Scalable Priors for Foundation Models

Sangwoo Mo

Research Fellow,
University of Michigan

aistudio.google.com

Gemini 1.5 Pro: Sherlock Jr

Save

Insert: Image Video File Folder

Chat

Sherlock_Jr.mp4 696,161 tokens

Find the moment when a piece of paper is removed from the person's pocket and tell me some key information on it, with the timecode.

Model

At 12:01, a piece of paper is removed from the person's pocket. It is a pawn ticket from I. Goldman & Co. Pawn Brokers, dated 10/23/1924, for a watch and chain pawned for $4.00. The ticket is signed by Will Smith.

Type something

Preview 696263 / 1048576

Sherlock_Jr.mp4





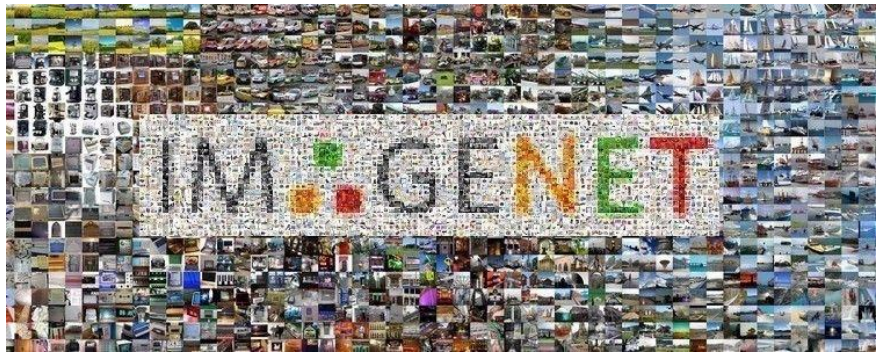




Is AI ready for real-world problems?

Is AI ready for real-world problems?

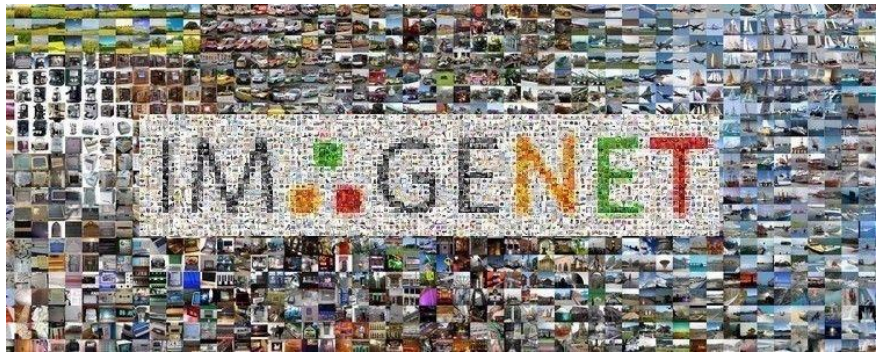
AI fails in practice 😕



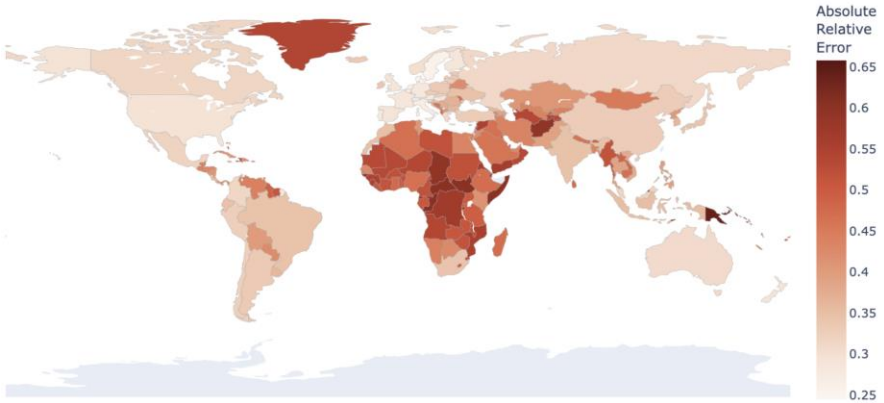


Is AI ready for real-world problems?

AI fails in practice 😕



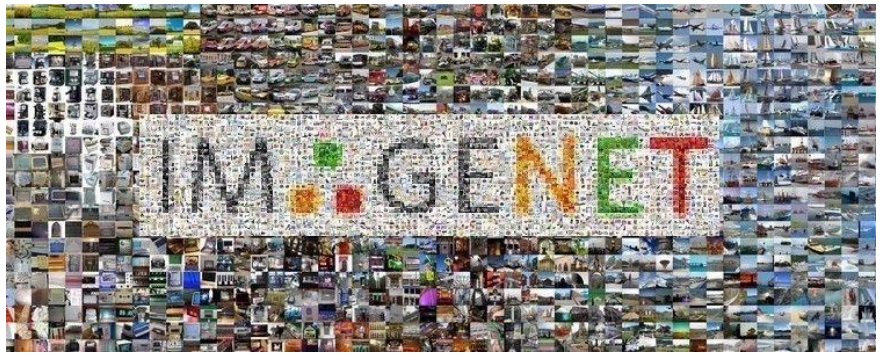
✅

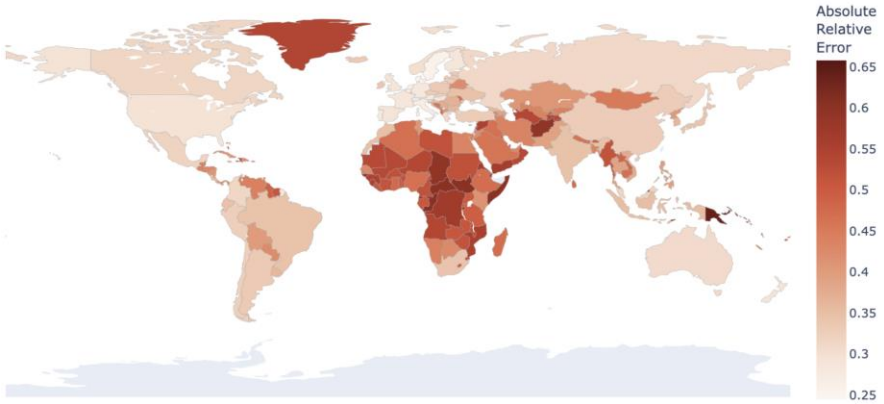


❌

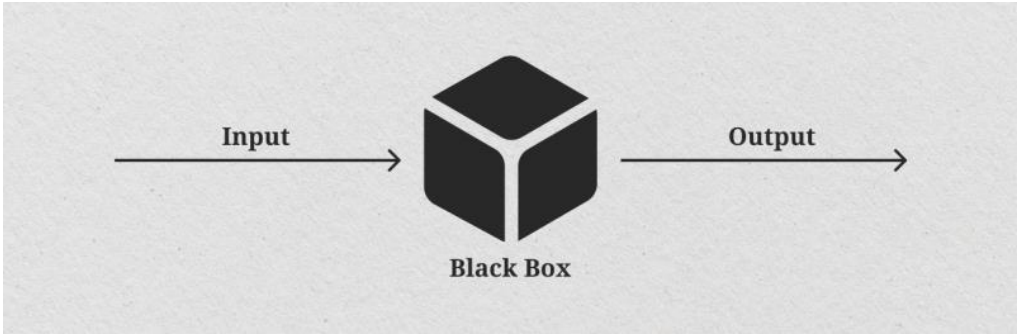
Is AI ready for real-world problems?

AI fails in practice 😕



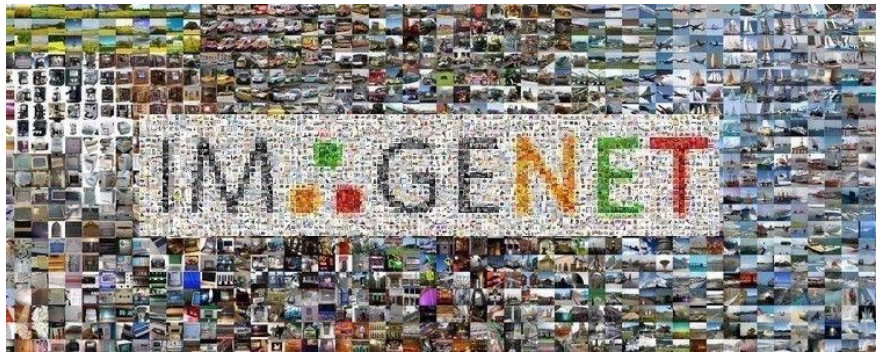


AI is black-box 😕

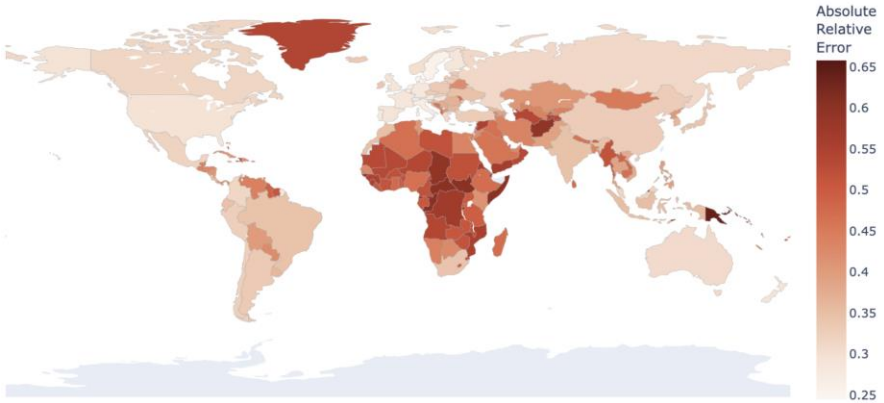


My Research: Scalable Priors for Foundation Models

AI fails in practice 😕

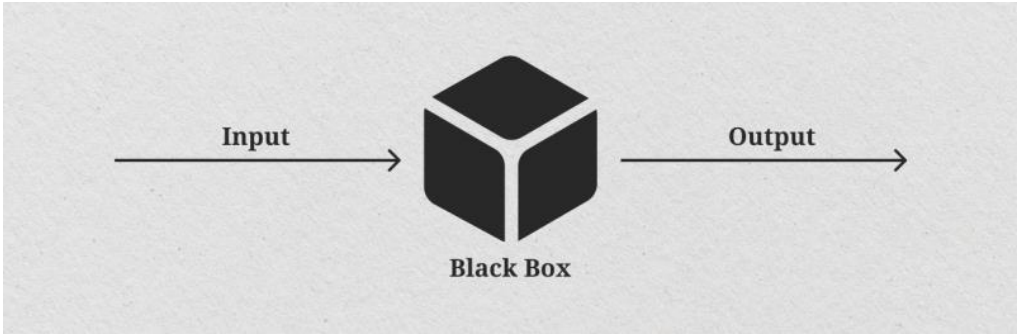


✅



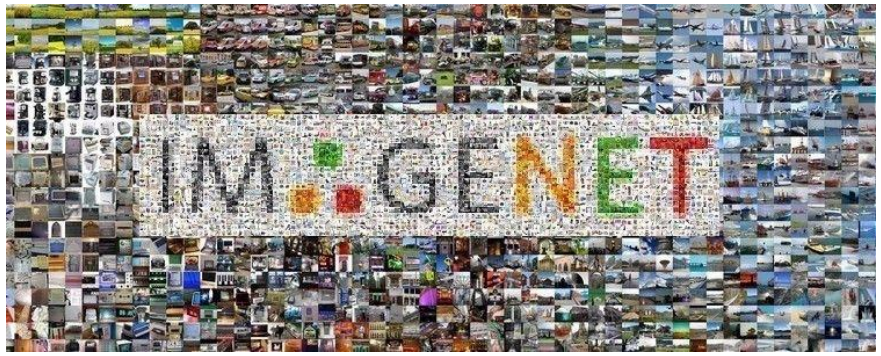
❌

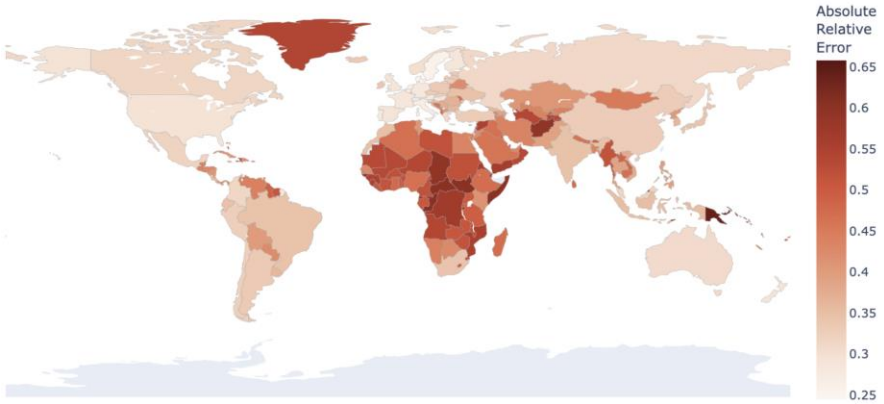
AI is black-box 😕



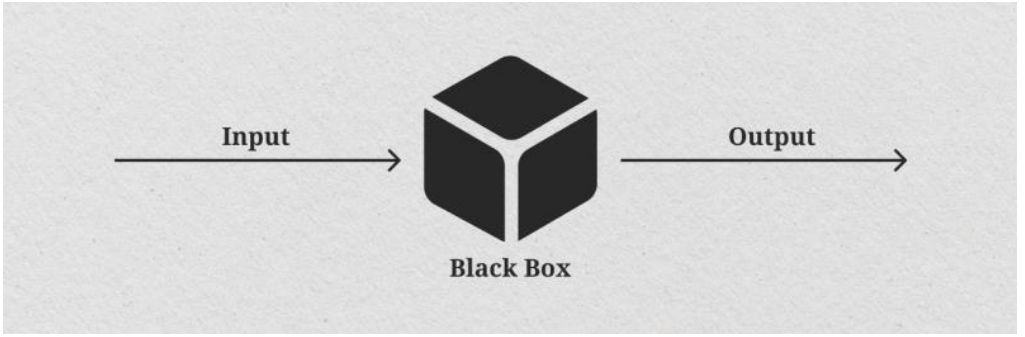
My Research: **Scalable Priors for Foundation Models**

AI fails in practice 😕





AI is black-box 😕

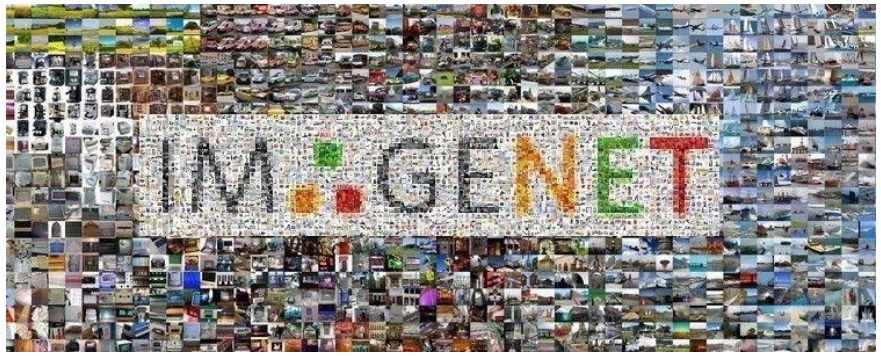


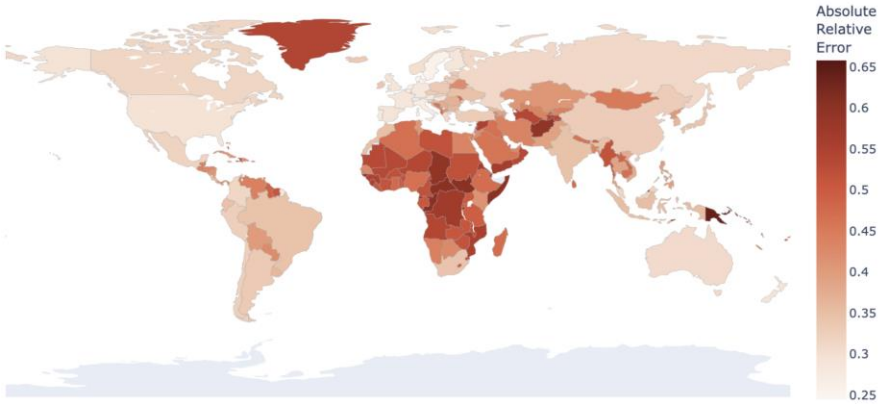
Statistical priors in data manifolds

→ Generalization & Reliability

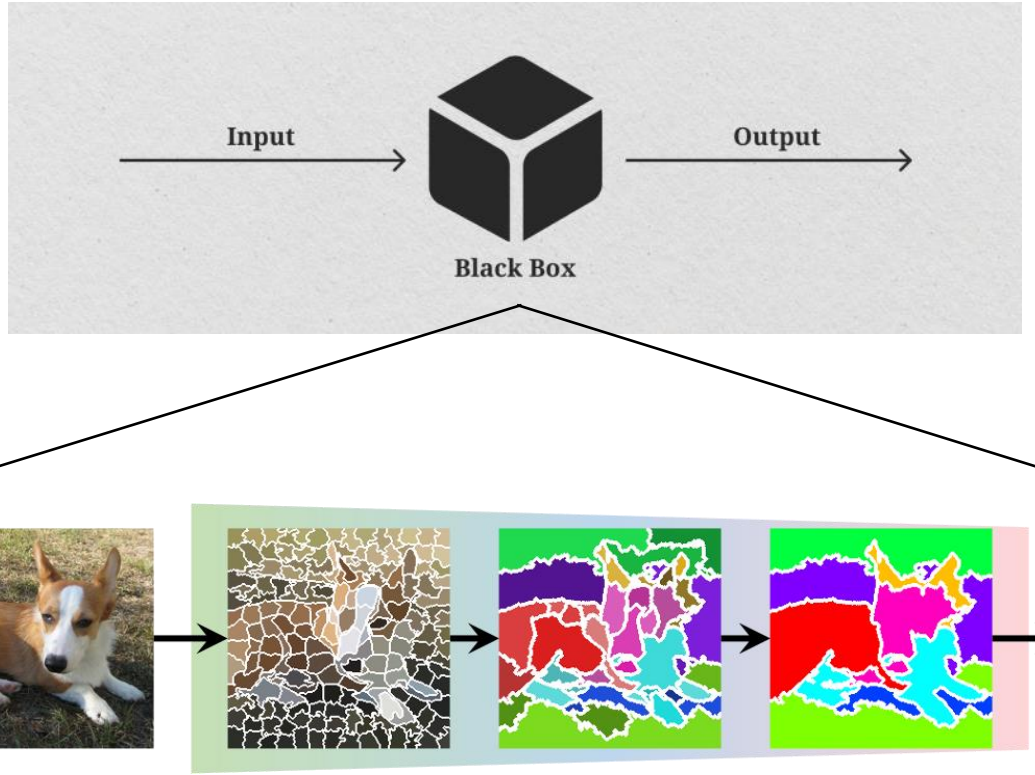
My Research: **Scalable Priors for Foundation Models**

