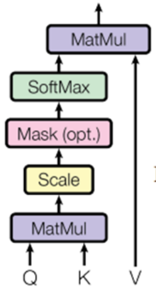
Invariant Representations for Visual Forensic Tasks

Shuren Qi

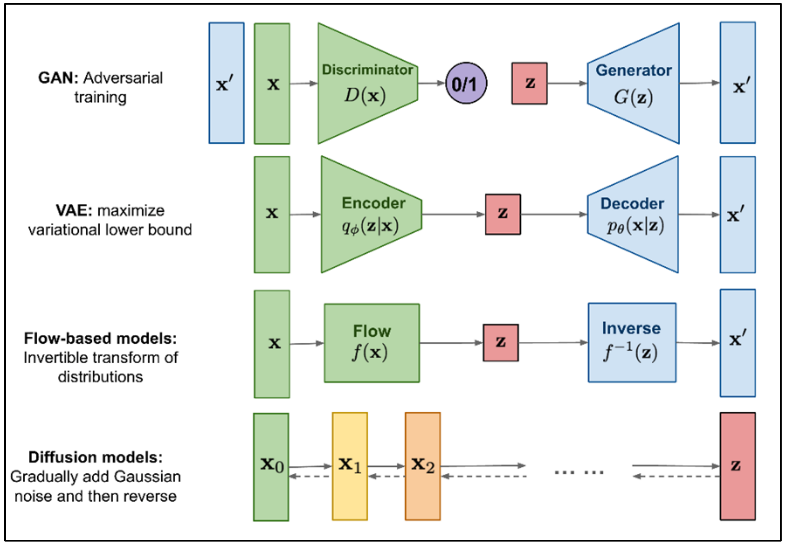
Center for Mathematical Artificial Intelligence
Department of Mathematics
The Chinese University of Hong Kong

The Era of AIGC

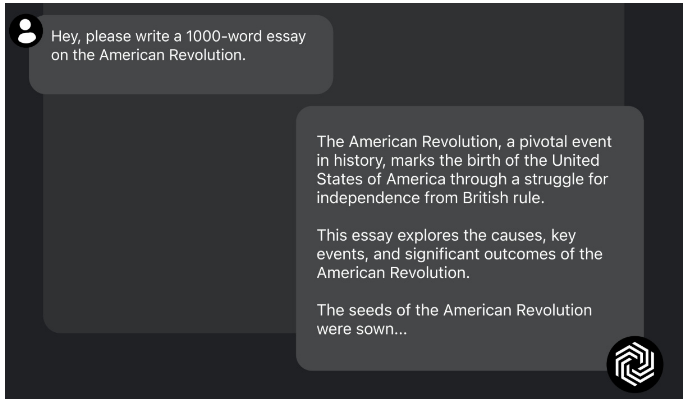


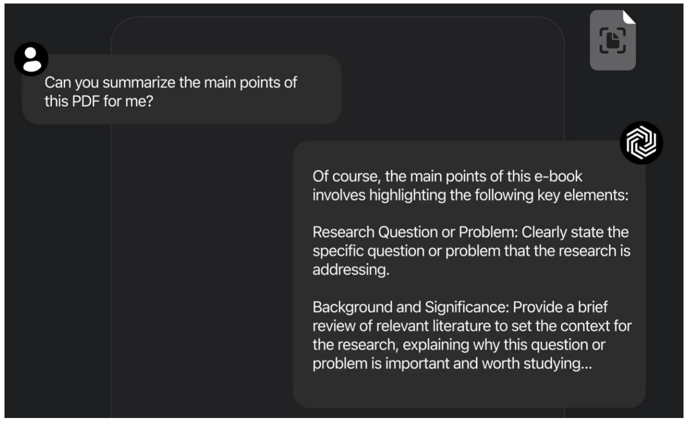
$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head_1}, ..., \mathrm{head_h})W^O$$
$$\text{where } \mathrm{head_i} = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

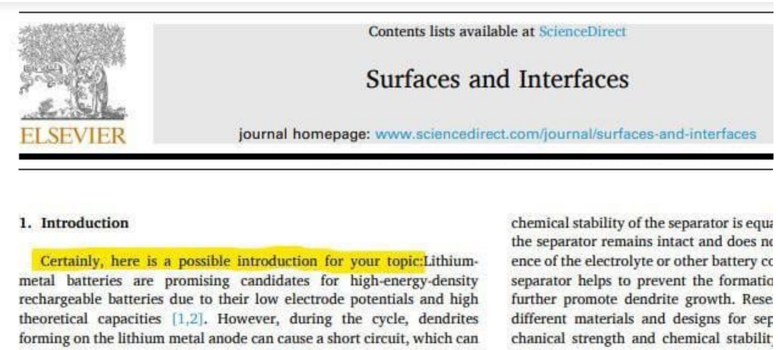


The Good


Hey, please write a 1000-word essay on the American Revolution.
The American Revolution, a pivotal event in history, marks the birth of the United States of America through a struggle for independence from British rule.
This essay explores the causes, key events, and significant outcomes of the American Revolution.
The seeds of the American Revolution were sown...


Can you summarize the main points of this PDF for me?
Of course, the main points of this e-book involves highlighting the following key elements:
Research Question or Problem: Clearly state the specific question or problem that the research is addressing.
Background and Significance: Provide a brief review of relevant literature to set the context for the research, explaining why this question or problem is important and worth studying...

The Bad and The Ugly


Contents lists available at ScienceDirect

Surfaces and Interfaces

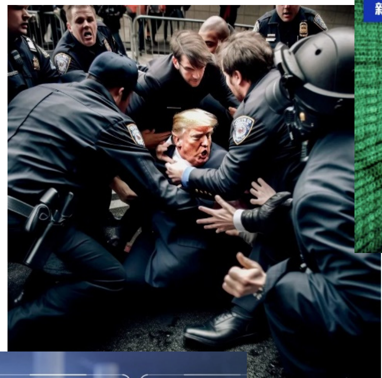
journal homepage: www.sciencedirect.com/journal/surfaces-and-interfaces

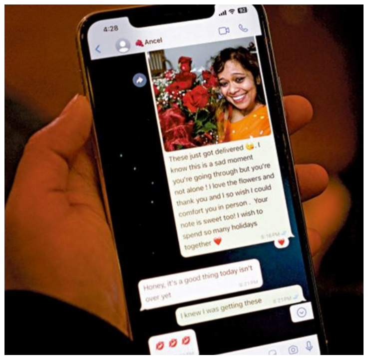
ELSEVIER

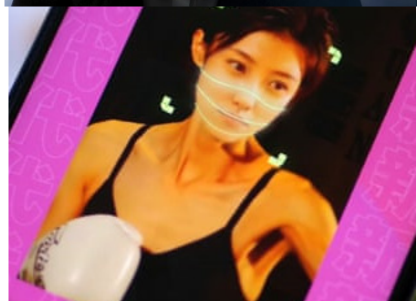
1. Introduction

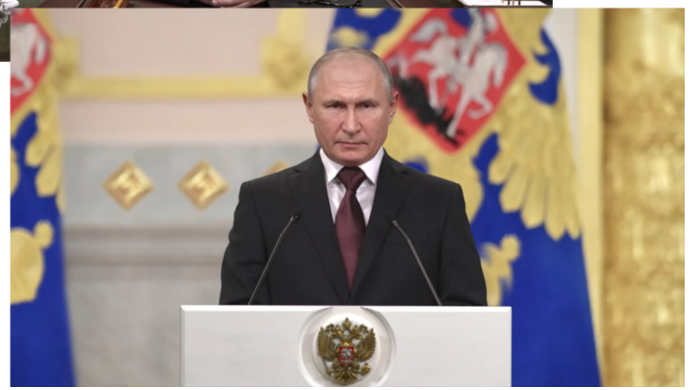
Certainly, here is a possible introduction for your topic:Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can

chemical stability of the separator is equa the separator remains intact and does no ence of the electrolyte or other battery co separator helps to prevent the formatio further promote dendrite growth. Rese different materials and designs for sep chanical strength and chemical stabilit

