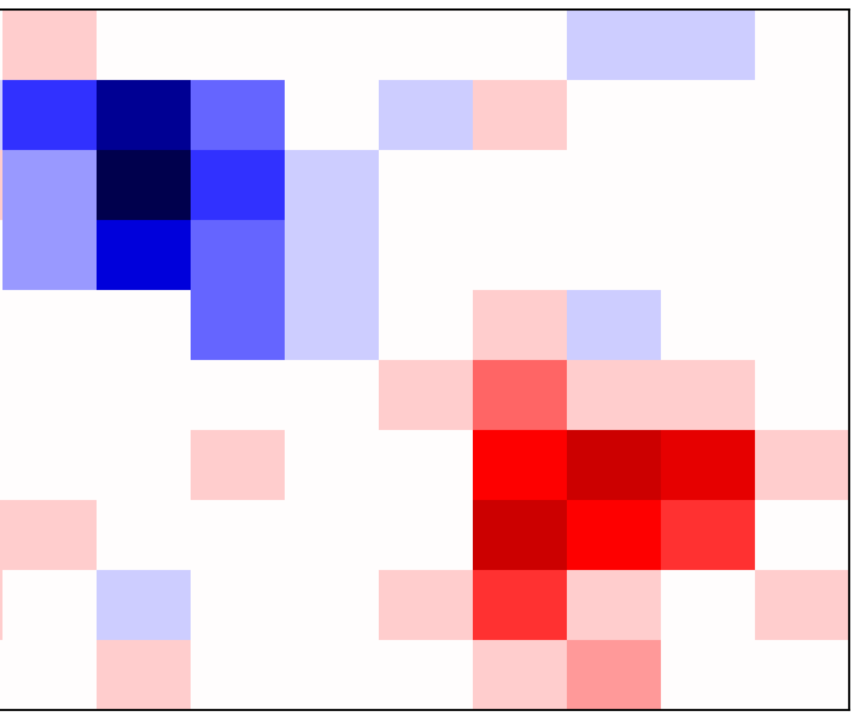
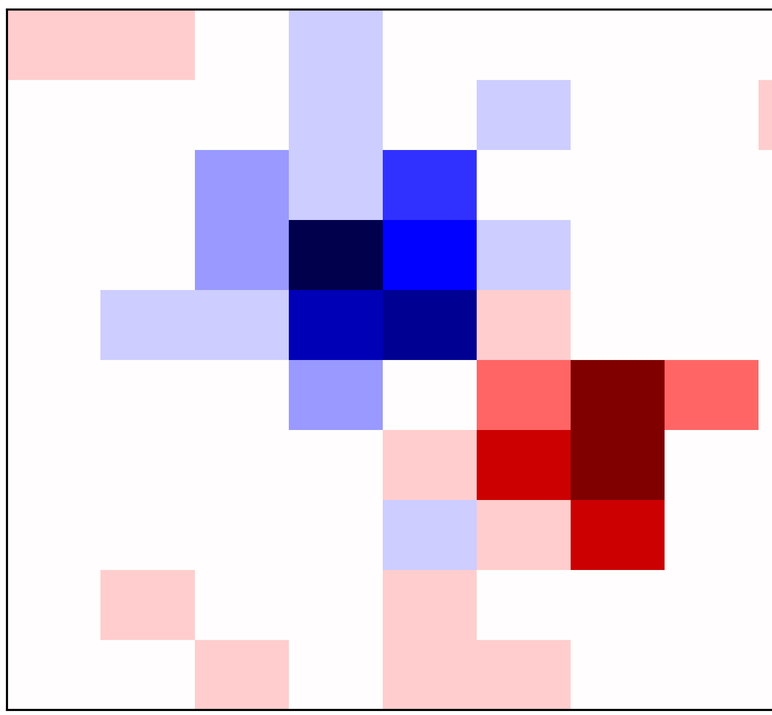
Relativistic Harmonic Networks



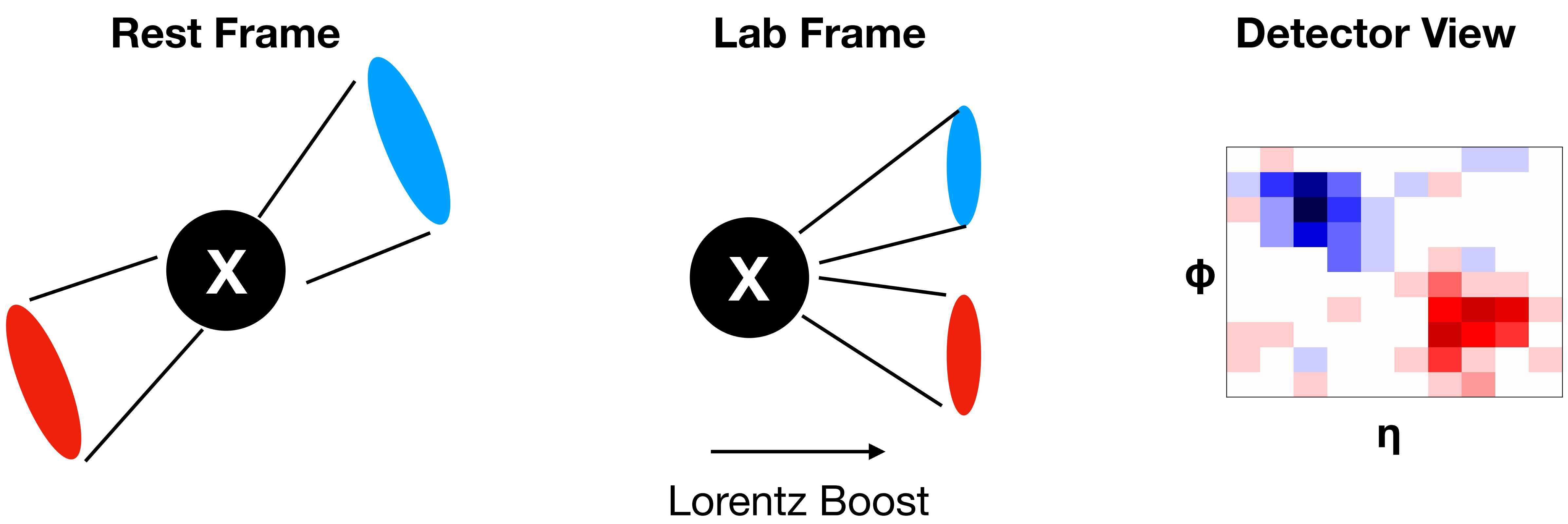
Chase Shimmin (Yale)
Mohammad Abdullah (TAMU)
Paul Tipton (Yale)