AI fails in practice 😕





AI is black-box 😕



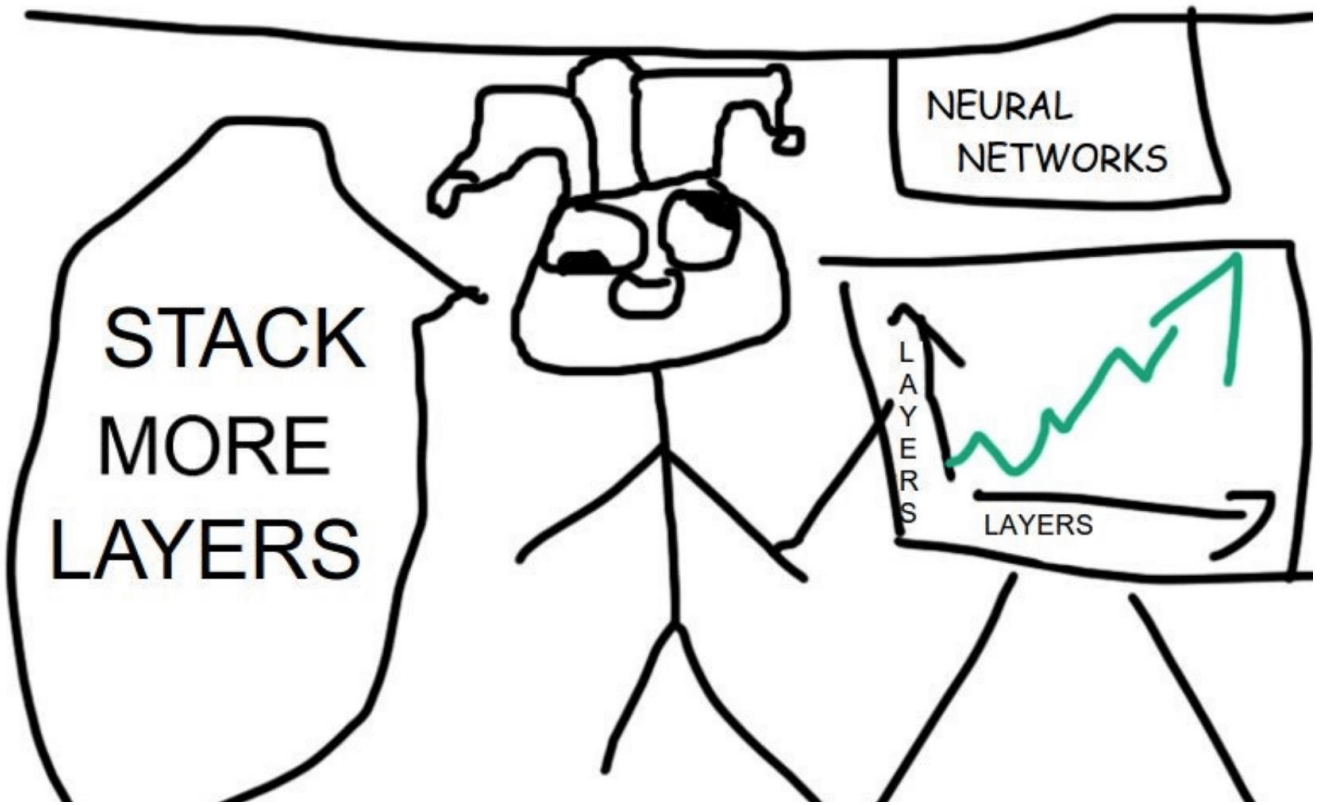
Statistical priors in data manifolds

→ Generalization & Reliability

Structural priors in data instances

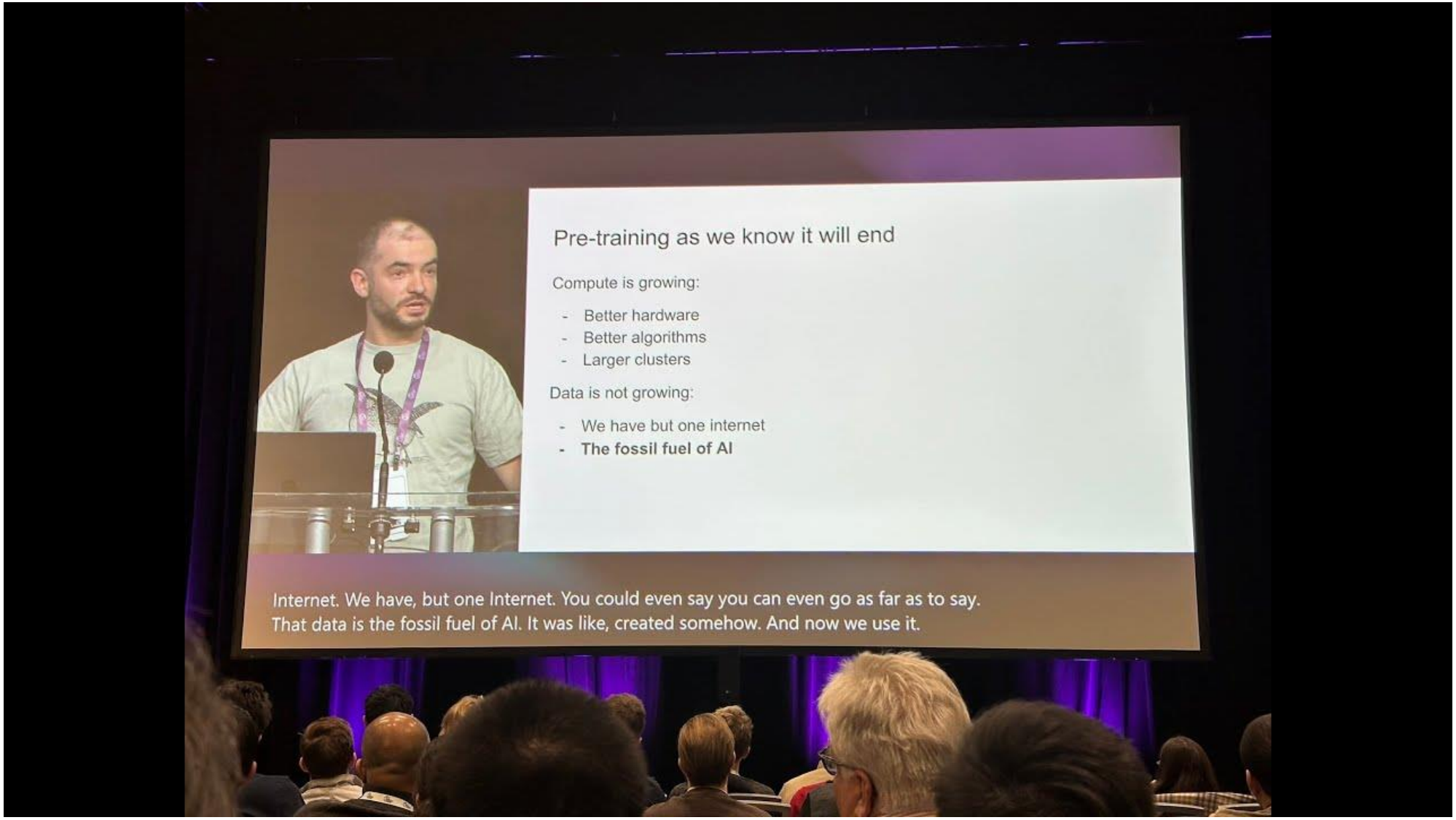
→ Performance & Interpretability

Q: Can Large Models Solve Everything?



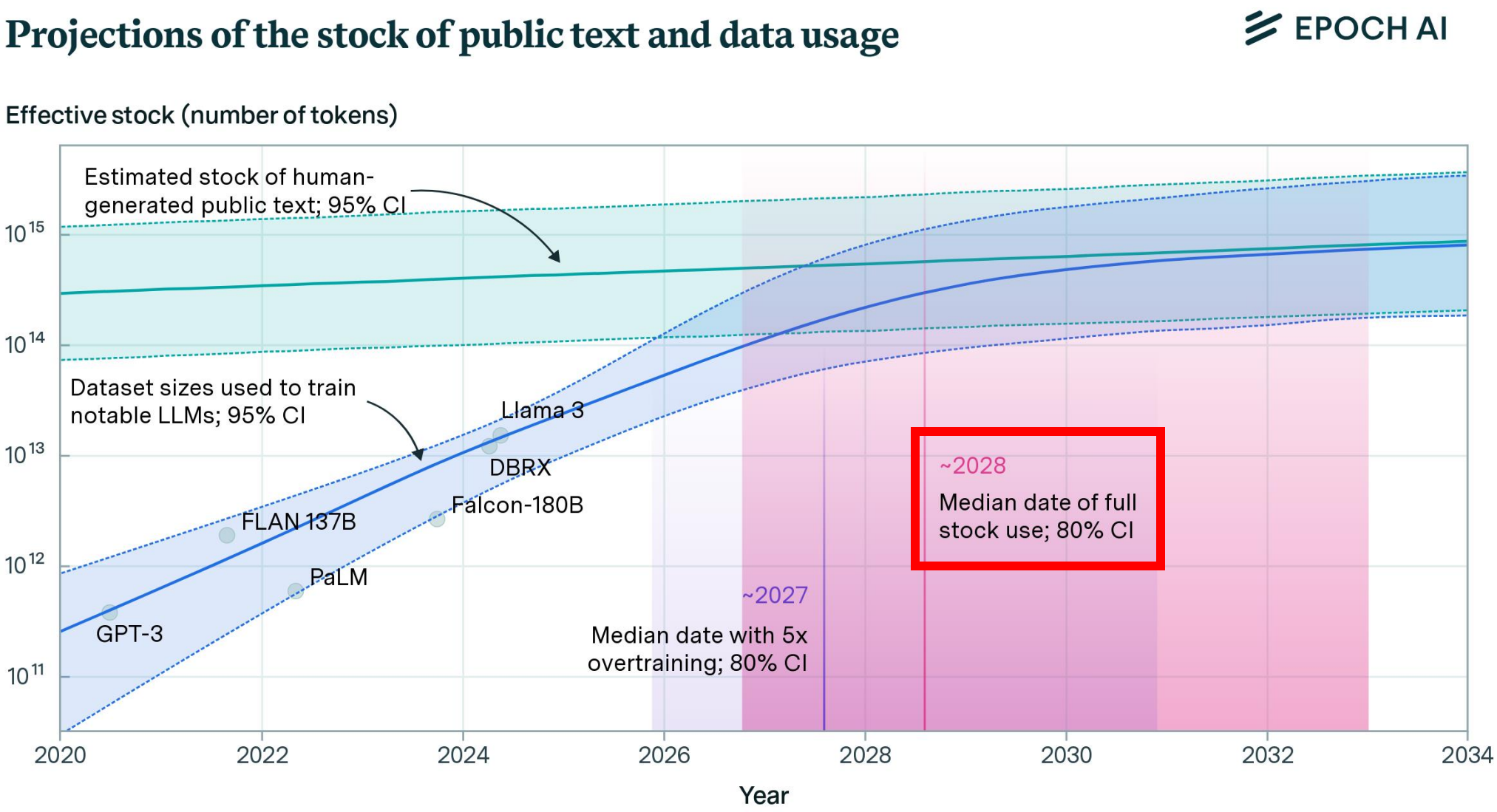
Source: https://www.reddit.com/r/ProgrammerHumor/comments/5si1f0/machine_learning_approaches

No - 1. Scaling Laws are Reaching their Limits



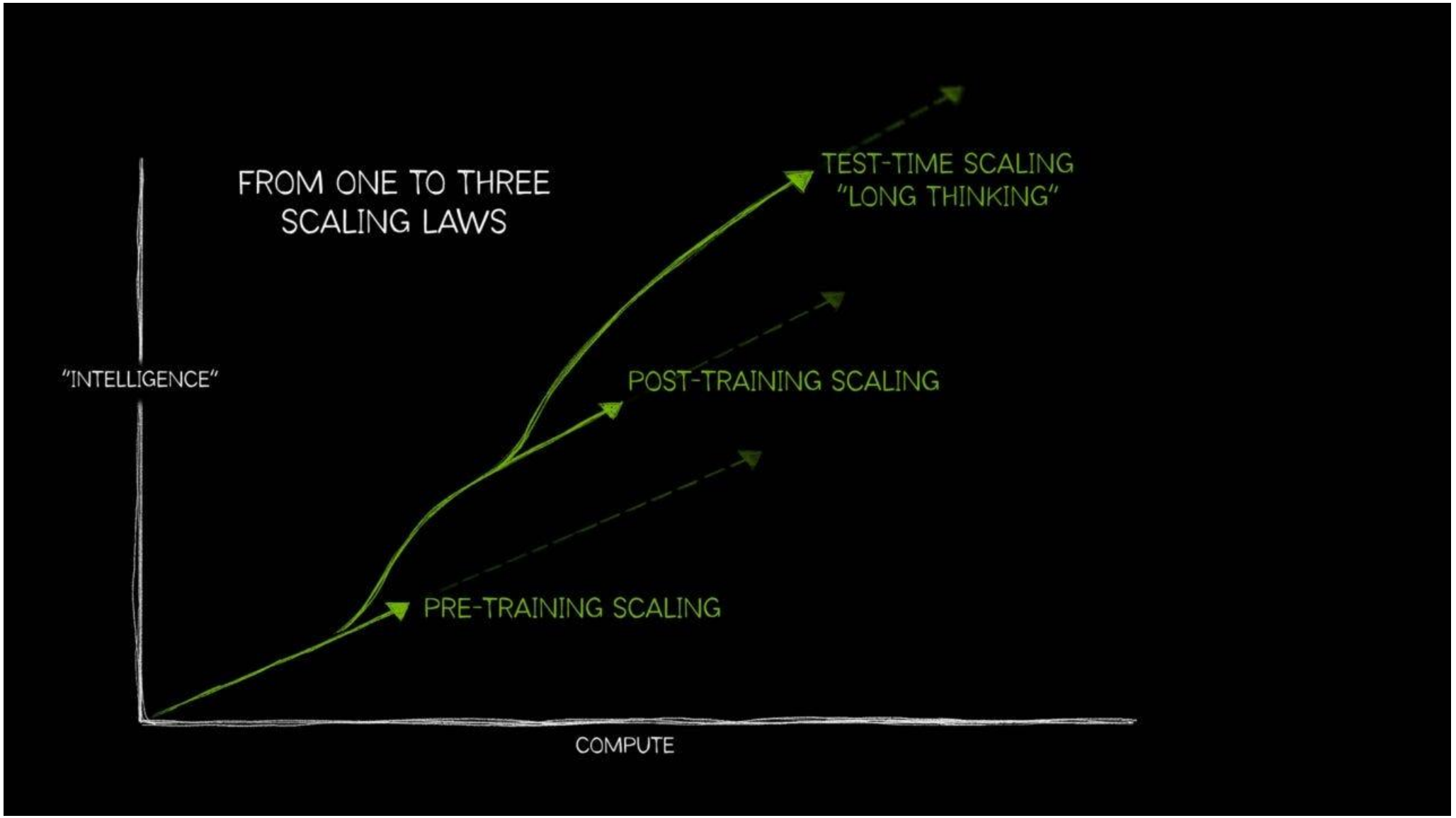
Source: https://www.reddit.com/r/singularity/comments/1hdrjvq/ilyas_full_talk_at_neurips_2024_pretraining_as_we/

No - 1. Scaling Laws are Reaching their Limits



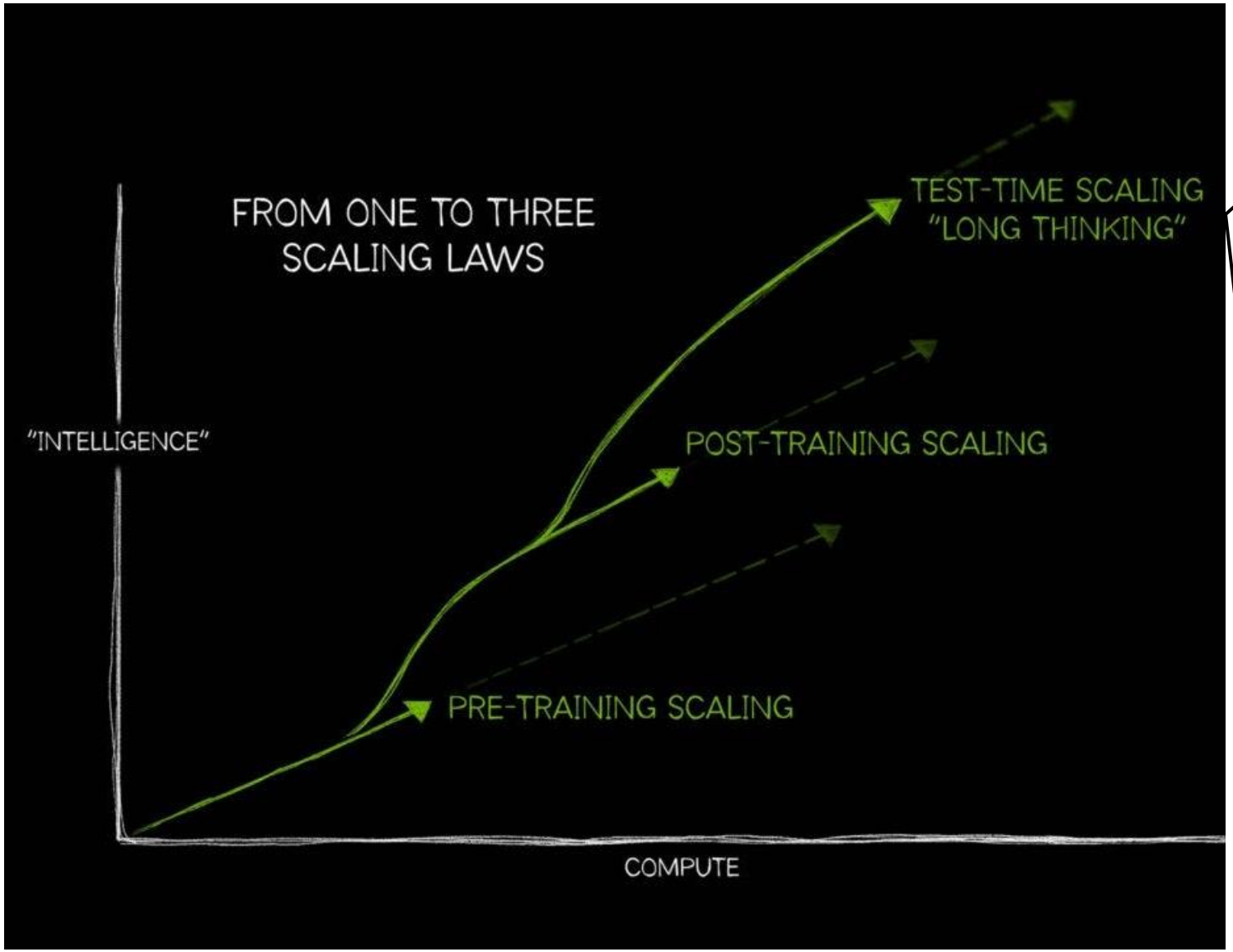
Source: https://www.reddit.com/r/singularity/comments/1hdrjvq/ilyas_full_talk_at_neurips_2024_pretraining_as_we/

No - 2. Priors Accelerate Scaling Laws



Source: https://blogs.nvidia.com/blog/ai-scaling-laws/

No - 2. Priors Accelerate Scaling Laws



Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

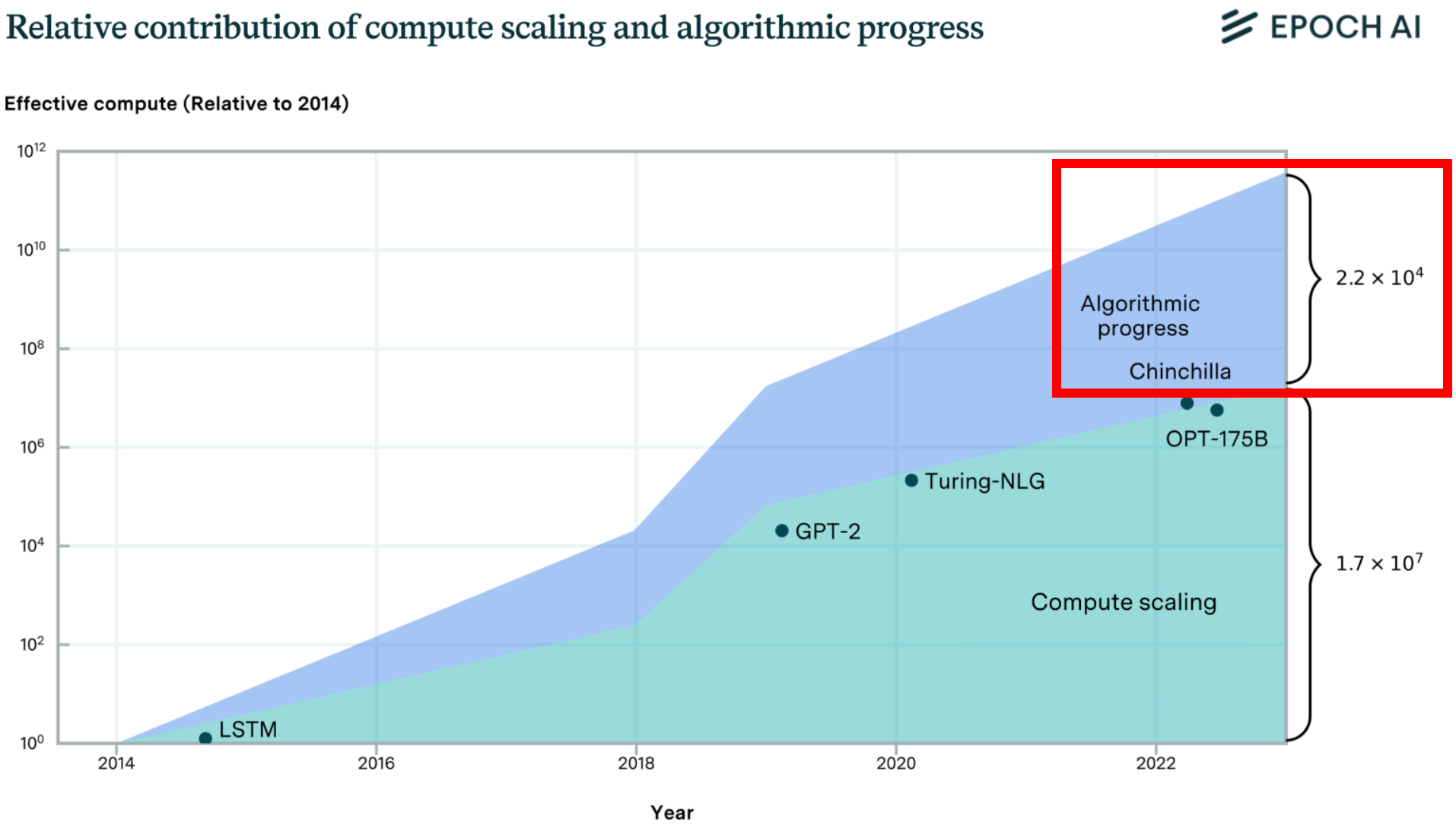
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Reasoning priors accelerate the scaling laws!

