Group Equivariant Subsampling

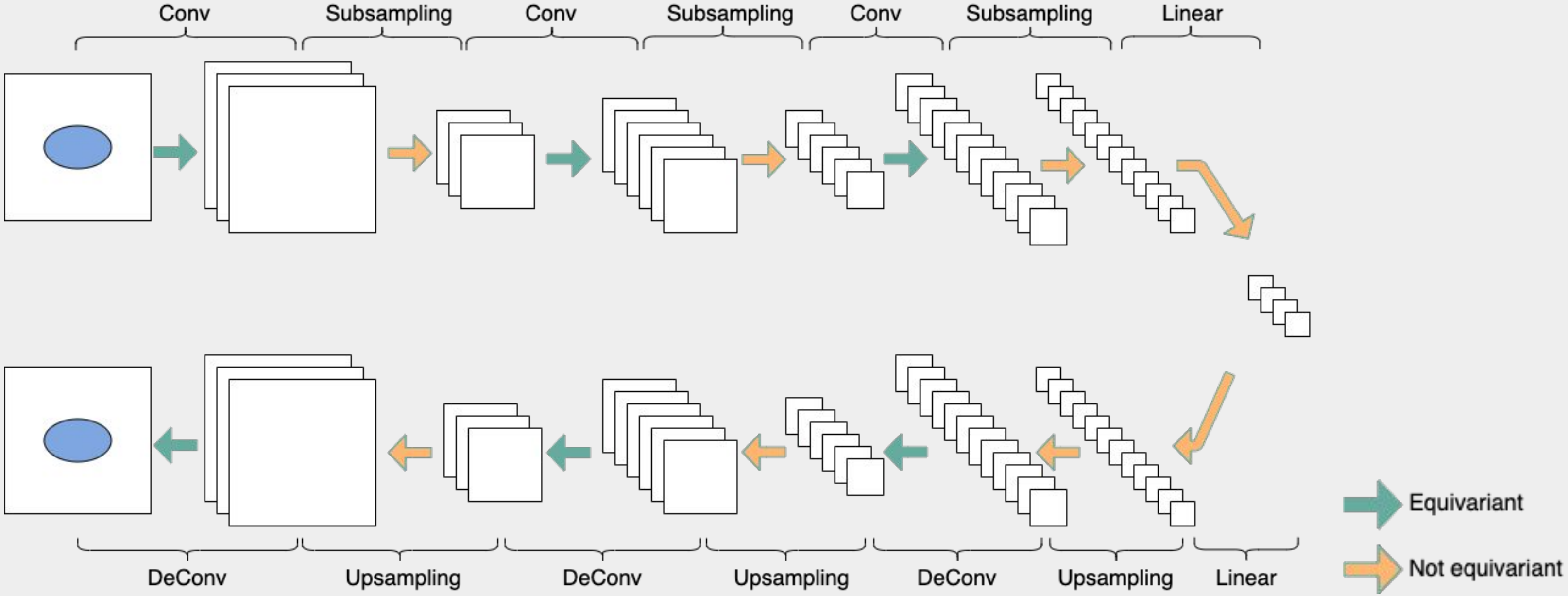
Jin Xu, Hyunjik Kim, Tom Rainforth, Yee Whye Teh

