Group Equivariant Convolutional Networks

From Discrete Symmetries to Harmonic Networks

Tristan Lecourtois   Léos Coutrot

ENS Paris-Saclay

January 28, 2026

Outline

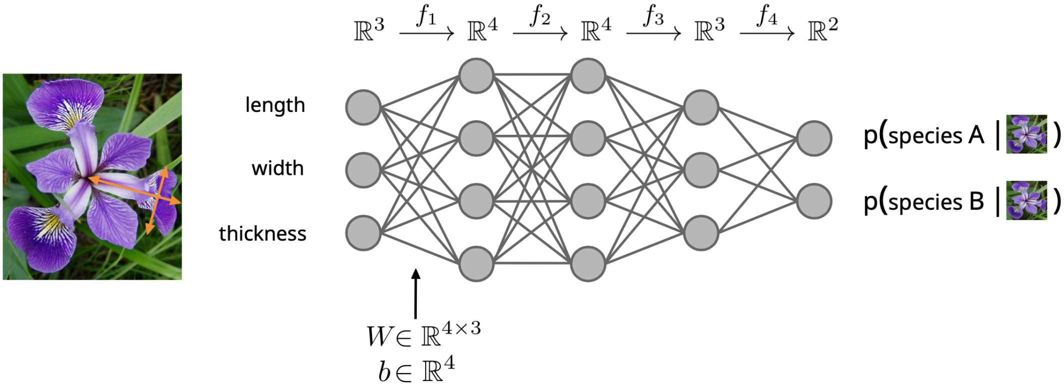
1. Introduction & Motivation
2. Discrete G-CNN Theory
3. Experiments: Discrete Equivariance
4. Continuous Symmetries: Harmonic Networks
5. Conclusion

Multi-Layer Perceptrons (MLPs)

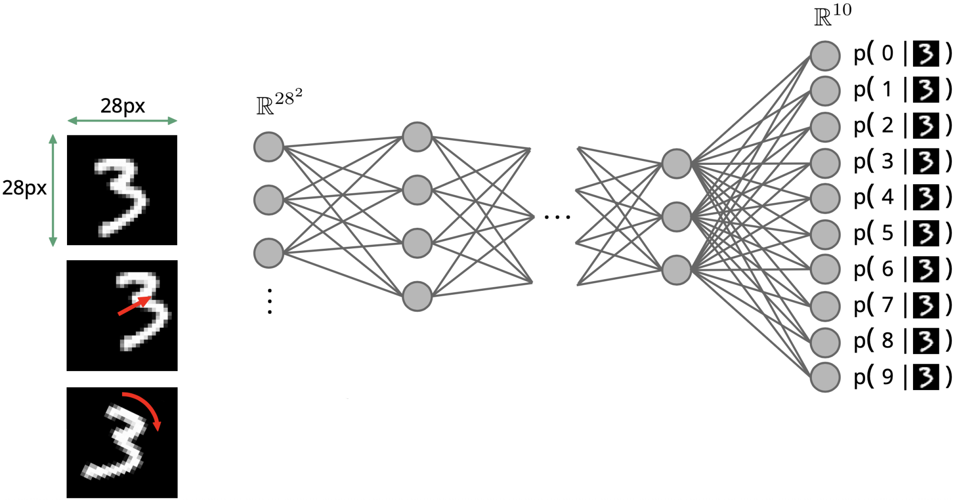
- Universal function approximators

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

- Composed of affine maps and nonlinearities: $\mathbf{x}_{i+1} = \sigma\left(W_i \mathbf{x}_i + \mathbf{b}_i\right)$



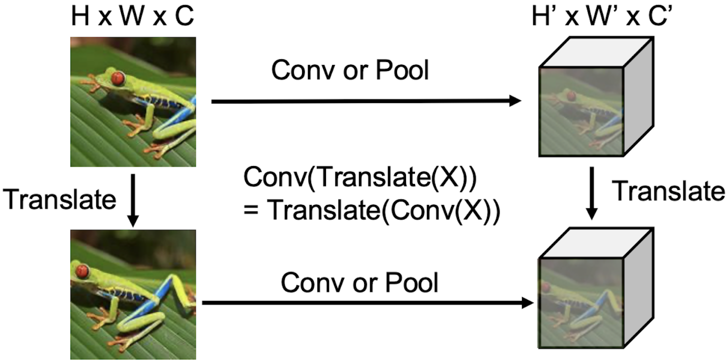
Using MLPs for image processing



- MLPs don't generalize over geometric transformations. They are ignorant of the geometric arrangements of pixels.