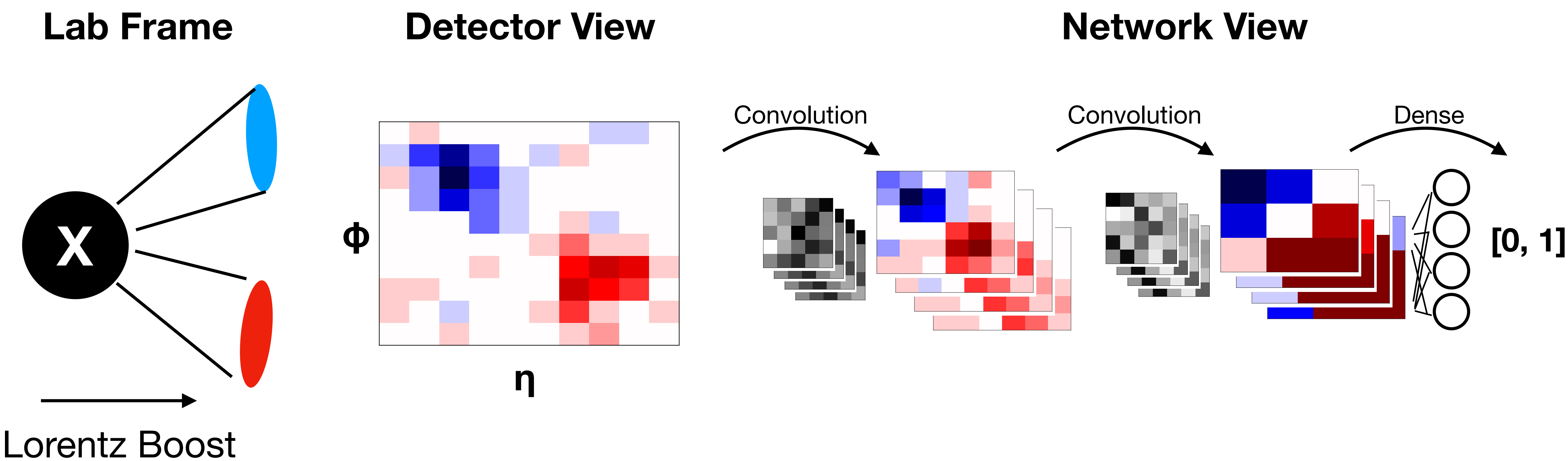
BOOST 2019

Yale University

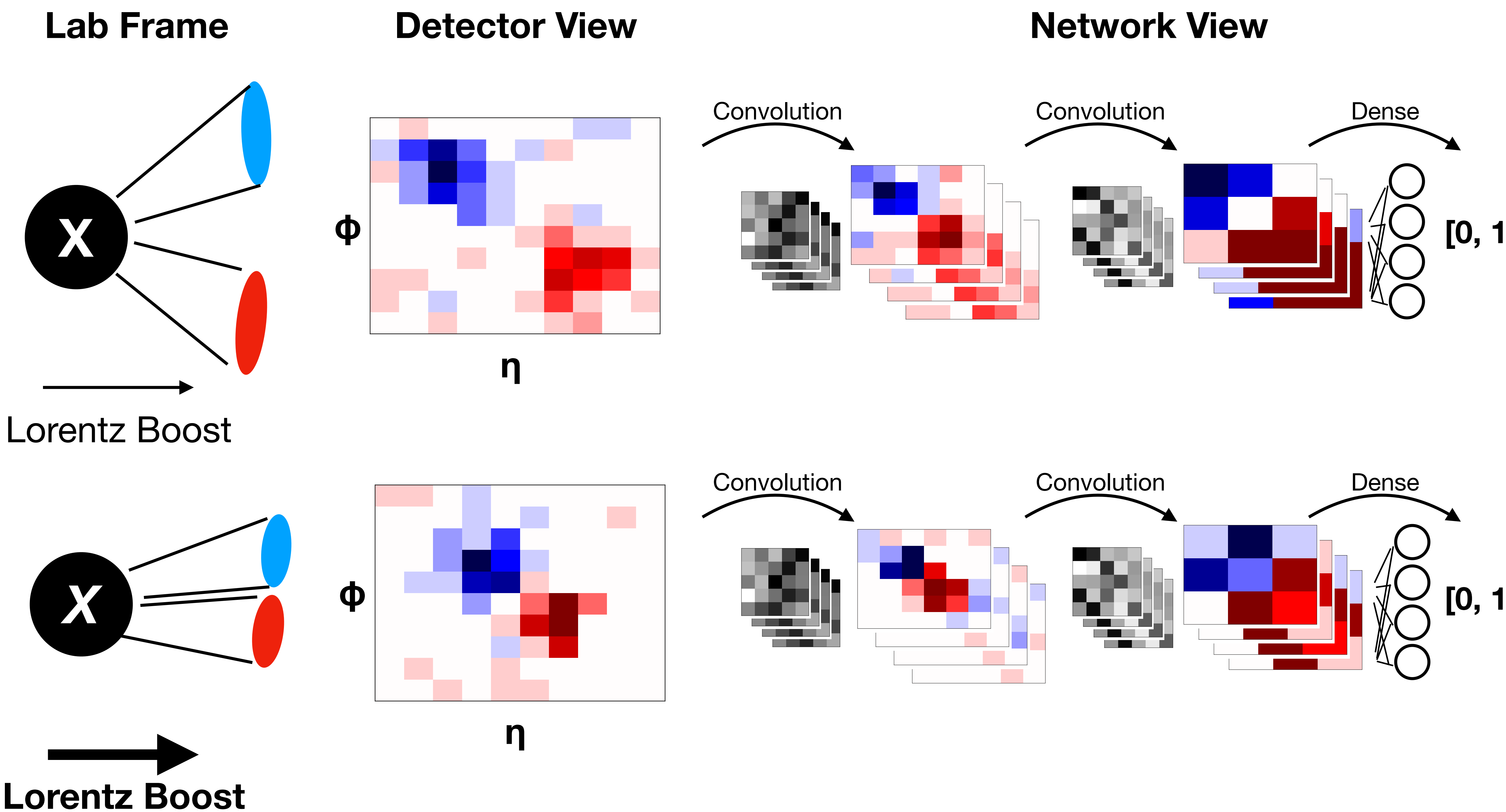
Boosted Jets and Neural 'Nets



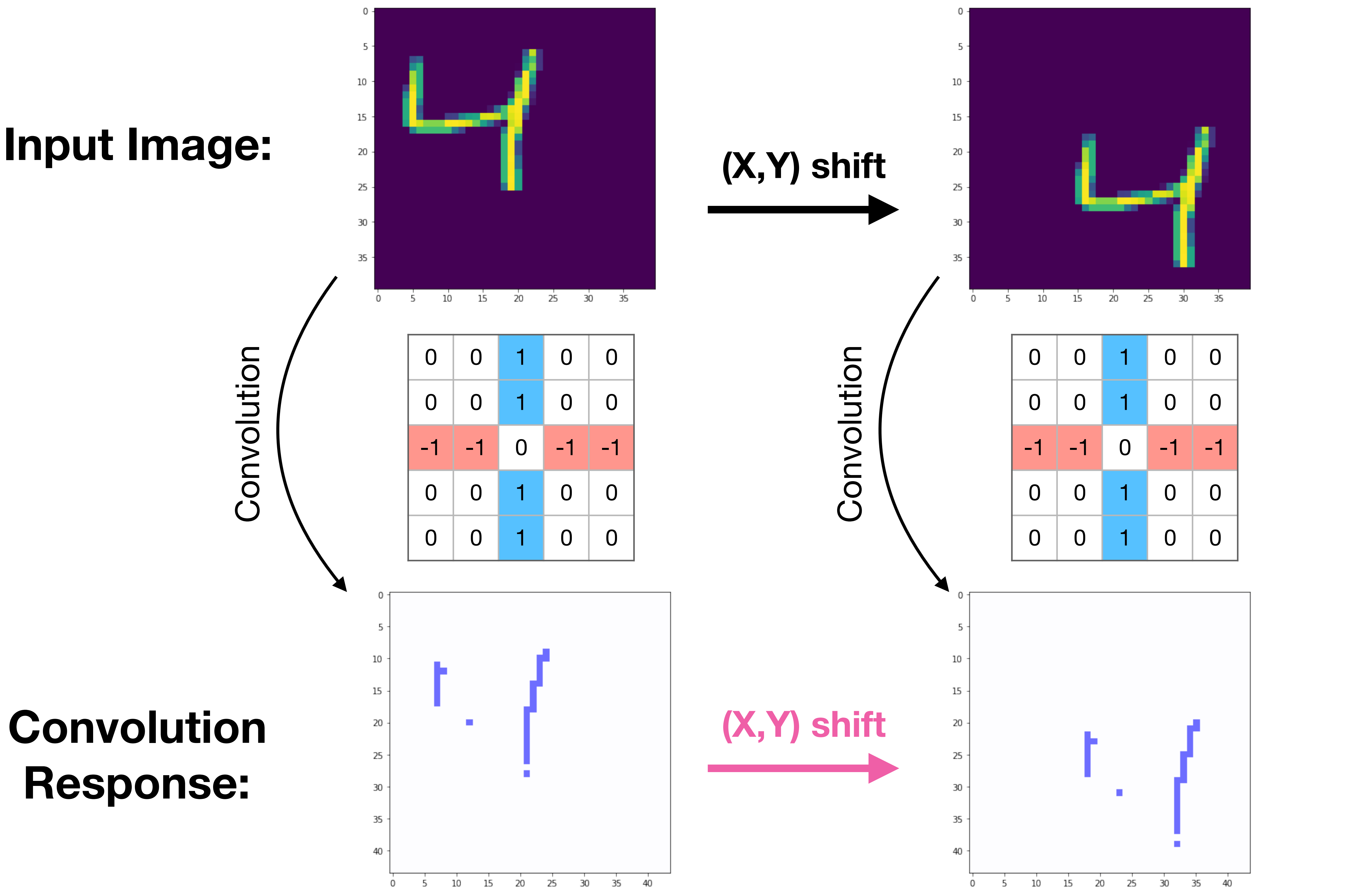
Boosted Jets and Neural 'Nets



Boosted Jets and Neural 'Nets

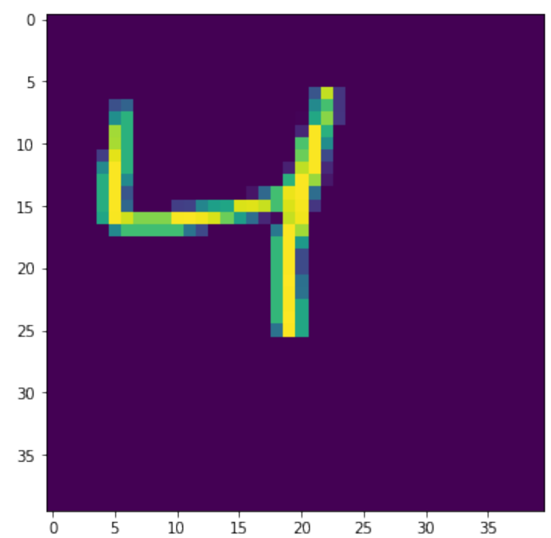


Equivariance by Example

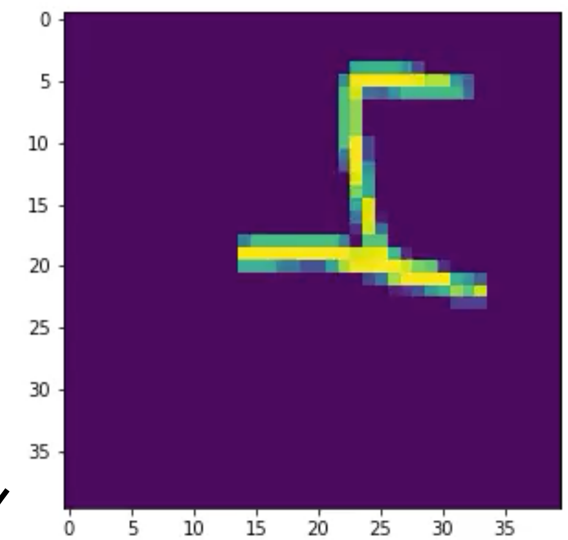


~~Equivariance~~ by Example

Input Image:



Rotation →



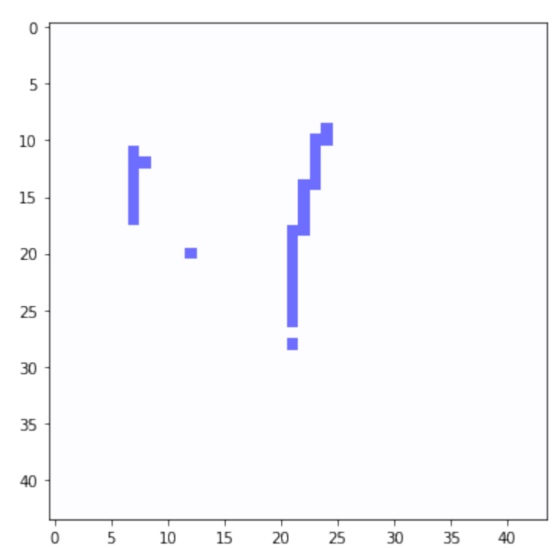
Convolution

0	0	1	0	0
0	0	1	0	0
-1	-1	0	-1	-1
0	0	1	0	0
0	0	1	0	0

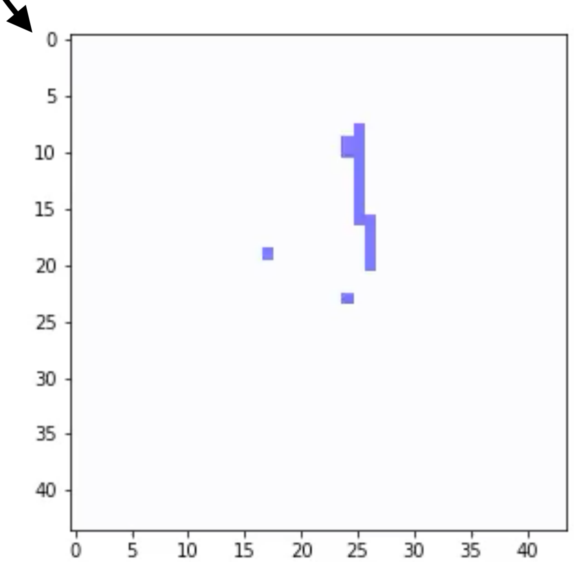
Convolution

0	0	1	0	0
0	0	1	0	0
-1	-1	0	-1	-1
0	0	1	0	0
0	0	1	0	0

Convolution Response:

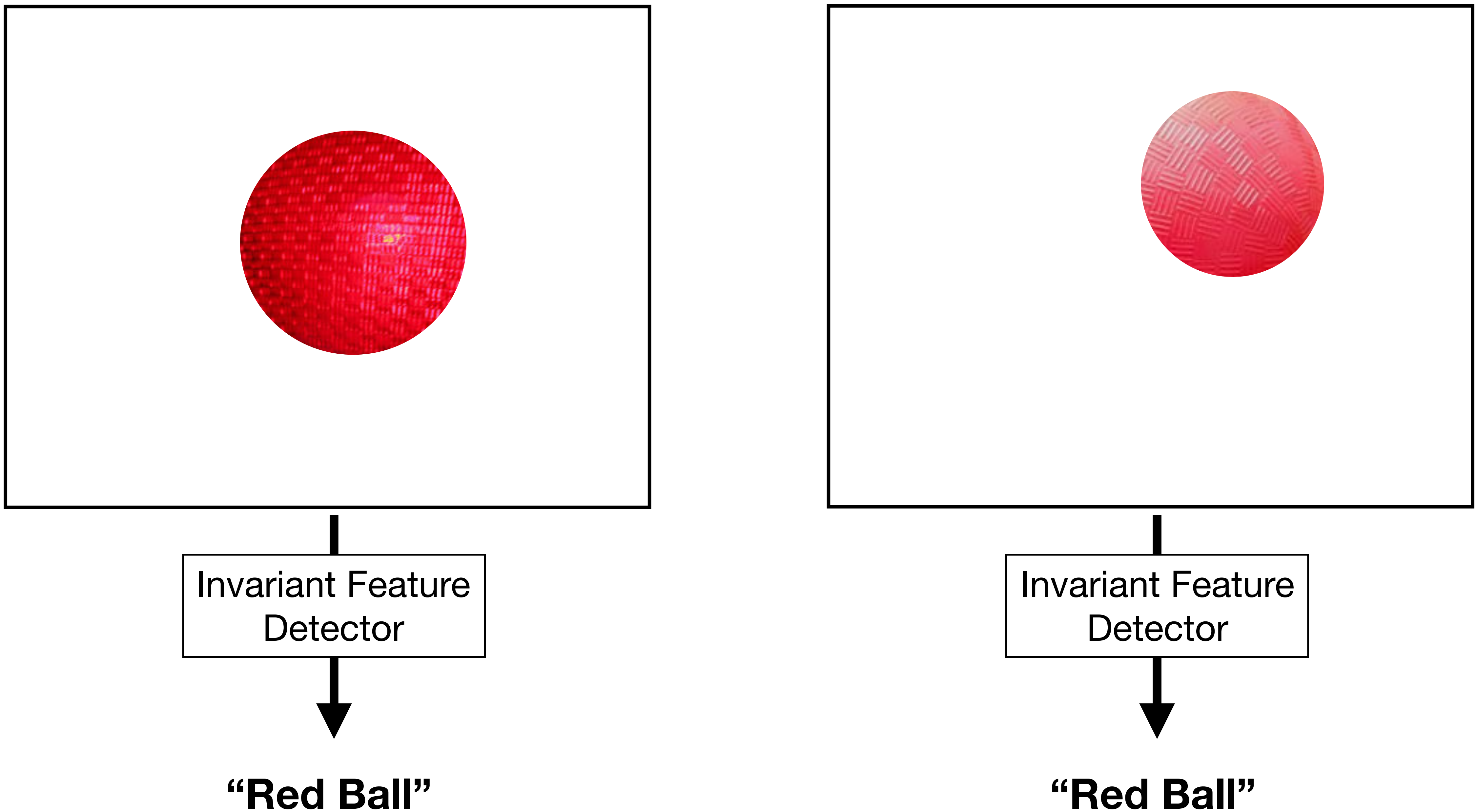


??? →



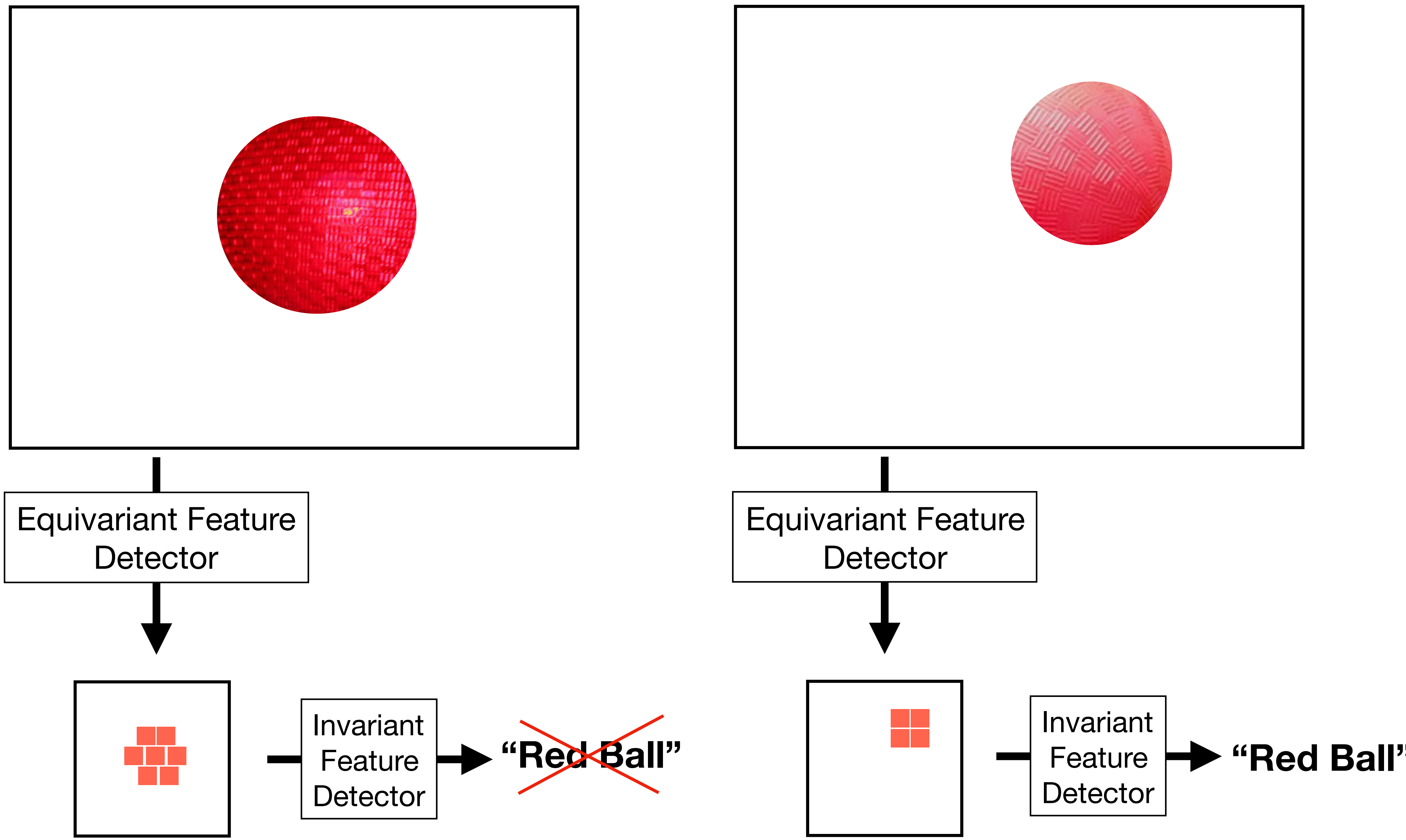
Why Equivariance?

Example: Categorization → Want *Invariant* Response



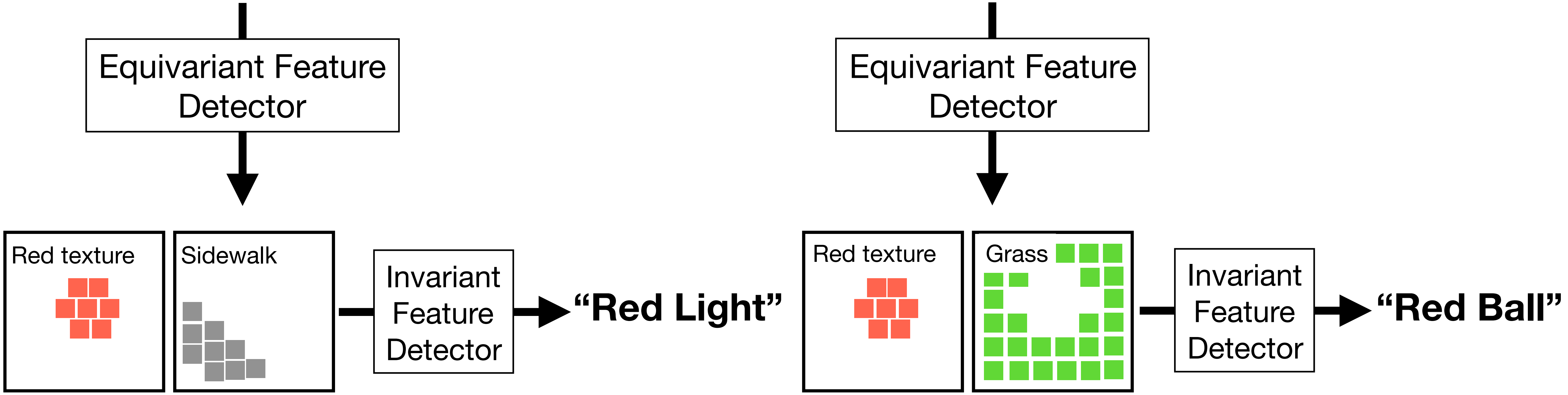
Why Equivariance?

Example: Categorization → Want *Invariant* Response



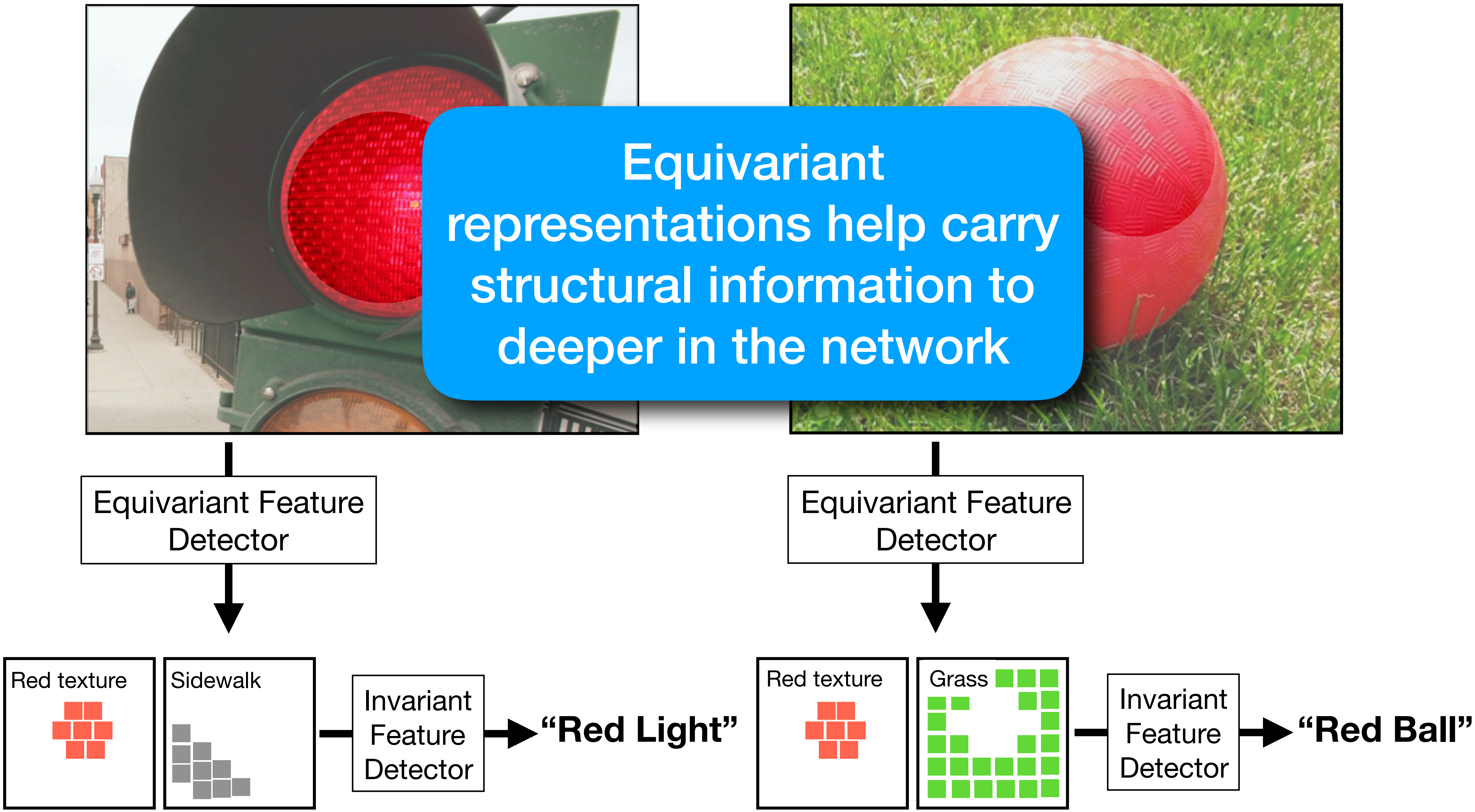
Why Equivariance?