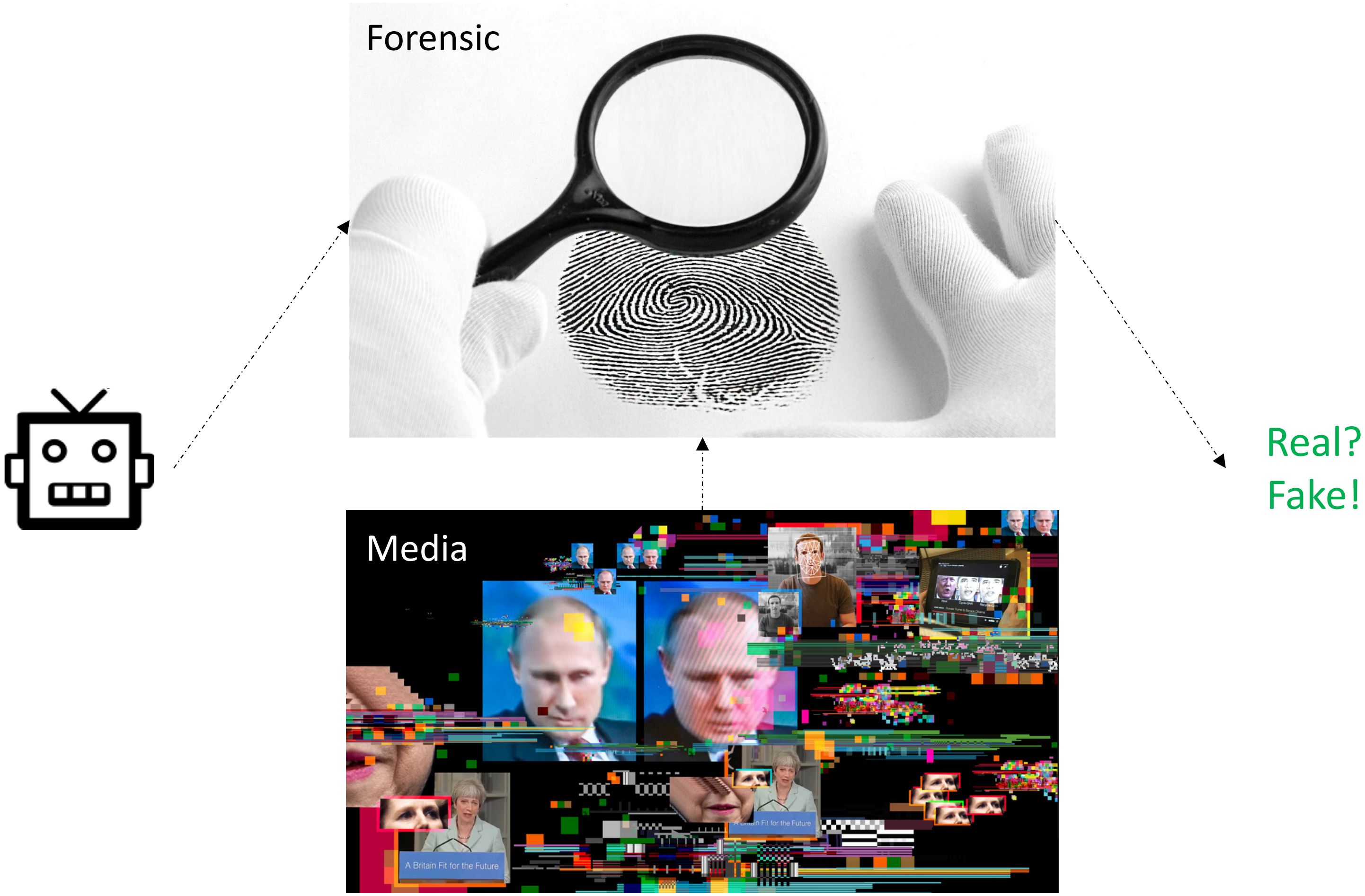








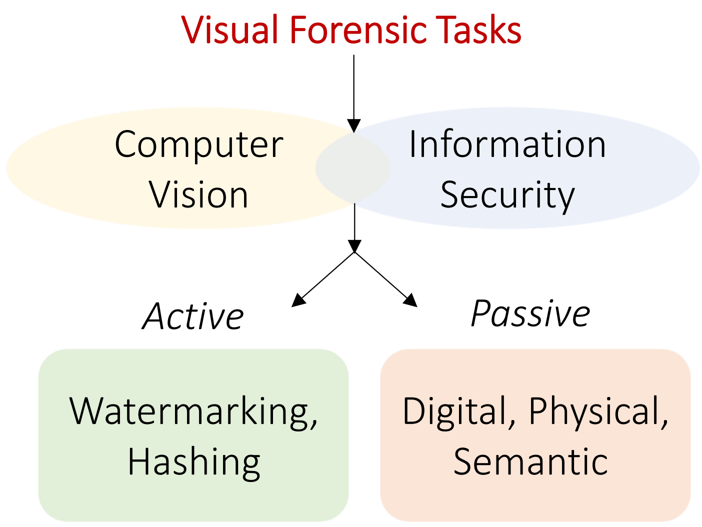
Fighting Against AIGC Abuse



Visual Forensic Tasks

at the intersection of vision and security

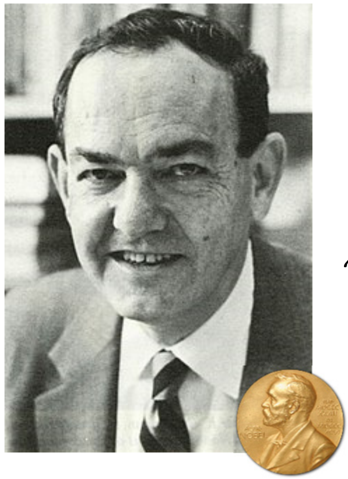
- Forensic research aims to check the authenticity of visual data, at the intersection of computer vision and information security.
- Forensic research is carried out in active and passive paths, depending on whether the action is taken before or after data distribution.
  - **Active forensics** are typically performed by embedding robust patterns in the image, i.e., digital *watermarking*, or extracting image fingerprints as registration, i.e., *hashing* and blockchain.
  - **Passive forensics** rely exclusively on the given image itself. They discover artifacts at *digital*, *physical*, and *semantic* levels, which are inevitably introduced by certain manipulations.



Visual Forensic Tasks
consistency with typical visual tasks

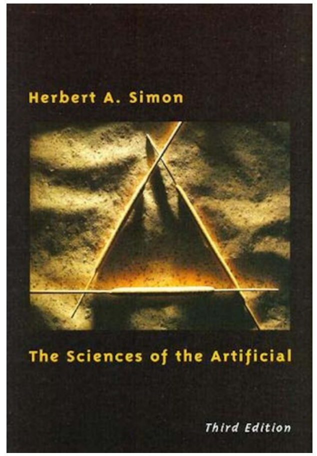
- Similar to typical visual tasks (e.g., image classification), the effectiveness of visual forensic is also strongly dependent on proper representations.

Visual data ⇨ Representation ⇨ Knowledge discovery ⇨ Application { Typical visual tasks / Forensic tasks }



H. Simon, 1969
The Sciences of the Artificial

"solving a problem simply means representing it so as to make the solution transparent"



Visual Forensic Tasks
differences from typical visual tasks

- Unlike typical visual tasks, forensic tasks show the basic features of information security research.
  - **Adversary**: there is always an adversary in forensics, so not only the natural variation, but also the active attack.
  - **Evidence and credibility**: forensics are required to provide judgmental evidence for debates, so not only the accuracy, but also the credibility.