Source: https://blogs.nvidia.com/blog/ai-scaling-laws/

No - 2. Priors Accelerate Scaling Laws



Source: https://epoch.ai/blog/algorithmic-progress-in-language-models

Part 1. **Statistical Priors in Data Manifolds**

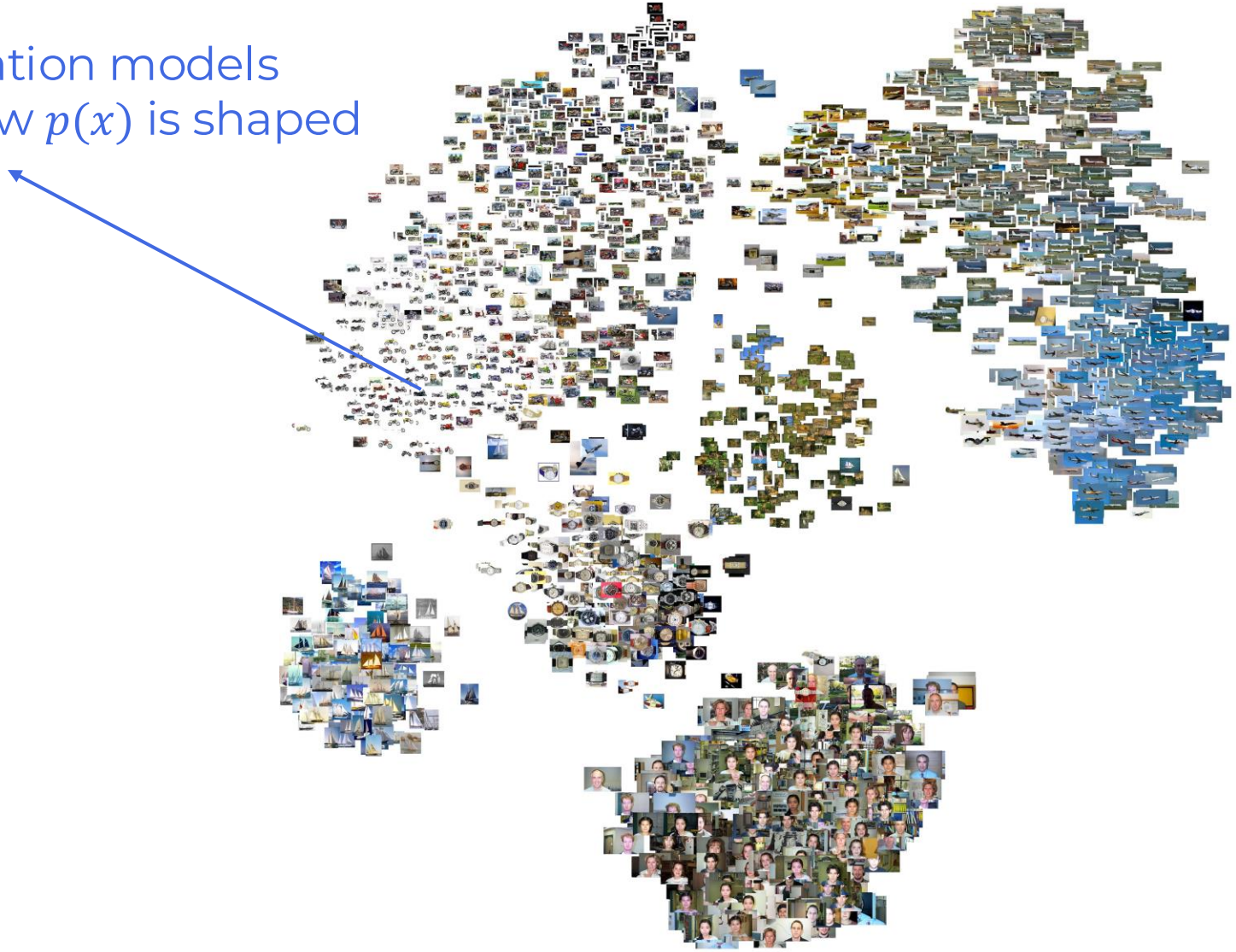
Statistical Prior: What Does the Data Manifold Look Like?



Source: https://towardsdatascience.com/visualizing-topic-models-with-scatterpies-and-t-sne-f21f228f7b02

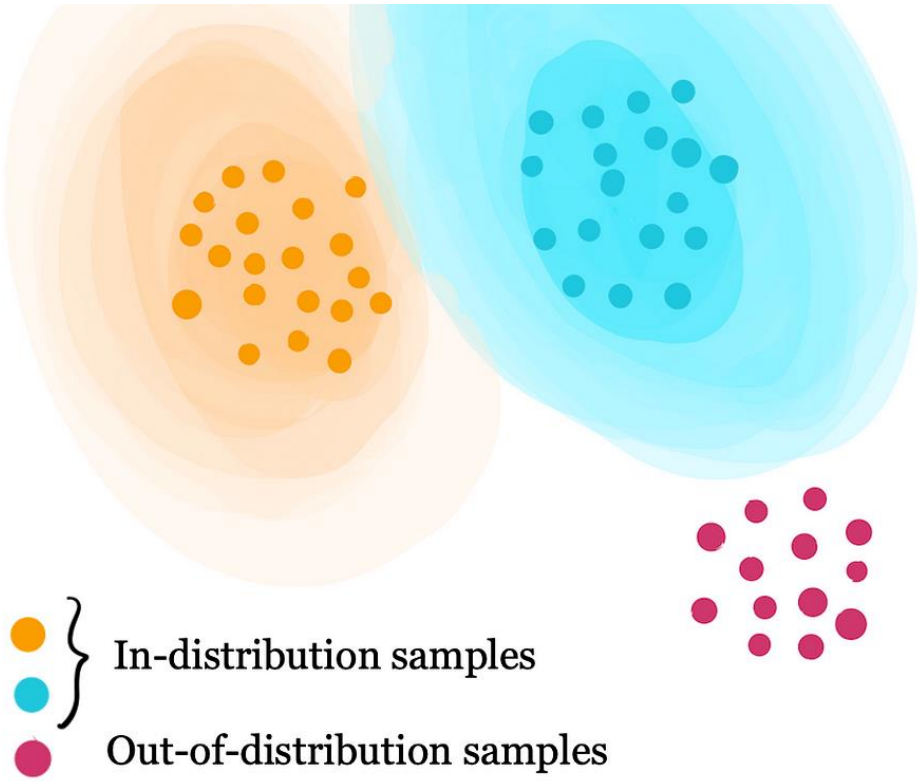
Statistical Prior: **What Does the Data Manifold Look Like?**

1. Foundation models
know how $p(x)$ is shaped



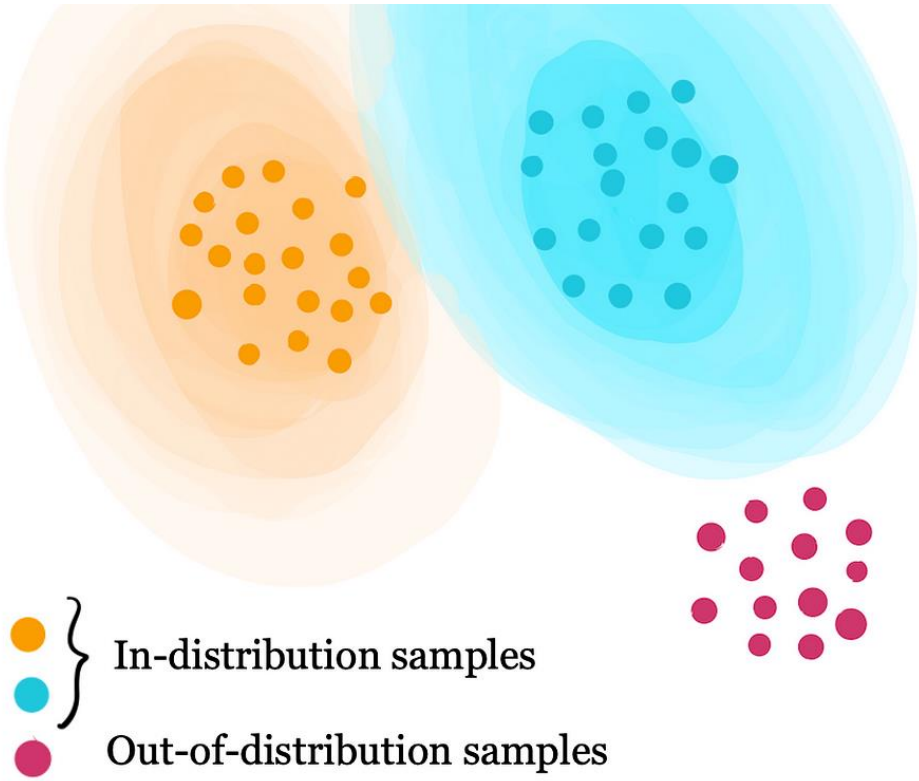
Source: https://towardsdatascience.com/visualizing-topic-models-with-scatterpies-and-t-sne-f21f228f7b02

1. Foundation models **know** how $p(x)$ is shaped



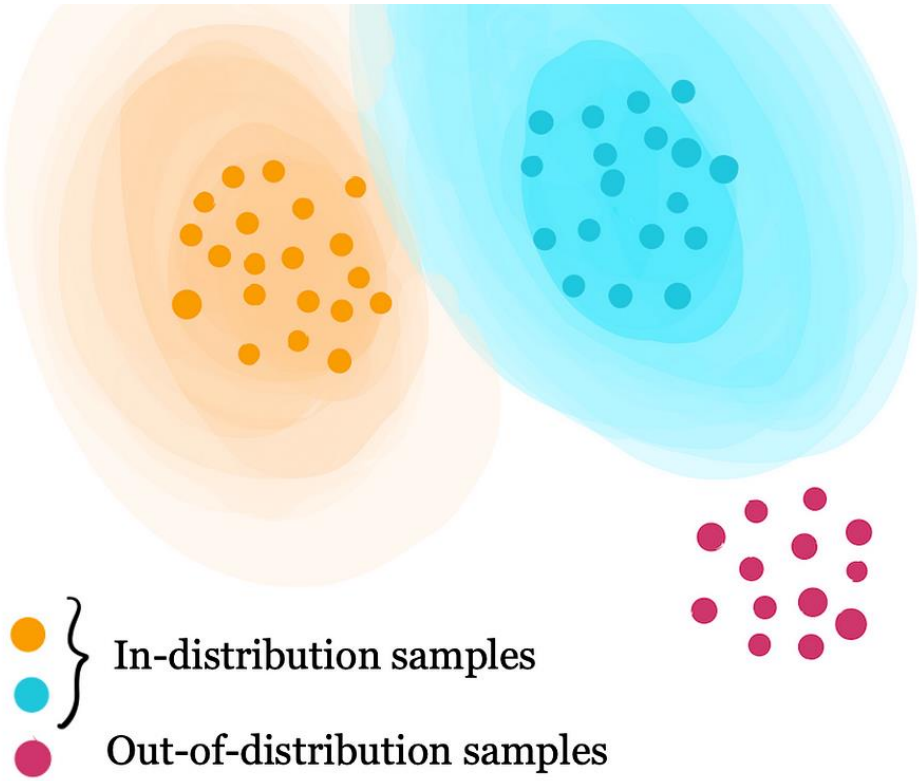
Source: https://medium.com/geekculture/out-of-distribution-detection-in-medical-ai-b638b385c2a3

1. Foundation models **know** how $p(x)$ is shaped



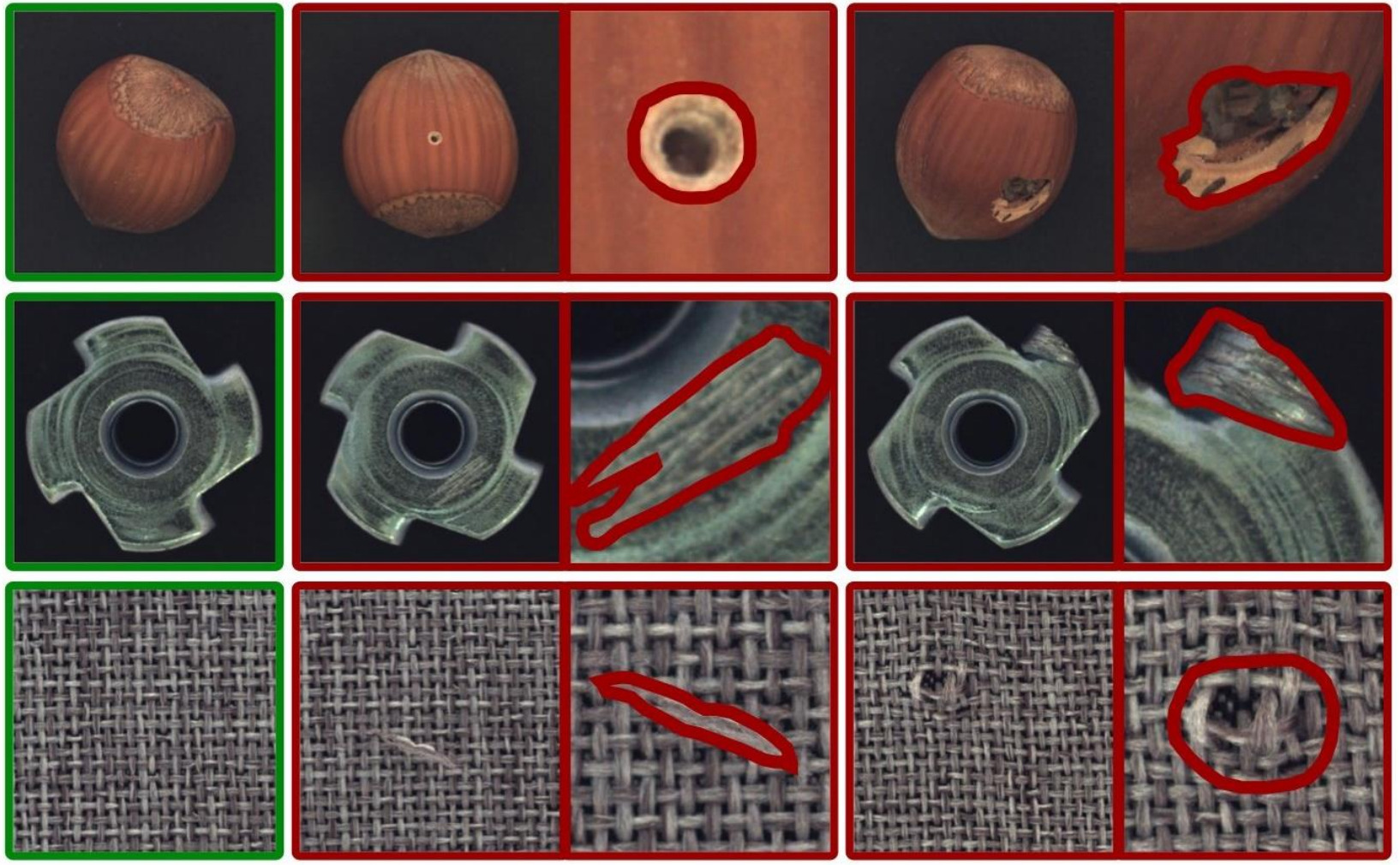
- SimCLR (contrastive learning) is an OOD detector for images
  – CSI (Tack* & Mo* et al., NeurIPS'20)
- BERT (masked language modeling) is an OOD detector for texts
  – MASKER (Moon* & Mo* et al., AAAI'21)

1. Foundation models **know** how $p(x)$ is shaped



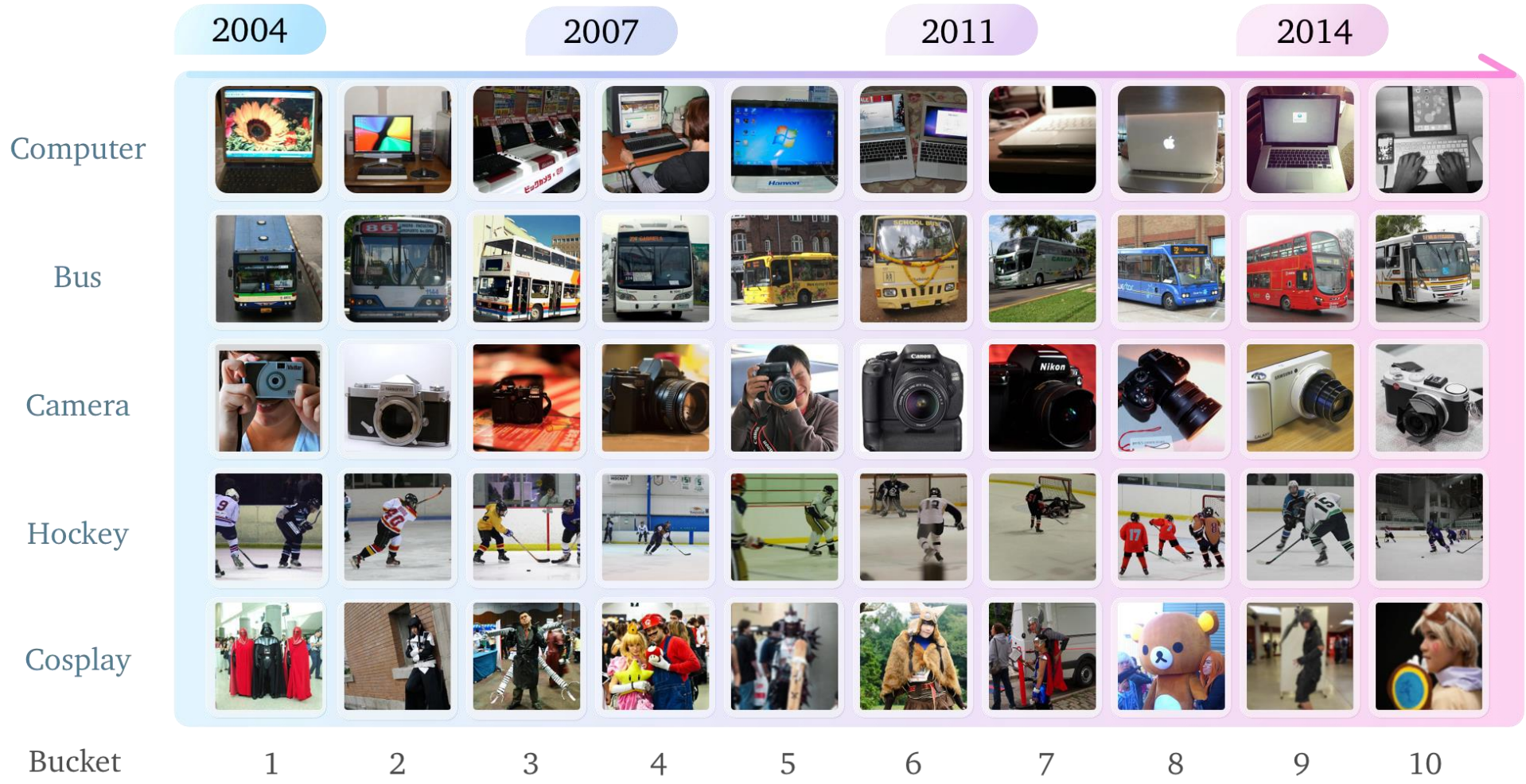
- SimCLR (contrastive learning) is an OOD detector for images
  – CSI (Tack* & Mo* et al., NeurIPS'20) → **730** citations, Qualcomm awards
- BERT (masked language modeling) is an OOD detector for texts
  – MASKER (Moon* & Mo* et al., AAAI'21) → 42 citations

