



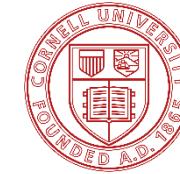
THE WEB
CONFERENCE ACM



Center
for
Mathematical
Artificial
Intelligence
CMAI



DEPARTMENT OF
MATHEMATICS
THE CHINESE UNIVERSITY OF HONG KONG



Rethink Deep Learning with Invariance in Data Representation

A Tutorial at The Web Conference 2025 in Sydney (WWW 2025)

Shuren Qi¹, Fei Wang², Tieyong Zeng¹, and Fenglei Fan³

¹ CMAI, IMS, Department of Mathematics, The Chinese University of Hong Kong, HK

²Weill Cornell Medicine, Department of Population Health Sciences, Cornell University, USA

³Department of Data Science, City University of Hong Kong, HK

13:30 - 15:00, Tuesday, April 29, 2025
Room C3.4, ICC Sydney, Australia

Organizer and Presenter



Shuren Qi

Postdoc, CUHK

shurenqi@cuhk.edu.hk few2001@med.cornell.edu zeng@math.cuhk.edu.hk fenglfan@cityu.edu.hk

Presenter



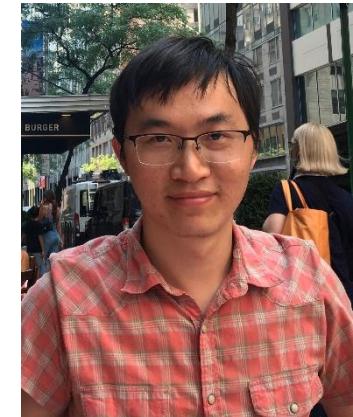
Fei Wang

Professor, Cornell



Tieyong Zeng

Professor, CUHK



Fenglei Fan

AP, CityU

Organizer

Main Organizer

Tutorial Homepage

Tutorial proposal, slides, reading list, video, and more materials available at
<https://shurenqi.github.io/wwwtutorial/>

Rethink Deep Learning with Invariance in Data Representation

Shuren Qi
shurenqi@uhk.edu.hk
The Chinese University of Hong Kong
Hong Kong, China

Fei Wang
few2001@med.cornell.edu
Cornell University
New York, United States

Tieyong Zeng
zeng@math.cuhk.edu.hk
The Chinese University of Hong Kong
Hong Kong, China

Fenglei Fan
ffan@math.cuhk.edu.hk
The Chinese University of Hong Kong
Hong Kong, China

Abstract
Integrating invariance into data representations is a principled design in intelligent systems and web applications. Representations play a fundamental role, where systems and applications are both built on meaningful representations of digital inputs (rather than the raw data). In fact, the proper design/learning of such representations relies on priors w.r.t. the task of interest. Here, the concept of symmetry plays a key role. In general, symmetry is a mathematical property; informally, a symmetry of a system is a transformation that leaves a certain property of the system invariant. Symmetry priors are ubiquitous, e.g., translation as a symmetry of the object classification, where object categories are invariant under translation. The quest for invariance has been a long history of the development of various representations in the era before deep learning, such as the SIFT. As we enter the early era of deep learning, the invariance principle is largely ignored and replaced by a data-driven paradigm, which is the CNN. However, this neglect did not last long before they faced serious setbacks regarding robustness, interpretability, efficiency, and so on. The invariance principle has returned in the era of rethinking deep learning, forming a new field known as Geometric Deep Learning (GDL).

In this tutorial, we will give a historical perspective of the invariance in data representations. More importantly, we will identify those research dilemmas, promising works, future directions, and web applications.

CCS Concepts
• Theory of computation → Theory and algorithms for application domains • Computing methodologies → Artificial intelligence

Keywords
Pattern Recognition, Data Mining, Invariance, Symmetry, Representation, Tutorial

This work is licensed under a Creative Commons Attribution 4.0 International License
WWW Companion '25, Sydney, NSW, Australia
© 2025 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8716-9/25/04
https://doi.org/10.1145/3570176.3713861

Towards Robust, Interpretable, and Efficient AI

Deep learning representations v.s. robustness, interpretability, and efficiency principles.

Symmetry and Invariance Priors

E. Noether, 1918: Bräuer-Polymer E. Noether's theorem H. Weyl, 1929: The Basis of Symmetry C. N. Yang & R. L. Mills, 1954: Yang-Mills Theory

Lei G., Deng J., Y. & Huang G., 2015: The Selectivity-Invariance Dilemma: "Representations that are selective to the features that are important for discrimination must also be invariant to represent objects"

Symmetry and Invariance are Ubiquitous

Image Classification: permutation, Remote Sensing: orientation, Self-driving Car: motion blurring, Speech Command: time warping, Point Cloud Analysis: noise, Prediction of Molecular Properties: permutation

Representations Equipped with Symmetry and Invariance

Geometric Deep Learning

Permutations: function regularity, CNN: translation, Group CNNs: translation-invariance, LSTM: time warping, Deformers: permutations, GNNs: permutations, Intrinsic CNNs: isometry-gauge invariance

Outline of the Tutorial

- Introduction (20 min)
- Preliminaries of invariance (20 min)
- Invariance in the era before deep learning (40 min)
- Invariance in the early era of deep learning (40 min)
- Invariance in the era of rethinking deep learning (40 min)
- Conclusions and discussions (20 min)

The good The ugly The bad

A Historical Perspective of Data Representation
Rethinking Deep Learning with Invariance

6

THEORY OF ALGEBRAIC INVARIANTS

David Hilbert

Geometric Invariant Theory

M. Mumford, J. Pogorelsky, F. Kawanou
Third Enlarged Edition

Scale-Space Theory in Computer Vision

Springer Science+Business Media, Berlin

Moments and Moment Invariants in Pattern Recognition

Jan Flusser, Tomáš Šuk, Barbora Zitová

EQUIVARIANT AND COORDINATE INDEPENDENT CONVOLUTIONAL NETWORKS

A GAUGE FIELD THEORY OF NEURAL NETWORKS

Maurice Weiler, Patrick Forré, Erik Verlinde, Max Welling

Tutorial Outline

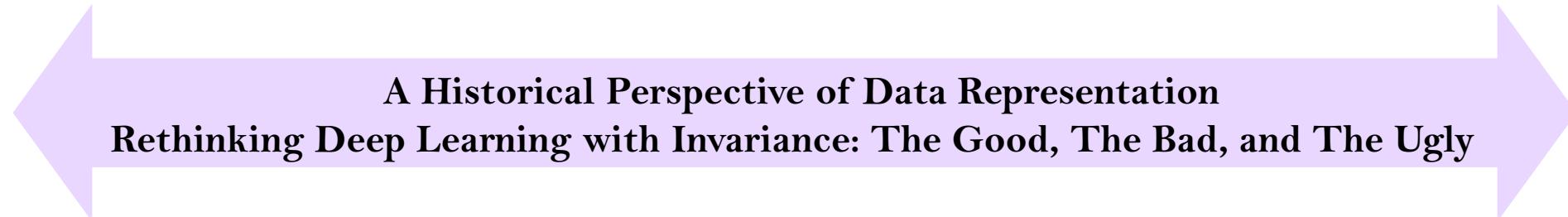
- **Part 1:** Background and challenges (20 min)
- **Part 2:** Preliminaries of invariance (20 min)
- *Q&A / Break (10 min)*
- **Part 3:** Invariance in the era before deep learning (30 min)
- **Part 4:** Invariance in the early era of deep learning (10 min)
- *Q&A / Coffee Break (30 min)*
- **Part 5:**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

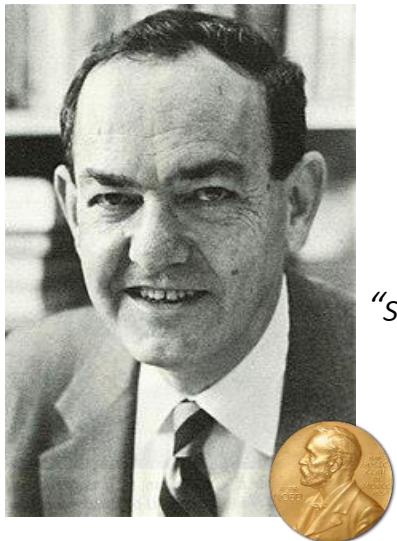
Tutorial Outline

- **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)*



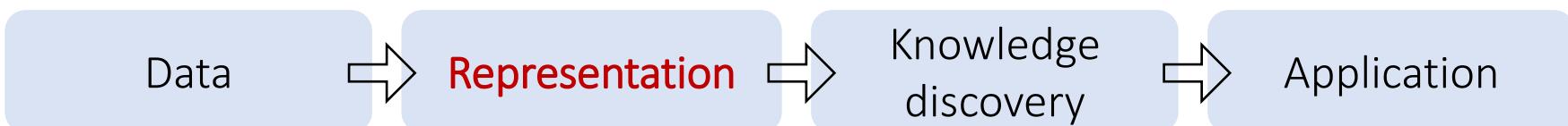
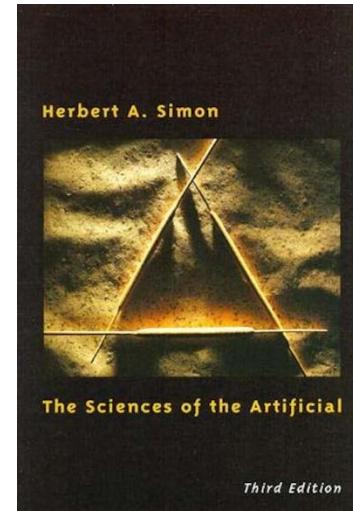
Deep (Representation) Learning,
A Big Bang Moment For AI

Data Representation



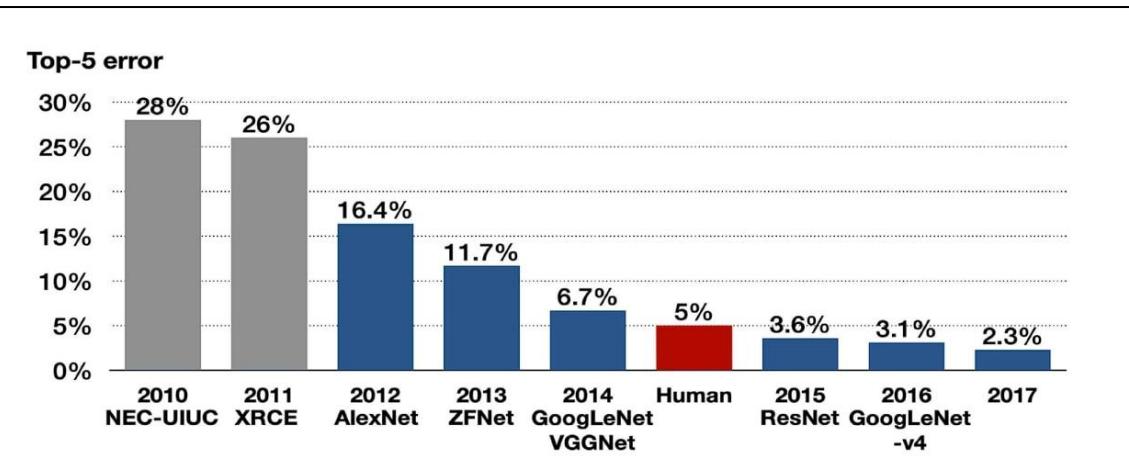
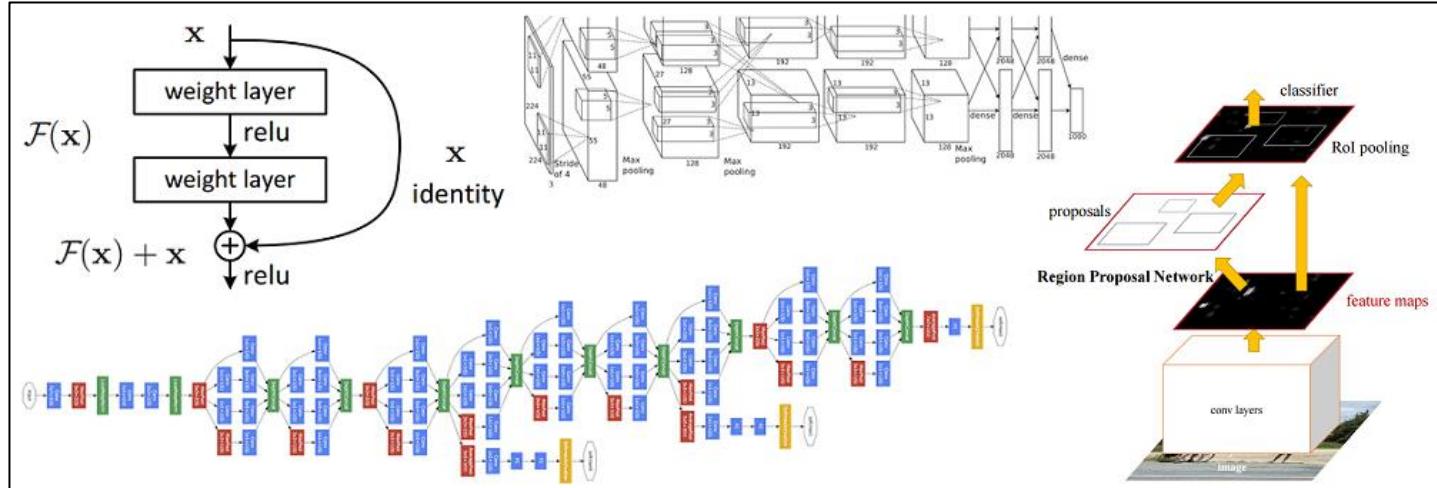
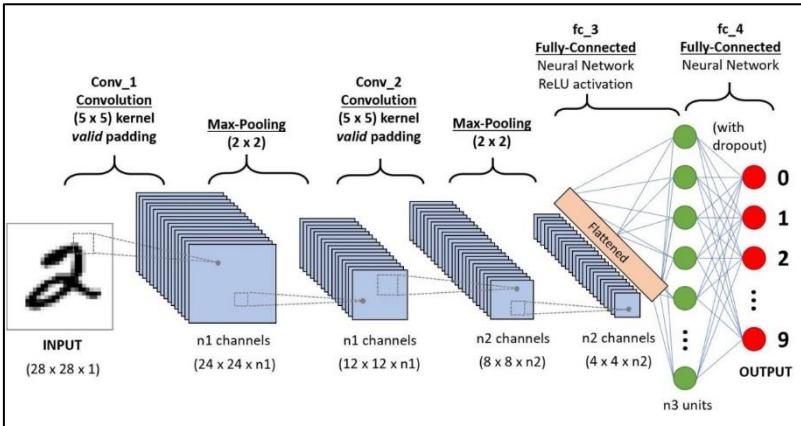
H. Simon, 1969
The Sciences of the Artificial

*“solving a problem simply means representing it
so as to make the solution transparent”*



- H Simon. *The Sciences of the Artificial (Third edition)*. MIT Press, 1996.

Processing Human Perceptual Information

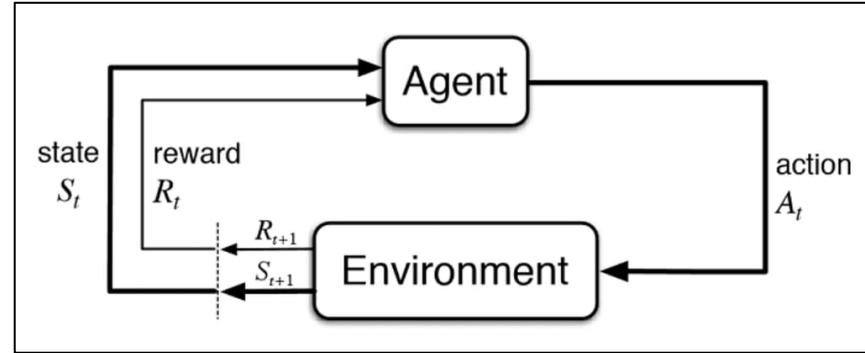
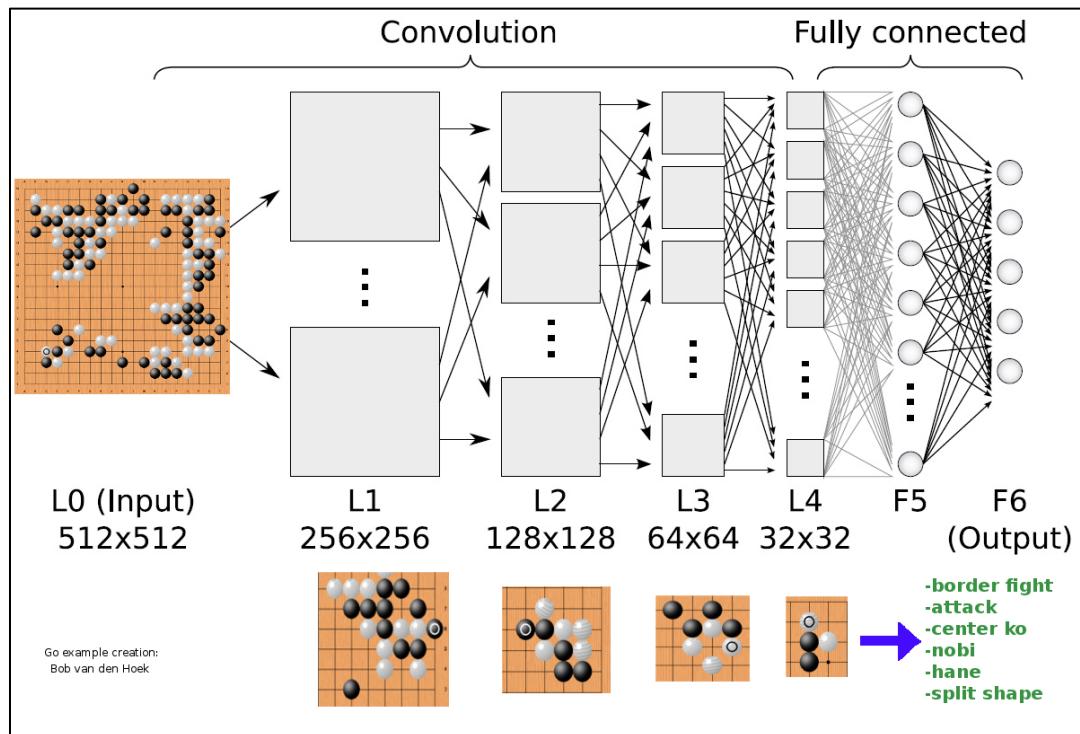


- J Deng, W Dong, R Socher, et al. ImageNet: A large-scale hierarchical image database. CVPR, 2009.

Playing Board Games

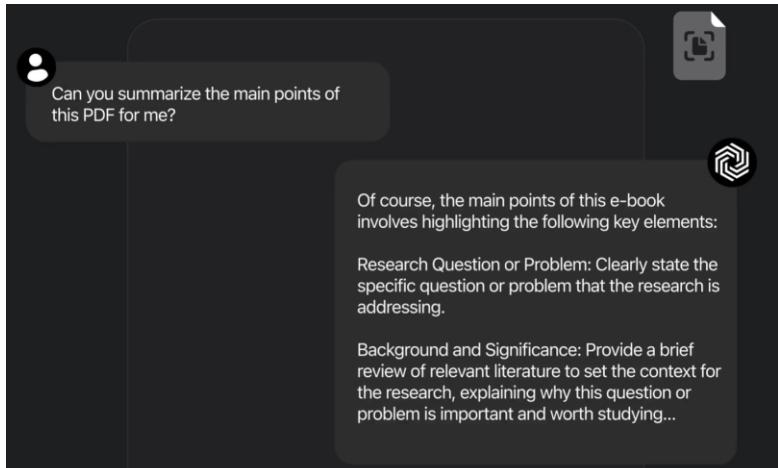
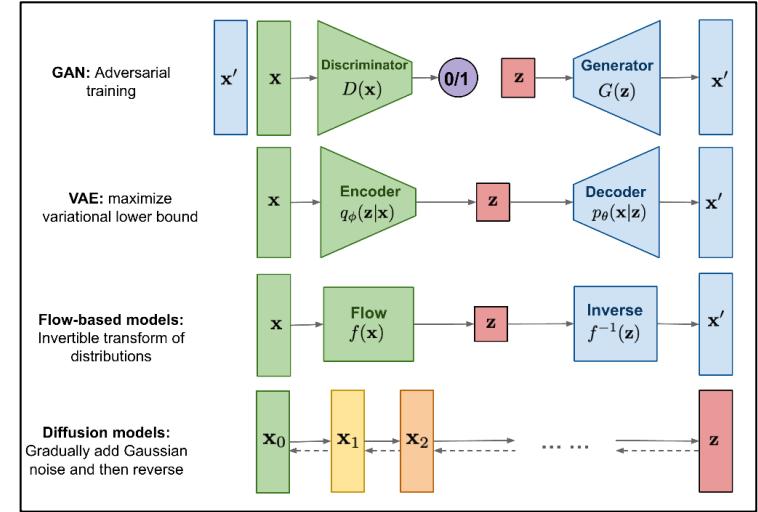
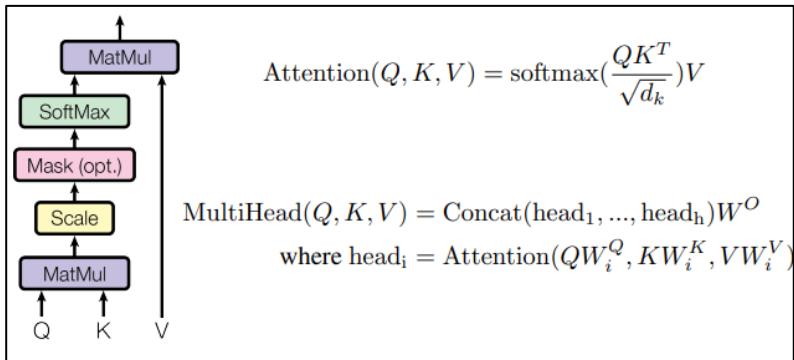


AlphaGo



- D Silver, J Schrittwieser, K Simonyan, et al. Mastering the game of go without human knowledge. *Nature*, 2017.

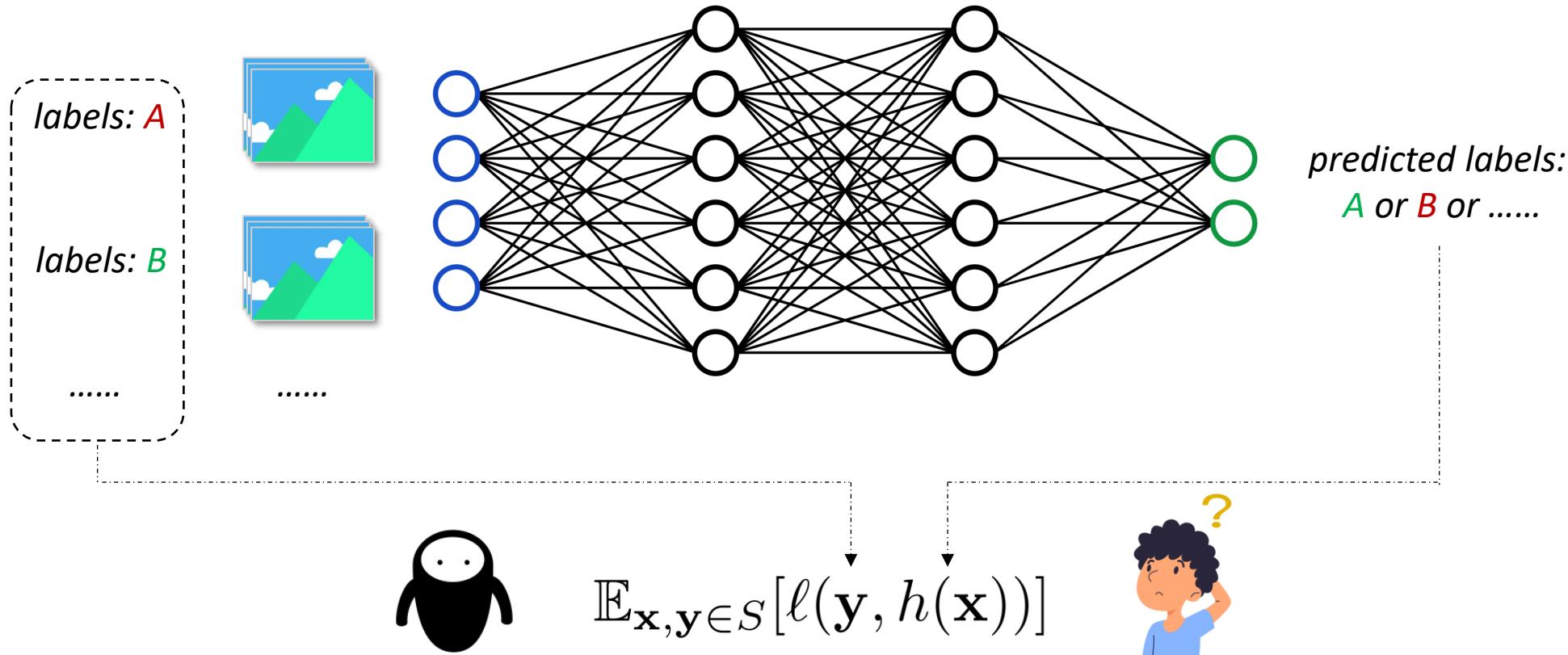
Generating Realistic Media



- Z Epstein, A Hertzmann, L. Herman, et al. Art and the science of generative AI. *Science*, 2023.

Empirical Risk Minimization (ERM),
Behind All These Successes

Empirical Risk Minimization



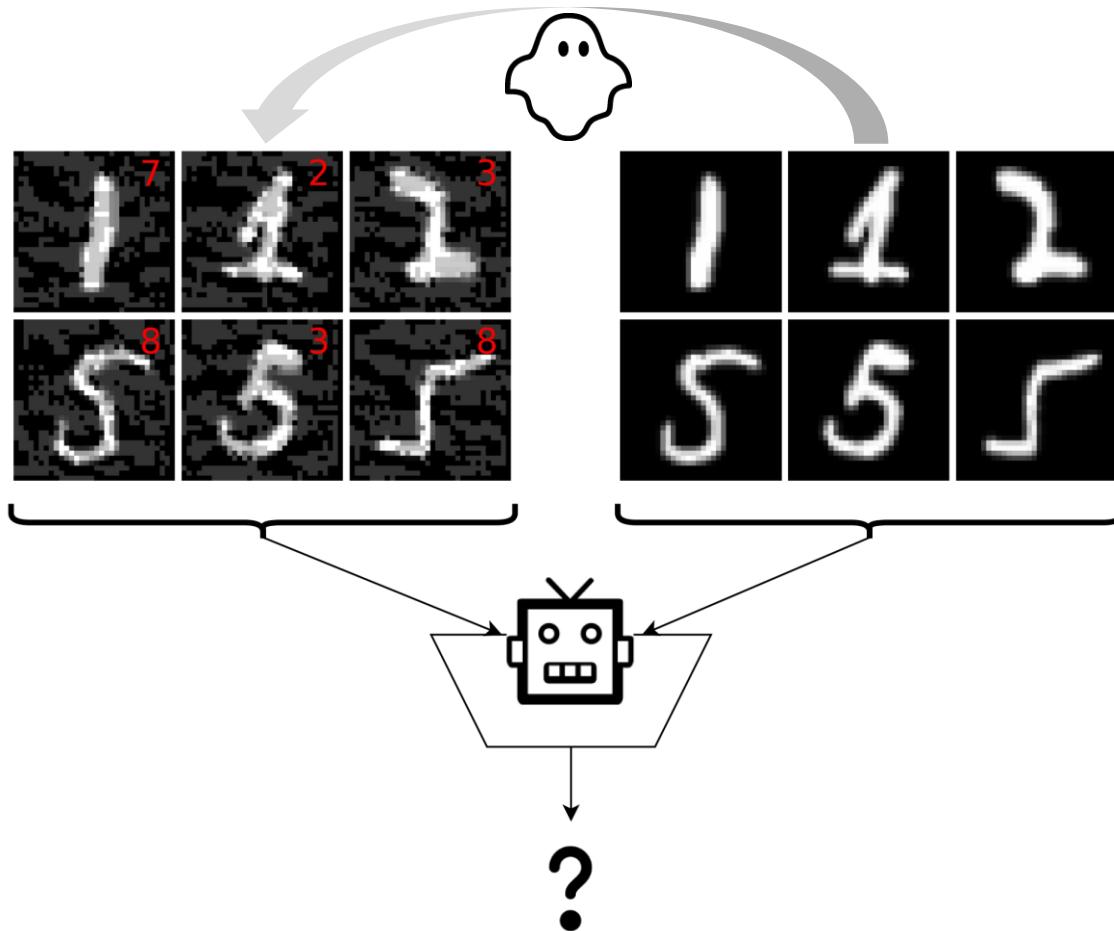
Empirical learning draws the
lines between categories.

But what about **robustness**,
interpretability, and **efficiency**?

- V Vapnik. Principles of risk minimization for learning theory. *NIPS*, 1991.

Robustness of Empirical Learning

- Robustness: the performance of a system is stable for **intra-class variations** on the input.



Universal



One-pixel



Physical



Color



Watermark

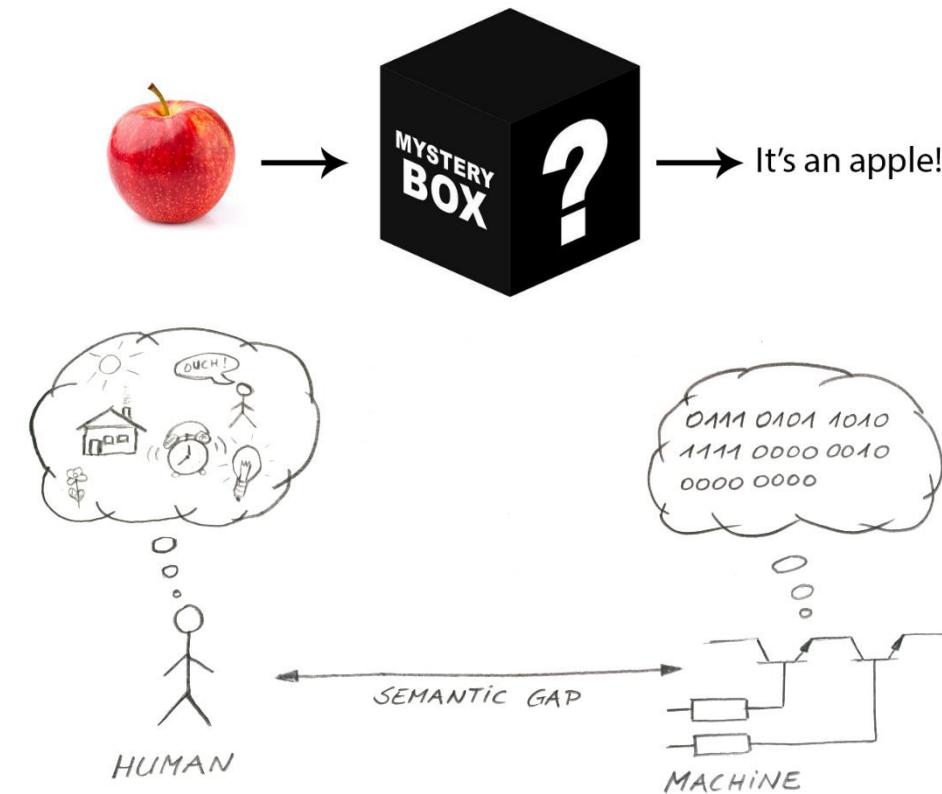
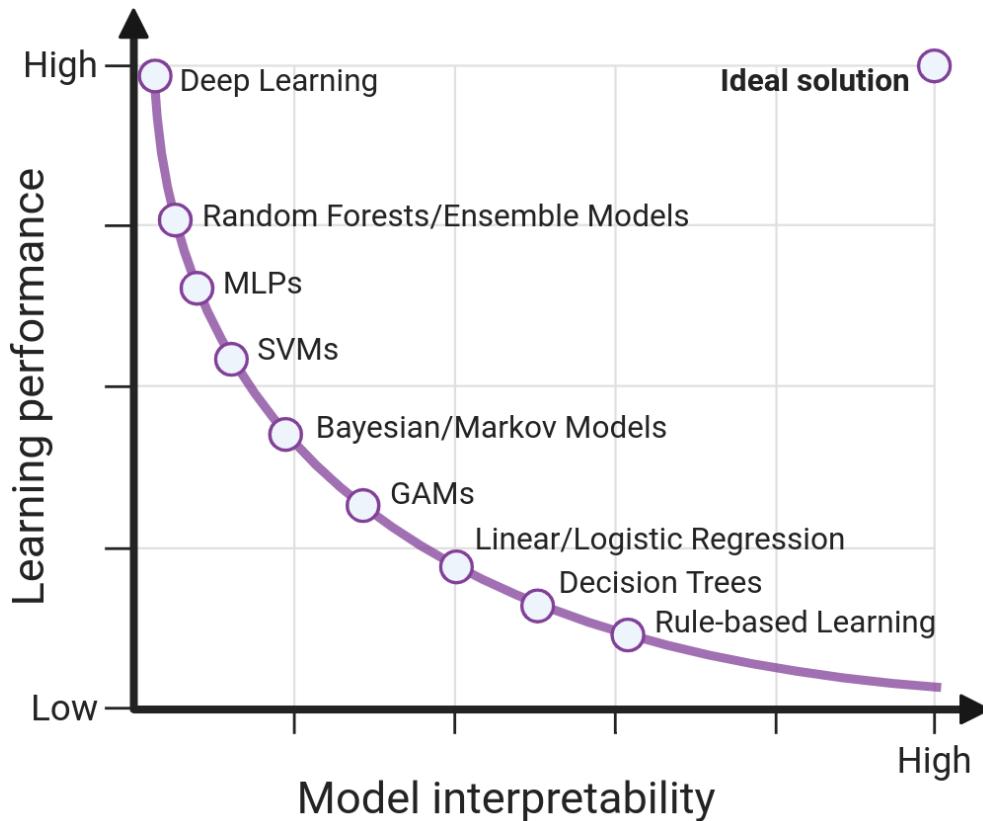


Weird

- C Buckner. Understanding adversarial examples requires a theory of artefacts for deep learning. *Nature Machine Intelligence*, 2020.