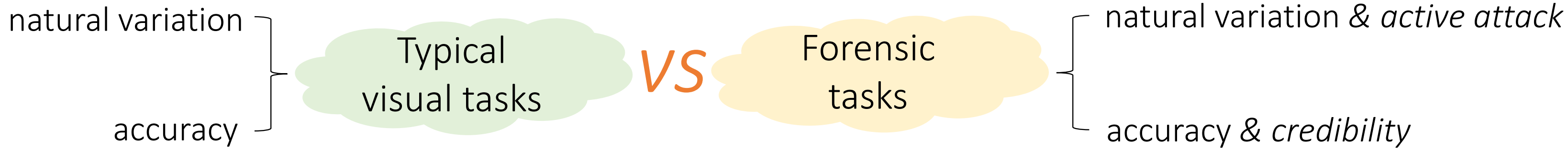
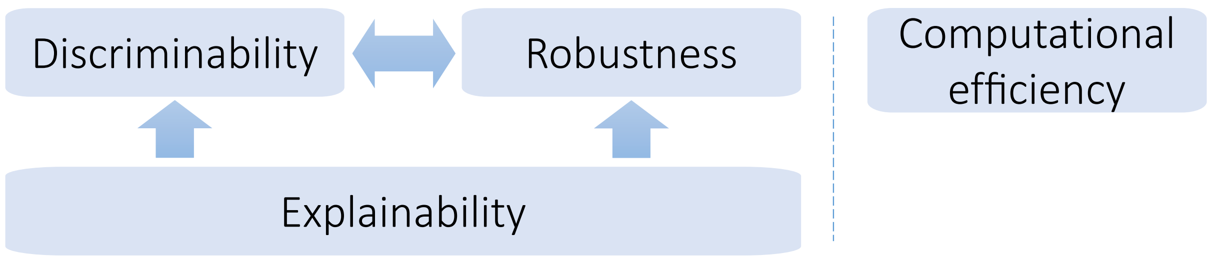


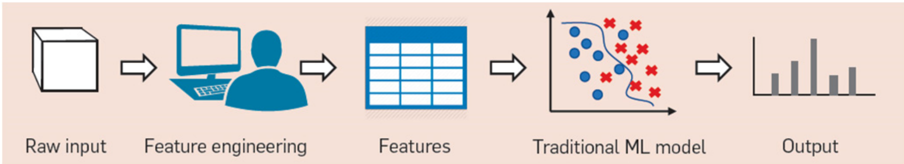
Image Representation for Visual Forensic Tasks

- According to the consistency and difference between typical visual tasks and forensic tasks, the principles that image representations in forensic should satisfy are summarized here.
  - **Discriminability:** the representation is sufficiently informative for distinguishing between real and fake data.
  - **Robustness:** the representation is not influenced by variations that may be introduced by the adversary.
  - **Explainability:** the representation should have reliable theory guarantees, implying that causation is more important than correlation, due to the role as evidence.
  - **Computational efficiency:** the representation should have a reasonable implementation.



Related Works
learning and hand-crafted representations

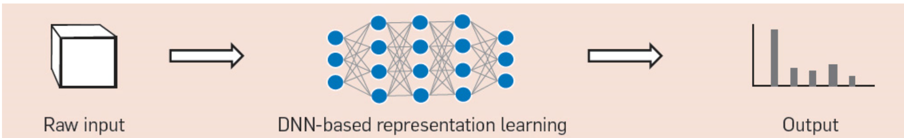
hand-crafted



- *Pros:* robustness and explainability
- *Cons:* discriminability

VS

learning



- *Pros:* discriminability
- *Cons:* robustness and explainability

LeCun, Y., Bengio, Y. & Hinton, G. , 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"

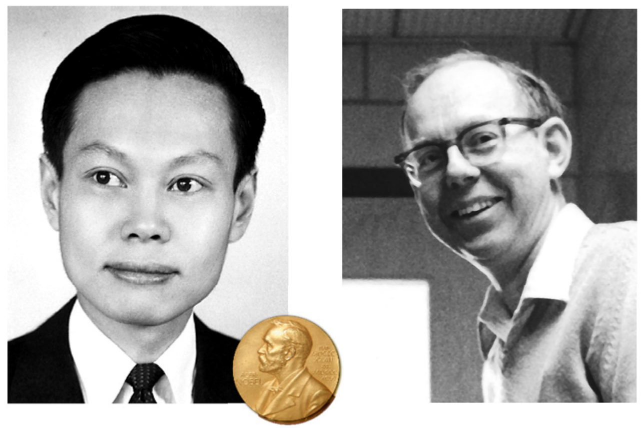
Invariance/Symmetry Priori

- In general, an AI system is a digital modeling of the physical systems in the natural world. Therefore, the exploitation level of natural priors determines the robustness and explainability level in the AI system.
- Among many priors, symmetry may be the most fruitful prior — informally, a symmetry of a system is a transformation that leaves a certain property of system invariant.

F. Klein, 1872
Erlangen Program

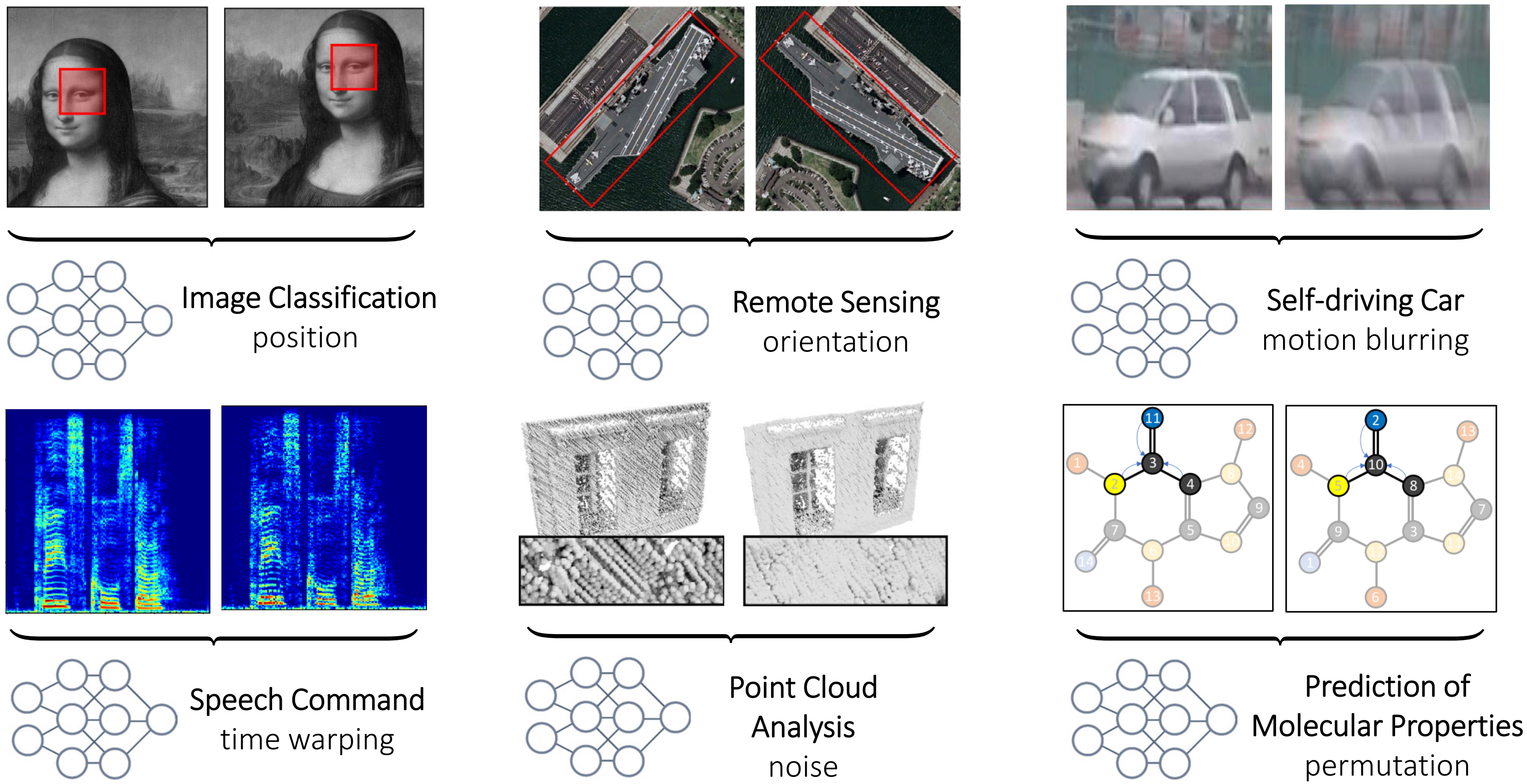
E. Noether, 1918
Noether's Theorem

H. Weyl, 1929
The Book of Symmetry

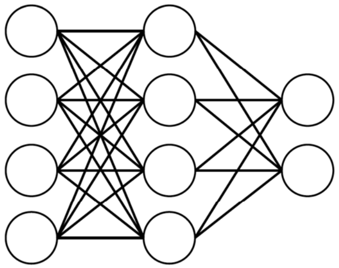


C. N. Yang & R. L. Mills, 1954
Yang-Mills Theory

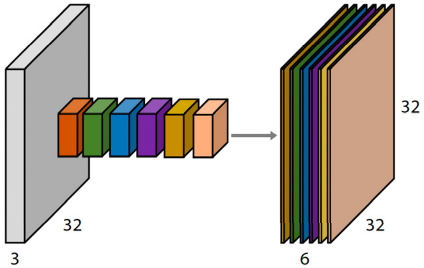
Invariance/Symmetry Priori is Ubiquitous



