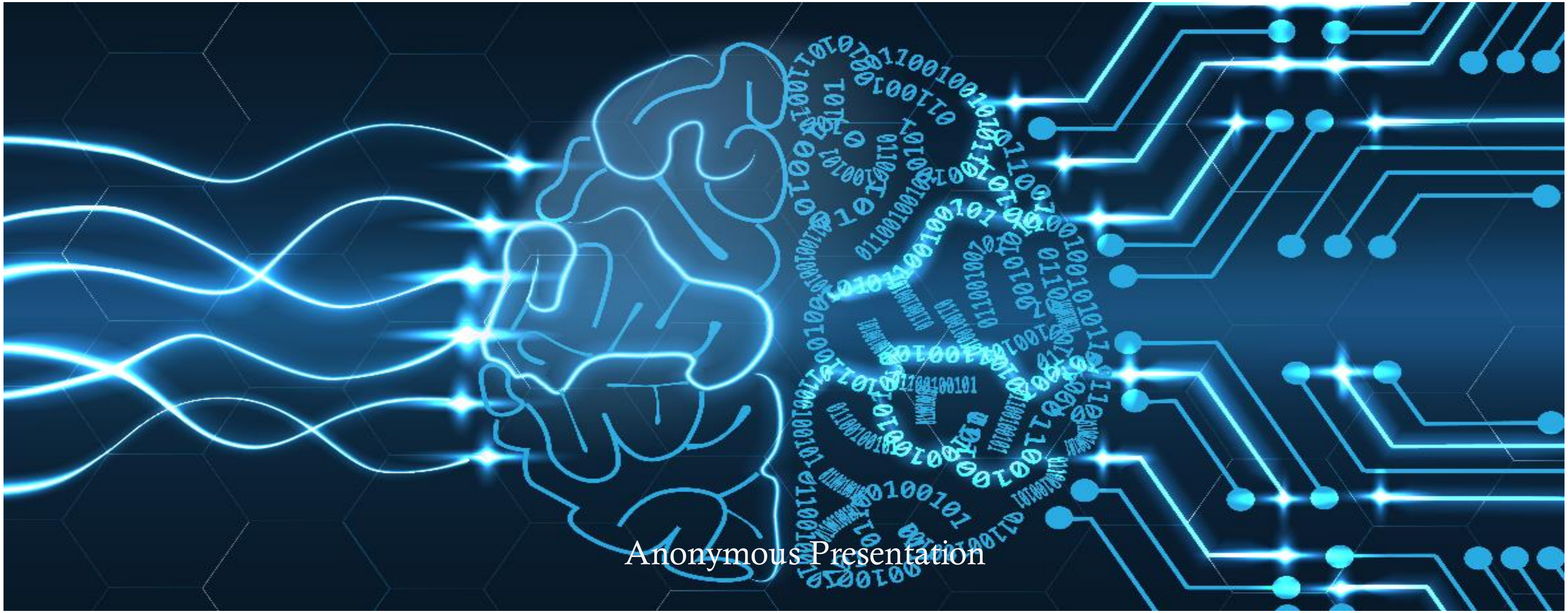
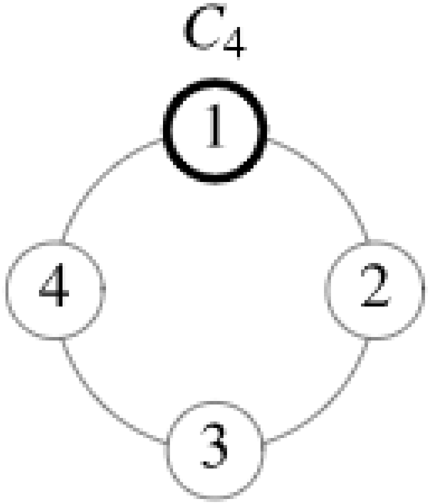
EFFICIENT EQUIVARIANT SUBSAMPLING