UNIVERSITY OF OXFORD

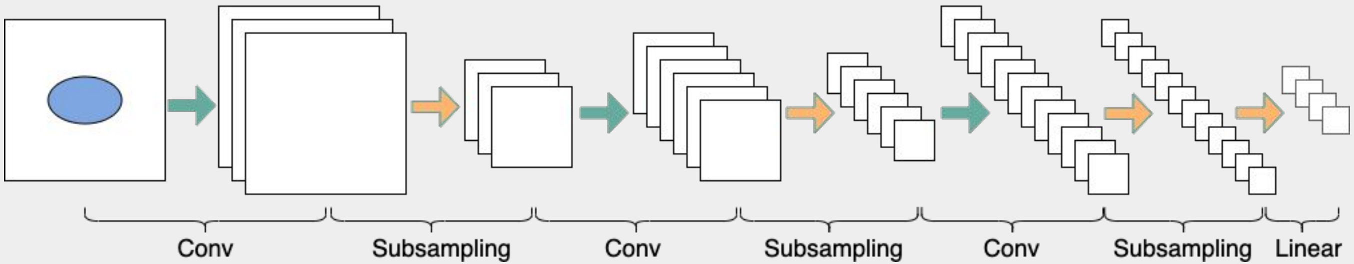
DeepMind

35th Conference on Neural Information Processing Systems

Convolutional Autoencoders

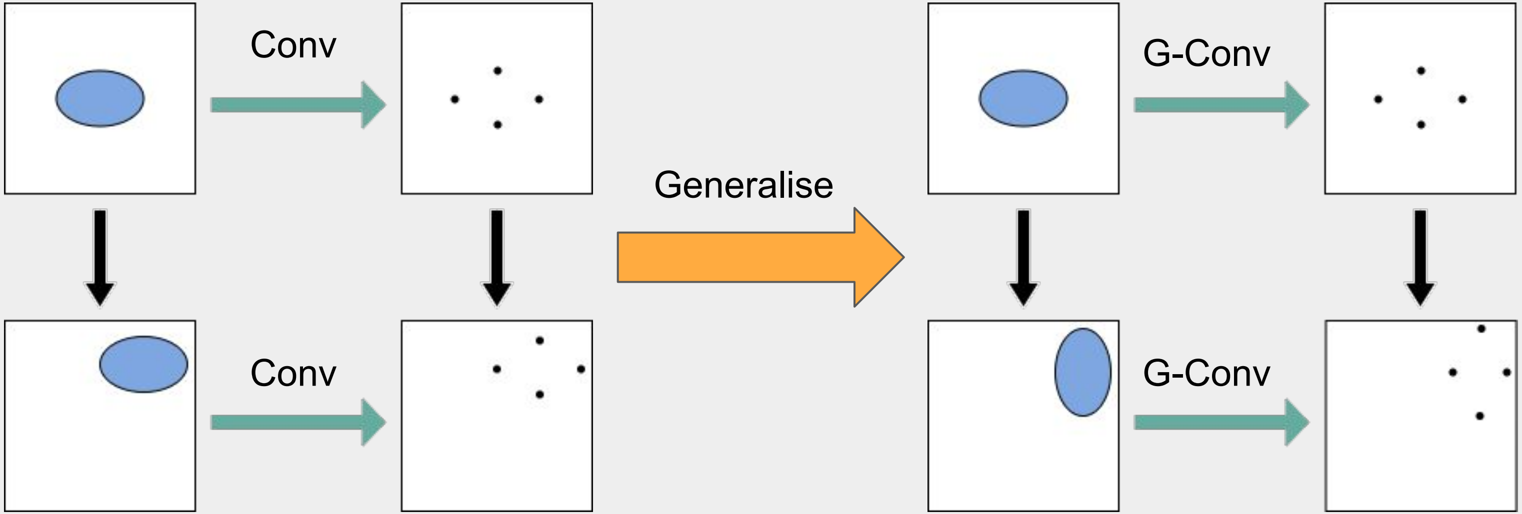


Convolutional Classifiers

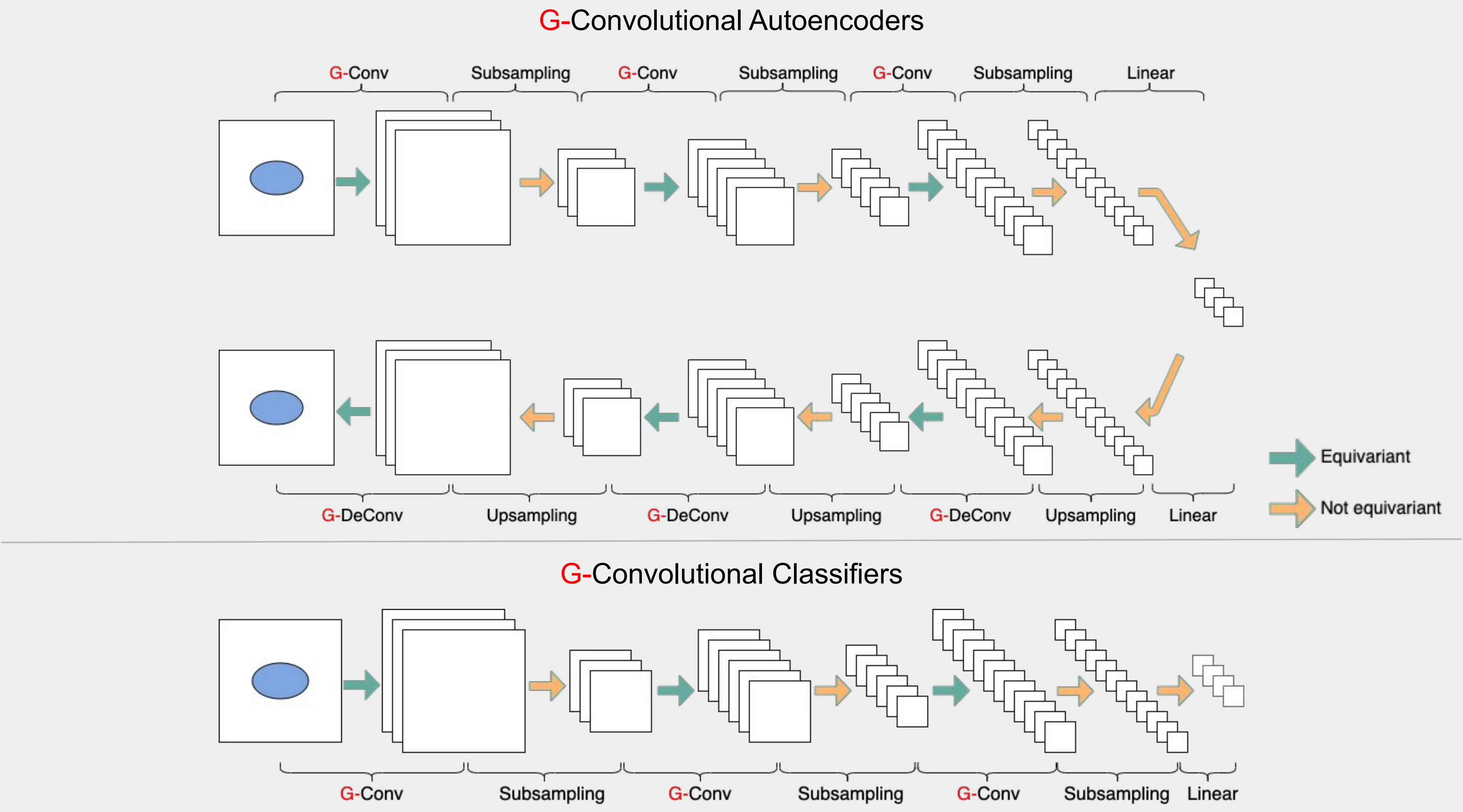


Translational Equivariance

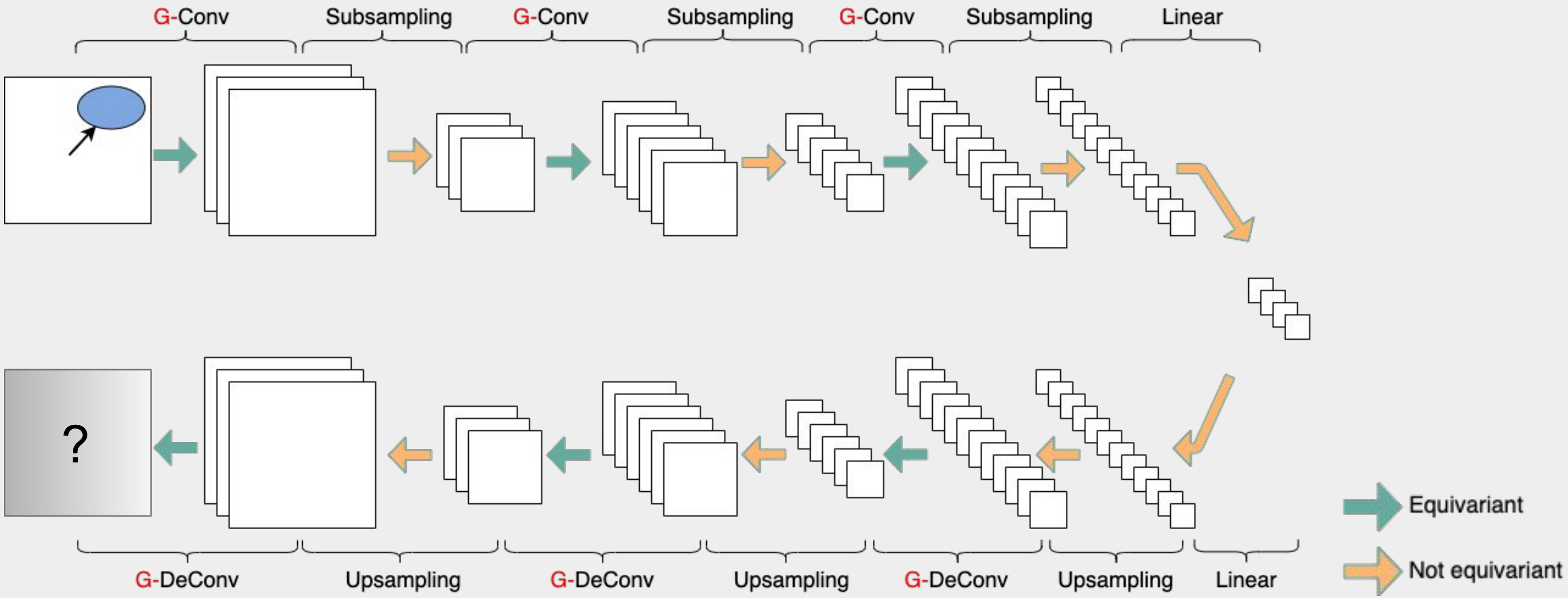
Group Equivariance



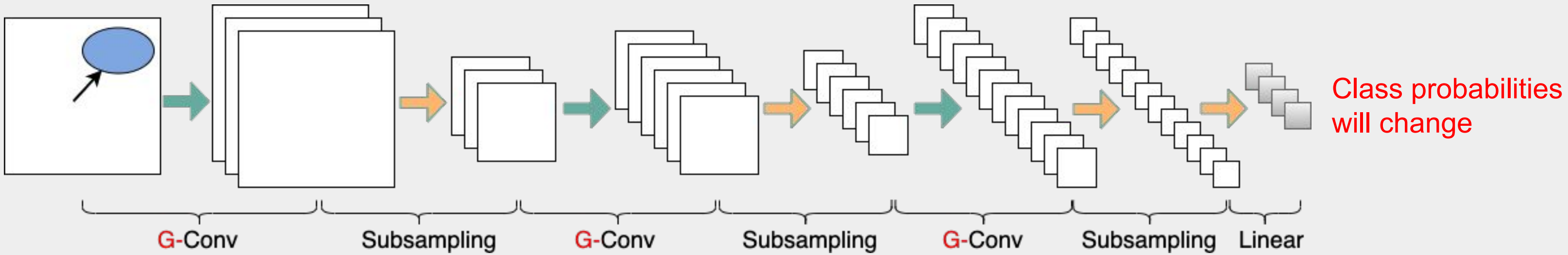
Cohen, T., & Welling, M. (2016, June). Group equivariant convolutional networks. In International conference on machine learning (pp. 2990-2999). PMLR.


