Learning Partial Equivariances From Data

David W. Romero[1]
Vrije Universiteit Amsterdam

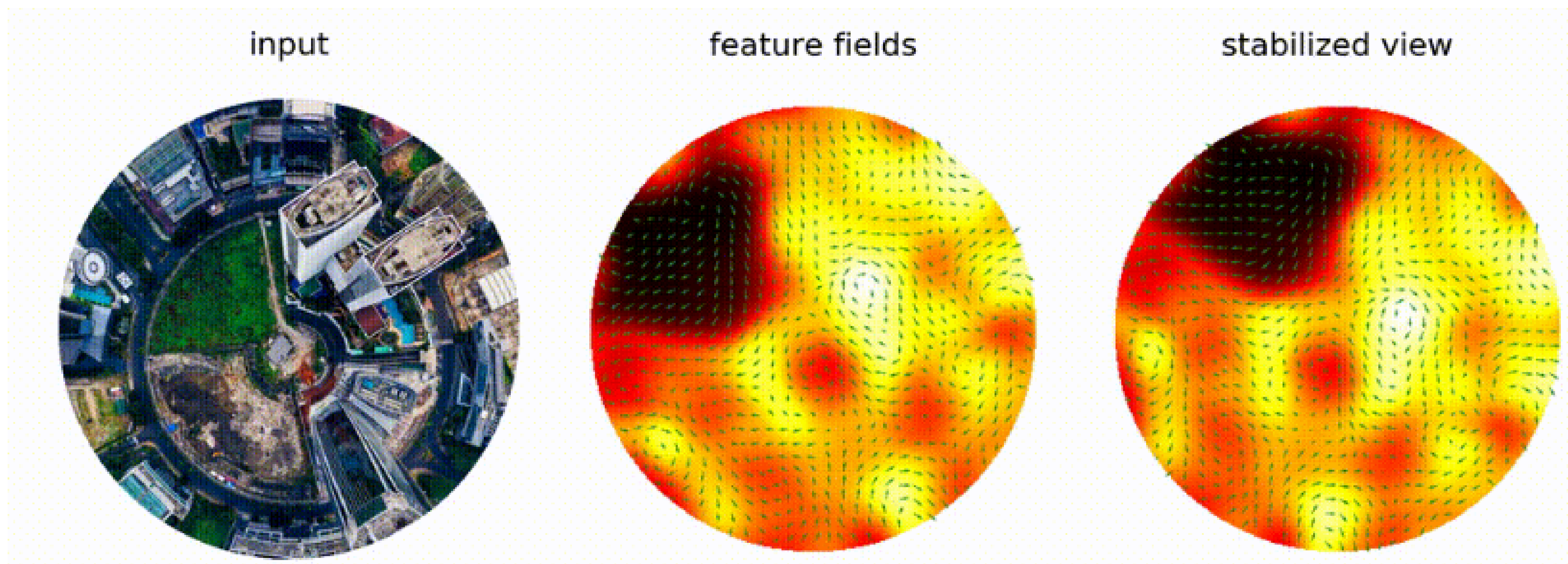
Suhas Lohit
Mitsubishi Electric Research Laboratories

[1]*Work done at Mitsubishi Electric Research Laboratories*

Partial Group Equivariant Neural Networks

GCNNs, Symmetries and Partial (Soft) Symmetries

Group convolutions restrict their convolutional kernels such that the convolution preserves the symmetries observed in data →Equivariance.