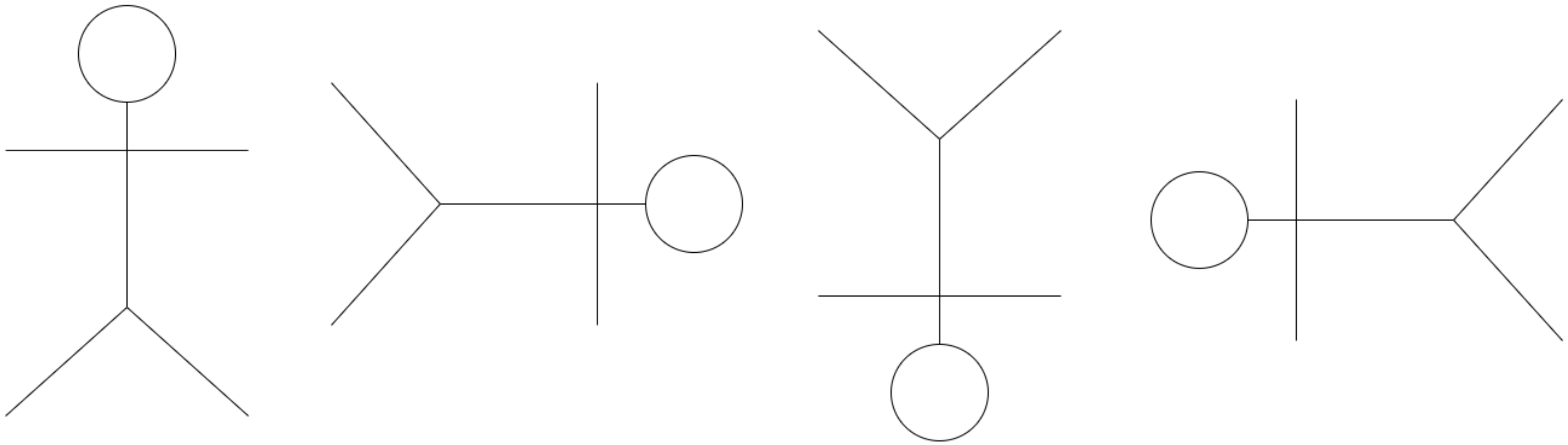


Anonymous Presentation

Definition: Groups and the Cyclic Group 4 (C4)

- A group is defined as a set and an operation that maintain closure and associativity over the group operation and possess an identity element and an inverse function
- Cyclic Group 4 (C4) contains the transforms for rotating images along 90-degree angles