G-Convolutional Autoencoders
G-Conv
Subsampling
G-Conv
Subsampling
G-Conv
Subsampling
Linear
G-DeConv
Upsampling
G-DeConv
Upsampling
G-DeConv
Upsampling
Linear
Equivariant
Not equivariant
G-Convolutional Classifiers
G-Conv
Subsampling
G-Conv
Subsampling
G-Conv
Subsampling
Linear

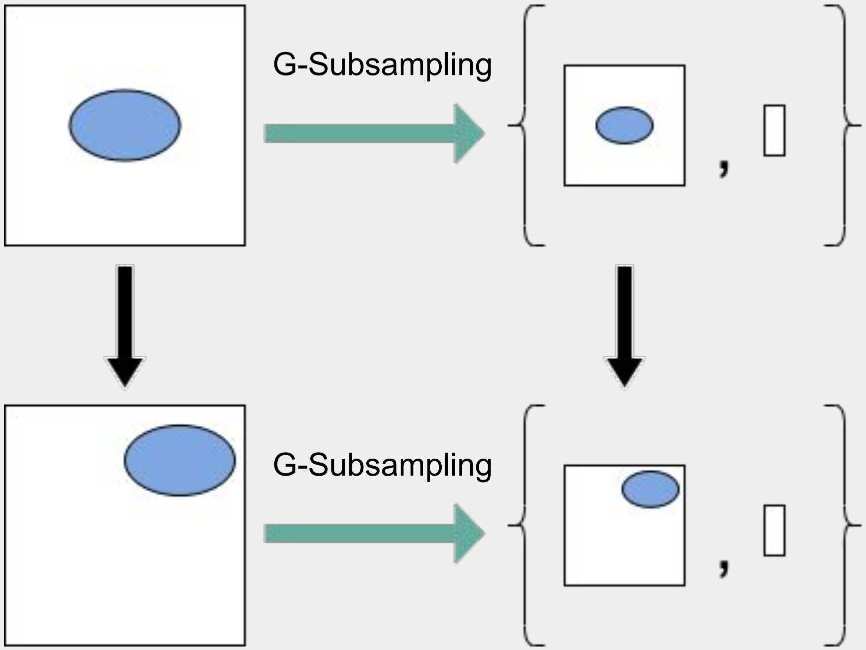
G-Convolutional Autoencoders



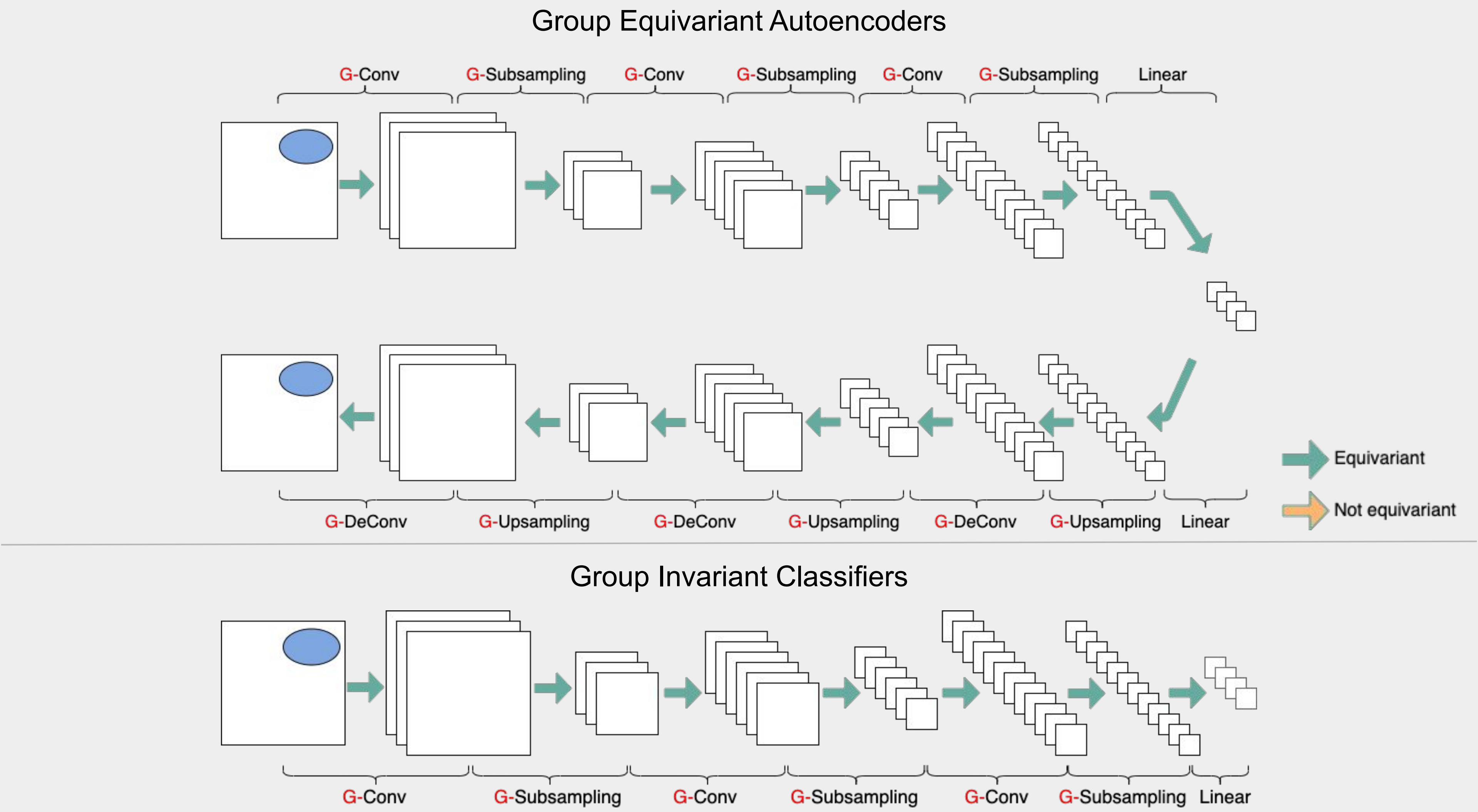
G-Convolutional Classifiers



Group Equivariant Subsampling/Upsampling

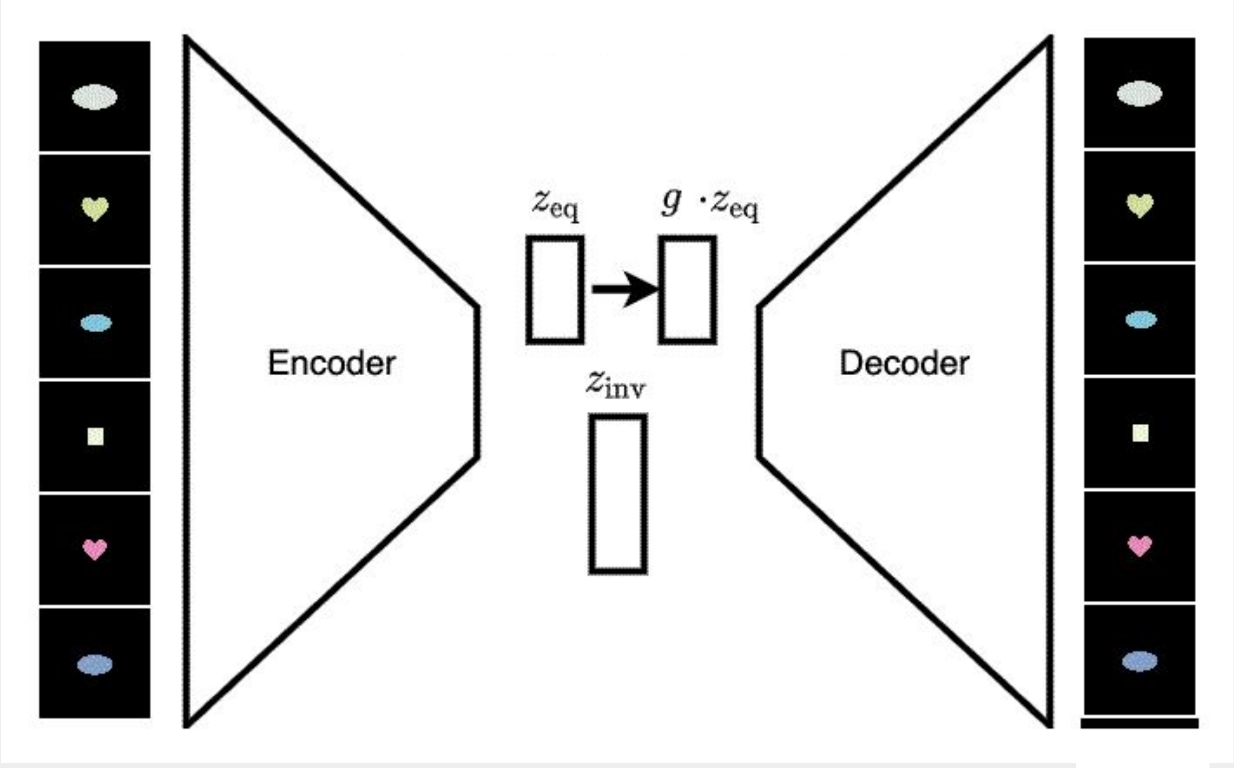


In this work, we propose group equivariant subsampling/upsampling (G-subsampling/G-upsampling) layers.



Group Equivariant Autoencoders
