k \star f](g)$$

Let $\rho(h^{-1}) = \frac{d\mu_x(g_y^{-1} \cdot x)}{d\mu_x(x)} = 1$

Starting with the constraint of [2], we have:

$$[k \star f](x) = \rho(h^{-1})K(h^{-1}x)$$

$$\Leftrightarrow K(h^{-1}g^{-1}g') = \rho(h^{-1})K(g^{-1}g')$$

$$\Leftrightarrow \int_G K((gh)^{-1}g')f(g')dg' = \rho(h^{-1}) \int_G K(g^{-1}g')f(g')dg'$$

$$\Leftrightarrow [K \star f](gh) = \rho(h^{-1})[K \star f](g)$$

$\Leftrightarrow [K \star f]$ is a Mackey function

$\Leftrightarrow [K \star f]$ satisfies the equivariance constraint of [1]

# 3 Scale Group Convolution

In this section, I will be discussing about how existing models have achieved equivariance over scale group in the perspective of the ideas from section 1 and 2. For ease of reference, I will refer to the perspective of [1] as 'general equivariant convolution' and [2] as 'group equivariant convolution'

B-spline CNN over lie groups: In context of group equivariant convolution, $G = R^2 \times R^+$ and $H = R^+$; implying $G/H = R^2$. The kernels are transformed through the 2nd constraint specified in 2.2 in order to maintain equivariance over group.

Scale equivariant steerable Network achieves equivariance over scale through constructing the kernel as a linear combination of scale steerable basis (e.g. Hermite polynomial function). From the perspective of general equivariant convolution $E = Z^2 \times S, B = Z^2$. Operations such as pooling and constraints are analogous to the ideas specified in [1].

Deep Scale Spaces lifts the input image through a gaussian kernel then makes use of inter-scale dilated convolution to achieve scale equivariance. Assuming the gaussian kernel operation can be approximated by scaled B-spline basis functions $B_s^{R^d,n} = B^{R^d,n}(\frac{1}{s}x)$, we can view deep scale space as a scale group ($R^2 \times R^+$) equivariant convolution model. (Proof in appendix)

Similar to SESN, in [5] the kernel is expressed in terms of two function bases, namely eigenfunctions of Dirichlet Laplacian on unit disk and [-1,1]. This reduces the number of parameter as well as computational cost significantly.

# 4 Direction of improvement

In attempt of improving the existing models, there can be directions such as 1. reducing computational cost, 2. Improving performance through more expressive parameters. Since reducing computational cost through a more concise expression of convolution can be challenging, I believe that making changes to B-spline CNN on scale groups ($R^2 \times R^+$) can be a direction.

# 5 Appendix

## 5.1 Proof of constraints of [2]

Equivariance of the kernel can be expressed as $K \circ L_g^{G \to L_2(x)}(f) = (L_g^{G \to L_2(Y)} \circ K)(f)$

This is equivalent to:

$$\int_X \widetilde{K}(y,x) f(g^{-1}x) dx = \int_X \widetilde{K}(g^{-1}y, x) f(x) dx$$

$$= \int_X \widetilde{K}(g^{-1}y, g^{-1}x) f(g^{-1}x) d(g^{-1}x)$$

$$= \int_X \widetilde{K}(g^{-1}y, g^{-1}x) f(g^{-1}x) \tfrac{1}{|det(g)|} d(x)$$

$$\Rightarrow \widetilde{K}(y,x) = \tfrac{1}{|det(g)|} \widetilde{K}(g^{-1}y, g^{-1}x)$$

$$= \widetilde{K}(g_y y_0, x) (\because \text{ Since G is transitive over Y, } \forall_{y, y_0 \in Y} \exists_{g_y \in G} \text{ such that } y = g_y y_0)$$

$$= \tfrac{1}{|det(g)|} \widetilde{K}(y_0, g_y^{-1}(x)) = \tfrac{1}{|det(g)|} K(g_y^{-1} x) \qquad \text{[A]}$$

$y_0$ is the origin hence $\widetilde{K} \to K$ is a conversion from two argument to one argument kernel. Moreover making use of the fact that H is a stabalizer group of $y_0$, the constraint can be further specified as:

$$K(x) := \widetilde{K}(y_0, x) = \widetilde{K}(h \circ y_0, x) \quad (\because h \in H = Stab_G(y_0))$$

$$= \tfrac{1}{|det(h)|} K(h^{-1} \circ x) \;\; (\because \text{[A]})$$

We should note since Lebesgue measure ($d\mu_x$) on $R^d$ for affine lie group satisfies:

$$\tfrac{1}{|det(g_y)|} = \tfrac{d\mu_x(g_y^{-1} \cdot x)}{d\mu_x(x)}$$

If $d\mu_x$ is haar measure on $G$ for affine lie group (G) then $\frac{d\mu_x(g_y^{-1} \cdot x)}{d\mu_x(x)} = 1$

## 5.2 3 stages of group convolution model described in [2]

1. Lifting stage ($\mathbb{Z}^2 \to G$)

   Note that since $g \in G = \mathbb{R}^d \times \mathbb{H} \Rightarrow g = (x, h)$

   Furthermore, $K_h$ indicates the transformed kernel $K_h(x) = \frac{1}{|det(h)|}(\mathbb{L}_h^{H \to L_2(\mathbb{R}^d)})$

$$
\begin{aligned}
(Kf)(g) = (K \widetilde{\star} f)(g) &:= \frac{1}{|det(h)|}(L_g^{G \to L_2(\mathbb{R}^d)} K, f)_{L_2(R^d, dx)} \\
&= \frac{1}{|det(h)|}(L_x^{\mathbb{R}^d \to L_2(X)} L_h^{H \to L_2(X)} K, f) \\
&= \int_{\mathbb{R}^d} K_h(x - x') f(x') dx' \\
&= (K_h \star_{\mathbb{R}^d} f)(x)
\end{aligned}
$$

2. Group Convolution ($G \to G$)

$$
\begin{aligned}
(Kf)(g) = (K \star f)(g) &:= (L_g^{G \to L_2(\mathbb{R}^d)} K, F)_{L_2(G, d\mu)} \qquad (\because \frac{d\mu_x(g_y^{-1} \cdot x)}{d\mu_x(x)} = 1) \\
&= (L_x^{\mathbb{R}^d \to L_2(X)} L_h^{H \to L_2(X)} K, F)_{L_2(G, d\mu)} \\
&= (L_x^{\mathbb{R}^d \to L_2(X)} K_h, f)_{L_2(G, d\mu)} \\
&= \int_{\mathbb{R}^d} K_h(x - \widetilde{x}) F(\widetilde{x}) d\mu(\widetilde{x}) \\
&= (K_h \star_{\mathbb{R}^d} F)(x)
\end{aligned}
$$

3. Projection Layer ($G \to \mathbb{R}^d$)

$$
(KF)(x) := \int_H F(x, \widetilde{h}) \; d\mu(\widetilde{h})
$$