Transparent Vision: A Theory of Hierarchical Invariant Representations

ICCV HONOLULU HAWAII OCT 19-23, 2025

Shuren Qi, Yushu Zhang, Chao Wang, Zhihua Xia, Xiaochun Cao, and Fenglei Fan

Center for Mathematical Artificial Intelligence CMAI

0. TL;DR

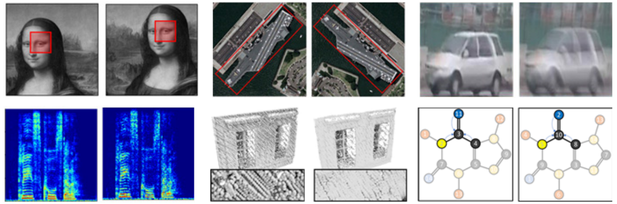
When deep nets get too complicated, I resort to toy models: a *designed* CNN with *hierarchical invariance.*

spherical cow

1. Why Invariance?

Empirical learning *v.s.* robustness, interpretability, efficiency…

- Integrating invariance into representations



Invariance is ubiquitous!

2. State of the Art and Motivation

Invariance helps deep nets

- MLP (no inv.) → CNN (translation inv. on grid) → Geometric Deep Learning (GDL, going beyond translation or grid)

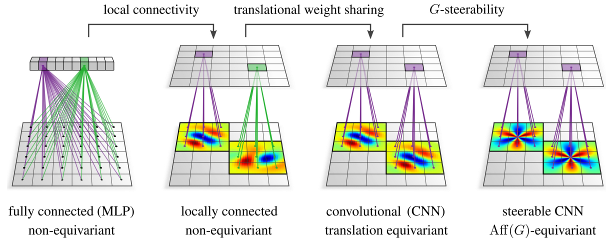
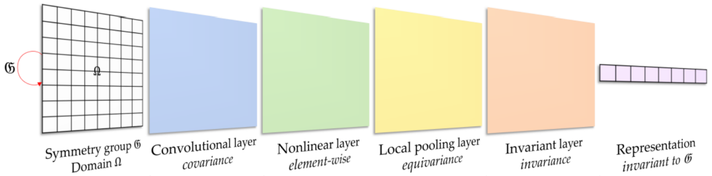


Image credit M. Weiler's book

- Equivariant networks: SOTA in GDL but need *discrete sampling*, with limited efficiency & invariance.

3. The Blueprint

Rethink CNN modules and formalize a hierarchical invariant blueprint.



Need Local Invariants | *Need Global Invariants*

[1] S. Qi et al. A Principled Design of Image Representation. *TPAMI*, 2023.

[2] J. Flusser et al. *Moments and Moment Invariants in Pattern Recognition*, 2009.

Image information can pass through each inter CNN layer in a *geometrically controllable manner*, and on the last layer, the invariants are allowed by compact designs with sufficient information.

4. Definitions

$\mathfrak{G}$-invariant representation $\mathcal{R}_p \triangleq \mathbb{I} \circ \mathbb{P}_{[L]} \circ \mathbb{S}_{[L]} \circ \mathbb{C}_{[L]} \circ \cdots \circ \mathbb{P}_{[1]} \circ \mathbb{S}_{[1]} \circ \mathbb{C}_{[1]}$

- $\mathfrak{G}$-covariant convolutional layer $\mathbb{C}M \triangleq \langle M, V_{nm}^{uvw} \rangle$ [1]
- Nonlinearity layer $\mathbb{S}M = \sigma(M(i,j)) \triangleq |M(i,j;k)|$
- Local pooling layer $\mathbb{P}M = M'$
- $\mathfrak{G}$-invariant layer $\mathbb{I}M = \mathcal{I}(\{\langle M(i,j;k), V_{nm}(x_i, y_j) \rangle\})$ [2]

Ω: 2D image grid; $\mathfrak{G}$: translation, rotation, flipping, scaling symmetry group over Ω

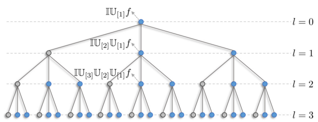
5. Properties

One-shot symmetry properties between image and representation

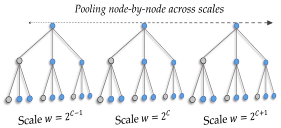
- $\mathfrak{G}_1$ equivariance $\mathbb{U}_{[L]} \circ \cdots \circ \mathbb{U}_{[2]} \circ \mathbb{U}_{[1]}(\mathfrak{g}_1 M) \equiv \mathfrak{g}_1 \mathbb{U}_{[L]} \circ \cdots \circ \mathbb{U}_{[2]} \circ \mathbb{U}_{[1]}(M)$
- $\mathfrak{G}_2$ covariance $\mathbb{U}^w_{[L]} \circ \cdots \circ \mathbb{U}^w_{[2]} \circ \mathbb{U}^w_{[1]}(\mathfrak{g}_2 M) \equiv \mathfrak{g}'_2 \mathbb{U}^w_{[L]} \circ \cdots \circ \mathbb{U}^w_{[2]} \circ \mathbb{U}^w_{[1]}(M)$
- $\mathfrak{G}_0$ hierarchical invariance $\mathbb{I}(\mathfrak{g}'_0 M)_{[L]} \equiv \mathbb{I}M_{[L]}$

$\mathfrak{G}_1$: translation, rotation, flipping symmetry group; $\mathfrak{G}_2$: scaling symmetry group, with scaling factor s; any $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \times \mathfrak{G}_2$ as the symmetry group of interest; $\mathbb{U} \triangleq \mathbb{P} \circ \mathbb{S} \circ \mathbb{C}$; $\mathfrak{g}'_2$ means acting also on the w of $\mathbb{U}$: $\mathfrak{g}'_2 \mathbb{U}^w \triangleq \mathfrak{g}_2 \mathbb{U}^{ws}$

6. Practices



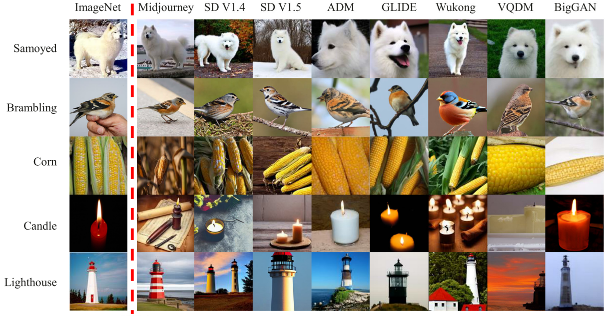
A Single-scale Practice with $\mathfrak{G}_1$ Invariance

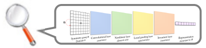


A Multi-scale Practice with $\mathfrak{G}_0$ Invariance

7. An Application Case: AIGC Detection

Similar discriminability to large CNN, with much fewer training samples and new robustness/interpretability.





baseline CNN *vs.* invariant CNN
74.10% *vs.* 94.73%

8. Useful Links

[code](#)	[tutorial](#)	[profile](#)	**Online**
this paper	bigger picture	homepage	**Q&A**