G-Conv
G-Subsampling
G-Conv
G-Subsampling
G-Conv
G-Subsampling
Linear
G-DeConv
G-Upsampling
G-DeConv
G-Upsampling
G-DeConv
G-Upsampling
Linear
Equivariant
Not equivariant
Group Invariant Classifiers
G-Conv
G-Subsampling
G-Conv
G-Subsampling
G-Conv
G-Subsampling
Linear

Application: Group Equivariant Autoencoders (GAEs)



Max-Pooling & Strided Convolutions

Kernel size

Strides

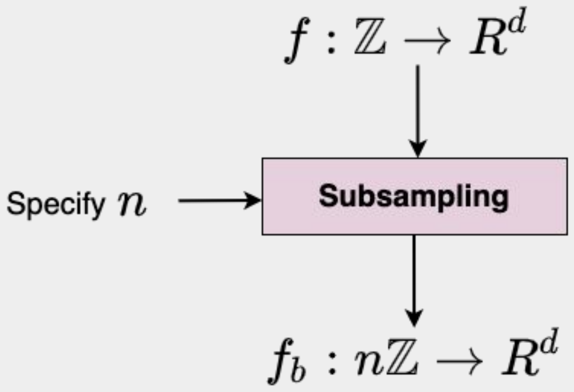
Scale factor

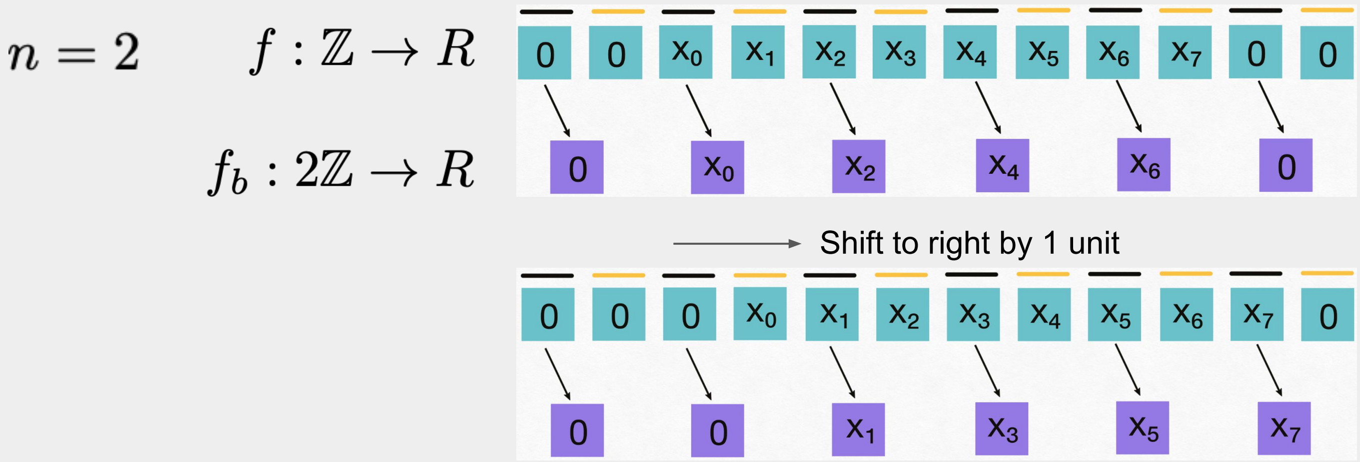
$$\mathrm{CONV}_k^c = \mathrm{SUBSAMPLING}^c \circ \boxed{\mathrm{CONV}_k^1}$$

$$\mathrm{MAXPOOL}_k^c = \mathrm{SUBSAMPLING}^c \circ \boxed{\mathrm{MAXPOOL}_k^1}$$

