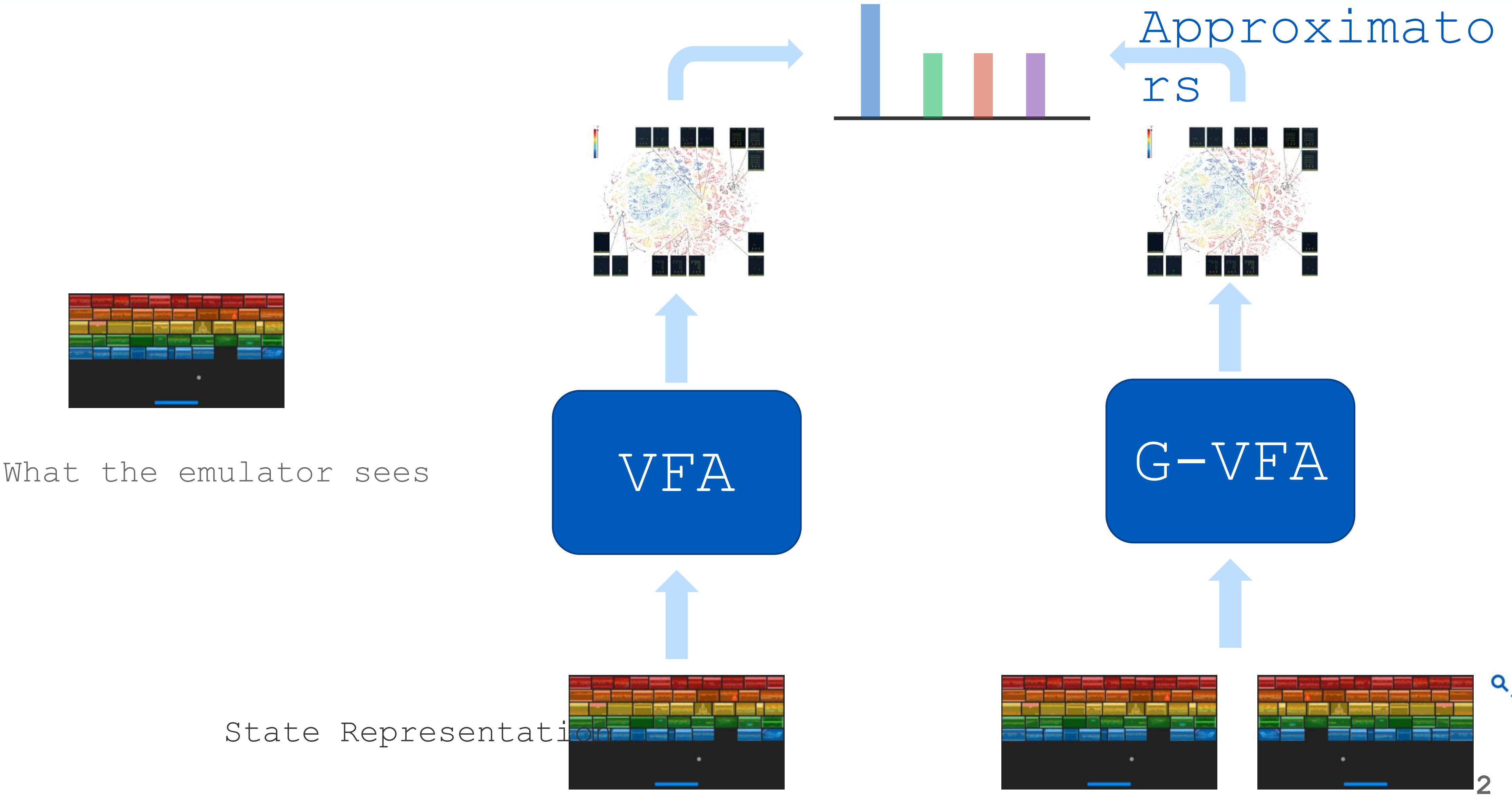
Group Equivariant Q-networks

Kiran Vaddi

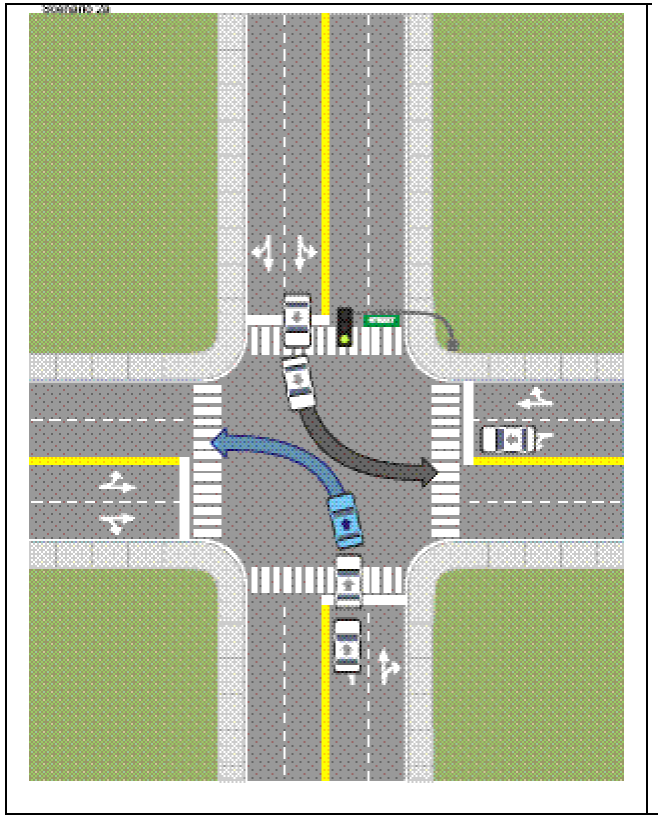
University at Buffalo
School of Engineering and Applied Sciences

Value function Approximators



Why do we need equivariance?

Blue and Black lines are symmetric



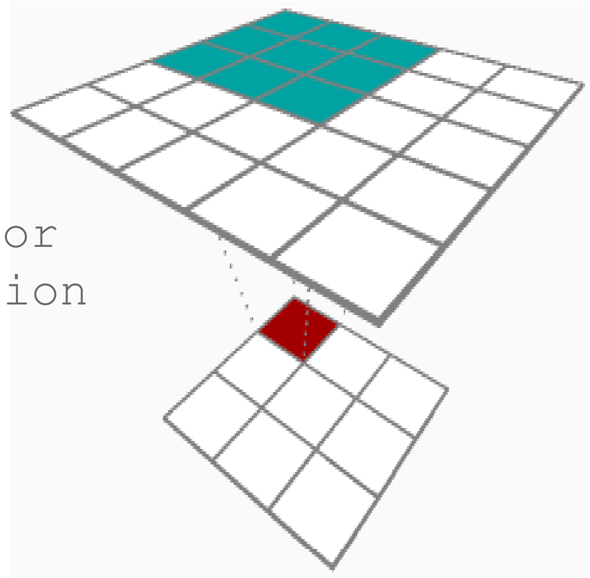
Agent can better estimate consequences:

- considers what it would if its on the symmetrically opposite path
- Doesn’t have to re-learn

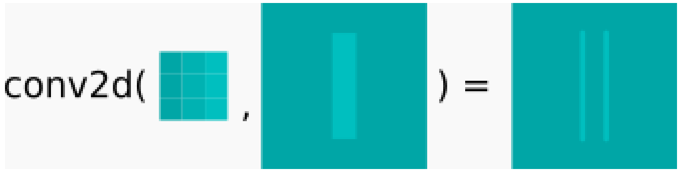
- Nature by default has some symmetry encoded into it
- Learning a symmetry valued representation is a better model of the environment
- Value function approximation learns a simple representation of model to make decision (refer t-sne plot in original DQN paper)