Application: **Anomaly Detection in Manufacturing Data**



Source: MVTec AD dataset

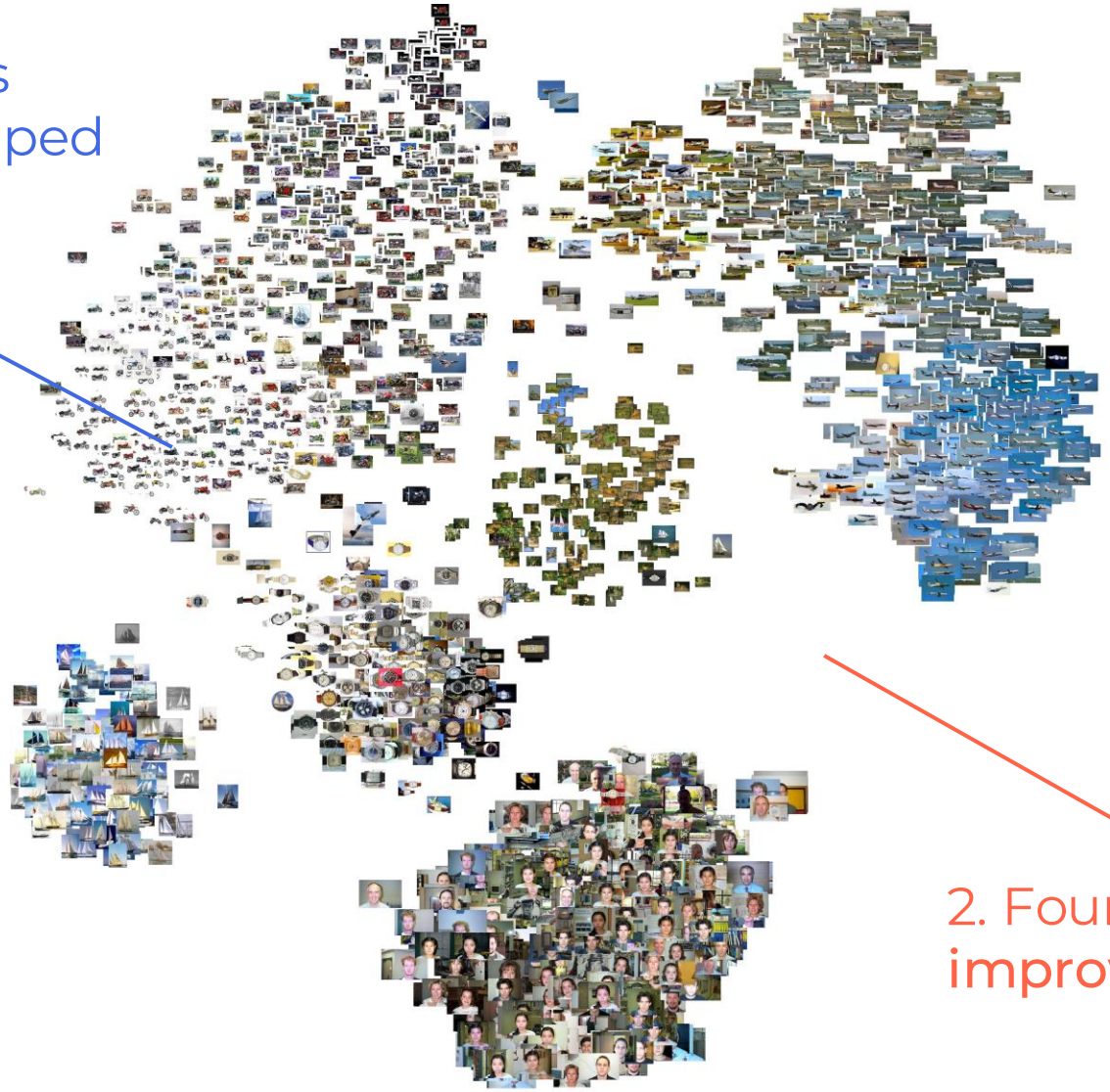
Application: Lifelong Learning in Evolving Worlds



Source: CLEAR paper

Statistical Prior: **What Does the Data Manifold Look Like?**

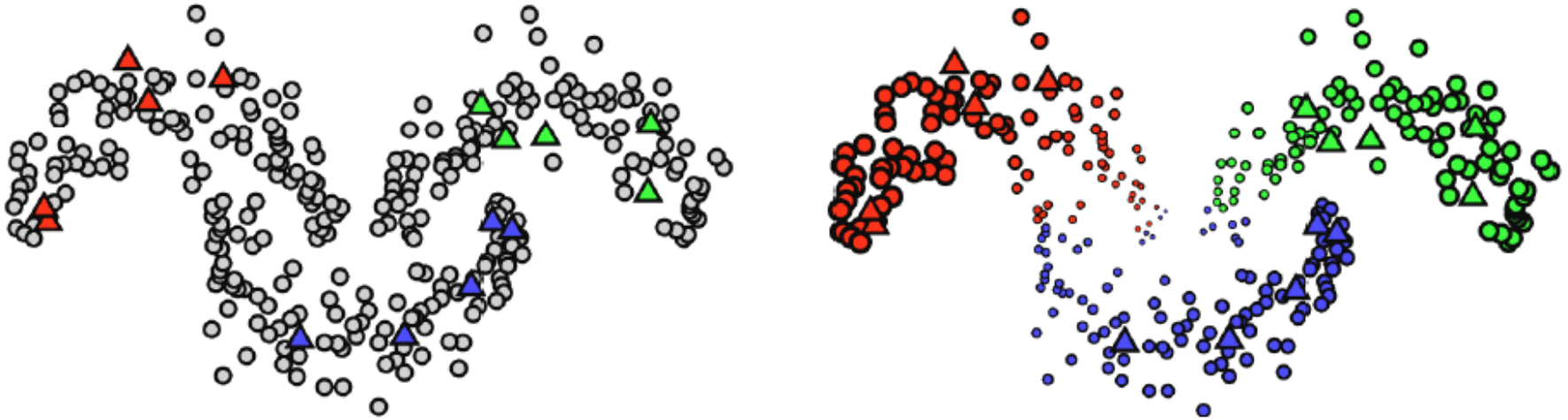
1. Foundation models **know** how $p(x)$ is shaped



2. Foundation models can be **improved** by leveraging $p(x)$

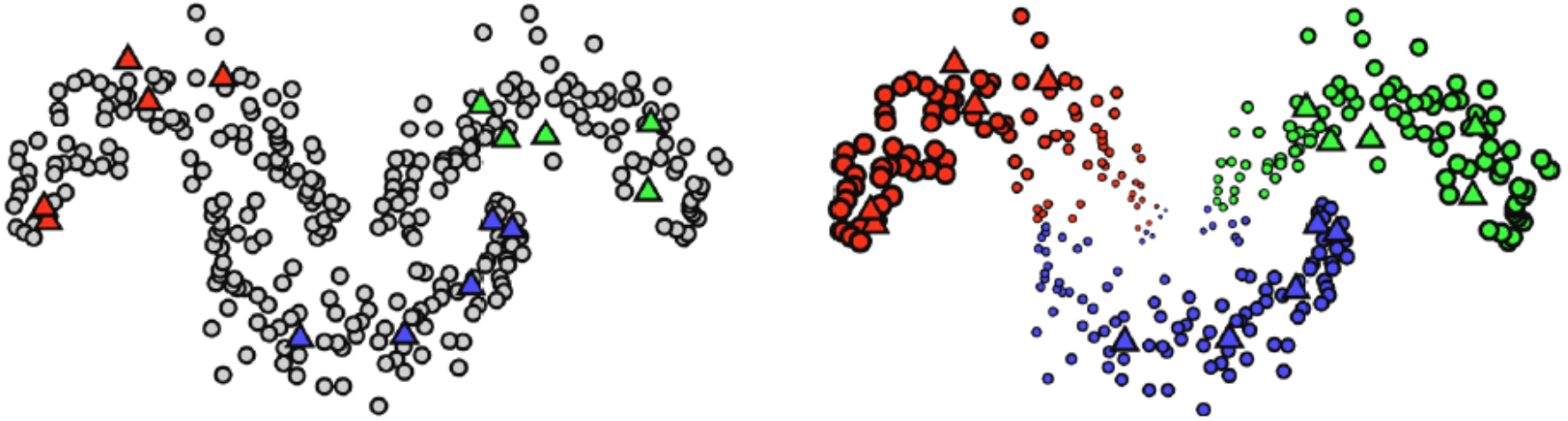
Source: https://towardsdatascience.com/visualizing-topic-models-with-scatterpies-and-t-sne-f21f228f7b02

2. Foundation models can be **improved** by leveraging $p(x)$



Source: https://sh-tsang.medium.com/review-label-propagation-for-deep-semi-supervised-learning-7b1233cc0ce

2. Foundation models can be **improved** by leveraging $p(x)$



- **Semi-supervised learning of image classifiers (DINO)**
  – RoPAWS (Mo et al., ICLR'23)
- **Semi-supervised learning of vision-language models (CLIP)**
  – S-CLIP (Mo et al., NeurIPS'23)

Application: Foundation Models for Various Domains



Source: FAME-ViL paper

S-CLIP: Semi-supervised Vision-Language Learning using Few Specialist Captions



Sangwoo Mo



Minkyu Kim



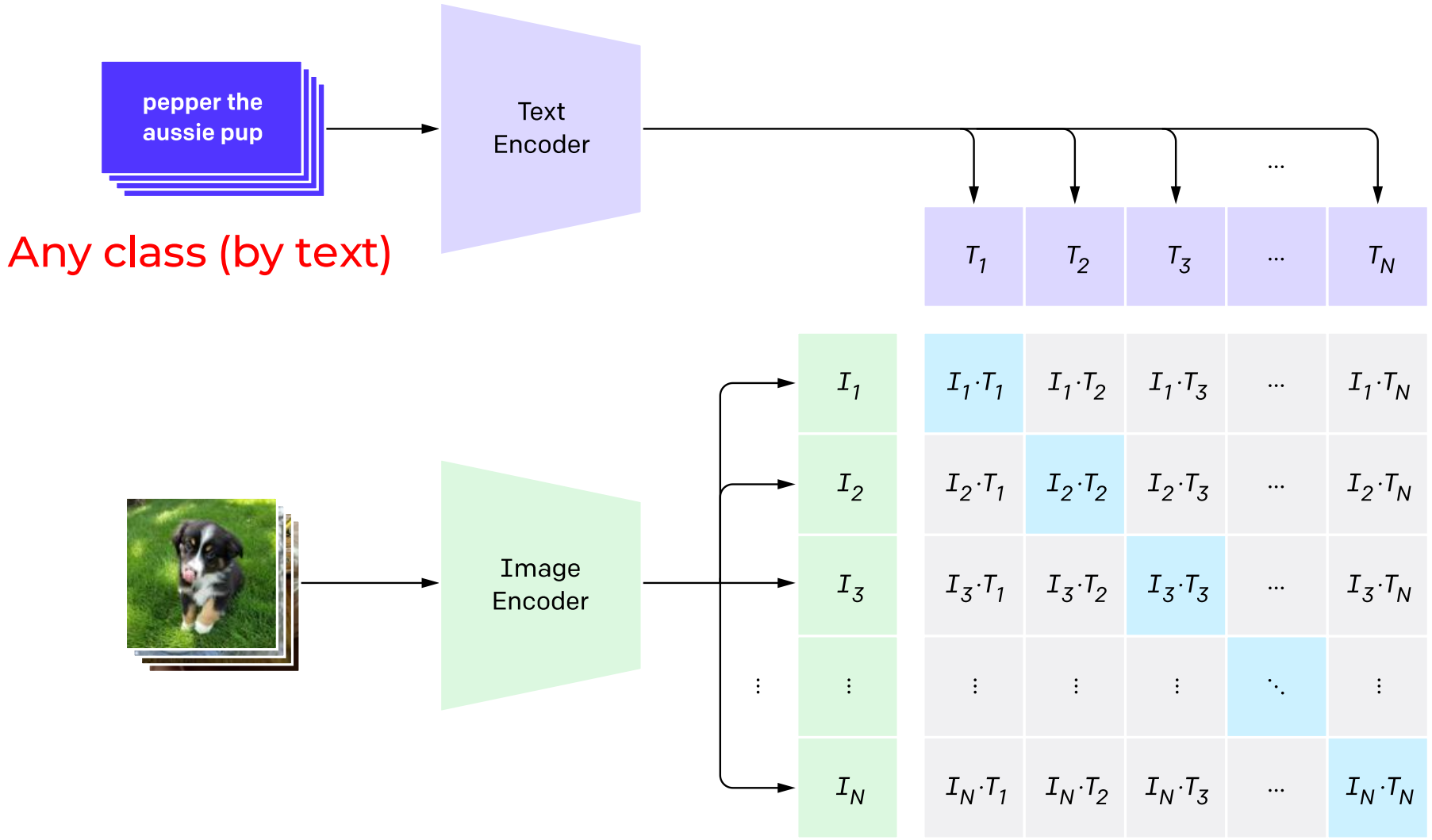
Kyungmin Lee



Jinwoo Shin

NeurIPS 2023

Vision-language models (e.g., CLIP) are amazing!



Vision-language models (e.g., CLIP) are amazing!





Source: CIRR and StyleCLIP papers

...but only for **standard domains** (i.e., LAION-like)

Backend url:
https://knn5.laior
Index:
laion_5B

Clip retrieval works by converting the text query to a CLIP embedding , then using that embedding to query a knn index of clip image embedddings

Display captions ☑
Display full captions ☐
Display similarities ☐
Safe mode ☑
Hide duplicate urls ☑
Hide (near) duplicate images ☑
Search over
image
Search with multilingual clip ☐

french cat


french cat


french cat


How to tell if your feline is french. He wears a b...


イケメン猫モデル「トキ・ナンタケット」がかっこいい - NAVER まとめ


Hilarious pics of funny cats! funnycatsgif.com


Hipster cat


網友挑戰「加幾筆畫出最創意貓咪圖片」，笑到岔氣之後我也手


cat in a suit Georgian sells tomatoes




French Bread Cat Loaf Metal Print

CLIP for **Specialist Domains?**



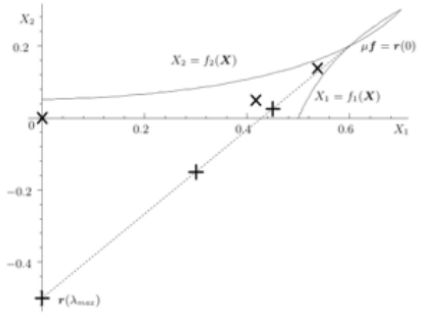
“many houses arranged neatly and some roads crossed them with some plants in the roadside”

(a) Remote sensing



“grey classic u-neck t-shirt. short sleeve t-shirt in heather grey. tonal logo flag at u-neck collar.”

(b) Fashion



“Fig. 2. variable 2-D initial probe distribution used for CFO runs reported here”

(c) Scientific figure



“simpsons wearing headphones and looking like a guy”

(d) Comics

CLIP for **Specialist Domains**?



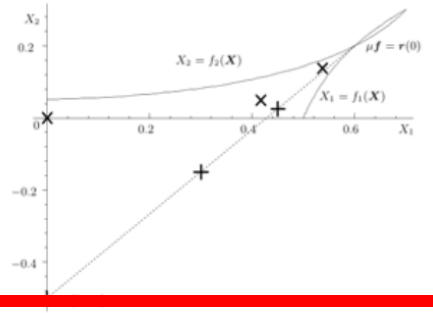
“many houses arranged neatly and some roads crossed them with some plants in the roadside”

(a) Remote sensing



“grey classic u-neck t-shirt. short sleeve t-shirt in heather grey. tonal logo flag at u-neck collar.”

(b) Fashion



“Fig. 2. variable 2-D initial probe distribution used for CFO runs reported here”

(c) Scientific figure

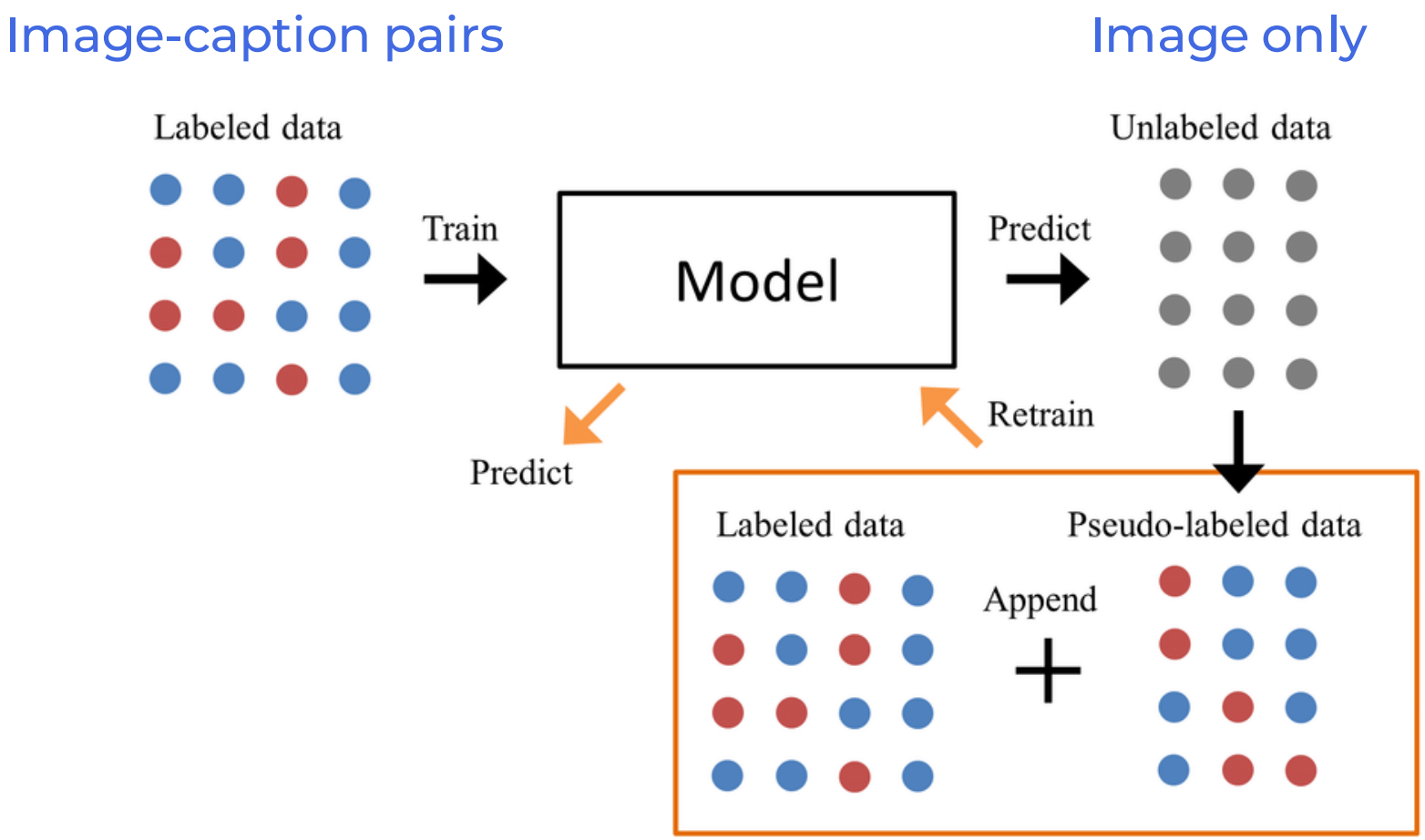


“simpsons wearing headphones and looking like a guy”

(d) Comics

Challenge: We don’t have many **detailed captions**

Solution: **Semi-supervised learning** (Semi-SL)



Source: https://www.researchgate.net/figure/Schematic-diagram-of-pseudo-labeling_fig2_341813397

Naïve **Pseudo-Labeling** (PL) Fails



"two tennis courts and some buildings are in two sides of a road"