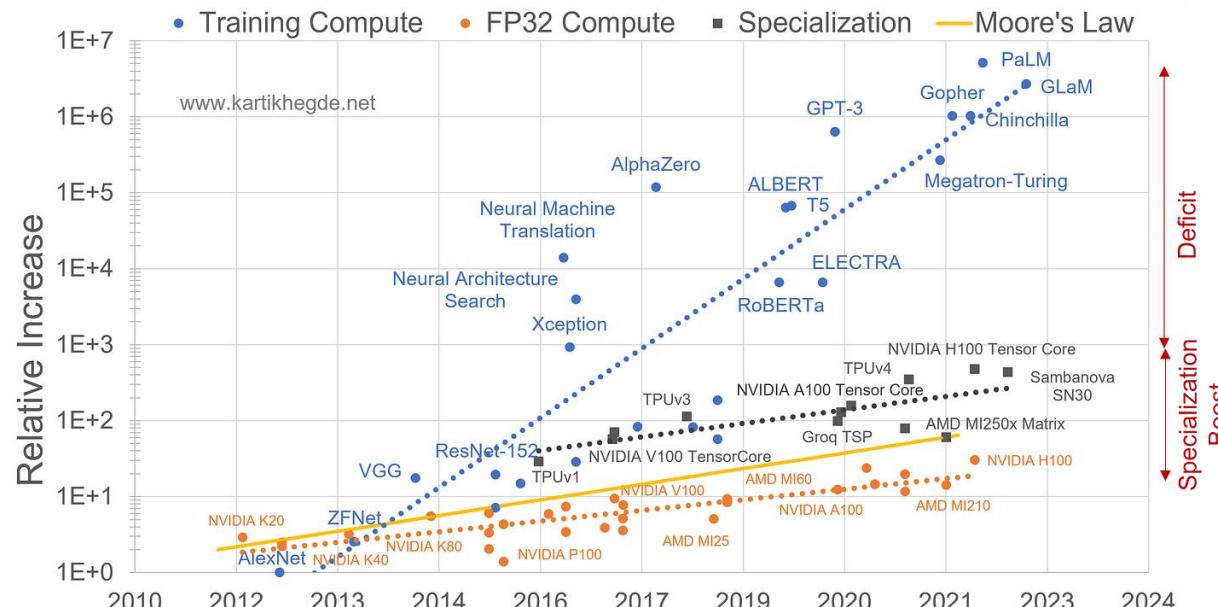
Interpretability of Empirical Learning

- Interpretability: the behavior of a system can be **understood** or **predicted** by humans.



Efficiency of Empirical Learning

- Efficiency: the **real-time availability** and **energy cost** during human-computer interaction.



Common carbon footprint benchmarks

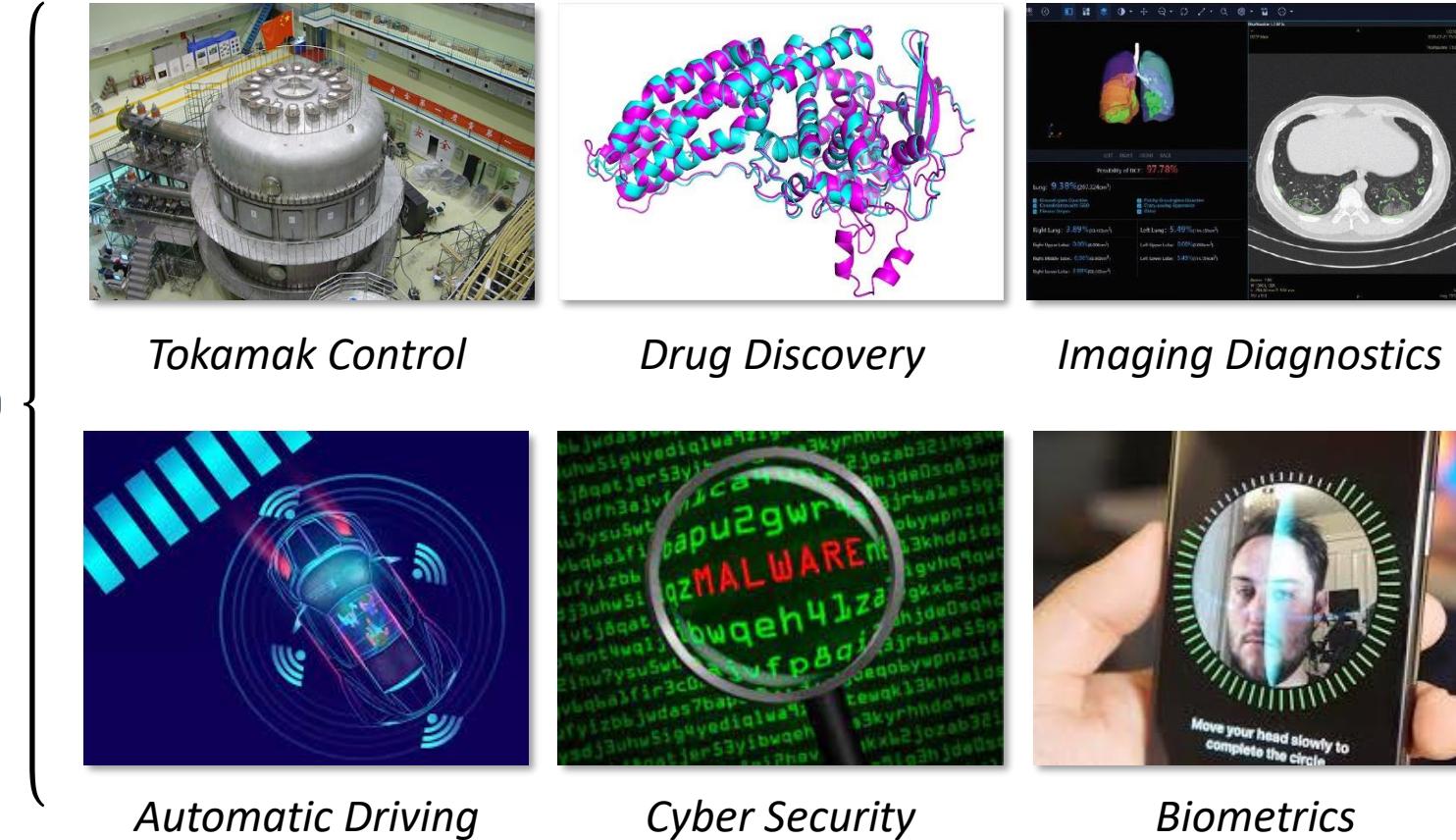
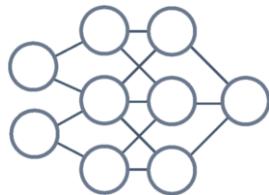
in lbs of CO₂ equivalent

Roundtrip flight b/w NY and SF (1 passenger)	1,984
Human life (avg. 1 year)	11,023
American life (avg. 1 year)	36,156
US car including fuel (avg. 1 lifetime)	126,000
Transformer (213M parameters) w/ neural architecture search	626,155



- E Strubell, A Ganesh, A McCallum, et al. Energy and policy considerations for modern deep learning research. AAAI, 2020.

When Moving Towards Trustworthy AI



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

A Fundamental Prior Underlying
Both Natural World And AI Systems

Invariance/Symmetry in Natural World

- A **symmetry** of a system is a transformation that leaves a certain property **invariant**.



F. Klein, 1872
Erlangen Program



E. Noether, 1918
Noether's Theorem



H. Weyl, 1929
The Book of Symmetry



C. N. Yang & R. L. Mills, 1954
Yang-Mills Theory



- F Klein. A comparative review of recent researches in geometry. *Bulletin of the American Mathematical Society*, 1893.
 - H Weyl. *Symmetry*. Princeton University Press, 2015.

Invariance/Symmetry in AI Systems

- An AI system is a digital modeling of the physical systems in the natural world.



Y. LeCun, Y. Bengio & G. Hinton, 2015,
Deep learning, Nature

The Selectivity–Invariance Dilemma:
“representations that are selective to the aspects that are important for discrimination, but that are invariant to irrelevant aspects”



- Y LeCun, Y Bengio, G Hinton. Deep learning. *Nature*, 2015.
- Y Bengio, A Courville, P Vincent. Representation learning: A review and new perspectives. *TPAMI*, 2013.

How Invariance/Symmetry Helps Robustness, Interpretability, Efficiency

- Perfect robustness — the performance of the AI system remains invariant with respect to the transformations of interest.
- Interpretable concept — humans and AI systems share a basic concept that allows humans to predict AI behavior on transformations of interest.
- Structural efficiency — AI systems no longer need to memorize non-discriminative data variants.



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

Invariance/Symmetry is Ubiquitous in AI Tasks

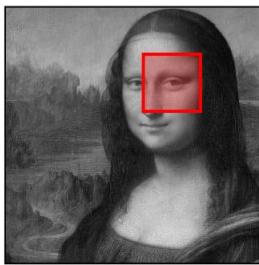
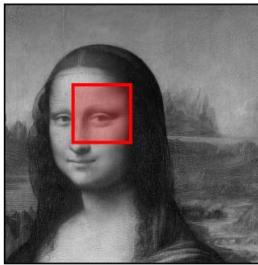
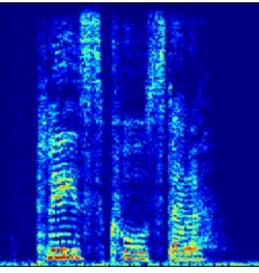
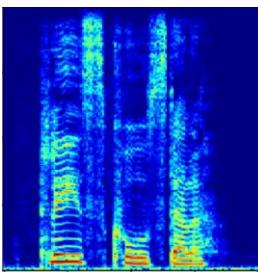


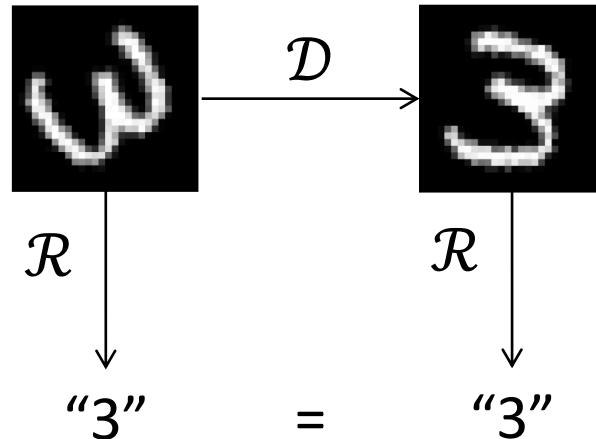
Image Classification
position



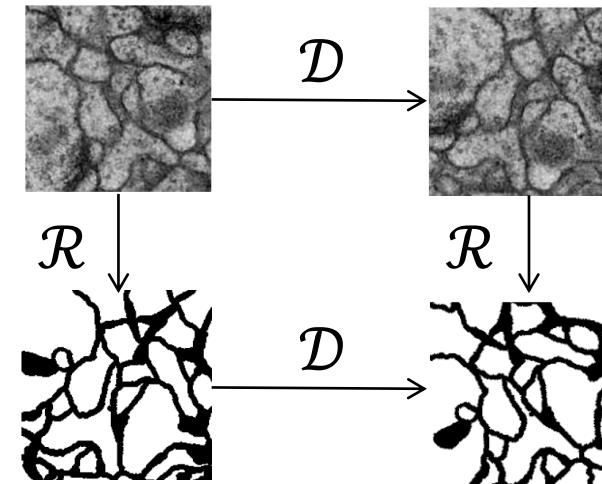
Speech Command
time warping

A Formalization of Invariance/Symmetry (in Representation)

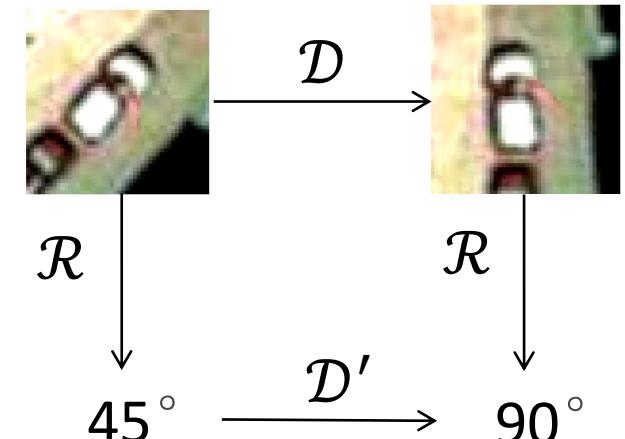
- Invariance: $\mathcal{R}(\mathcal{D}(f)) \equiv \mathcal{R}(f)$
 - Equivariance: $\mathcal{R}(\mathcal{D}(f)) \equiv \mathcal{D}(\mathcal{R}(f))$
 - Covariance: $\mathcal{R}(\mathcal{D}(f)) \equiv \mathcal{D}'(\mathcal{R}(f))$
 - \mathcal{R} is a representation, \mathcal{D} is a degradation, and invariance/equivariance is a special case of covariance with $\mathcal{D}' = \text{id}/\mathcal{D}$
- K Lenc, A Vedaldi. Understanding image representations by measuring their equivariance and equivalence. CVPR, 2015.



Invariance
 \mathcal{D} = rotate, \mathcal{R} = classifier



Equivariance
 \mathcal{D} = rotate, \mathcal{R} = detector



Covariance
 \mathcal{D} = rotate, \mathcal{R} = estimator

A History of Invariance/Symmetry (in Representation)

Algebraic
Invariants

Geometric
Invariants

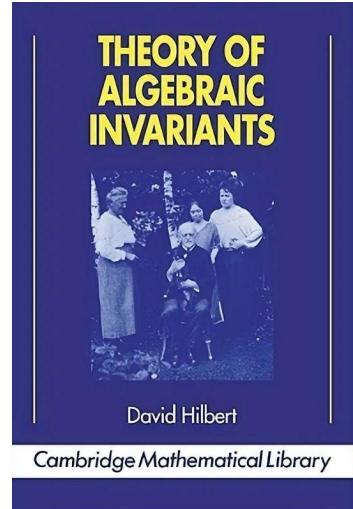
Moment
Invariants

Multiscale
and Wavelet

CNN to Geometry
Deep Learning

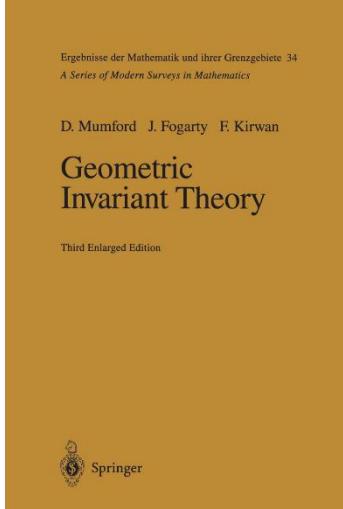
1840s

Hilbert Cayley Klein...



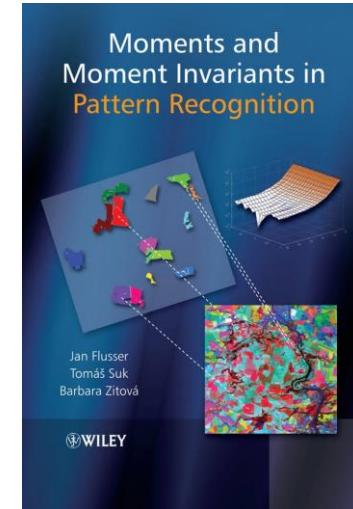
1960s

Mumford



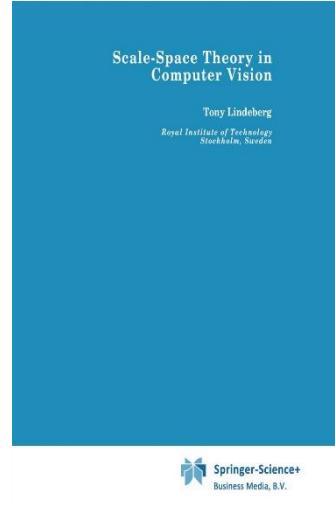
2000s

Flusser



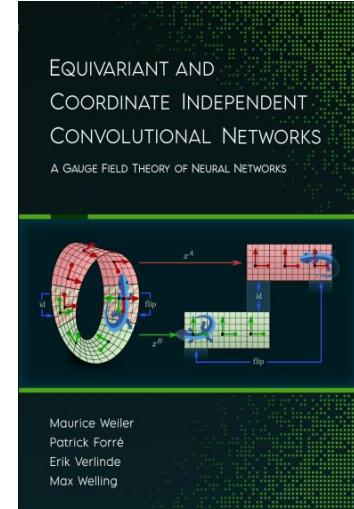
2010s

Lowe Lindeberg Mallat...



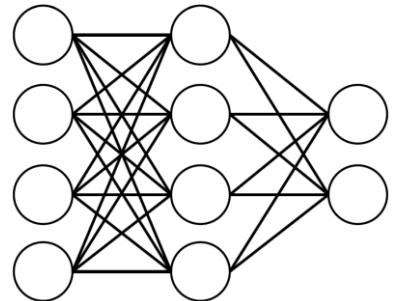
2020s

LeCun Bronstein...

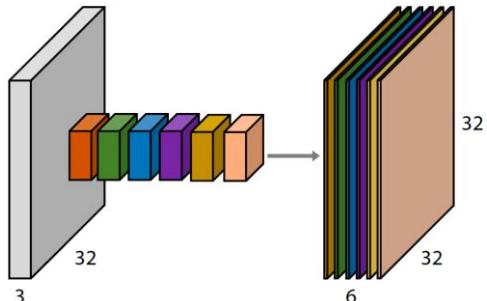


Rethinking Representations by Invariance/Symmetry

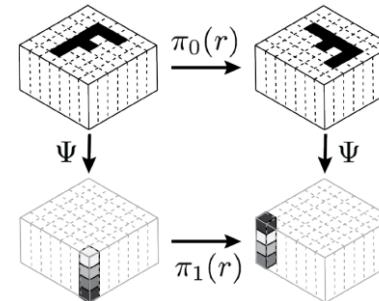
Geometric Deep Learning



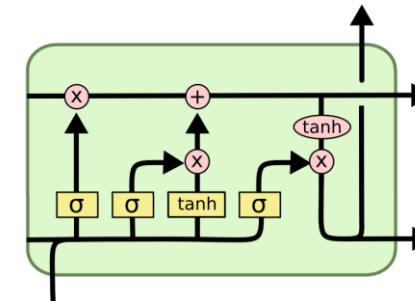
Perceptrons
function regularity



CNNs
translation



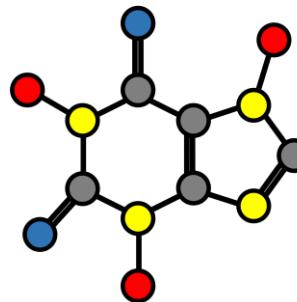
Group-CNNs
translation+rotation



LSTMs
time warping



DeepSets /Transformers
permutation



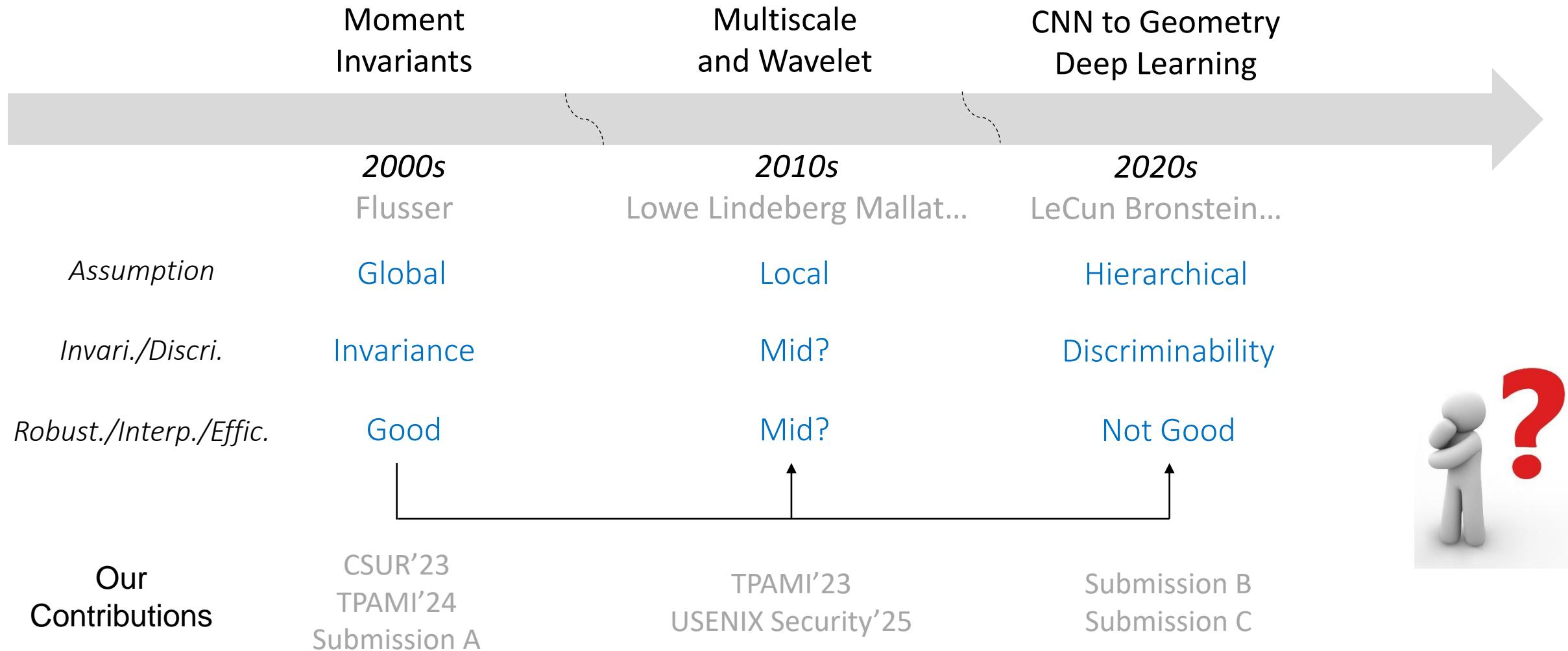
GNNs
permutation



Intrinsic CNNs
isometry/gauge choice

What I Did With My Collaborators
In The Process Of Invariance?

Our Contributions



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



Time for
a Break!!

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

Invariance in The Era Before Deep Learning

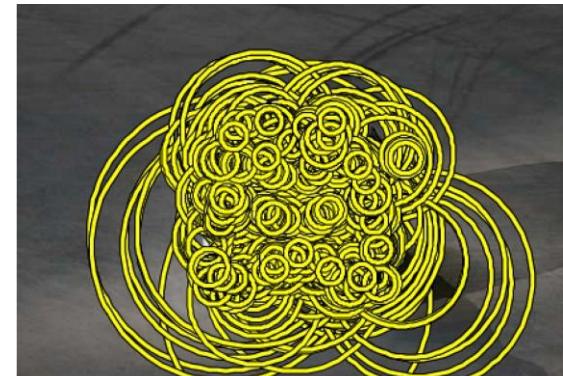
- In the era before deep learning, data representations were almost always designed by experts manually, driven by **knowledge** in math, physics, signal processing, and computer vision.
- Depending on the **spatial scope** of the action, these representations can be classified as **global**, **locally sparse** and **locally dense**. Such assumptions are different and lead to different realizations of invariance.



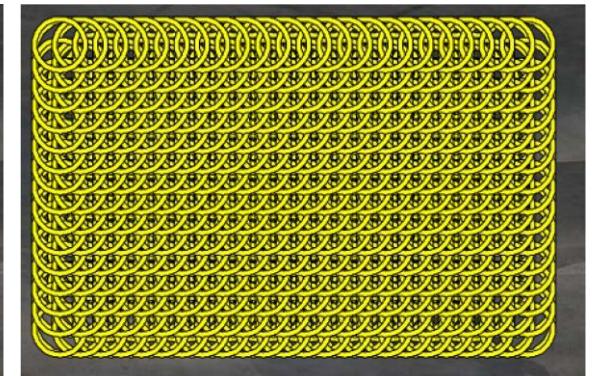
Original Image



Global Representation



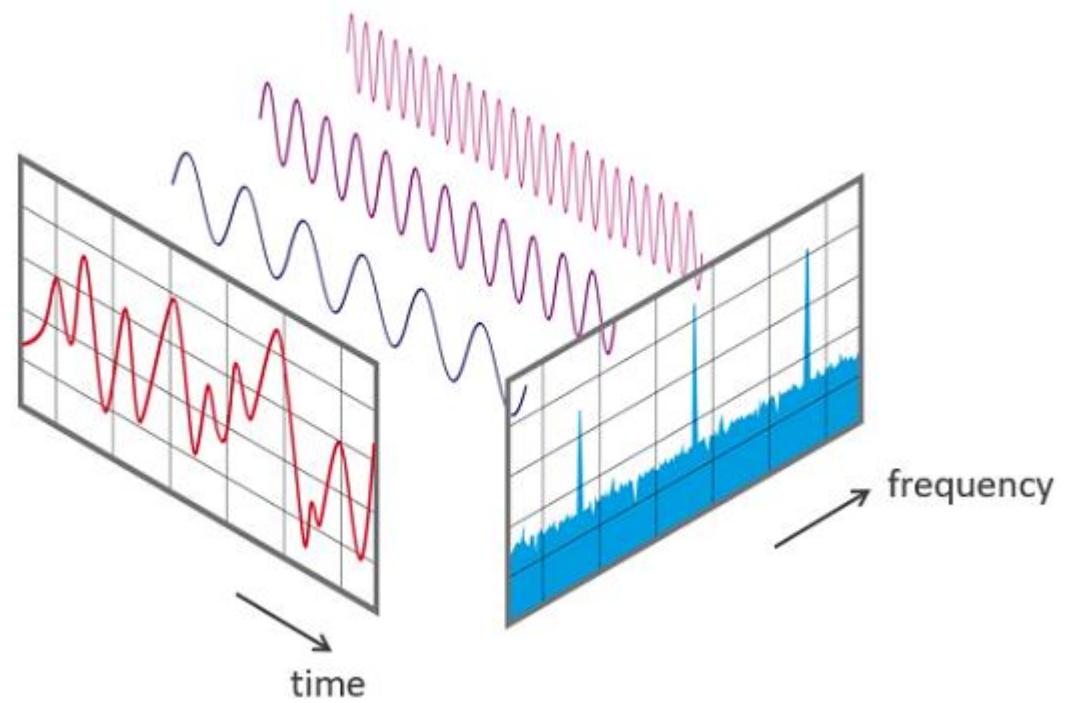
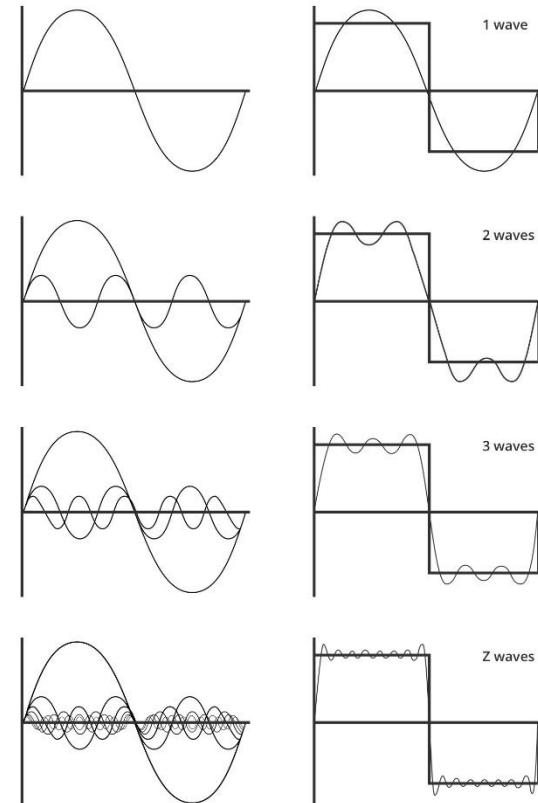
*Locally Sparse
Representation*



*Locally Dense
Representation*

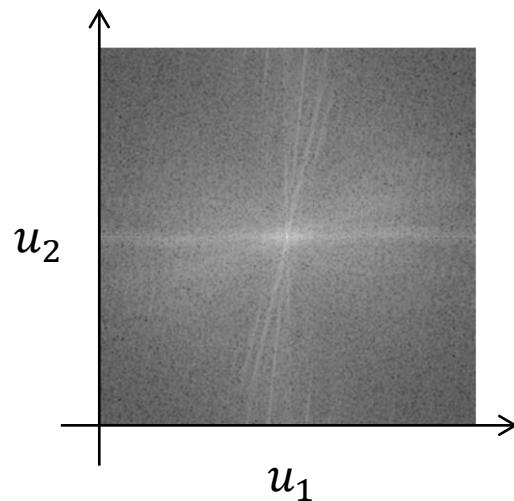
Global Representations: Fourier Transform

- **Fourier Transform** is a tool that rewrite a (continuous and smooth) function as a (coefficient-weighted) sum of sine/cosine functions.



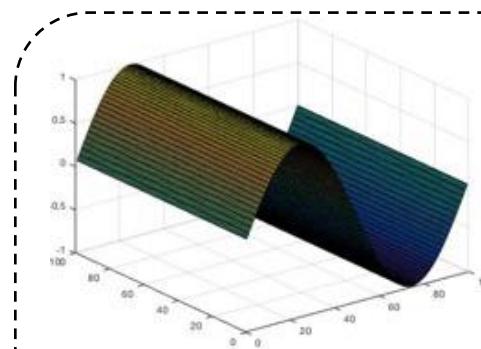
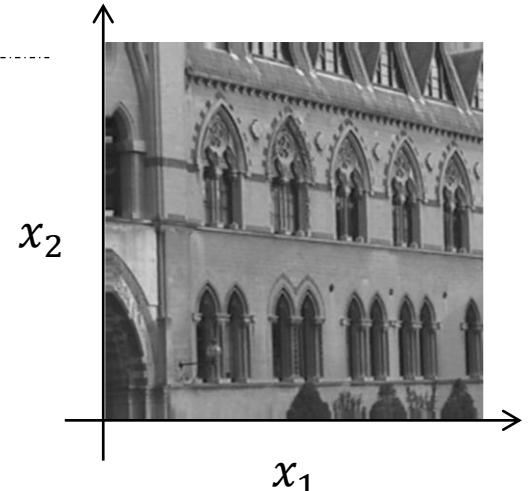
Global Representations: Fourier Transform

- Image, as a 2D function, can also be rewritten as a sum of 2D sine/cosine functions:

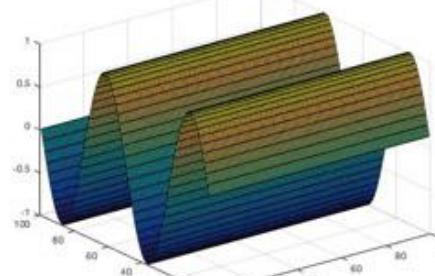


$$\mathcal{F}(f)(\mathbf{u}) \equiv F(\mathbf{u}) = \int_{\mathbb{R}^d} [e^{-2\pi i \mathbf{u} \cdot \mathbf{x}}] f(\mathbf{x}) d\mathbf{x}$$

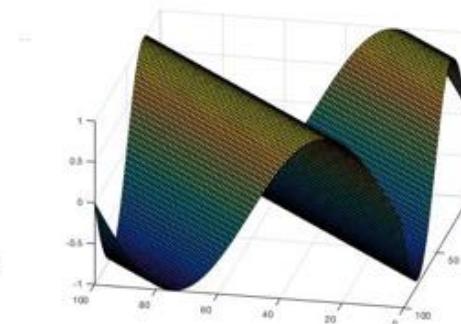
$$\mathcal{F}^{-1}(F)(\mathbf{x}) = \int_{\mathbb{R}^d} e^{2\pi i \mathbf{u} \cdot \mathbf{x}} F(\mathbf{u}) d\mathbf{u} = f(\mathbf{x})$$



$$u_1 = 1, u_2 = 0$$



$$u_1 = 0, u_2 = 2$$



$$u_1 = 1, u_2 = 1$$

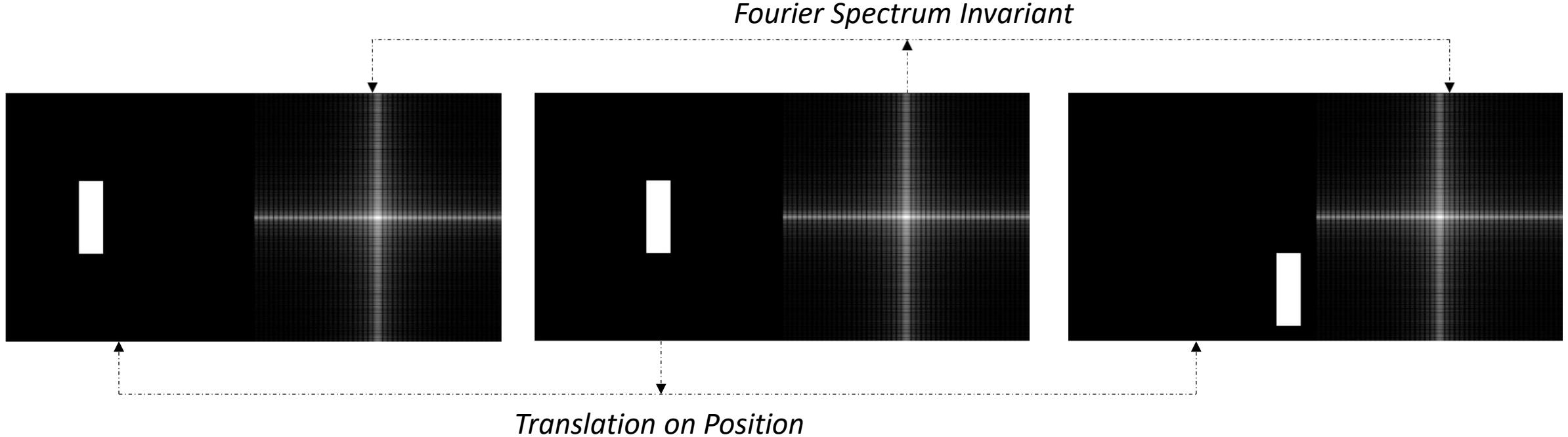
So, How About Invariance?

Translation Invariance of Fourier Transform

- Translating the function leads to multiplying the Fourier transform by a phase factor:

$$\mathcal{F}(f(\mathbf{x} - \mathbf{t}))(\mathbf{u}) = [e^{-2\pi i \mathbf{u} \cdot \mathbf{t}}] \mathcal{F}(f(\mathbf{x}))(\mathbf{u})$$

- As a consequence, the absolute values of Fourier transform are invariant to translation.



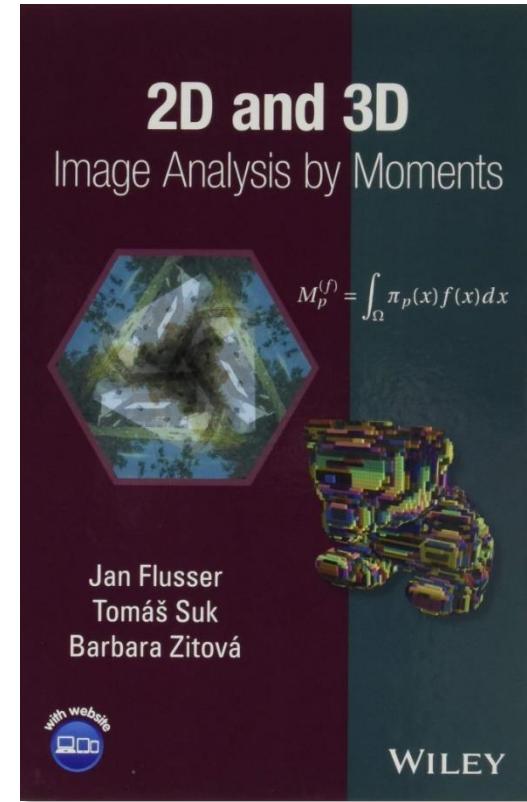
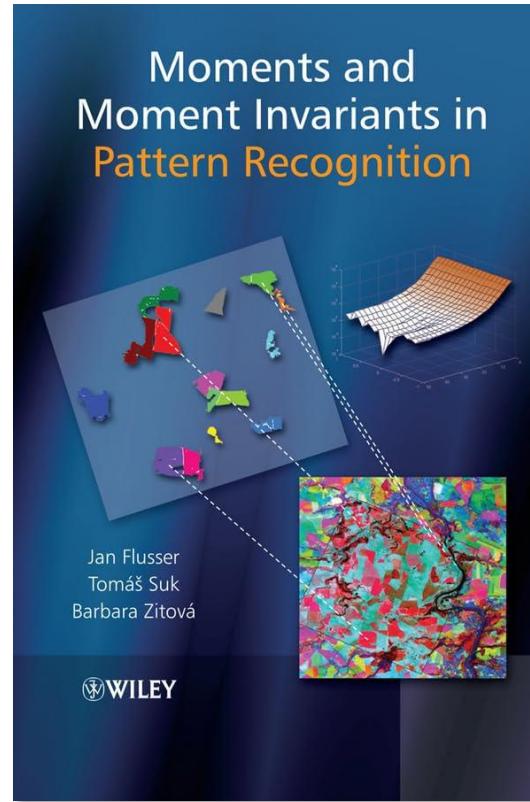
Can Global Invariance Be Generalized To Other
Geometric Transformations?