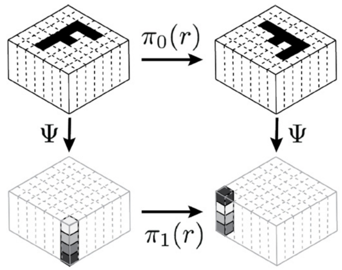
Representations Equipped with Symmetry/Invariance
Geometric Deep Learning



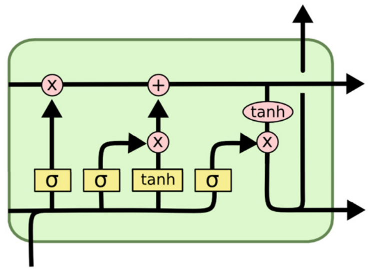
Perceptrons
function regularity



CNNs
translation

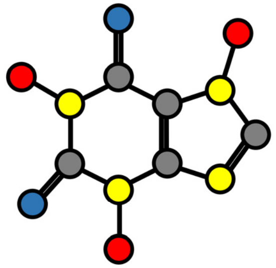


Group-CNNs
translation+rotation



LSTMs
time warping

DeepSets /Transformers
permutation

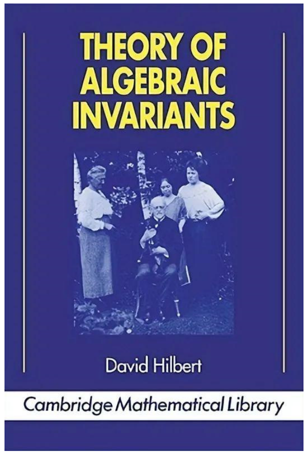


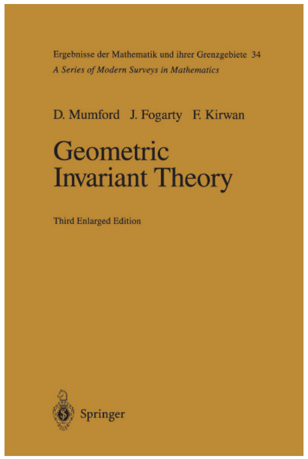
GNNs
permutation

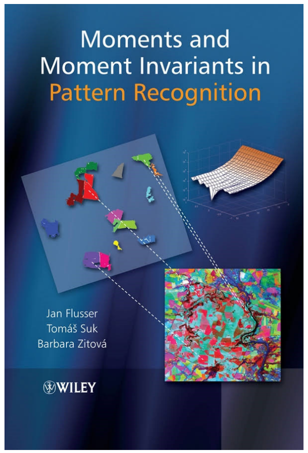
Intrinsic CNNs
isometry/gauge choice

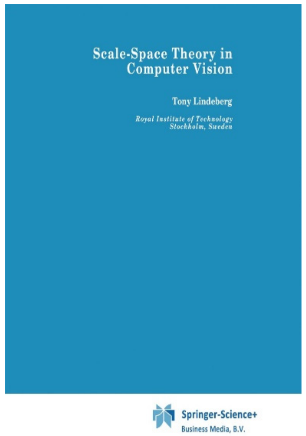
A History of Invariance/Symmetry (in Representation)

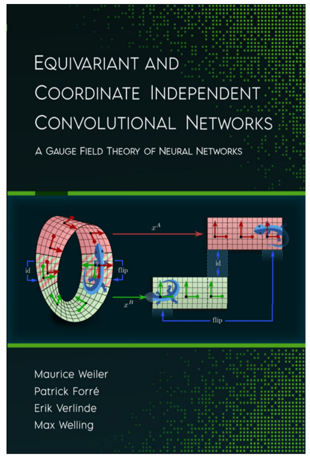
Algebraic Invariants	Geometric Invariants	Moment Invariants	Multiscale and Wavelet	CNN to Geometry Deep Learning
1840s	*1960s*	*2000s*	*2010s*	*2020s*
Hilbert Cayley Klein…	Mumford	Flusser	Lowe Lindeberg Mallat…	LeCun Bronstein…











My Contributions to the Representations

	Moment Invariants	Multiscale and Wavelet	CNN to Geometry Deep Learning
	2000s	*2010s*	*2020s*
	Flusser	Lowe Lindeberg Mallat…	LeCun Bronstein…
Assumption	Global	Local	Hierarchical
Invari. or Discri.	Invariance	Mid?	Discriminability
Robust. & Interp.	Good	Mid?	Not Good
My Contributions	CSUR'23, TPAMI'24, CVPR review	TPAMI'23, USENIX Security'25	IJCV review, TIFS review

My Contributions to the Forensics

Paths	Passive Forensic	Active Forensic
Tasks	AIGC Detection	AIGC Watermarking/Hashing
Motivations	Robustness	Interpretability
My Contributions	TPAMI’23, IJCV review, TIFS review	TPAMI’24, CVPR review, USENIX Security’25

My Research Overview

Trustworthy AI as background

Symmetry priors in the natural world as principles

Expanding invariant representations at theoretical and practical levels

Invariant Representation

Invariant Theory
CSUR’23

Moment Invariants	Multiscale and Wavelet	Geometric Deep Learning
Refining global invariants TPAMI’24	**Designing local invariants** TPAMI’23	**Exploring hierarchical invariants** IJCV review

