

Tutorial Outline

- **Part 1:** Background and challenges (20 min)
- **Part 2:** Preliminaries of invariance (20 min)
- *Q&A / Break (10 min)*
- **Part 3:** Invariance in the era before deep learning (30 min)
- **Part 4:** Invariance in the early era of deep learning (10 min)
- *Q&A / Coffee Break (30 min)*
- **Part 5: Invariance in the era of rethinking deep learning (50 min)**
- **Part 6:** Conclusions and discussions (20 min)
- *Q&A (10 min)*

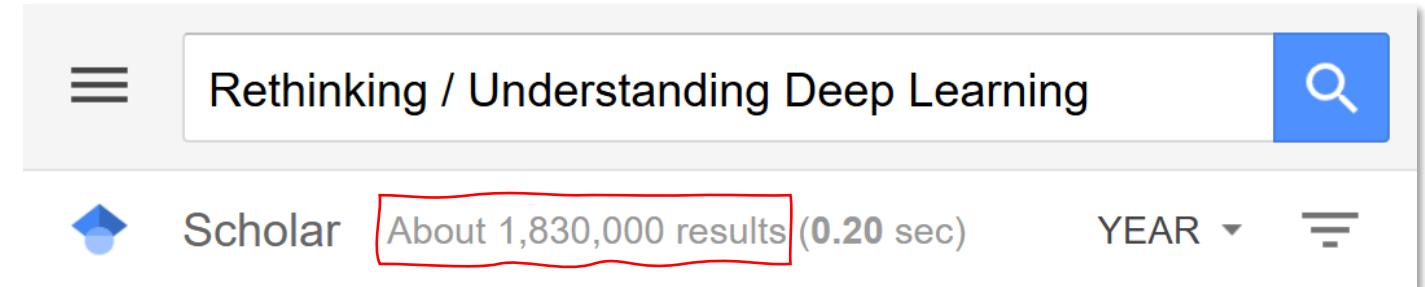
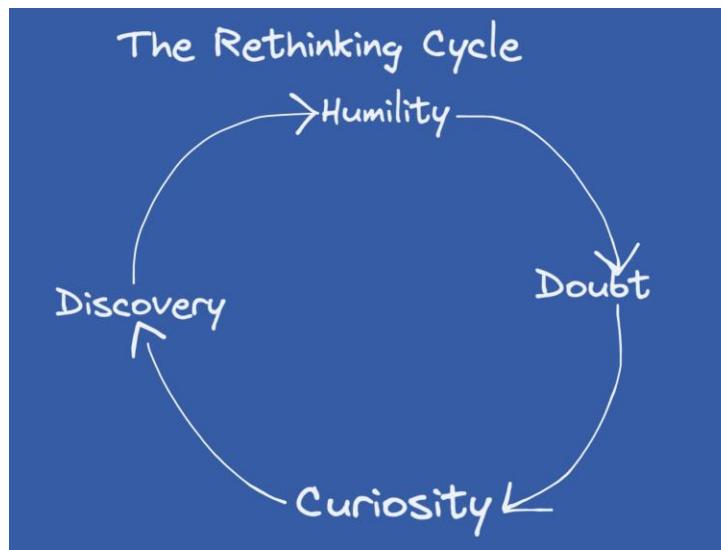


A Historical Perspective of Data Representation
Rethinking Deep Learning with Invariance: The Good, The Bad, and The Ugly

Why Rethink Deep Learning? Realistic Needs And Theoretical Extensions

Invariance in The Era of Rethinking Deep Learning

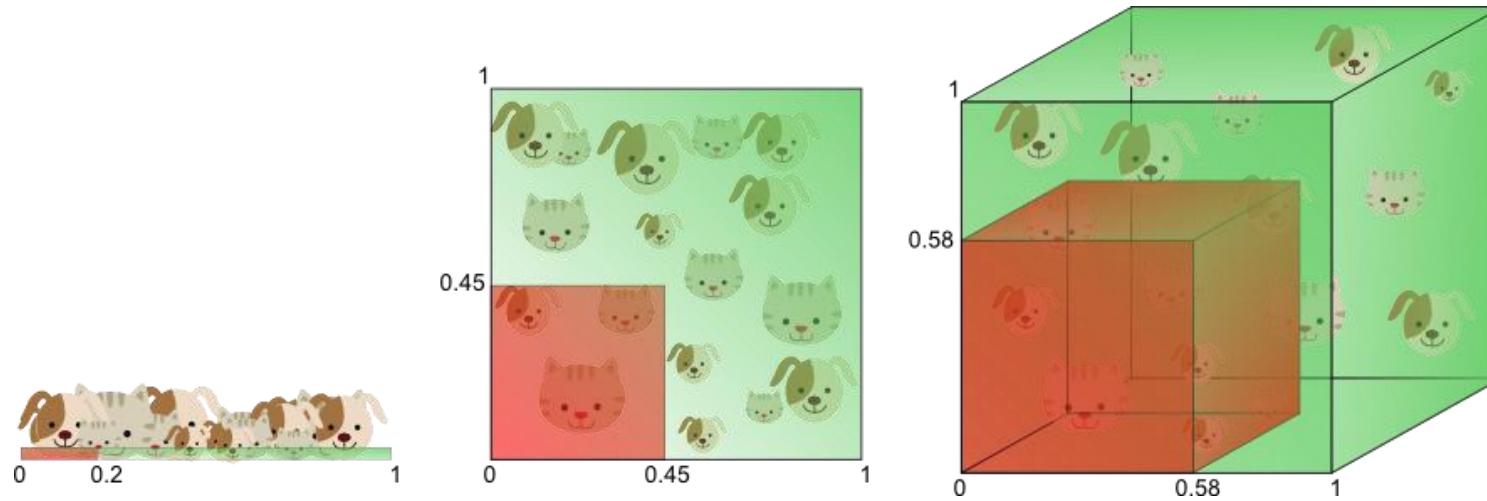
- Realistic Needs: Bottlenecks in **robustness, interpretability, and efficiency**.
- Theory Extensions: A unified theory perspective to avoid getting into **endless experimental designs**.



Why Invariance Is A Must For The Rethinking? Universal Approximation vs. The Curse Of Dimensionality

Universal Approximation vs. The Curse of Dimensionality

- Universal Approximation: “Any 2-layer perceptron can approximate a continuous function to any desired accuracy”.
- The Curse of Dimensionality: The required number of learning samples increases sharply with the dimension, until it is out of feasibility.
- Invariance: Inherent structure of the data, reducing the need for unnecessary and impractical learning, just like from NN to CNN.



Geometry Deep Learning

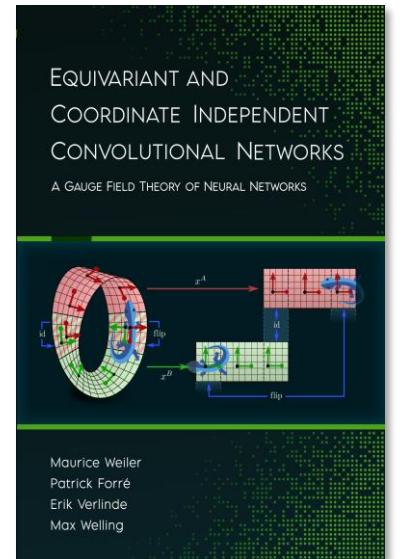
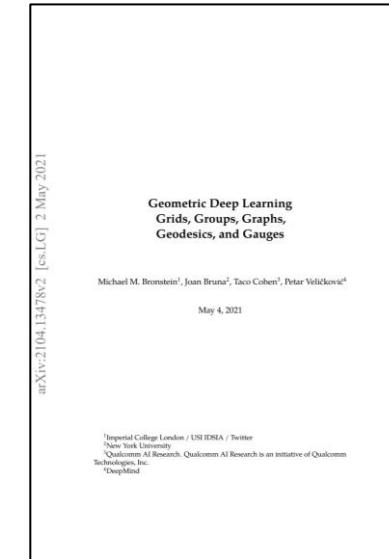
Rethinking From The Lens Of Invariance

Geometric Deep Learning

- **Geometric Deep Learning** is a way to **rethink deep learning from the lens of invariance**:
 - Extending hierarchical invariance to transformations beyond translations. CNN is already invariant to translation, how to generalize this success to rotation, scaling
 - Extending hierarchical invariance to data beyond images. CNN works well on images, how to generalize this success to sets, graphs, surfaces
 - Harmonizing the existing learning architectures with invariance-principled designs. CNN has translation invariance, how about invariance of LSTM, GNN, Transformer

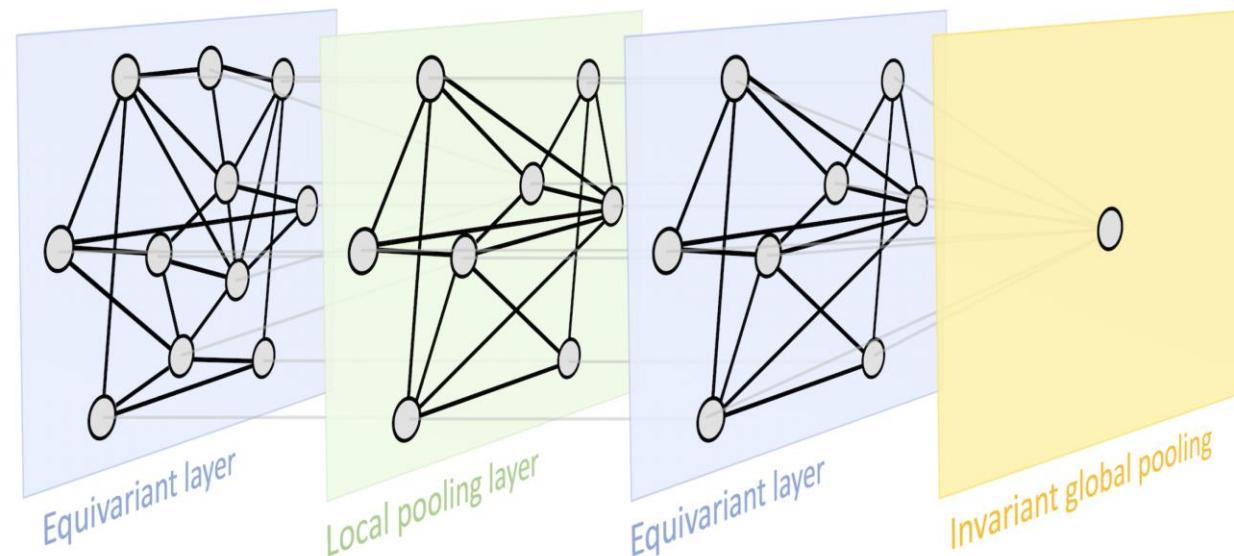


M. M. Bronstein, J. Bruna, T. Cohen, & P. Veličković, 2017
Geometry Deep Learning



Blueprint: High-level Intuition

- A blueprint for achieving a unified hierarchical invariance over different transformations, architectures, and data types.
- Equivariant local representations as the **inter-links**; invariant global representations as the **final-links**, which together form a **hierarchical invariant representation**.



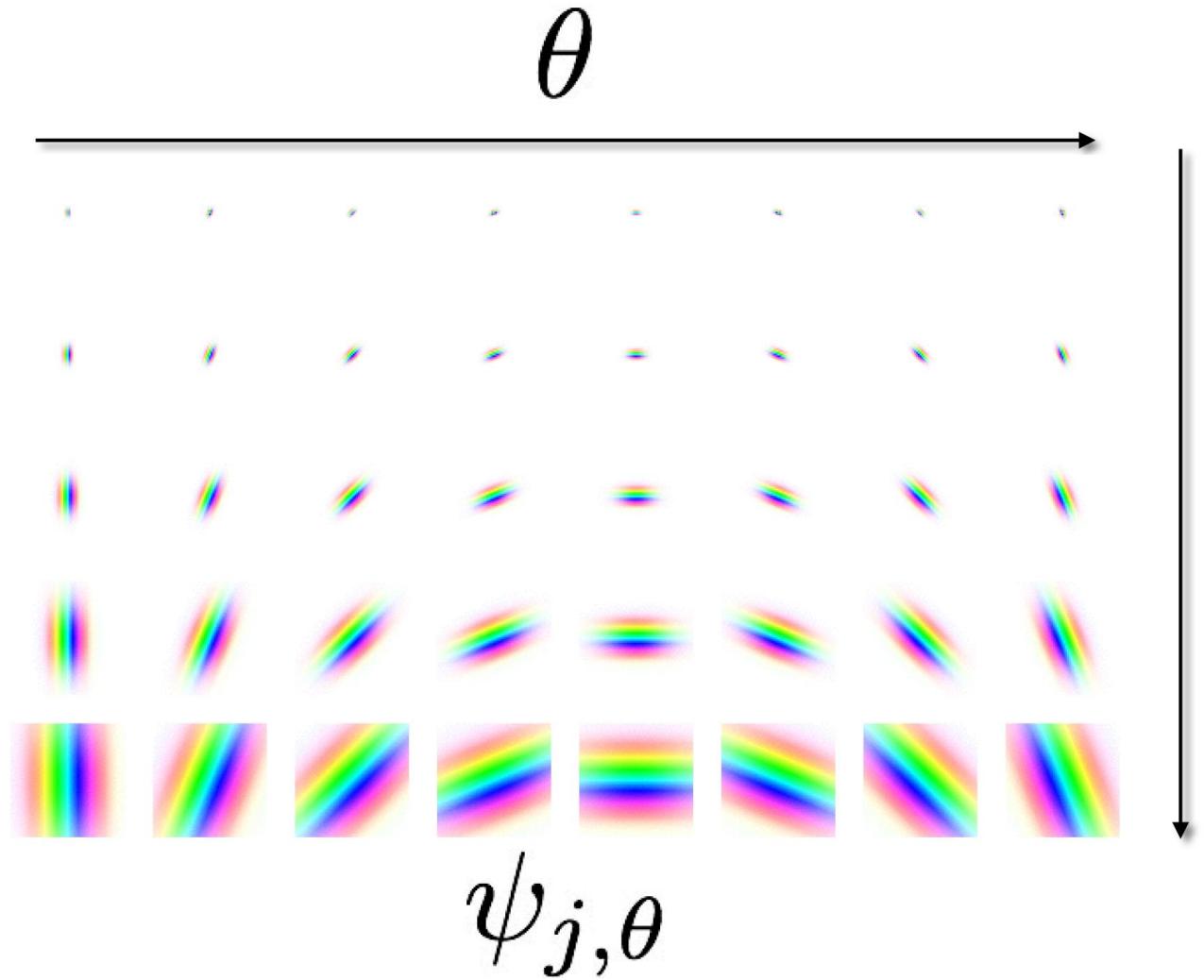
Blueprint: Formalization

- Let Ω be domain and \mathfrak{G} a symmetry group over Ω .
 - \mathfrak{G} -equivariant layer B :** $X(\Omega, C) \rightarrow X(\Omega', C')$, $B(g \cdot x) = g \cdot B(x)$ for all $g \in \mathfrak{G}$ and $x \in X(\Omega, C)$.
 - Nonlinearity σ :** $C \rightarrow C'$ applied element-wise as $(\sigma(x))(u) = \sigma(x(u))$.
 - Local pooling P :** $X(\Omega, C) \rightarrow X(\Omega', C)$, such that $\Omega' \subseteq \Omega$ as a compact version of Ω .
 - \mathfrak{G} -invariant layer A :** $X(\Omega, C) \rightarrow Y$, $A(g \cdot x) = A(x)$ for all $g \in \mathfrak{G}$ and $x \in X(\Omega, C)$.
 - \mathfrak{G} -invariant functions f :** $X(\Omega, C) \rightarrow Y$, $f = A \circ \sigma_N \circ B_N \circ \dots \circ P_1 \circ \sigma_1 \circ B_1$