Global Representations: Moment Invariants

- **Moment Invariants** are similar to Fourier transforms in that they also rewrite the function as a (coefficient-weighted) sum of basis functions, but with a different purpose — more generalized invariants.



J. Flusser, B. Zitova, & T. Suk, 2009
Moment Invariants



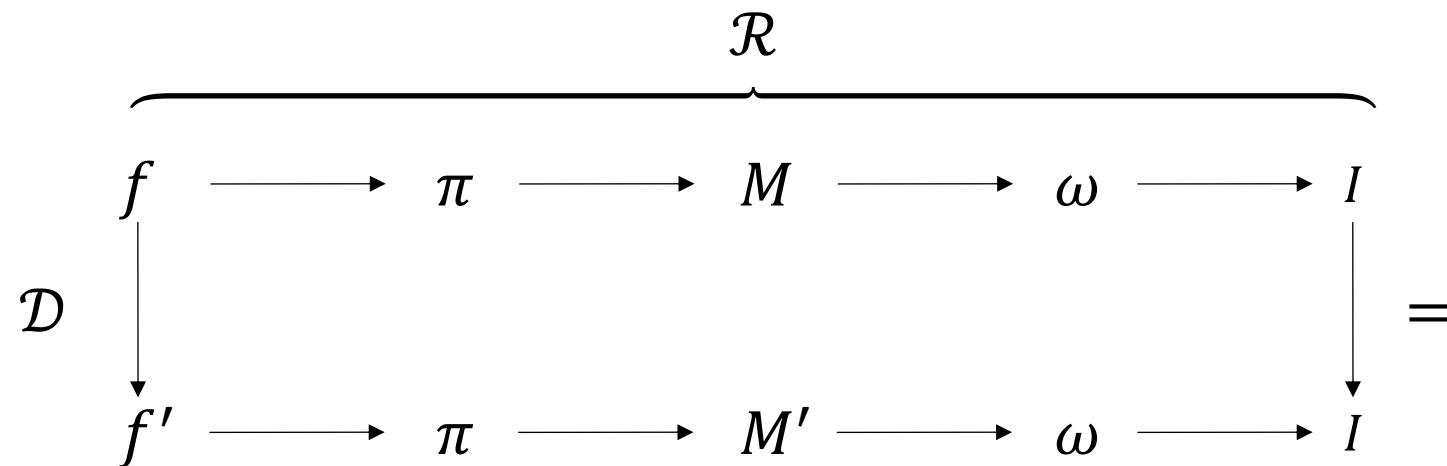
- J. Flusser, B. Zitova, T. Suk. *Moments and Moment Invariants in Pattern Recognition*. John Wiley & Sons, 2009.

Moments as a Generic Form of Global Representation

- Fundamentally, **moments** have a very simple definition, and is in fact a **generic form of the global representation**:

$$M_p^{(f)} = \int_{\Omega} \pi_p(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

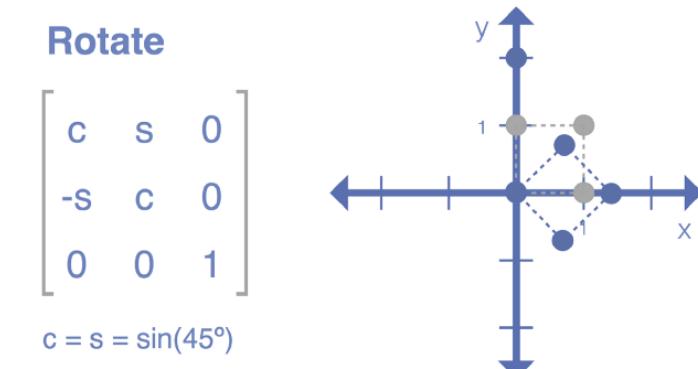
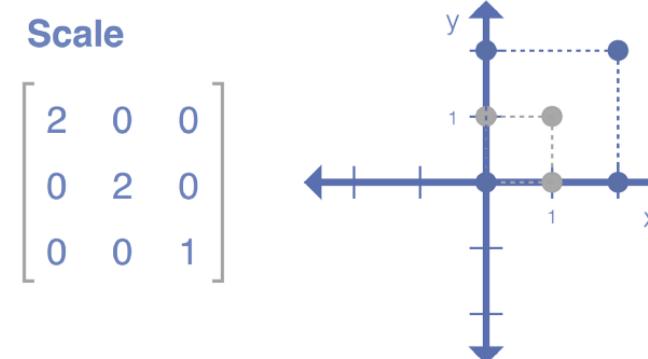
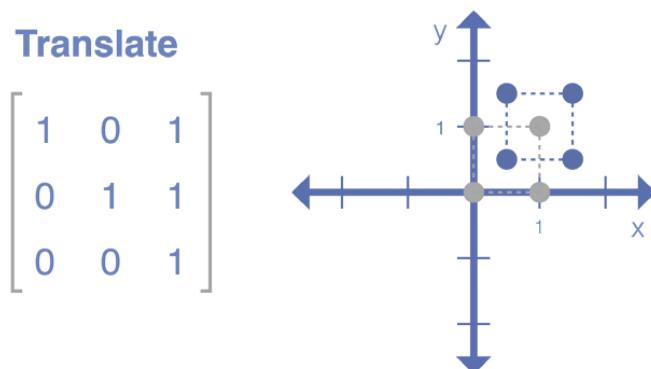
- Here, the core is how such basis functions π are designed so that more generalized invariants I can be derived from the corresponding moments M by a certain cancellation ω .



Geometric Transformations and Geometric Moments

- Let us consider the basic geometric transformations, including **translation, rotation and scaling**, which can be modeled as:

$$\mathbf{x}' = s\mathbf{R}_\alpha \mathbf{x} + \mathbf{t} \quad \mathbf{R}_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$



- We can also define the so-called **geometric moments** with very simple basis functions:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

Translation and Scaling Invariants

- With the above definitions, **translation invariants μ** can be derived from the geometric moments:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy \quad x_c = m_{10}/m_{00}, \quad y_c = m_{01}/m_{00}$$

- where (x_c, y_c) should be considered as the **centroid** of the image. The invariance is achieved by aligning the coordinate origin of the basis functions with the centroid.
- Let us further consider **scaling invariants v** , which again can be derived from geometric moments, by normalizing the scaling factor on moments:

$$\mu'_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x/s, y/s) dx dy \quad v_{pq} = \frac{\mu_{pq}}{\mu_{00}^w} \quad w = \frac{p+q}{2} + 1$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^p (x - x_c)^p s^q (y - y_c)^q f(x, y) s^2 dx dy = \boxed{s^{p+q+2} \mu_{pq}} \quad v'_{pq} = \frac{\mu'_{pq}}{(\mu'_{00})^w} = \frac{s^{p+q+2} \mu_{pq}}{(s^2 \mu_{00})^w} = v_{pq}$$

Rotation Invariants by Hu and Hilbert

- Are **rotation invariants ϕ** equally derivable from geometric moments? Yes, **Hu** gives 7 invariants based on **Hilbert's algebraic invariants**, which seems very complex. But it makes sense, due to the nonlinear action of the rotations on x and y .

$$\phi_1 = m_{20} + m_{02},$$

$$\phi_2 = (m_{20} - m_{02})^2 + 4m_{11}^2,$$

$$\phi_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2,$$

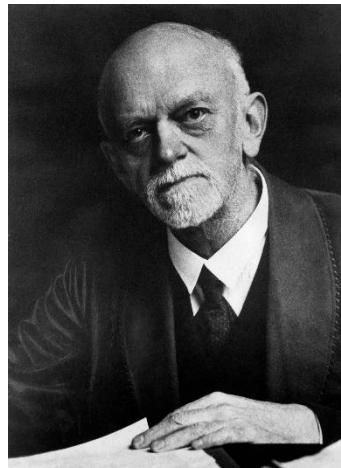
$$\phi_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2,$$

$$\begin{aligned}\phi_5 = & (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ & + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2),\end{aligned}$$

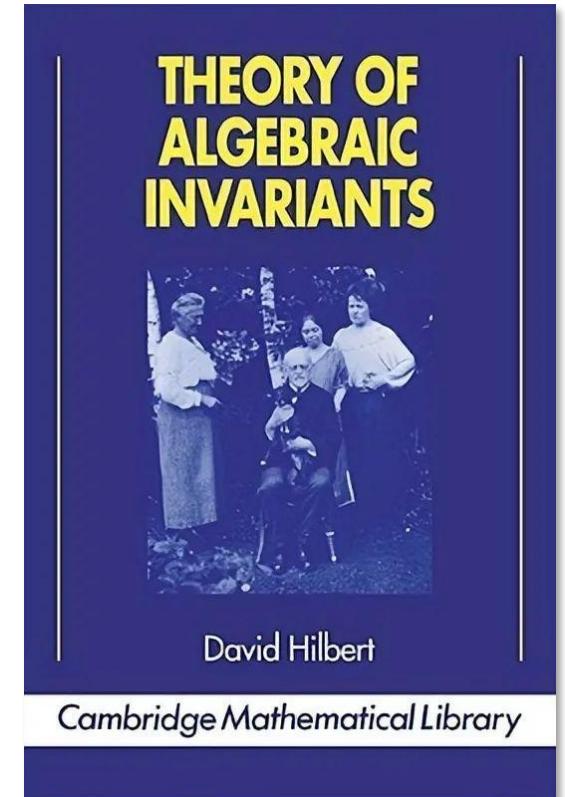
$$\begin{aligned}\phi_6 = & (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) \\ & + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03}),\end{aligned}$$

$$\begin{aligned}\phi_7 = & (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ & - (m_{30} - 3m_{12})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2).\end{aligned}$$

- MK Hu. Visual pattern recognition by moment invariants. *TIT*, 1962.



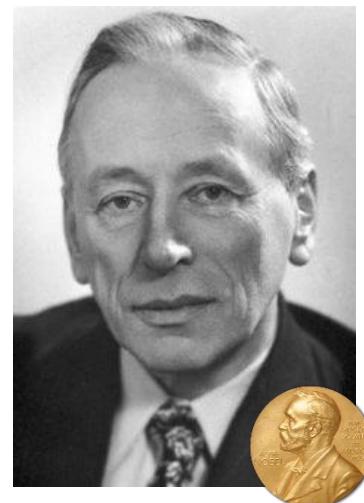
D. Hilbert, 1897
Algebraic Invariants



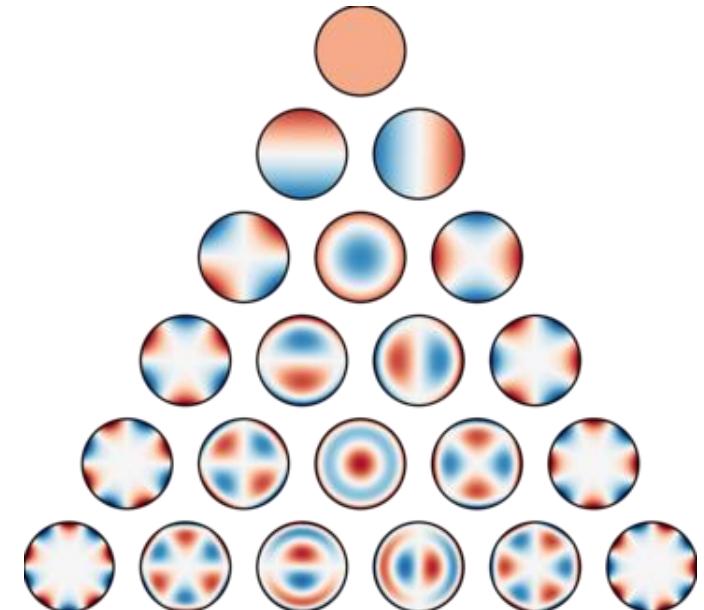
Rotation Invariants by Zernike

- Can rotation invariants be derived more simply? Let us define the basis functions in **polar** coordinates, where the effects caused by rotations are more easily managed, by leveraging the translation theorem of the Fourier transform in an angular form.
- In this respect, **Zernike polynomials** are typical — they are complete orthogonal bases on the unit circle and easily realize rotation invariance, from **Zernike's optical research**.

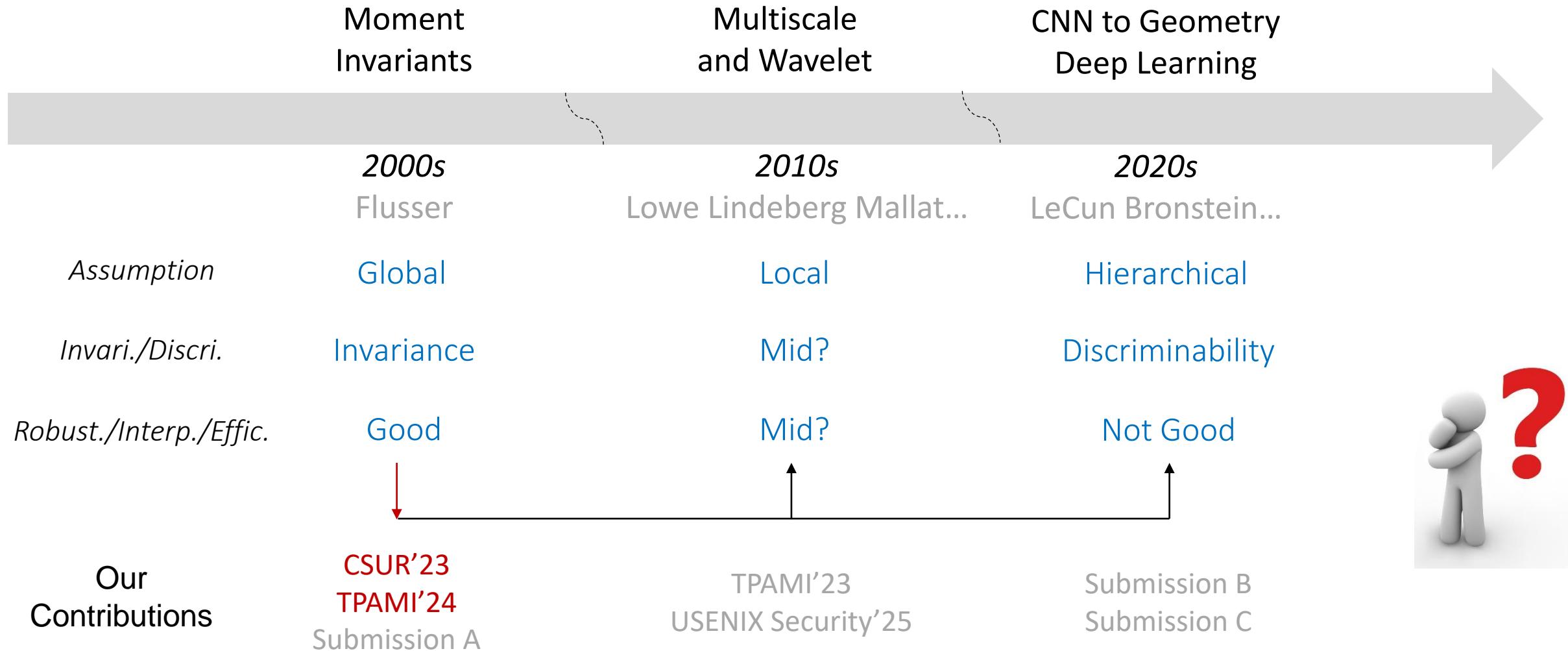
$$\begin{aligned} C_{pq} &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)\theta} f(r, \theta) r d\theta dr \\ C'_{pq} &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)\theta} f(r, \theta + \alpha) r d\theta dr \\ &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)(\theta-\alpha)} f(r, \theta) r d\theta dr \\ &= \underbrace{\left[e^{-i\xi(p,q)\alpha} \right]}_{\text{---}} C_{pq}. \end{aligned}$$



F. Zernike, 1934
Zernike Polynomials

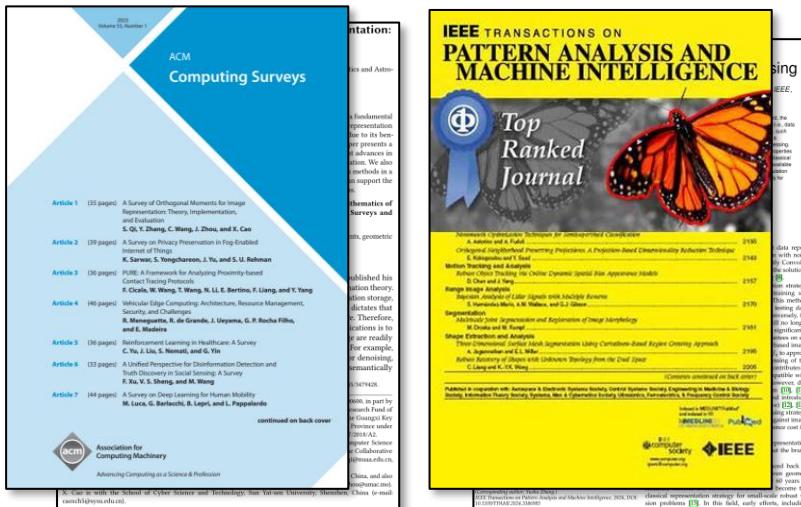


Our Contributions

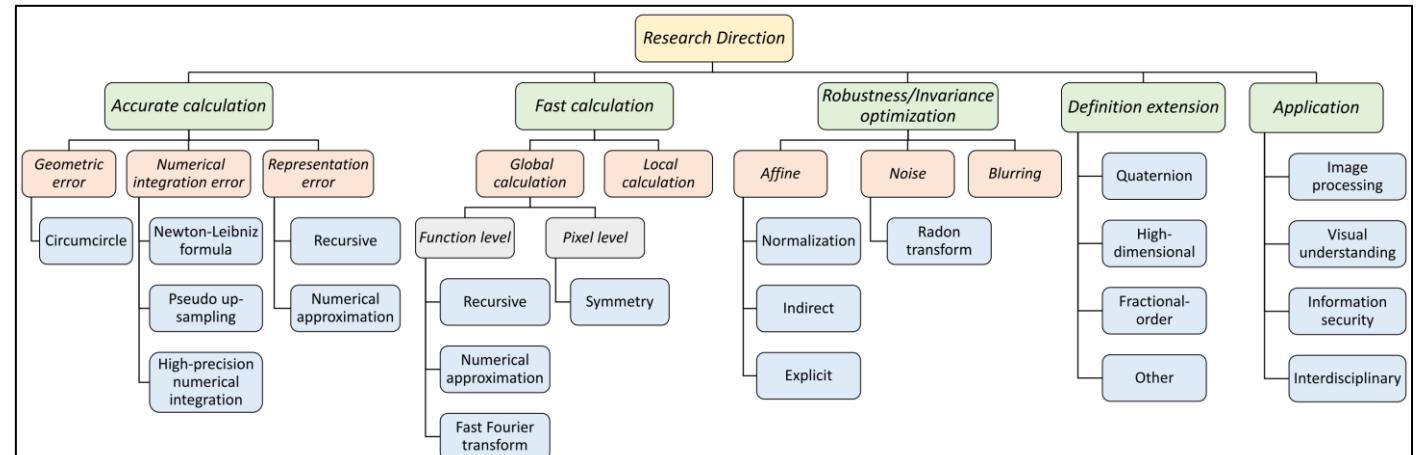


Refining Global Invariants

- We give papers on the practical aspects of moments for refining global invariants, covering **numerical analyses**, **software implementations**, **benchmark evaluations**, and **recent advances**.



- S. Qi, Y. Zhang, C. Wang, et al. A Survey of Orthogonal Moments for Image Representation: Theory, Implementation, and Evaluation. *ACM Computing Surveys (CSUR)*, 2023, 55(1): 1-35.
- S. Qi, Y. Zhang, C. Wang, et al. Representing Noisy Image Without Denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2024, 46(10): 6713 - 6730

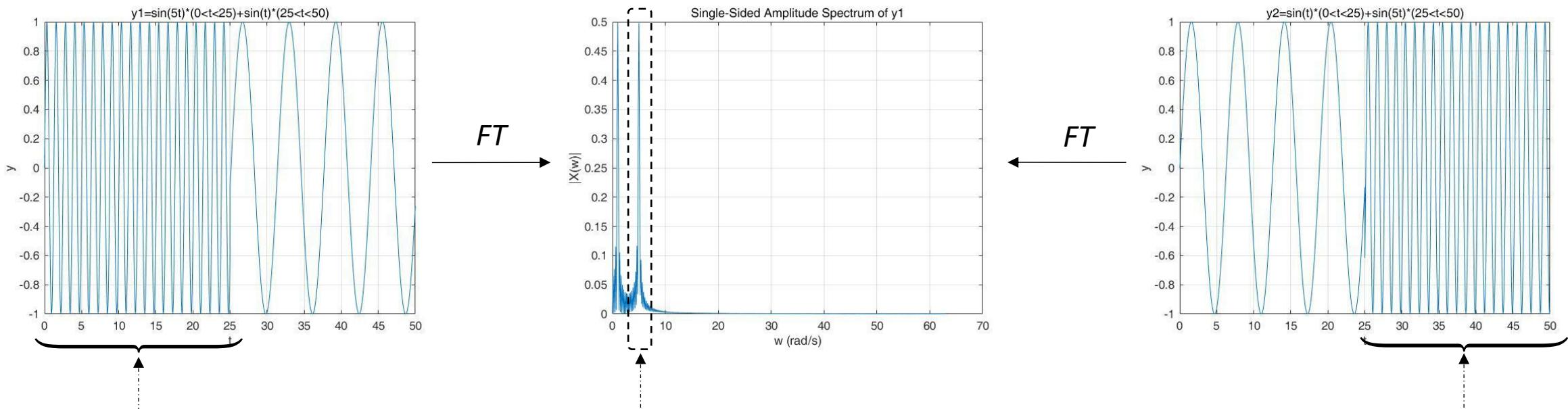


A screenshot of a GitHub repository named 'MomentToolbox' (Public). The repository has 1 branch and 0 tags. The README.md file was updated 4 years ago by ShurenQi. The repository contains three main folders: 1.Moment Calculation (Accuracy and Complexity), 2.Image Reconstruction (Representation Capabilities), and 3.Pattern Recognition (Robustness and Invariance). The URL of the repository is <https://github.com/ShurenQi/MomentToolbox>. The repository has 4 forks and 28 stars. It is categorized under image-processing, pattern-recognition, image-analysis, and orthogonal-moments.

From Global To Local

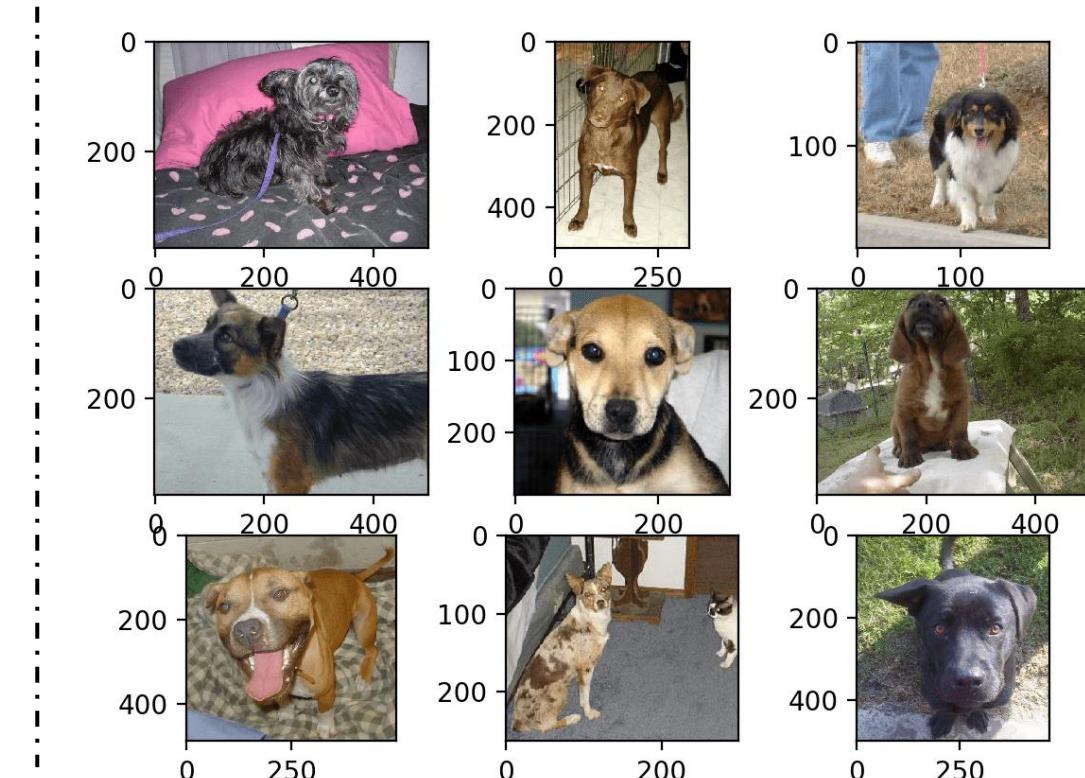
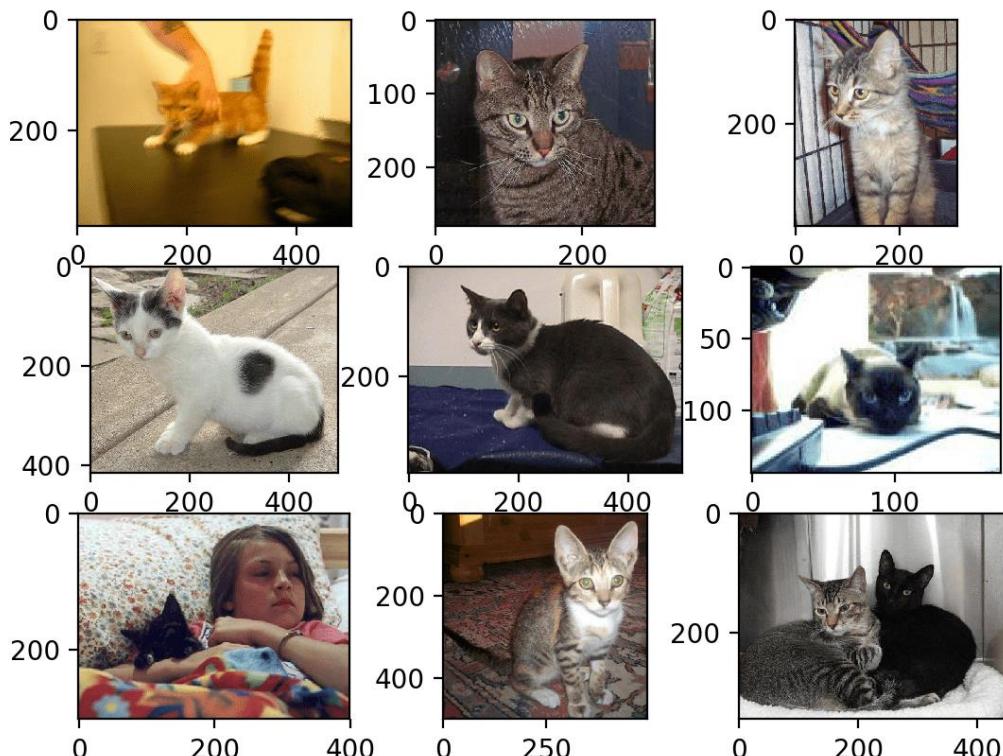
Why We Need Local Representations

- Fourier transform-like global representations are typically **(under)-complete** and are just designed for **low-level processing**, struggling to express high-level semantics with over-completeness.
- As a toy example, the Fourier transform cannot even distinguish the order in which the two signals appear.



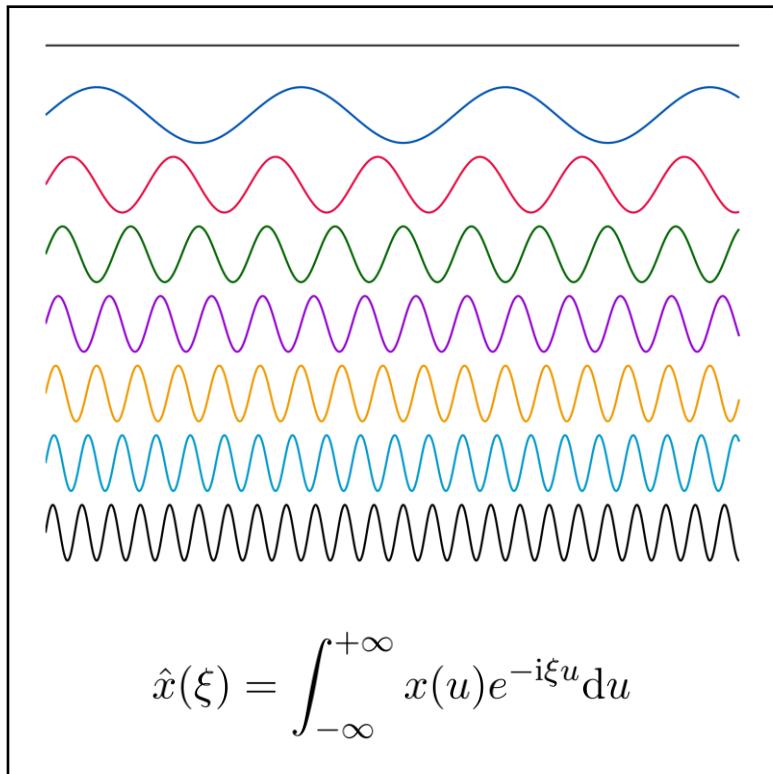
Why We Need Local Representations

- Under realistic considerations, there are too many tasks concerned with local semantic properties — **recognition and classification** (distinguishing images of cats and dogs), where global representations are likely unable to provide enough information to support discriminability.

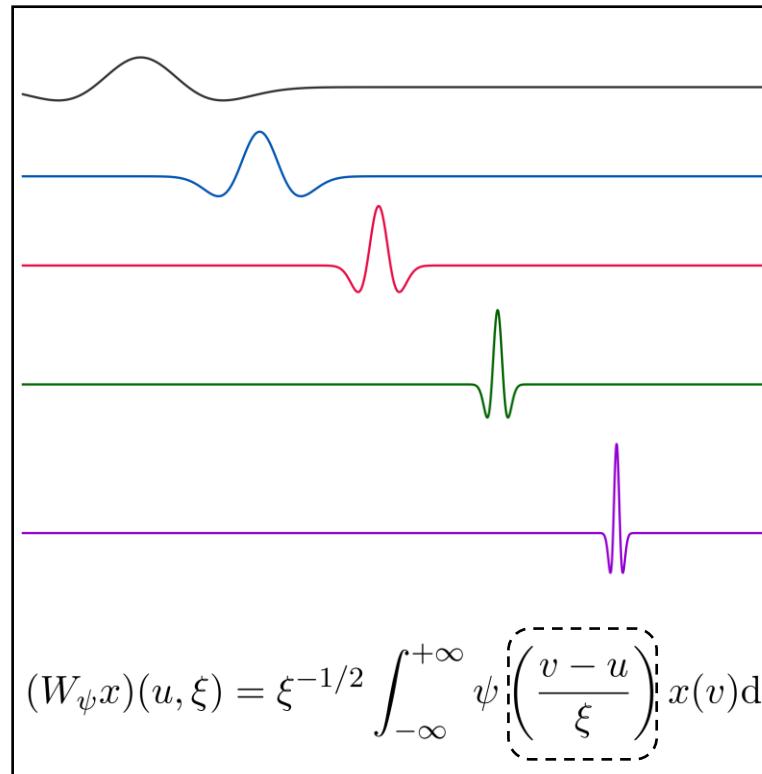


Local Representations: Wavelet Transform

- Different from Fourier, basis functions of **Wavelet Transform** are **local** and **multi-scale**.



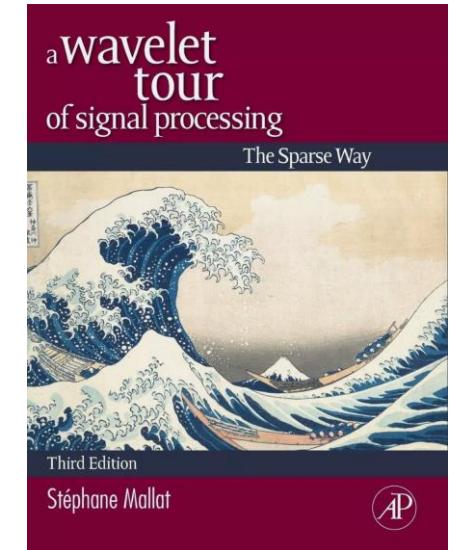
Fourier



Wavelets



S. Mallat, 1999
Wavelets



Local Representations: Wavelet Transform

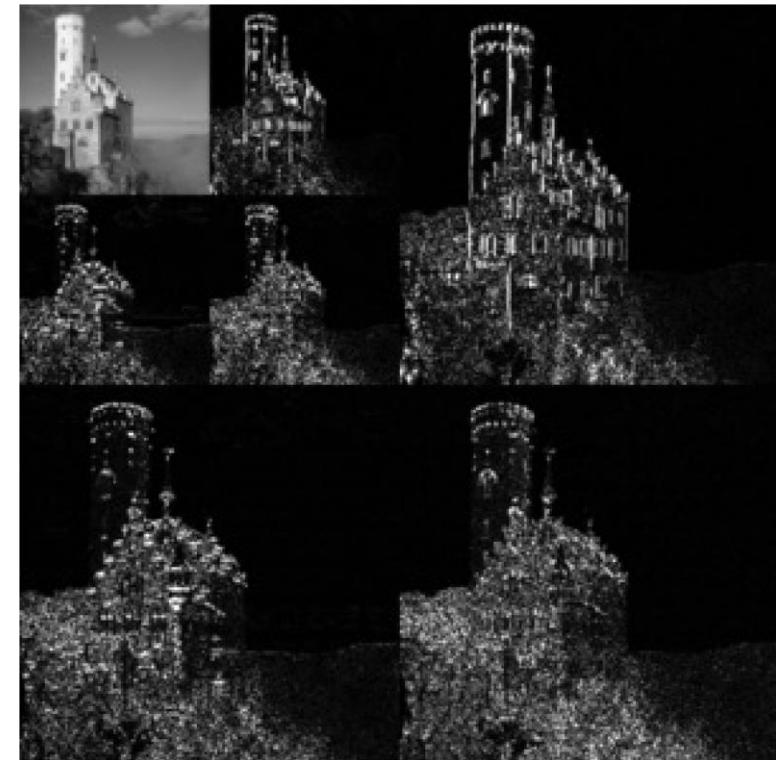
- Wavelet transform can capture local information, with **better discriminative properties** — time-frequency discriminability and over-completeness.