Nearest image in the batch

Their captions are different! (naive pseudo-labeling fails)



"the huge rectangle contains a baseball field, a swimming pool, two basketball courts and eight tennis courts"

Idea 1. **Caption-level PL**

Pseudo-label = **Combination of the captions**
(from visually similar images in the batch)



=

0.4 ×



"…tennis courts…"

+

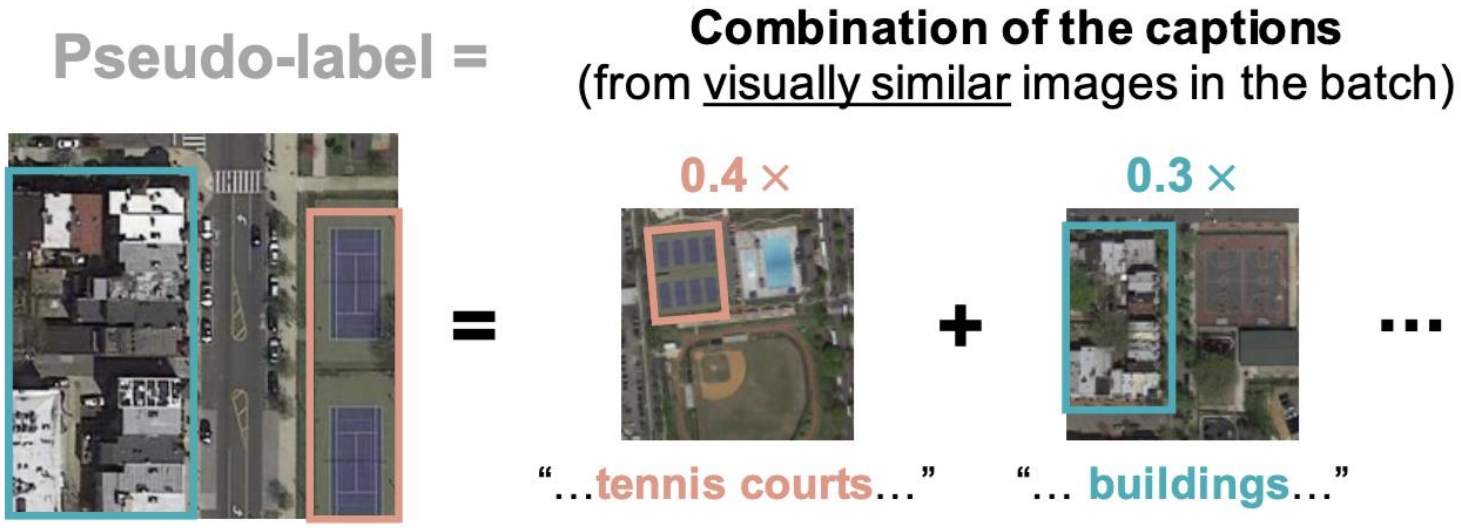
0.3 ×



"… buildings…"

…

Idea 1. **Caption-level PL**



Optimal Transport (OT) between labeled & unlabeled images

1) measure distance by visual similarity

2) avoid over-confidence using OT

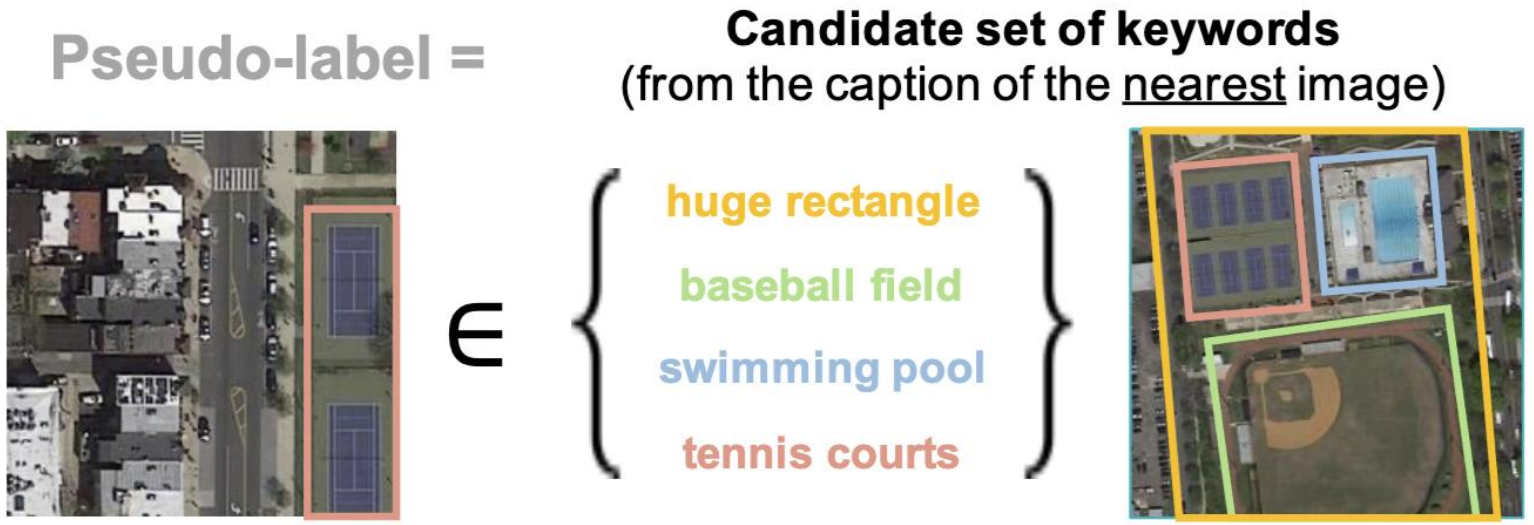
Idea 2. **Keyword-level PL**

Pseudo-label =  ∈ { huge rectangle, baseball field, swimming pool, tennis courts }

Candidate set of keywords
(from the caption of the nearest image)



Idea 2. **Keyword-level PL**

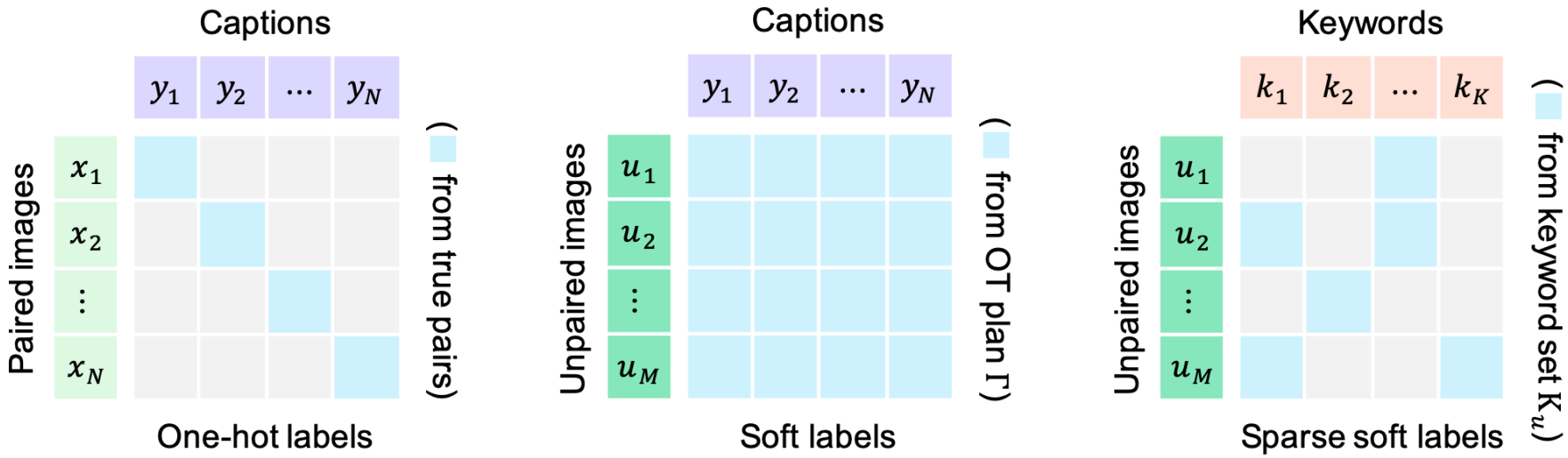


Formulated as a **Partial Label Learning** (PLL) problem

1) estimate the likelihood of each element in the set

2) update the model with the current estimates and repeat 1

Training Objective of S-CLIP



(a) CLIP loss $\mathcal{L}_{\text{CLIP}}$ (b) Caption loss $\mathcal{L}_{\text{caption}}$ (c) Keyword loss $\mathcal{L}_{\text{keyword}}$

Better Performance

Zero-shot classification

Method	Data	RSICD-CLS	UCM-CLS	WHU-RS19	RSSCN7	AID
CLIP (original)	-	45.3	50.5	65.5	58.9	47.8
CLIP (fine-tune)	L	58.3±0.3	63.5±3.4	76.5±3.2	61.9±1.2	63.1±1.3
Hard-PL [16]	L=U	56.6±3.5 (-1.7)	61.6±2.2 (-1.9)	78.1±2.5 (+1.6)	63.9±2.1 (+2.0)	63.2±2.6 (+0.1)
Soft-PL [21]	L=U	62.5±0.8 (+4.2)	65.7±2.7 (+2.2)	83.7±2.7 (+7.2)	65.7±0.6 (+3.8)	68.0±0.7 (+4.9)
S-CLIP (ours)	L=U	**66.9**±1.7 (+8.6)	**66.7**±1.6 (+3.2)	**86.9**±2.0 (+10.4)	**66.2**±1.1 (+4.3)	**73.0**±0.3 (+9.9)

Image-text retrieval

		Image→text R@5			Text→image R@5		
Method	Data	RSICD	UCM	Sydney	RSICD	UCM	Sydney
CLIP (original)	-	9.4	34.3	36.2	10.1	24.8	51.7
CLIP (fine-tune)	L	15.4±1.7	41.3±1.8	47.1±6.5	15.1±1.0	40.9±1.6	56.1±2.4
Hard-PL [16]	L=U	16.1±0.2	40.8±2.9	43.1±3.0	15.7±0.7	40.5±3.0	47.7±5.3
Soft-PL [21]	L=U	17.0±0.9	43.2±3.9	42.0±4.3	16.5±0.1	42.9±3.3	50.2±4.9
S-CLIP (ours)	L=U	**18.4**±0.6	**45.7**±1.4	**50.0**±3.0	**16.8**±1.2	**43.5**±1.5	**55.1**±2.0

Better Robustness

Labeled & unlabeled images come from **different** datasets

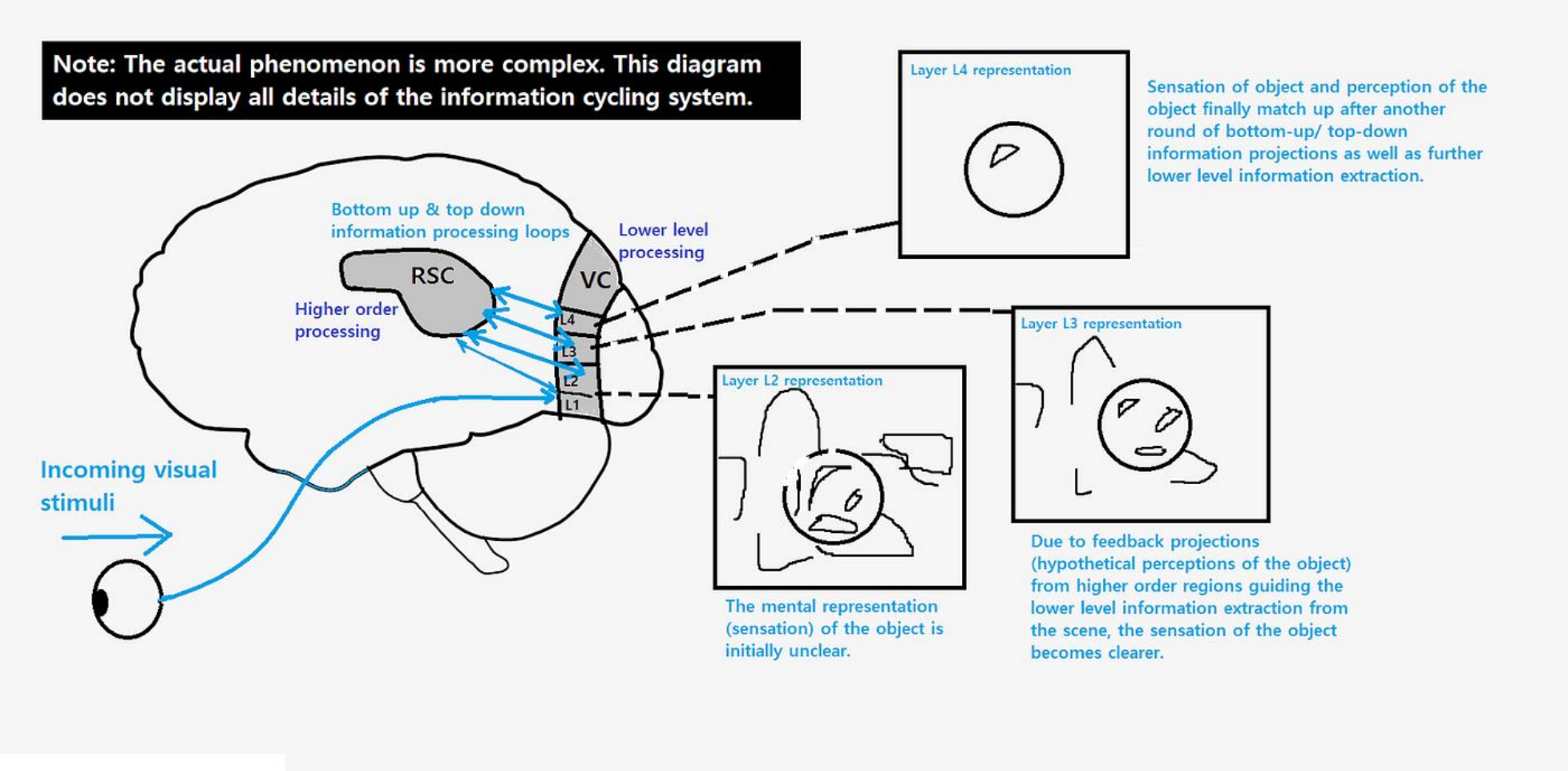
Method	Data	RSICD-CLS	UCM-CLS	WHU-RS19	RSSCN7	AID
CLIP (original)	-	45.3	50.5	65.5	58.9	47.8
CLIP (fine-tune)	L	58.3±0.3	63.5±3.4	76.5±3.2	61.9±1.2	63.1±1.3
Hard-PL [16]	L≠U	55.8±1.2 (-2.5)	61.0±6.2 (-2.5)	76.3±1.1 (-0.2)	62.5±2.7 (+0.6)	62.3±1.4 (-0.8)
Soft-PL [21]	L≠U	56.1±3.5 (-2.2)	62.4±1.8 (-1.1)	79.5±3.6 (+3.0)	63.3±2.6 (+1.4)	62.4±2.1 (-0.7)
S-CLIP (ours)	L≠U	**65.3**±2.1 (+7.0)	**66.0**±0.4 (+2.5)	**86.3**±1.1 (+9.8)	**68.3**±2.2 (+6.4)	**70.8**±2.5 (+7.7)