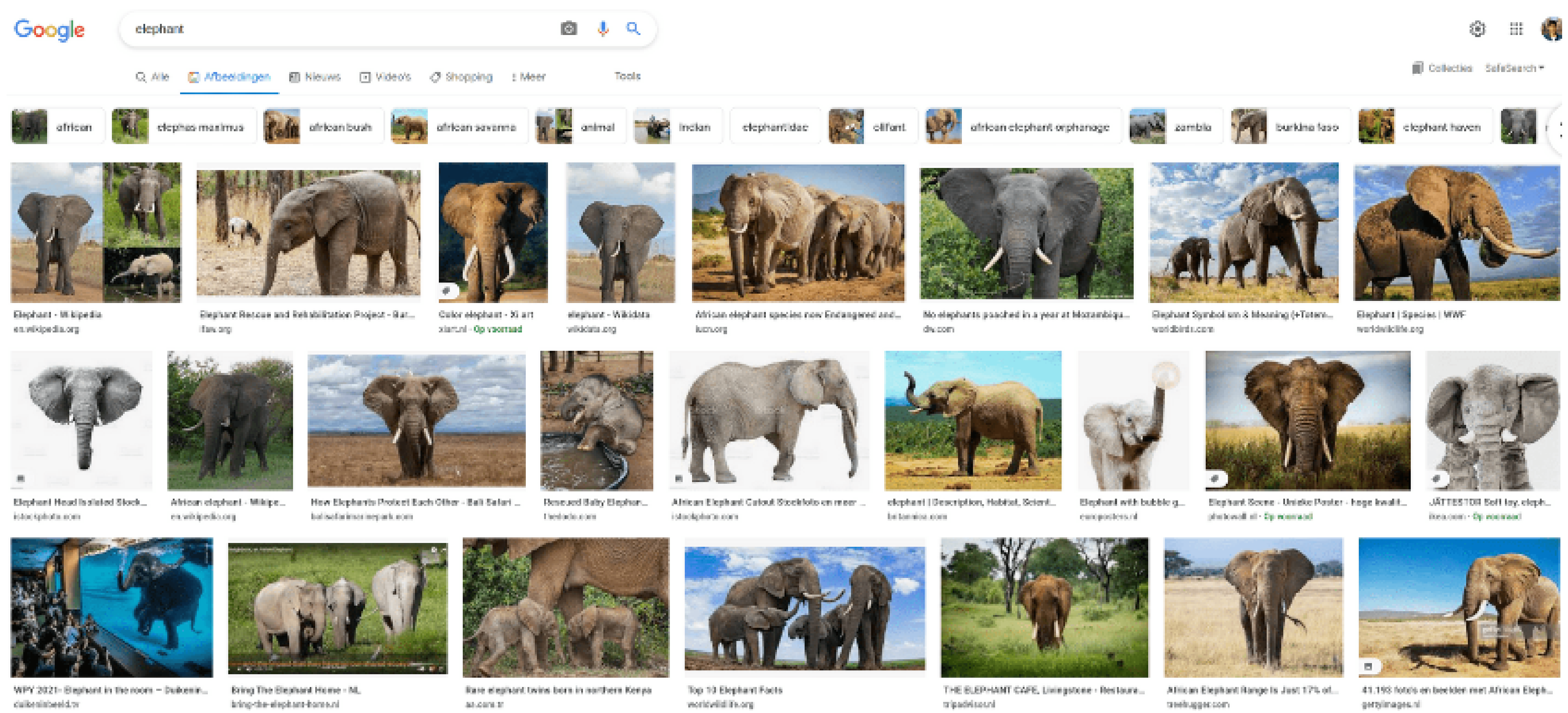
Group Convolutional Neural Networks (G-CNNs) are very data efficient, as the network does not need to learn these symmetries by itself.

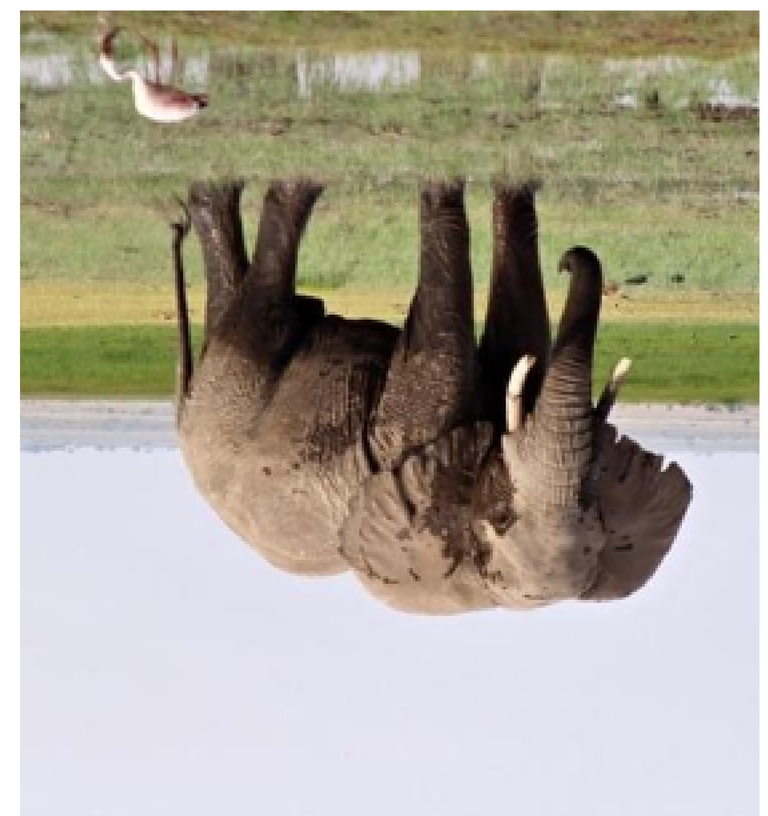
GCNNs are ubiquitous in data-scarse ML → Often SOTA in medical imaging, etc