CNNs = MLPs + Equivariance

- **Local connectivity**: each neuron connects only to a local spatial neighborhood
- **Translational weight sharing**: the same filters are applied across spatial locations, making CNNs **translation equivariant**



The Problem with Standard CNNs

- **Geometric Symmetries:** Data often possesses natural symmetries (rotation, reflection).
- **Standard CNNs:** Only possess **translation equivariance**.
- **Consequence:** Models must learn rotations via *Data Augmentation*.
- **Cost:**
  - Increased parameter complexity.
  - Longer training times.
  - Redundant feature learning.

Goal

Design an architecture that is **equivariant by construction**, reducing sample complexity.

Background : Symmetry Groups

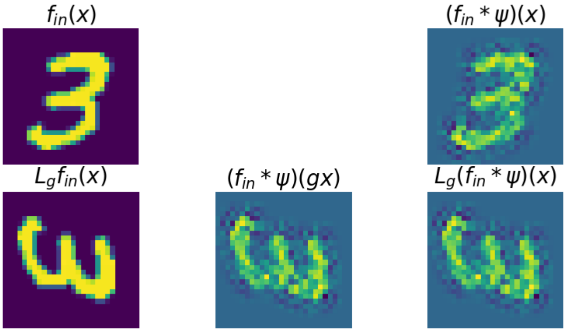
- A **group** is a non-empty set G equipped with a binary operation

$$\cdot : G \times G \to G$$

  satisfying:
  - **Associativity:** $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
  - **Identity:** $\exists\, e \in G$ s.t. $a \cdot e = e \cdot a = a$
  - **Inverse:** $\forall\, a \in G, \exists\, a^{-1} \in G$ s.t. $a \cdot a^{-1} = a^{-1} \cdot a = e$
- **Examples of symmetry groups in vision:**
  - $\mathbb{Z}^2$: 2D integer translations
  - $p4$ (p_n): translations + rotations by $\pi/2$ ($2\pi/n$)
  - $p4m$: translations, mirror reflections, and rotations by $\pi/2$

Background: Group Equivariance

- Let G be a symmetry group acting on the input space via $L_g : \mathcal{X} \to \mathcal{X}$.
- A function $f : \mathcal{X} \to \mathcal{Y}$ is G-**equivariant** if $f(L_g x) = L'_g f(x)$ for all $g \in G$, where L'_g denotes the induced action of G on the output space.



The Lifting Convolution (Layer 1)

Maps an image ($\mathbb{Z}^2$) to the Group (G).

$$[f * \psi](g) = \sum_{\mathbf{y} \in \mathbb{Z}^2} f_{in}(\mathbf{y})\, \psi(g^{-1}\mathbf{y})$$

- Filter ψ is applied in every orientation $r \in C_4$.
- **Output:** 4 feature maps, one for each rotation.

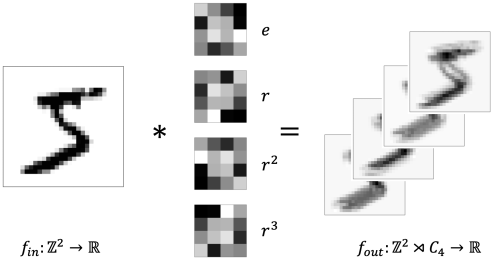


Figure 1: Input image convolved with 4 rotated filters.

Convolving signals already defined on the group G.

$$[f * \psi](g) = \sum_{h \in G} f(h)\,\psi(g^{-1}h)$$

- Input has spatial + orientation dims.
- Filters are now 3D (Spatial + Group).
- Preserves equivariance structure.

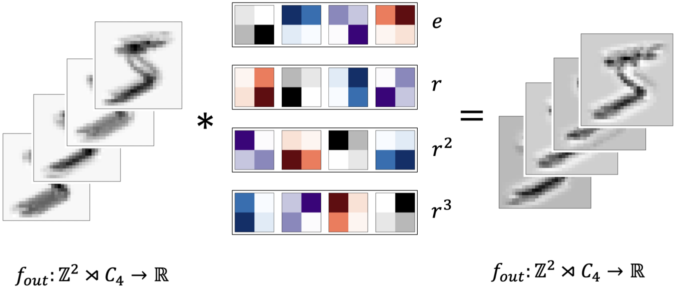


Figure 2: Previous layer maps (4 orientations) convolved to

Experimental Results: Accuracy

Comparison of baseline CNN vs. G-CNNs (P_n denotes equivariance to n rotations).

Dataset	**CNN**	**P4CNN**	**P8CNN**	**P16CNN**
MNIST	98.5	99.0	99.2	**99.3**
Rotated MNIST	92.0	96.0	97.2	**97.5**
CIFAR-10	81.3	83.0	83.5	**84.0**
Aug. CIFAR-10	82.0	84.2	85.0	**85.5**

Observation

Increasing the symmetry group ($P_2 \rightarrow P_{16}$) consistently improves accuracy, especially on Rotated MNIST.