Part 2. **Structural Priors in Data Instances**

Structural Prior: **What Does the Data Instance Look Like?**



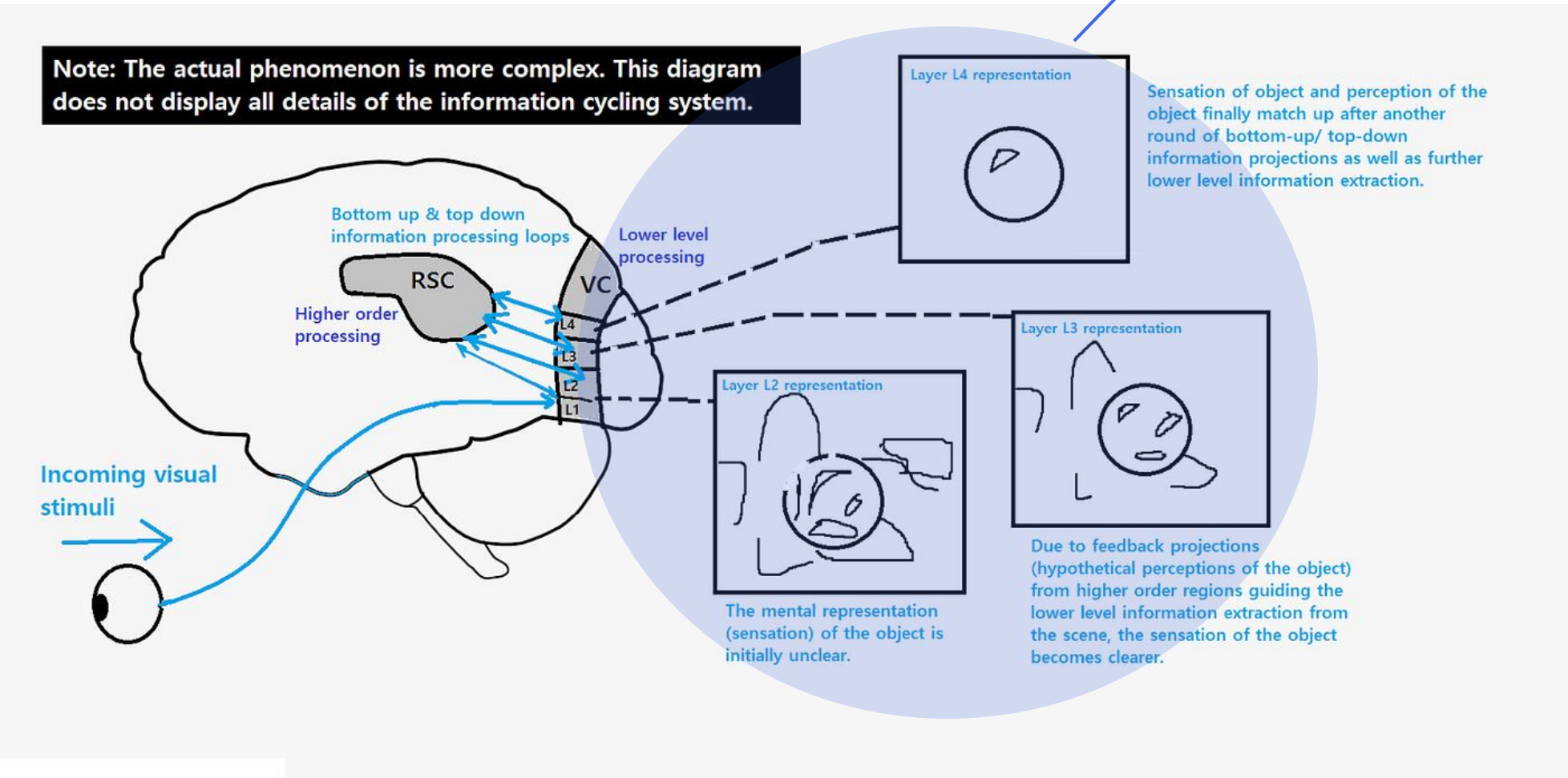
Structures in Vision



Source: https://math.hws.edu/graphicsbook/c2/s4.html

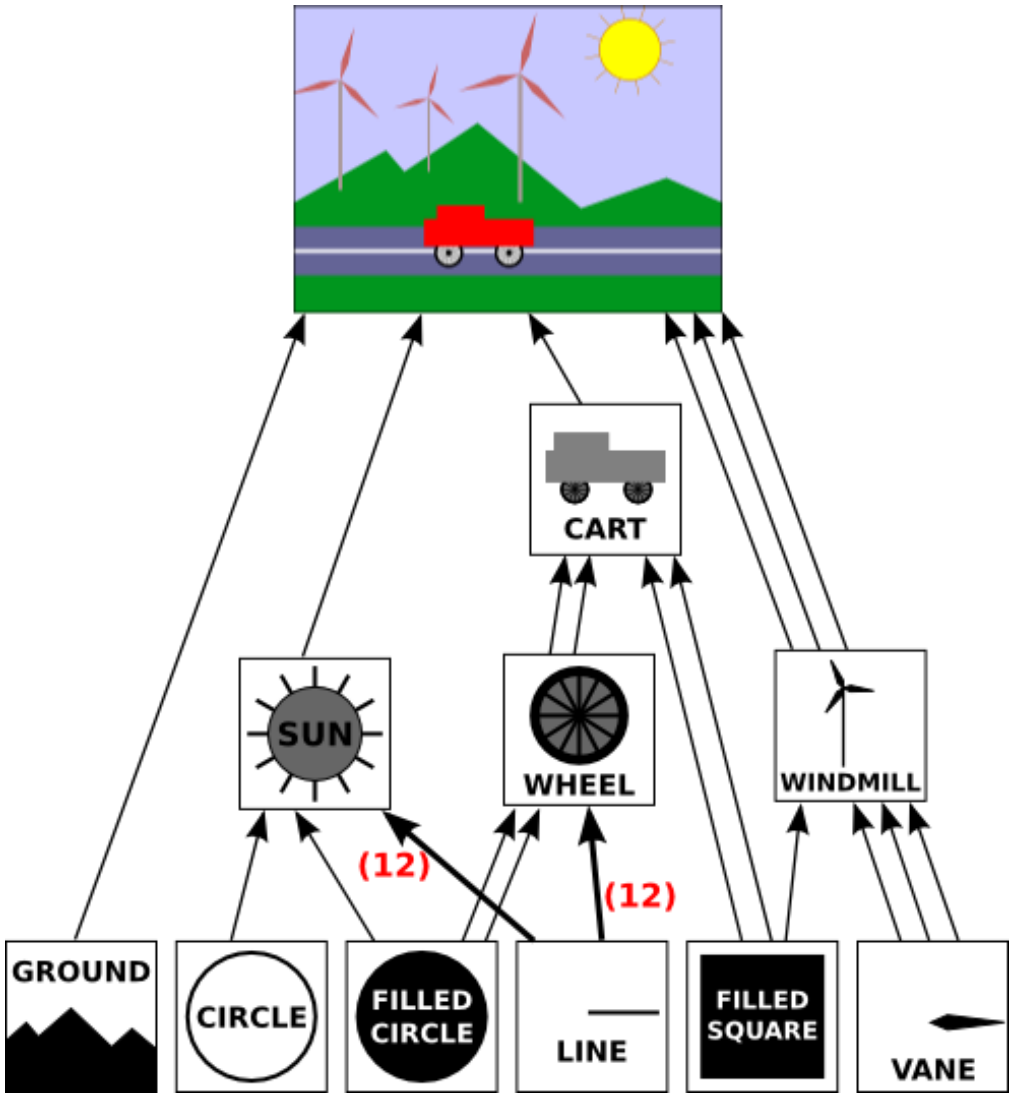
Structures in Vision

1. Bottom-up: Images follow a scene hierarchy



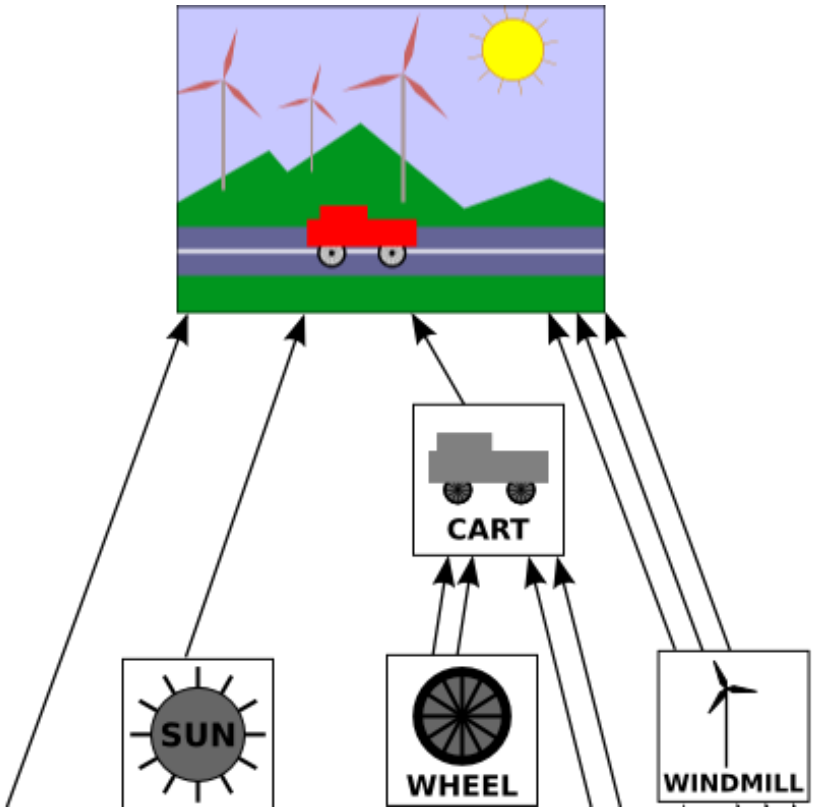
Source: https://math.hws.edu/graphicsbook/c2/s4.html

1. Bottom-up: Images follow a **scene hierarchy**



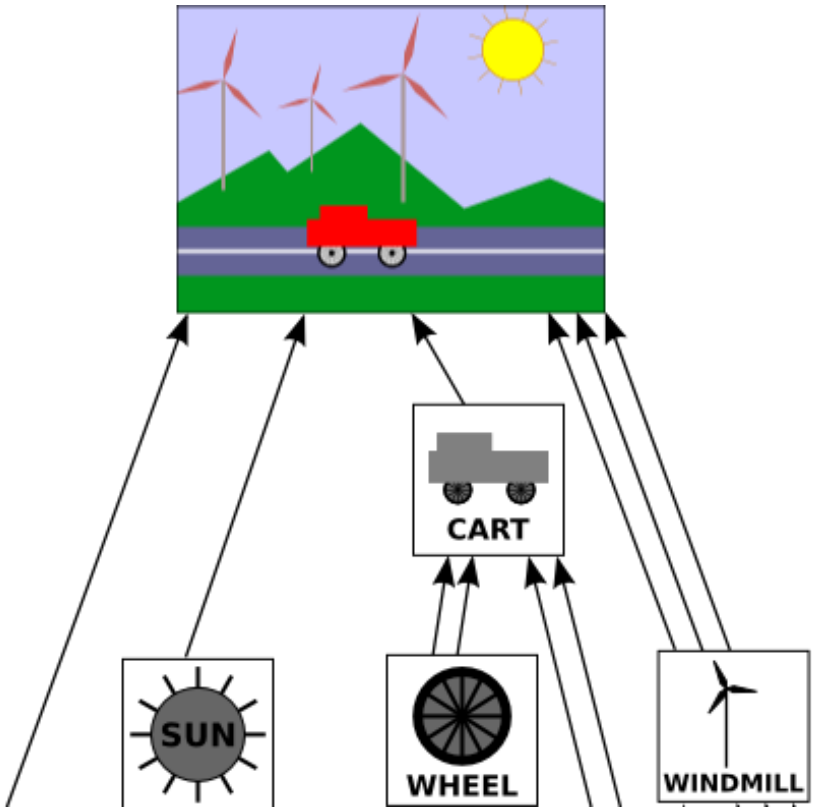
Source: https://math.hws.edu/graphicsbook/c2/s4.html

1. Bottom-up: Images follow a **scene hierarchy**



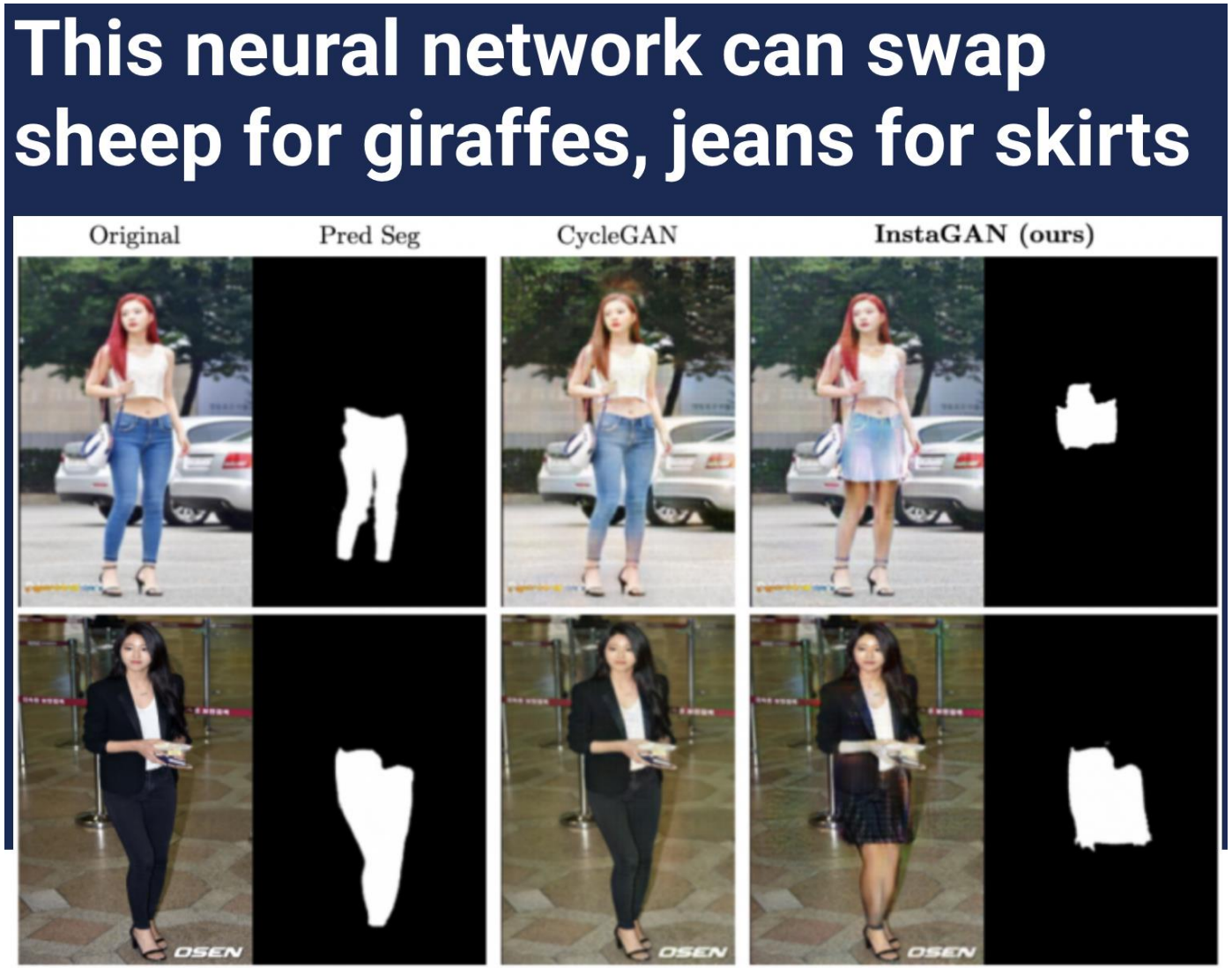
- Scene hierarchy enhances image generation (GAN)
  – InstaGAN (Mo et al., ICLR'19)
- Scene hierarchy enhances representation learning (MoCo)
  – OACon (Mo* & Kang* et al., NeurIPS'21)

1. Bottom-up: Images follow a **scene hierarchy**



- Scene hierarchy enhances image generation (GAN)
  – InstaGAN (Mo et al., ICLR'19) → 217 citations, Samsung awards
- Scene hierarchy enhances representation learning (MoCo)
  – OACon (Mo* & Kang* et al., NeurIPS'21) → 51 citations, Qualcomm awards

Application: Fine-grained Image Manipulation



A neural network -- InstaGAN -- swaps clothing.
Image Credit: Mo, Cho, and Shin

Application: Fine-grained Image Recognition

Question: Tell me the number of that player who is shooting.

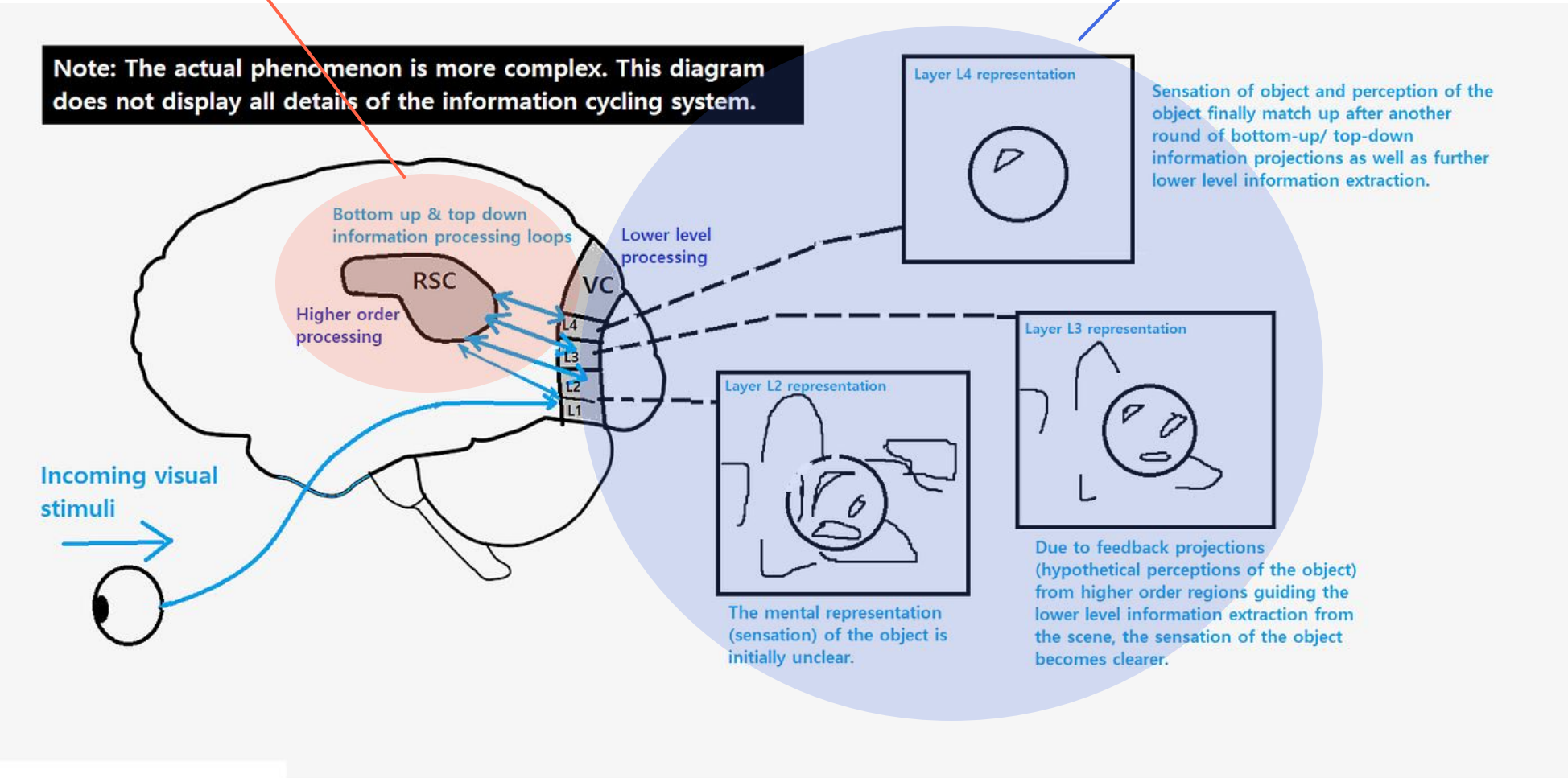


Source: VStar paper

Structures in Vision

2. Top-down: Perception is **guided** by context

1. Bottom-up: Images follow a **scene hierarchy**



Source: https://math.hws.edu/graphicsbook/c2/s4.html

2. Top-down: Perception is **guided** by context



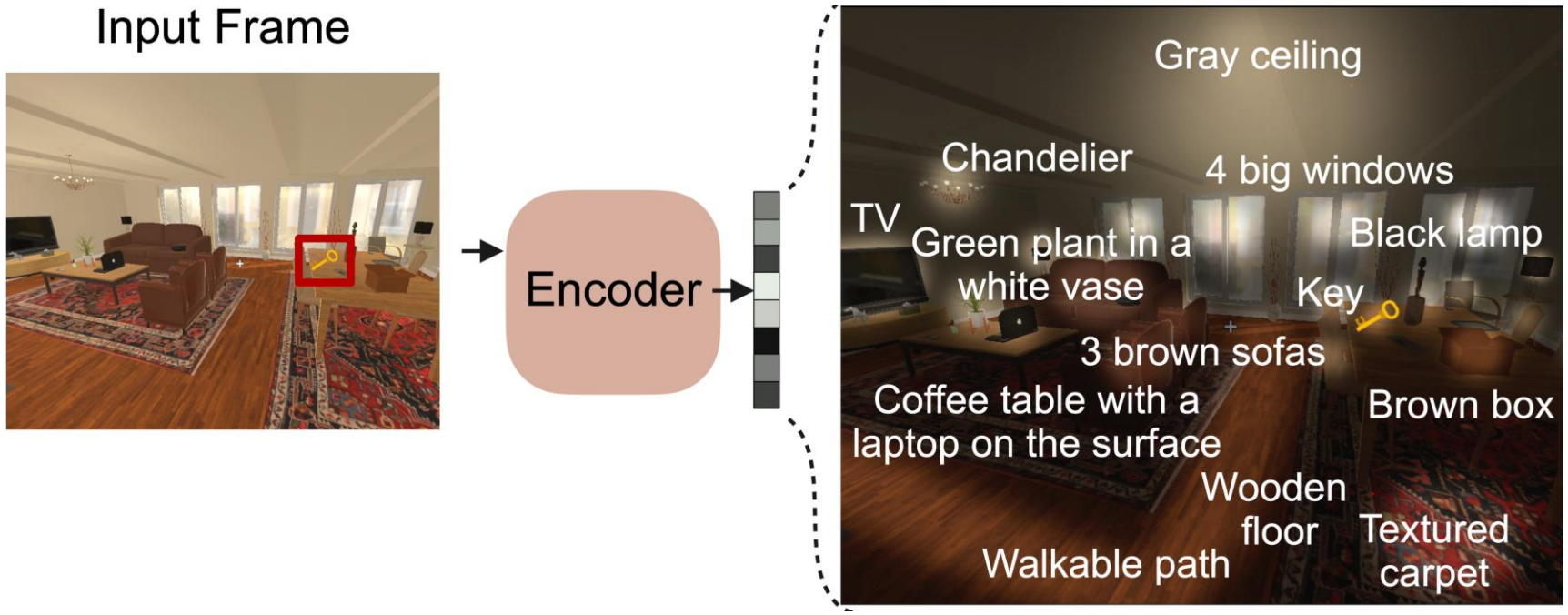
Source: https://www.popsci.com/story/diy/dalmatian-illusion/

2. Top-down: Perception is **guided** by context



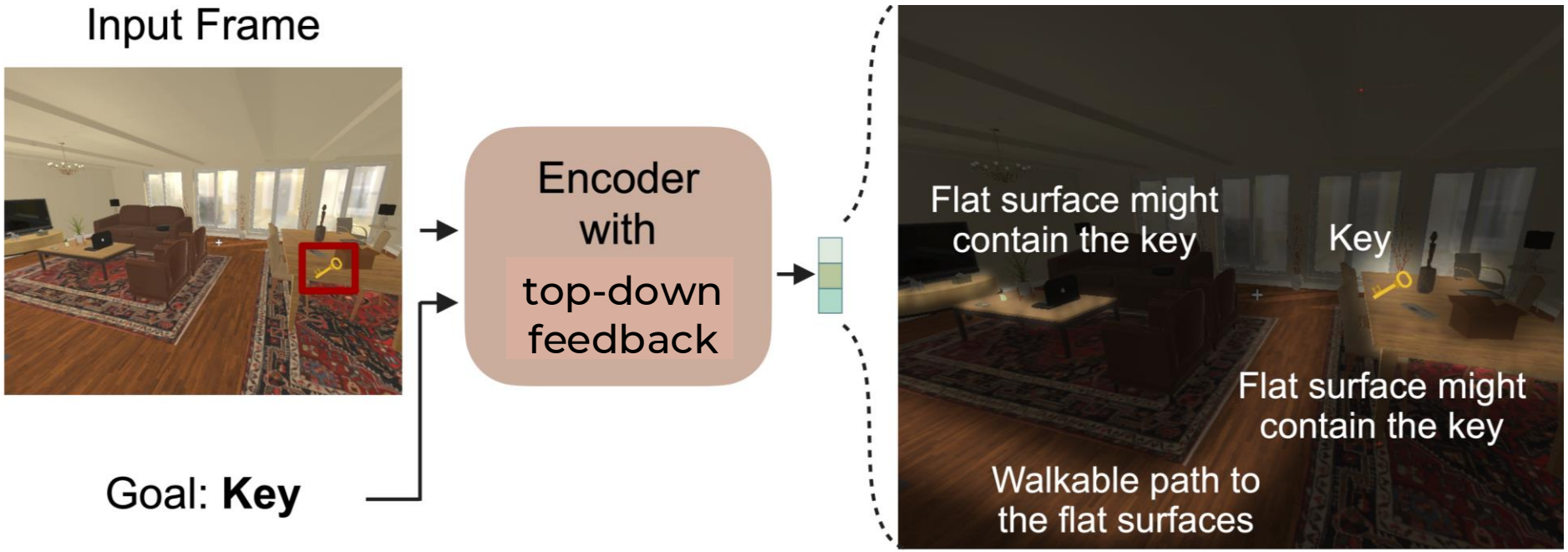
- **Top-down feedback guides how to parse images (ViT)**
  – CAST (Ke* & Mo* et al., ICLR'24)
- **Top-down feedback guides where to focus in context (CLIP)**
  – OAK (Wang* & Mo* et al., CVPR'25)