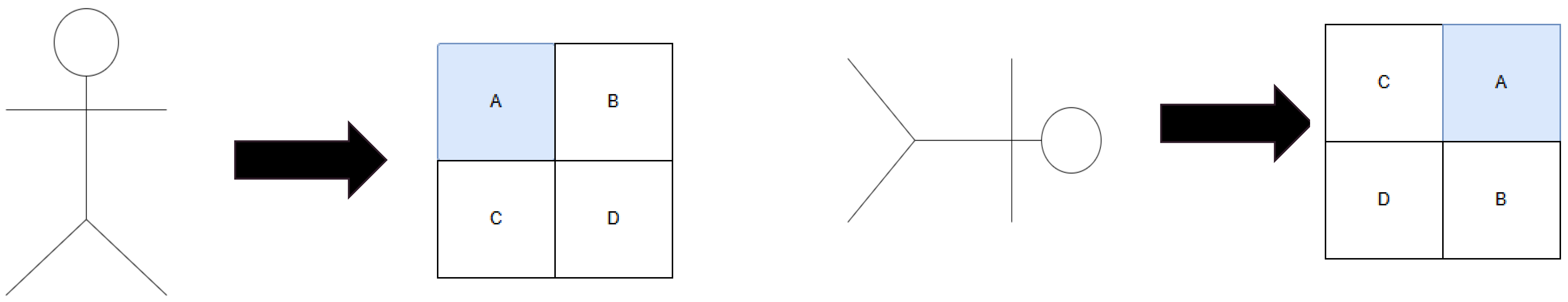




Definition: Group Equivariance

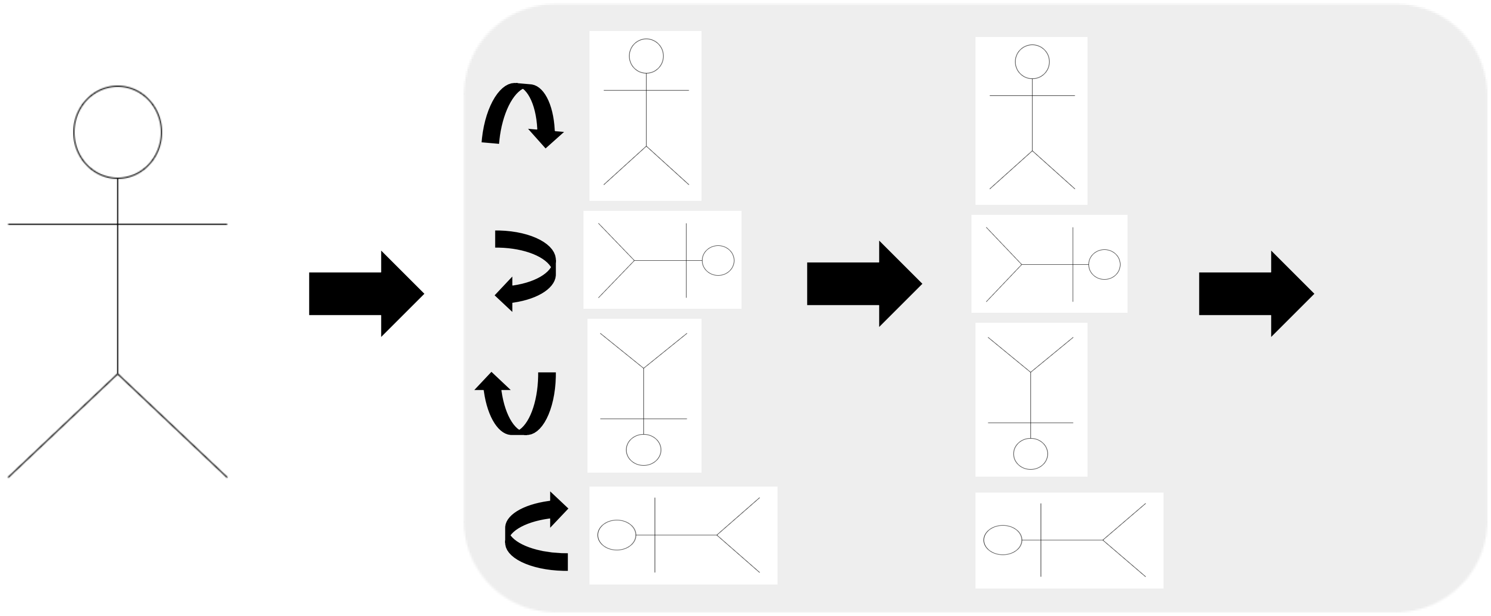
- Applying the group action to a feature before feeding it to a network should produce the same results as providing the input to the network and then applying the group action to the output.

$$\forall u \in \mathcal{G}, \qquad \Psi\left[\mathcal{T}_u[\mathbf{f}]\right] = \mathcal{T}'_u[\Psi[\mathbf{f}]]$$