Architecture	Domain Ω	Symmetry group \mathfrak{G}
CNN	Grid	Translation
<i>Spherical CNN</i>	Sphere / $SO(3)$	Rotation $SO(3)$
<i>Intrinsic / Mesh CNN</i>	Manifold	Isometry $Iso(\Omega)$ / Gauge symmetry $SO(2)$
GNN	Graph	Permutation Σ_n
<i>Deep Sets</i>	Set	Permutation Σ_n
<i>Transformer</i>	Complete Graph	Permutation Σ_n
LSTM	1D Grid	Time warping

Geometric Deep Learning For Different Transformations

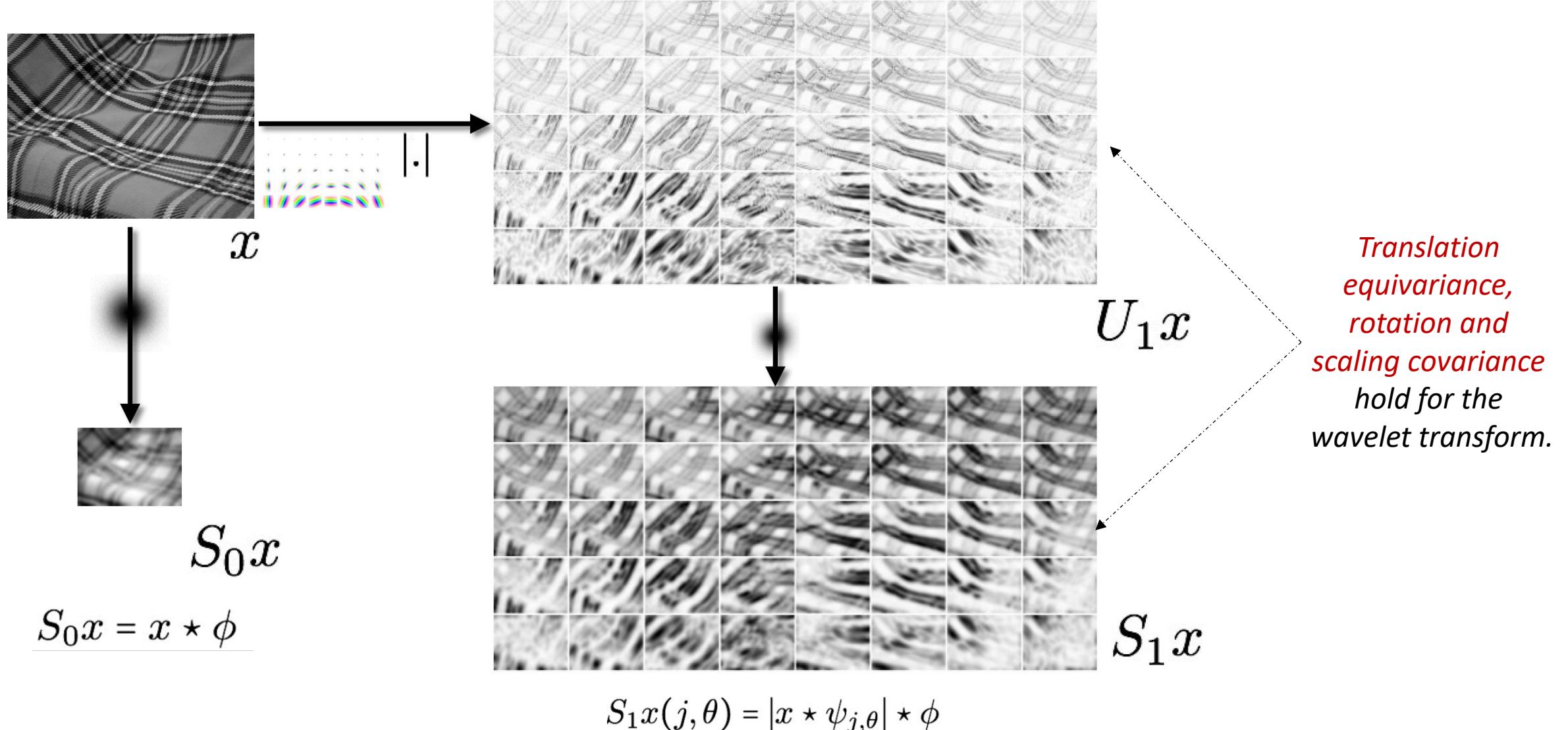
Beyond Translations: Wavelet Scattering Networks



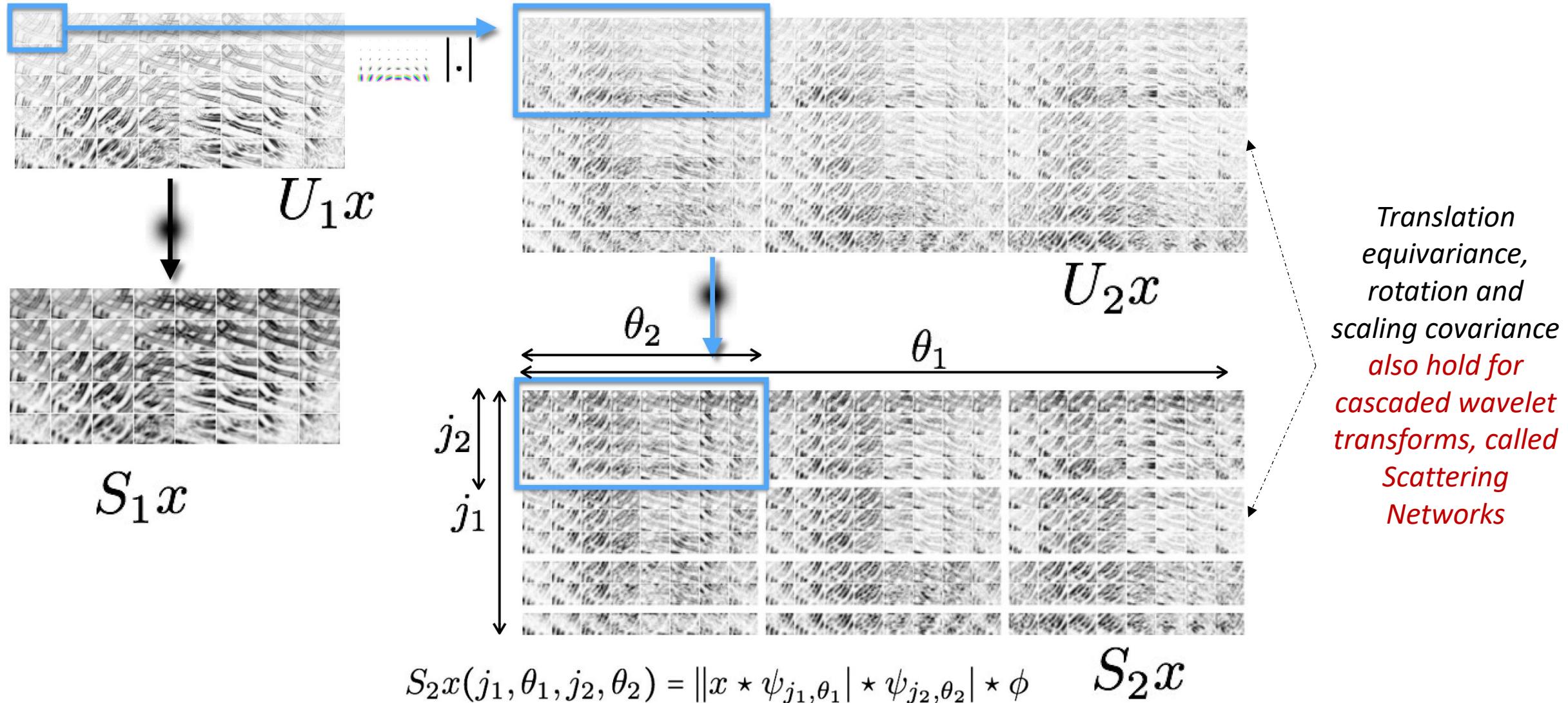
*By fully considering **the translation, rotation and scaling symmetry group**, the wavelet basis functions can be well designed to achieve invariance.*

- J Bruna, S Mallat. Invariant scattering convolution networks. *TPAMI*, 2013.
- J Bruna, S Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. *CVPR*, 2013.

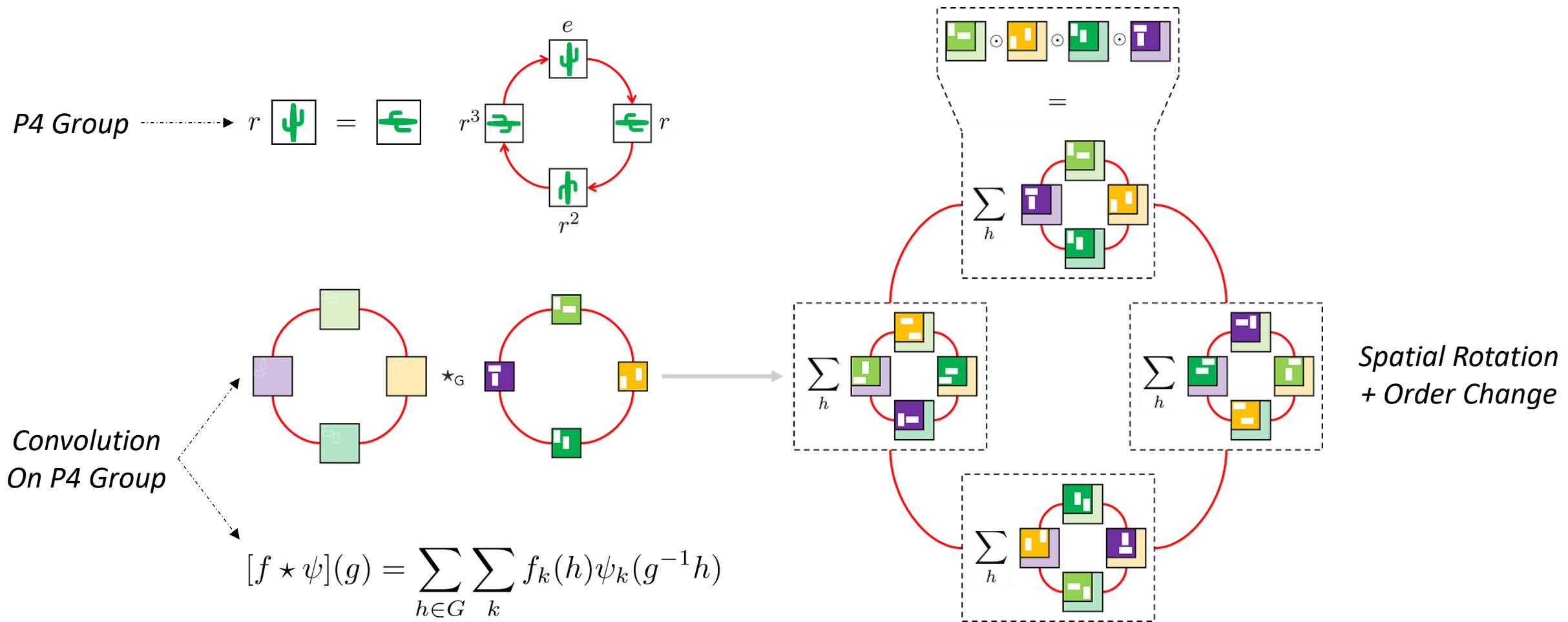
Beyond Translations: Wavelet Scattering Networks



Beyond Translations: Wavelet Scattering Networks



Beyond Translations: Group Equivariant Networks



Beyond Translations: Group Equivariant Networks

$$[f \star \psi^i](x) = \sum_{\substack{y \in \mathbb{Z}^2 \\ k=1}} \sum_{k=1}^{K^l} f_k(y) \psi_k^i(y - x)$$

Convolution

$$\begin{aligned} [[L_t f] \star \psi](x) &= \sum_y f(y-t) \psi(y-x) \\ &= \sum_y f(y) \psi(y+t-x) \\ &= \sum_y f(y) \psi(y-(x-t)) \\ &= [L_t[f \star \psi]](x). \end{aligned}$$

Translation Equivariance

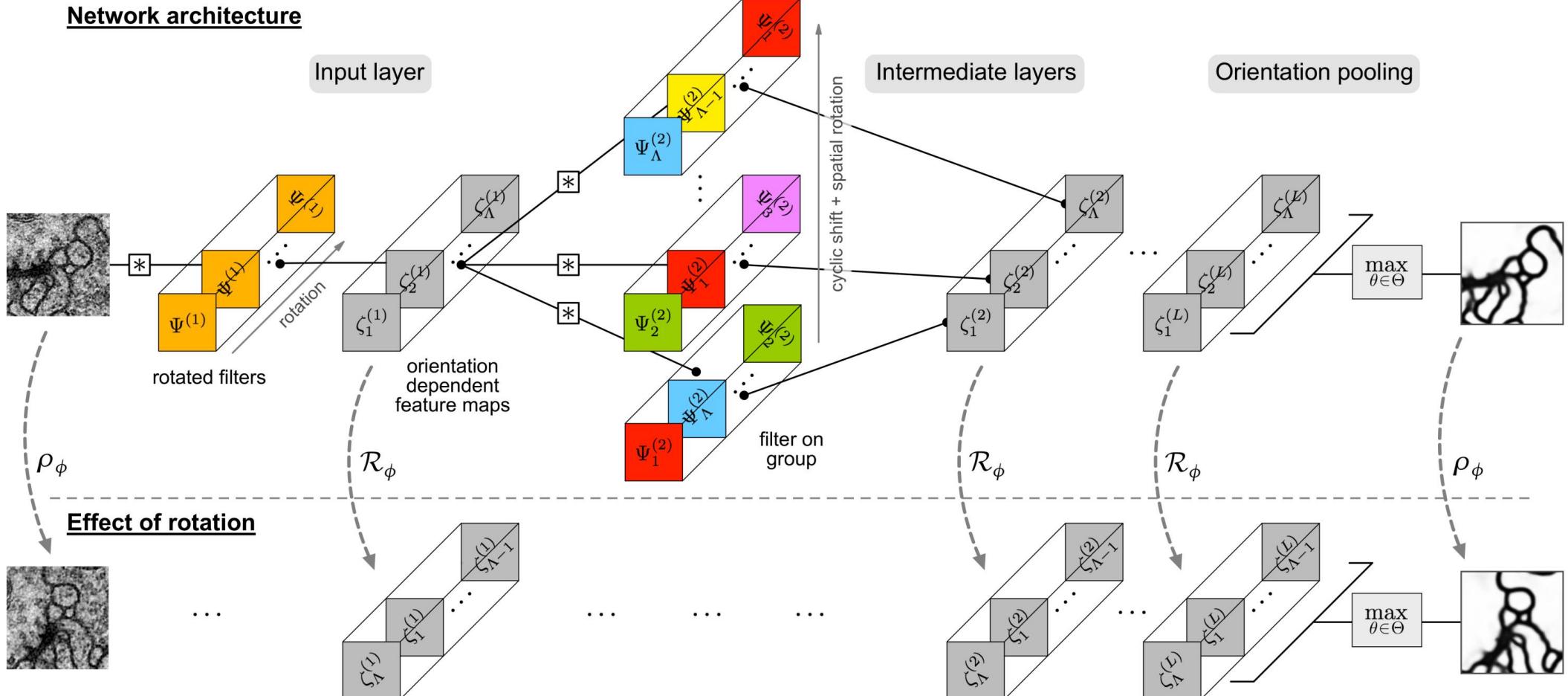
$$[f \star \psi](g) = \sum_{\substack{h \in G \\ k}} f_k(h) \psi_k(g^{-1}h)$$