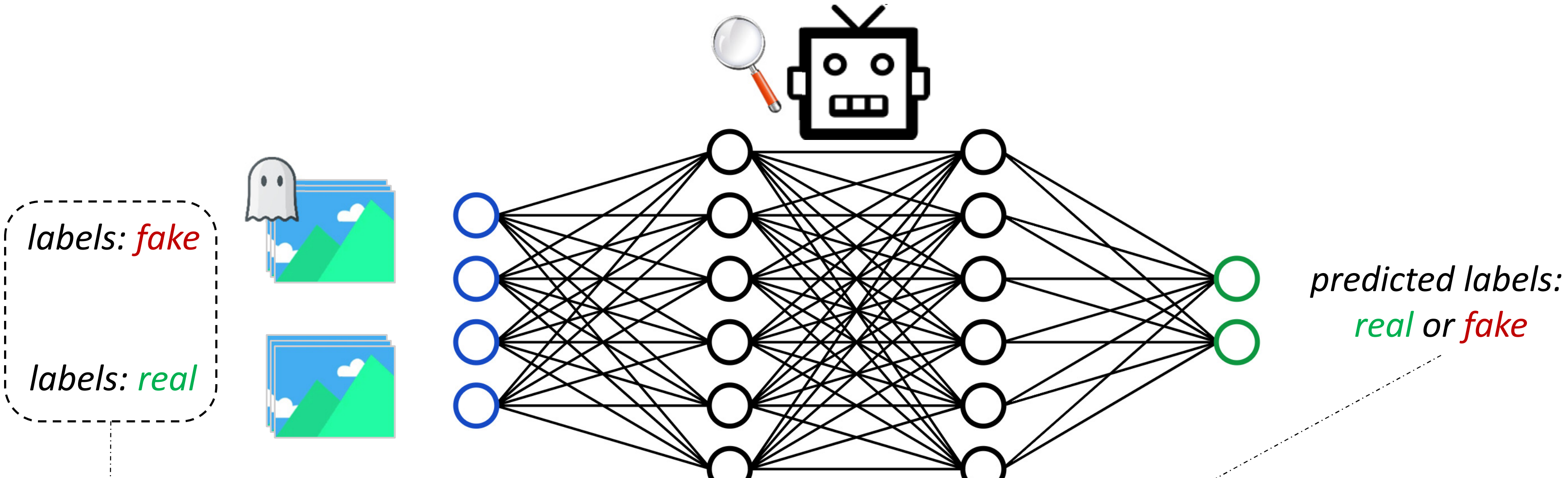
Forensic Tasks

AIGC Detection
TIFS review

AIGC Watermarking
CVPR review

AIGC Hashing
USENIX Security’25

AIGC Detection

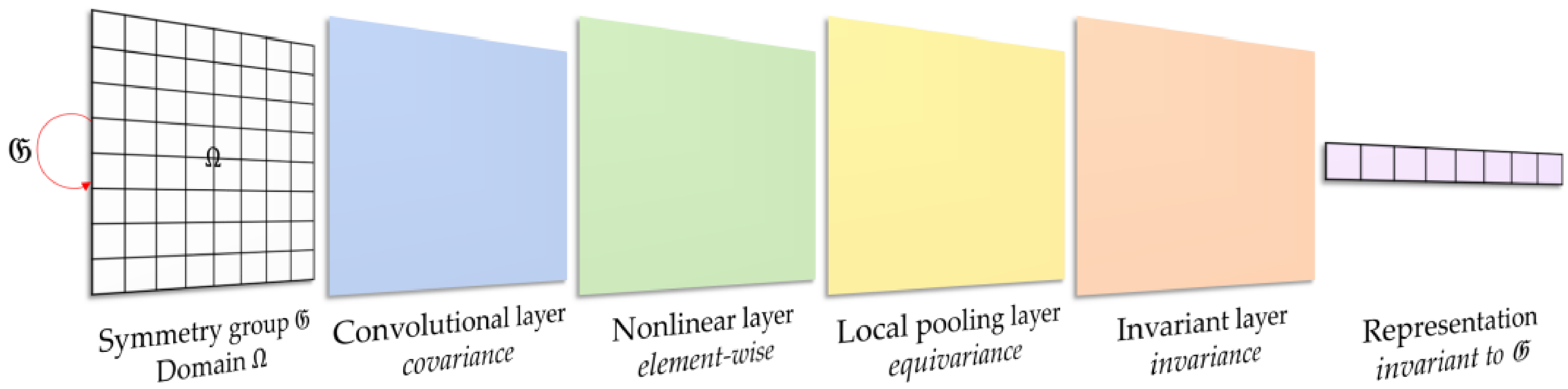


$$L_{\mathcal{S}}(h) = \frac{1}{n}\sum_{i=1}^{n} \ell(h(\mathbf{x}_i), y_i)]$$

Where is the Robustness and Interpretability?

AIGC Detection

- The main idea is to generalize the fundamental theory of global and local invariants to the hierarchical case.
- We propose hierarchical invariant representation, by rethinking the typical modules of CNN.



AIGC Detection

- We formalize a blueprint for hierarchical invariance and define new modules with their compositions to fulfill the blueprint. The group theory shows the continuous and one-shot equivariance at each intermediate layer.

Property 1. ***(Equivariance for translation, rotation, and flipping).*** *For a representation unit $\mathbb{U} \triangleq \mathbb{P} \circ \mathbb{S} \circ \mathbb{C}$ with arbitrary parameters λ (for the convolutional layer), any composition of $\mathbb{U}$ satisfy the joint equivariance for translation, rotation, and flipping (ignoring edge effects and resampling errors), i.e., the following identity holds:*

$$\mathbb{U}_{[L]} \circ \cdots \circ \mathbb{U}_{[2]} \circ \mathbb{U}_{[1]}(\mathfrak{g}_1 M) \equiv \mathfrak{g}_1 \mathbb{U}_{[L]} \circ \cdots \circ \mathbb{U}_{[2]} \circ \mathbb{U}_{[1]}(M). \quad (16)$$

for any composition length $L \geq 1$, any $\mathfrak{g}_1 \in \mathfrak{G}_1$ and $M \in X$, where $\mathfrak{G}_1$ is the translation/rotation/flipping symmetry group.

Property 2. ***(Covariance for scaling).*** *For a representation unit $\mathbb{U}$, where the scale parameter of its convolutional layer is specified as w with a notation $\mathbb{U}^w \triangleq \mathbb{P} \circ \mathbb{S} \circ \mathbb{C}^w$, any composition of $\mathbb{U}^w$ satisfy the covariance for scaling (ignoring edge effects and resampling errors), i.e., the following identity holds:*

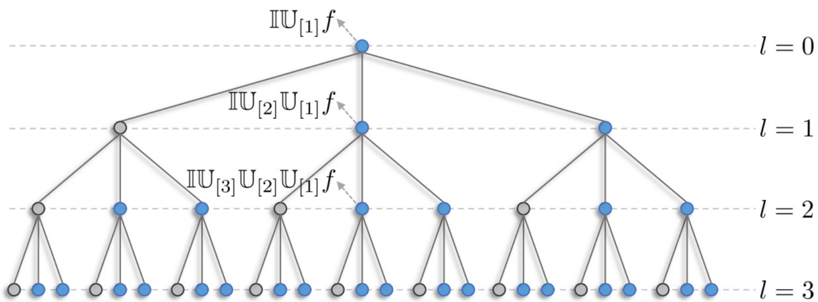
$$\begin{aligned} &\mathbb{U}^w_{[L]} \circ \cdots \circ \mathbb{U}^w_{[2]} \circ \mathbb{U}^w_{[1]}(\mathfrak{g}_2 M) \\ &\equiv \mathfrak{g}'_2 \mathbb{U}^w_{[L]} \circ \cdots \circ \mathbb{U}^w_{[2]} \circ \mathbb{U}^w_{[1]}(M) \\ &= \mathfrak{g}_2 \mathbb{U}^{ws}_{[L]} \circ \cdots \circ \mathbb{U}^{ws}_{[2]} \circ \mathbb{U}^{ws}_{[1]}(M), \end{aligned} \quad (18)$$

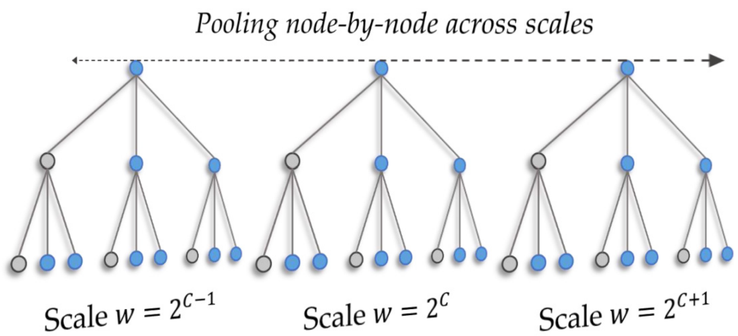
for any composition length $L \geq 1$, any $\mathfrak{g}_2 \in \mathfrak{G}_2$ and $M \in X$, where $\mathfrak{g}'_2$ is a predictable operation corresponding to $\mathfrak{g}_2$ with explicit form $\mathfrak{g}'_2 \mathbb{U}^w \triangleq \mathfrak{g}_2 \mathbb{U}^{ws}$, s is the scaling factor w.r.t. $\mathfrak{g}_2$, and $\mathfrak{G}_2$ is the scaling symmetry group.

Property 3. ***(Hierarchical invariance).*** *For any composition of representation unit $\mathbb{U}$, it is practical to design a global invariant map $\mathcal{I}$ w.r.t. the symmetry group of interest $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \times \mathfrak{G}_2$, due to the predictable geometric symmetries between the input image and deep feature map (at each intermediate layer) guaranteed by Properties 1 and 2. More specifically, with the Definition 4, we assume that there exists a $\mathcal{I}$ such that $\mathbb{I}(\mathfrak{g}'_0 M) = \mathbb{I}(M)$ for any $\mathfrak{g}_0 \in \mathfrak{G}_0$ and $M \in X$, i.e., invariance holds on one layer, where $\mathfrak{g}'$ is a predictable operation corresponding to $\mathfrak{g}$ and $\mathbb{U}$. Then we have following invariance:*

$$\mathbb{I}(\mathfrak{g}'_0 M)_{[L]} \equiv \mathbb{I} M_{[L]}, \quad (20)$$

holds for any composition length $L \geq 1$.