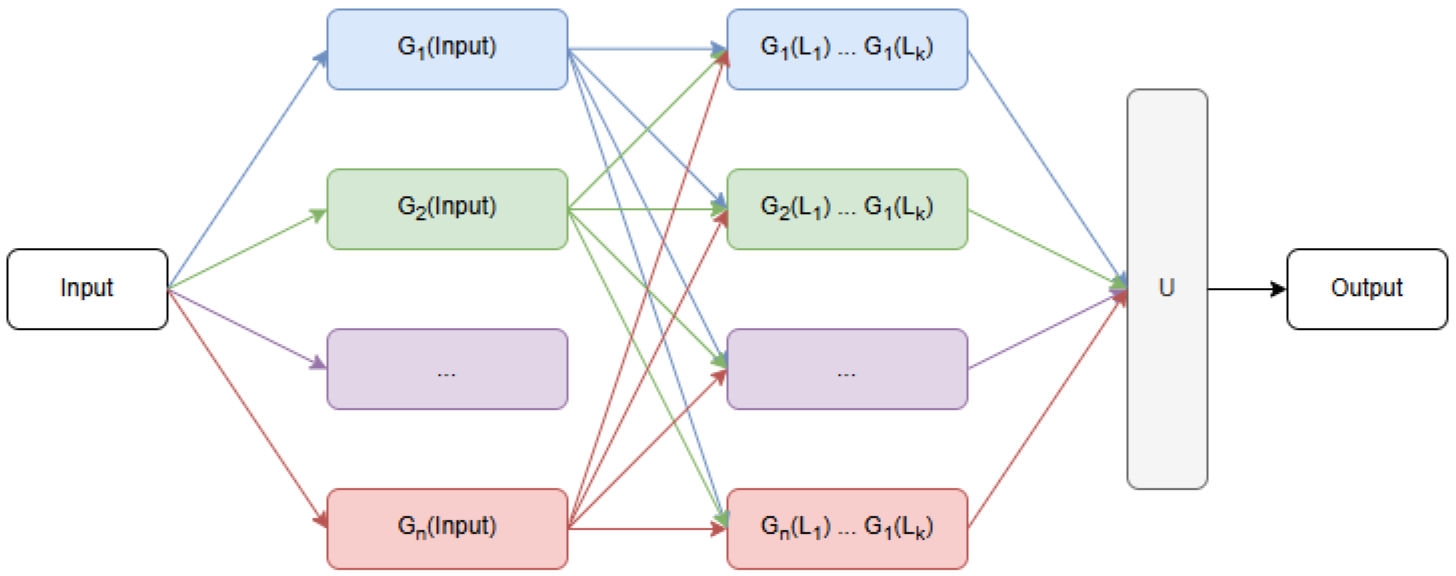
Background: Group Equivariant Networks

- Group Equivariant networks map the group representation of an input from layer to layer across the network
- Lifting operations add a new dimension that contains every representation of the input across the group
- The final layers break equivariance as the merge to form a classification output



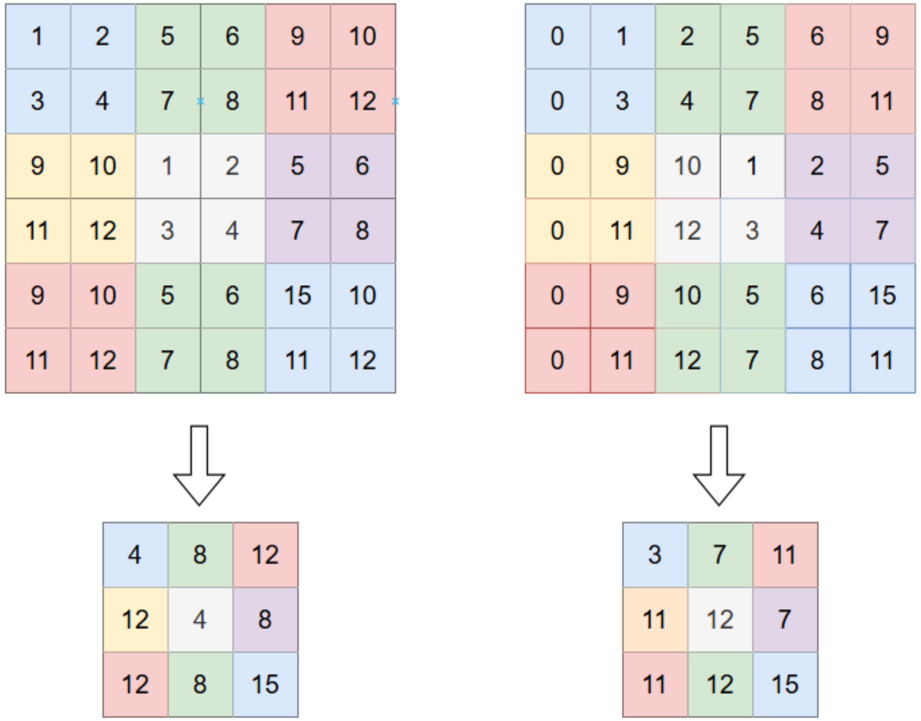
Background: Group Equivariant Networks Drawbacks and Limitations