Original Image



Fourier Representations

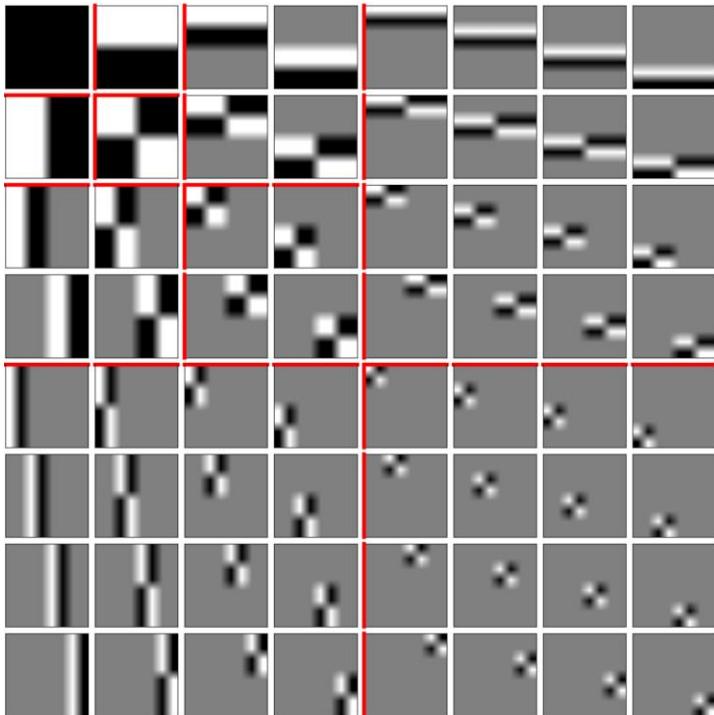


Wavelet Representations

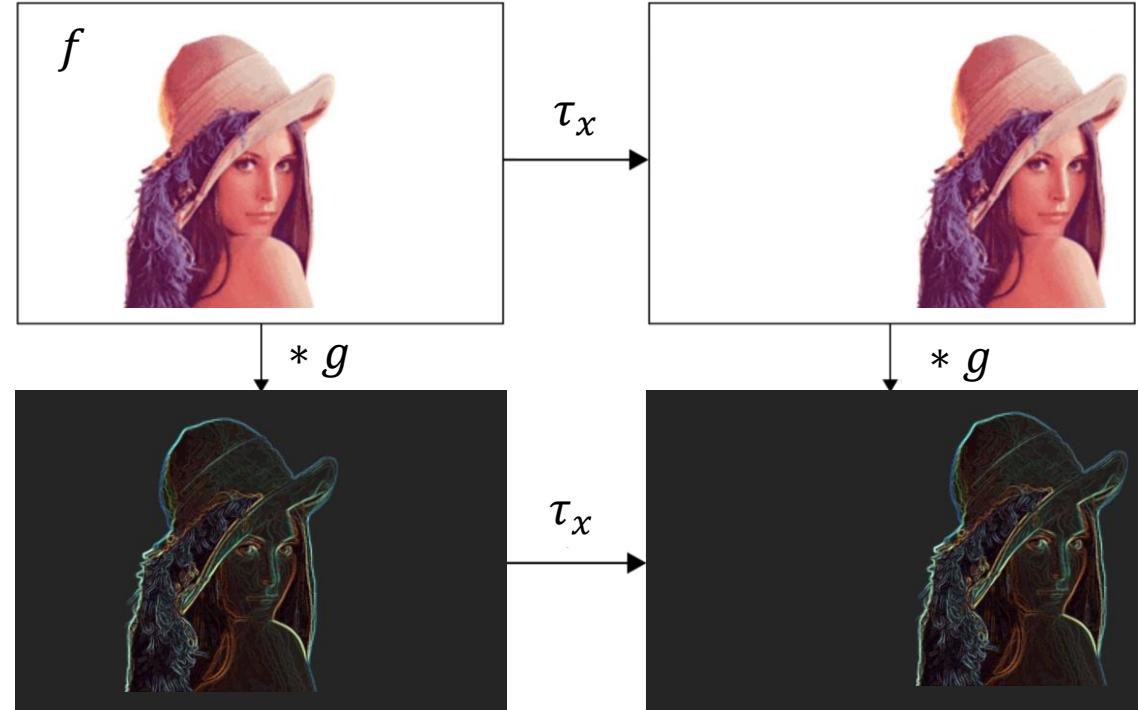
So, How About Invariance?

Translation Equivariance of Wavelet Transform

- The wavelet basis functions define **convolution operators g** — the wavelet transform of an image f means the convolution of f and g . Therefore, the wavelet transform has a translation equivalence with the convolution.



Wavelet Convolutional Operators g



$$(\tau_x f) * g = \tau_x(f * g)$$

Can Local Invariance Be Generalized To Other
Geometric Transformations?

Local Representations: SIFT

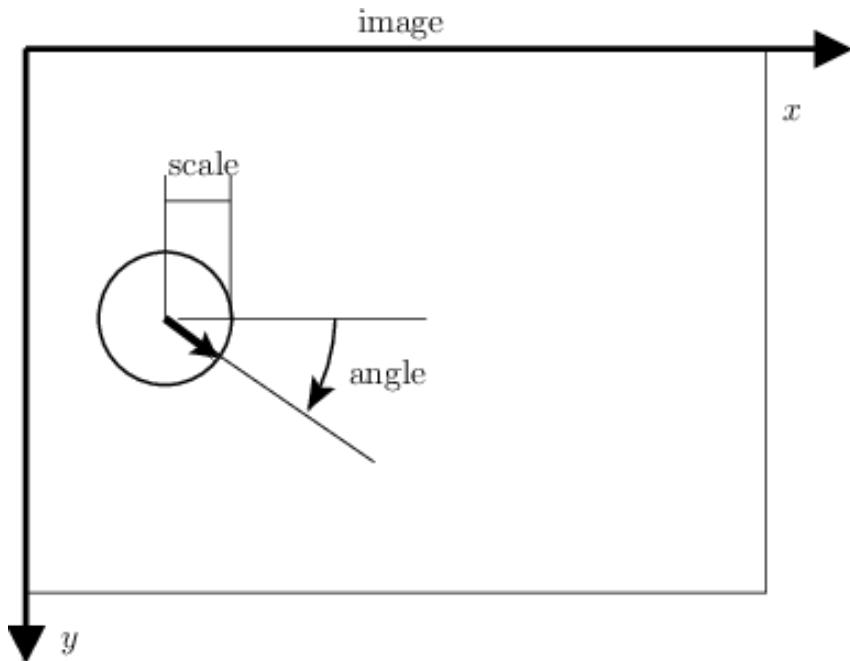
- The local and multiscale concepts of the wavelet transform were **followed** by later local representations.
- For example, the well-known **Scale-Invariant Feature Transform (SIFT)** aims at the **local invariance of rotation and scaling in multiscale spaces**.



- DG Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

Local Representations: SIFT

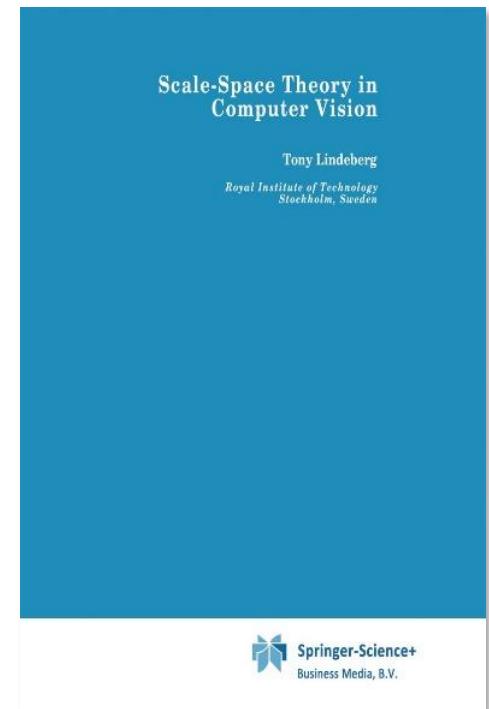
- SIFT describes **local regions that have their own scale and orientation**, with the **scale space theory** as a foundation.
- Here, once the scale and orientation of the regions can be evaluated stably, then invariant features can be constructed by **normalizing** the scale and orientation.



T. Lindeberg, 1993
Scale Space Theory



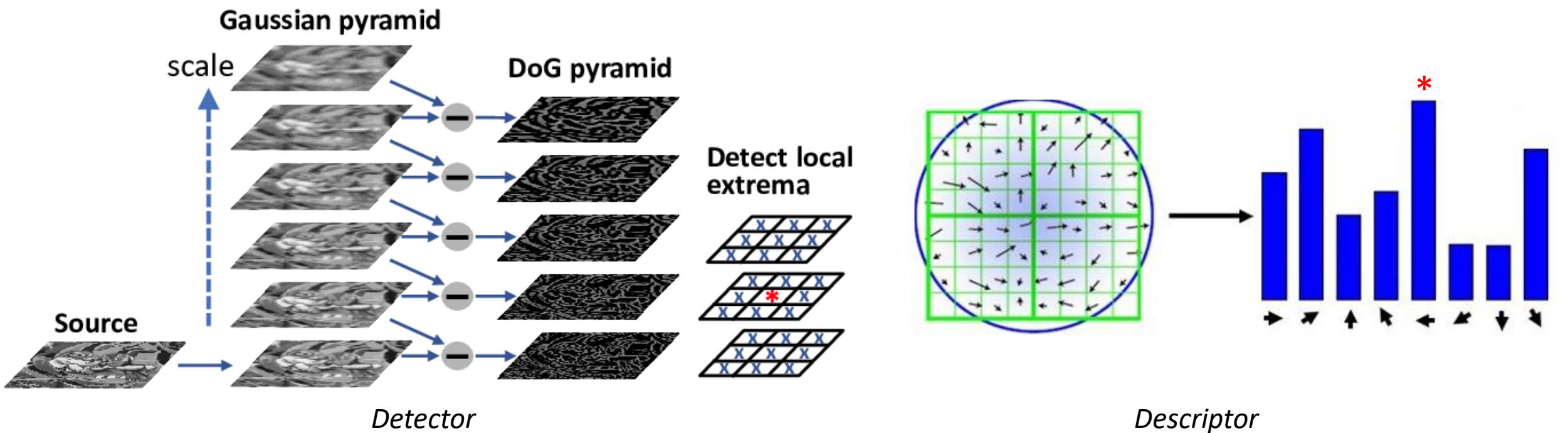
D. Lowe, 1999
SIFT



- T Lindeberg. *Scale-space Theory in Computer Vision*. Springer Science & Business Media, 1993.

Local Representations: SIFT

- SIFT has two main components: **detector** and **descriptor**.
- The detector is responsible for finding the interest point with evaluated scale to achieve **scaling invariance**. The descriptor is responsible for describing the interest point with evaluated orientation to further achieve **rotation invariance**.



Scale is evaluated by finding the extreme in the scale space Orientation is evaluated by computing the histogram of gradients

From Sparse To Dense

Why We Need Dense Representations

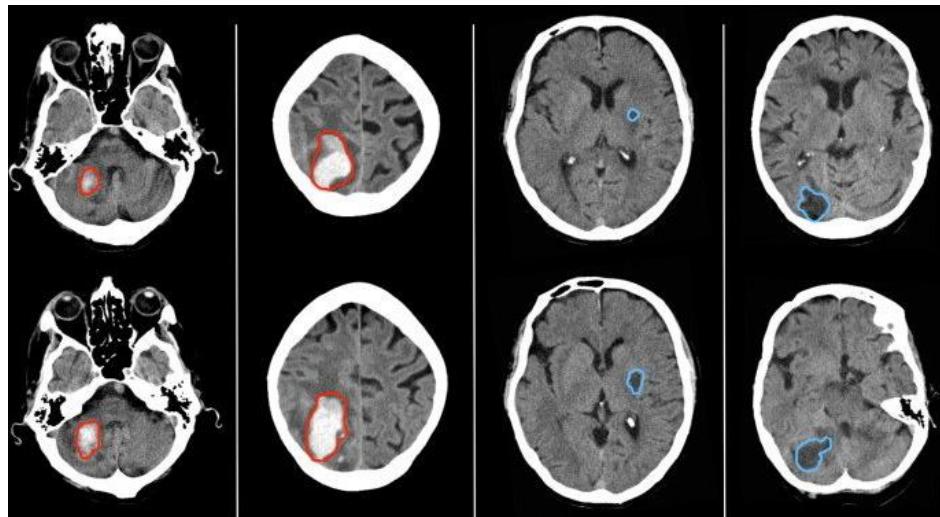
- SIFT-like interest points are **sparse** in the image and are designed to **focus only on the main subject (ignoring all other regions)**.



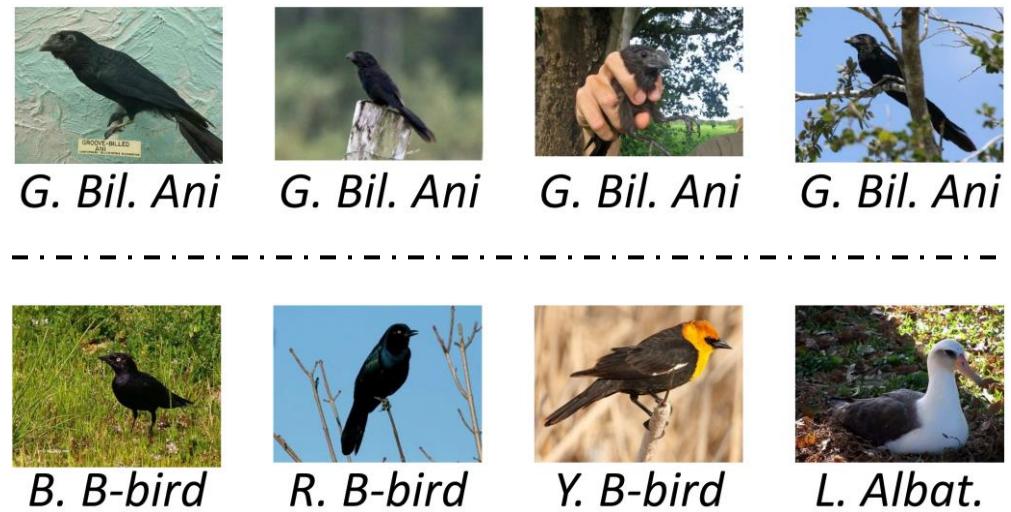
• A Iscen, G Tolias, PH Gosselin, et al. A comparison of dense region detectors for image search and fine-grained classification. *TIP*, 2015.

Why We Need Dense Representations

- Under realistic considerations, there are too many tasks concerned with dense semantic properties — **detection/localization** (detecting lesions in CT images), **fine-grained classification** (distinguishing large-scale bird images), where sparse interest points are likely to miss potentially important local information.



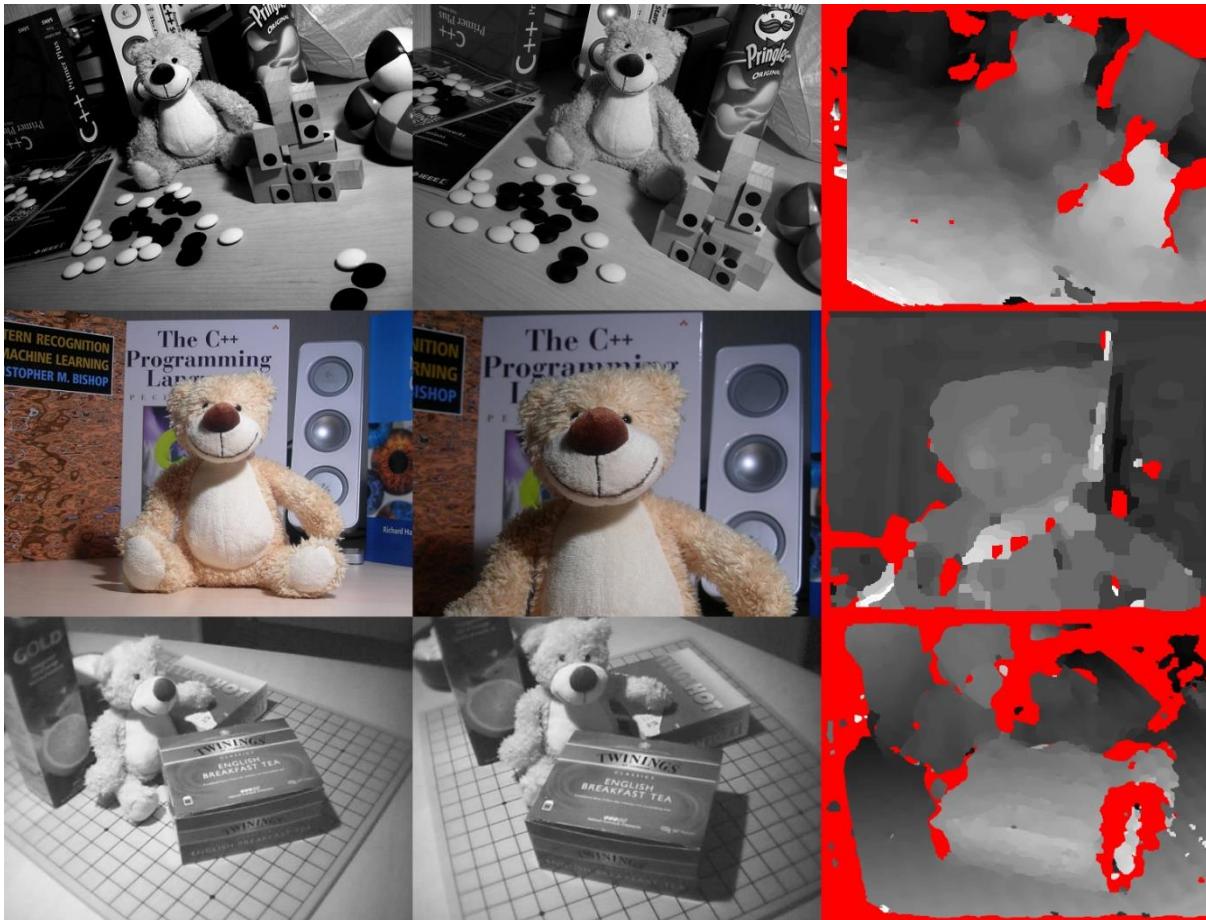
Detection/Localization



Fine-grained Classification

Local Representations: DAISY

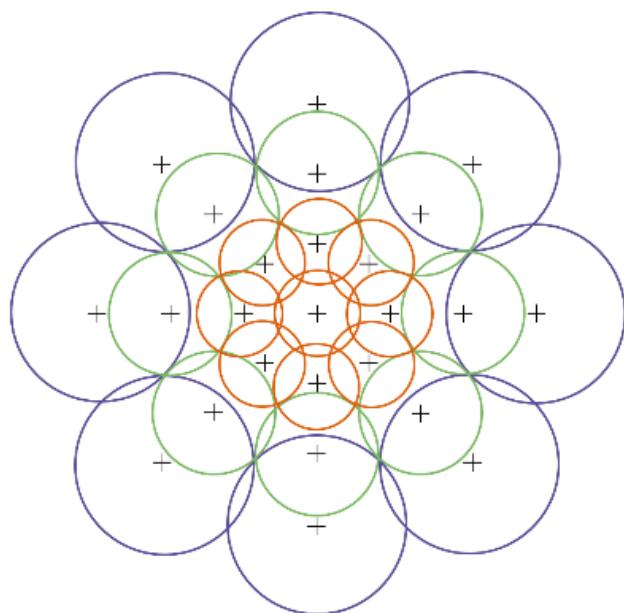
- DAISY aims to extend SIFT from sparse to dense, achieving local invariance of rotation and scaling for each pixel position.



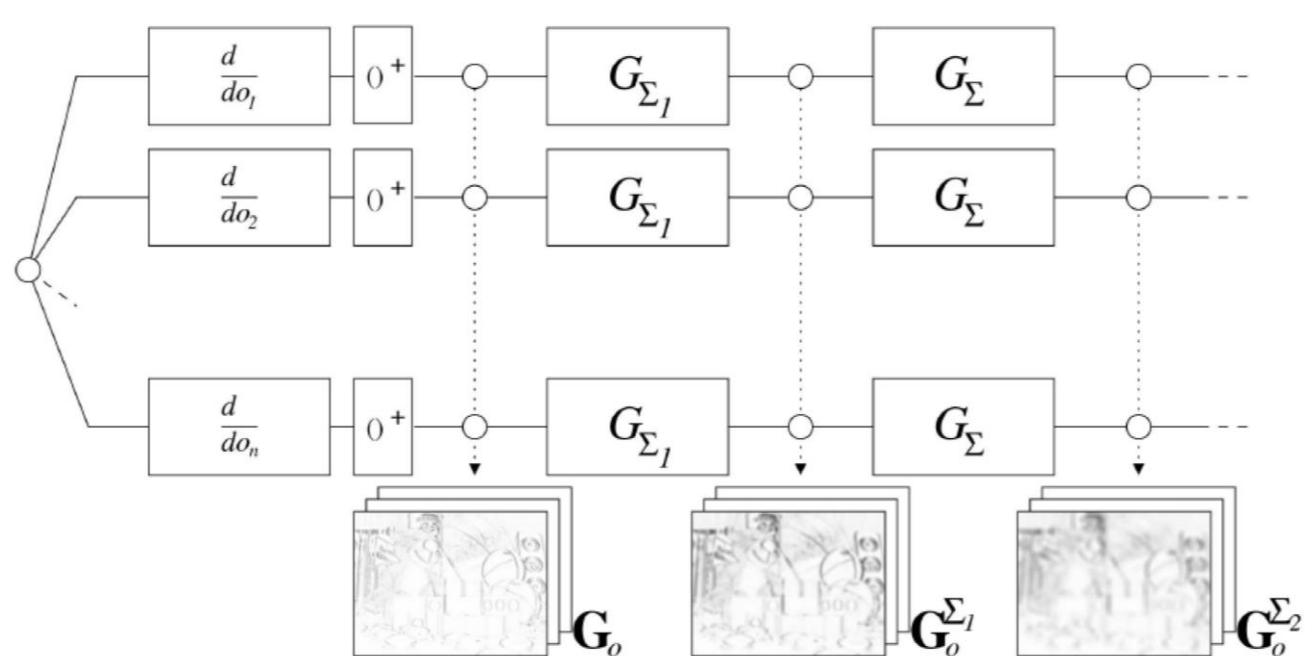
- E Tola, V Lepetit, P Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *TPAMI*, 2009.

Local Representations: DAISY

- The main difficulty is that the complex operations of SIFT in scale and orientation evaluation **cannot be performed directly for dense positions**, due to high complexity.
- Therefore, DAISY introduces a series of simplified designs for scale and orientation, but at the same time invariance is reduced.

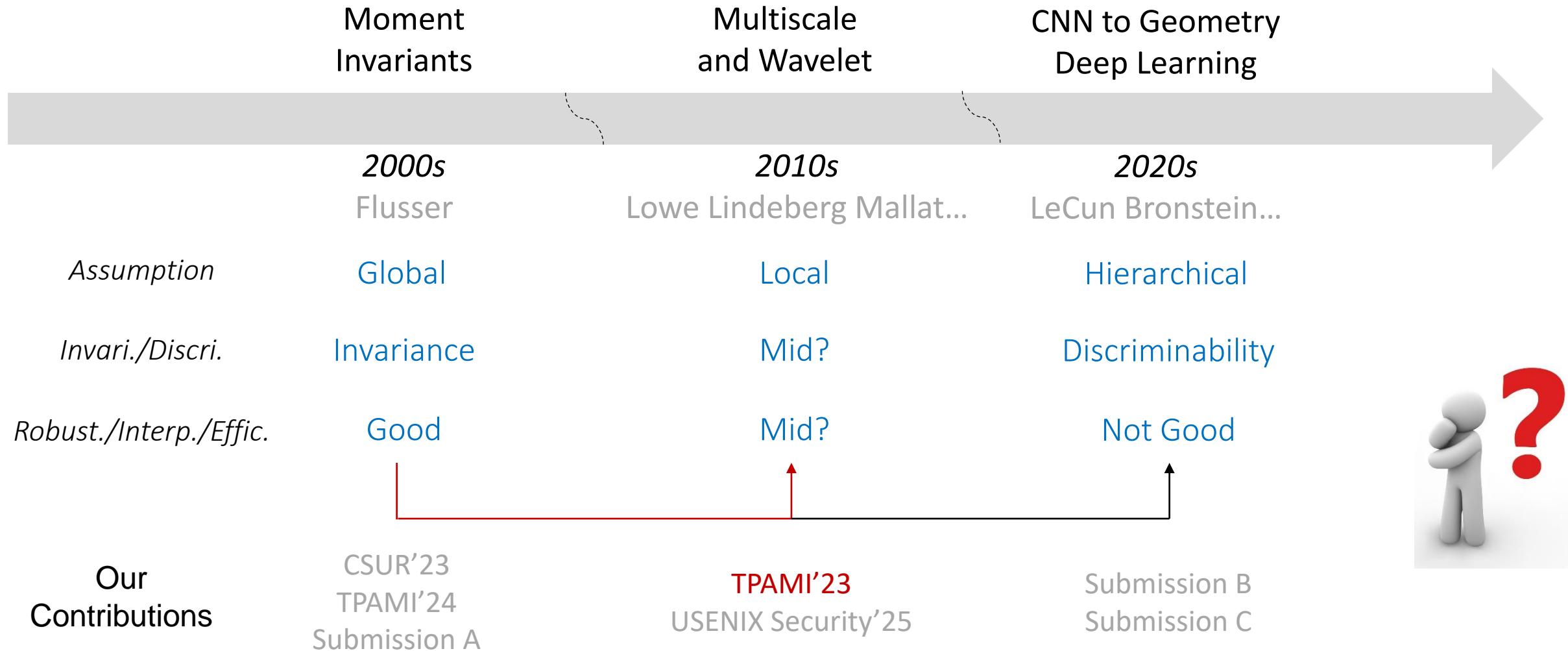


DAISY Descriptor



Simplified Designs for Scale and Orientation

Our Contributions

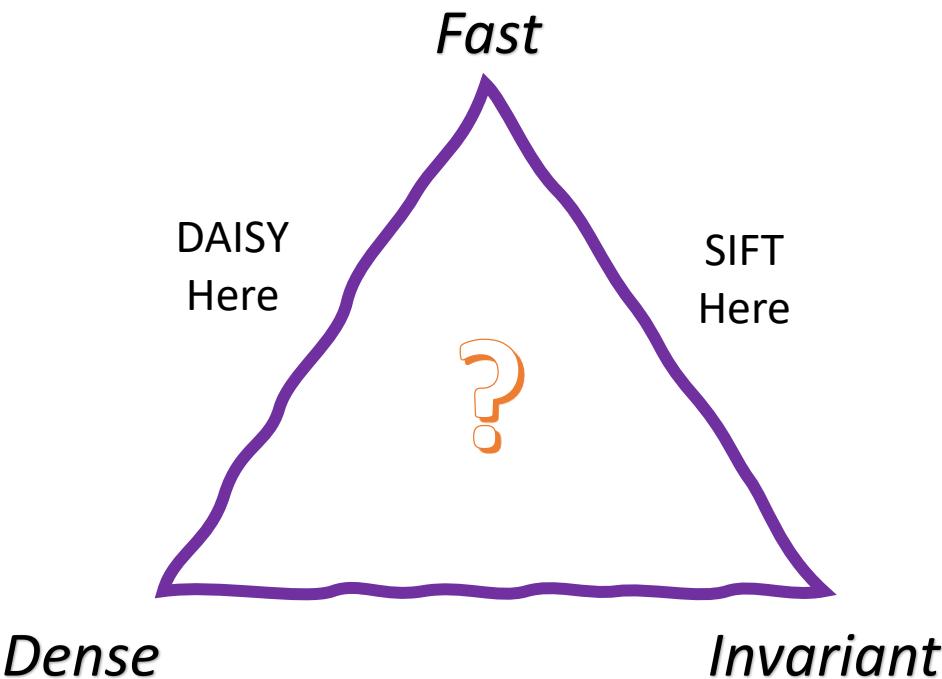


Designing Local Invariants

- Reviewing the above local invariants, one can note **a gap**: SIFT is fast and invariant, but not suitable for dense tasks; DAISY is fast and dense, but largely compresses invariance.
- We tried to define **truly dense invariants while being fast enough**. We achieved this goal by exploring the potential of **classical moment invariants**.



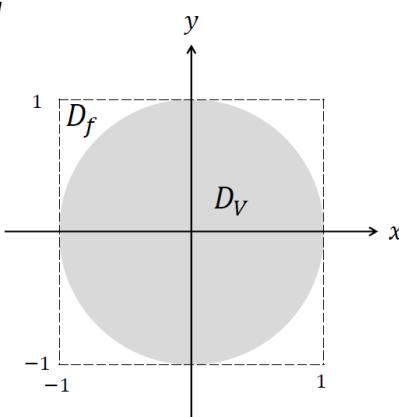
- S. Qi, Y. Zhang, C. Wang, et al. A Principled Design of Image Representation: Towards Forensic Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, 45(5): 5337 - 5354



Moments: From Global to Local

- First, we extend the definition of classical moments from the global to the local with scale space. Here, local coordinate system (x', y') is a translated and scaled version of the global coordinate system (x, y) , with translation offset (u, v) and scale factor w .
- Two interesting properties: generic nature and local representation capability.

Global



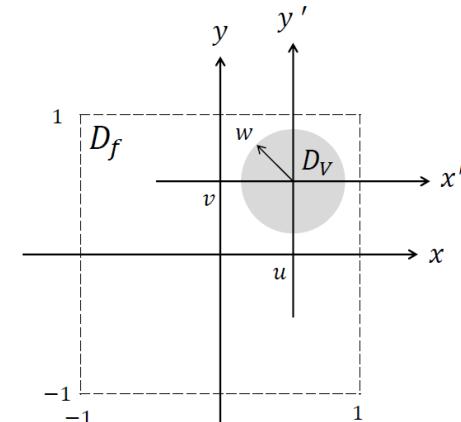
$$\langle f, V_{nm} \rangle = \iint_D R_n^*(r) A_m^*(\theta) f(r, \theta) r dr d\theta$$

Our Transformations

$$(x', y') = \frac{(x, y) - (u, v)}{w}$$

$$\left\{ \begin{array}{l} r' = \sqrt{(x')^2 + (y')^2} = \frac{1}{w} \sqrt{(x-u)^2 + (y-v)^2} \\ \theta' = \arctan\left(\frac{y'}{x'}\right) = \arctan\left(\frac{y-v}{x-u}\right) \end{array} \right.$$

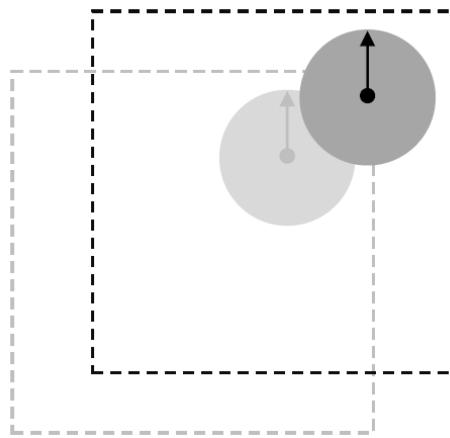
Local



$$\langle f, V_{nm}^{uvw} \rangle = \iint_D \underbrace{R_n^*\left(\frac{\sqrt{(x-u)^2 + (y-v)^2}}{w}\right)}_{(V_{nm}^{uvw}(x,y))^*} \underbrace{A_m^*\left(\arctan\left(\frac{y-v}{x-u}\right)\right)}_{\theta'} f(x, y) dx dy$$

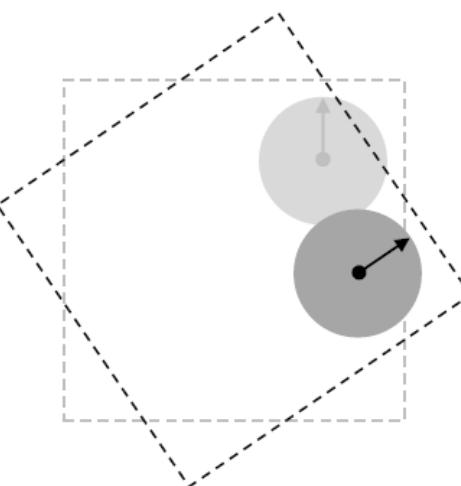
Moment Invariants: From Global to Local

- Then, we found the symmetry properties of the local definition for several geometric transformations.
- Therefore, rotation and flipping invariants can be obtained by taking the absolute values; translation and scaling invariants can be obtained by pooling over the $(u, v)/w$.