Training Convergence

Equivariance leads to faster convergence as each sample is more "informative".

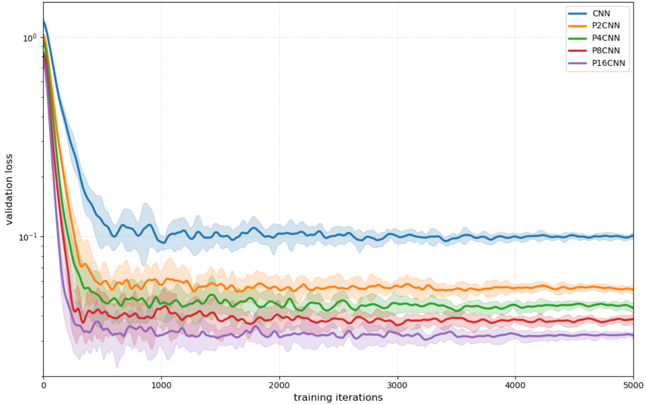


Figure 3: Validation Loss on Rotated MNIST. Equivariant models drop loss much earlier than standard CNNs.

Generalization Under Reduced Data

Hypothesis: Weight sharing across rotations acts as implicit data augmentation.

- **1000 samples:** G-CNN $> 90\%$, CNN $< 80\%$.
- **100 samples:** G-CNN $\approx 20\%$, CNN $\approx 10\%$ (random guessing).

G-CNNs effectively see $4\times$ the data due to symmetry constraints.

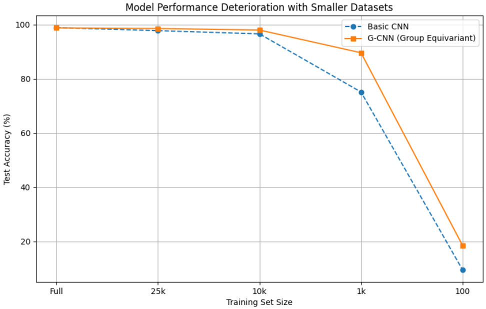


Figure 4: Accuracy vs. Training Set Size.

From Discrete to Continuous

Limitation of G-CNN

P_4 or P_{16} only handle discrete rotations. Real-world objects rotate continuously ($SO(2)$).

Solution: Harmonic Networks Use complex-valued filters W with a specific phase structure:

$$W_{\Delta m}(r, \phi) = R(r)\, e^{i(\Delta m\, \phi + \beta)}$$

Convolution with a rotated input results in a predictable phase shift:

$$W_{\Delta m} * (R_\theta F) = e^{i\Delta m \theta}(W_{\Delta m} * F)$$

Visualizing Harmonic Filters

The network learns radial profiles $R(r)$ while the angular part is fixed analytically.

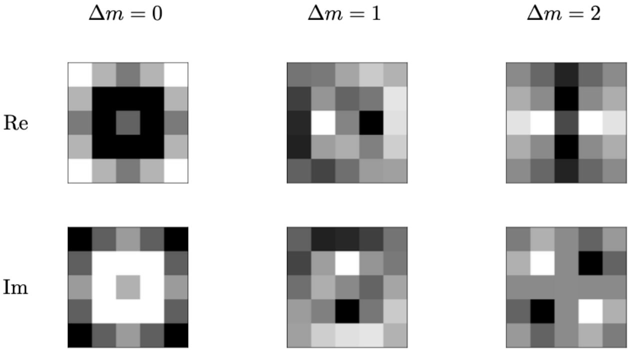


Figure 5: Learned filters. $\Delta m = 0$ (isotropic), $\Delta m = 1$ (edges), $\Delta m = 2$ (complex patterns).

Results: Harmonic vs Discrete

Harmonic Networks (HN) capture continuous rotation, outperforming fixed discrete groups.

Dataset	**P4CNN**	**HN (2 streams)**	**HN (3 streams)**
MNIST	99.0	99.2	99.4
Rotated MNIST	96.0	**98.3**	98.3
CIFAR-10	83.0	84.7	**85.8**

Takeaway: Continuous equivariance (HN) yields better generalization than discrete equivariance (P4).

Summary

1. **Equivariance by Design:** Reduces parameter count and improves robustness without data augmentation.
2. **Discrete G-CNNs:**
   - Drastically improve sample efficiency (works with 10x less data).
   - Improves significantly the overall accuracy of the models.
3. **Harmonic Networks:**
   - Extend this to continuous $SO(2)$ rotations.
   - Achieve the best performance out of all the models tested.

Thank You