Group Convolution

$$\begin{aligned} [[L_u f] \star \psi](g) &= \sum_{h \in G} \sum_k f_k(u^{-1}h) \psi(g^{-1}h) \\ &= \sum_{h \in G} \sum_k f(h) \psi(g^{-1}uh) \\ &= \sum_{h \in G} \sum_k f(h) \psi((u^{-1}g)^{-1}h) \\ &= [L_u[f \star \psi]](g) \end{aligned}$$

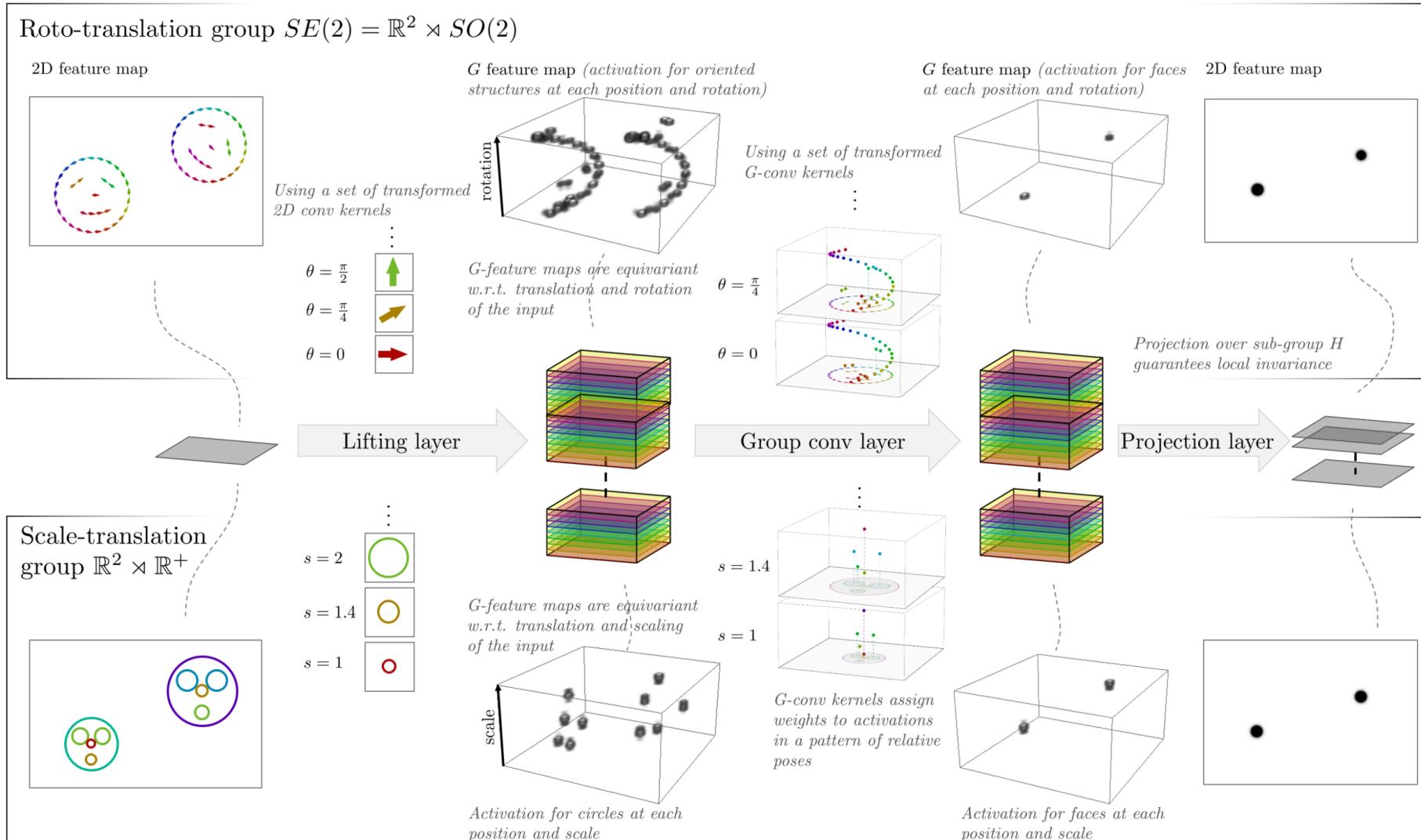
Group Equivariance

Beyond Translations: Group Equivariant Networks



- M Weiler, FA Hamprecht, M Storath. Learning steerable filters for rotation equivariant CNNs. *CVPR*, 2018.

Beyond Translations: Group Equivariant Networks



Geometric Deep Learning For Different Architectures And Data Types

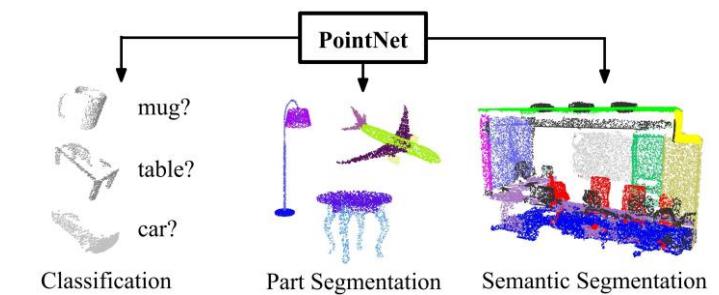
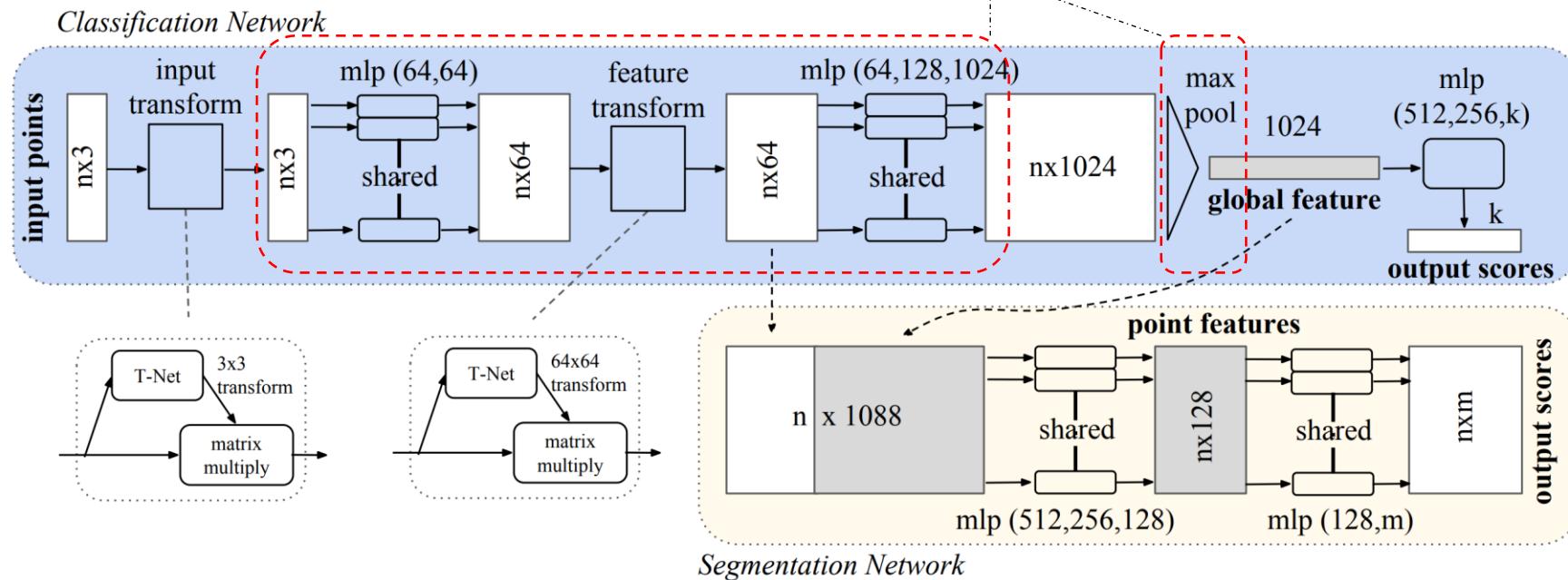
Beyond Images: Deep Sets and PointNet

$$f(\{x_1, \dots, x_M\}) = f(\{x_{\pi(1)}, \dots, x_{\pi(M)}\}) \quad \textit{Permutation Invariance on Set}$$

$$\mathbf{f}([x_{\pi(1)}, \dots, x_{\pi(M)}]) = [f_{\pi(1)}(\mathbf{x}), \dots, f_{\pi(M)}(\mathbf{x})] \quad \textit{Permutation Equivariance on Set}$$

$$\rho \left(\sum_{x \in X} \phi(x) \right)$$

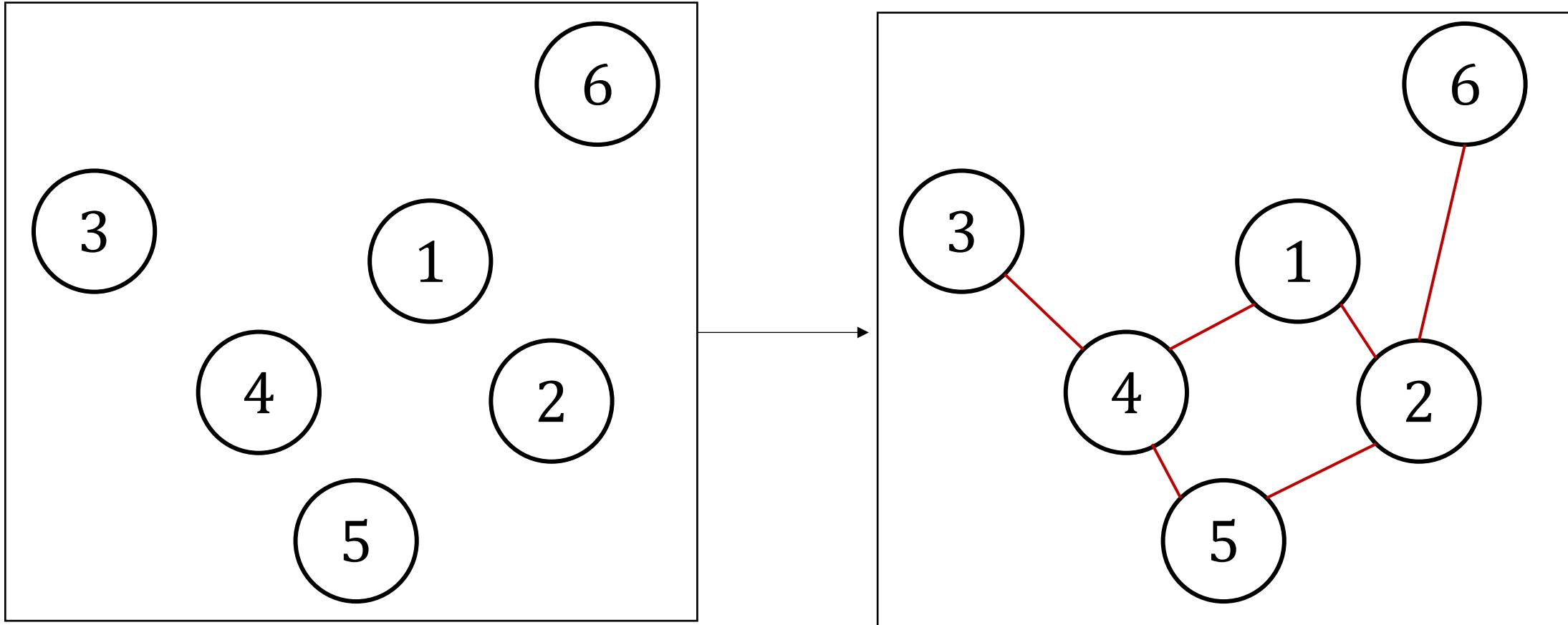
Deep Sets, Point-wise ϕ for Permutation Equivariance and Global Pooling Σ for Permutation Invariance



PointNet: A Practice of Deep Sets

- M Zaheer, S Kottur, S Ravanbakhsh. Deep sets. *NIPS*, 2017.
- CR Qi, H Su, K Mo, et al. PointNet: Deep learning on point sets for 3D classification and segmentation. *CVPR*, 2017.