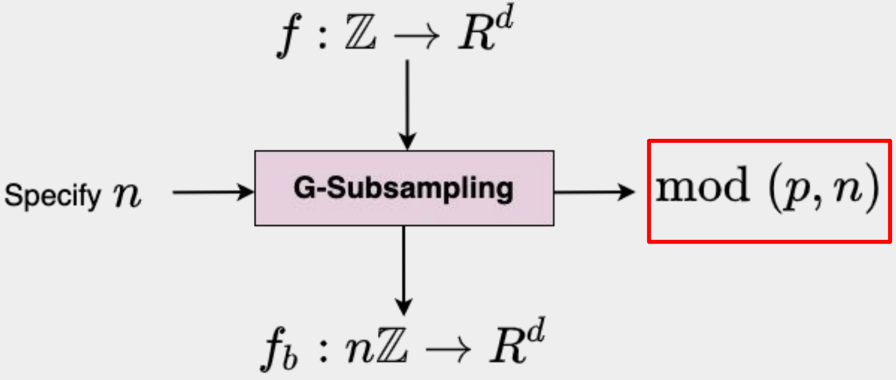
Both are equivariant operations

1D Translation Case of Conventional Subsampling

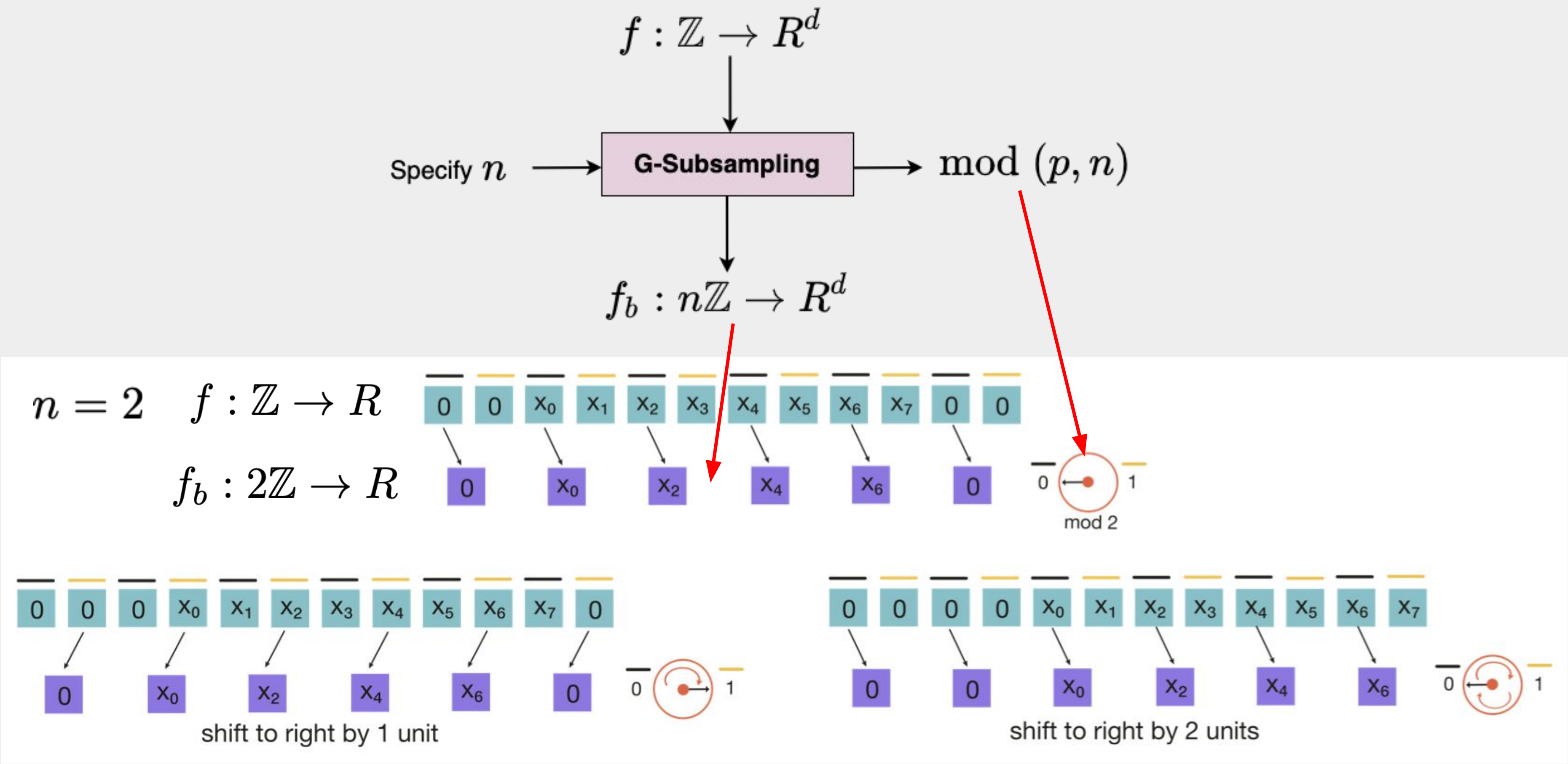




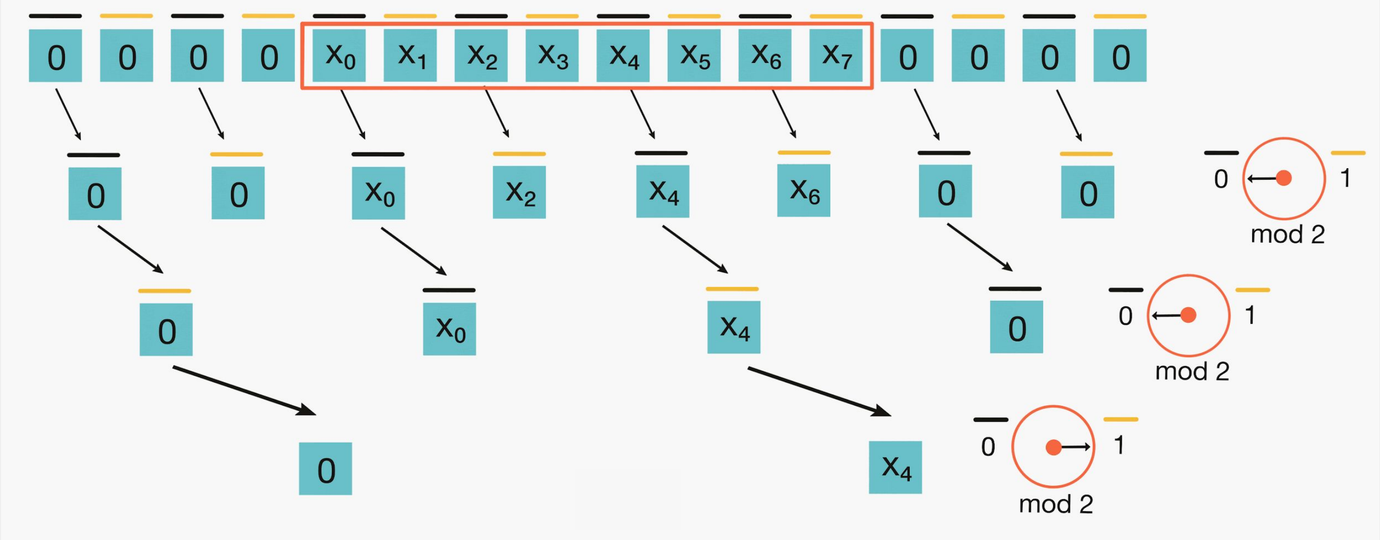
1D Translation Case of G-Subsampling



1D Translation Case of G-Subsampling

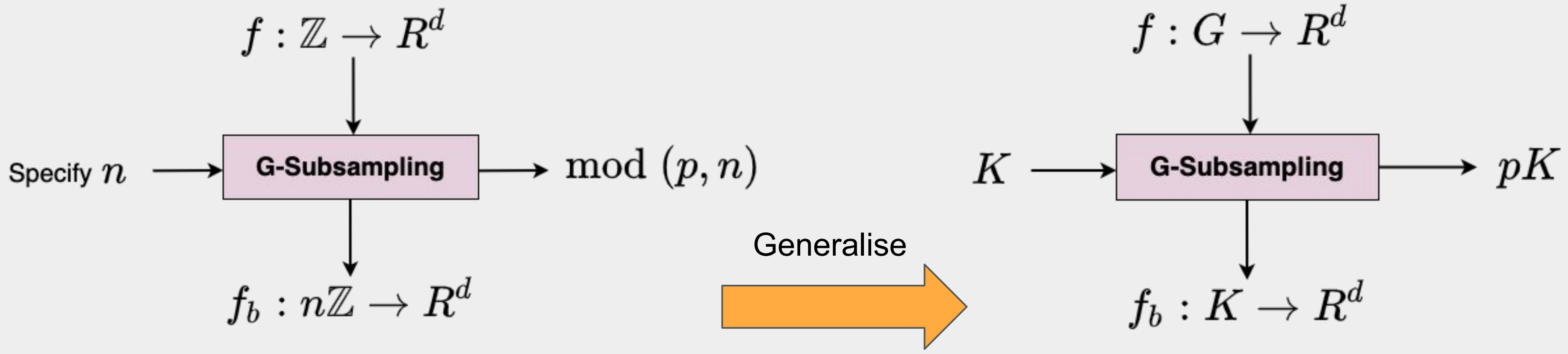


Multiple Layers of G-Subsampling