Partial Group Equivariant Neural Networks

GCNNs, Symmetries and Partial (Soft) Symmetries

Nevertheless, several phenomena and tasks are better described with partial symmetries.





?

For natural images, full rotation equivariance is **overly** restrictive!

Partial Group Equivariant Neural Networks

GCNNs, Symmetries and Partial (Soft) Symmetries

Nevertheless, several phenomena and tasks are better described with partial symmetries.

What about this classification problem?

6 / 9

G-CNNs cannot distinghish among group transformations of the input.

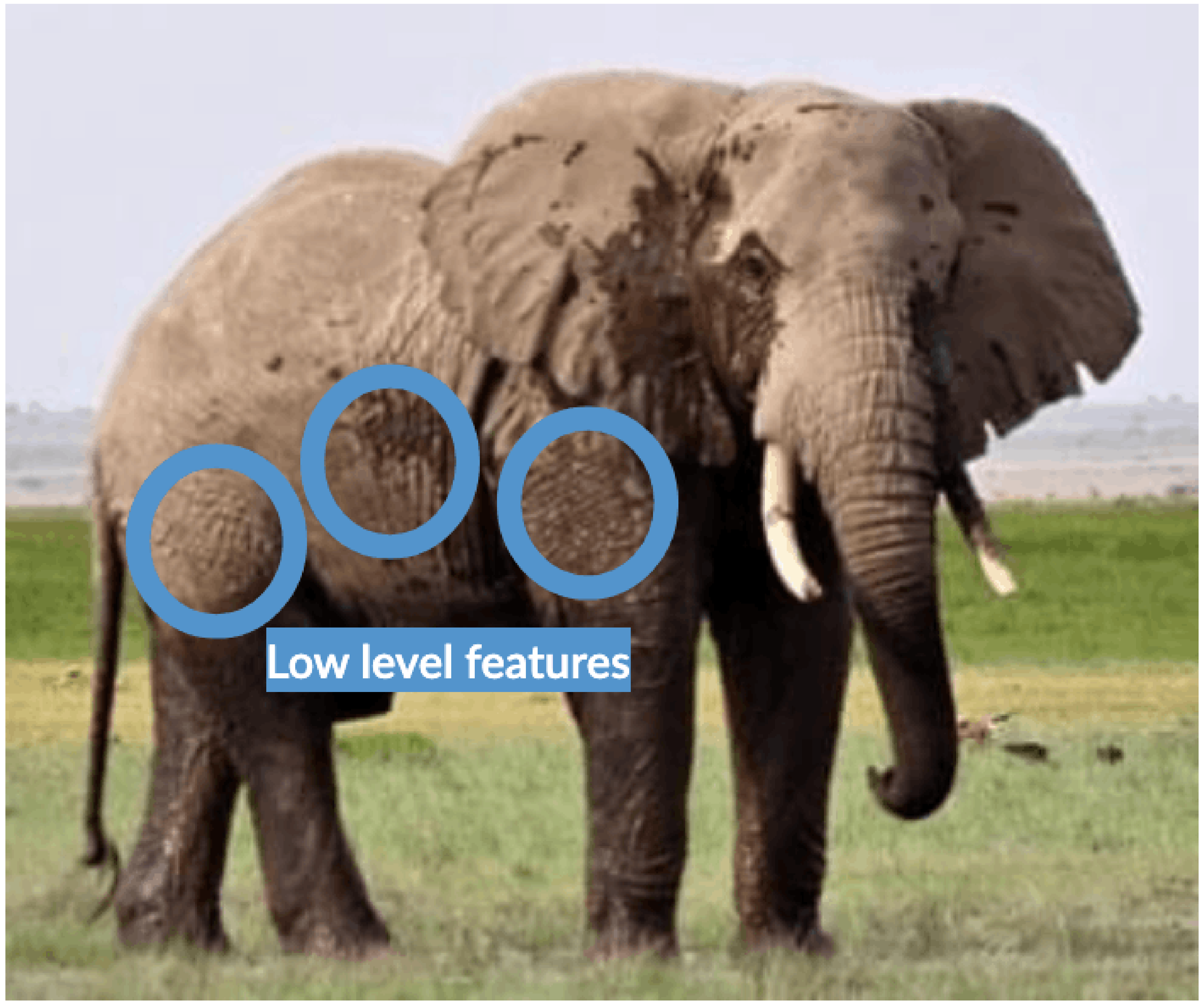
A rotation equivariant CNNs **cannot** dsintinguish between these digits.