CNN vs G-CNN

CNN

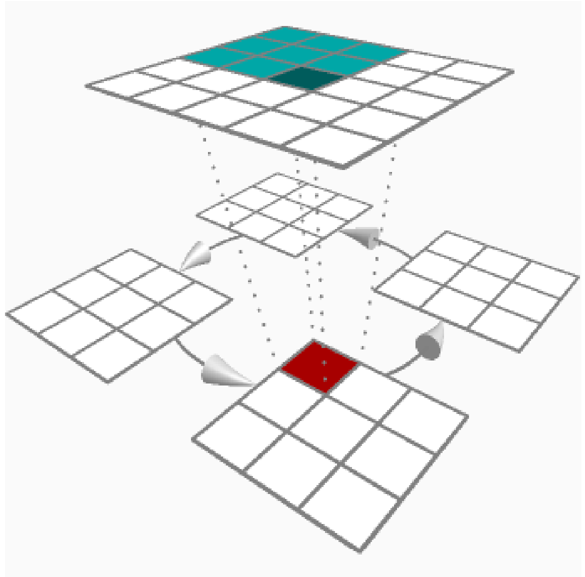


One channel for each translation of filter

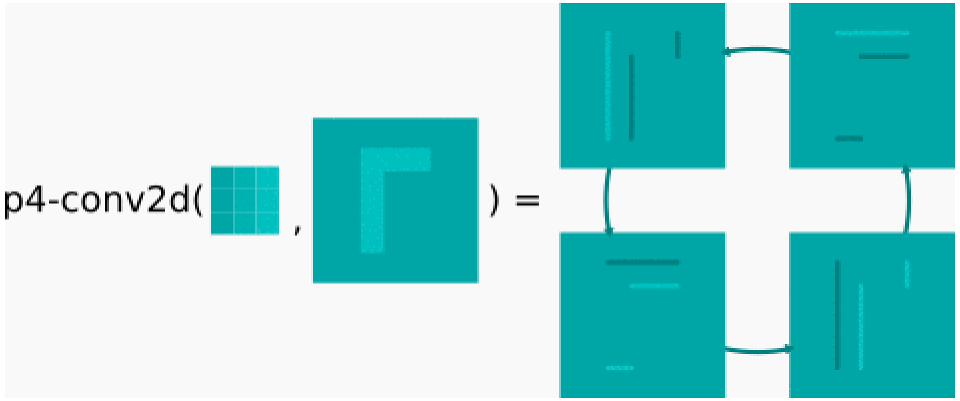


Features are different for bar when rotated

P4-CNN



One channel for each rotation + translation of filter



Features are same when rotated

http://scyfer.nl/2016/12/13/data-efficient-deep-

Implementing Gconv2d

For any $(u, v) \in \mathbb{Z}/2\mathbb{Z}$

Compute augmented filter bank:

$$g(r, u, v) = \begin{bmatrix} \cos(r\pi/2) & -\sin(r\pi/2) & u \\ \sin(r\pi/2) & \cos(r\pi/2) & v \\ 0 & 0 & 1 \end{bmatrix}$$

for $r \in \{0, 1, 2, 3\}$

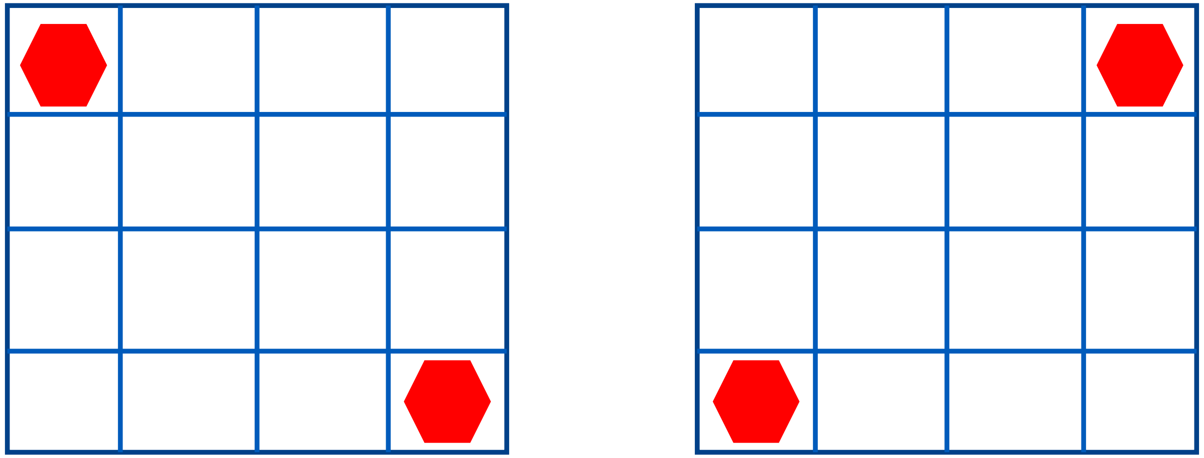
Now compute convolution for augmented filter bank and store it in corresponding dime

For 4x84x84 channel image input, you get 4xKxNxN kernel as output

Cohen, Taco, et.al, 2016.
Finzi, Marc, et alarXiv preprint arXiv:2002.12880 (2020).

