Learning Symmetric Embeddings for Equivariant World Models

Northeastern University
University of Amsterdam

Jung Yeon Park[1*], Ondrej Biza[1*], Linfeng Zhao[1], Jan-Willem van de Meent[1,2], Robin Walters[1]

[*] Equal contribution [1] Northeastern University, Boston, MA, USA [2] University of Amsterdam, Netherlands

Motivation

- Neural networks equivariant to symmetries, such as rotation and translation, are more generalizable and sample-efficient.
- **Symmetries in natural data are difficult to express** analytically, limiting the use of equivariant networks.

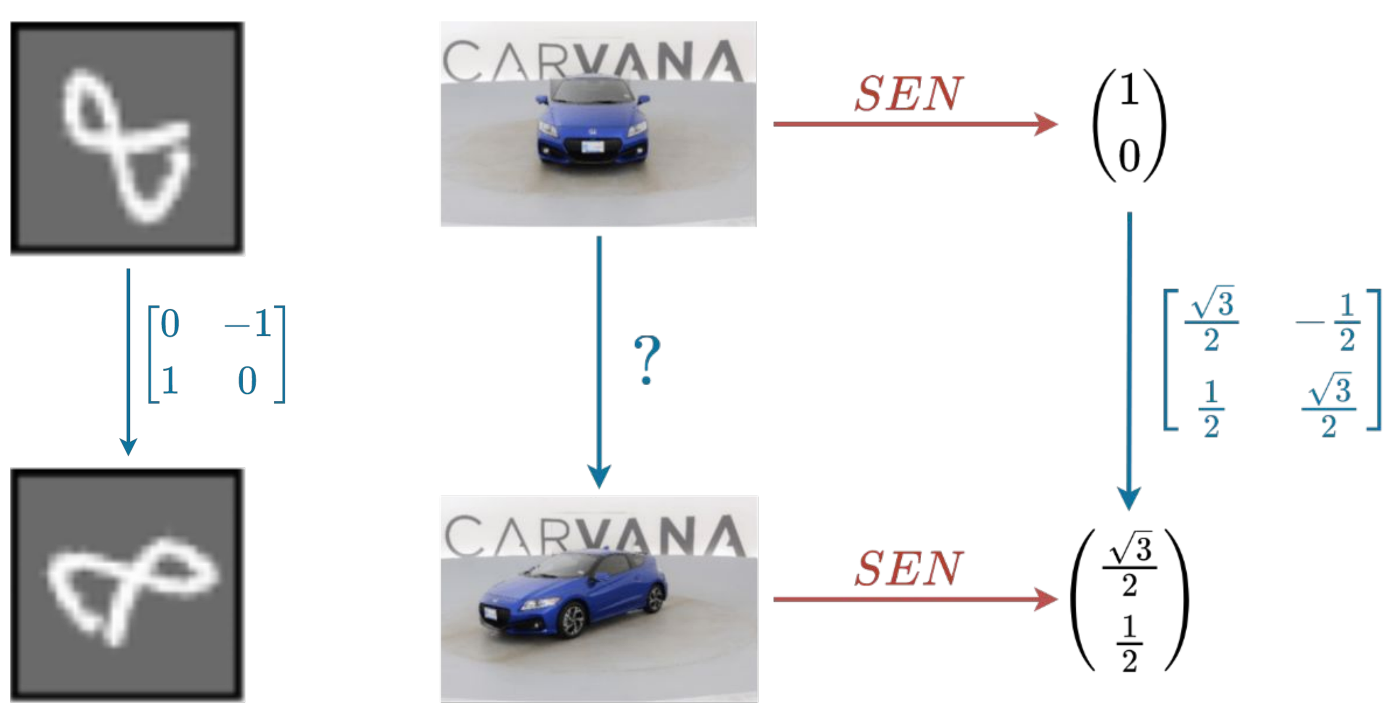


Figure 1: 2D rotation of an object can be expressed analytically for pixels, whereas a 3D rotation is difficult to compute.