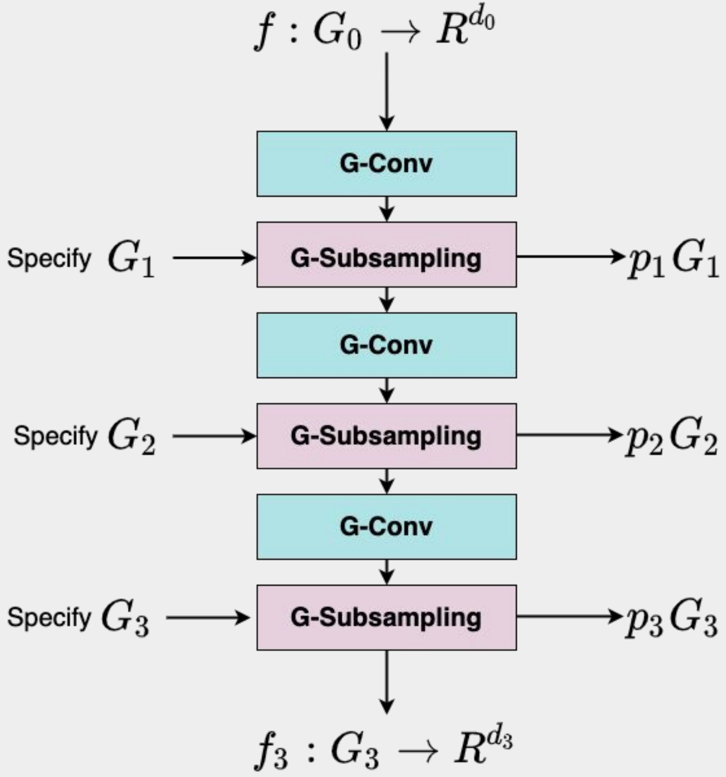


(Equivariant convolutional layers inserted between subsampling layers are omitted in this illustration.)

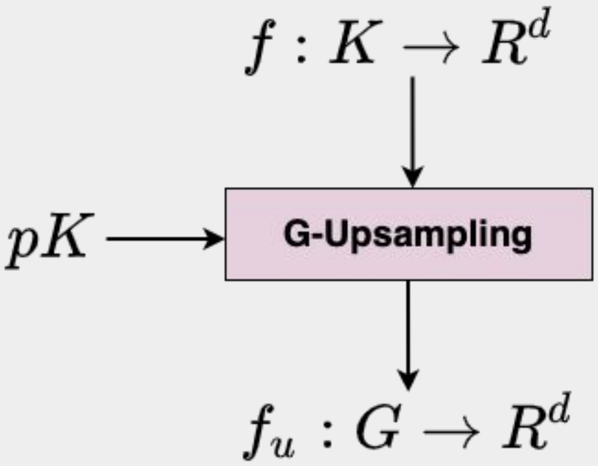
General Case of G-Subsampling



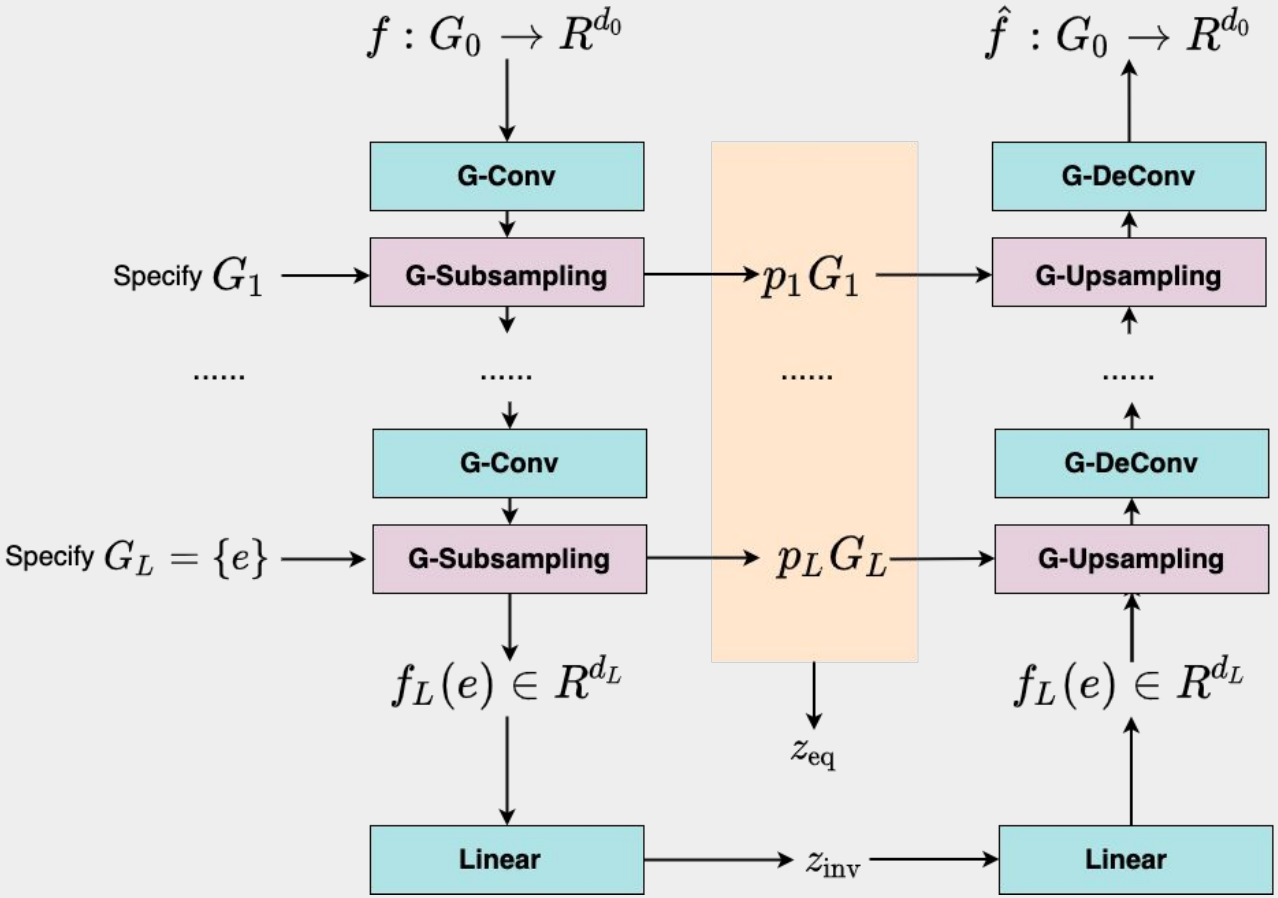
Multiple Layers of G-Subsampling



G_0, G_1, G_2, G_3 is a sequence of nested subgroups.