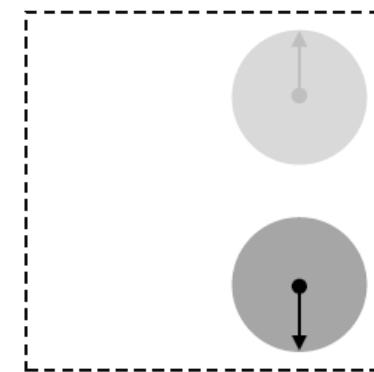
$$\begin{aligned} &\langle f(x + \Delta x, y + \Delta y), V_{nm}^{uvw}(x, y) \rangle \\ &= \langle f(x, y), V_{nm}^{(u+\Delta x)(v+\Delta y)w}(x, y) \rangle \end{aligned}$$

Translation Equivariance
w.r.t. (u, v)



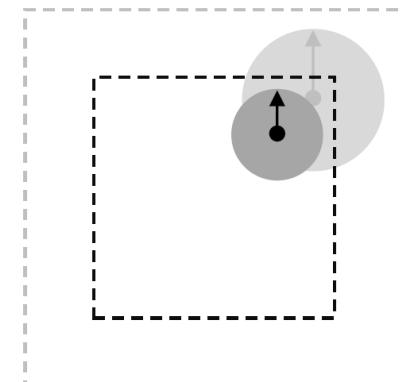
$$\begin{aligned} &\langle f(r, \theta + \phi), V_{nm}^{uvw}(r', \theta') \rangle \\ &= \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle A_m^*(-\phi) \end{aligned}$$

Rotation Invariance
w.r.t. absolute values



$$\begin{aligned} &\langle f(r, -\theta), V_{nm}^{uvw}(r', \theta') \rangle \\ &= (\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle)^* \end{aligned}$$

Flipping Invariance
w.r.t. absolute values

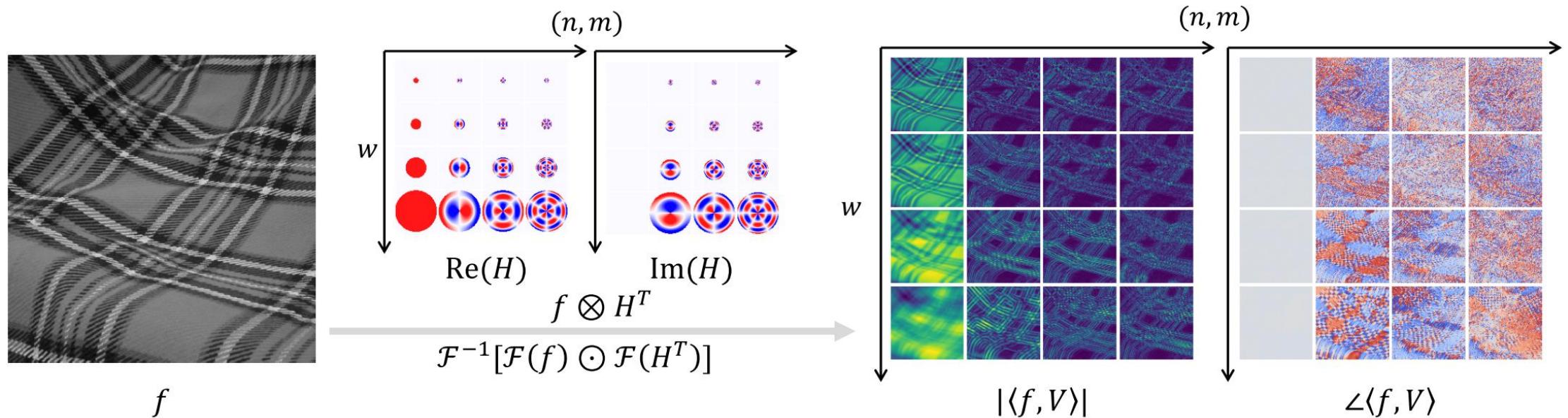


$$\begin{aligned} &\langle f(sx, sy), V_{nm}^{uvw}(x, y) \rangle \\ &= \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle \end{aligned}$$

Scaling Covariance
w.r.t. w

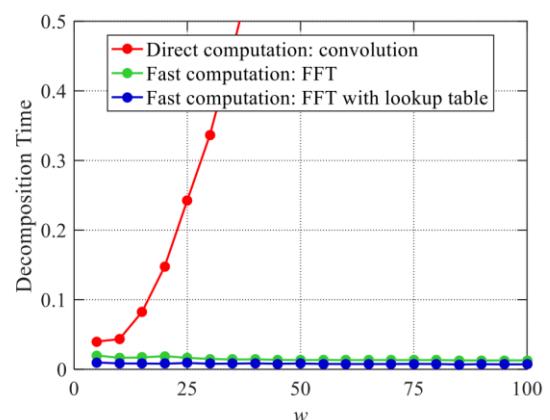
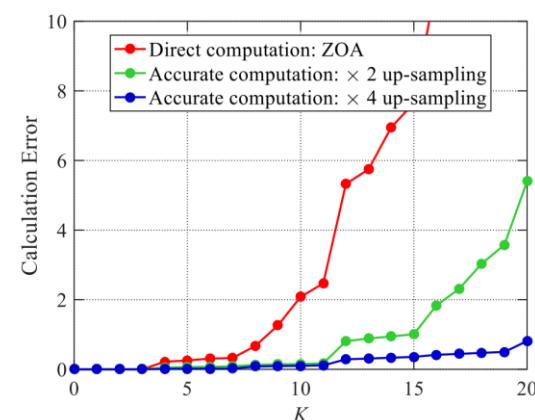
Fast Implementation

- Finally, we give a fast implementation by the **convolution theorem**.



$$\mathcal{O}(w_{\max}^2 \#_{uv} \#_w) \quad \text{VS} \quad \mathcal{O}(\#_w \#_{uv} \log \#_{uv})$$

w_{\max}^2 **VS** $\log(\#_{uv})$



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)*

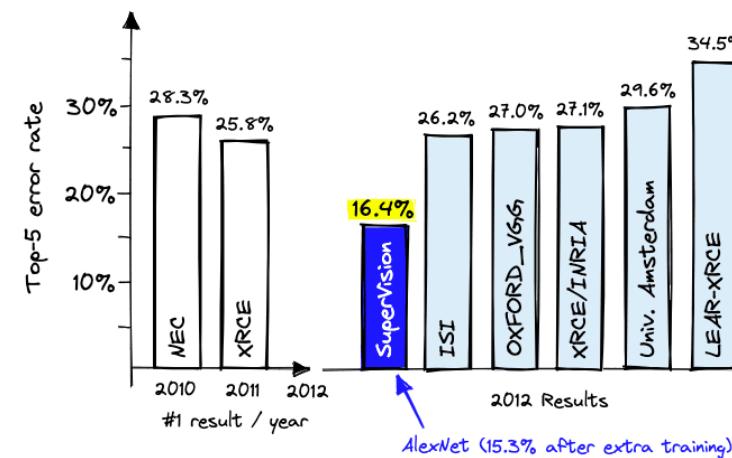
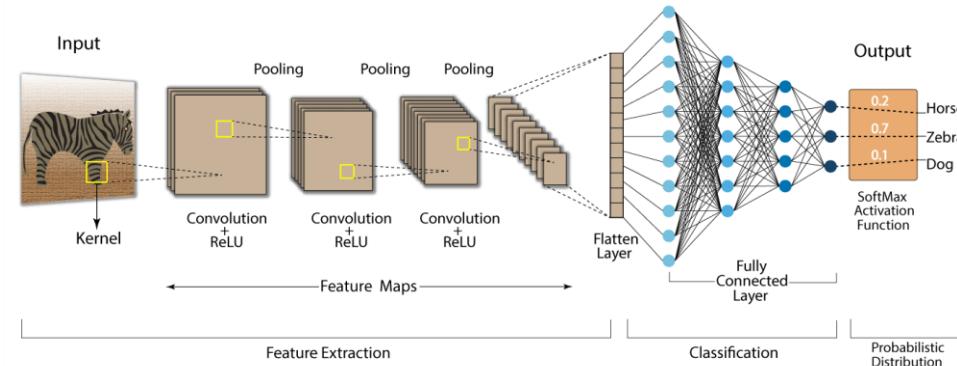


A Historical Perspective of Data Representation
Rethinking Deep Learning with Invariance: The Good, The Bad, and The Ugly

From Knowledge Driven To Data Driven

Invariance in The Early Era of Deep Learning

- Knowledge Driven: Despite decades of research, these hand-crafted representations still **fail to provide sufficient discriminability for large-scale tasks**, especially in the discrimination of real-world semantic content.
- Data Driven: As we enter the early era of deep learning, **convolutional neural networks** achieve strong discriminative power for large-scale tasks, known as *ImageNet moment*.



*The Success of Deep Learning
AlexNet Wins ILSVRC 2012 Competition*

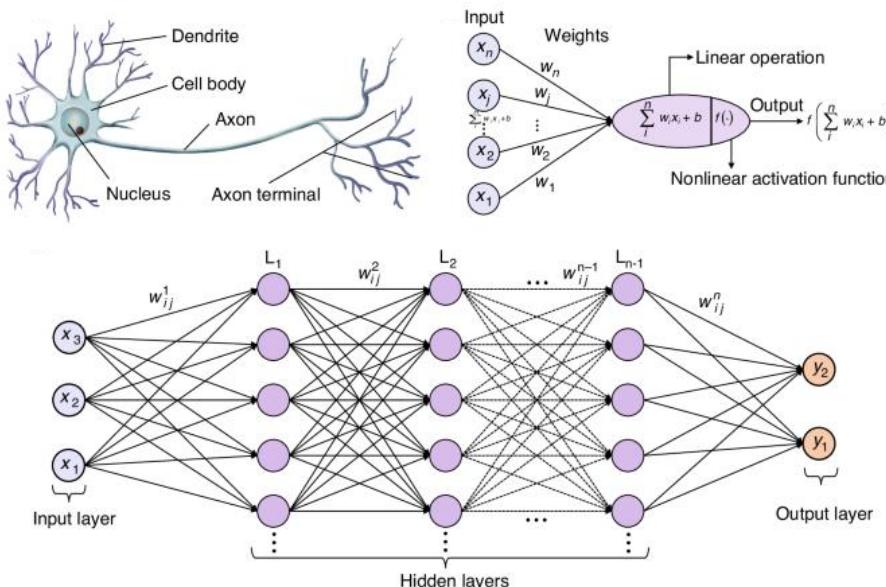
A. Krizhevsky, 2012
AlexNet

- A. Krizhevsky, I. Sutskever, G.E. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*, 2012.



A Huge Span of Time

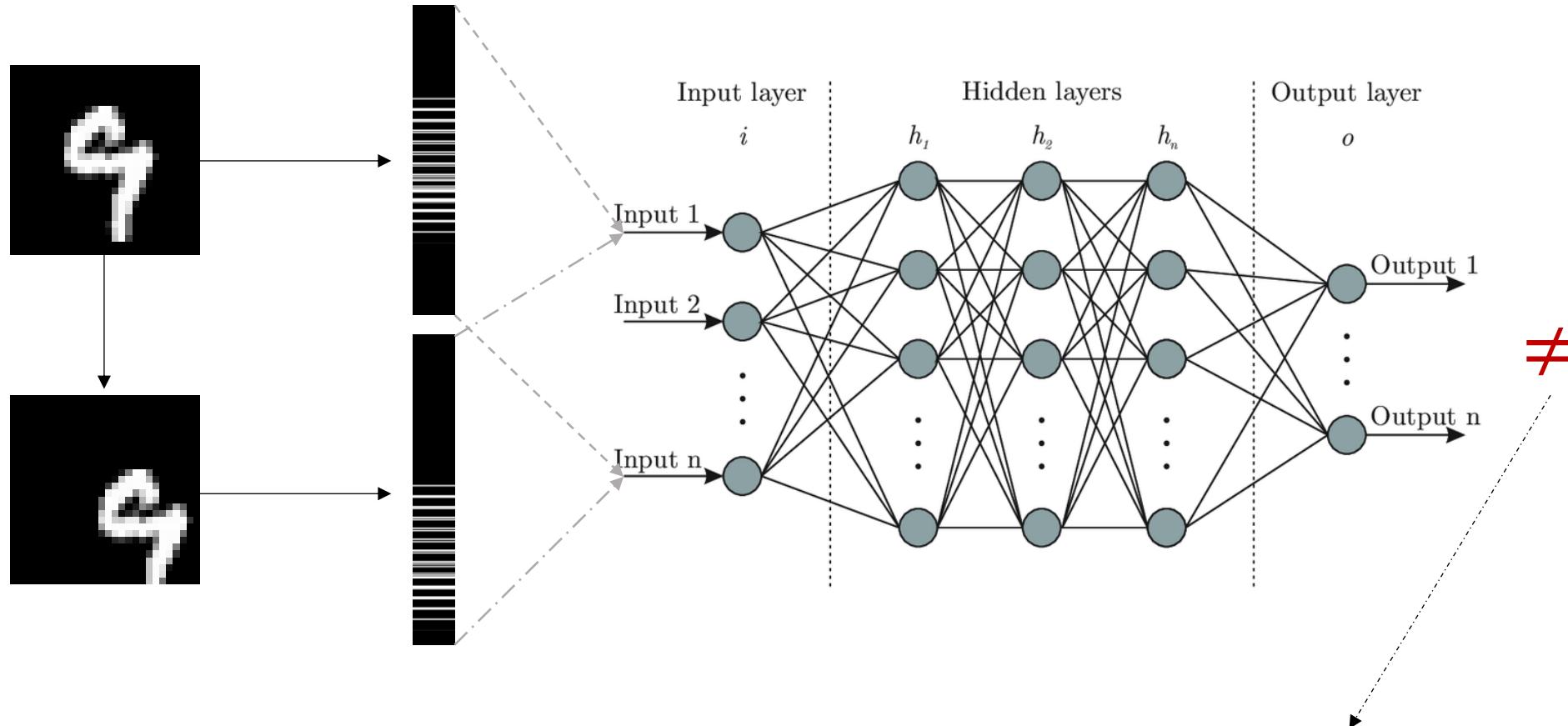
- Neural networks were proposed quite early, dating back to 1950s for the perceptron; but it was not until AlexNet in 2012 that the remarkable achievement was realized.
- What is the missing key?



F. Rosenblatt, 1958
Perceptron

- F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958.

Translations on Neural Networks



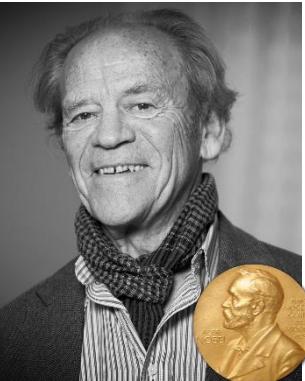
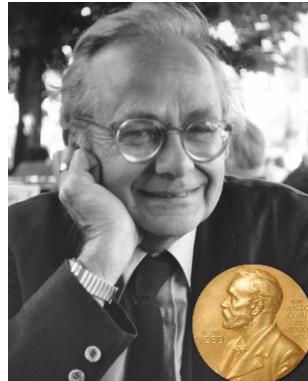
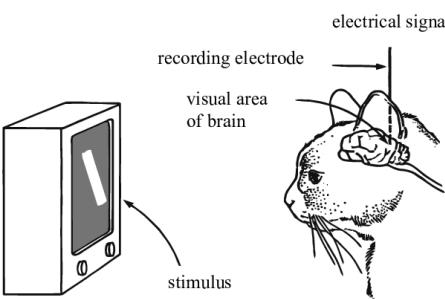
"The response of [Perceptrons] was severely affected by the shift in position [...] of the input patterns. Hence, their ability for pattern recognition was not so high." — Fukushima

- K Fukushima, S Miyake. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, 1982.

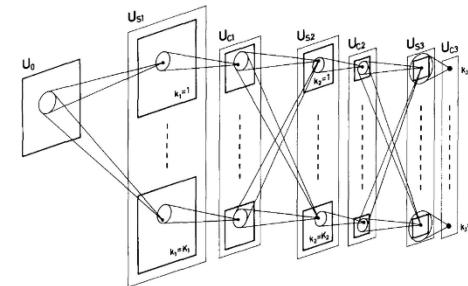
Convolution And Translation Equivariance

Convolutional Neural Networks

- Convolution with its Translation Equivariance (see also Wavelet Transform) are key to enabling neural networks successful in visual tasks.
 - First, local structures was discovered in the biological vision by Hubel and Wiesel.
 - Then, convolution was introduced into neural networks by Fukushima.
 - Finally, such networks were equipped with learnability and backpropagation by LeCun.
- Invariance still plays an important role, even in the rise of the learning paradigm.



D. Hubel & T. Wiesel, 1959
Structure of Visual Cortex



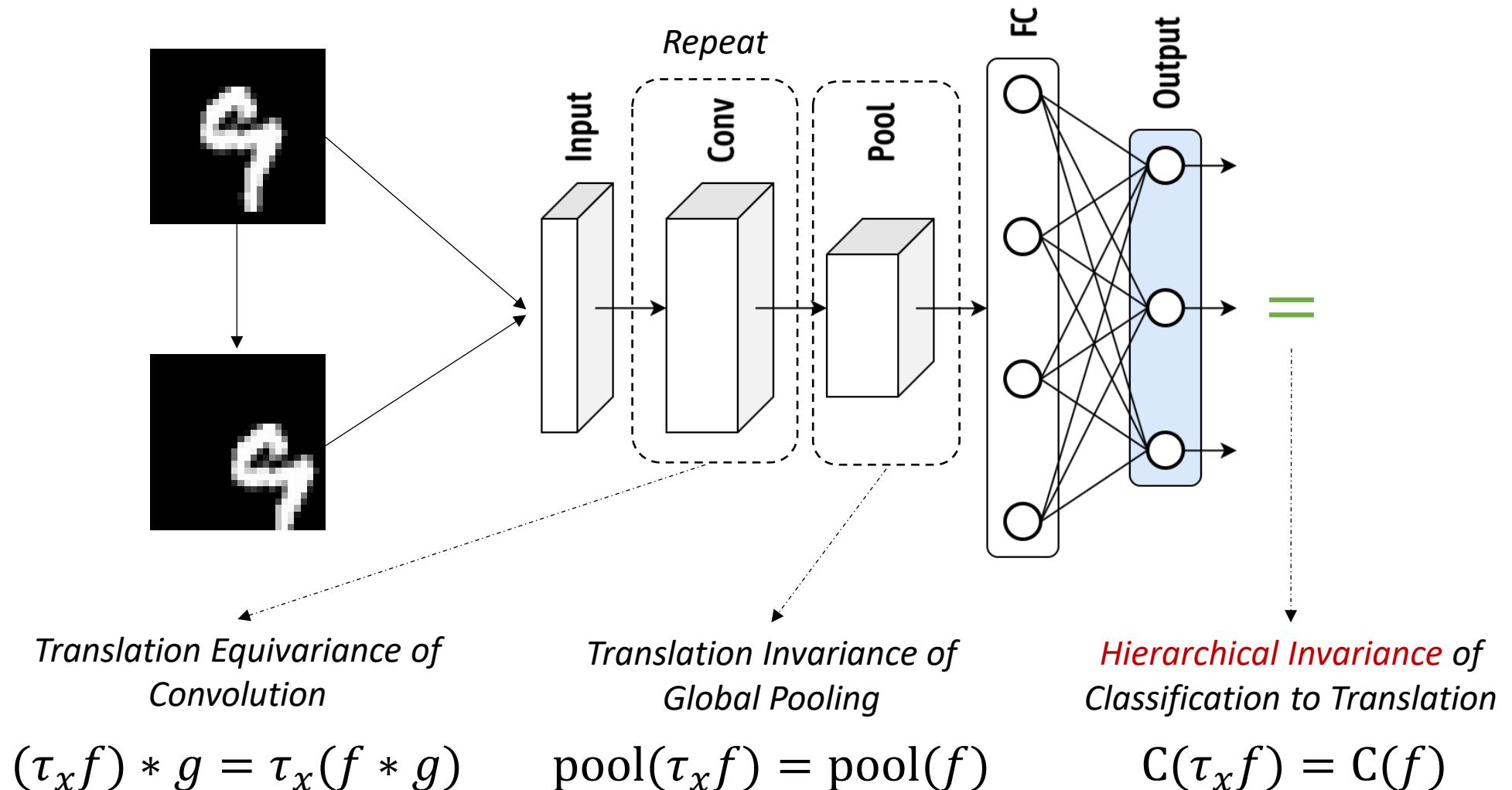
K. Fukushima, 1982
Neocognitron



Y. LeCun, 1989
LeNet

- Y LeCun, B Boser, J Denker, et al. Handwritten digit recognition with a back-propagation network. *NIPS*, 1989.

Translation Equi/In-variance of Convolutional Neural Networks



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



Time for
a Break!!

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)*

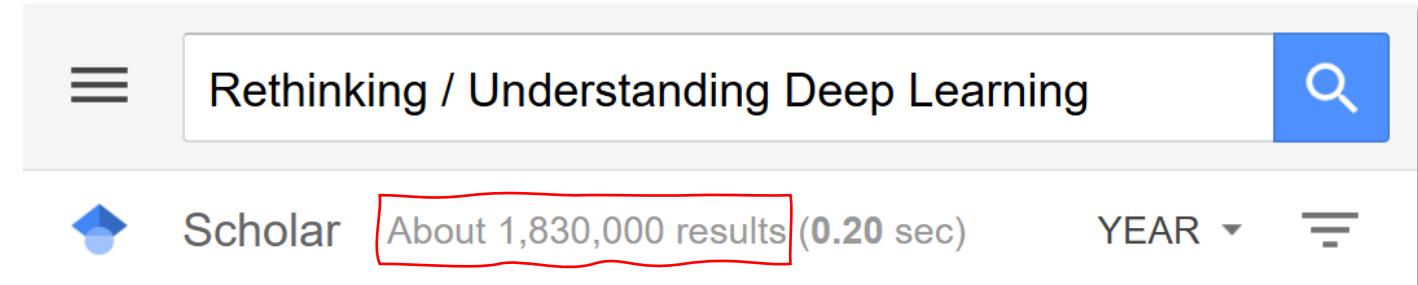
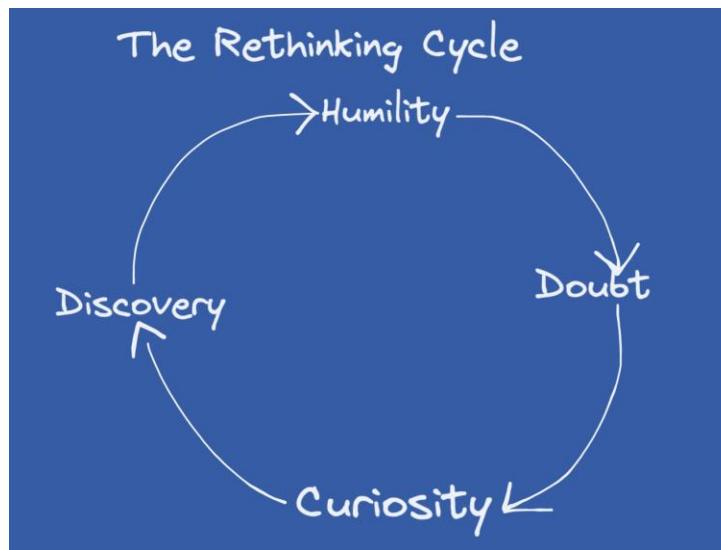


A Historical Perspective of Data Representation
Rethinking Deep Learning with Invariance: The Good, The Bad, and The Ugly

Why Rethink Deep Learning? Realistic Needs And Theoretical Extensions

Invariance in The Era of Rethinking Deep Learning

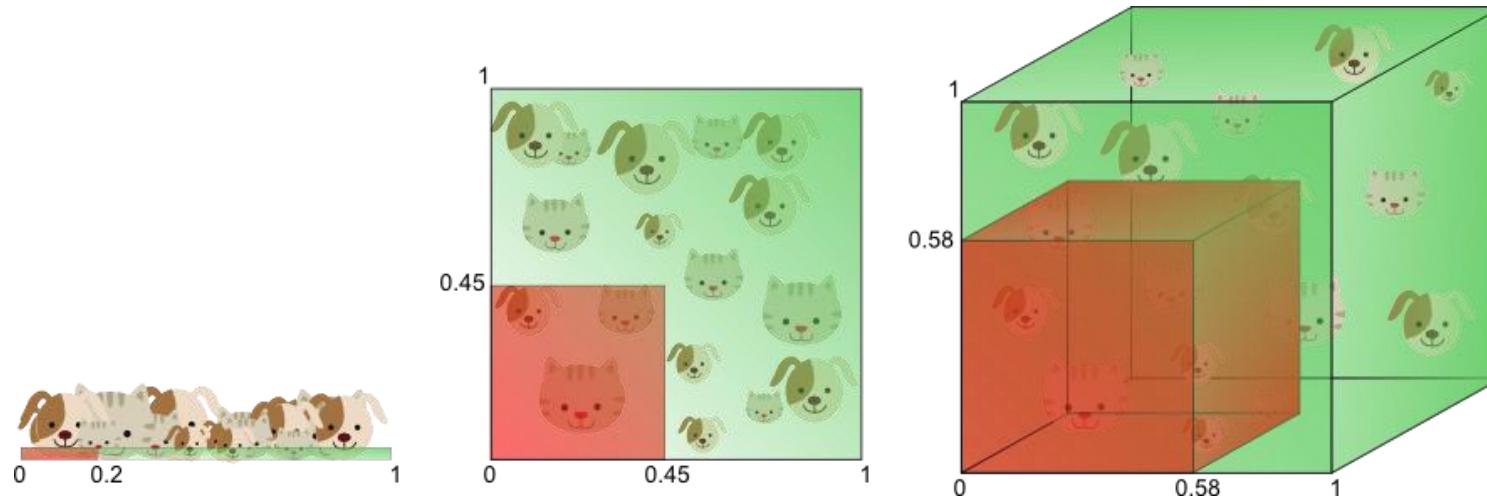
- Realistic Needs: Bottlenecks in **robustness, interpretability, and efficiency**.
- Theoretical Extensions: A unified theory perspective to avoid getting into **endless experimental designs**.



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 and the system, reducing the need for unnecessary and impractical learning, just like from NN to CNN.

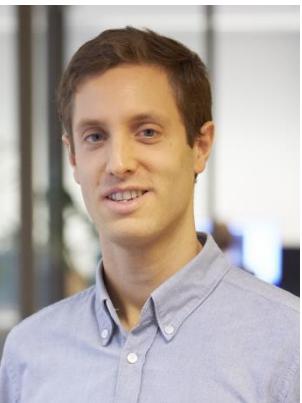


Geometry Deep Learning

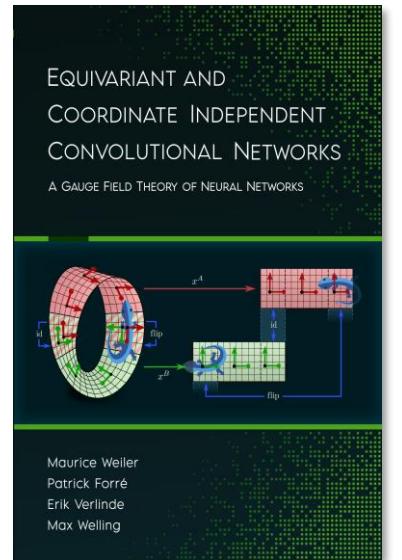
Rethinking From The Lens Of Invariance

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

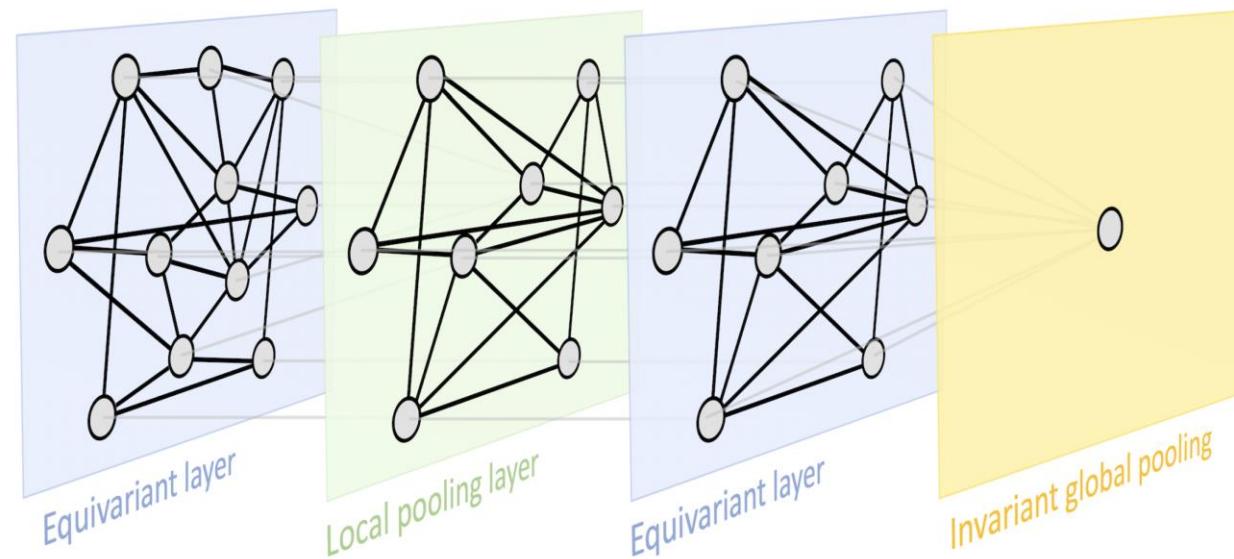


M. M. Bronstein, J. Bruna, T. Cohen, & P. Veličković, 2017
Geometry Deep Learning



Blueprint: High-level Intuition

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



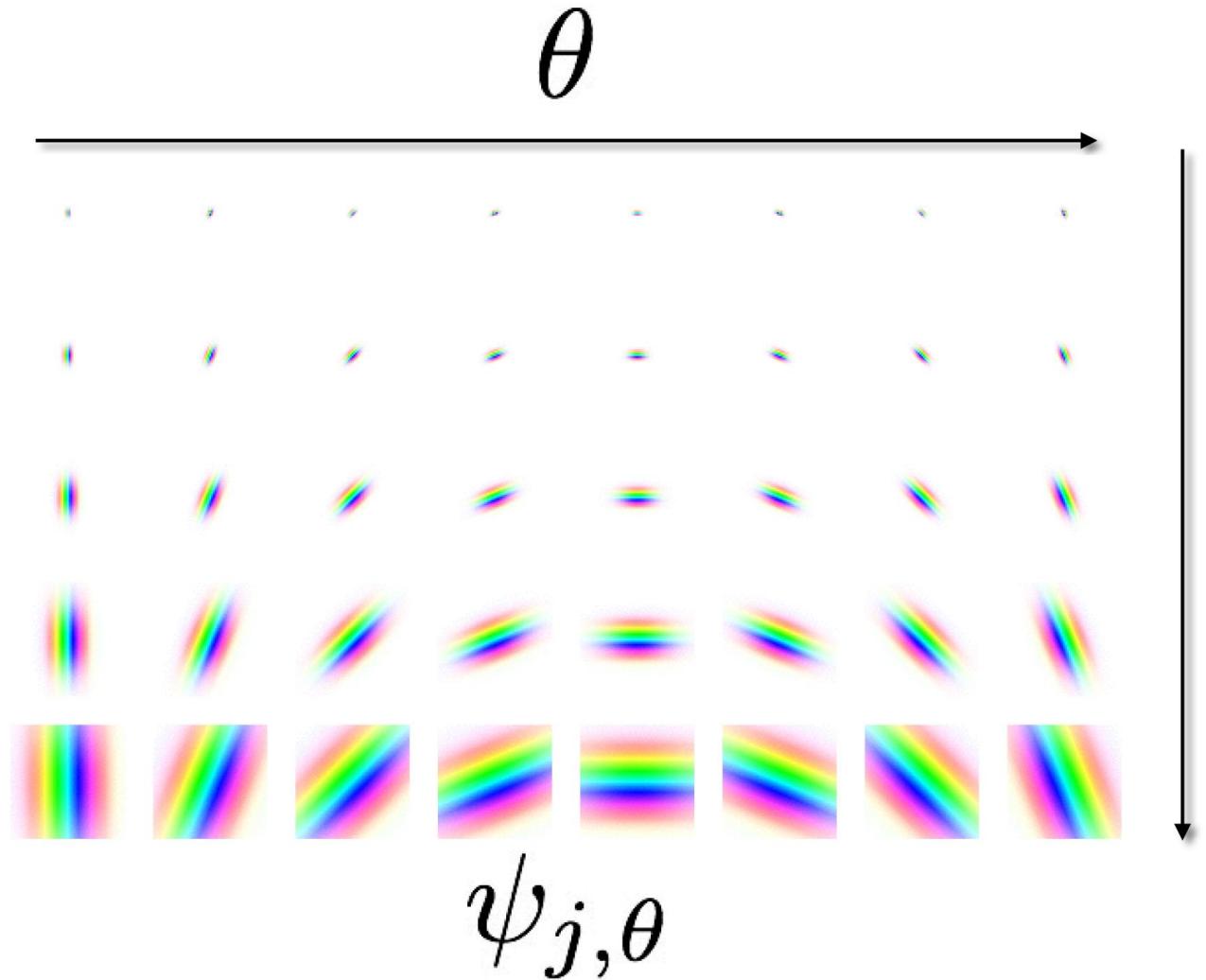
Blueprint: Formalization

- 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 $\text{Iso}(\Omega)$ / 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

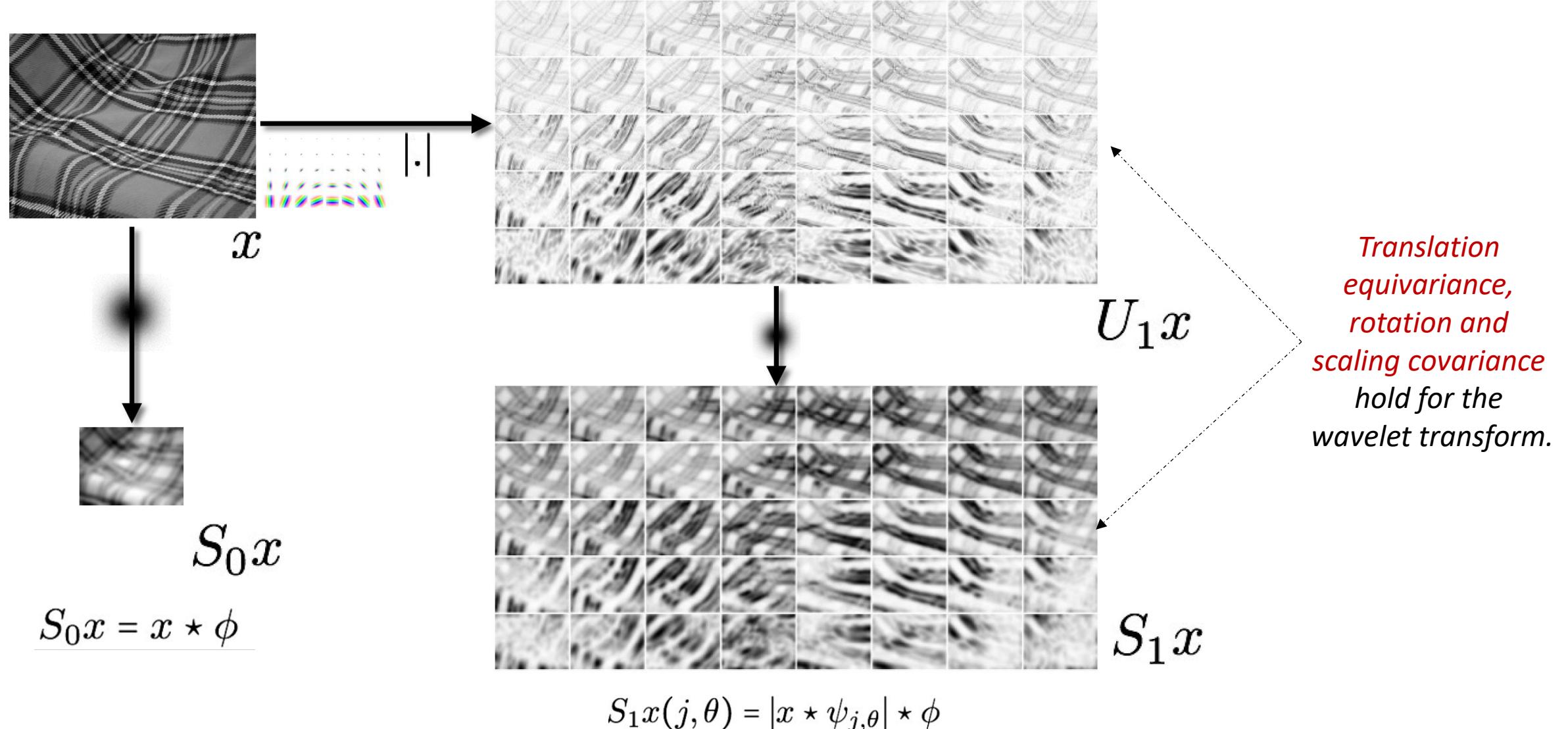
Beyond Translations: Wavelet Scattering Networks



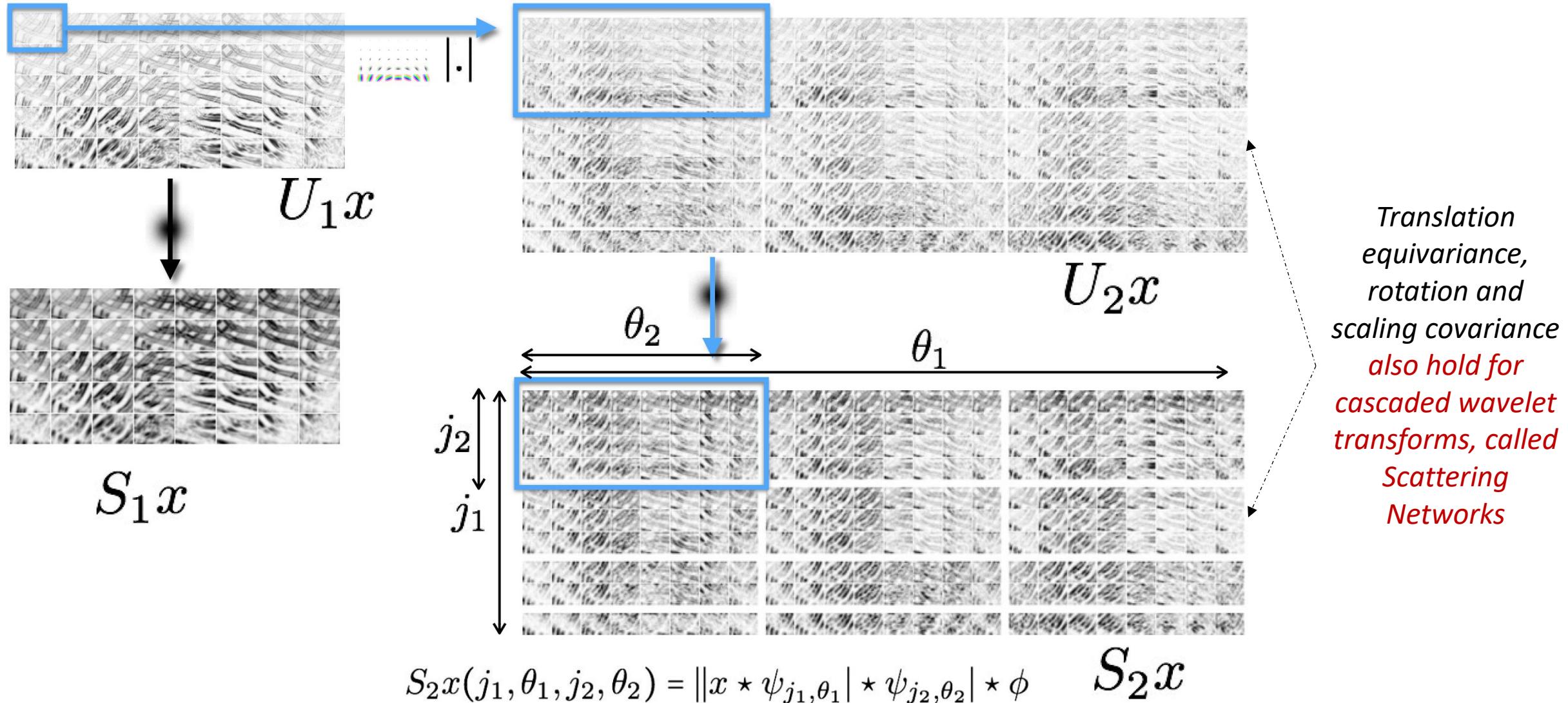
*By fully considering **the translation, rotation and scaling symmetry group**, the wavelet basis functions can be well designed to achieve invariance.*

- 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.