Simple Example

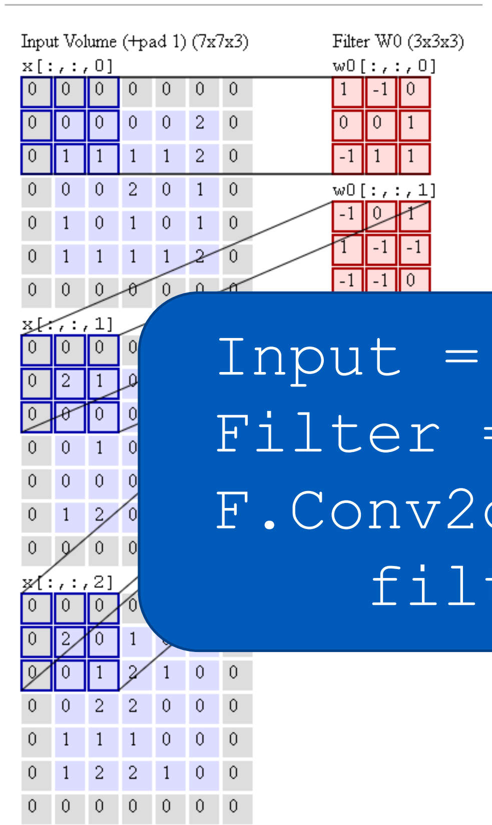
- State representations are important for MDP formulation
- Equivariant representations are important for Model free RL



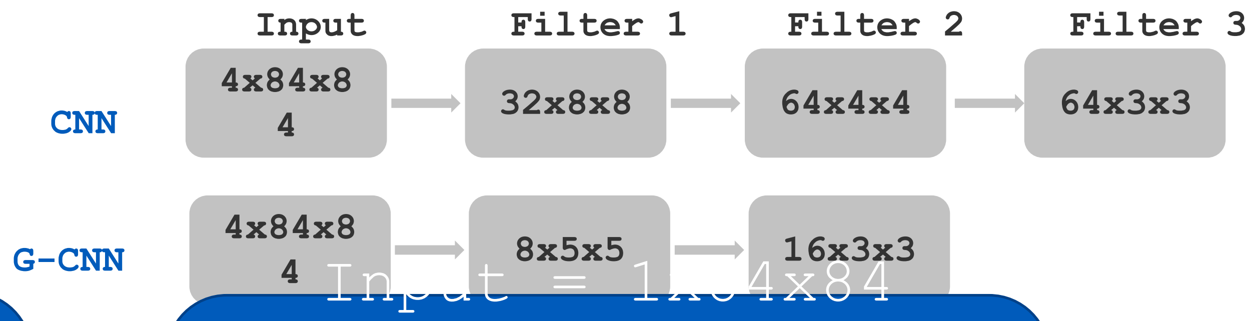
- **CNN : Get to the goal states in left and right corners**
- **G-CNN : get to goal states in any of the opposite corner**

- G-CNN has same representation for both grid worlds

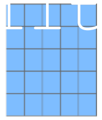
CNN vs G-CNN structure



Input = 1x84x84
Filter = 32x8x8
F.Conv2d(input, filter)