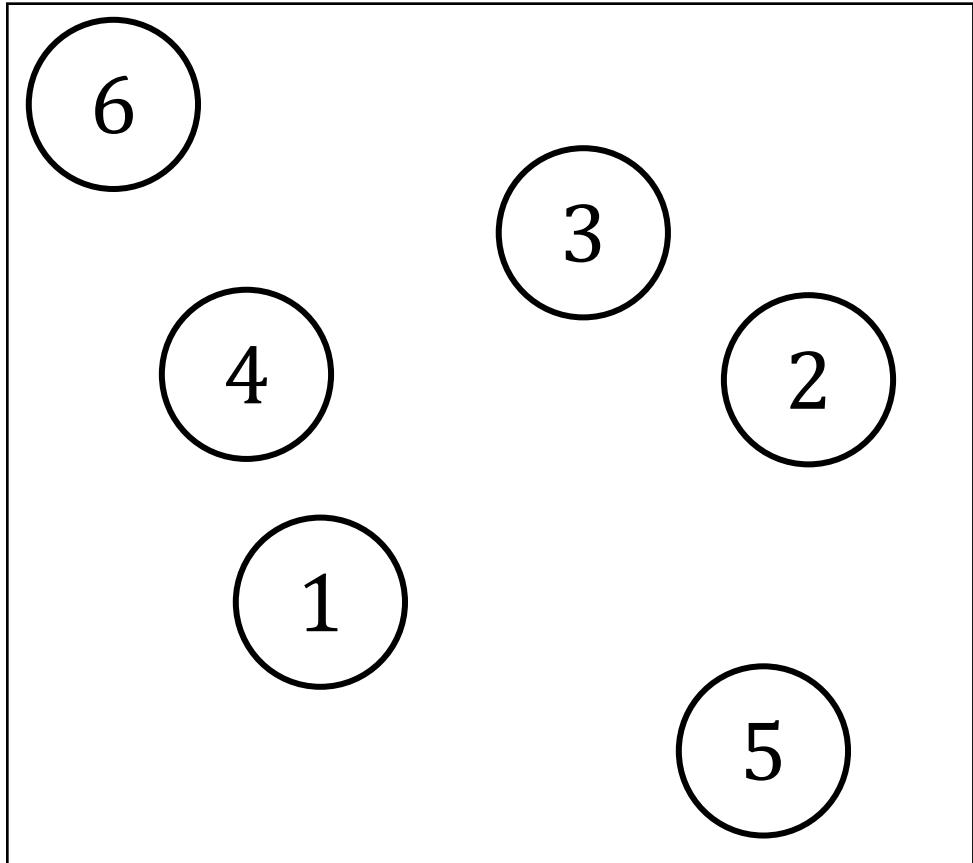
Beyond Images: Graph Networks



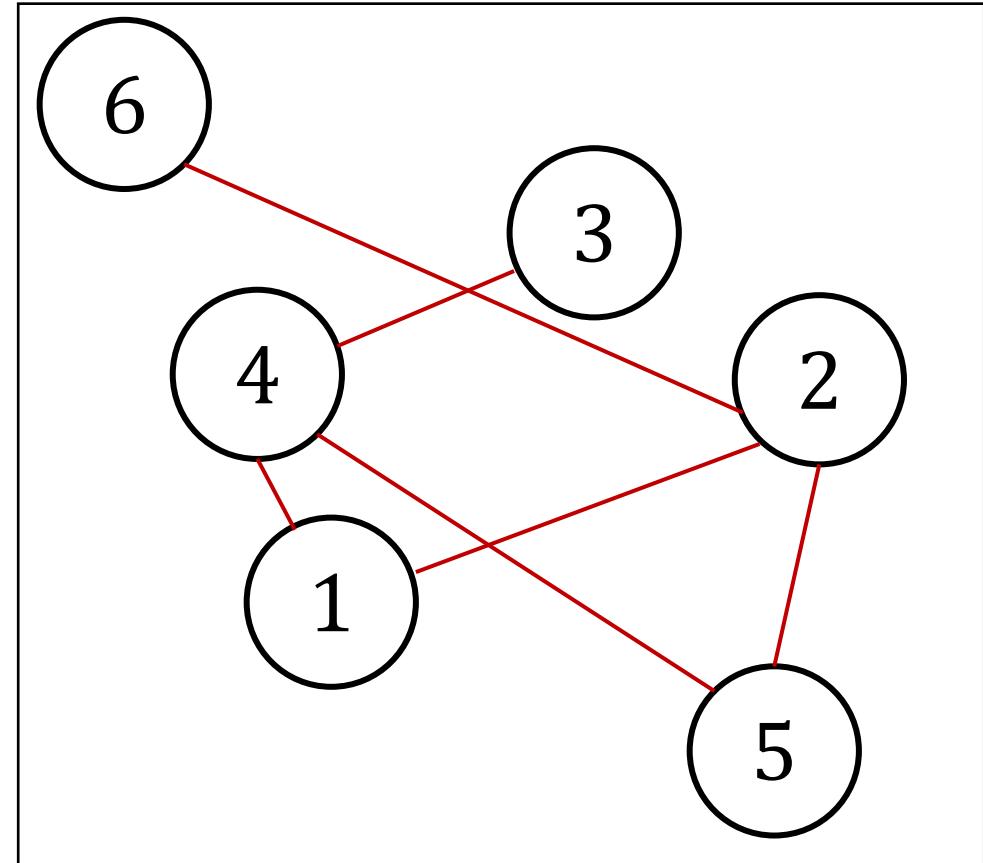
Set: $S = \{1, 2, \dots, 6\}$

Graph: $G = \{X, A\}$,
 $X = S, A = \{\{1, 2\}, \dots, \{2, 6\}\}$

Beyond Images: Graph Networks

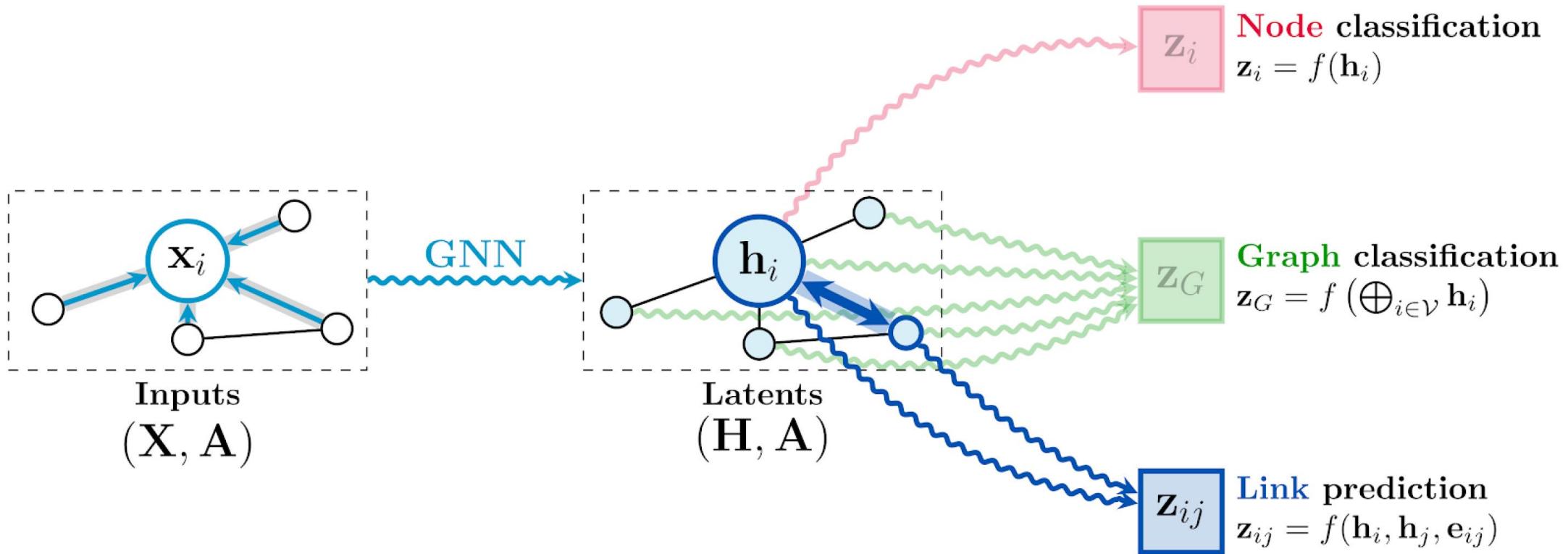


Set Permutation
Just order changes



Graph Permutation
Node order changes with edge order changes

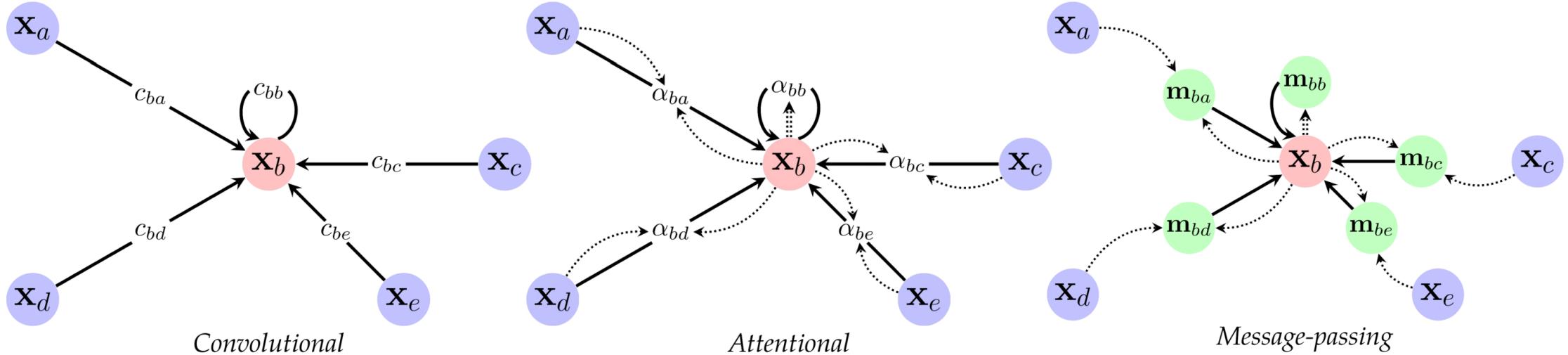
Beyond Images: Graph Networks



Local and Global Applications of GNN

- TN Kipf, M Welling. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.
 - P Veličković, G Cucurull, A Casanova, et al. Graph attention networks. *ICLR*, 2018.
- J Gilmer, SS Schoenholz, PF Riley, et al. Neural message passing for quantum chemistry. *ICML*, 2017.

Beyond Images: Graph Networks



$$\mathbf{h}_i = \phi \left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} c_{ij} \psi(\mathbf{x}_j) \right)$$

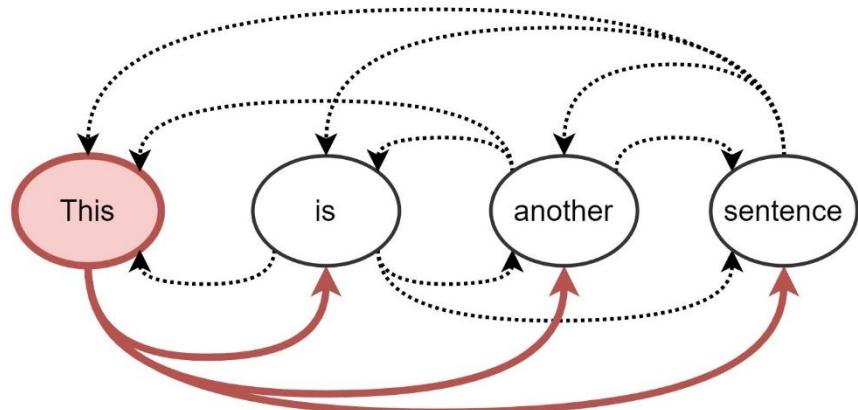
$$\mathbf{h}_i = \phi \left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j) \psi(\mathbf{x}_j) \right)$$

$$\mathbf{h}_i = \phi \left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} \psi(\mathbf{x}_i, \mathbf{x}_j) \right)$$

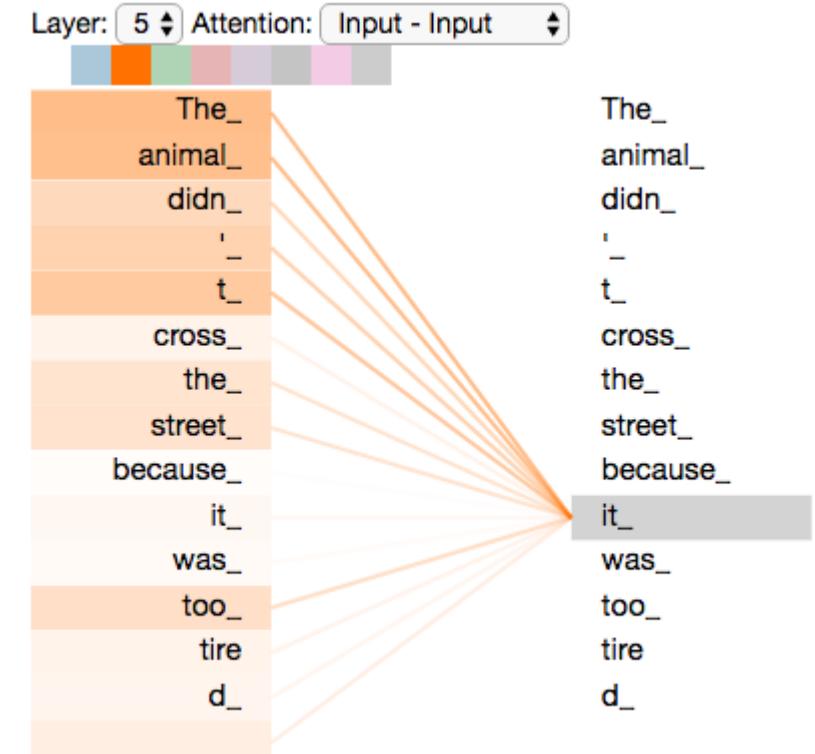
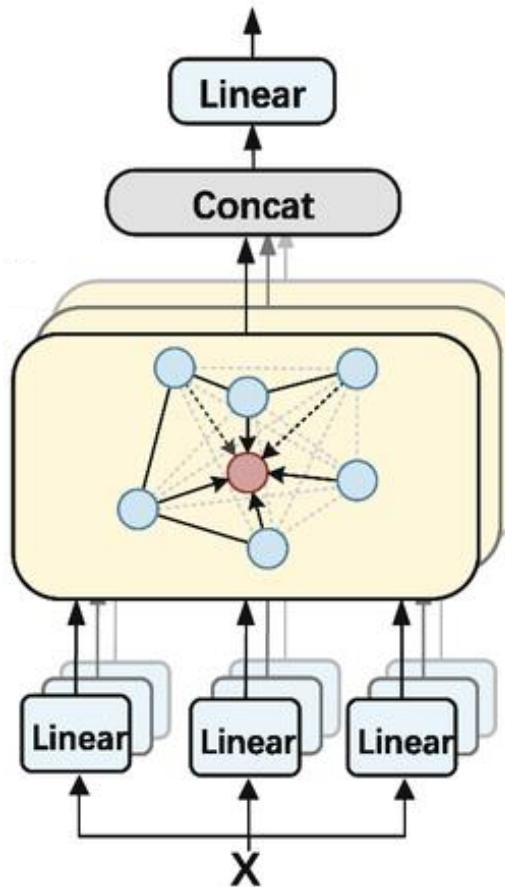
Permutation Equivariant GNN Layers

- TN Kipf, M Welling. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.
 - P Veličković, G Cucurull, A Casanova, et al. Graph attention networks. *ICLR*, 2018.
- J Gilmer, SS Schoenholz, PF Riley, et al. Neural message passing for quantum chemistry. *ICML*, 2017.