Application: **Efficient Visual Processing**



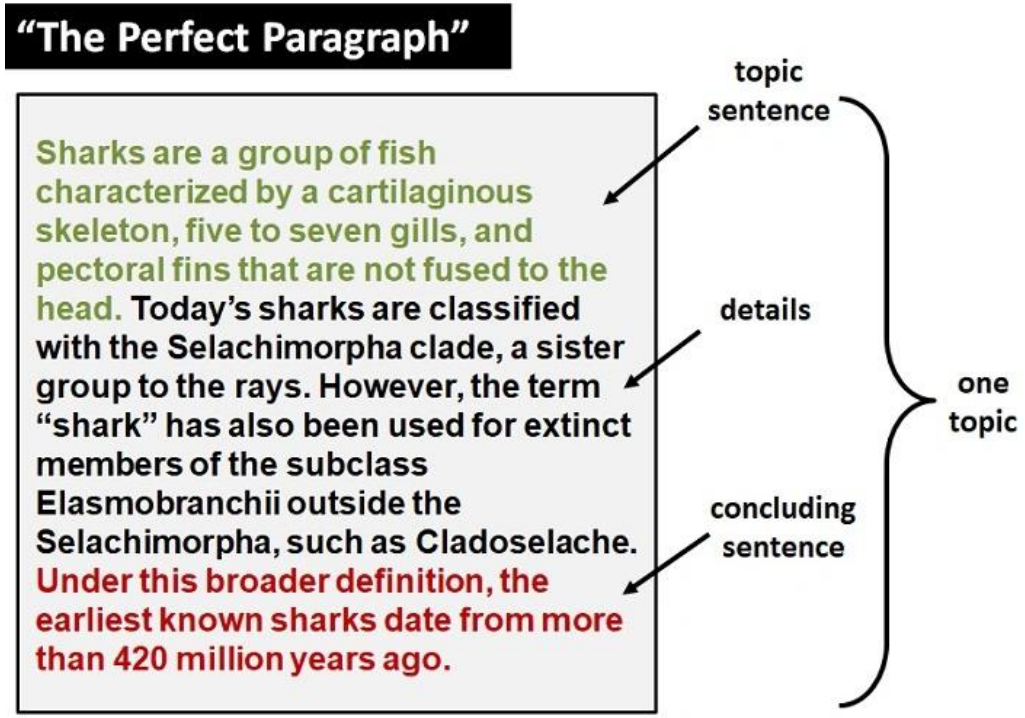
Source: Embodied Codebook paper

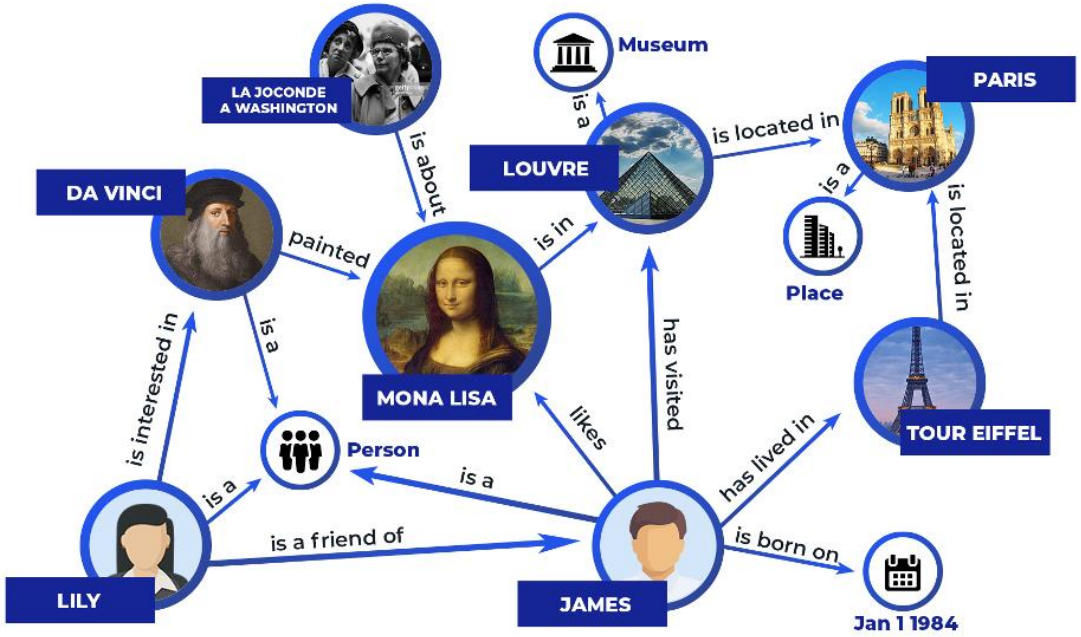
Application: **Efficient Visual Processing**



Source: Embodied Codebook paper

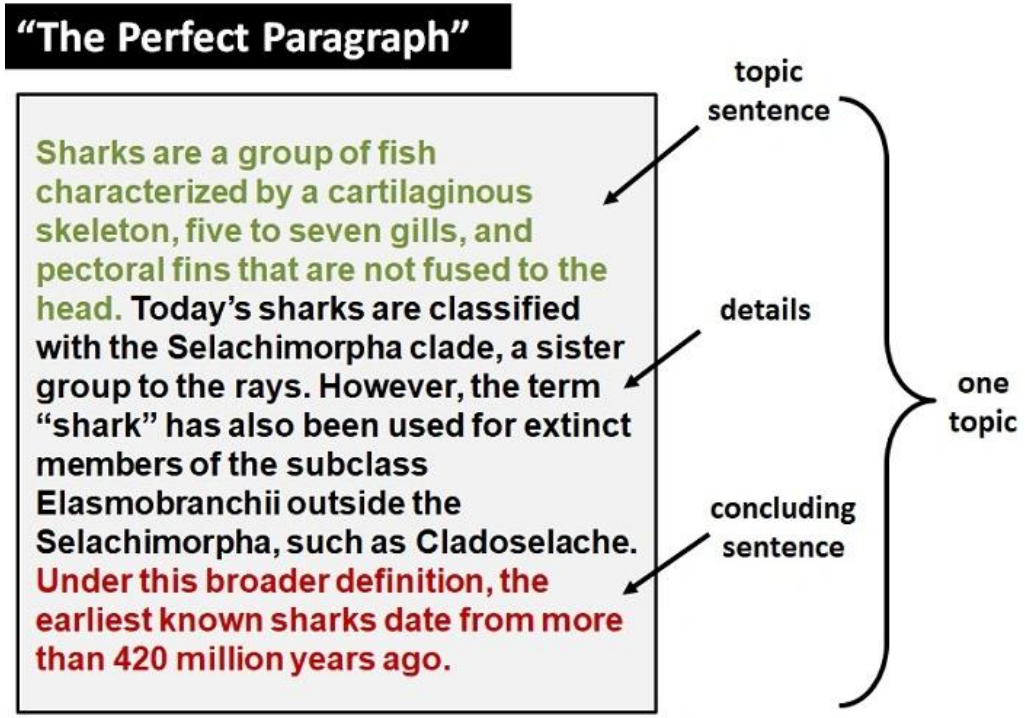
Structures are Everywhere

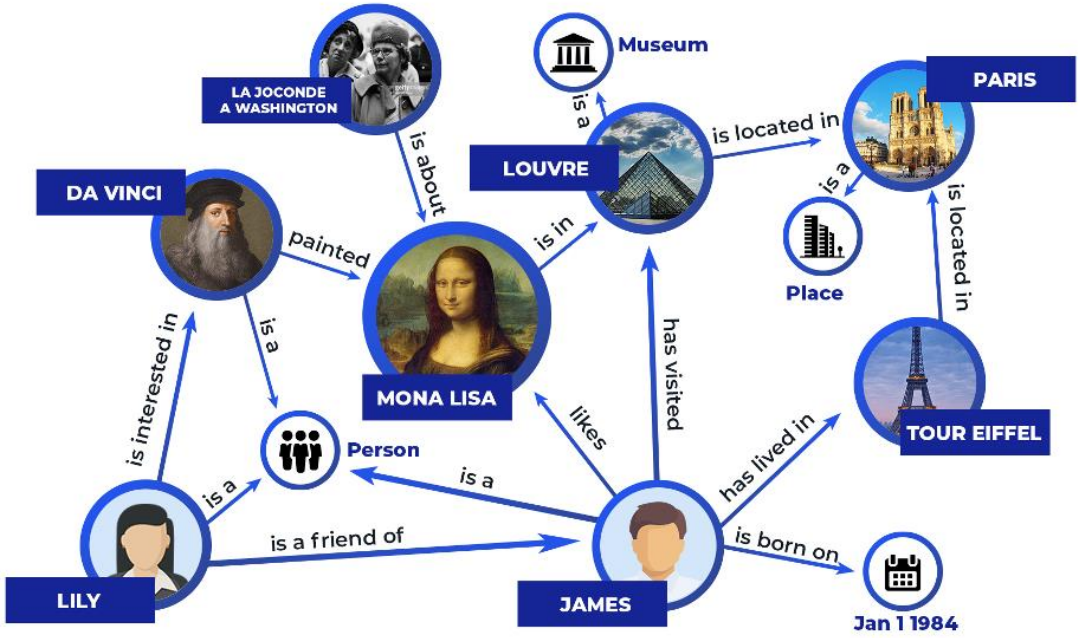




Source: https://www.grammar-monster.com/glossary/paragraph.htm, https://zilliz.com/learn/what-is-knowledge-graph

Structures are Everywhere





- Hierarchical abstraction aids long-context understanding
  – HOMER (Song et al., ICLR'24)
- Structured rationale aids knowledge-based reasoning
  – SuRe (Kim et al., ICLR'24)

Open Ad-hoc Categorization with Contextualized Feature Learning



Zilin Wang*



Sangwoo Mo*



Stella X. Yu



Sima Behpour



Liu Ren

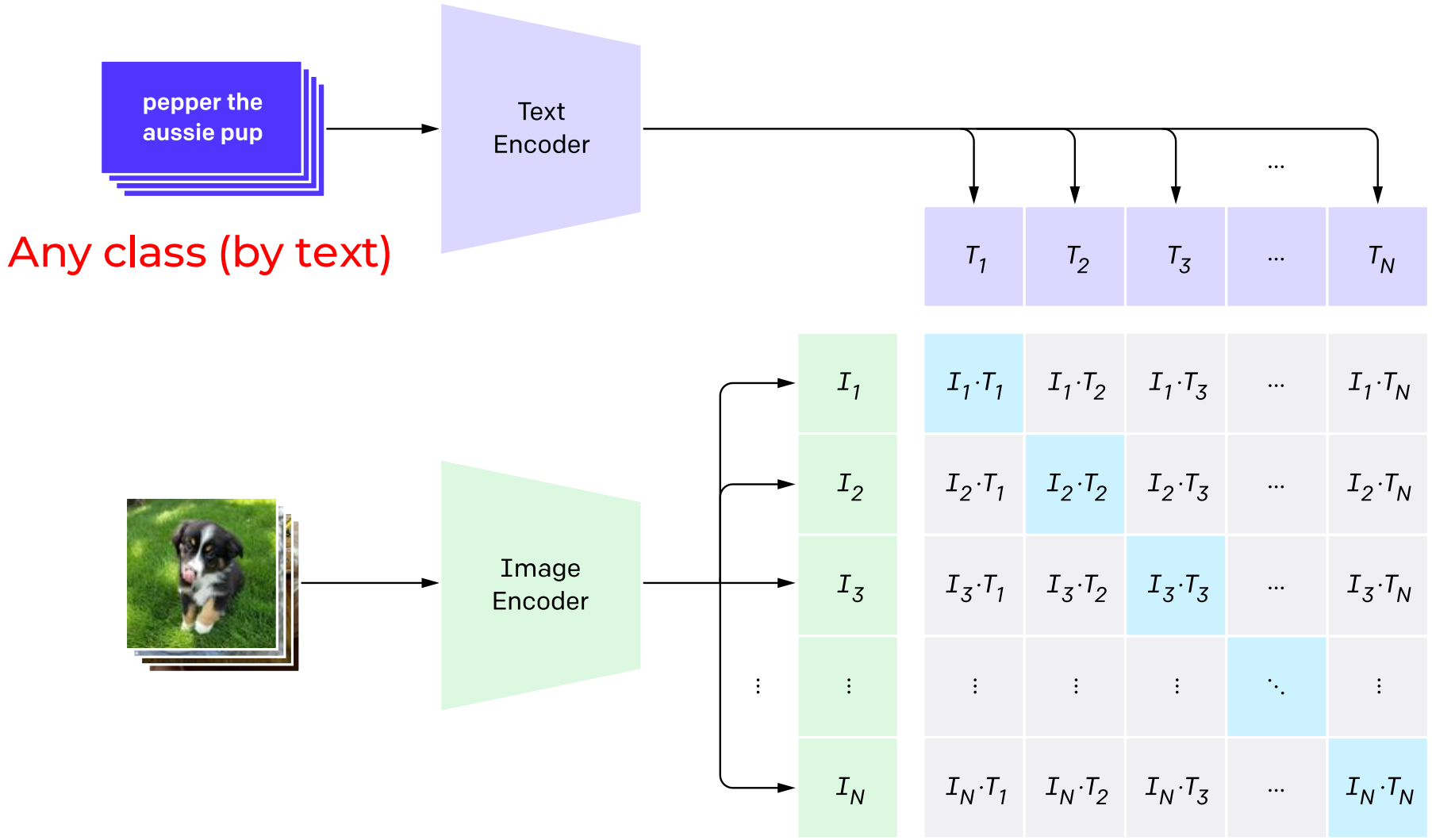


CVPR 2025

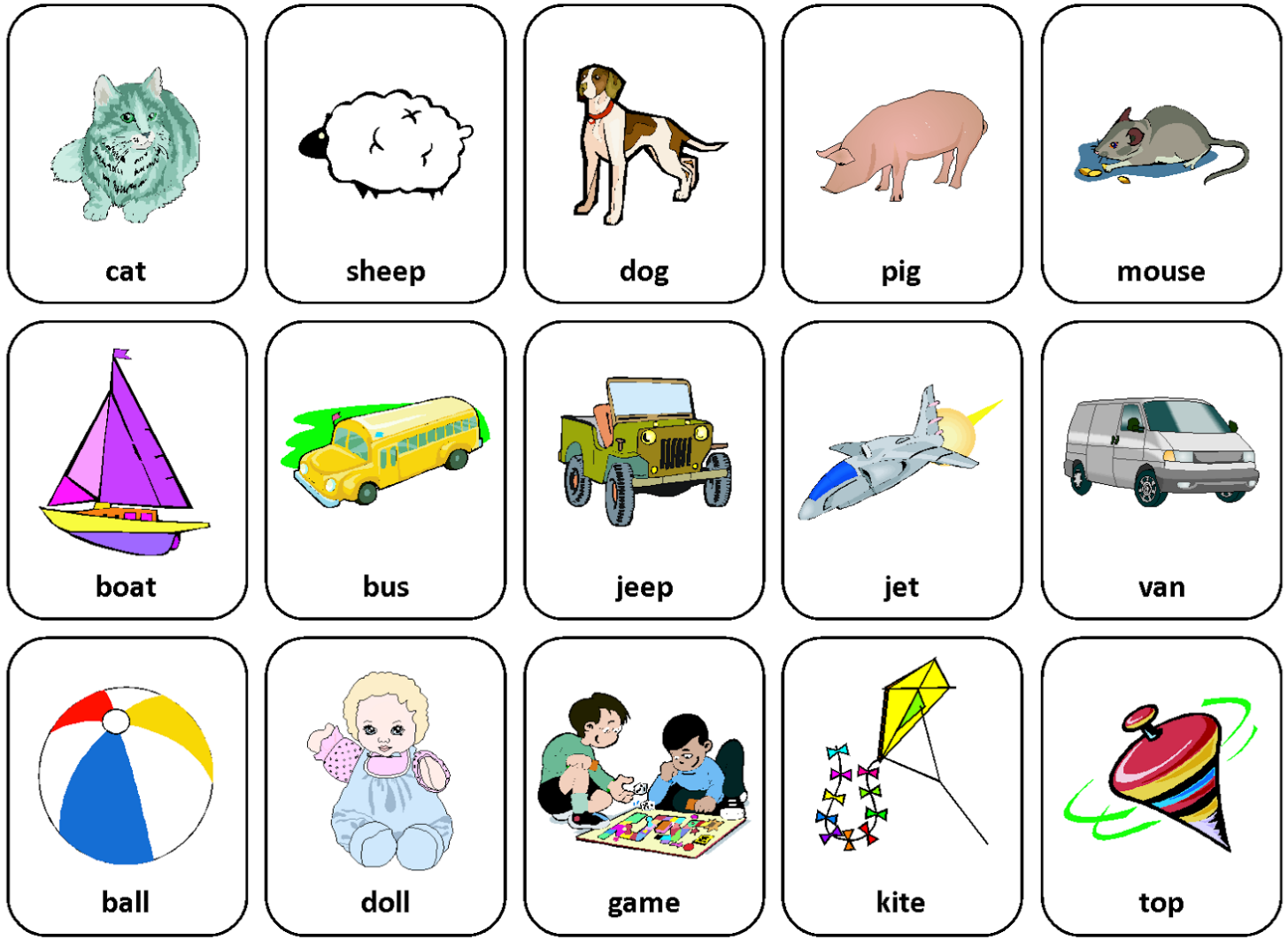
* Equal contribution



Is Categorization **Solved**?