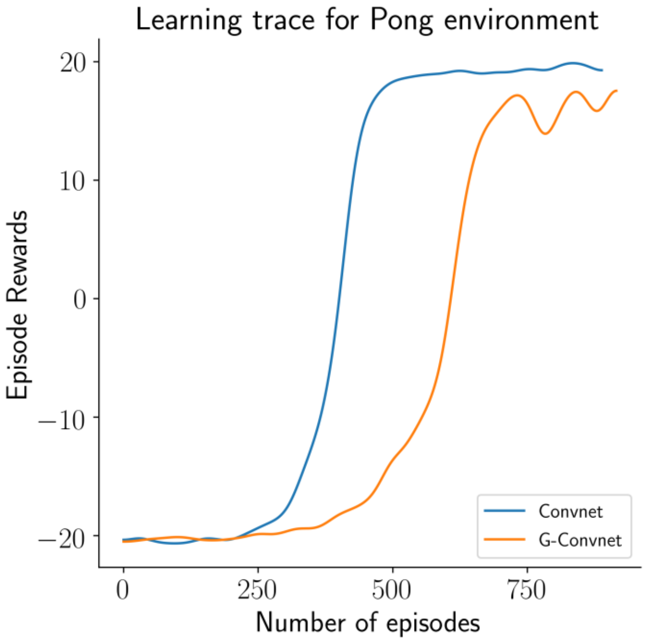


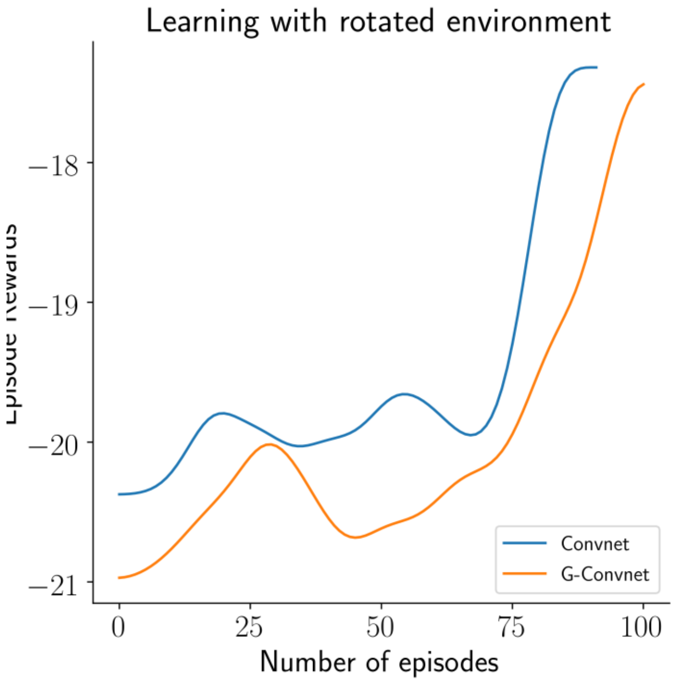
F1 = 8x5x5
Filter = g(F1) = 32x5x5
F.Conv2d(input,