BASE GROUP	DATASET	G-CNN
SE(2)	MNIST6-180	50.0
Mirroring	MNIST6-M	50.0
E(2)	MNIST6-180	50.0
	MNIST6-M	50.0

Partial Group Equivariant Neural Networks

GCNNs, Symmetries and Partial (Soft) Symmetries

Question: Should we then avoid using rotation equivariant group convolutions?



For **high-level features**, depends on the data.

For **low-level features**, also depends, but chances are equivariance will improve data efficiency.

Partial Group Equivariant Neural Networks

GCNNs, Symmetries and Partial (Soft) Symmetries

Question: Can we construct a NN which has the data-efficiency advantages of G-CNNs, but that is able to adapt its restrictions to model the data correctly?

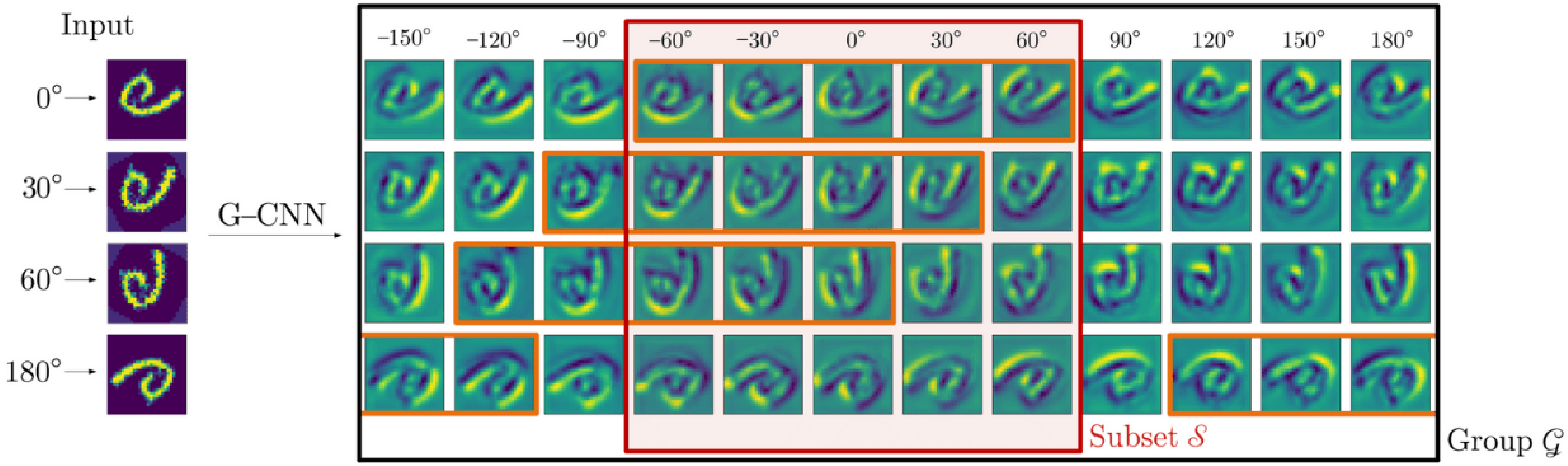
That is a NN that can **learn the level of *partial* equivariance** at each layer?

Partial Group Equivariant Neural Networks

Partial Group Equivariant Neural Networks

Partial Equivariance

Main Idea: Let each group conv. layer learn a subset of the group to which it is equivariant.