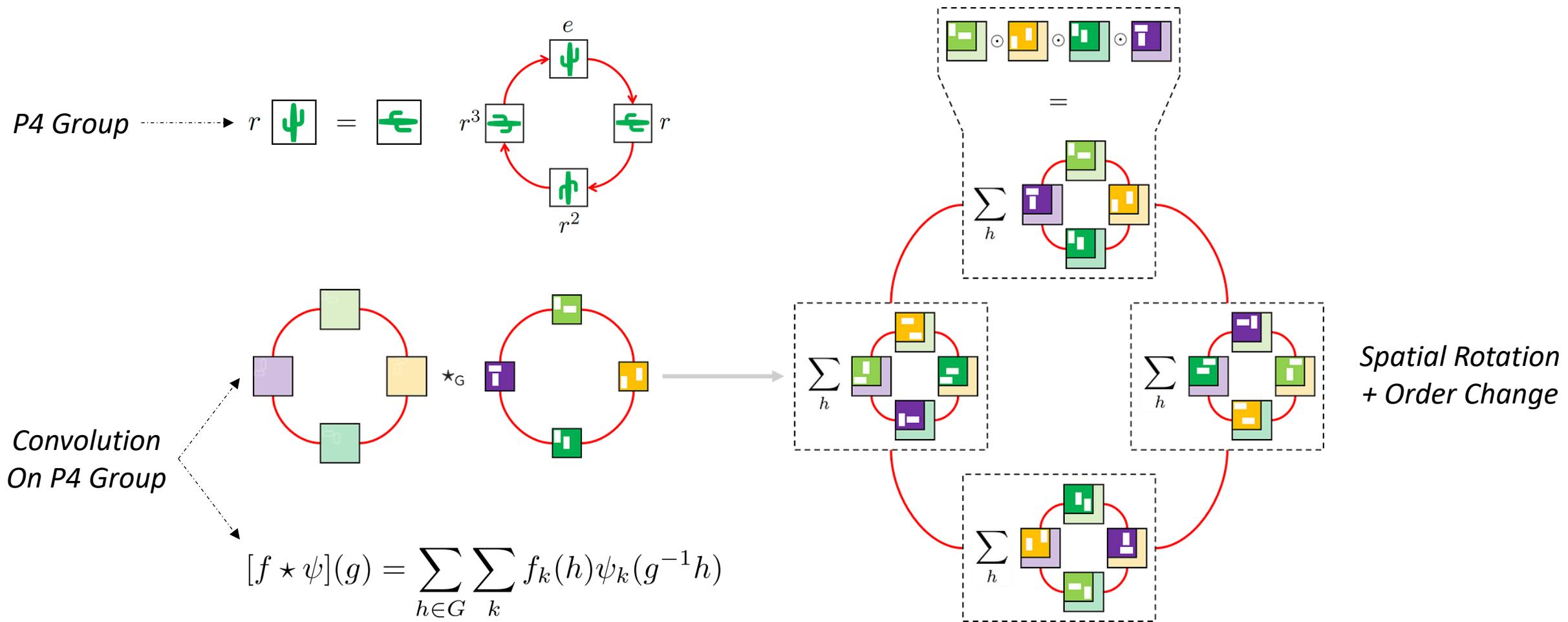
Beyond Translations: Wavelet Scattering Networks



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

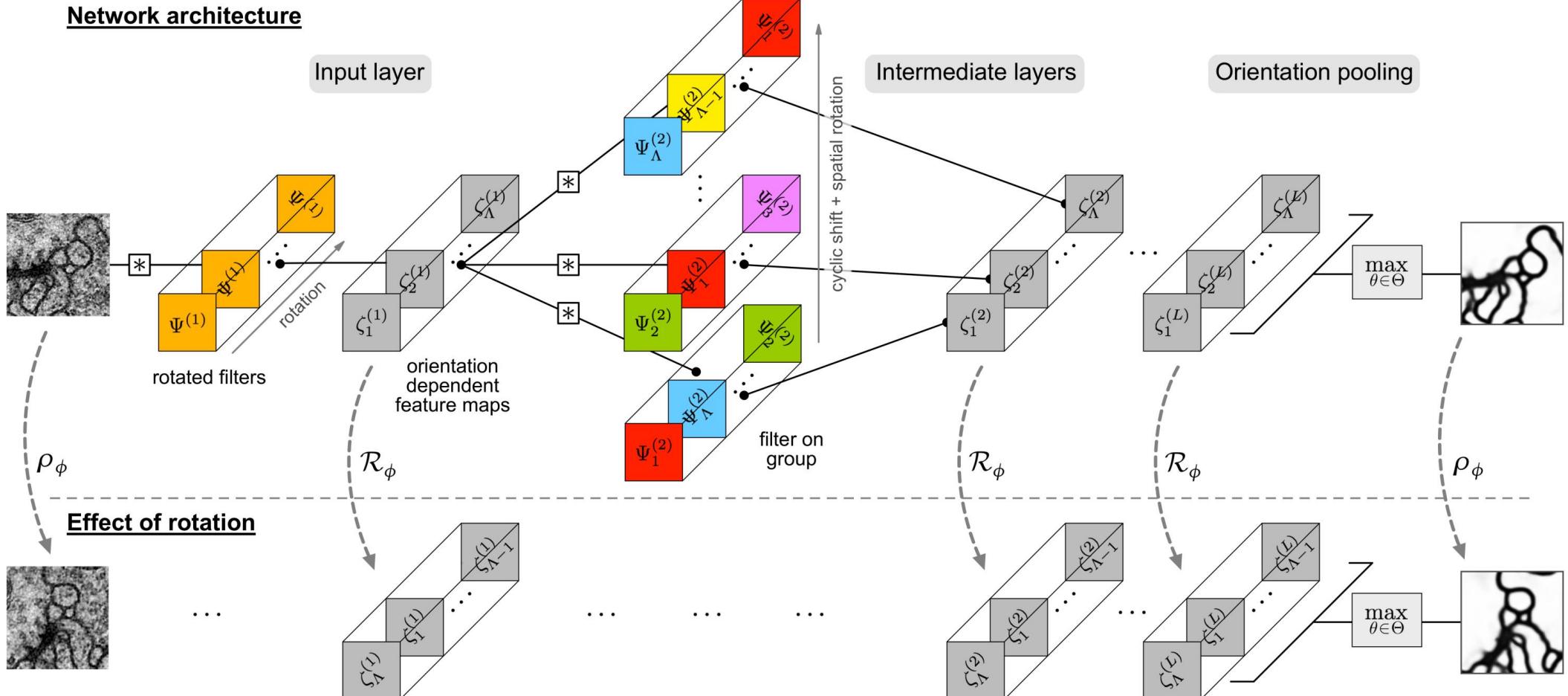
$$[f \star \psi](g) = \sum_{\substack{h \in G \\ k}} f_k(h) \psi_k(g^{-1}h)$$

Group Convolution

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

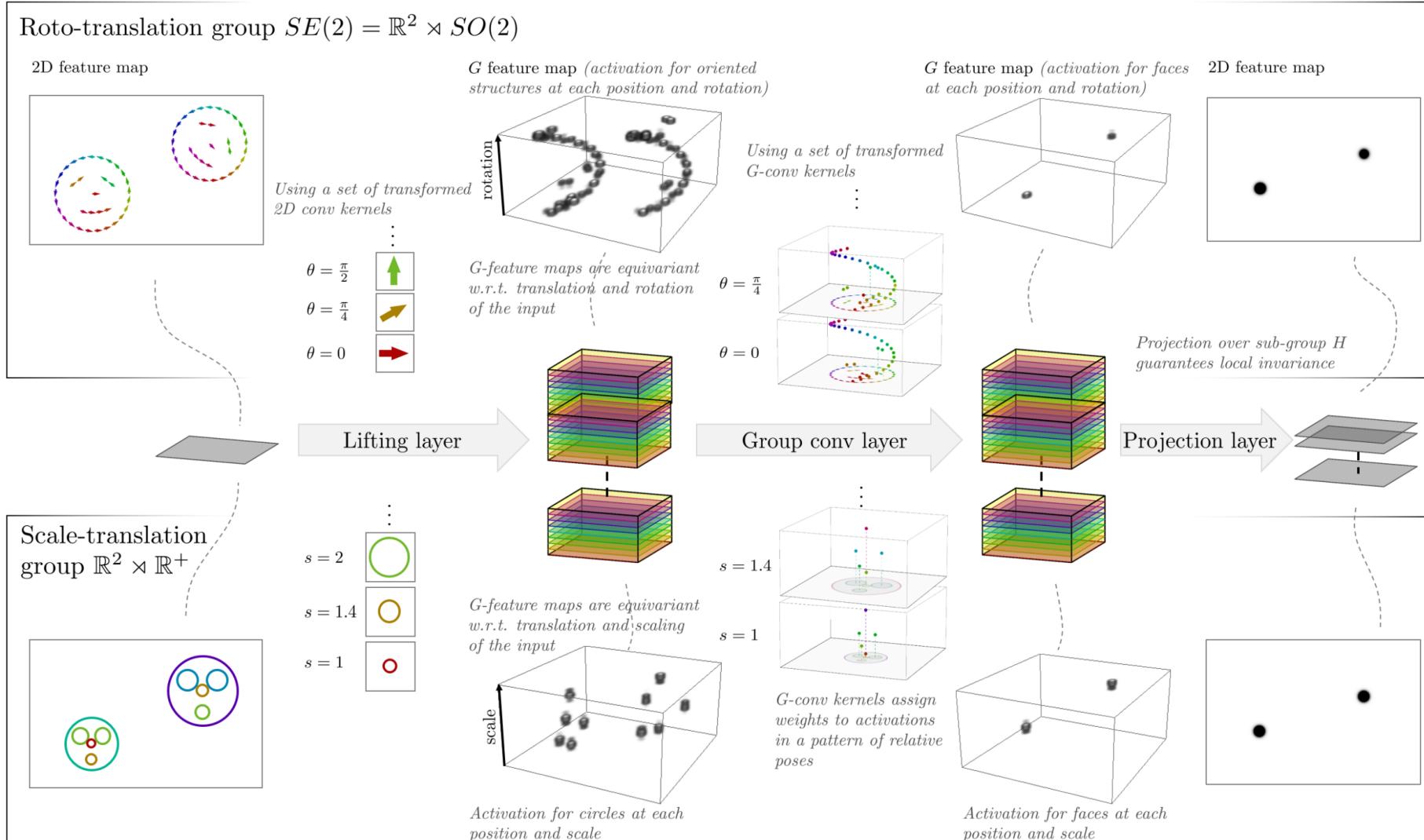
Group Equivariance

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

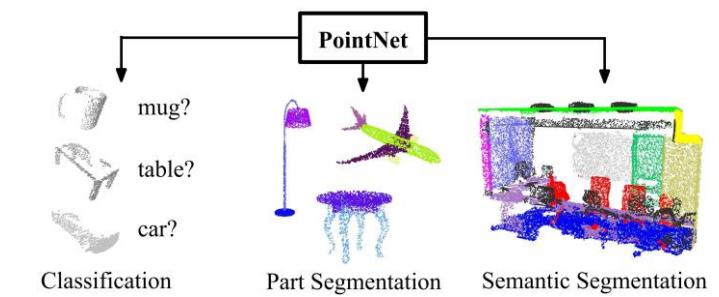
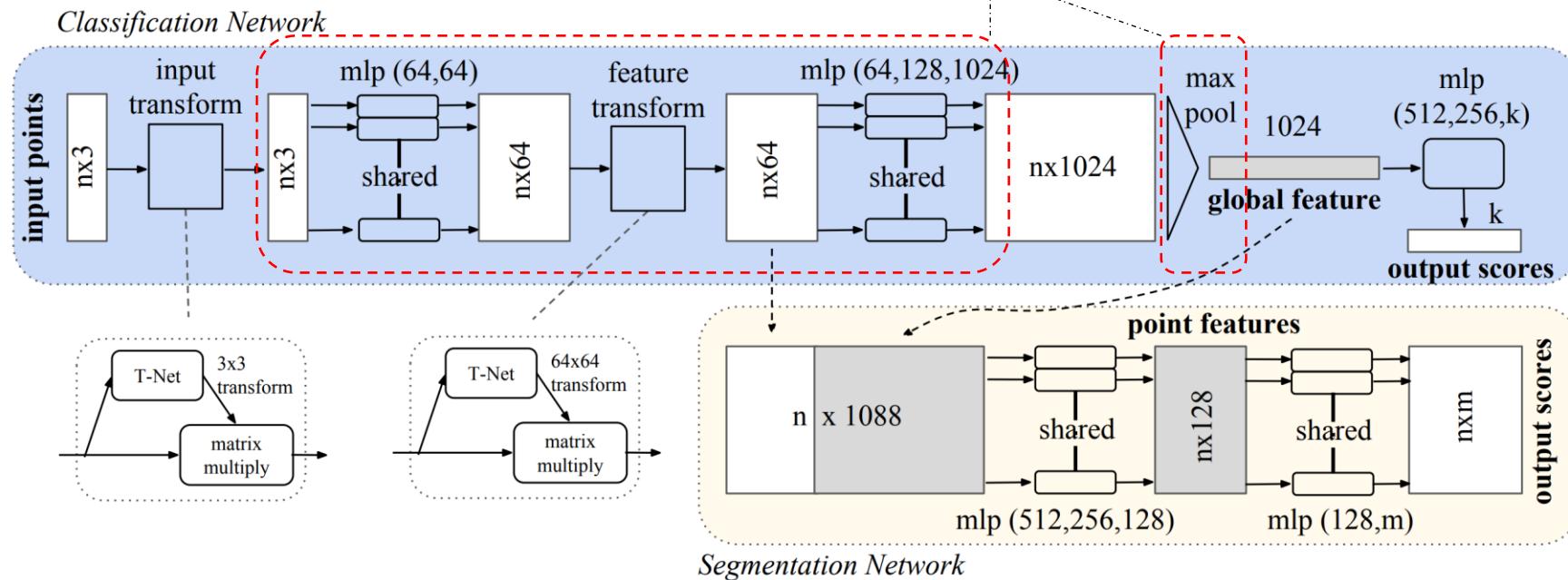
Beyond Images: Deep Sets and PointNet

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

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

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

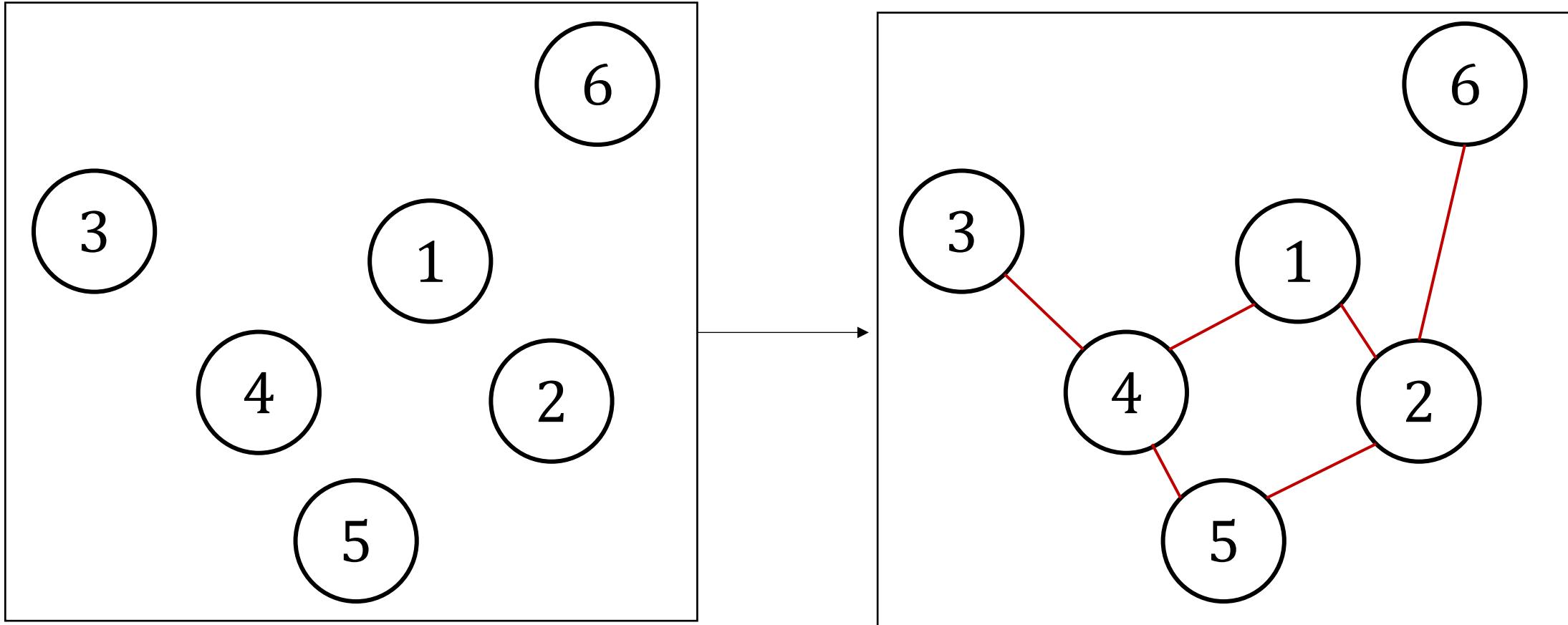
Deep Sets, Point-wise ϕ for Permutation Equivariance and Global Pooling Σ for Permutation Invariance



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.

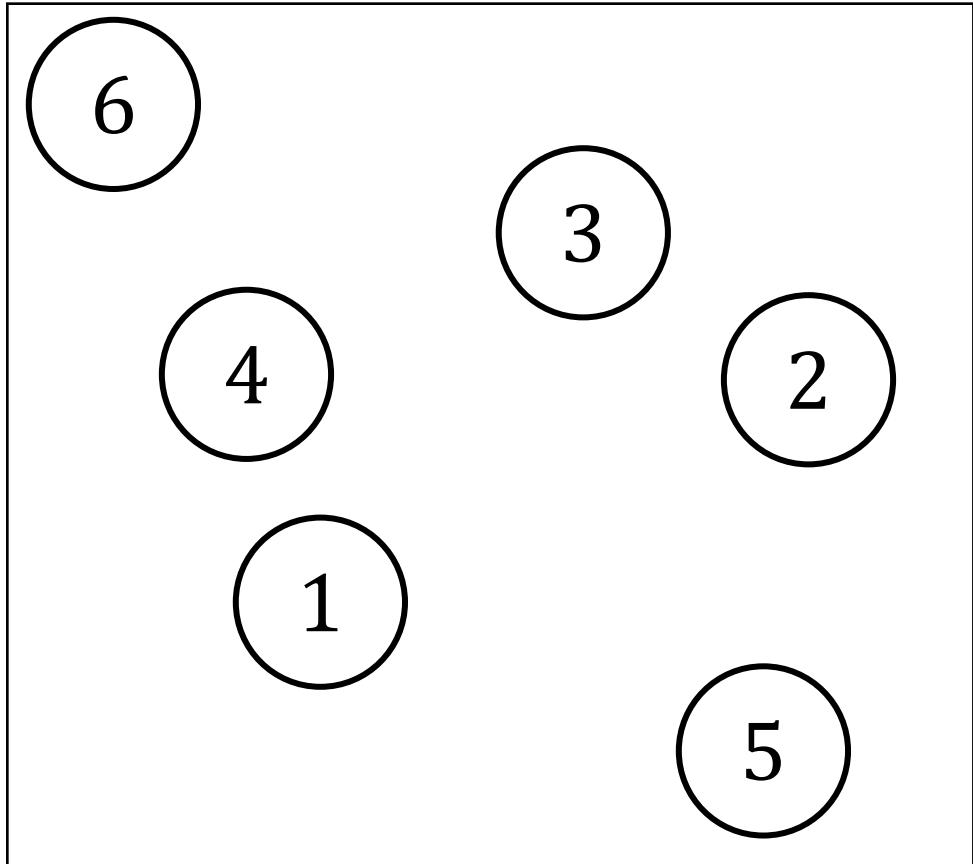
Beyond Images: Graph Networks



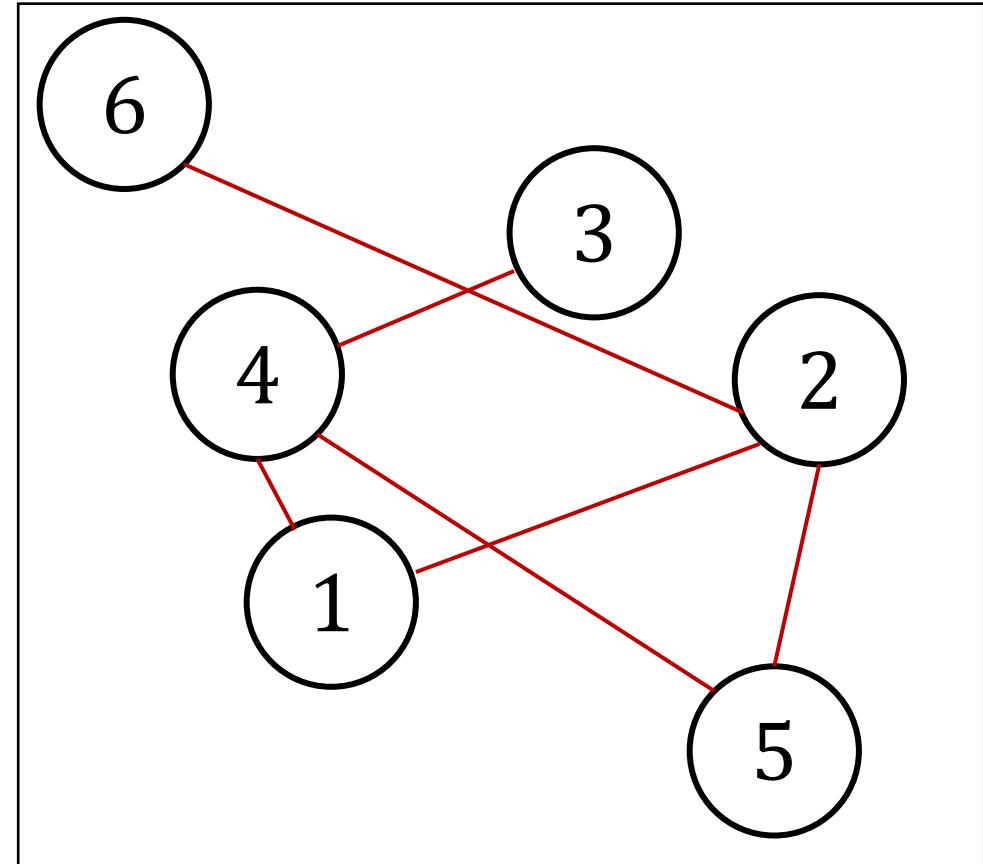
Set: $S = \{1, 2, \dots, 6\}$

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

Beyond Images: Graph Networks

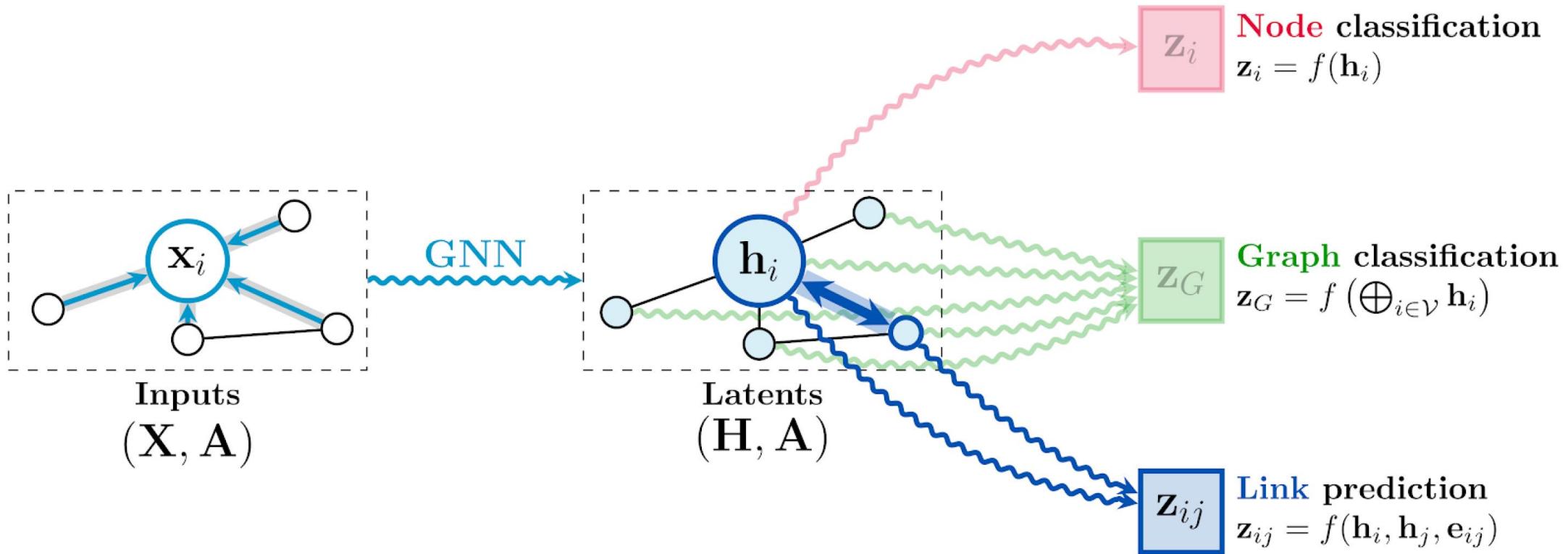


Set Permutation
Just order changes



Graph Permutation
Node order changes with edge order changes

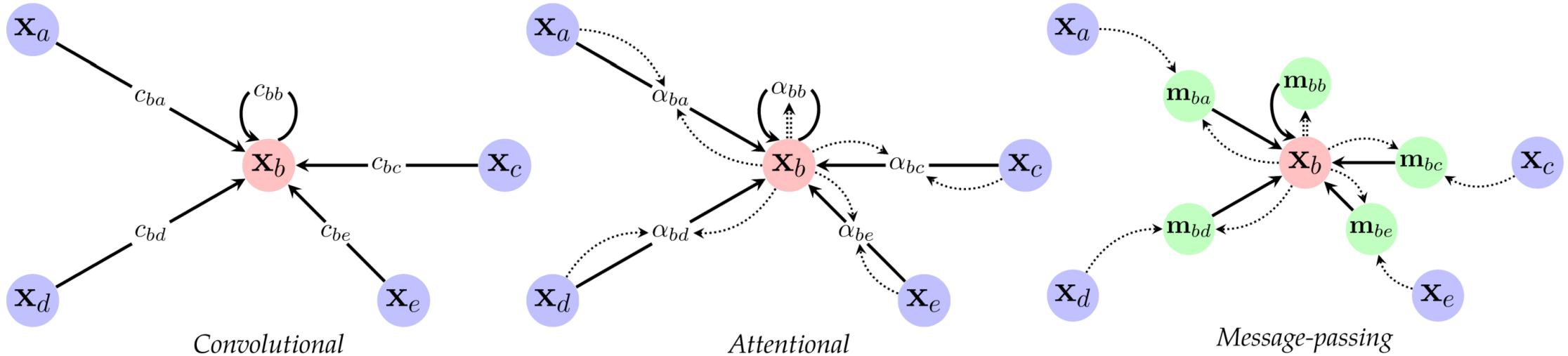
Beyond Images: Graph Networks



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



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

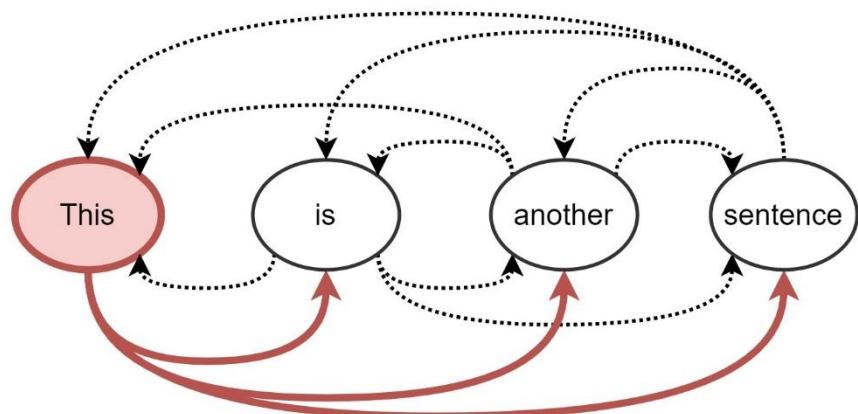
$$\mathbf{h}_i = \phi \left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j) \psi(\mathbf{x}_j) \right)$$

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

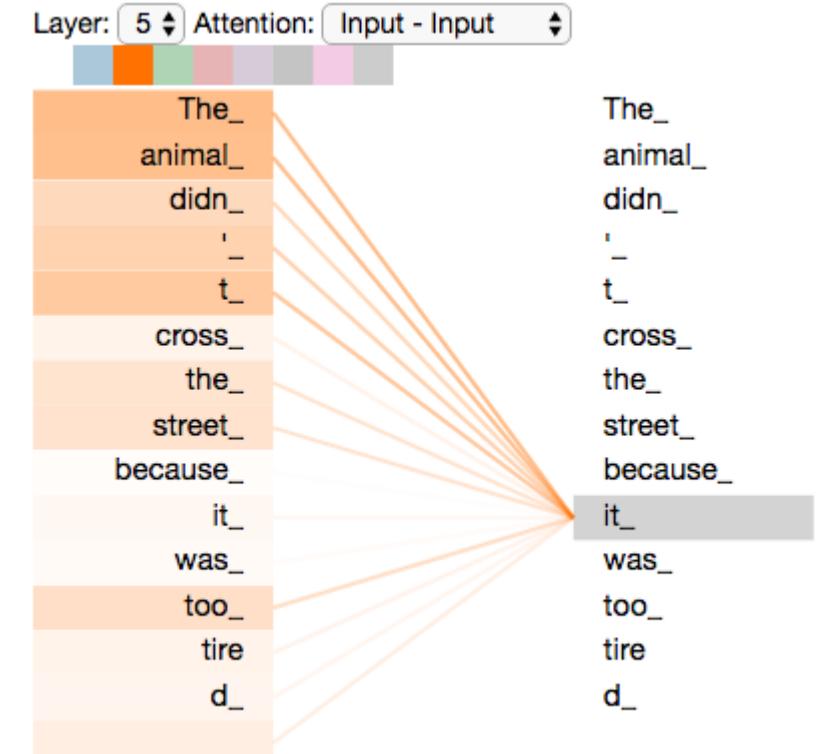
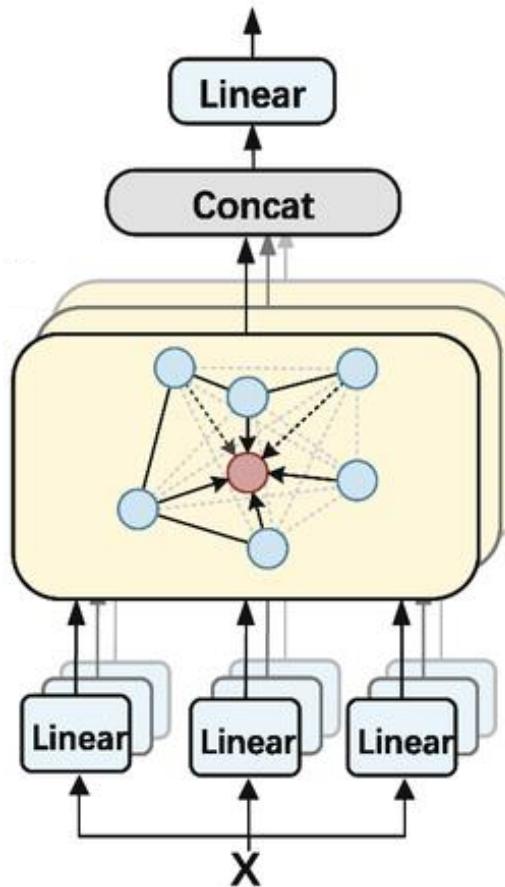
Permutation Equivariant GNN Layers