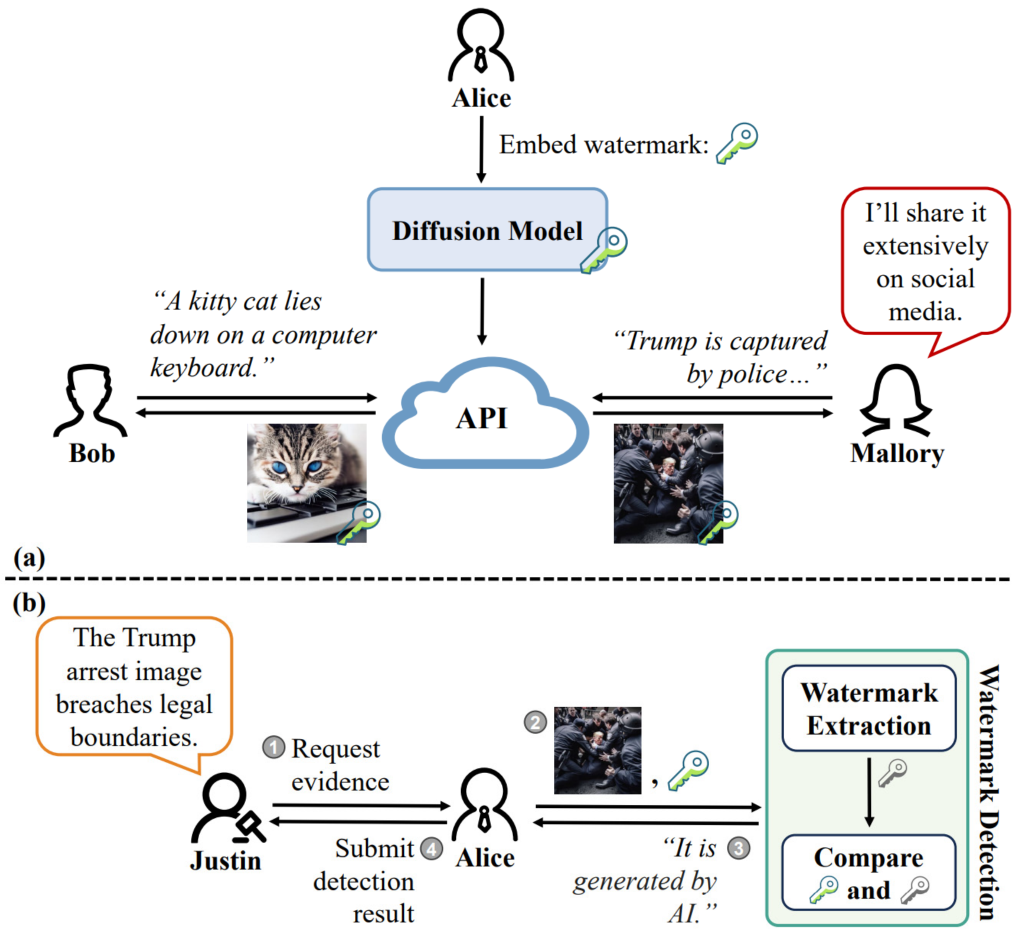


AIGC Detection

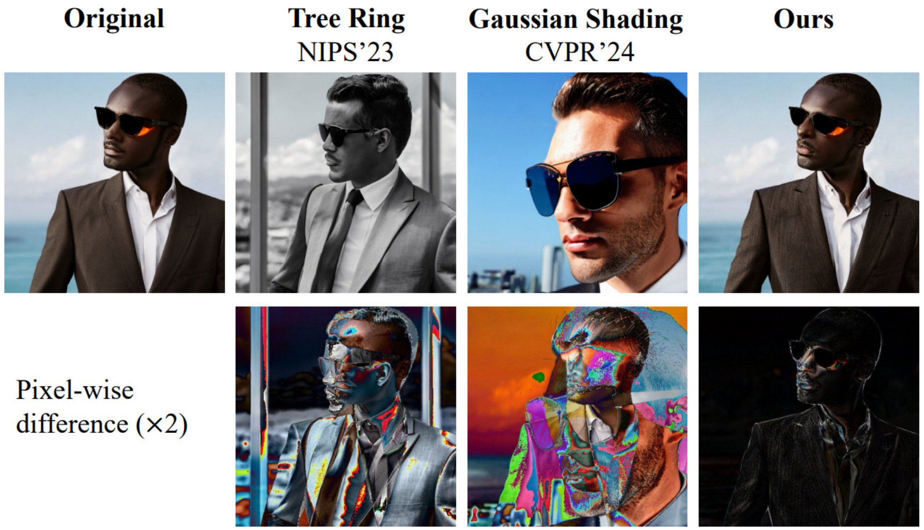
- Such efforts led to a new theory of hierarchical invariants, with better trade-off between invariance and discriminability than traditional invariants and CNN in larger-scale vision tasks and AI-generated forgery forensics, also showing explainability and efficiency benefits.

Method	Train./Test. = 5/5			Train./Test. = 1/9		
	Pre.	Rec.	F1	Pre.	Rec.	F1
Classical:						
Cosine NN	0.00	0.00	0.00	0.00	0.00	0.00
Cosine SVM	94.95	94.57	94.76	94.36	91.06	92.68
Wavelet NN	48.70	94.17	64.20	48.69	94.13	64.18
Wavelet SVM	94.03	94.57	94.30	83.55	93.48	88.24
Kraw. NN	0.00	0.00	0.00	0.00	0.00	0.00
Kraw. SVM	75.24	74.77	75.00	71.56	68.57	70.03
Learning:						
SimpleNet	61.79	40.70	49.08	56.40	60.48	58.37
AlexNet	80.76	77.63	79.16	71.83	72.50	72.17
VGGNet	84.75	86.67	85.70	72.45	72.37	72.41
GoogLeNet	74.15	80.40	77.15	67.84	68.83	68.33
ResNet	85.10	83.03	84.06	76.88	73.67	75.24
DenseNet	86.83	85.23	86.02	76.84	75.37	76.10
InceptionNet	82.69	86.63	84.62	68.62	68.56	68.59
MobileNet	81.54	82.47	82.00	68.52	68.57	68.55
Invariant:						
Scatter. NN	83.68	83.73	83.71	79.37	79.70	79.53
Scatter. SVM	90.31	85.17	87.67	85.28	79.70	82.40
HIR NN	96.79	96.47	96.63	95.66	93.04	94.33
HIR SVM	96.92	96.37	96.64	95.21	94.26	94.73

AIGC Watermarking



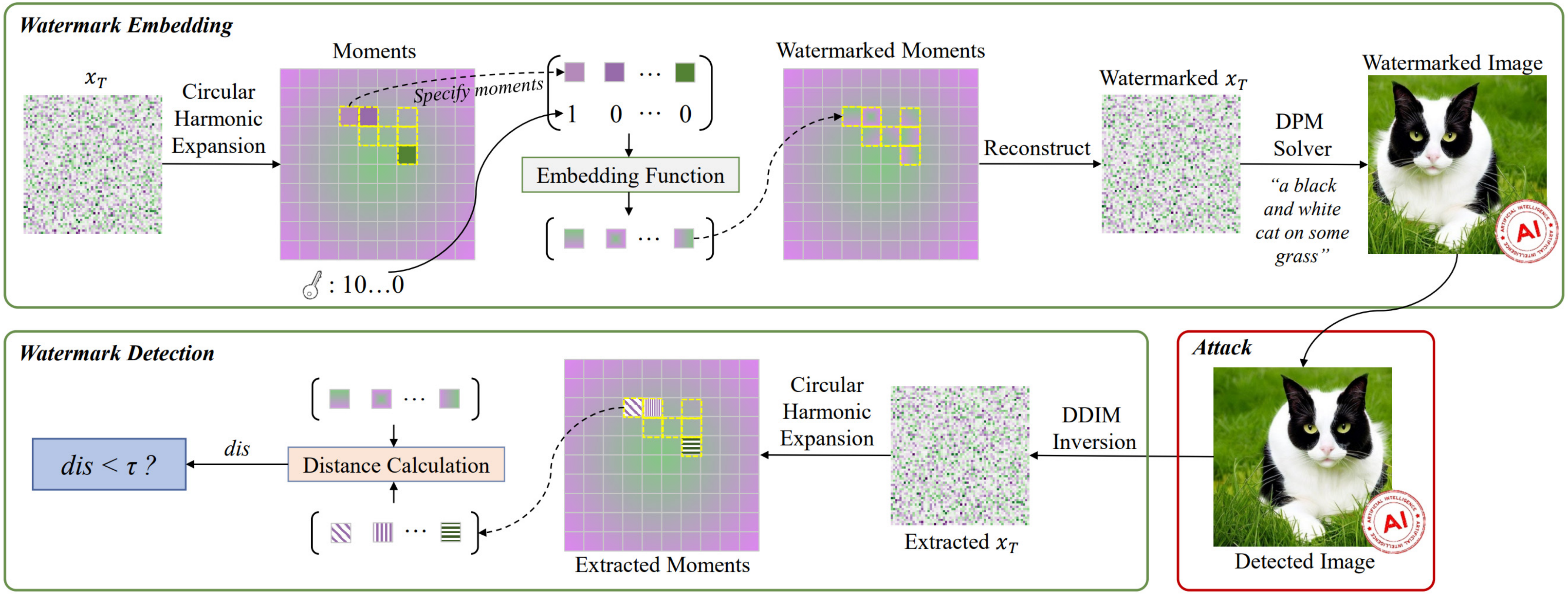
AIGC Watermarking



Is there a balance between robustness and imperceptibility?