The group convolution will be (approx.) equivariant to some transformations, but ***<u>not all</u>***.

Partial Group Equivariant Neural Networks

Monte Carlo Approximations

How can we learn these subsets during training?

Monte Carlo approximation of group convolutions *[Finzi et al. 2020].*

$$(\psi * f)(u) = \int_{\mathcal{G}} \psi(v^{-1}u) f(v) \, \mathrm{d}\mu_{\mathcal{G}}(v),$$

We can approximate the continuous operation by sampling points in the domain of the input and output of the convolution, and evaluating the (continuous) functions on these points.

$$(\psi \hat{*} f)(u_i) = \sum_j \psi(v_j^{-1}u_i) f(v_j) \bar{\mu}_{\mathcal{G}}(v_j),$$

If points are sampled uniformly (on the Haar measure) from the group, the approximation is equivariant (in expectation).

Base Group	Dataset	G-CNN
SE(2)	MNIST6-180	50.0
Mirroring	MNIST6-M	50.0
E(2)	MNIST6-180	50.0
	MNIST6-M	50.0

Note: *It must be possible to evaluate the conv kernel at arbitrary positions →* ***continuous convolutional kernels!***

Partial Group Equivariant Neural Networks

Monte Carlo Approximations

How can we learn these subsets during training?

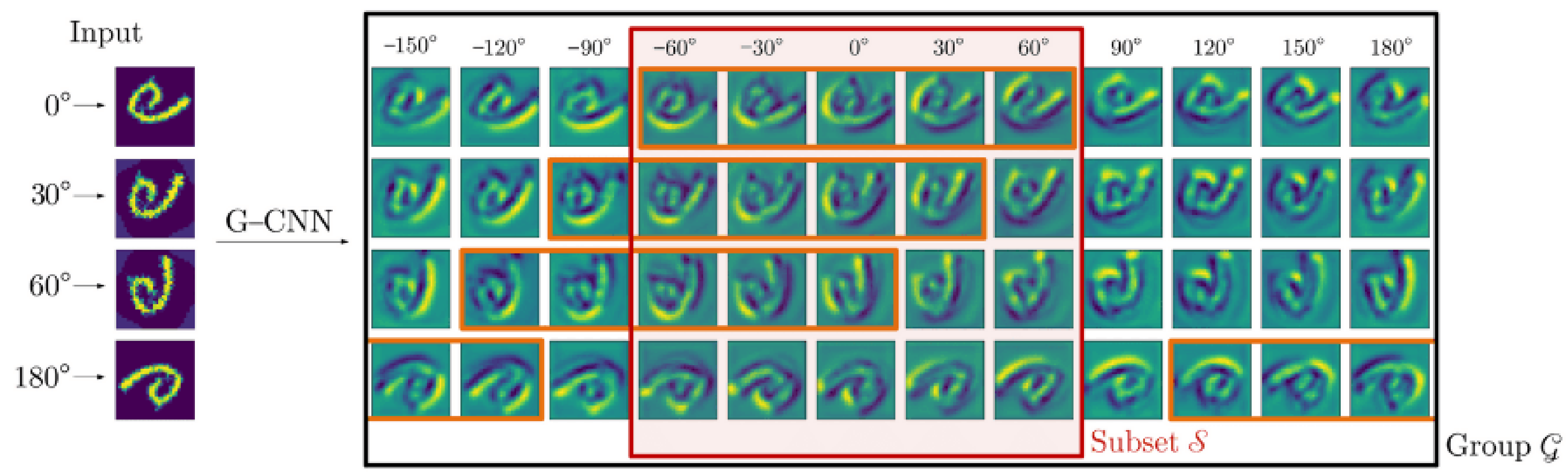
What if we don't sample uniformly from the entire group, but ***only*** from a part of the group?

$$(\psi * f)(u) = \int_{\mathcal{G}} \psi(v^{-1}u) f(v)\, \mathrm{d}\mu_{\mathcal{G}}(v), \quad \rightarrow \quad (\psi * f)(u) = \int_{\mathcal{G}} \mathrm{p}(u)\psi(v^{-1}u) f(v)\, \mathrm{d}\mu_{\mathcal{G}}(v);\ \mathrm{p}(u) \neq 0 \textit{ iff } u \in \mathcal{S}$$

Then, when we use the Monte Carlo approximation:

$$(\psi \hat{*} f)(u_i) = \sum_j \psi(v_j^{-1} u_i) f(v_j) \bar{\mu}_{\mathcal{G}}(v_j),$$

We effectly sample from a subset of the group! 😀



Partial Group Equivariant Neural Networks

Learning Distributions on the Group

Continuous groups

We want a distribution which is uniform in some part of the group and zero otherwise.