Example: Categorization → Want *Invariant* Response

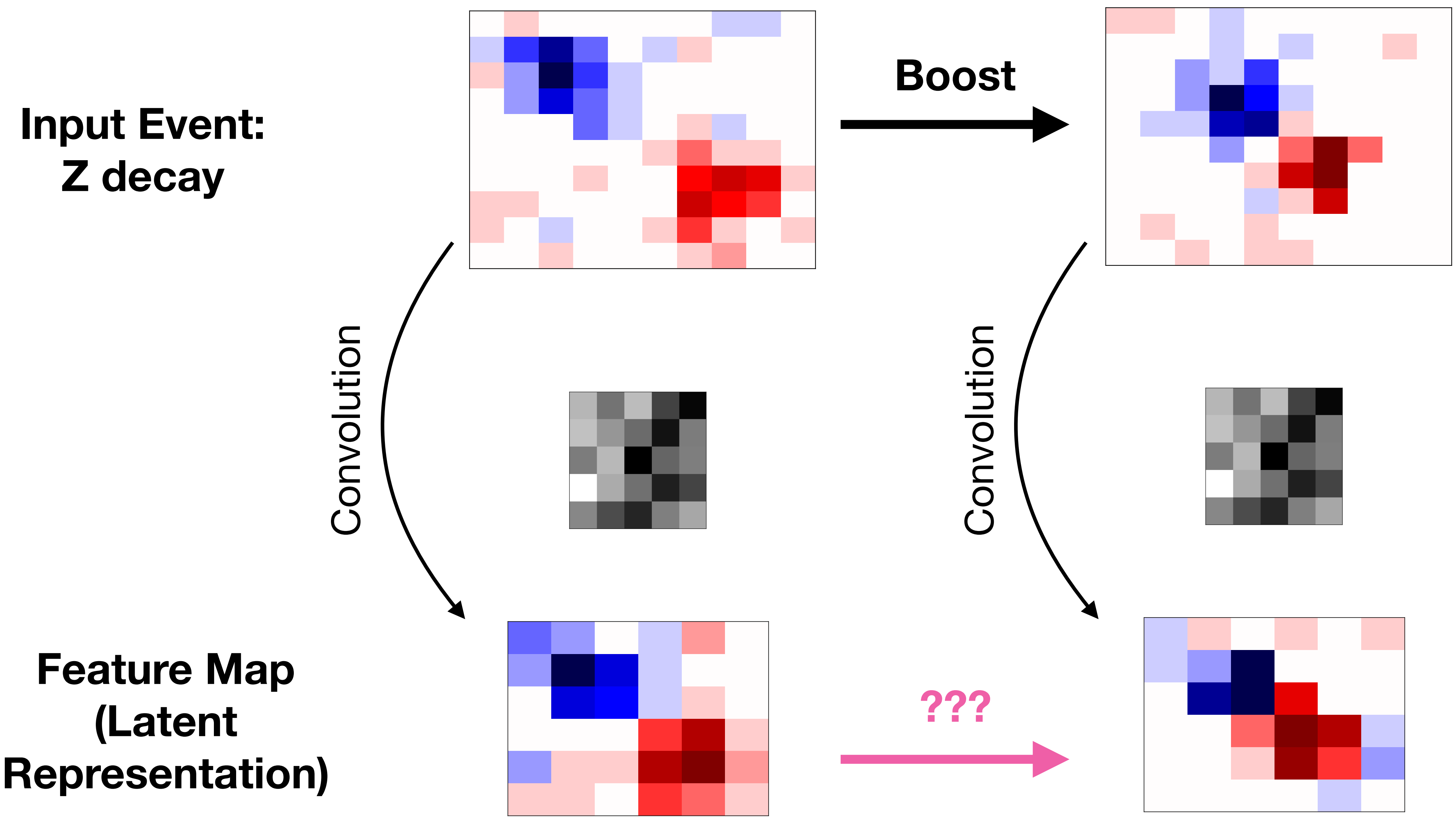


Why Equivariance?

Example: Categorization → Want *Invariant* Response



Equivariance for Jets?



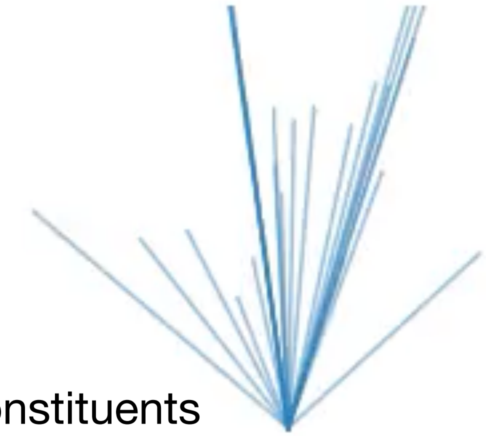
Group Convolutions

Warm up: SO(3) convolutions

Consider a jet to be a function on the 2-Sphere:

$$j(\theta,\phi) = \sum_i E_i \delta(\theta-\theta_i)\delta(\phi-\phi_i)$$

$$\approx \sum_{\ell,m}^{\ell_{\max}} a_{\ell m} Y_{\ell m}(\theta,\phi)$$



Jet constituents

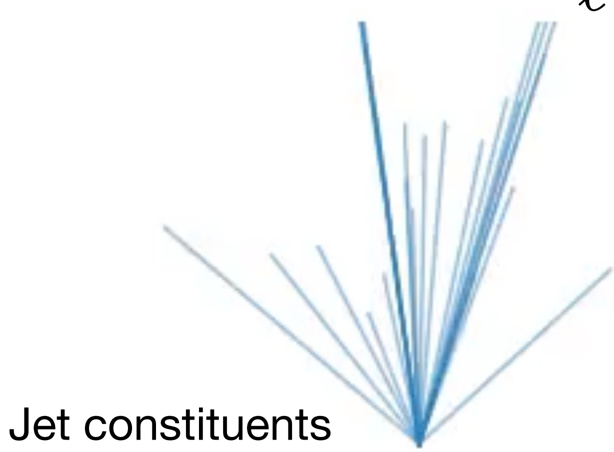
Group Convolutions

Warm up: SO(3) convolutions

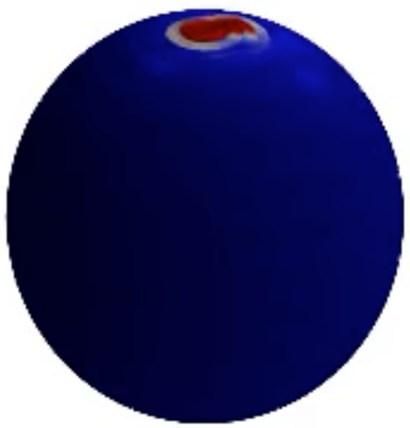
Consider a jet to be a function on the 2-Sphere:

$$j(\theta, \phi) = \sum_i E_i \delta(\theta - \theta_i)\delta(\phi - \phi_i)$$

$$\approx \sum_{\ell,m}^{\ell_{\max}} a_{\ell m} Y_{\ell m}(\theta, \phi)$$



Spherical Basis



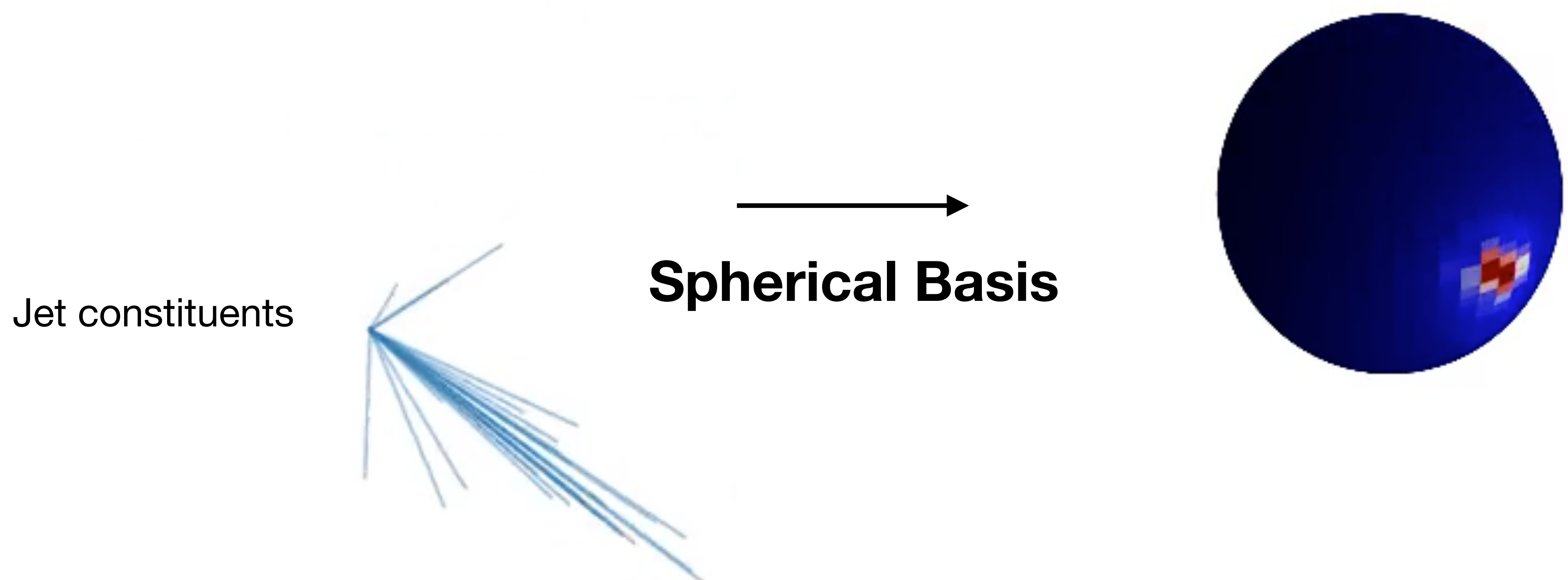
Group Convolutions

Warm up: SO(3) convolutions

Consider a jet to be a function on the 2-Sphere:

$$j(\theta, \phi) = \sum_i E_i \delta(\theta - \theta_i)\delta(\phi - \phi_i)$$

$$\approx \sum_{\ell,m}^{\ell_{\max}} a_{\ell m} Y_{\ell m}(\theta, \phi)$$



Group Convolutions

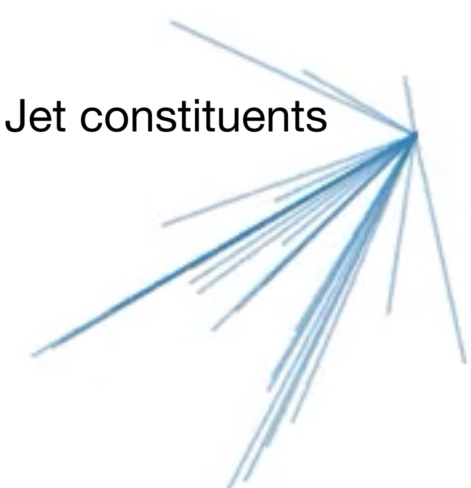
Warm up: SO(3) convolutions

Consider a jet to be a function on the 2-Sphere:

$$j(\theta, \phi) = \sum_i E_i \delta(\theta - \theta_i)\delta(\phi - \phi_i)$$

$$\approx \sum_{\ell,m}^{\ell_{\max}} a_{\ell m} Y_{\ell m}(\theta, \phi)$$

Jet constituents



Spherical Basis

Group Convolutions

Warm up: SO(3) convolutions

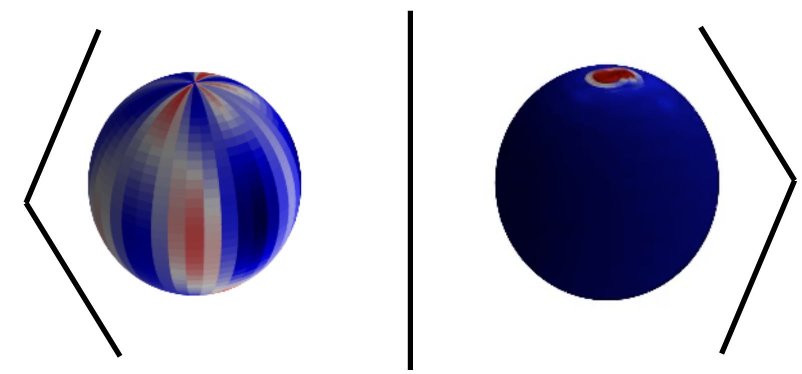
Construct an arbitrary (learnable) linear combination of spherical harmonics (i.e., a **filter**):

$$\psi(\theta, \phi) = \sum_{\ell, m}^{\ell_{\max}} b_{\ell m} Y_{\ell m}(\theta, \phi)$$

Projections are just a number:

$$\langle \psi | j \rangle \in \mathbb{C}$$

$$[\psi \star j](R) =$$



But if we consider the projection **at all possible orientations:**

$$[\psi \star j](R) := \langle \psi | L_R j \rangle \, , R \in SO(3)$$

$$[\psi \star f] : SO(3) \to \mathbb{C}$$

Group Convolutions

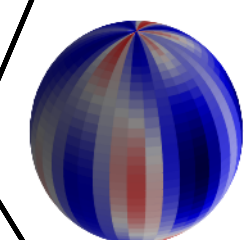
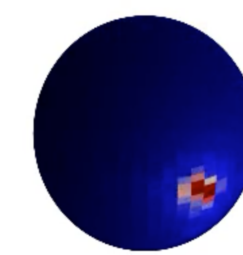
Warm up: SO(3) convolutions

Construct an arbitrary (learnable) linear combination of spherical harmonics (i.e., a **filter**):

$$\psi(\theta, \phi) = \sum_{\ell, m}^{\ell_{\max}} b_{\ell m} Y_{\ell m}(\theta, \phi)$$

Projections are just a number:

$$\langle \psi | j \rangle \in \mathbb{C}$$

$[\psi \star j](R)$ = $\Big\langle$  $\Big|$  $\Big\rangle$

But if we consider the projection **at all possible orientations:**

$$[\psi \star j](R) := \langle \psi | L_R j \rangle \, , R \in SO(3)$$

$$[\psi \star f] : SO(3) \rightarrow \mathbb{C}$$

Group Convolutions

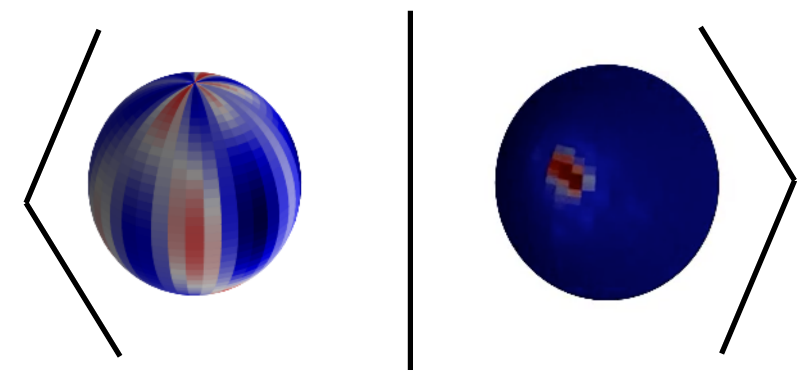
Warm up: SO(3) convolutions

Construct an arbitrary (learnable) linear combination of spherical harmonics (i.e., a **filter**):

$$\psi(\theta, \phi) = \sum_{\ell, m}^{\ell_{\max}} b_{\ell m} Y_{\ell m}(\theta, \phi)$$

Projections are just a number:

$[\psi \star j](R)$ =



$$\langle \psi | j \rangle \in \mathbb{C}$$

But if we consider the projection **at all possible orientations:**

$$[\psi \star j](R) := \langle \psi | L_R j \rangle \, , R \in SO(3)$$

$$[\psi \star f] : SO(3) \to \mathbb{C}$$

Group Convolutions

Warm up: SO(3) convolutions

In practice, can either rotate the “input", or the “filter”:

$$\langle \psi | L_R j \rangle = \langle L_{R^{-1}} \psi | j \rangle$$

Where we use the Wigner D-Matricies to rotate harmonics:

$$L_R \, |\ell, m\rangle = \sum_{m'} \mathcal{D}^{\ell}_{mm'}(R) \, |\ell, m'\rangle$$

This allows us to build deeper networks,
by applying rotations to internal representations.

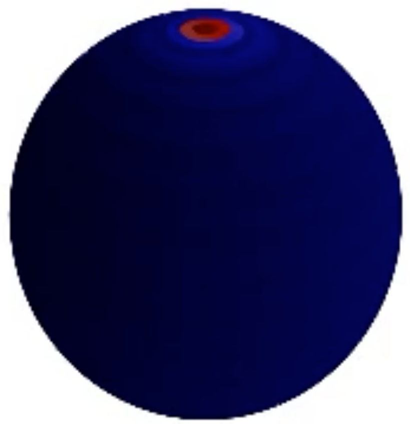
See also: arxiv:1801.10130, spherical CNNs

Boost Equivariance

How does this representation transform with an arbitrary boost?

Constituents under boost:

Harmonics under boost:



Problem 1: Rotations leave particle energy invariant!

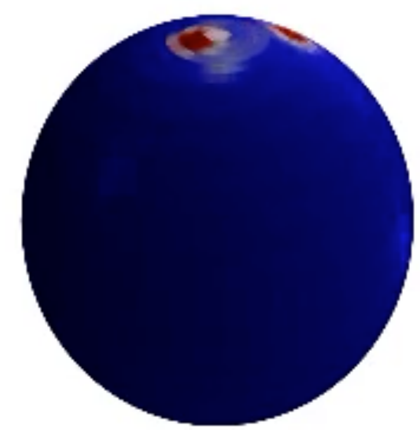
Problem 2: Rotations cannot mix different l-modes

Boost Equivariance

How does this representation transform with an arbitrary boost?

Constituents under boost:

Harmonics under boost:



Problem 1: Rotations leave particle energy invariant!