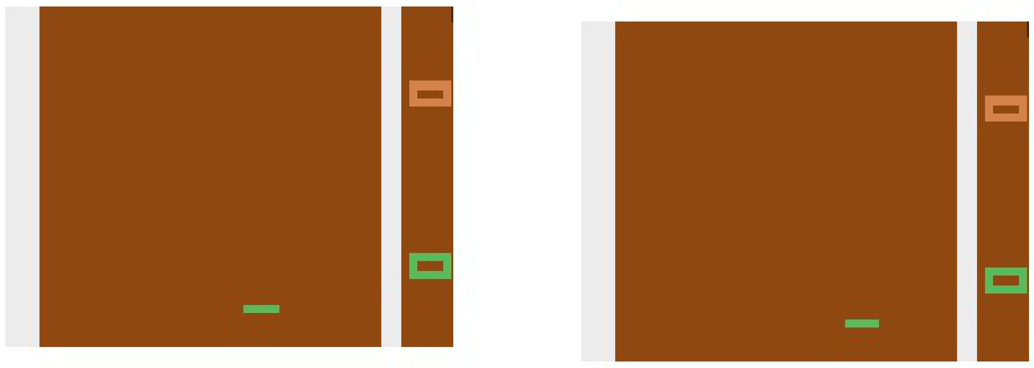


A convolution is computed by convolving filter channel by channel

Results on Atari Pong



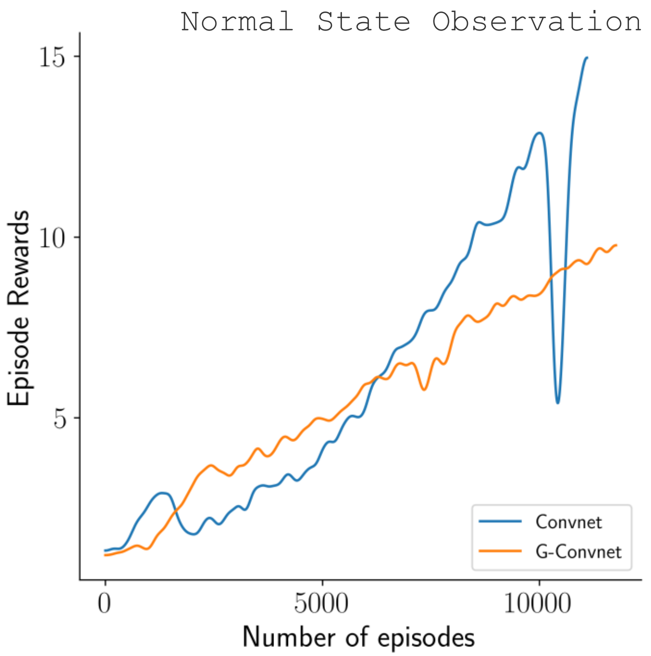


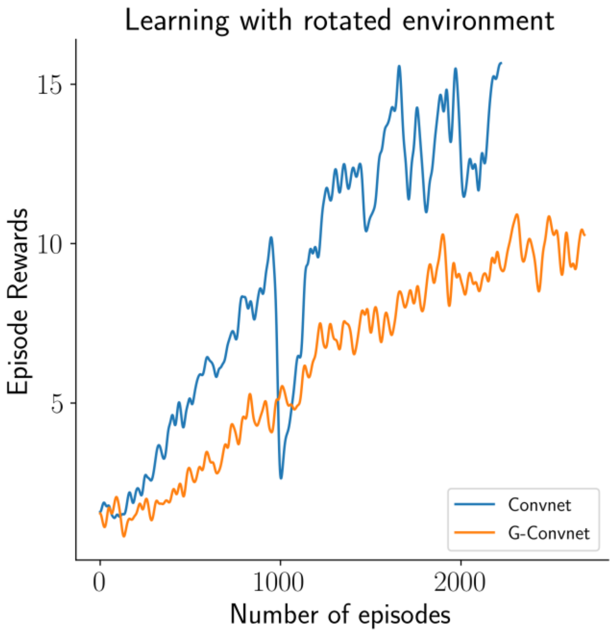


CNN G-CNN

G-convnet can learn using a much shallower network

Results on Atari Breakout





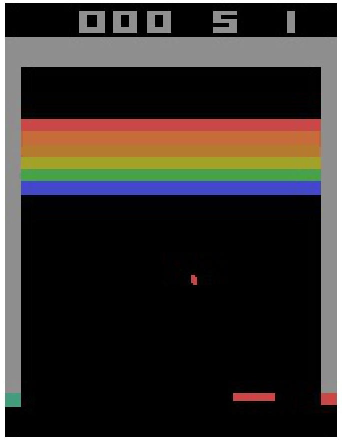
Learning trace on Atari breakout

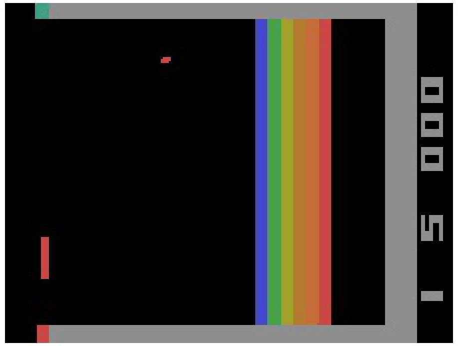
- From DQN paper : "The average total reward metric tends to be very noisy because the distribution of states changes with the policy visits"
- Equivariance in learned filters can be useful in obtaining a better state distribution thus less noisy total reward metric

Results: Continued

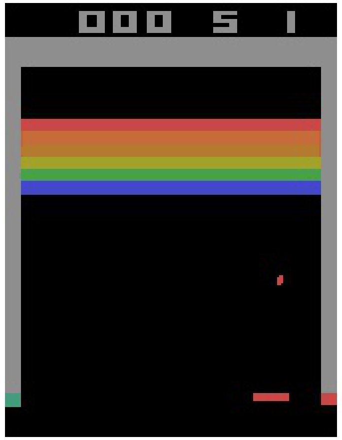
10 M frames | 1 M frames

Convnet





G-Convnet



Thank you

Questions?

University at Buffalo
School of Engineering and Applied Sciences