We use the reparameterization trick to learn a distribution on the Lie algebra, which is then mapped to the group via the exponential map.

$$\mathrm{p}(g) = \mathcal{U}(\theta \cdot [-1, 1))$$

Discrete groups

We can use the Gumbel Softmax trick to learn a Bernoulli distribution over each group element. This defines a distribution on the discrete group.

$$\mathrm{p}(e, g_1, \ldots, g_n) = \prod_{i=1}^{n} \mathrm{p}(g_i)$$

MITSUBISHI ELECTRIC Changes for the Better VU

Results

Partial Group Equivariant Neural Networks

Results

Partial G-CNNs adjust their level of equivariance based on the data.

They restrict equivariance if full equivariance is harmful

But, they learn to stay fully equivariant if full equivariance is advantageous.

Table 2. Test accuracy on vision benchmark datasets.

Base Group	No. Elems	Partial Equiv.	Classification Accuracy (%) RotMNIST	CIFAR10	CIFAR100
T(2)	1	-	97.23	83.11	47.99
SE(2)	4	✗	99.10	83.73	52.35
		✓	**99.13**	**86.15**	**53.91**
	8	✗	99.17	86.08	55.55
		✓	**99.23**	**88.59**	**57.26**
	16	✗	**99.24**	86.59	51.55
		✓	99.18	**89.11**	**57.31**
E(2)	8	✗	**98.14**	85.55	54.29
		✓	97.78	**89.00**	**55.22**
	16	✗	98.35	88.95	57.78
		✓	**98.58**	**90.12**	**61.46**

Table 1. Test accuracy on MNIST6-180 and MNIST6-M.

Base Group	Dataset	G-CNN	Partial G-CNN
SE(2)	MNIST6-180	50.0	**100.0**
Mirroring	MNIST6-M	50.0	**100.0**
E(2)	MNIST6-180	50.0	**100.0**
	MNIST6-M	50.0	**100.0**