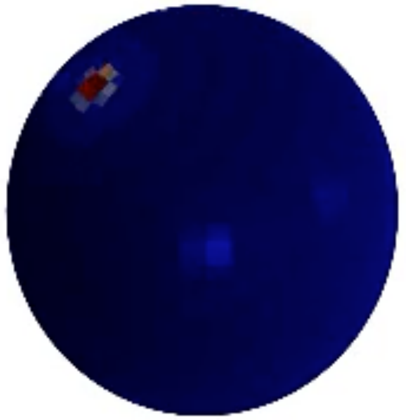
Problem 2: Rotations cannot mix different l-modes

Boost Equivariance

How does this representation transform with an arbitrary boost?

Constituents under boost:

Harmonics under boost:



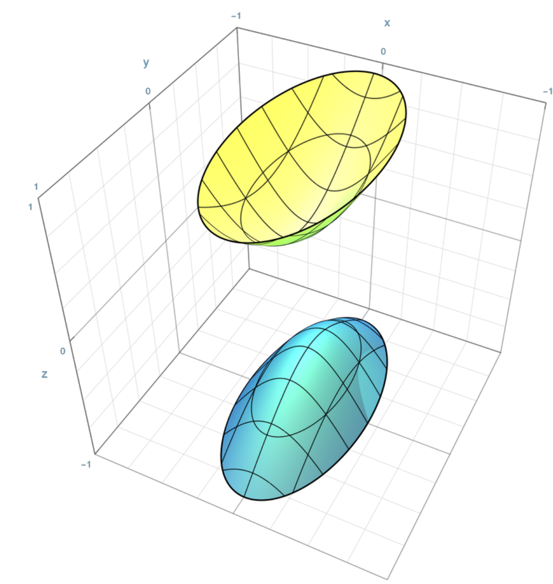
Problem 1: Rotations leave particle energy invariant!

Problem 2: Rotations cannot mix different l-modes

Spacetime Observables

There is a unitary representation for the Lorentz group, for functions on Minkowski hyperboloid:

$$p_{hyp} = (m\cosh(\xi), \vec{p}(\theta, \phi))$$



As before, write observations as deltas:

$$j(\theta, \phi) \to j(\xi, \theta, \phi) = \sum_i \delta(\xi - \xi_i)\delta(\theta - \theta_i)\delta(\phi - \phi_i)$$

Spherical harmonics become:

$$Y_{\ell m}(\theta, \phi) \to \Psi_{N\ell m}(\xi, \theta, \phi)$$

$$\Psi_{N\ell m}(\xi, \theta, \phi) = \frac{1}{M_\ell}\Pi_\ell(N, \xi)Y_{\ell m}(\theta, \phi) \quad N \in [0, \inf)$$

Relativistic Filters

As before, we can construct arbitrary, **learnable** filters:

$$\psi(\xi,\theta,\phi) = \int_0^{N_{\max}} dN \sum_{\ell,m}^{\ell_{\max}} a_{\ell m}(N)\Psi_{N\ell m}(\xi,\theta,\phi)$$

$a_{\ell m}(N)$ must be sampled discretely

The input delta functions make it easy to compute projections:

$$j(\xi,\theta,\phi) = \sum_i \delta(\xi-\xi_i)\delta(\theta-\theta_i)\delta(\phi-\phi_i)$$

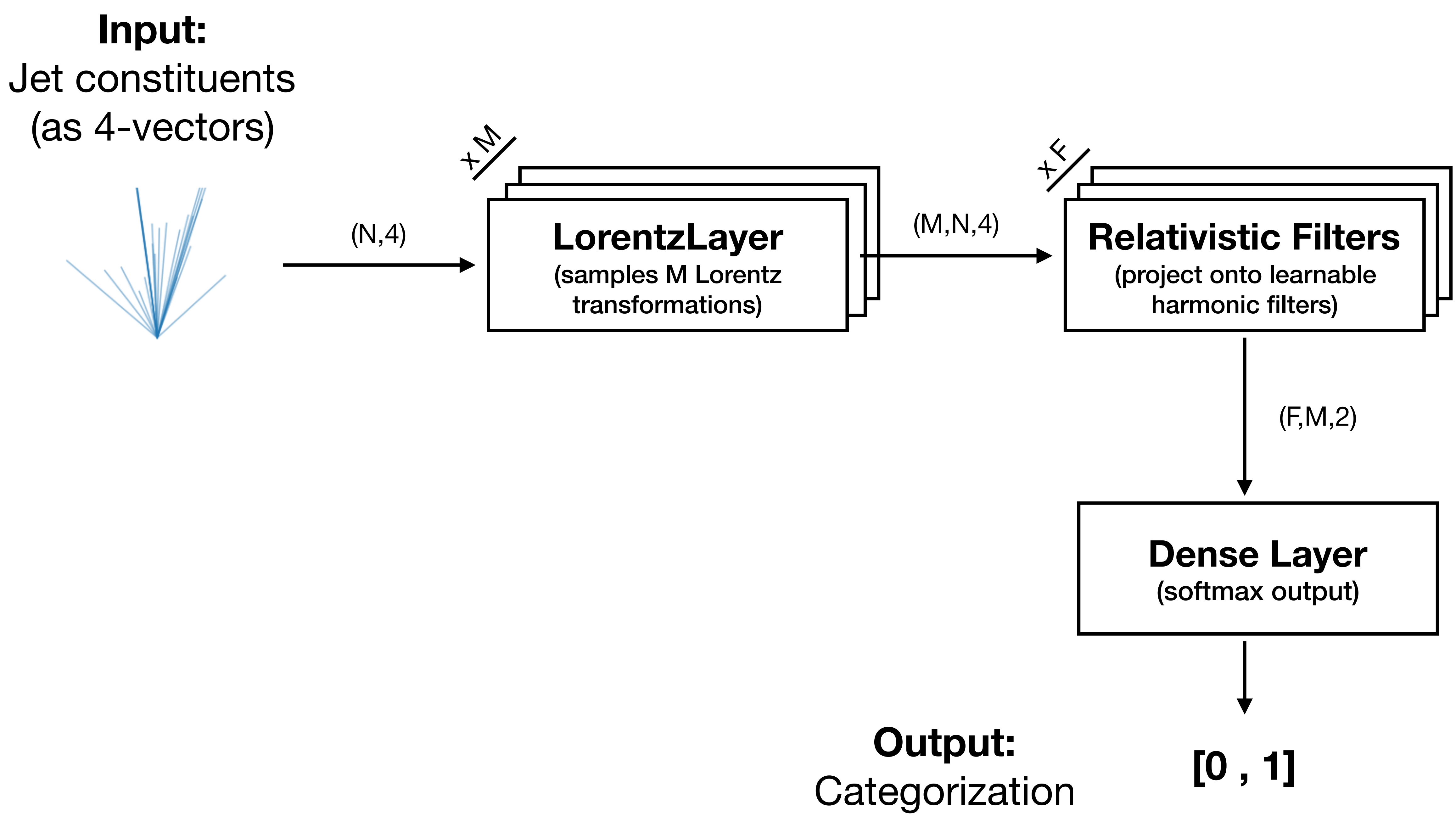
$$\langle N\ell m|j\rangle = \sum_i \Psi_{N\ell m}(\xi_i,\theta_i,\phi_i)$$

And, since we know how to apply Lorentz transformations to an arbitrary jet, **we can evaluate relativistic convolutions:**

$$[\psi \star j](\Lambda) := \langle \psi | L_\Lambda j\rangle \, ; \Lambda \in SO(3,1)^+$$

Single Layer Network

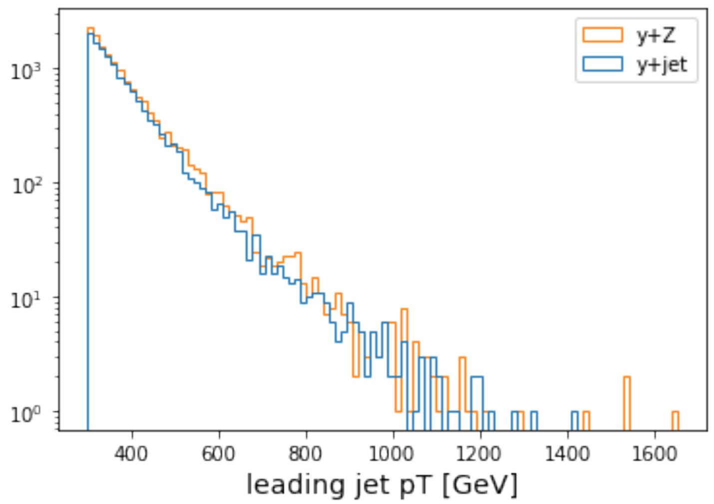
(work only with collinear boost subgroup)