Beyond Images: Transformers

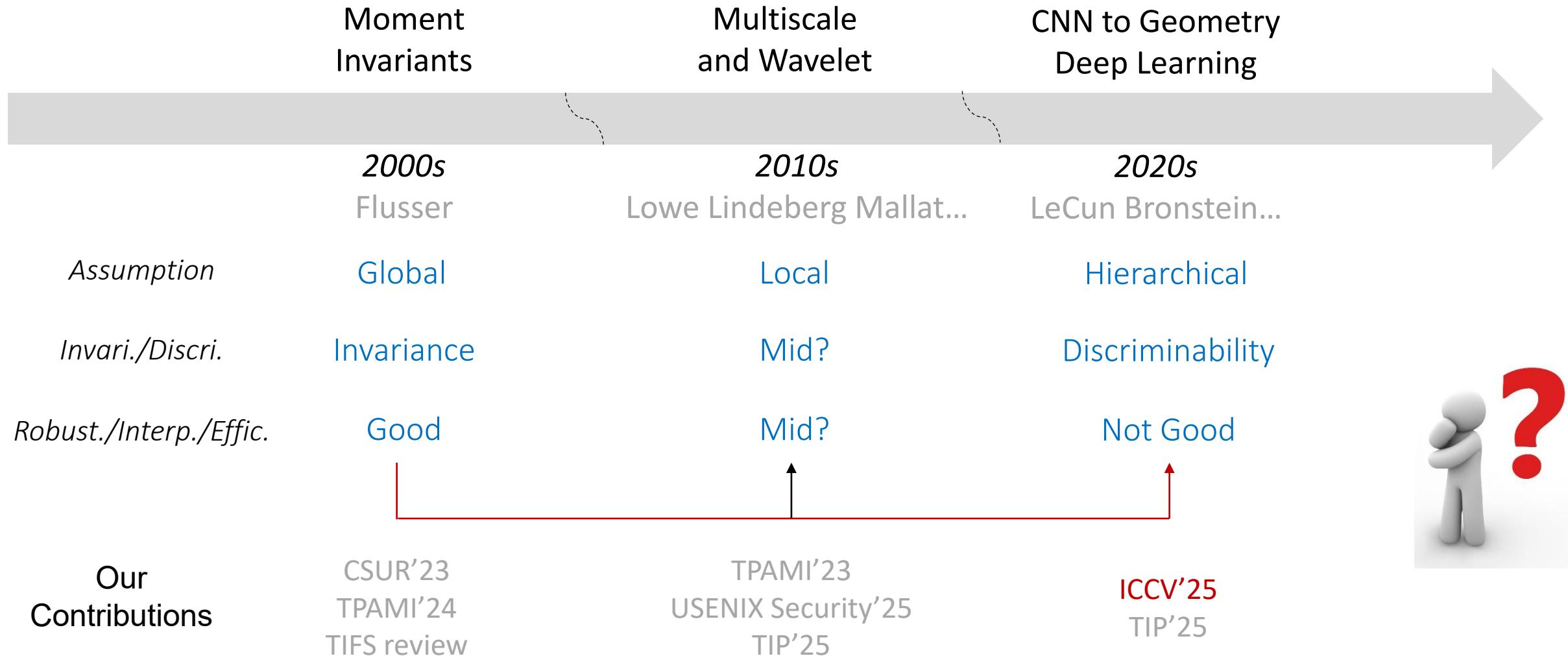


$$\mathbf{h}_u = \phi \left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{V}} a(\mathbf{x}_u, \mathbf{x}_v) \psi(\mathbf{x}_v) \right)$$



Transformers are GNNs on Fully-connected Graph

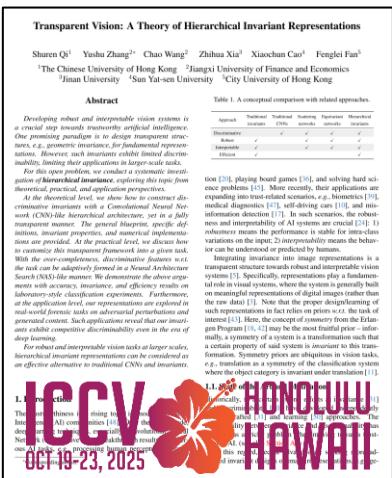
Our Contributions



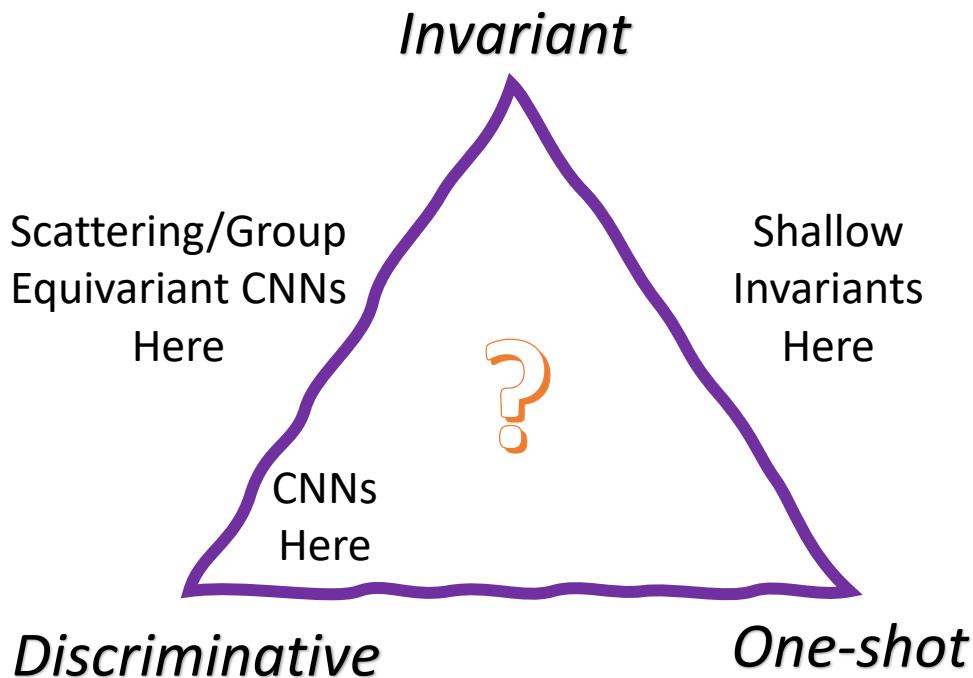
From Global And Local To Hierarchical

Exploring Hierarchical Invariants

- Reviewing the above hierarchical invariants, one can note **a gap**: equivariant CNNs are discriminative and invariant, but are implemented by sampling of symmetry, with limited efficiency and invariance, especially for joint invariance.
- We tried to define **hierarchical (discriminative) invariants while being one-shot**. We achieved this goal by exploring the potential of **classical moment invariants**.

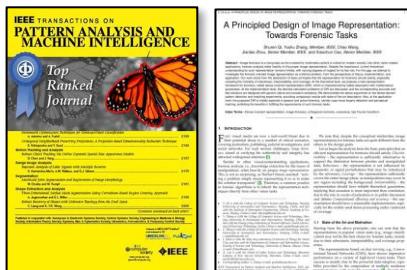
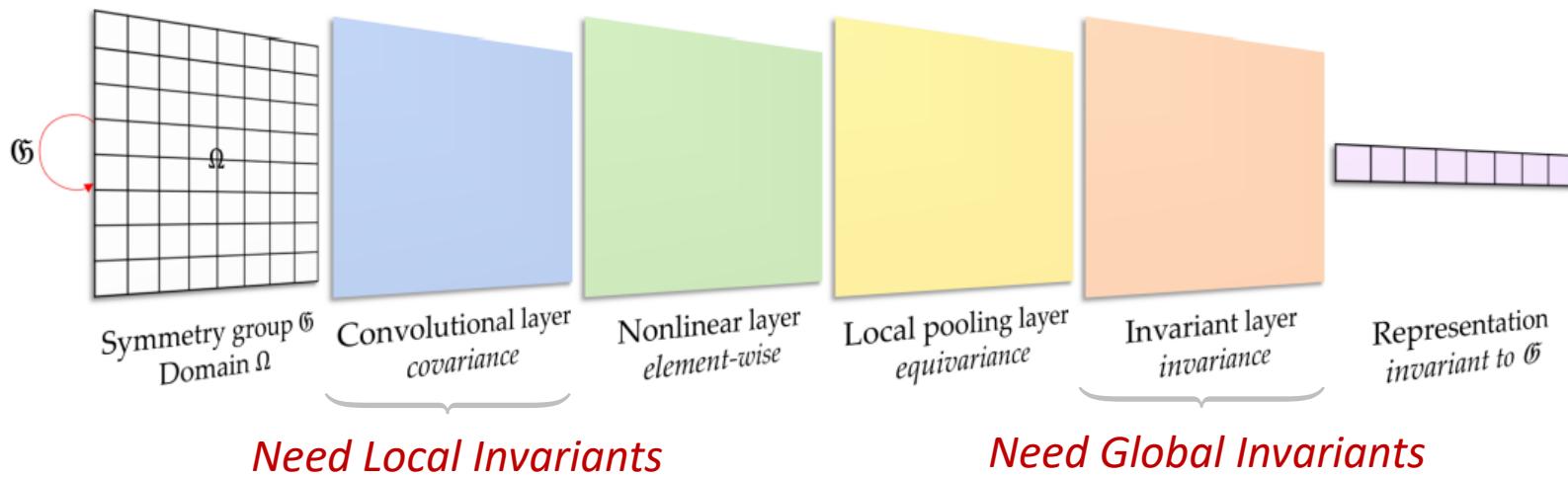


- S. Qi, Y. Zhang, C. Wang, et al. Transparent Vision: A Theory of Hierarchical Invariant Representations. *ICCV*, 2025.



Blueprint

- First, we rethink the typical modules of CNN, unifying the theory of global and local invariants into a hierarchical network.



Recalling the geometric deep learning blueprint, we are surprised that we already have the components to form the hierarchical invariance, we just have not yet assembled them.

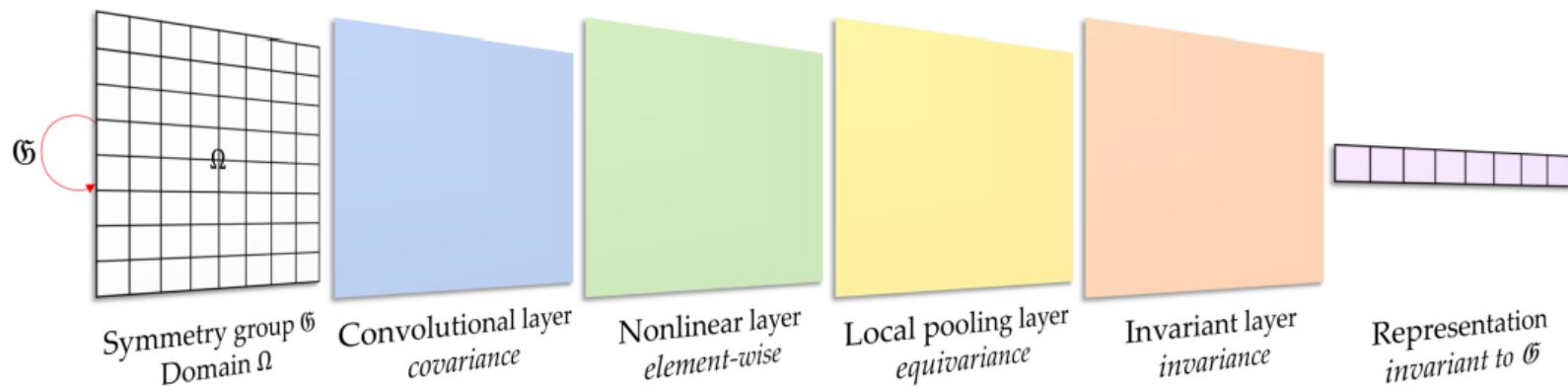
- S. Qi, Y. Zhang, C. Wang, et al. A Principled Design of Image Representation: Towards Forensic Tasks. *TPAMI*, 2023.



- J. Flusser, B. Zitova, T. Suk. *Moments and Moment Invariants in Pattern Recognition*. John Wiley & Sons, 2009.

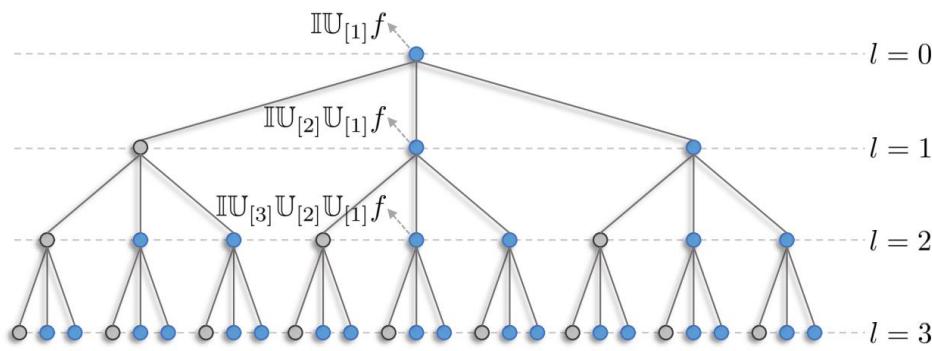
Definition

- Then, we can define new modules with their cascades to fulfill the blueprint:
 - Ω is 2D grid for images; \mathfrak{G} is a translation, rotation, flipping, and scaling symmetry group over Ω .
 - **\mathfrak{G} -covariant convolutional layer:** $\mathbb{C}M \triangleq \langle M, V_{nm}^{uvw} \rangle = M(i, j; k) \otimes (H_{nm}^w(i, j))^T$
 - **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:** $\mathbb{I}M = \mathcal{I}(\{\langle M(i, j; k), V_{nm}(x_i, y_j) \rangle\})$
 - **\mathfrak{G} -invariant representation:** $\mathcal{R}_p \triangleq \mathbb{I} \circ \mathbb{P}_{[L]} \circ \mathbb{S}_{[L]} \circ \mathbb{C}_{[L]} \circ \dots \circ \mathbb{P}_{[1]} \circ \mathbb{S}_{[1]} \circ \mathbb{C}_{[1]}$

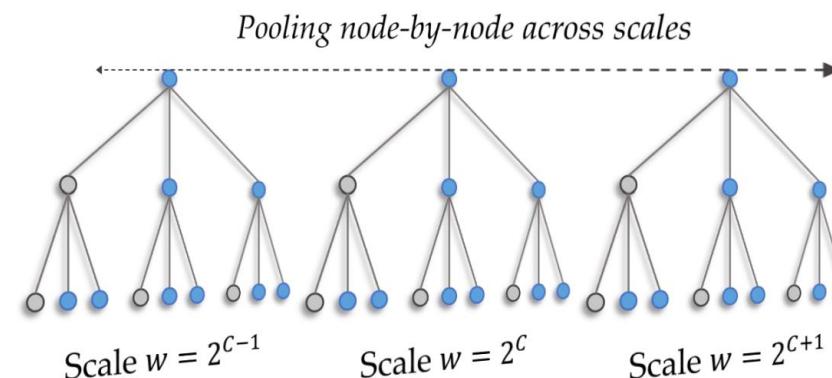


Property

- The group theory shows the one-shot symmetry property at each inter layer:
 - \mathfrak{G}_1 is the translation, rotation, and flipping symmetry group; \mathfrak{G}_2 is a 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. A representation unit denoted as $\mathbb{U} \triangleq \mathbb{P} \circ \mathbb{S} \circ \mathbb{C}$.
 - \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:** $\mathbb{U}_{[L]}^w \circ \cdots \circ \mathbb{U}_{[2]}^w \circ \mathbb{U}_{[1]}^w(\mathfrak{g}_2 M) \equiv \mathfrak{g}'_2 \mathbb{U}_{[L]}^w \circ \cdots \circ \mathbb{U}_{[2]}^w \circ \mathbb{U}_{[1]}^w(M) \quad \mathfrak{g}'_2 \mathbb{U}^w \triangleq \mathfrak{g}_2 \mathbb{U}^{ws}$
 - \mathfrak{G}_0 Hierarchical Invariance:** $\mathbb{I}(\mathfrak{g}'_0 M)_{[L]} \equiv \mathbb{I}M_{[L]}$