- Group to group mappings are computationally expensive
- Each element requires a mapping function to every element of the next layer in the group
- 4 groups means 16 separate maps
- Each map includes the typical channel dimensions for encoding different features of the input
- Number of Group Elements * Number of Group Elements * Channels



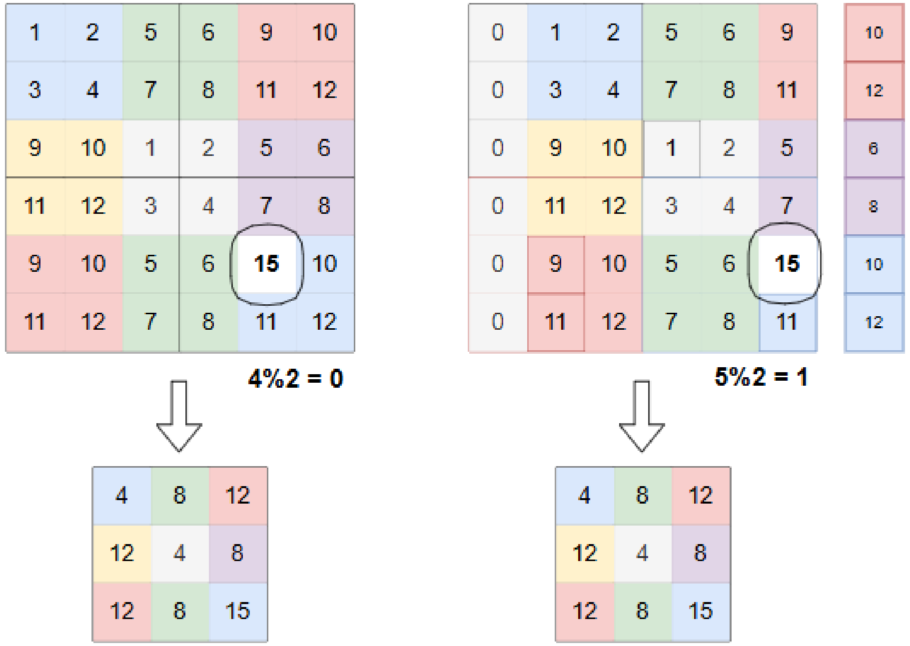
Background: Sub Sampling

- Subsampling functions map blocks of the input feature to a reduction function (avg, max, etc.)
- Subsampling breaks translation equivariance



Background: Equivariant Subsampling

- Find a reference point on an input image to base the location of pooling kernel
- When the reference point shifts by i, the kernels are shifted by i



$$i = \Phi_c(f) = \mathrm{mod}(\arg\max_{x \in \mathbb{Z}} \|f(x)\|_1, c)$$

- Argmax function provides equivariance across linear translations. Use i as the offset index to the pooling kernels
- Block[x,y] = max(input[i + block_idx: i + block_idx + 2]