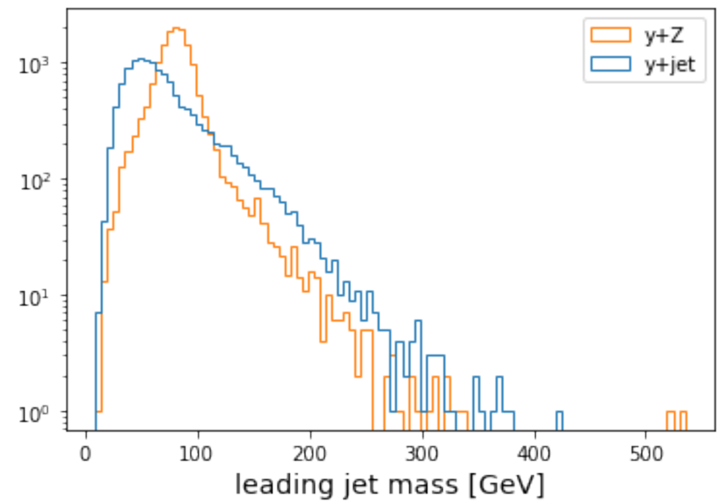


Benchmark Test

Task: discriminate leading jets from γ+Z (signal) and γ+jet (background)

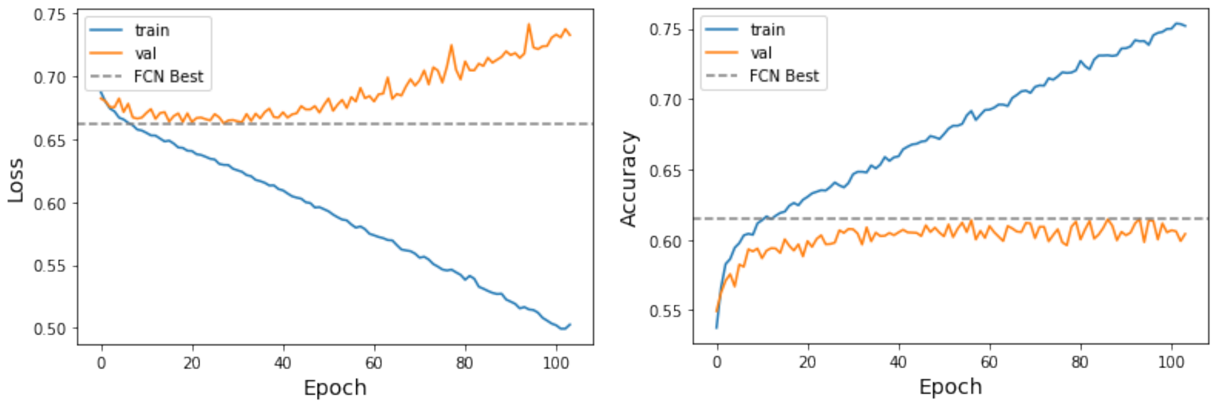
- AntiKt(R=1.0) clustering after hadronization
- Limit inputs to 32 leading-pT constituents
- pT(jet)>300 GeV





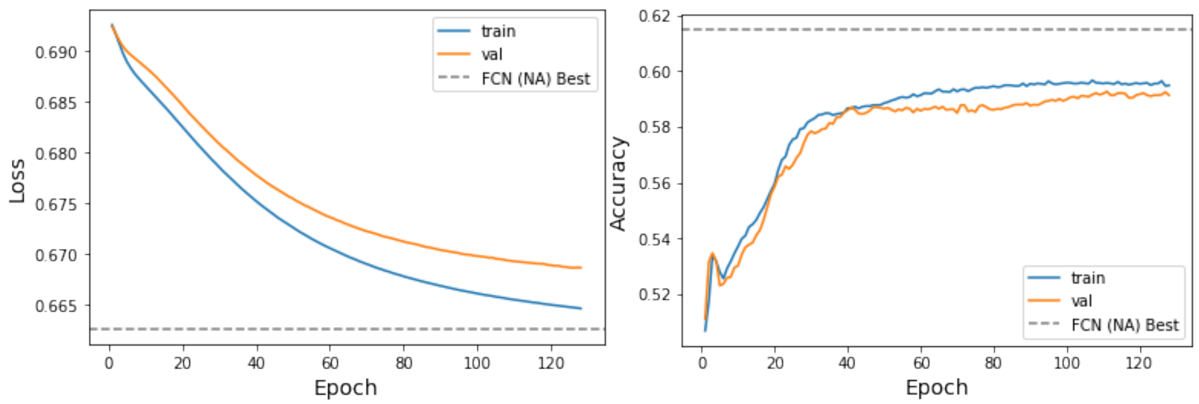
Network 1: Fully Connected
4 layers, 90k parameters

Accuracy: ~62%

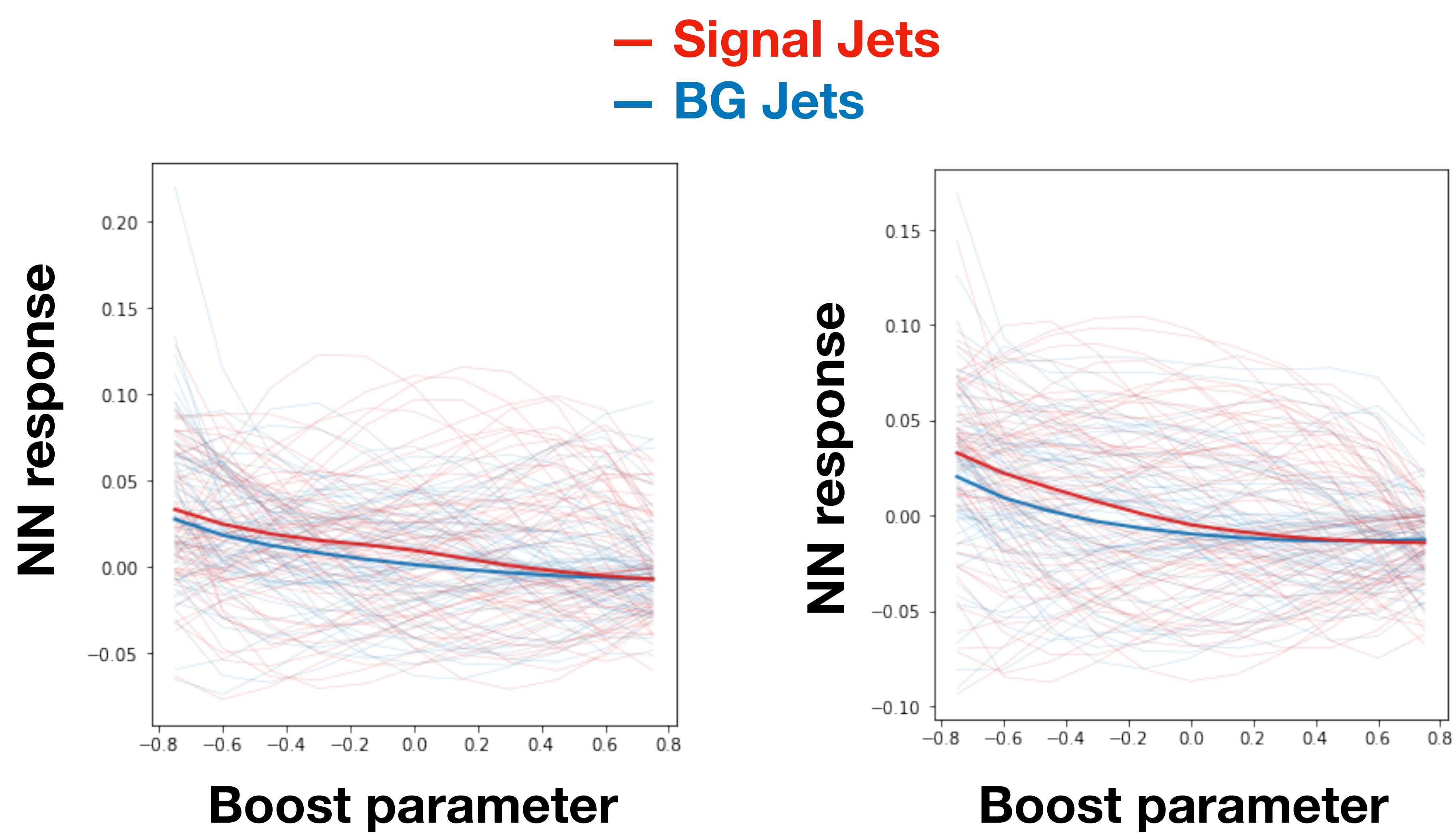


Network 2: Relativistic Harmonic Net (l_{max}=0), 596 parameters

Accuracy: ~59%



Output of Harmonic Layers



Conclusions

- A first attempt at enforcing equivariance relevant for jet physics
- Currently only works for **single-layer networks**….
  - But should generalize to N-layer networks soon!
- Naive benchmark shows nearly-comparable performance to naive models
  - Need to build deeper networks to do state-of-the-art comparisons

Thank you!