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

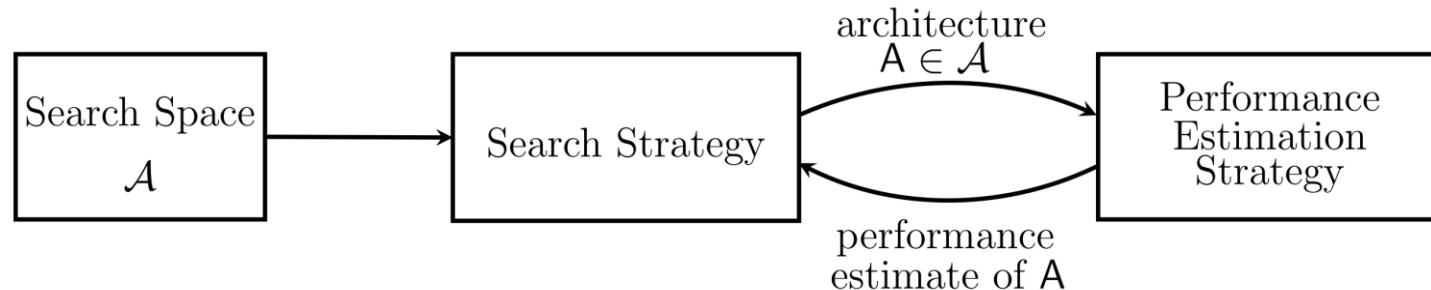


A Multi-scale Practice with \mathfrak{G}_0 Invariance

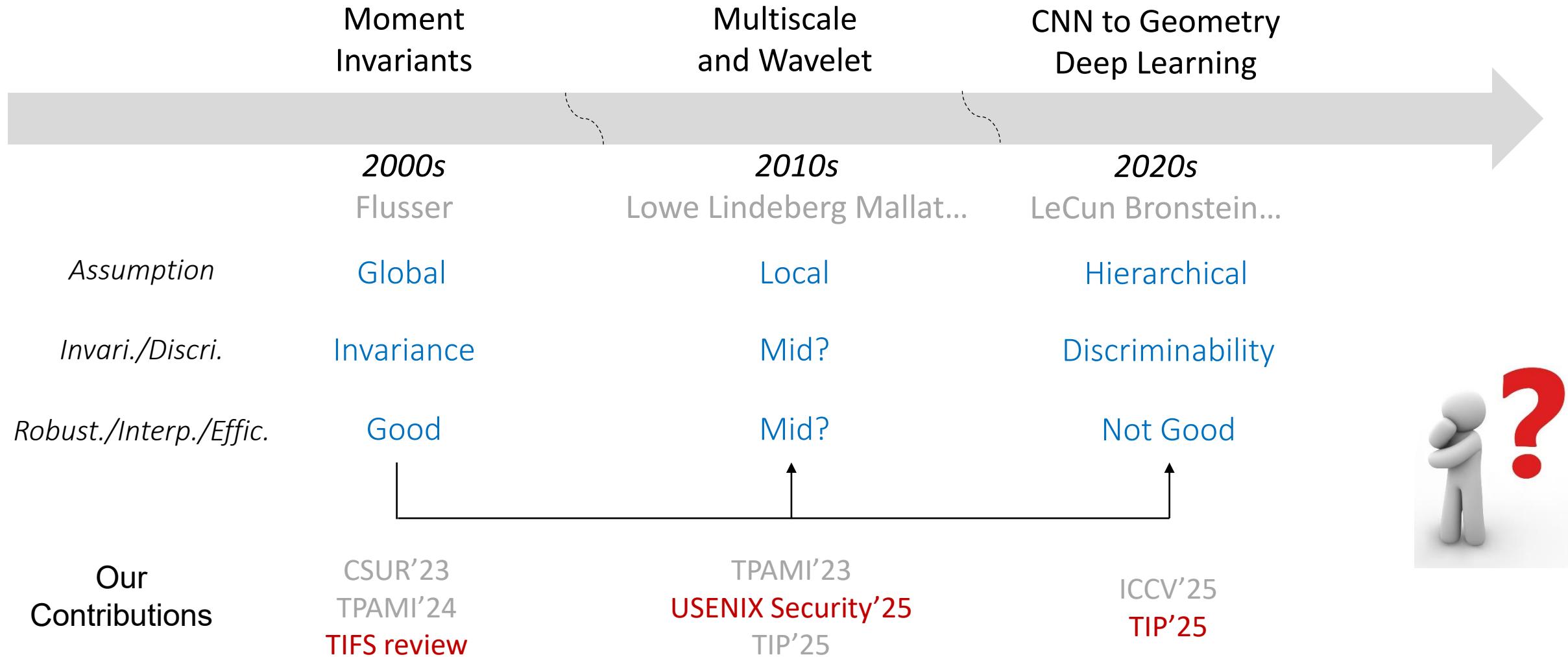
Practice

How to select task-discriminative features from such a huge feature space?

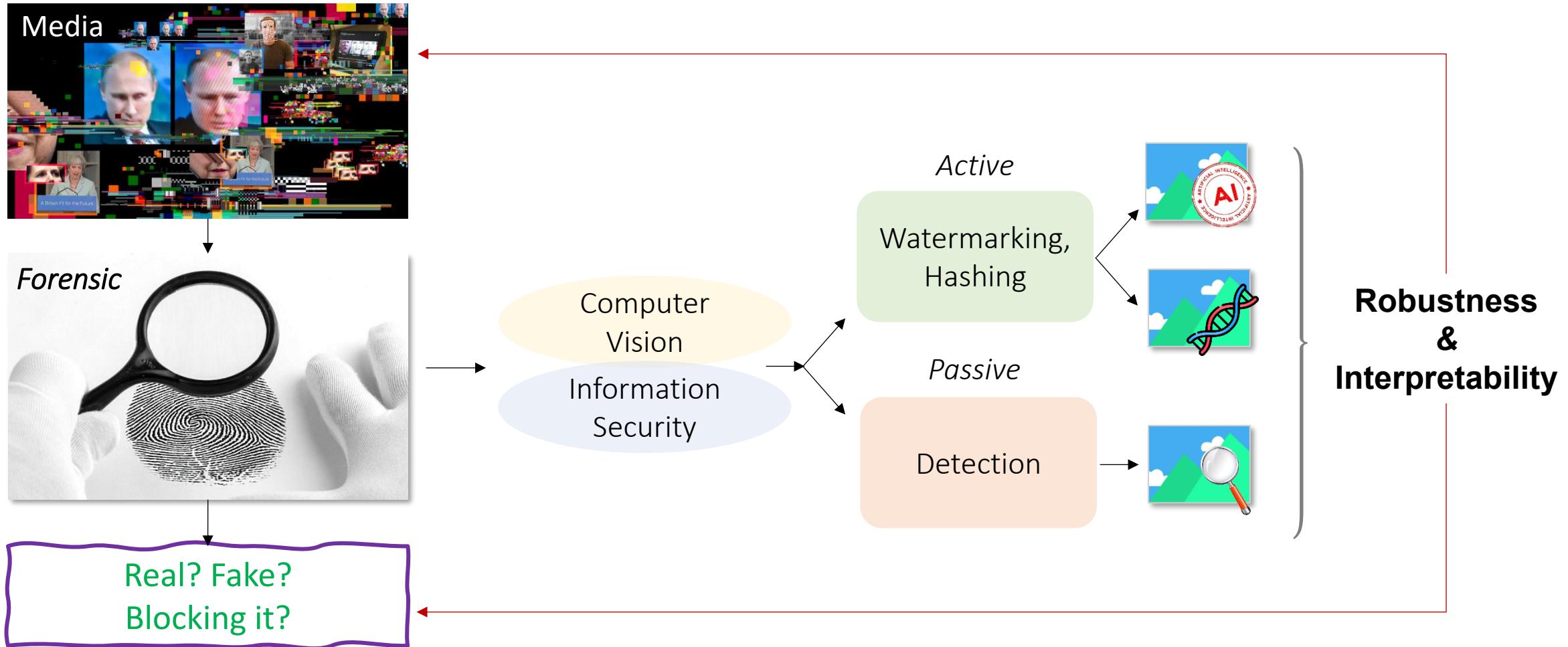
- Feature/Architecture Selection, inspired by Neural Architecture Search (NAS)
 - **Super network** covers preferred and sufficiently diverse parameters.
 - **Correlation analysis** for filtering the most relevant features.
 - **Concise network** by resampling the super network for most relevant features.
- Cascading Learning Module, inspired by Hybrid Representation Learning (HRL)
 - Replacing shallow layers of learning CNNs with our layers, such that discriminative features are formed in a space with geometric symmetries.