Illustrative Example

- **Simple supervised sequence labeling task** to test whether we can learn the input transformation
- Compose a fully connected (FC) layer with 1D convolutional layers and compare against only FC

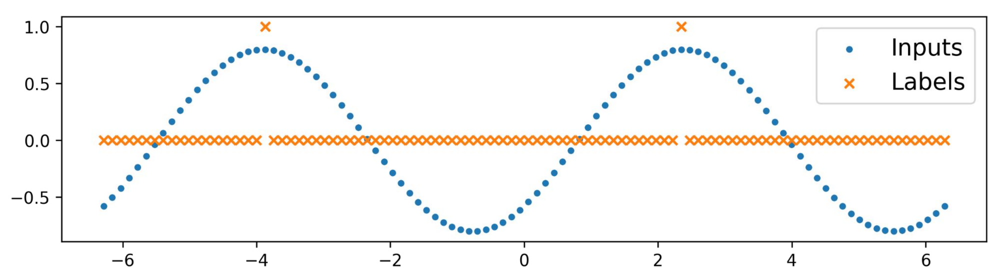


Figure 2: Training sample

FC+Conv | FC only | Equivariant FC

Figure 3: First FC layer weights. FC+Conv network learns shift equivariance

Symmetric Embedding Network

Key idea: Pair SEN with a downstream equivariant network and train end-to-end.

Meta-architecture: We propose a template for learning equivariant world models with input images for various downstream network architectures and symmetries (**Figure 4**).

SEN is an unconstrained network that compresses inputs to feature space for which group action is known. Downstream equivariant networks (**Encoder** and **Transition Model**) provide inductive bias.

Learning is done end-to-end using a contrastive loss.