	Passive detection	Pixel-level watermarking			Content-level watermarking			
Method	AEROBLADE [32] *CVPR'24*	DwtDct [7] *DW&S'07*	RivaGAN [46] *Arxiv'19*	Stable Signature [12] *ICCV'23*	Tree Ring [42] *NIPS'23*	Gaussian Shading [44] *CVPR'24*	AquaLoRA [10] *ICML'24*	Ours
Detection confidence		✓	✓	✓	✓	✓	✓	✓
Invariance and robustness					✓	✓	✓	✓
Imperceptibility	✓	✓	✓	✓			✓	✓
Plug-and-play	✓	✓	✓		✓	✓		✓

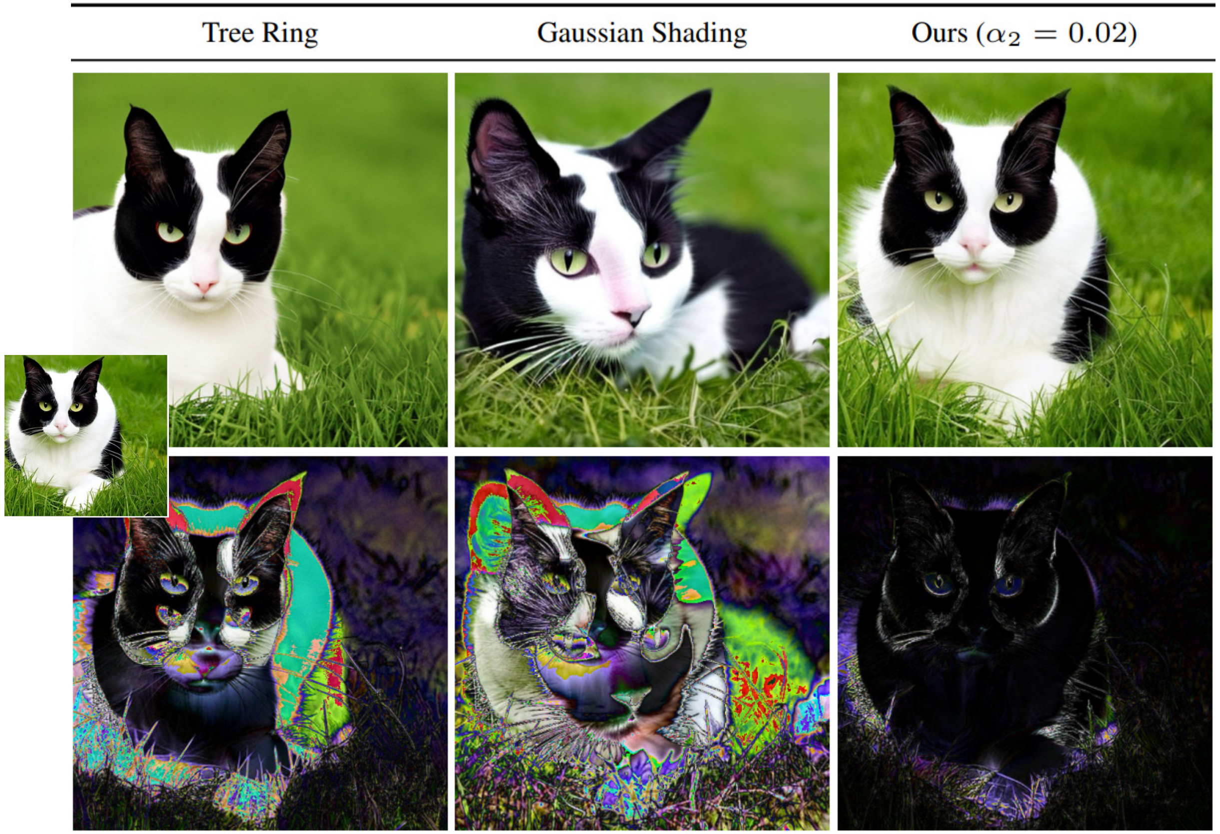
AIGC Watermarking



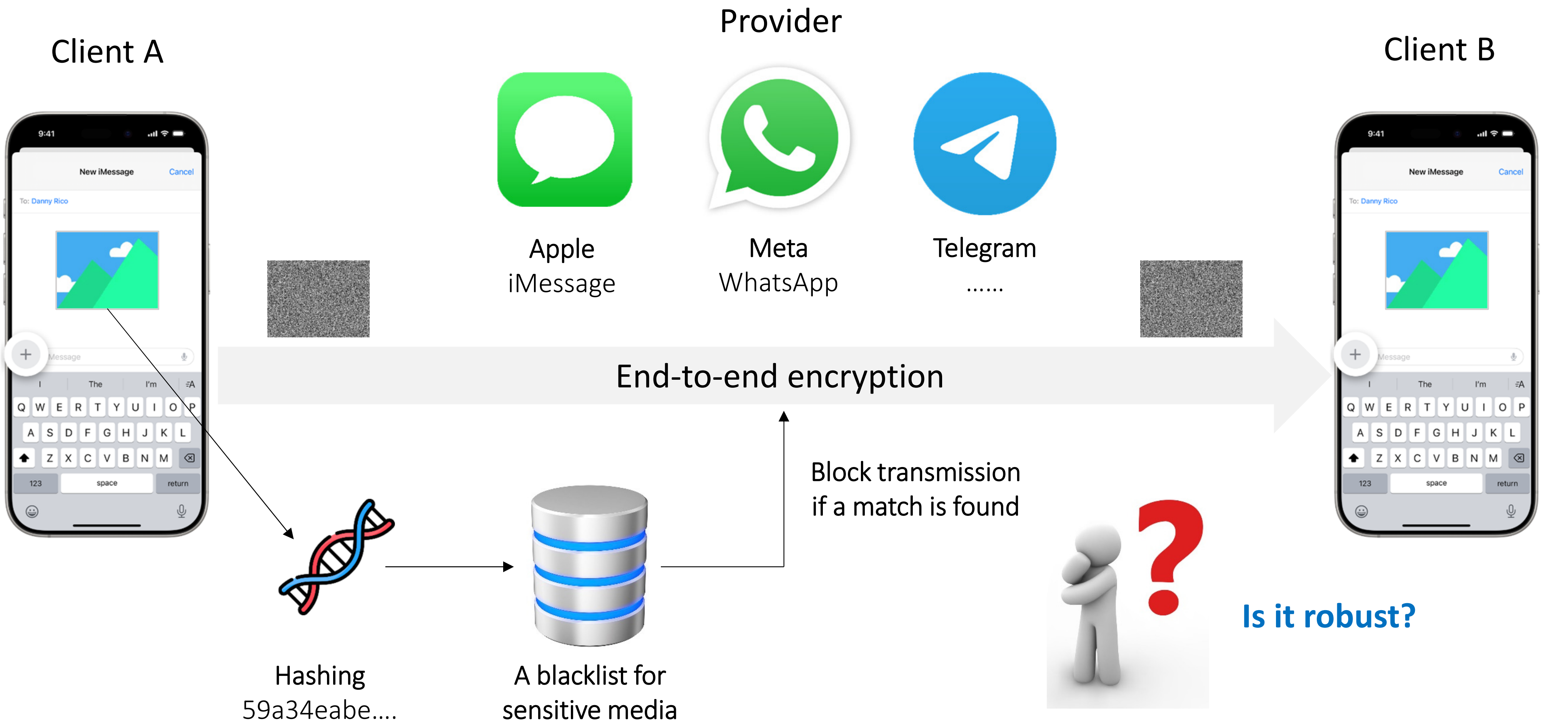
AIGC Watermarking

	VAE based		DM based	
Method	Bmshj'18	Cheng'20	SDv2.1	**Average**
Pixel-level				*0.165*
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
Content-level				*0.987*
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

	Metrics		
Method	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



AIGC Hashing

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

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

with notations of

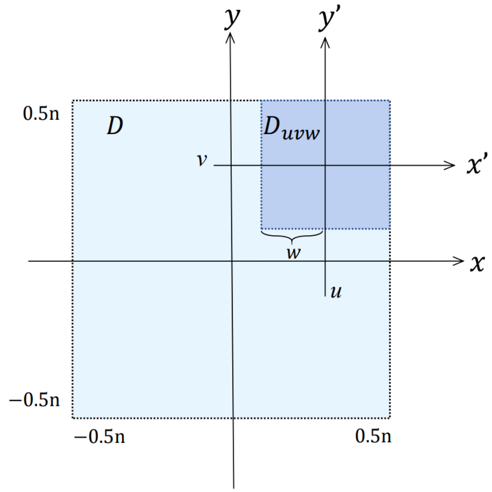
$$\mathcal{F}(X) = \langle X, V_{nm}^{uvw}\rangle = \iint_D (V_{nm}^{uvw}(x,y))^* X(x,y)dxdy, \tag{4}$$