- 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

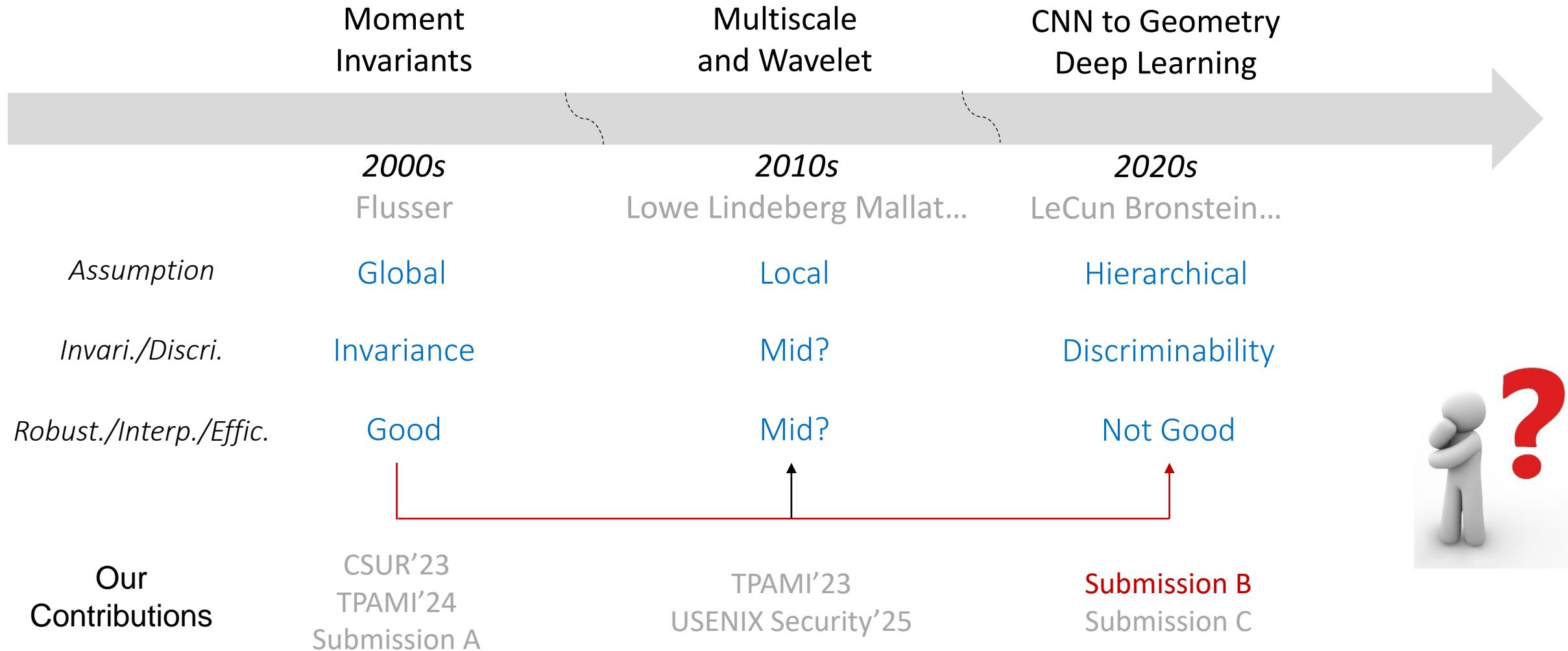


$$\mathbf{h}_u = \phi \left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{V}} a(\mathbf{x}_u, \mathbf{x}_v) \psi(\mathbf{x}_v) \right)$$



Transformers are GNNs on Fully-connected Graph

Our Contributions



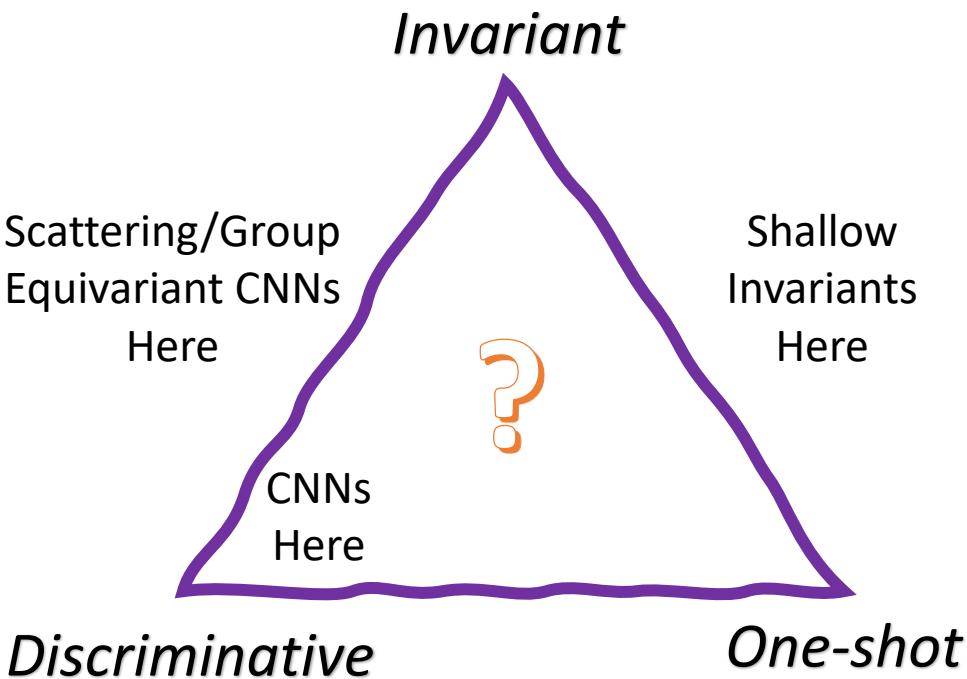
From Global And Local To Hierarchical

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**.

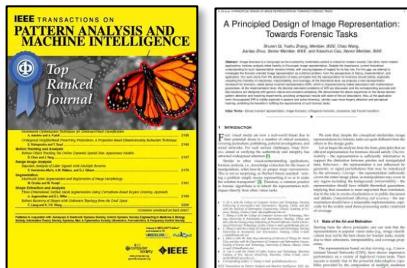
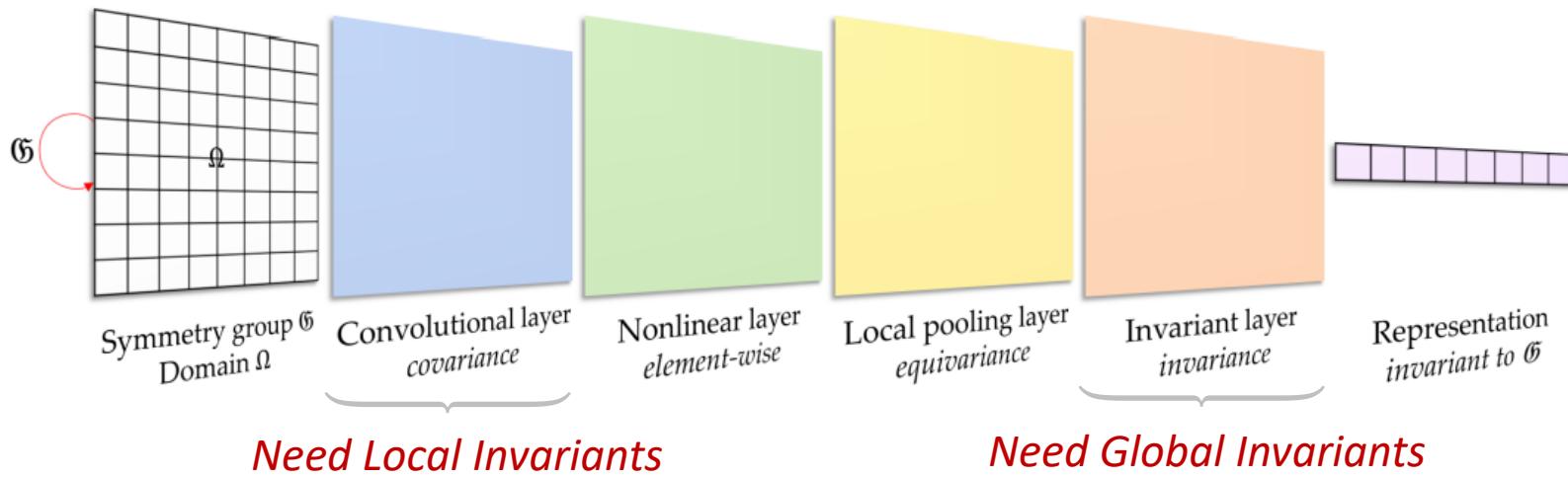


- S. Qi, Y. Zhang, C. Wang, et al. Hierarchical Invariance for Robust and Interpretable Vision Tasks at Larger Scales. *A Preprint/Submission*, 2025.



Blueprint

- First, we rethink the typical modules of CNN, unifying the fundamental 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.

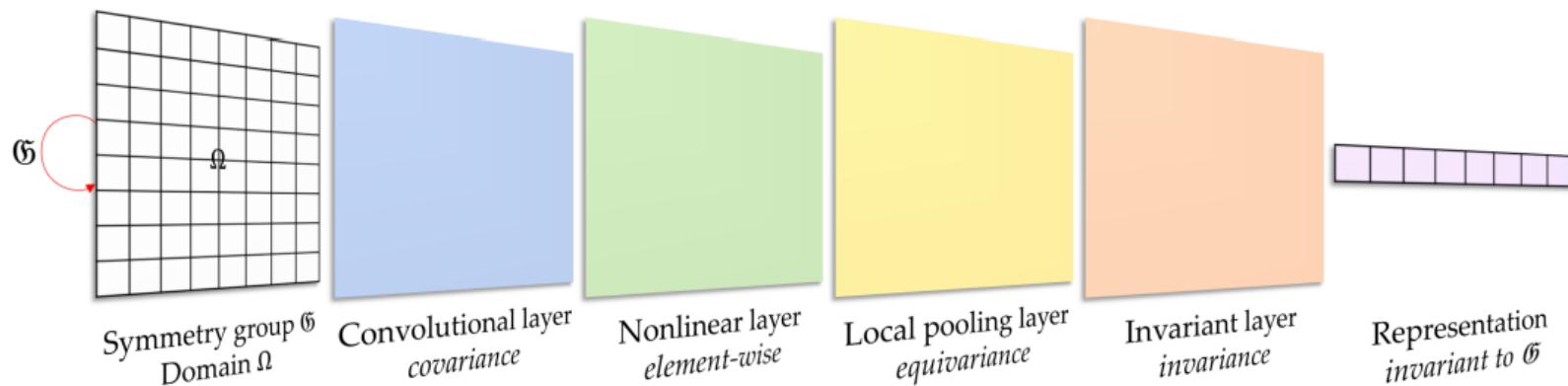
- S. Qi, Y. Zhang, C. Wang, et al. A Principled Design of Image Representation: Towards Forensic Tasks. *TPAMI*, 2023.



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

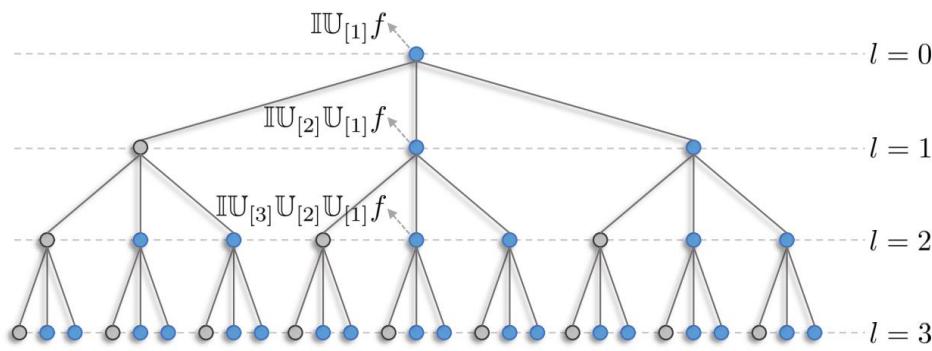
Definition

- 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:** $\mathcal{R}_p \triangleq \mathbb{I} \circ \mathbb{P}_{[L]} \circ \mathbb{S}_{[L]} \circ \mathbb{C}_{[L]} \circ \dots \circ \mathbb{P}_{[1]} \circ \mathbb{S}_{[1]} \circ \mathbb{C}_{[1]}$

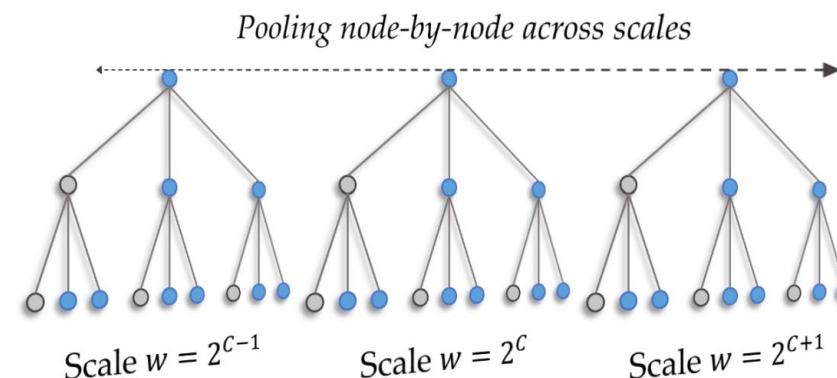


Property

- 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

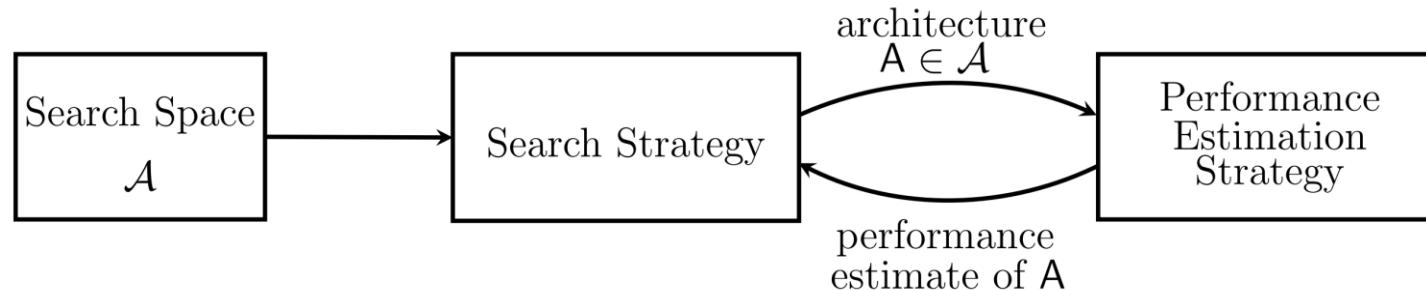


A Multi-scale Practice with \mathfrak{G}_0 Invariance

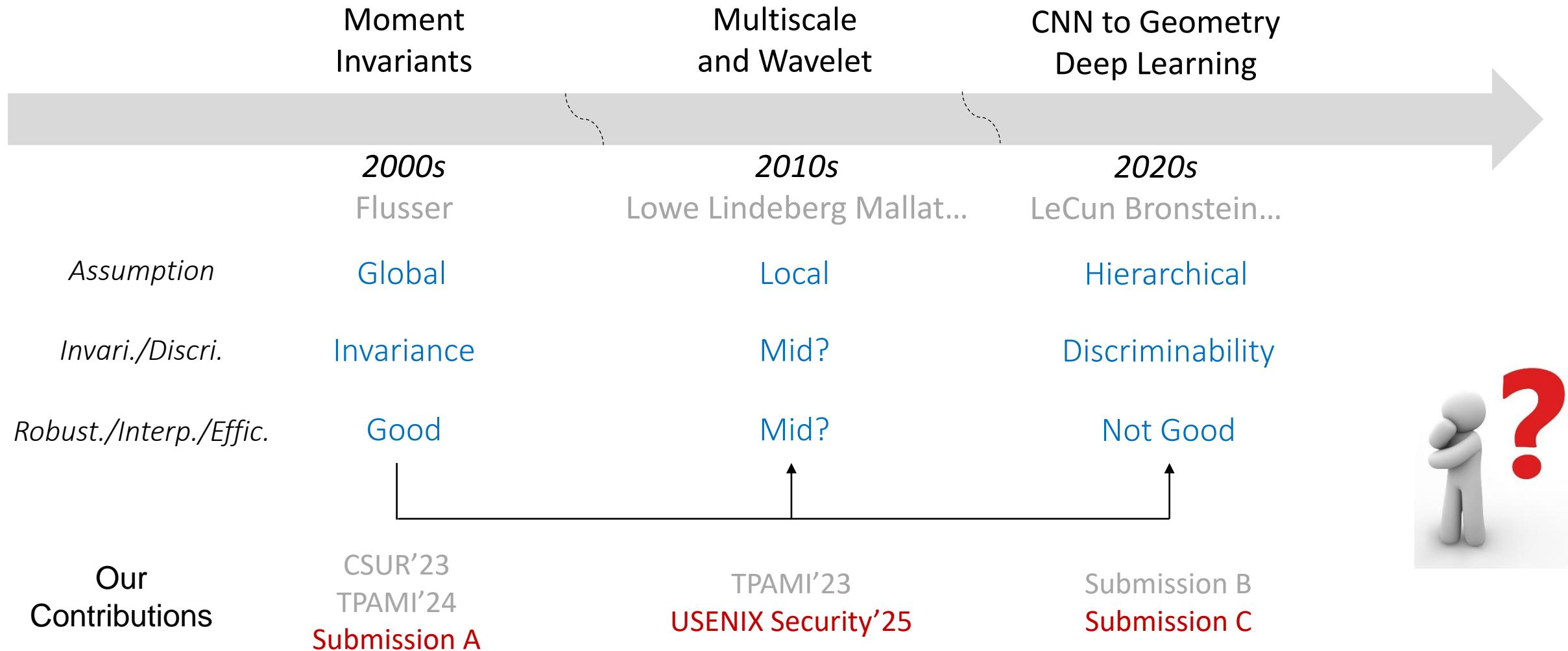
Practice

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

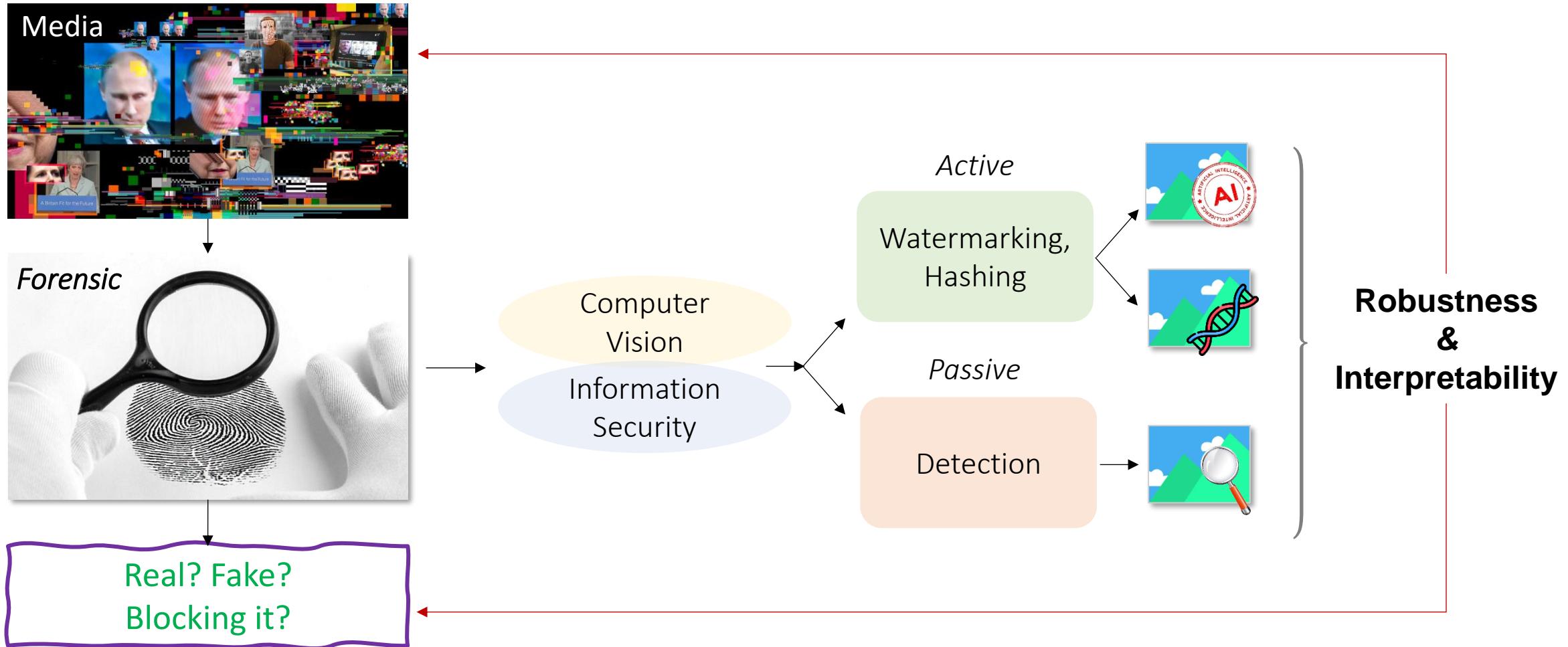
- Feature/Architecture Selection, inspired by Neural Architecture Search (NAS)
 - **Super network** covers preferred and sufficiently diverse parameters.
 - **Correlation analysis** for filtering the most relevant features.
 - **Concise network** by resampling the super network for most relevant features.
- 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.



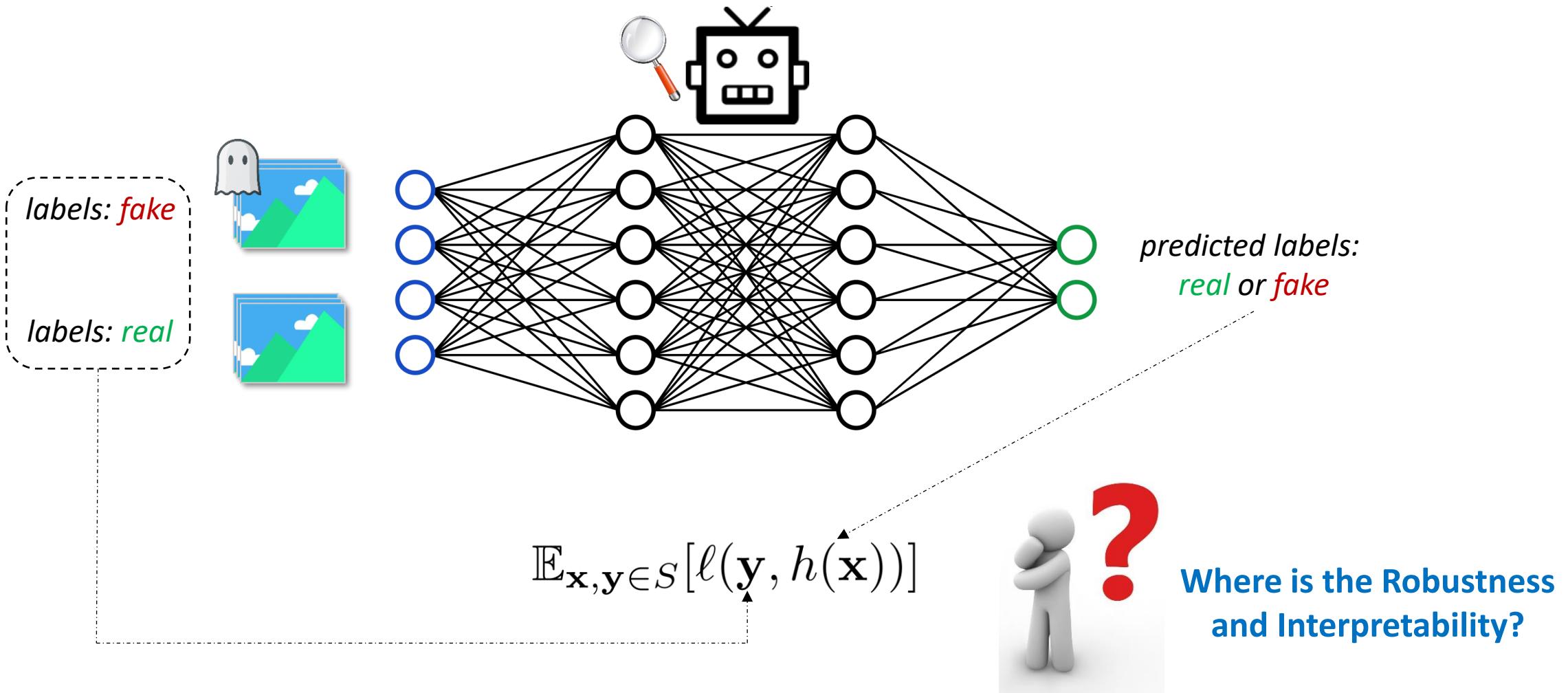
Our Contributions



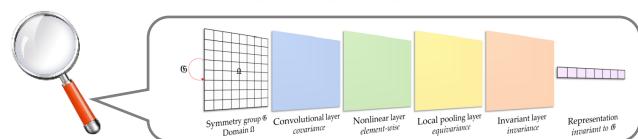
Forensic, Fighting Against AIGC Abuse



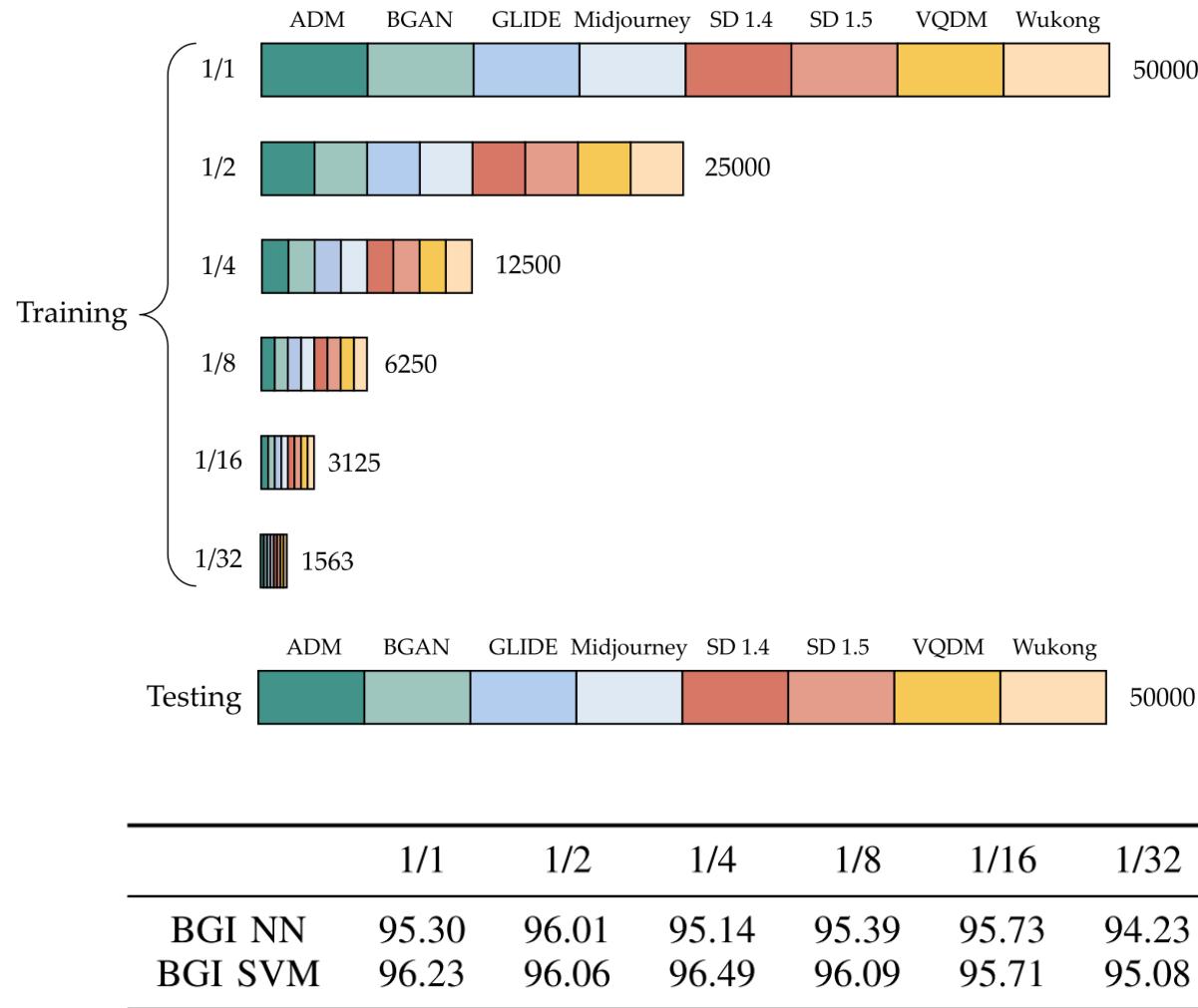
AIGC Detection: Motivations



AIGC Detection: Ideas

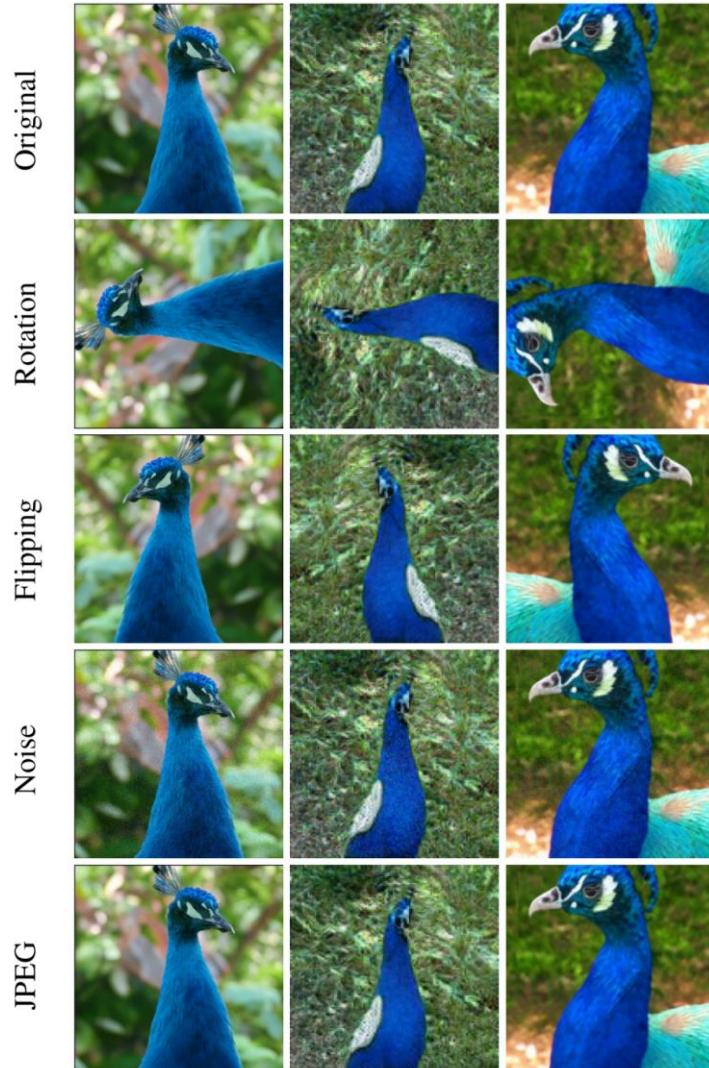


AIGC Detection: Training and Sample Efficiency



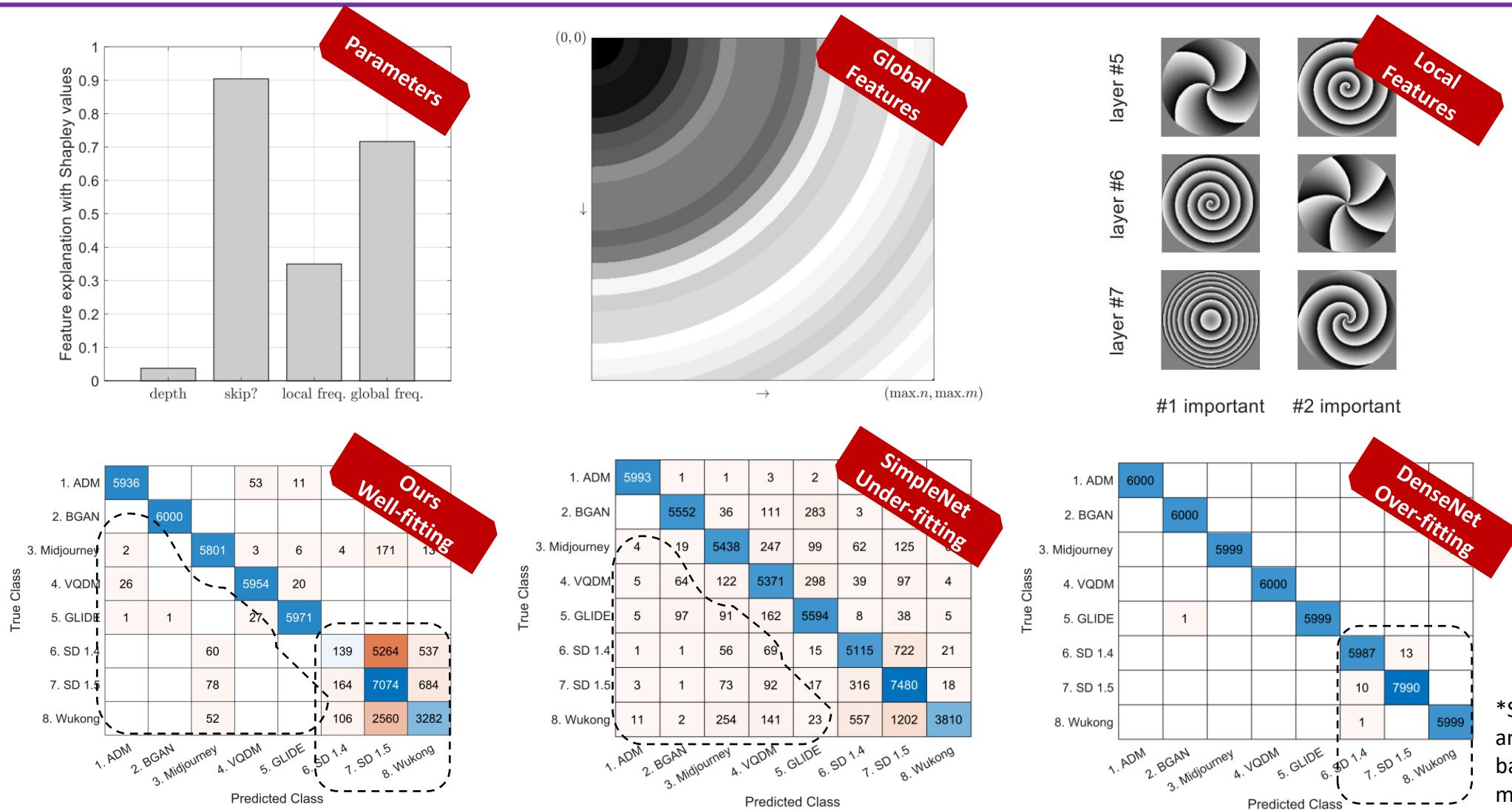
	1/8		
	Precision	Recall	F1
<i>Handcrafted</i>			
DCT NN	35.82	40.41	37.98
DCT SVM	95.00	94.98	94.99
DWT NN	49.98	99.72	66.59
DWT SVM	96.75	93.61	95.15
ScatterNet NN	86.12	77.46	81.56
ScatterNet SVM	86.75	91.67	89.14
<i>Learning</i>			
SimpleNet	75.76	36.49	49.25
ResNet	82.11	86.66	84.32
DenseNet	89.67	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