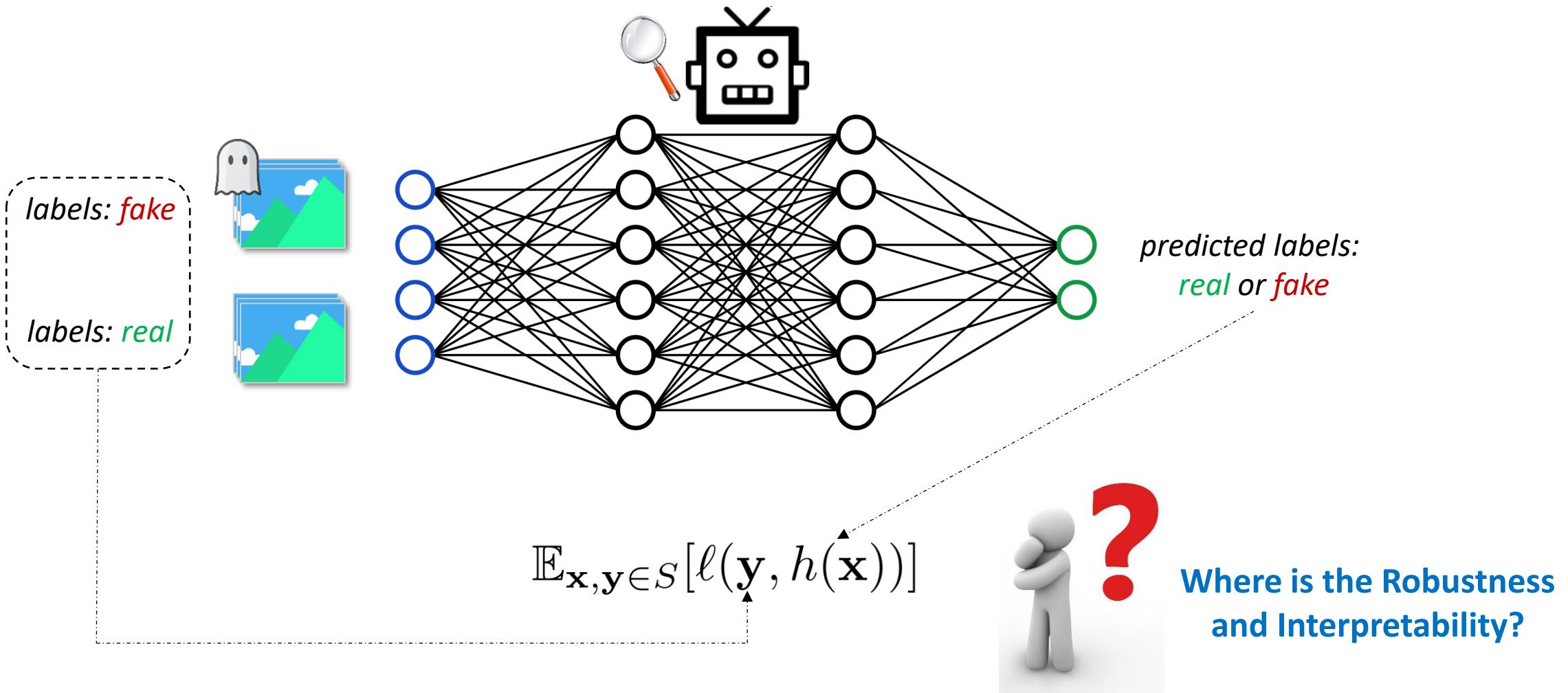
Our Contributions



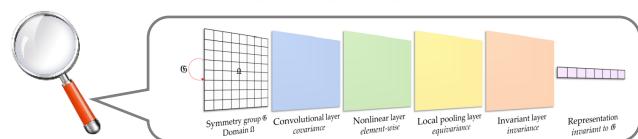
Forensic, Fighting Against AIGC Abuse



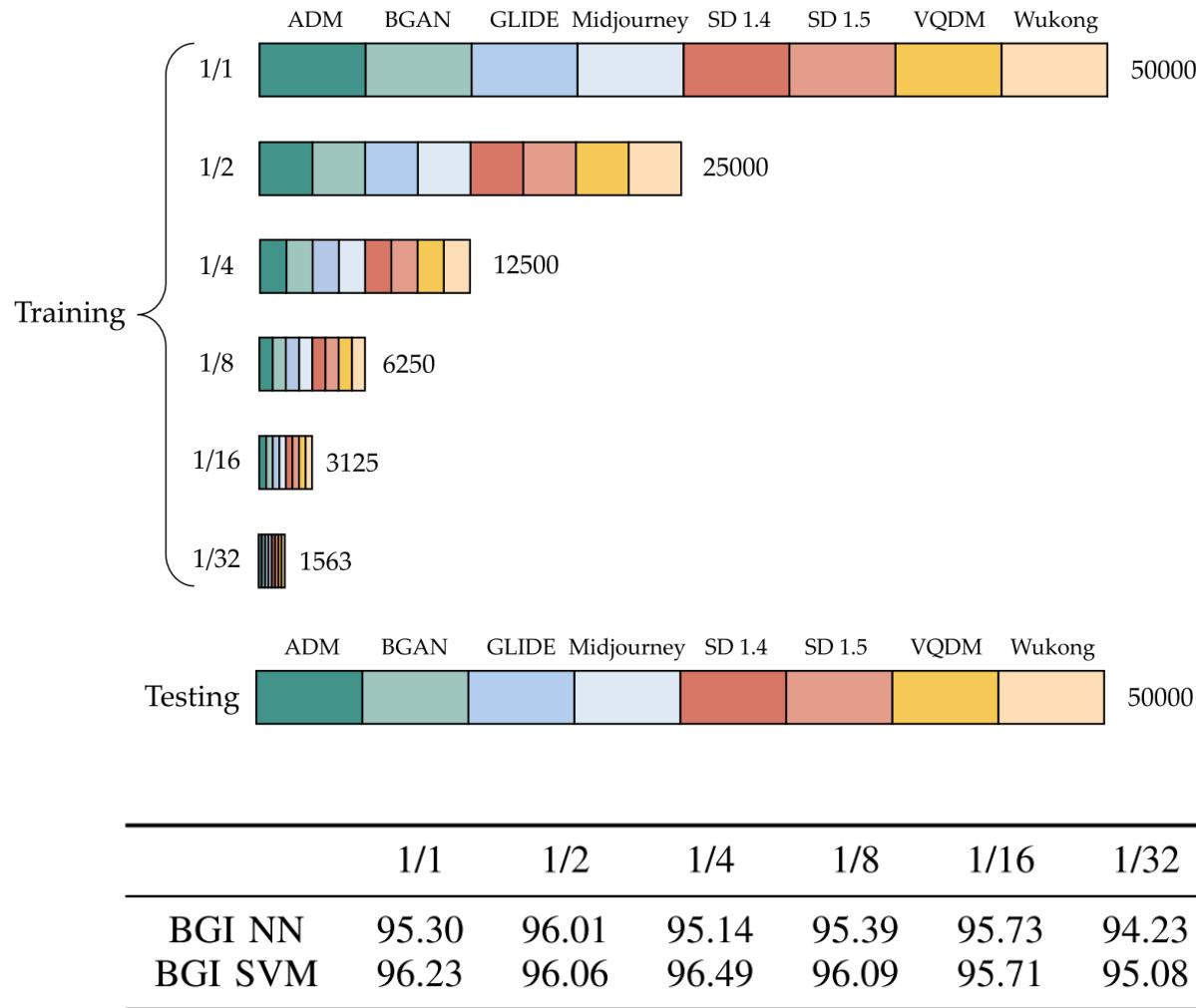
AIGC Detection: Motivations



AIGC Detection: Ideas

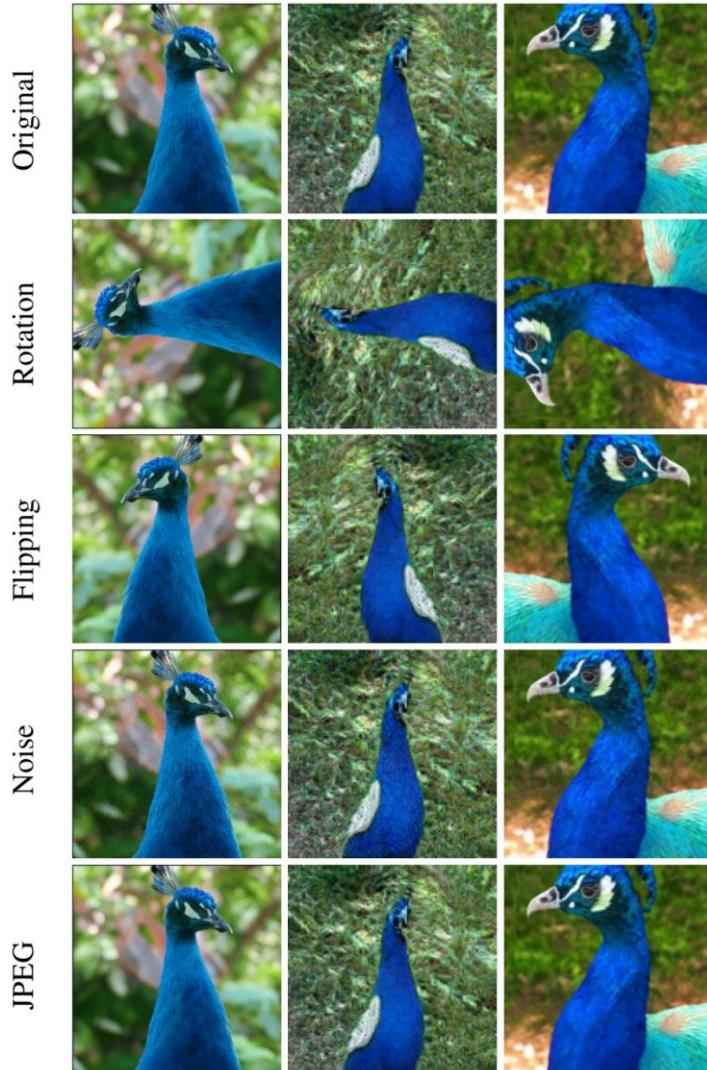


AIGC Detection: Training and Sample Efficiency



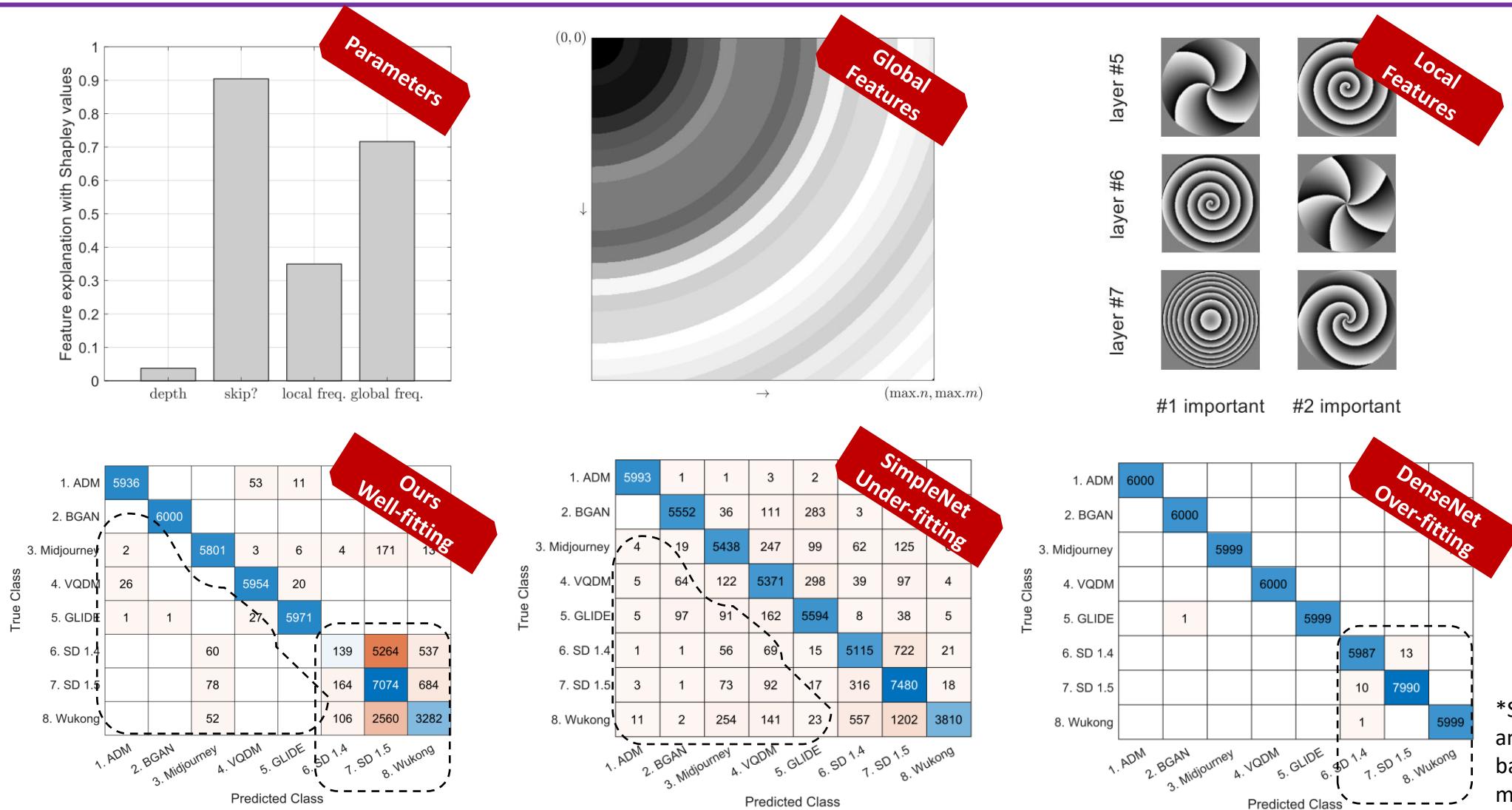
	1/8		
	Precision	Recall	F1
<i>Handcrafted</i>			
DCT NN	35.82	40.41	37.98
DCT SVM	95.00	94.98	94.99
DWT NN	49.98	99.72	66.59
DWT SVM	96.75	93.61	95.15
ScatterNet NN	86.12	77.46	81.56
ScatterNet SVM	86.75	91.67	89.14
<i>Learning</i>			
SimpleNet	75.76	36.49	49.25
ResNet	82.11	86.66	84.32
DenseNet	89.67	90.34	90.01
InceptionNet	83.39	79.15	81.22
<i>Forensic</i>			
CNN Spot	66.57	82.93	73.74
F3Net	80.43	79.54	79.67
<i>Ours</i>			
BGI NN	94.67	96.40	95.39
BGI SVM	94.59	97.73	96.09

AIGC Detection: Geometric and Signal Robustness

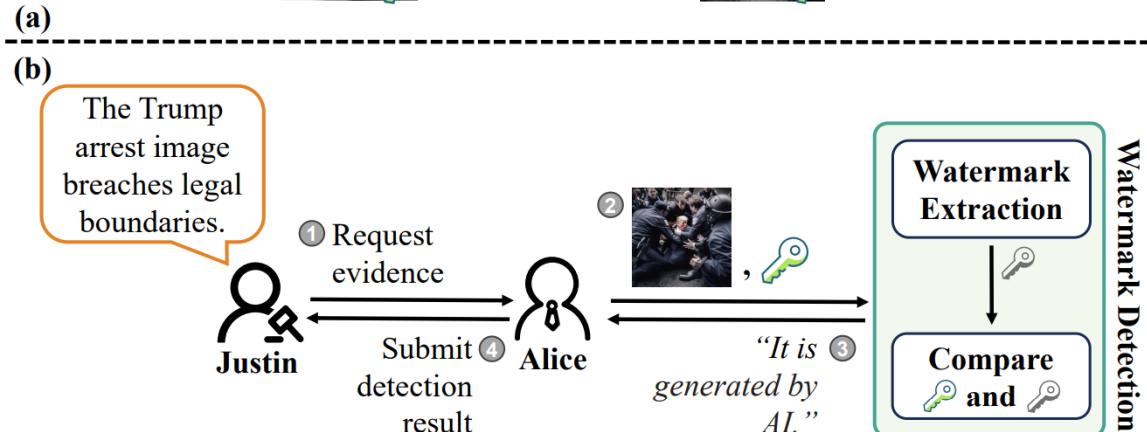
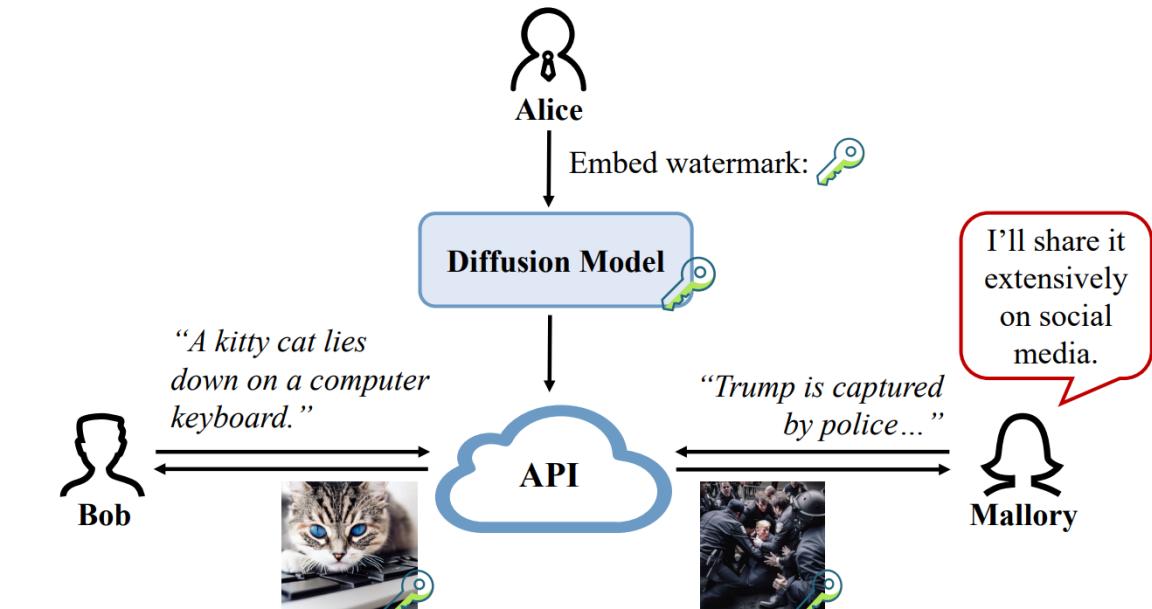


		Geometric Degradation			Signal Degradation		
		Precision	Recall	F1	Precision	Recall	F1
<i>Handcrafted</i>							
	DCT NN	0	0	0	0	0	0
	DCT SVM	80.86	95.03	87.38	78.50	91.35	84.44
	DWT NN	50.62	25.59	33.99	52.40	26.11	34.86
	DWT SVM	79.70	94.75	86.58	81.47	96.65	88.41
	ScatterNet NN	69.34	88.58	77.79	67.97	95.97	79.58
	ScatterNet SVM	90.67	80.23	85.13	92.37	90.38	91.36
<i>Learning</i>							
	SimpleNet	65.03	85.90	74.02	66.13	92.61	77.16
	ResNet	91.70	83.85	87.60	94.54	89.56	91.98
	DenseNet	96.02	89.92	92.87	98.78	90.01	94.19
	InceptionNet	92.00	92.24	92.12	96.77	84.06	89.97
<i>Forensic</i>							
	CNN Spot	83.12	80.64	81.51	68.35	59.32	63.14
	F3Net	79.83	77.57	77.96	80.96	74.83	77.13
<i>Ours</i>							
	BGI NN	96.84	92.01	94.36	90.03	95.10	92.50
	BGI SVM	96.45	93.40	94.90	92.52	95.84	94.15

AIGC Detection: Visualization and Interpretability

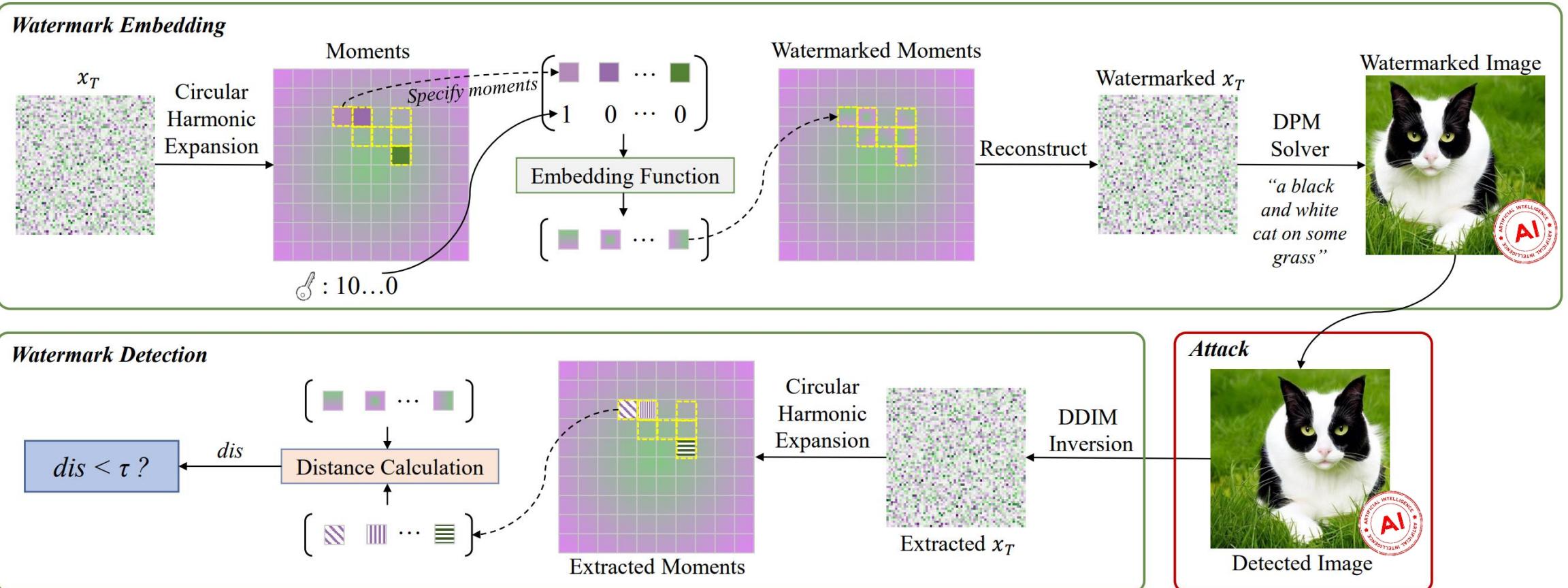


AIGC Watermarking: Motivations



Is there a balance
between robustness and
imperceptibility?