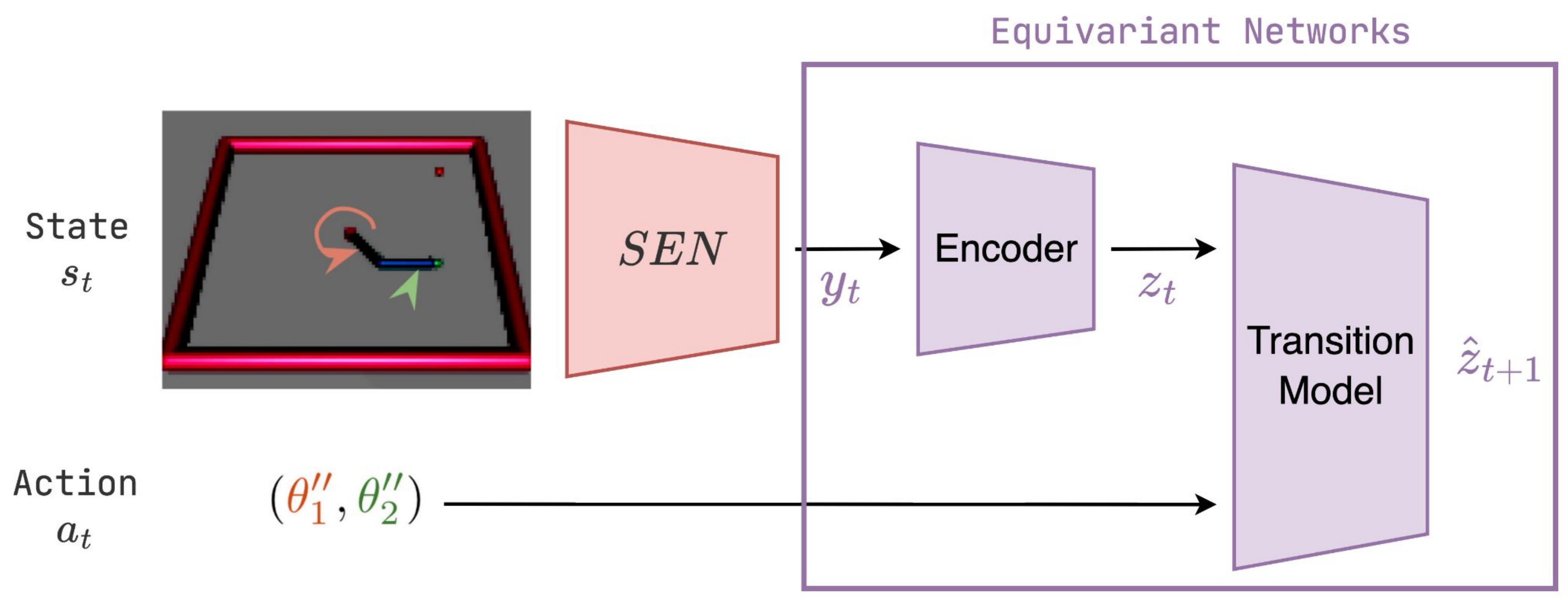
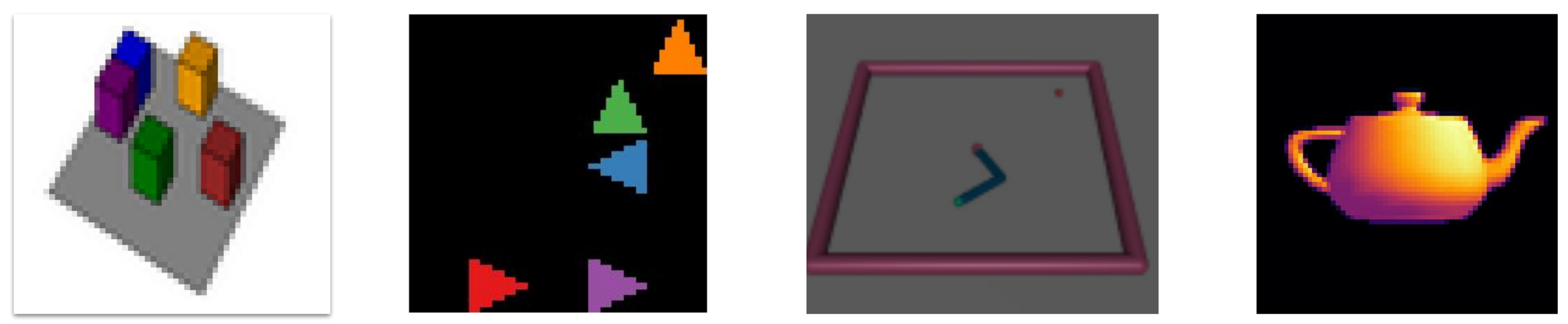


Figure 4: Meta-architecture with downstream equivariance.

Domains and Symmetries



2D & 3D Shapes	Rush Hour	Reacher	3D Teapot
π/2 rotation, object permutation	π/2 rotation, object permutation	π/2 rotation, flip, translation	3D rotation
C_4- equivariant MPNN	C_4- equivariant MPNN	E(2)-steerable CNN, equivariant MLP	Matrix Multiplication

SEN Contributions

- **Learn implicitly** how to map from pixel space with unknown group action to feature space with a known symmetry.
- **Flexible** meta-architecture for variety of different symmetry groups.
- **Extend** advantages of equivariant neural network architectures to previously inaccessible domains.

Results

- Develop two measures of equivariance Equivariance Error (EE) and Distance Invariance Error (DIE).

	Model	TH@1 (yaw, %)	TH@1 (pitch, %)	TH@1 (roll, %)	HH@1 (1 step, %)	EE(S)	DIE (1 step, $\times 10^{-2}$)
3D Teapot	Homeomorphic VAE	6.7	10.0	3.3	0.9	2.41	0.68
	None	6.7±6.7	60±40	86.7±6.6	93.9±2.2	2.38±0.04	3.41±0.16
	Ours (MatMul)	100±0.0	100±0.0	100±0.0	100±0.0	0.05±0.0	0.45±0.01

Table 1: Teapot - SEN achieves lower EE and DIE.