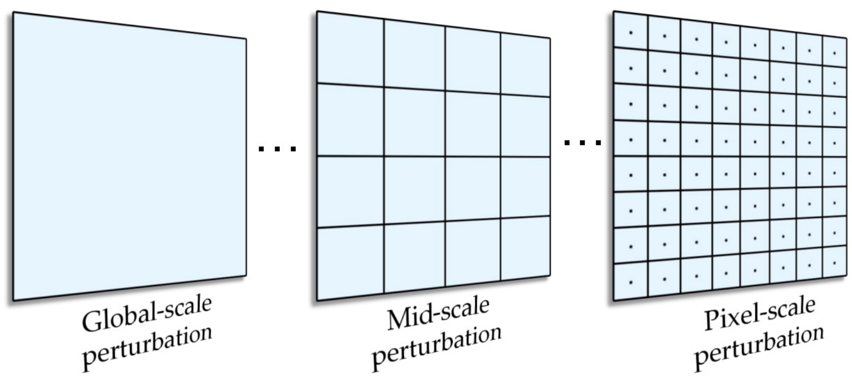
and

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

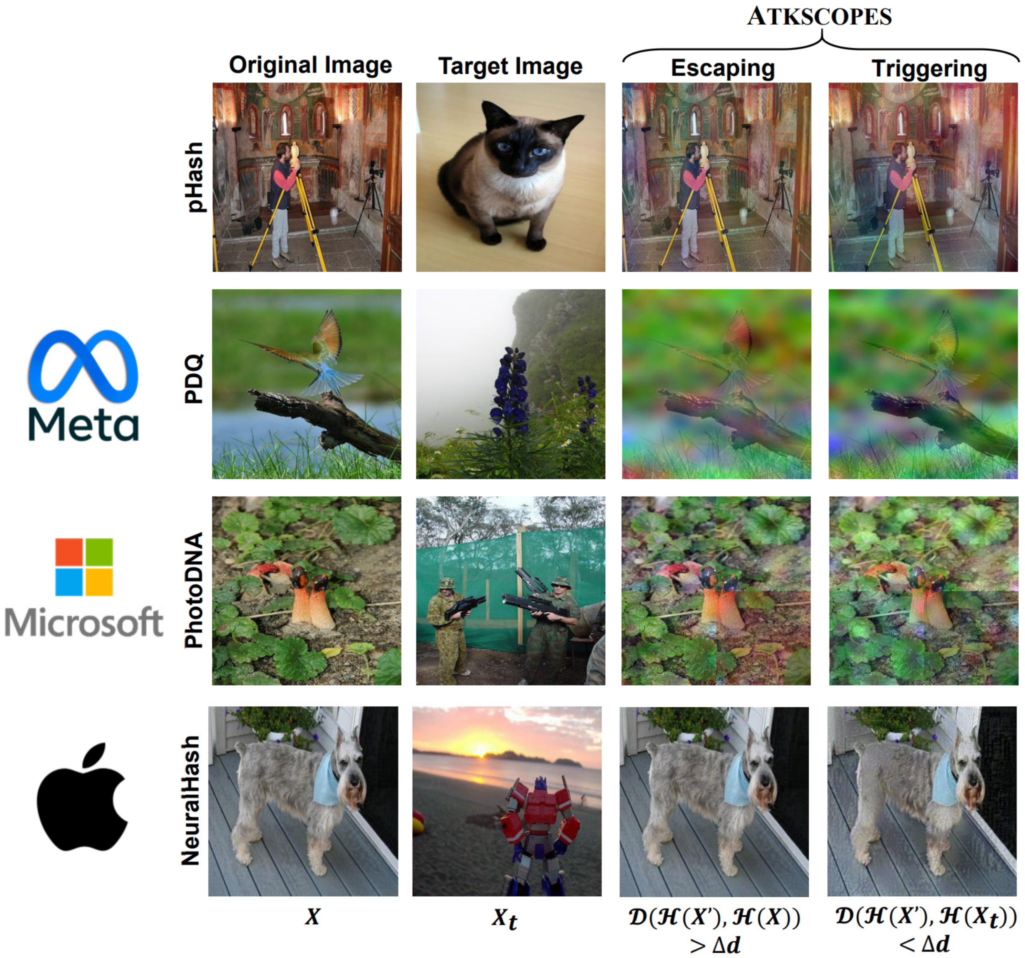
where $\mathcal{F}$ denotes the local orthogonal transformation [39], with image $X(x,y)$ on domain $(x,y)\in D$. The local orthogonal basis function V_{nm}^{uvw} is defined on the domain D_{uvw} with the order parameters $(n,m)\in\mathbb{Z}^2$, converting D to D_{uvw} by the translation offset (u,v) 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:

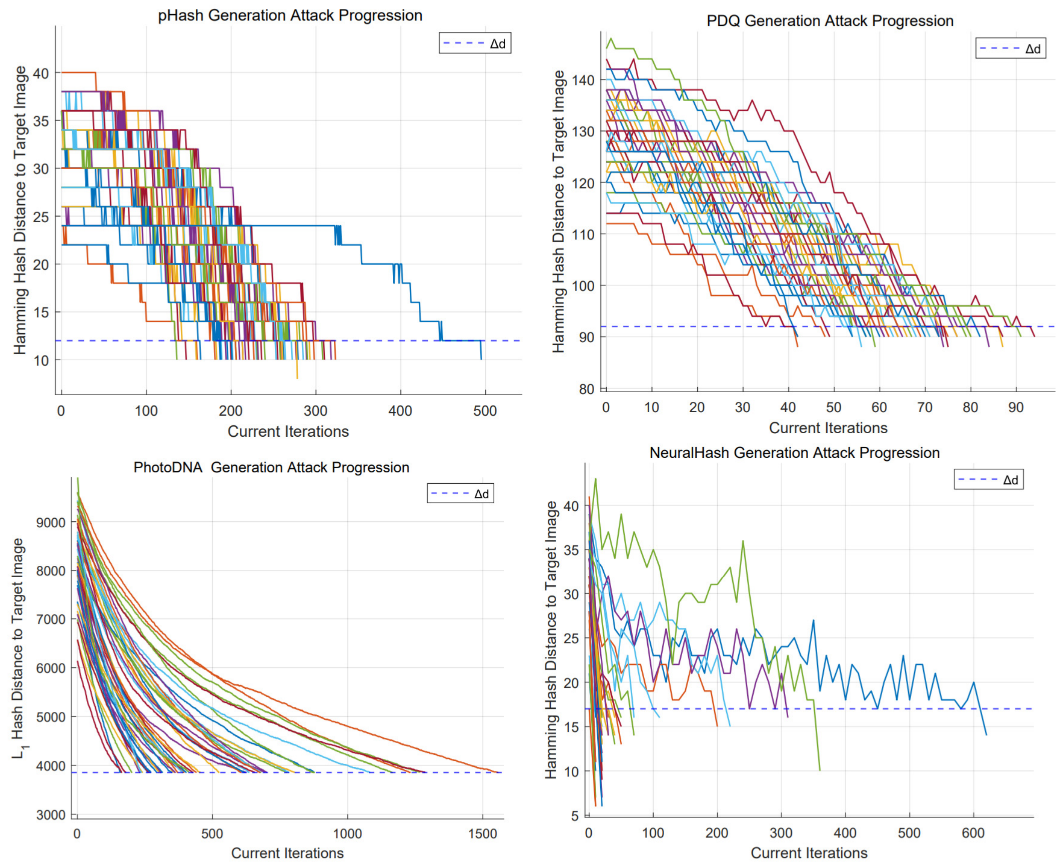
$$V_{nm}^{uvw}(x,y) = V_{nm}(x',y') = V_{nm}\left(\frac{x-u}{w}, \frac{y-v}{w}\right). \tag{6}$$





AIGC Hashing





§ Q & A

by Shuren Qi
shurenqi@cuhk.edu.hk | shurenqi.github.io