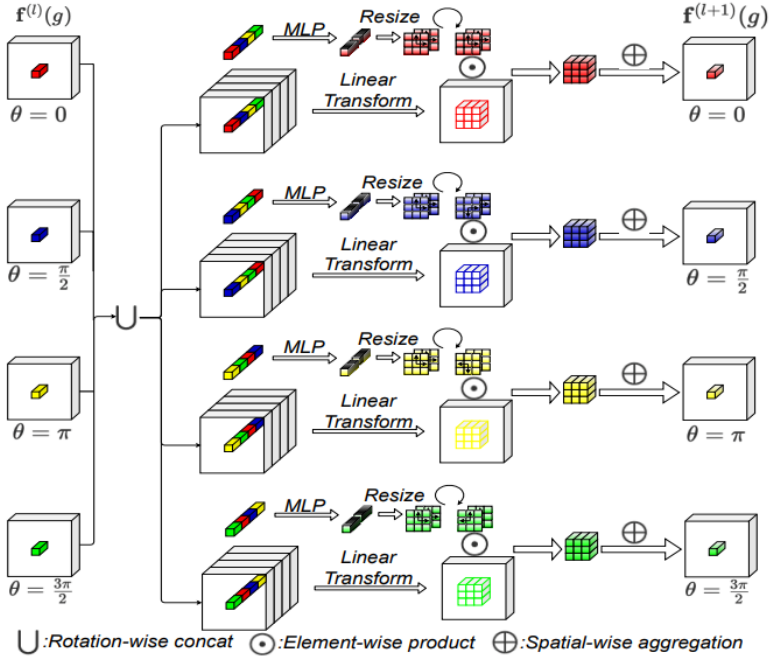
E4 Net

- H can represent the mapping from group element g to group element g'
- The E4 Net decomposes H into a kernel generator K and an encoder V, both of which are MLPs
- The kernel is transformed along the group action allowing the parameters to be reused across the group

$$\mathbf{f}^{(l+1)}(g) = \sum_{\widetilde{g}\in\mathcal{G}} \widetilde{H}_{g^{-1}\widetilde{g}}(\mathbf{f}^{(l)}(g), \mathbf{f}^{(l)}(\widetilde{g}))$$

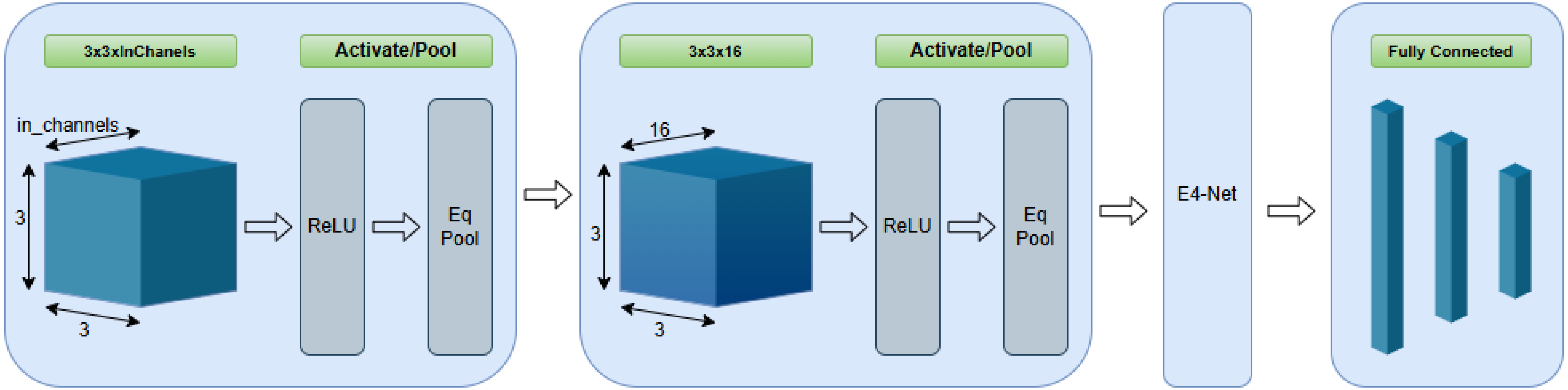
$$\mathbf{f}^{(l+1)}(g) = \sum_{\widetilde{g}\in\mathcal{G}} \widetilde{H}_{g^{-1}\widetilde{g}}(\mathcal{F}_{\mathcal{N}_1(g)}, \mathcal{F}_{\mathcal{N}_2(\widetilde{g})})$$

$$\forall \hat{g} \in \mathcal{G}, \quad \widetilde{H}_{\hat{g}}(x, y) = K_{\hat{g}}(x) \odot V(y)$$



ES4 Net

- Efficient Equivariant Subsampling network combines techniques from traditional CNNs, rotational equivariant networks, and translational equivariant sampling functions
- Two 3x3 convolutional networks are combined with equivariant maxpooling to reduce the input features to a small E4 Network



ES4-Net

Activate/Pool
in_channels
ReLU
Eq Pool
16
E4-Net
Fully Connected
Bird?
Plane?
???

Thank You!