$f : K \to R^d$
pK
G-Upsampling
$f_u : G \to R^d$

Application: Group Equivariant Autoencoders (GAEs)



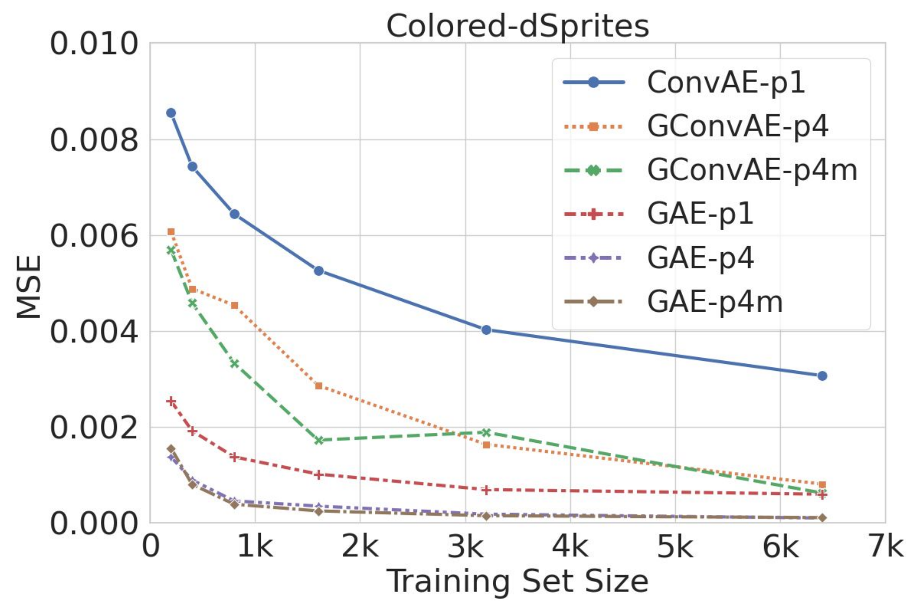
Variants of Autoencoders in Comparison

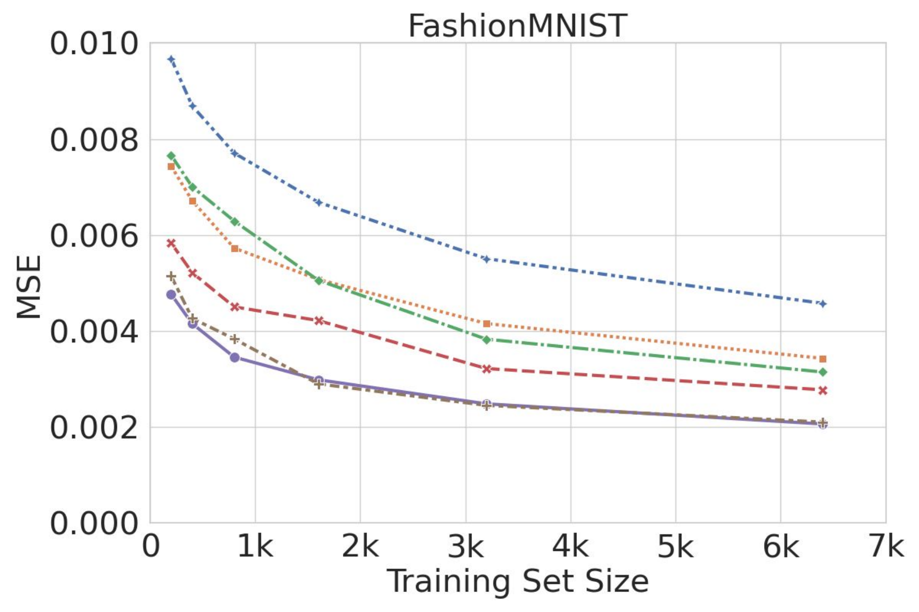
Model	ConvAE-p1	GConvAE-p4	GConvAE-p4m	GAE-p1	GAE-p4	GAE-p4m
Conv Layer	Conv (p1)	GConv (p4)	GConv (p4m)	Conv (p1)	GConv (p4)	GConv (p4m)
Subsampling Layer	Strides in Conv	Strides in Conv	Strides in Conv	G-Subsampling (p1)	G-Subsampling (p4)	G-Subsampling (p4m)

Wallpaper groups:

- p1: All 2D integer translations
- p4: All compositions of 2D integer translations and rotations by a multiple of 90 degrees.
- p4m: All compositions of elements in p4 and the mirror reflection

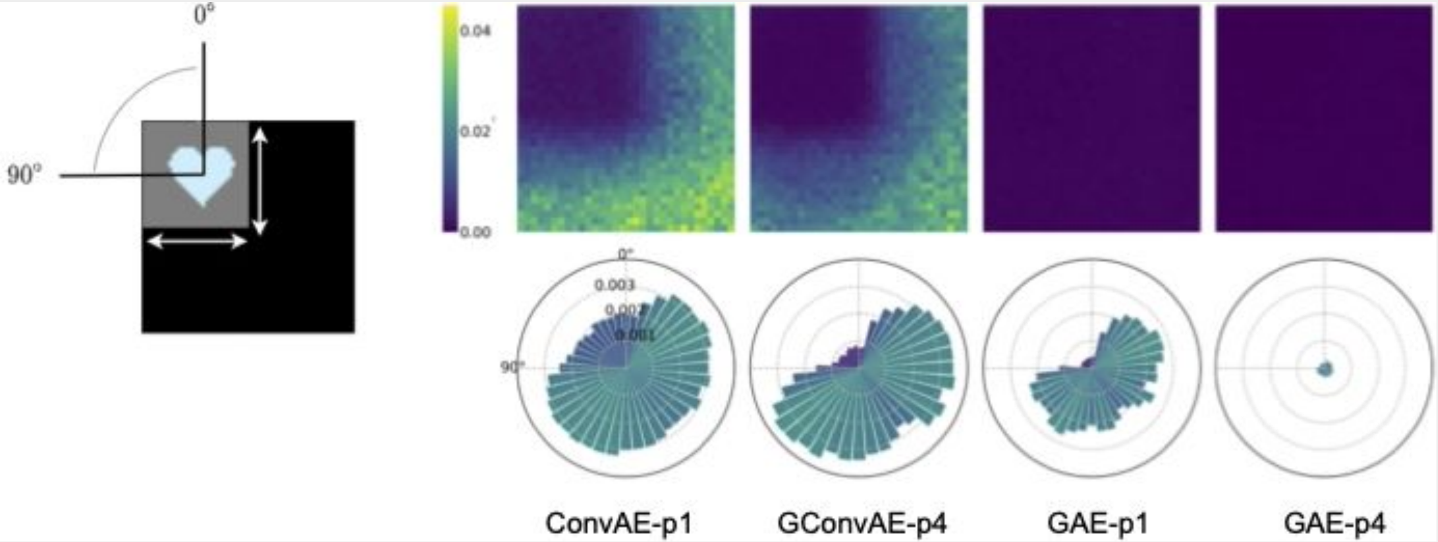
GAE Reconstruction Error on Single Object Datasets