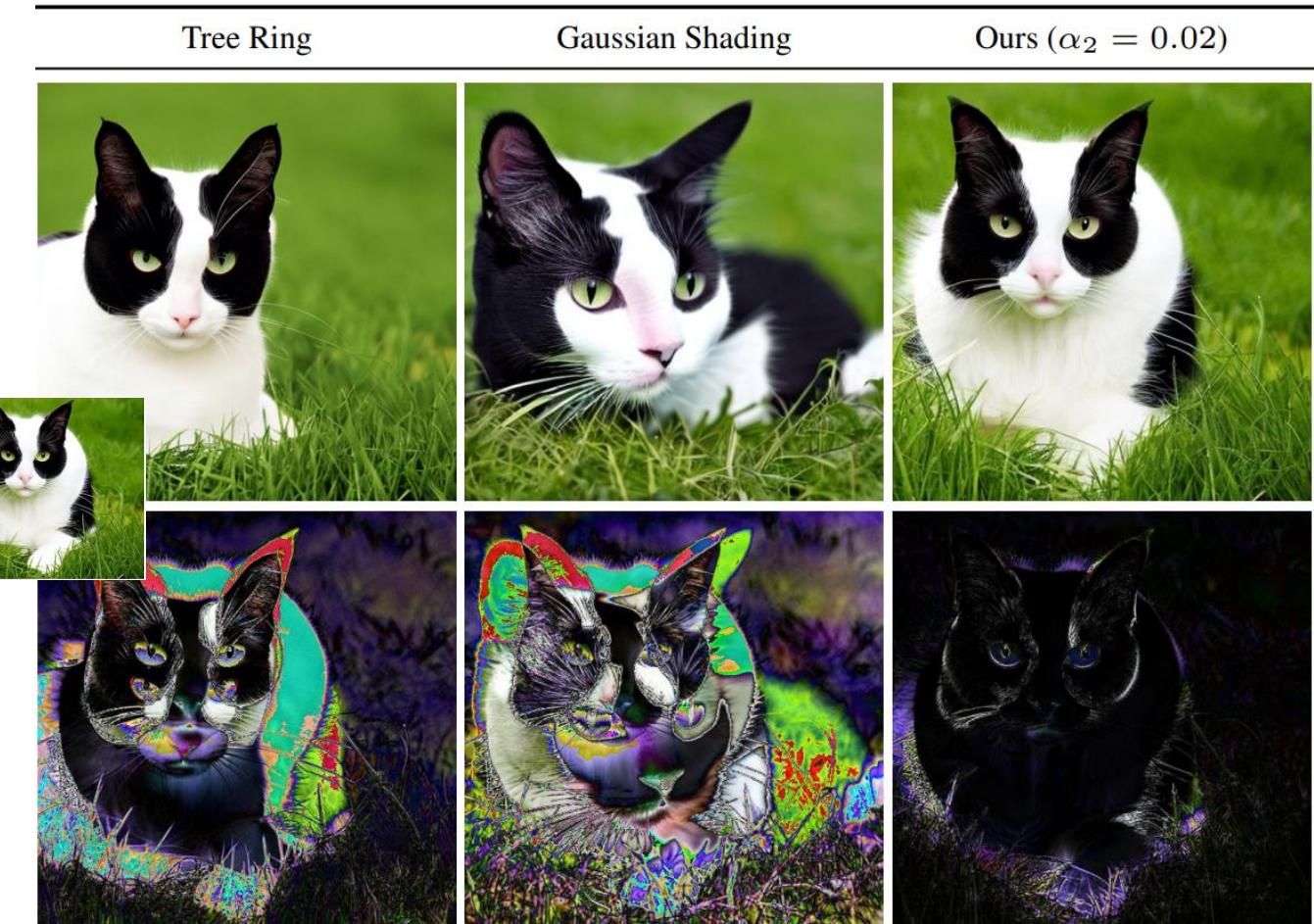
AIGC Watermarking: Ideas



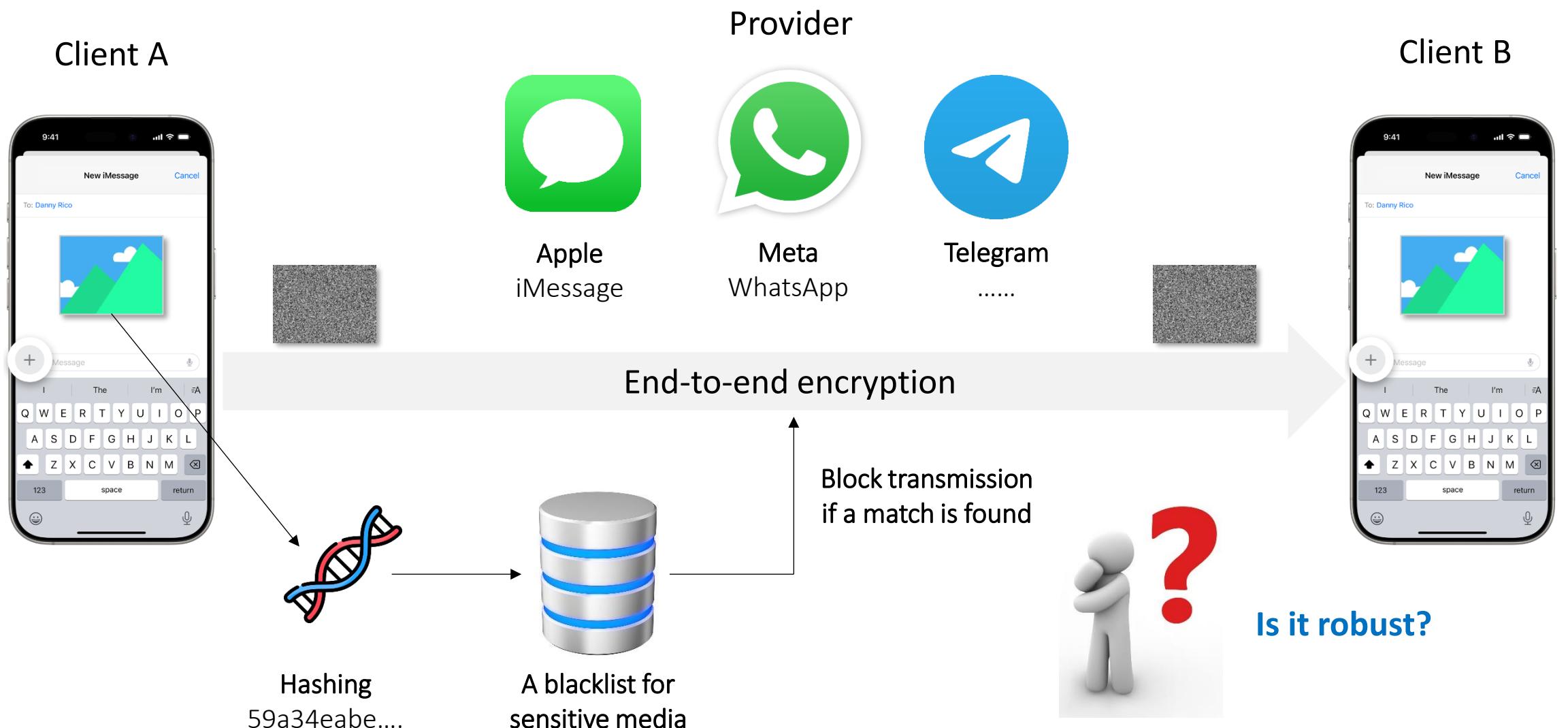
AIGC Watermarking: Robustness and Imperceptibility

Method	VAE based		DM based	
	Bmshj'18	Cheng'20	SDv2.1	Average
<i>Pixel-level</i>				
DwtDct	0.005	0.002	0.003	0.003
DwtDctSvd	0.103	0.124	0.230	0.152
RivaGAN	0.014	0.017	0.123	0.051
Stable Signature	0.541	0.813	0.003	0.452
<i>Content-level</i>				
Tree Ring	0.976	0.993	0.943	0.971
Gaussian Shading	1.000	1.000	1.000	1.000
Ours	0.990	0.983	1.000	0.991

Method	Metrics		
	SSIM↑	LPIPS↓	WO-FID↓
Tree Ring	0.47	0.50	43.81
Gaussian Shading	0.20	0.74	48.32
Ours ($\alpha_2 = 0.02$)	0.75	0.20	26.50
Ours ($\alpha_2 = 0.04$)	0.62*	0.31*	35.02*



AIGC Hashing: Motivations



AIGC Hashing: Ideas

Definition 1. (Multiresolution perturbation). The addition of multiresolution perturbation is defined as follows:

$$X'_{(x,y) \in D_{uvw}} = \mathcal{F}^{-1}(\mathcal{F}(X) + \delta), \quad (3)$$

with notations of

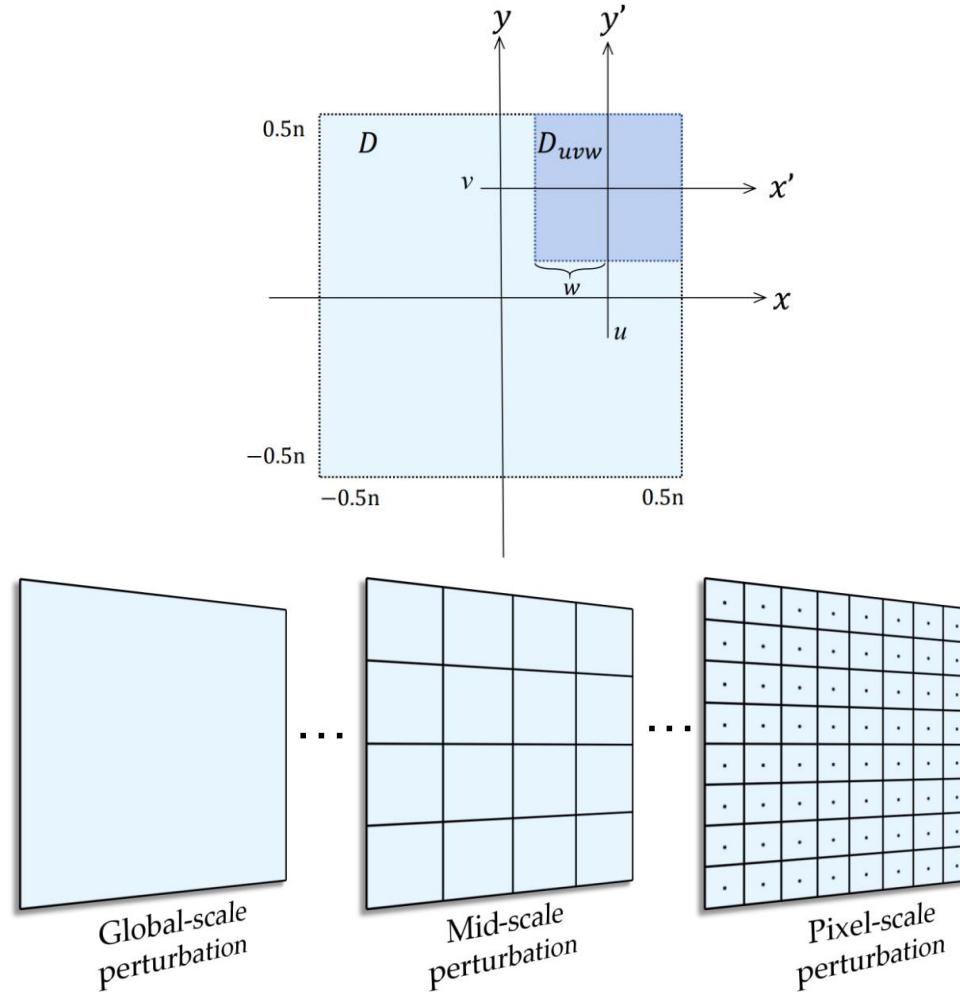
$$\mathcal{F}(X) = \langle X, V_{nm}^{uvw} \rangle = \iint_D (V_{nm}^{uvw}(x,y))^* X(x,y) dx dy, \quad (4)$$

and

$$\mathcal{F}^{-1}(\mathcal{F}(X)) = \sum_{n,m} V_{nm}^{uvw}(x,y) \mathcal{F}(X), \quad (5)$$

where \mathcal{F} denotes the local orthogonal transformation [39], with image $X(x,y)$ on domain $(x,y) \in D$. The local orthogonal basis function V_{nm}^{uvw} is defined on the domain D_{uvw} with the order parameters $(n,m) \in \mathbb{Z}^2$, converting D to D_{uvw} by the translation offset (u,v) and the scaling factor w , as illustrated in Figure 2. Note that the local orthogonal basis function V_{nm}^{uvw} can be defined from any global orthogonal basis function V_{nm} , e.g., a family of harmonic functions, with following form:

$$V_{nm}^{uvw}(x,y) = V_{nm}(x',y') = V_{nm}\left(\frac{x-u}{w}, \frac{y-v}{w}\right). \quad (6)$$



AIGC Hashing: Uniform, Fast, and Successful Attacks