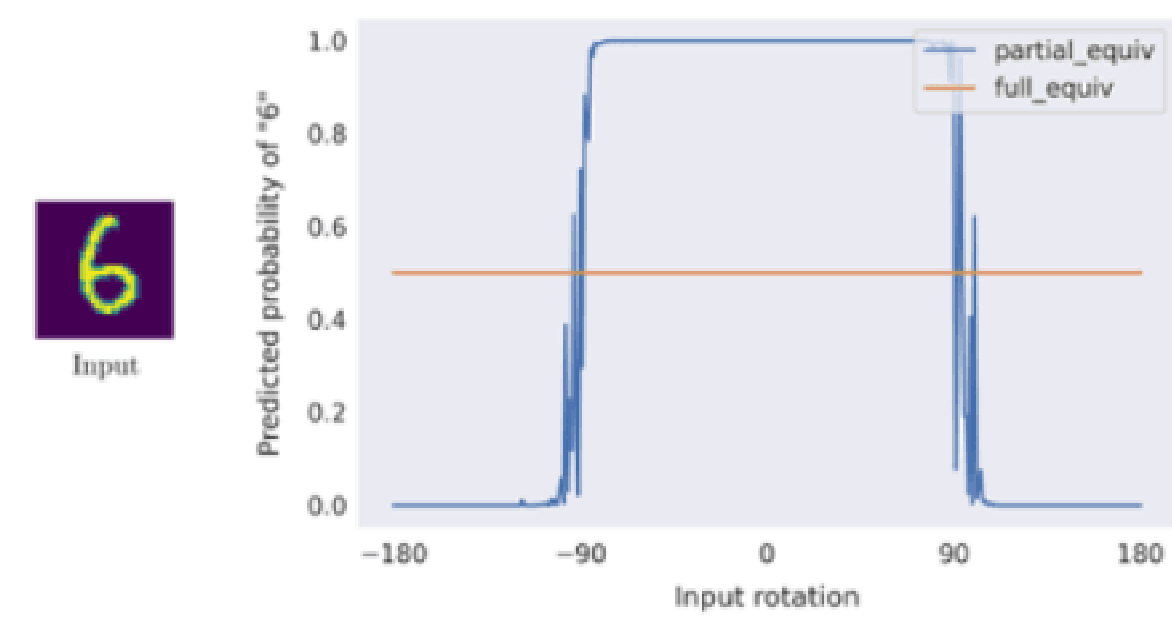


Figure 4. Learned equivariances on MNIST6-180. Partial G-CNNs learn to become equivariant to rotations on the semi-circle in order to solve the task. Regular G-CNNs, on the other hand, are unable to solve this task as it required setting group transformations apart.

Partial Group Equivariant Neural Networks

Learned subsets

We can look at the subsets learned by Partial G-CNNs

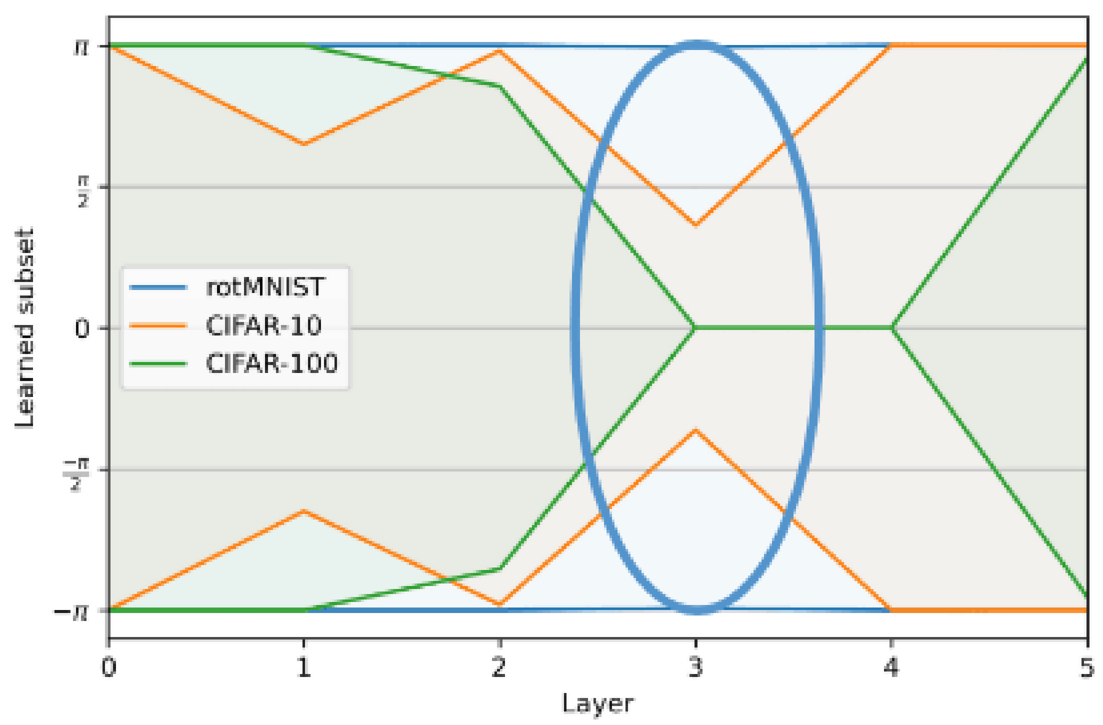


Table 2. Test accuracy on vision benchmark datasets.

Base Group	No. Elems	Partial Equiv.	Classification Accuracy (%) RotMNIST	CIFAR10	CIFAR100
T(2)	1	-	97.23	83.11	47.99
SE(2)	4	✗	99.10	83.73	52.35
		✓	**99.13**	**86.15**	**53.91**
	8	✗	99.17	86.08	55.55
		✓	**99.23**	**88.59**	**57.26**
	16	✗	**99.24**	86.59	51.55
		✓	99.18	**89.11**	**57.31**
E(2)	8	✗	**98.14**	85.55	54.29
		✓	97.78	**89.00**	**55.22**
	16	✗	98.35	88.95	57.78
		✓	**98.58**	**90.12**	**61.46**