- SEN learns consistent representations

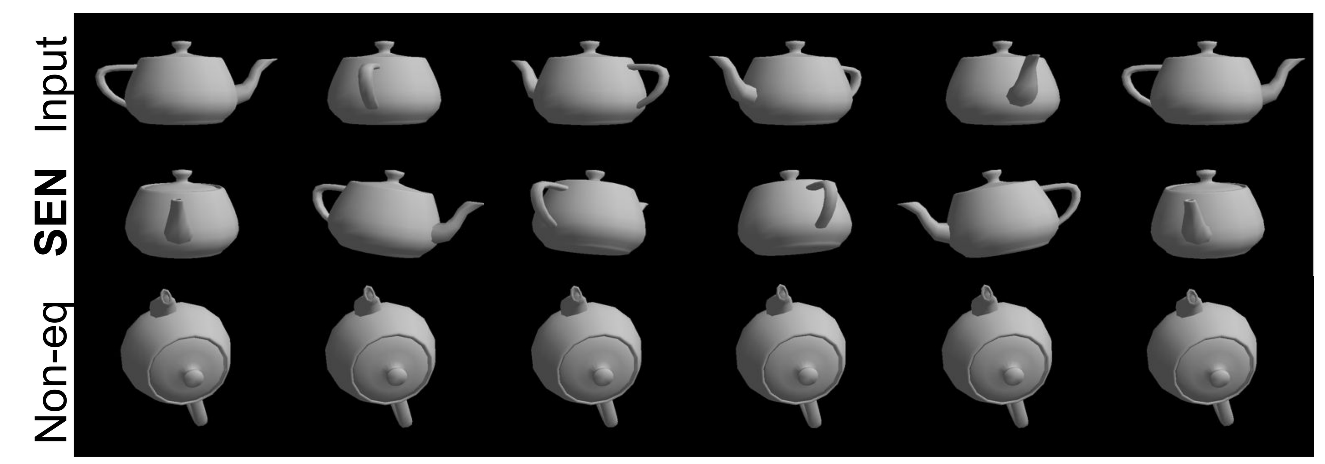


Figure 5: Teapot - SEN learns rotations in latent space correctly.

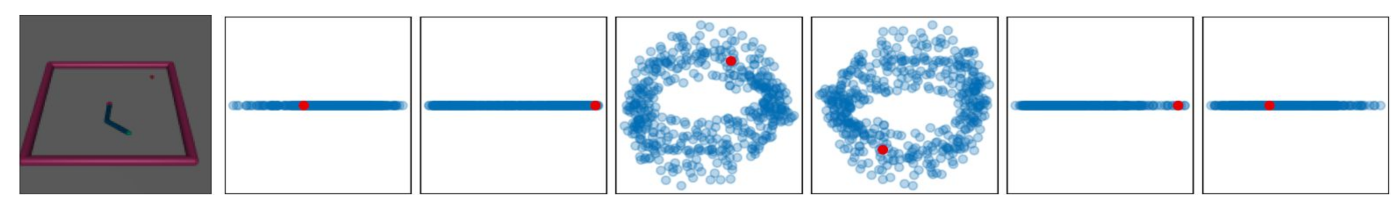


Figure 6: Reacher - learned latent factors. Middle factors are consistent with arm joint rotations.

- SEN generalizes to actions unseen during training.

	Limited Actions $\mathcal{A}'$	Model	H@1 (10 step, %)	MRR (10 step, %)	EE(S)	DIE (10 step, $\times 10^{-3}$)
2D Shapes	{up}	CNN	2.8±0.6	5.3±0.4	0.00±0.0	0.19±0.0
		Ours/Full	100±0.0	99.9±0.0	0.00±0.0	0.00±0.0
3D Blocks	{up,right,down}	None	52.3±14.	61.8±13.	0.98±0.2	181±79.
		Full	83.7±36.	86.0±31.	0.81±0.5	15±9.1
		Ours	99.9±0.0	100±0.0	0.96±0.3	5±4.7

Table 2: Next state prediction results for unseen actions.