Model	ConvAE-p1	GConvAE-p4	GConvAE-p4m	GAE-p1	GAE-p4	GAE-p4m
Conv Layer	Conv	GConv (p4)	GConv (p4m)	Conv	GConv (p4)	GConv (p4m)
Subsampling Layer	Strides in Conv	Strides in Conv	Strides in Conv	G-Subsampling (p1)	G-Subsampling (p4)	G-Subsampling (p4m)

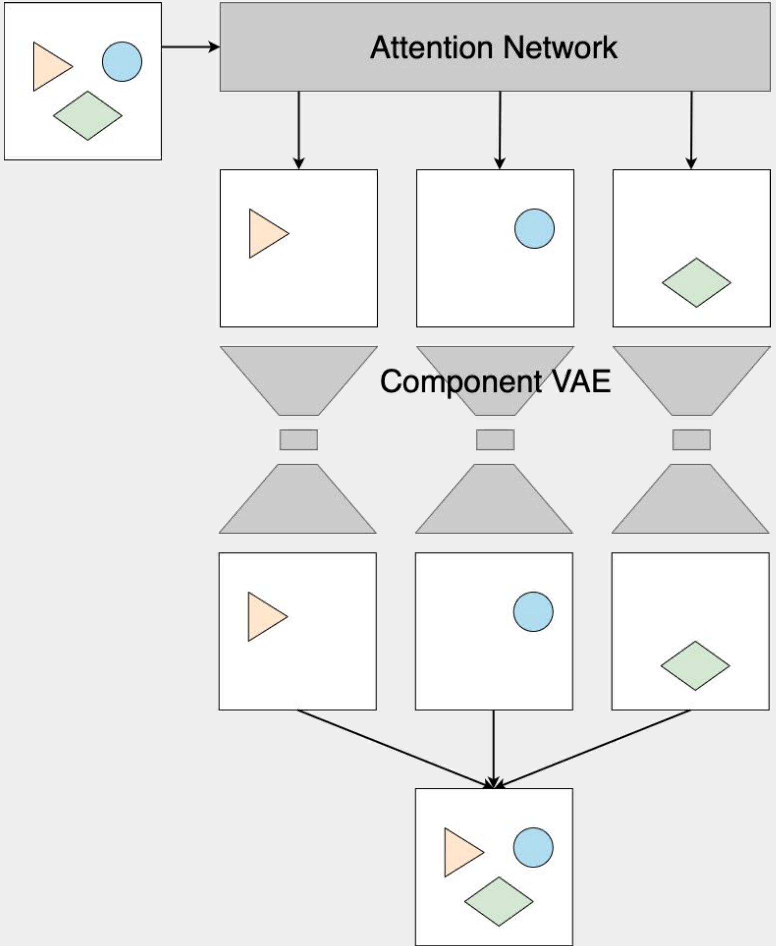
Generalisation to out-of-distribution object locations and poses



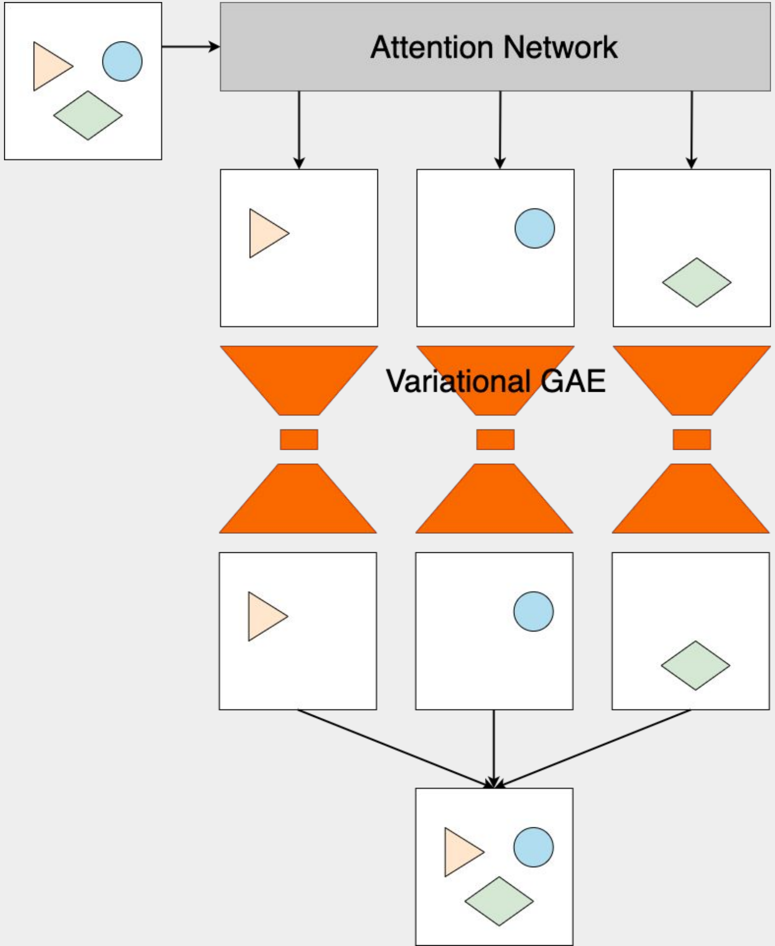
During training, we constrain shapes to be in the top-left quarter, and the orientation to be always less than 90 degrees.

On the right, we compare the error of reconstructions of different models generalise on objects at unseen locations in the first row, and how they generalise to unseen orientations in the second row.

Unsupervised Scene Decomposition with MONet



Unsupervised Scene Decomposition with MONet



Experiments: Multiple Objects

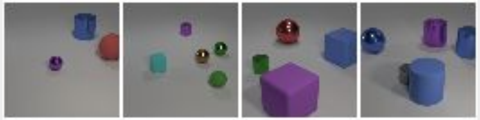


Table 1: Reconstruction error MSE ($\times 10^{-3}$) (mean(stddev) across 5 seeds) on multi-object datasets

Dataset	Multi-dSprites			CLEVR6		
Training Set Size	3200	6400	12800	3200	6400	12800
MONet	2.661(0.382)	1.385(0.235)	0.326(0.076)	0.673(0.059)	0.562(0.057)	0.546(0.056)[1]
MONet-GAE-$p1$	0.659(0.103)	0.359(0.025)	0.264(0.042)	0.473(0.064)	0.432(0.052)	0.388(0.016)
MONet-GAE-$p4$	0.563(0.195)	0.317(0.060)	0.231(0.067)	0.461(0.025)	0.414(0.022)	0.413(0.018)

Table 2: Foreground segmentation performance in terms of ARI (mean(stddev) across 5 seeds)

Dataset	Multi-dSprites			CLEVR6		
Training Set Size	3200	6400	12800	3200	6400	12800
MONet	0.597(0.022)	0.747(0.049)	0.891(0.009)	0.829(0.055)	0.878(0.023)	0.865(0.033)[1]
MONet-GAE-$p1$	0.762(0.049)	0.823(0.042)	0.889(0.013)	0.921(0.015)	0.917(0.032)	0.920(0.025)
MONet-GAE-$p4$	0.753(0.089)	0.833(0.072)	0.902(0.025)	0.878(0.055)	0.914(0.012)	0.910(0.011)

MSE: Measure the overall reconstruction quality, the lower the better.
ARI: Measure the (foreground) object segmentation performance, the higher the better.

Conclusions

- We have proposed subsampling/upsampling operations that preserve group equivariance.
- We have used the proposed layers to construct exact group equivariant autoencoders.
- We have shown that the proposed subsampling/upsampling layers can improve sample efficiency in both single-object and multi-object representation learning models.