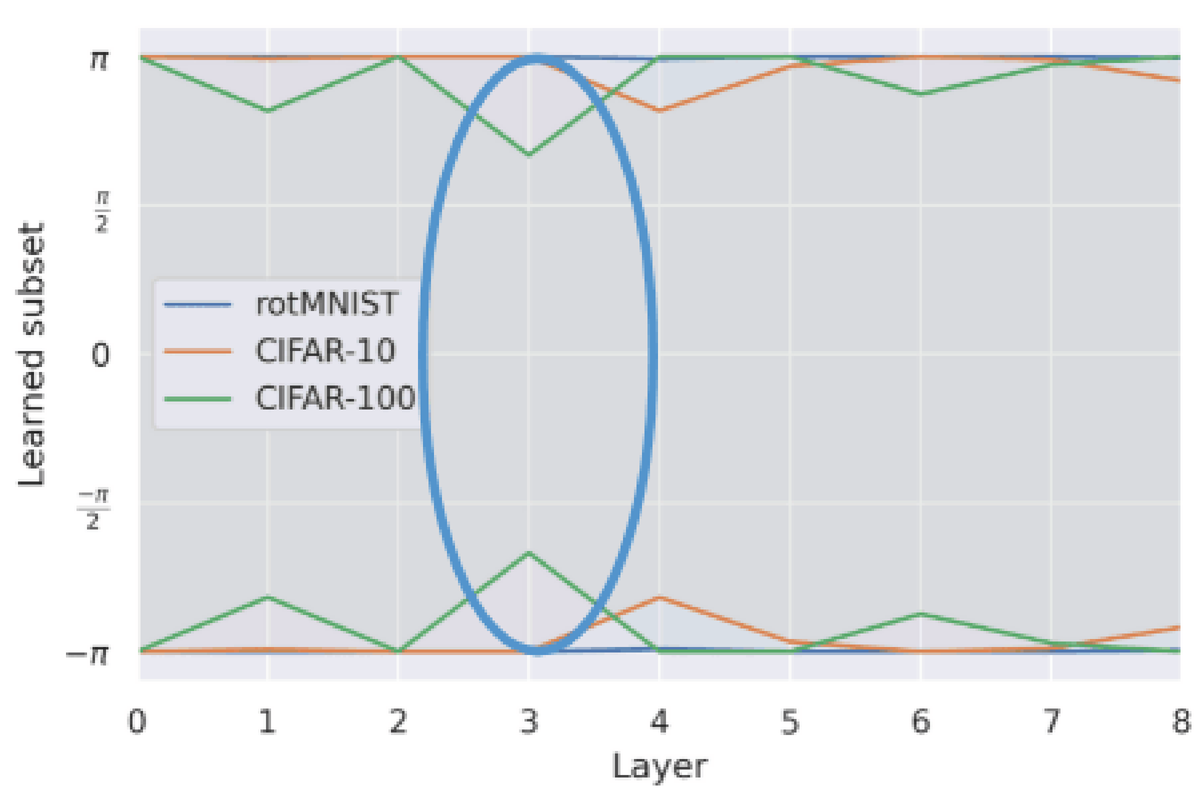


Table 5: Accuracy on vision benchmark datasets with (partial) group equivariant 13-layer CNNs [29].

Base Group	No. Elems	Partial Equiv.	Augerino	Classification Accuracy (%) RotMNIST	CIFAR10	CIFAR100
T(2)	1	-	-	96.90	91.21	67.14
SE(2)	2	✗	✗	98.70	85.51	62.06
		✗	✓	**98.94**	87.78	65.79
		✓	-	98.72	**92.48**	**66.72**
	4	✗	✗	98.43	89.73	65.97
		✗	✓	**98.94**	91.66	68.99
		✓	-	98.78	**92.28**	**69.83**
	8	✗	✗	98.54	90.55	67.70
		✗	✓	**99.28**	89.96	69.66
		✓	-	98.77	**91.99**	**70.80**

It is better to disrupt equivariance in the middle of the network!

Partial Group Equivariant Neural Networks

Conclusion

We presented a simple method with which the layer-wise partial equivariances can be learned.

It boils down to learning a probability distribution on the group, from which group elements are sampled during the group convolution.

We observe that Partial G-CNNs beat G-CNNs when equivariances are misspecified, and match them when these are correctly defined.

Thank you for your attention!

Questions?