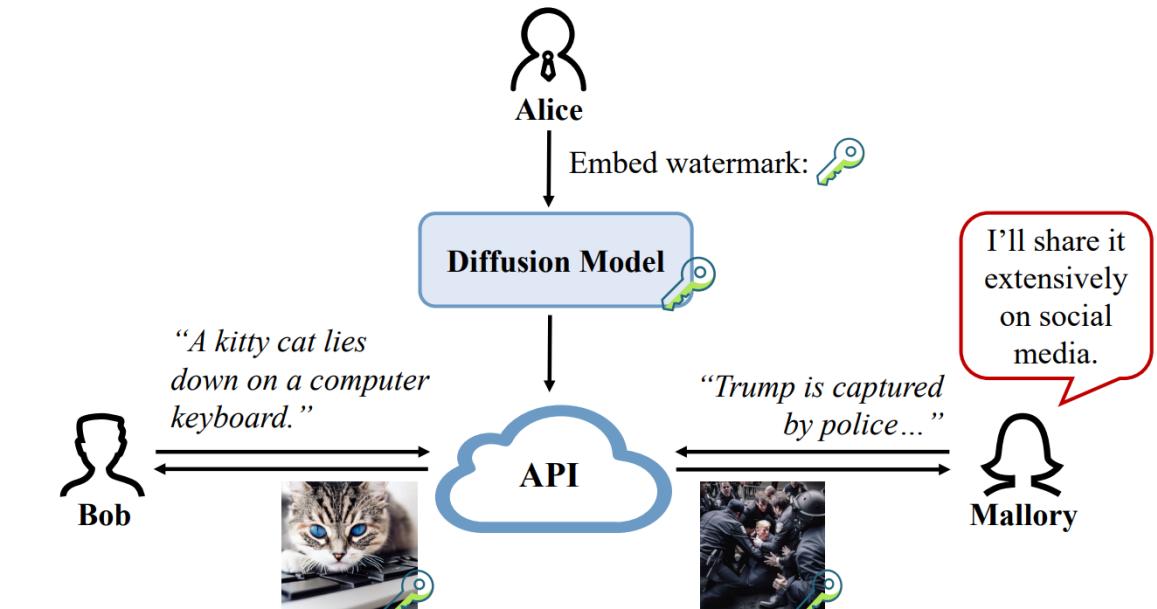


		Geometric Degradation			Signal Degradation		
		Precision	Recall	F1	Precision	Recall	F1
<i>Handcrafted</i>							
	DCT NN	0	0	0	0	0	0
	DCT SVM	80.86	95.03	87.38	78.50	91.35	84.44
	DWT NN	50.62	25.59	33.99	52.40	26.11	34.86
	DWT SVM	79.70	94.75	86.58	81.47	96.65	88.41
	ScatterNet NN	69.34	88.58	77.79	67.97	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

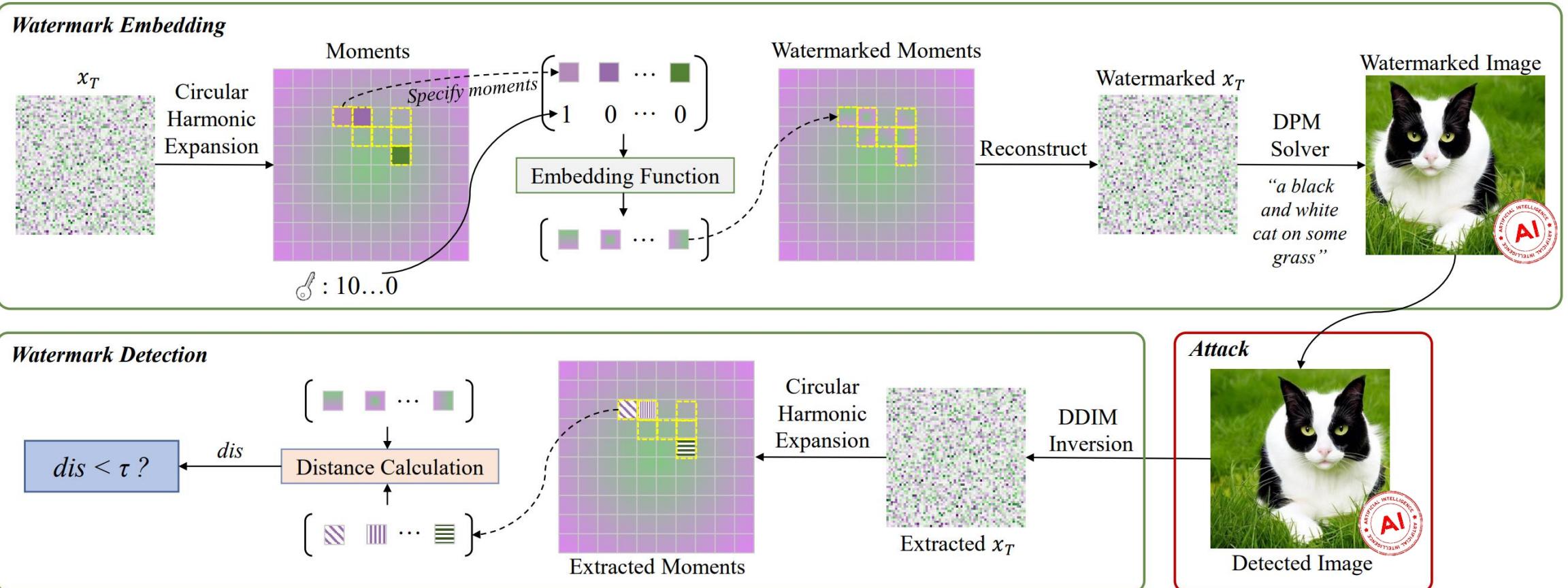


AIGC Watermarking: Motivations



Is there a balance
between robustness and
imperceptibility?

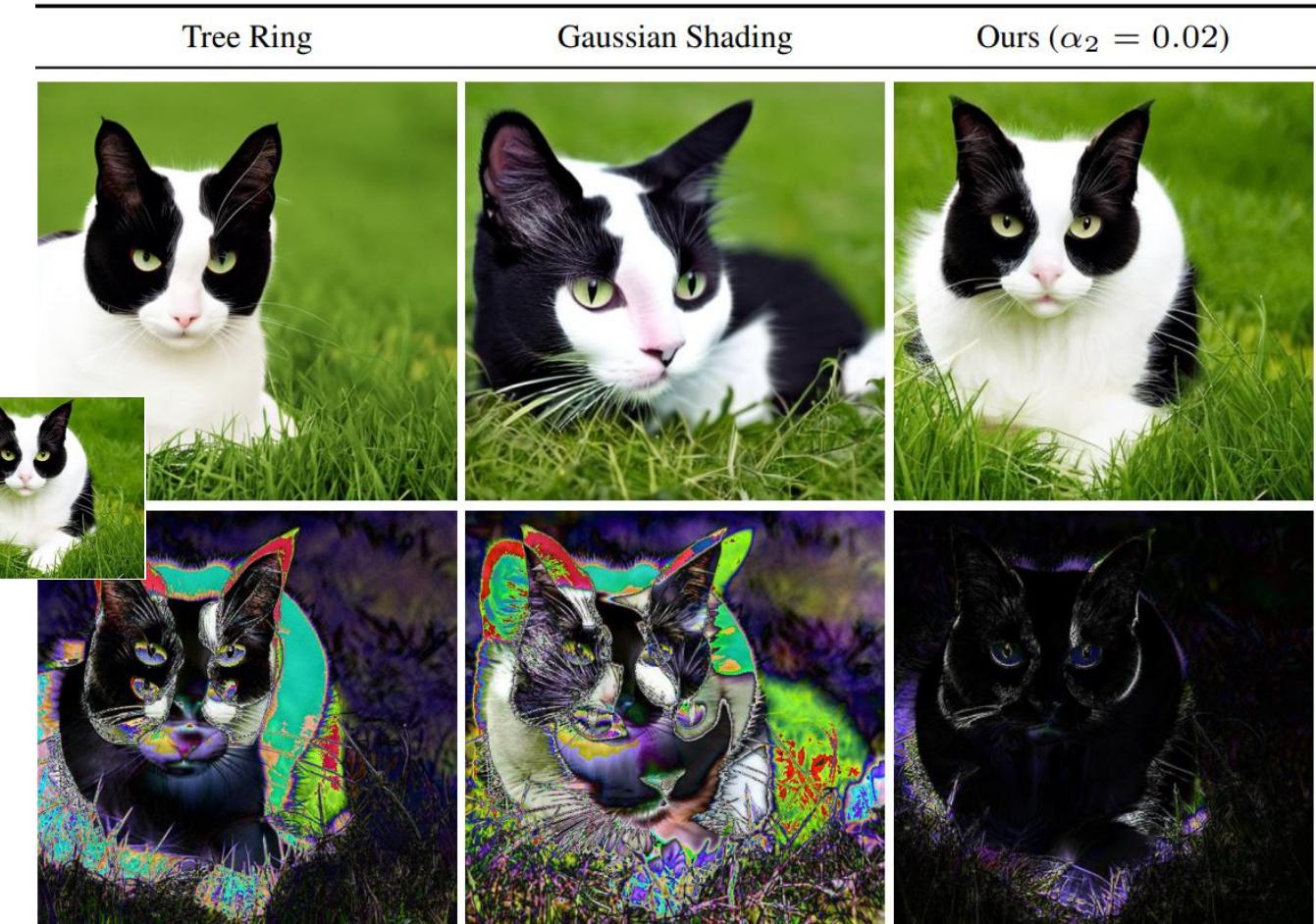
AIGC Watermarking: Ideas



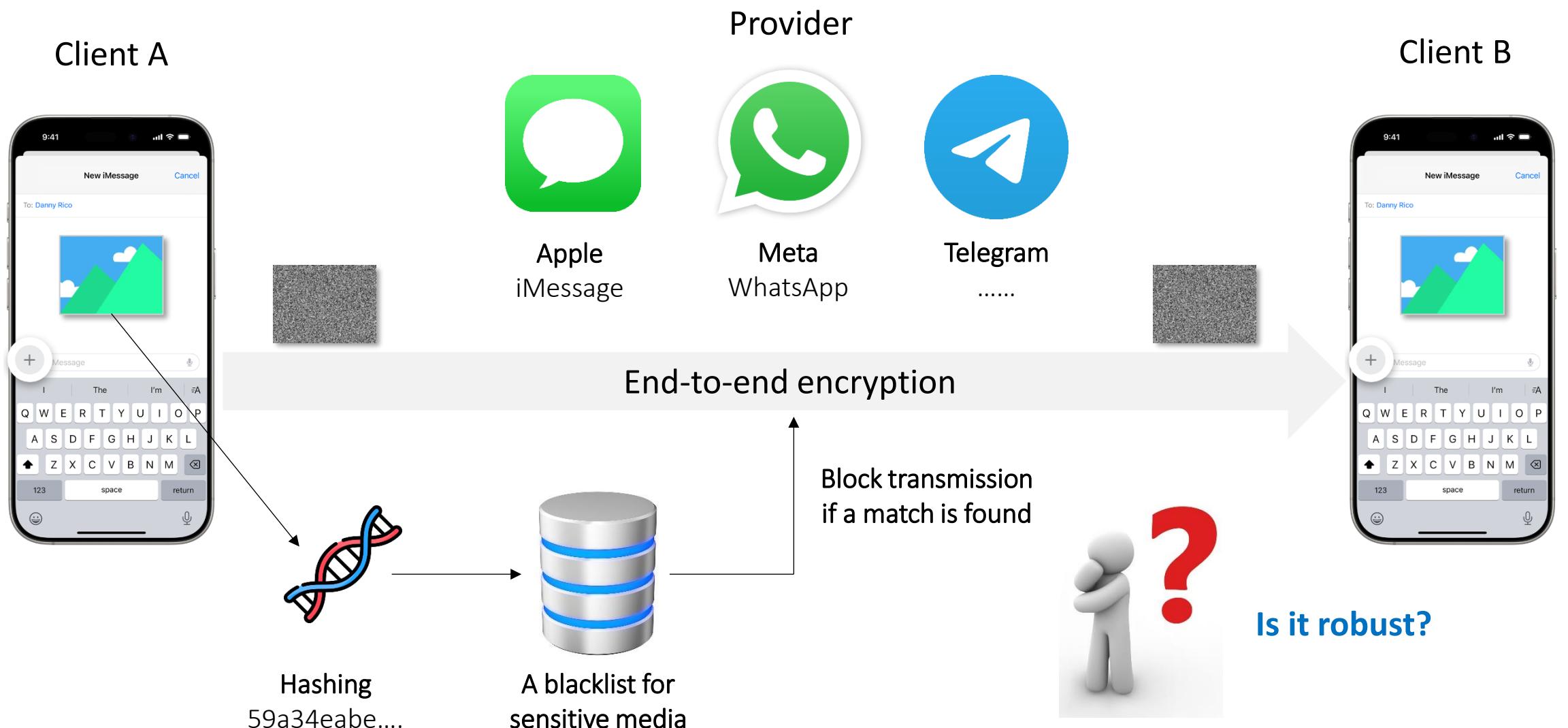
AIGC Watermarking: Robustness and Imperceptibility

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

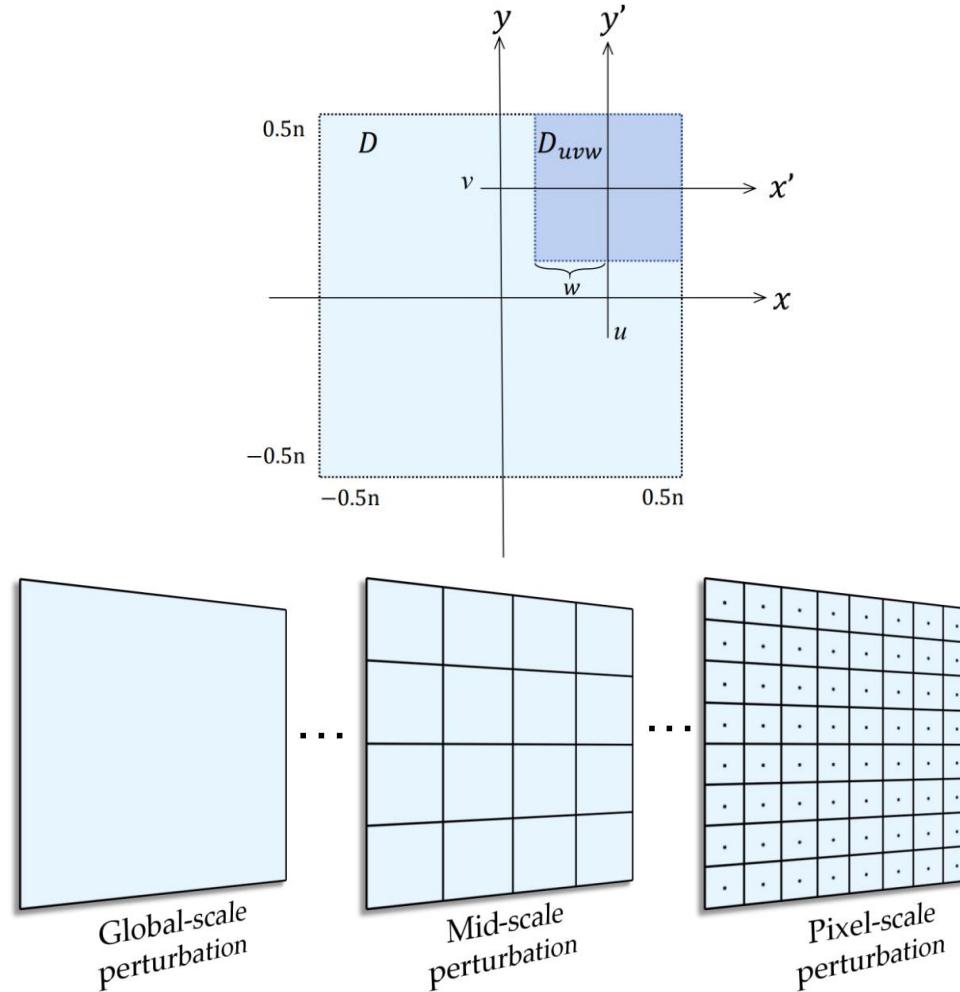
$$\mathcal{F}(X) = \langle X, V_{nm}^{uvw} \rangle = \iint_D (V_{nm}^{uvw}(x,y))^* X(x,y) dx dy, \quad (4)$$

and

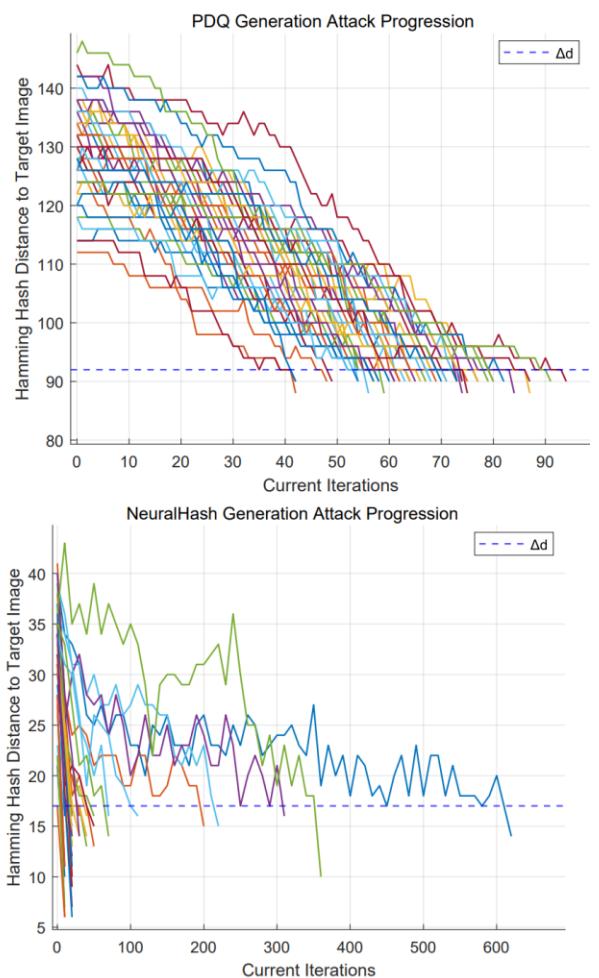
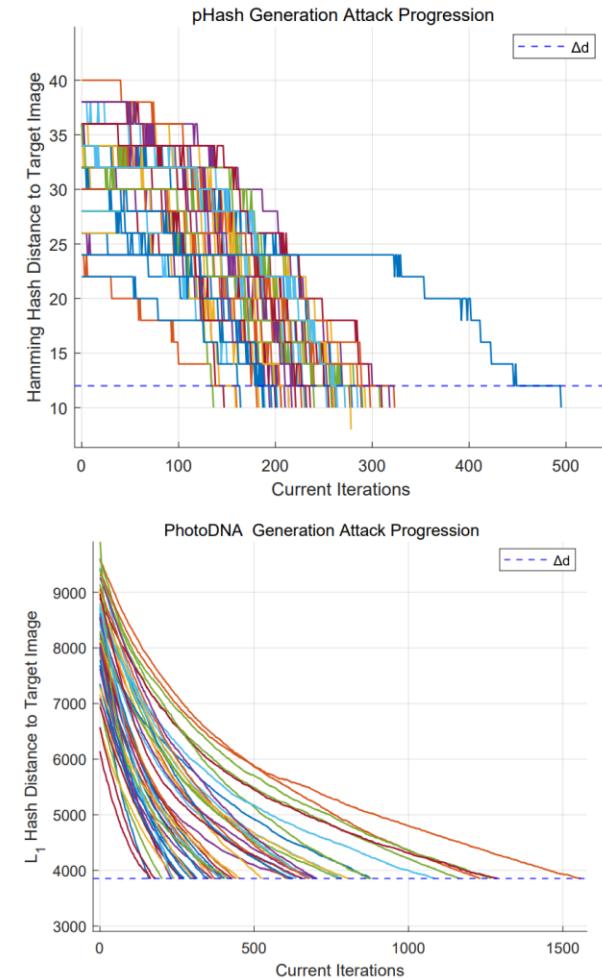
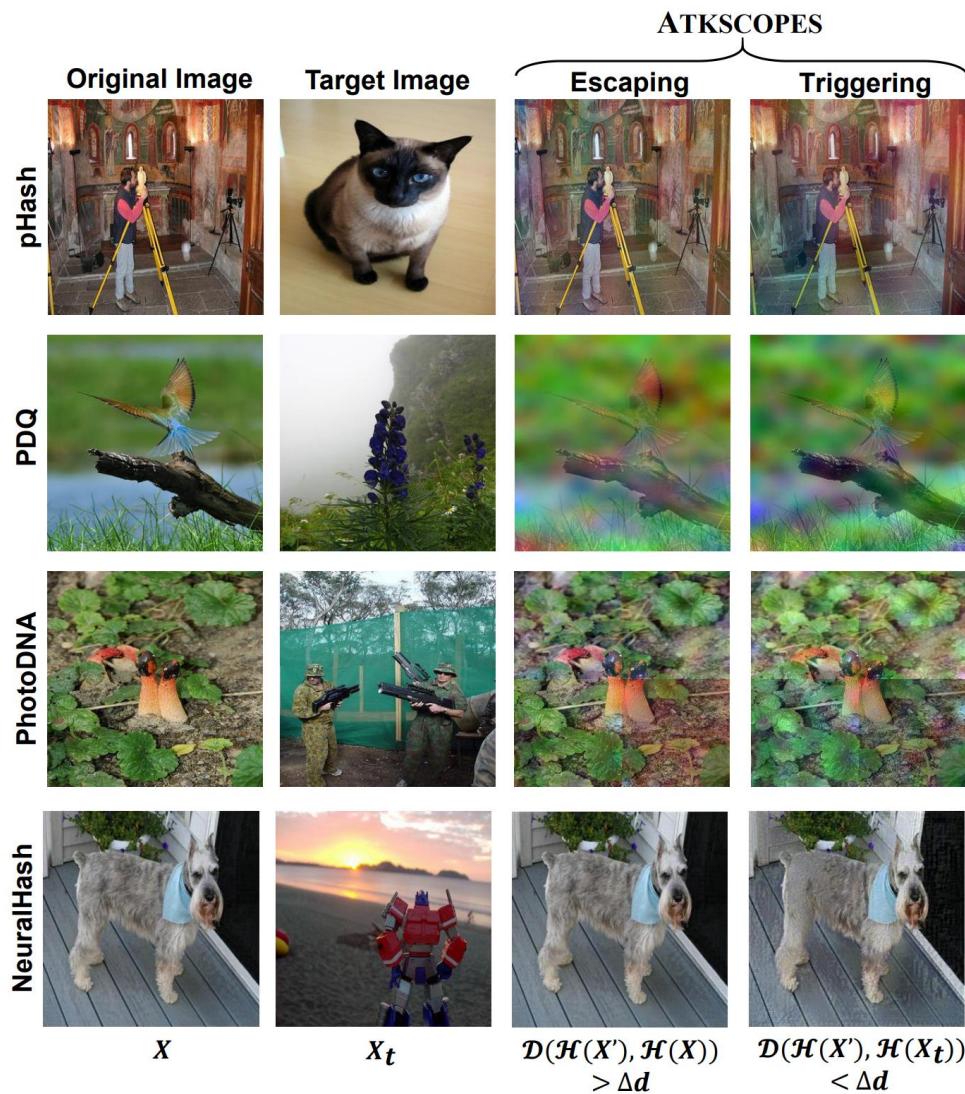
$$\mathcal{F}^{-1}(\mathcal{F}(X)) = \sum_{n,m} V_{nm}^{uvw}(x,y) \mathcal{F}(X), \quad (5)$$

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



Tutorial Outline

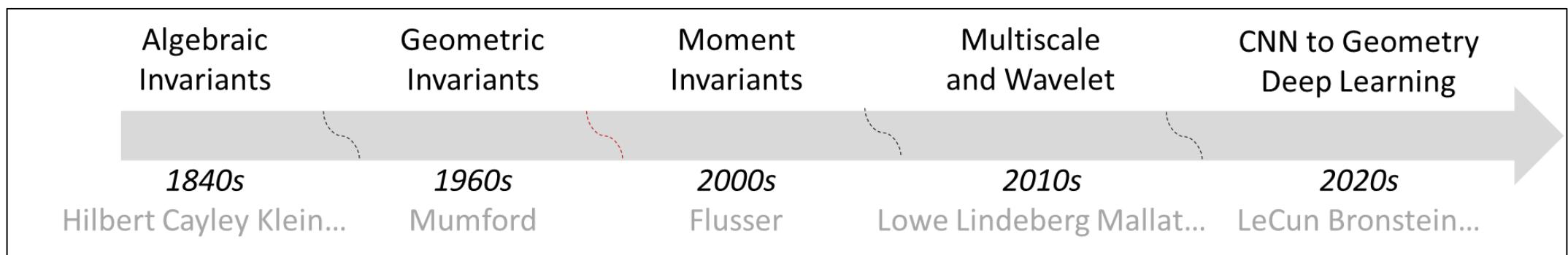
- **Part 1:** Background and challenges (20 min)
- **Part 2:** Preliminaries of invariance (20 min)
- *Q&A / Break (10 min)*
- **Part 3:** Invariance in the era before deep learning (30 min)
- **Part 4:** Invariance in the early era of deep learning (10 min)
- *Q&A / Coffee Break (30 min)*
- **Part 5:**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

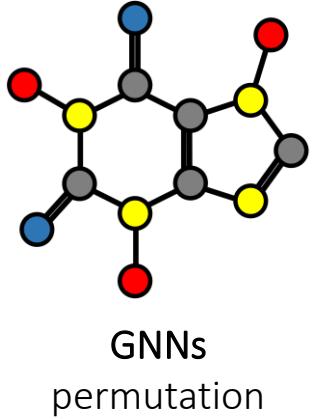
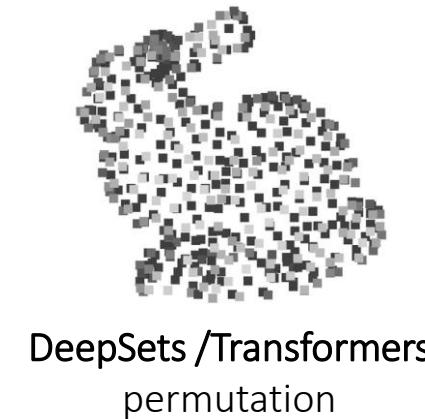
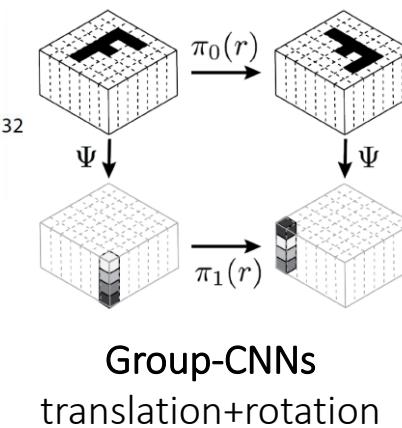
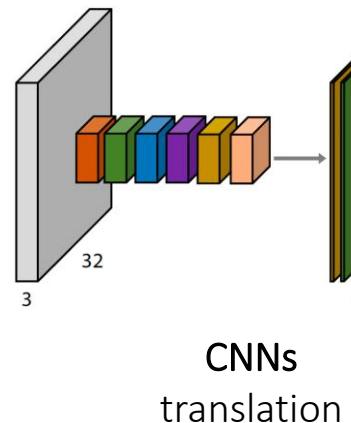
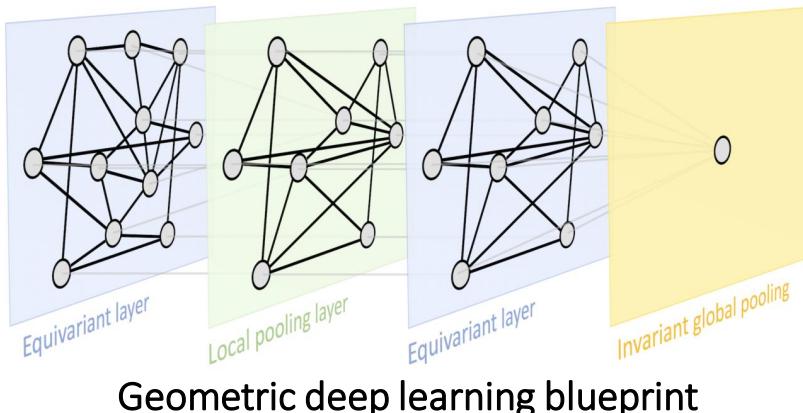
Conclusion 1: A Historical Perspective of Invariance

- A long history, from group theory, geometry, and physics
- In the era before deep learning: cornerstone
 - globally for the whole image (moment invariants), or locally for local parts of image (SIFT, DAISY, ...).
- In the early era of deep learning: largely ignored
 - CNN vs. perceptron.
- In the era of rethinking deep learning: returned, geometric deep learning
 - locally and hierarchically (CNN, equivariant CNN, equivariant NN for group set, graph...).



Conclusion 2: Rethinking Deep Learning by Invariance

- Robust, interpretable and efficient (representation) learning
 - Perfect robustness, interpretable concept, and structural efficiency.
- CNN vs. perceptron on image data
 - Translation equi/in-variance.
- Geometric Deep Learning
 - For different transformations: wavelet scattering networks, group equivariant networks.
 - For different architectures and data types: deep sets/pointnet, graph networks, transformers.

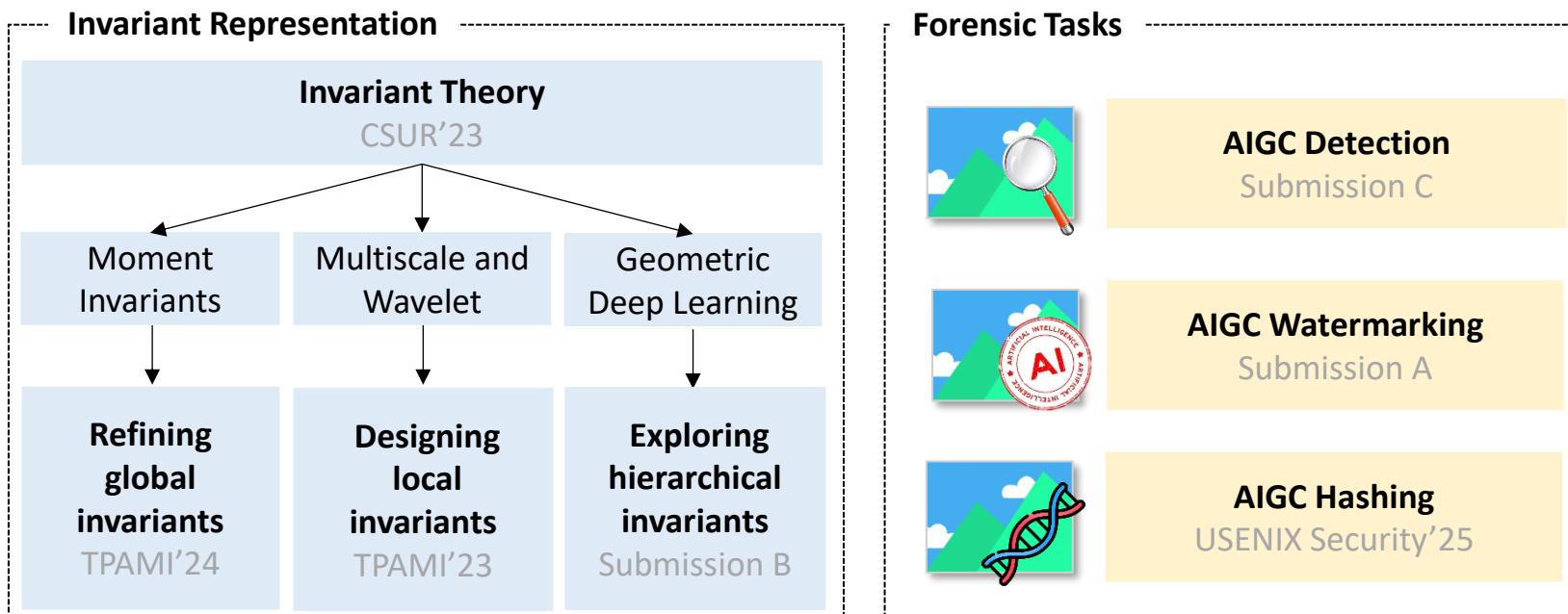


Conclusion 3: Our Works for Invariance

Trustworthy AI as [background](#)

Symmetry priors in the natural world as [principles](#)

Expanding invariant representations at theoretical and practical levels



Open Problem 1: Exploring the Limits of Handcrafted Invariants

- **The Good:**
 - Embedding knowledge; good interpretability, robustness, and efficiency.
- **The Bad and The Ugly:**
 - Discriminability, adaptivity.
- **Open Problem:**
 - Upper bound of discriminability?
 - Data-driven learning, a must?
 - If for a specific task, handcrafted invariants always sufficient?
- **Research Opportunity:**
 - Overcomplete designs of invariants, e.g., time-frequency, multi-scale, hierarchical.
 - Feature selection and explanation, from over-complete to task-discriminative.

Open Problem 2: More Flexible Designs for Learning Invariants

- **The Good:**
 - Discriminability, adaptivity.
- **The Bad and The Ugly:**
 - Limited invariance, inefficient implementation, especially for joint invariance.
- **Open Problem:**
 - Group convolution (symmetry sampling), uniformly good?
 - Element-wise operations and global pooling, sufficient for graphs/sets?
- **Research Opportunity:**
 - Continuous and high-order designs for local-equivariant and global-invariant representations.
 - Specific designs of equi/in-variance for different data types.

Open Problem 3: Real-world Impact and Application Considerations

- **The Good:**
 - Many low-level processing, some high-level tasks; AI for Science, e.g. AlphaFold.
- **The Bad and The Ugly:**
 - Real-world impact in broader applications.
- **Open Problem:**
 - Invariance, somewhat limit adaptivity?
 - Invariance, designed for generic tasks?
- **Research Opportunity:**
 - Designing high-performance invariants for specific tasks, i.e., specific data assumptions and knowledges.
 - Easy-to-use software, environment, and document.

There Is No Royal Road To Geometry

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

§ Thank you!

by Shuren Qi

shurenqi@cuhk.edu.hk | shurenqi.github.io