CLIP works well for **Basic Categories**



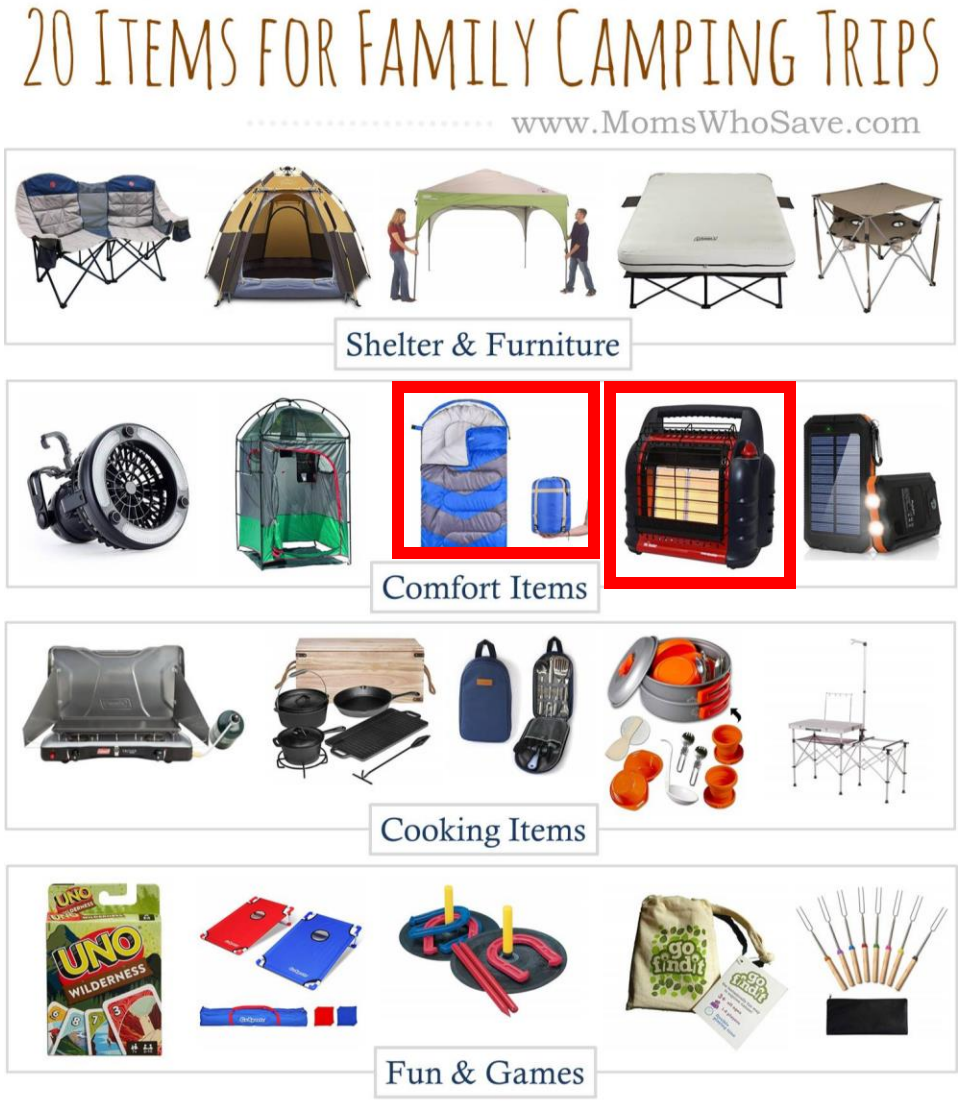
Source: https://testyyettrying.blogspot.com/2013/07/sorting-by-category-free-speech-therapy.html

However, Categorization is often Ad-hoc



Source: https://momswhosave.com/family-camping-trip-checklist/

However, Categorization is often Ad-hoc



Source: https://momswhosave.com/family-camping-trip-checklist/

Categorization is often Open



What else?

?

Source: https://momswhosave.com/family-camping-trip-checklist/

Open Ad-hoc Categorization



What else?

?

Source: https://momswhosave.com/family-camping-trip-checklist/

Open Ad-hoc Categorization

Semantics



What else?

?

Source: https://momswhosave.com/family-camping-trip-checklist/

Open Ad-hoc Categorization

Semantics



Visuals

What else?

?

Source: https://momswhosave.com/family-camping-trip-checklist/

Open Ad-hoc Categorization

Semantics (CLIP)



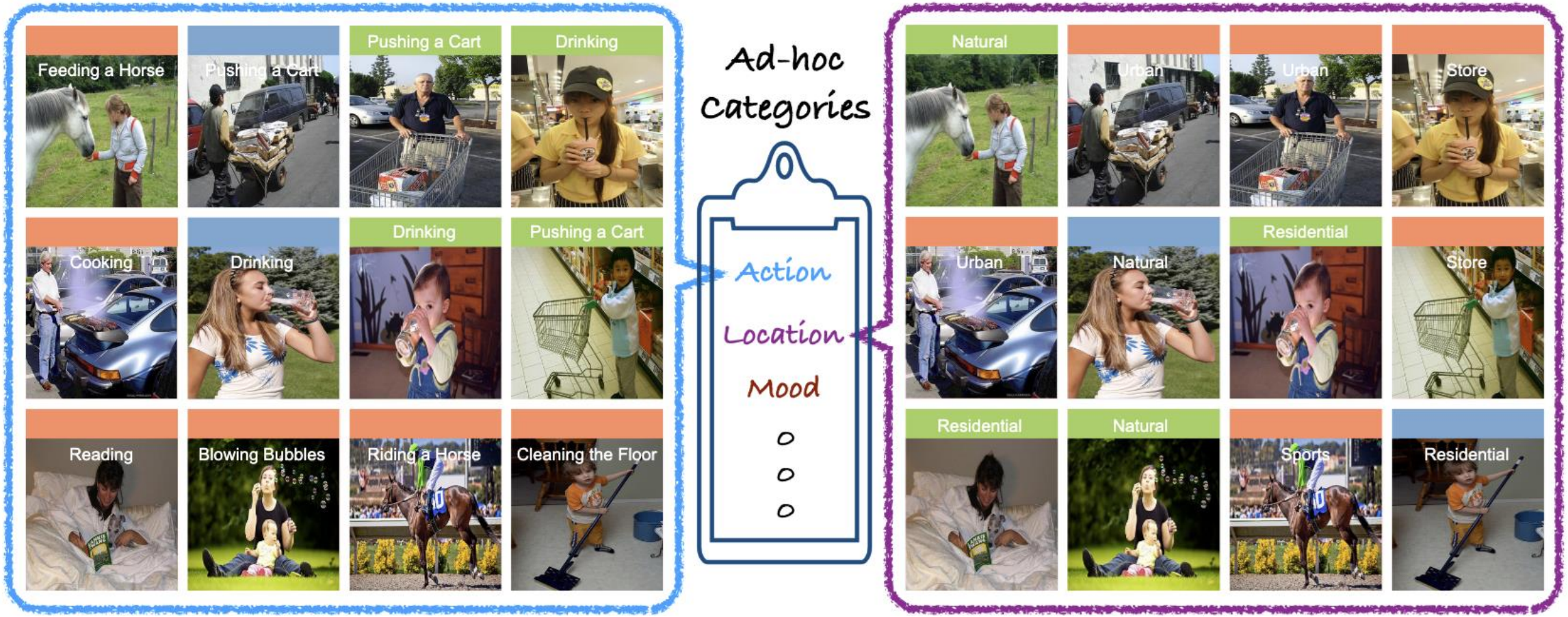
Visuals (GCD)

What else?

?

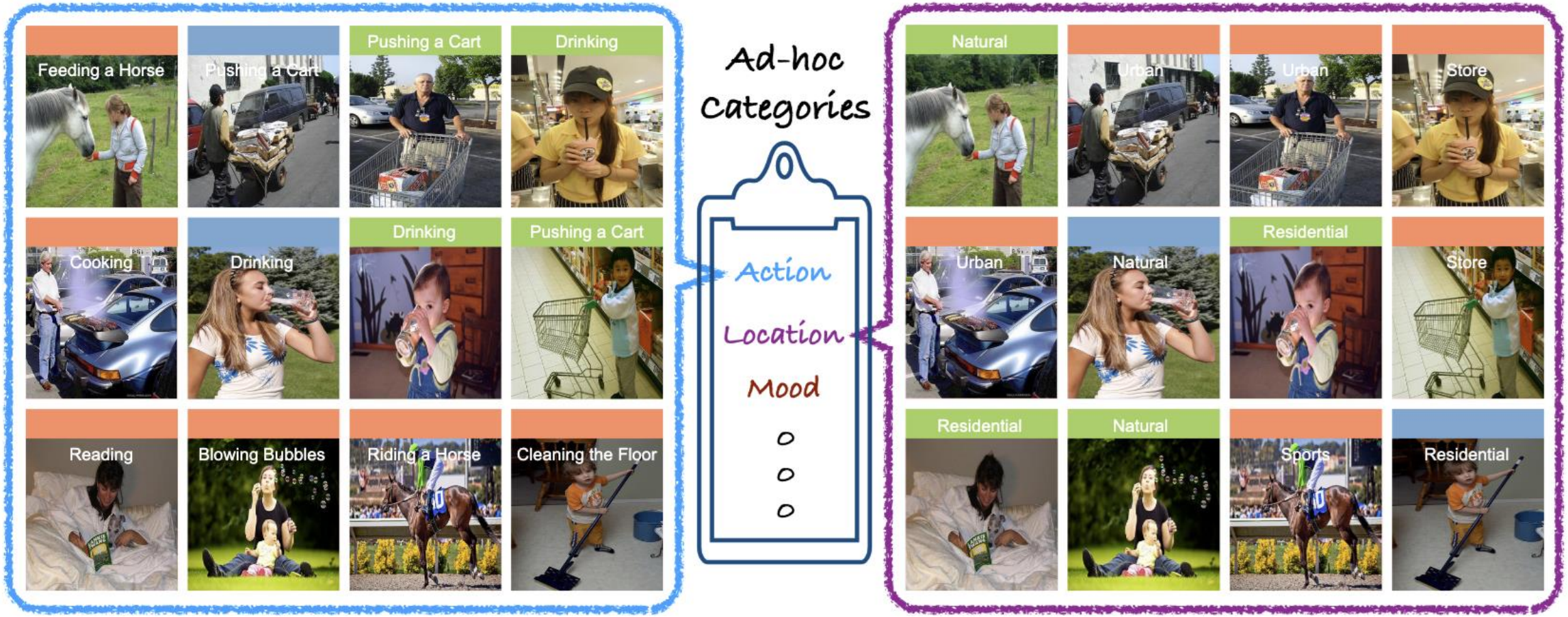
Source: https://momswhosave.com/family-camping-trip-checklist/

Problem Setup



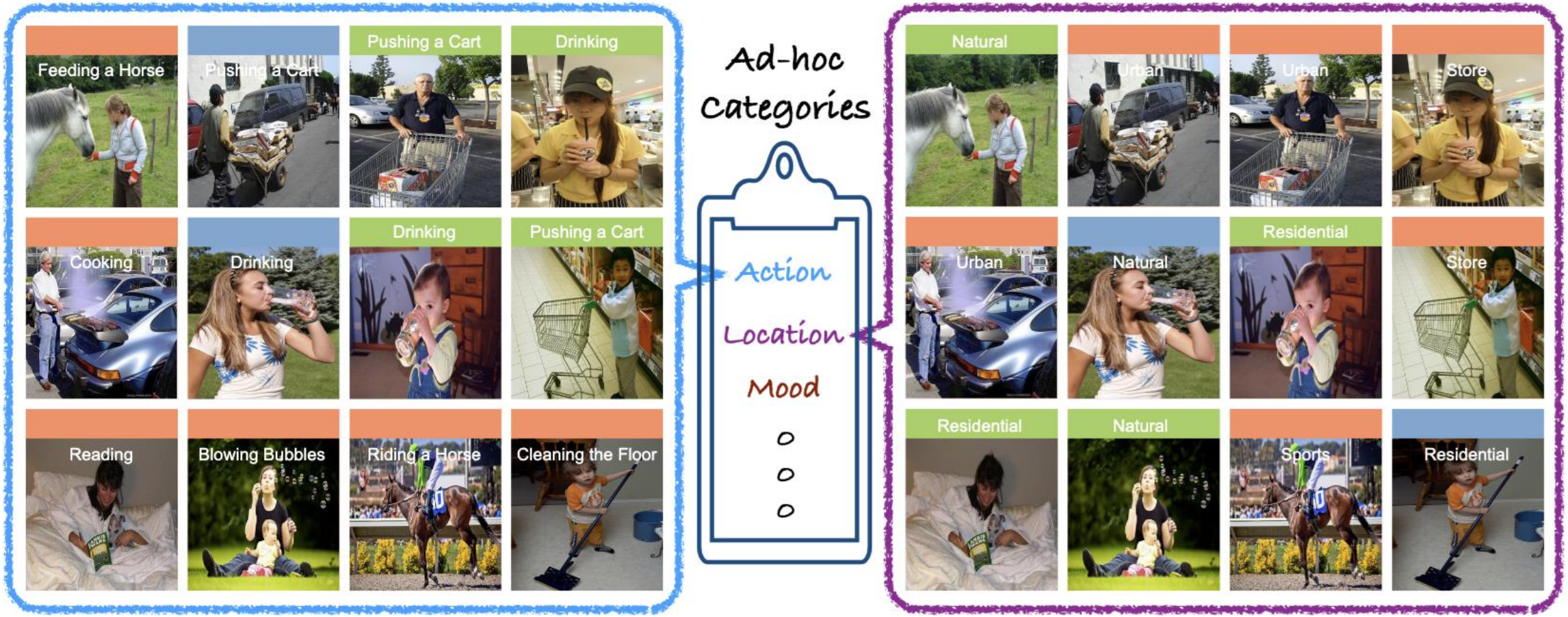
Problem Setup

Labeled



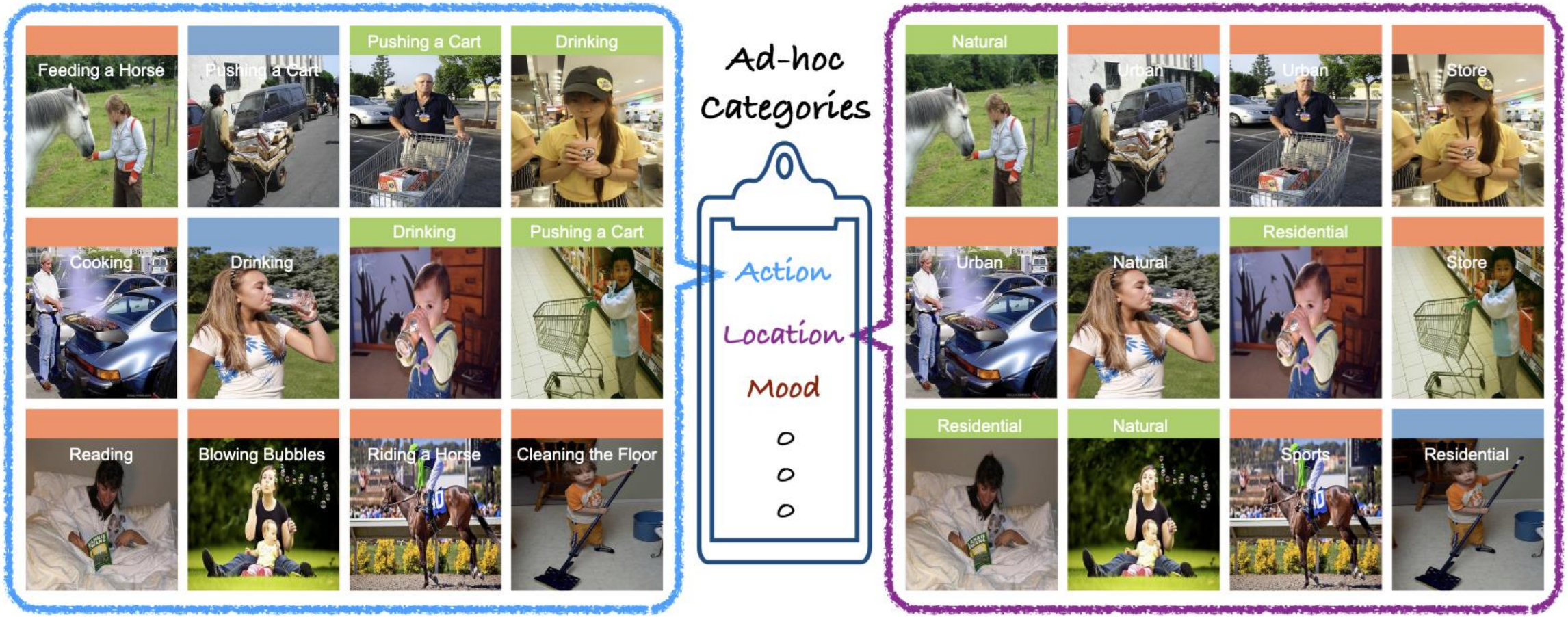
Problem Setup

Labeled
Unlabeled known

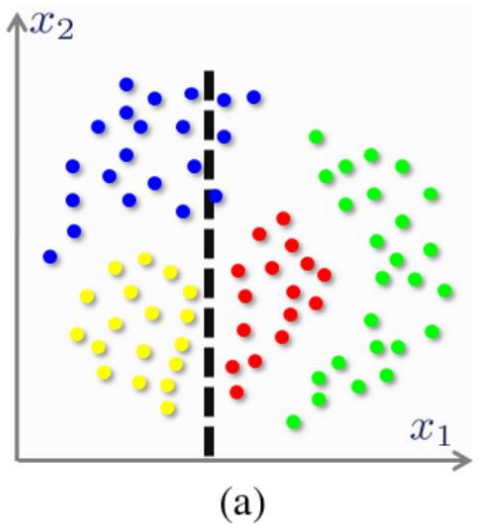


Problem Setup

Labeled
Unlabeled known
Unlabeled unknown

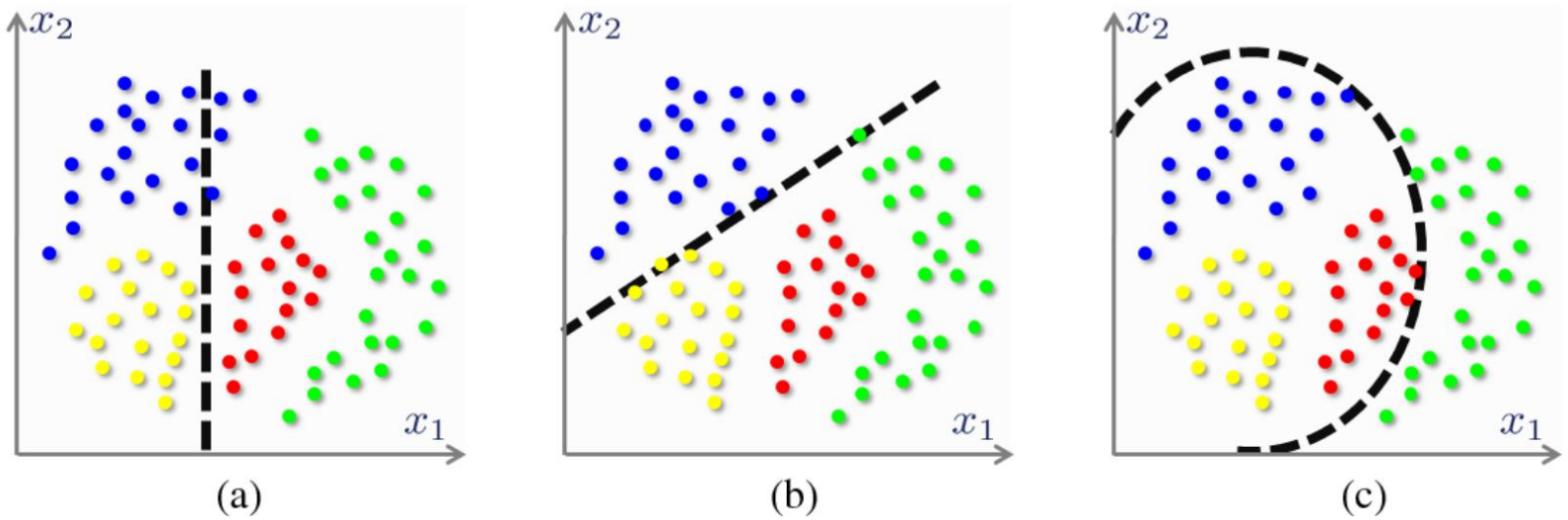


No Single Representation Rules Them All



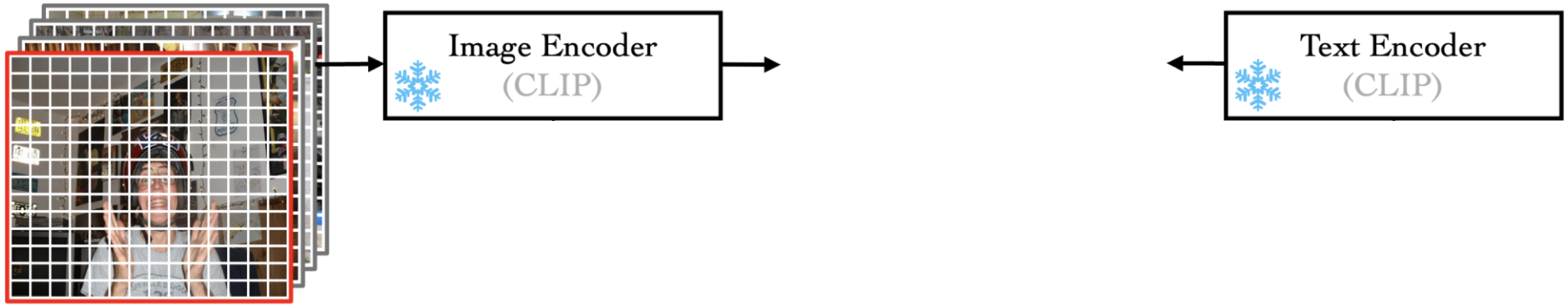
Source: https://datascience.stackexchange.com/questions/100015/orthogonal-decision-boundary-in-decision-tree

No Single Representation Rules Them All



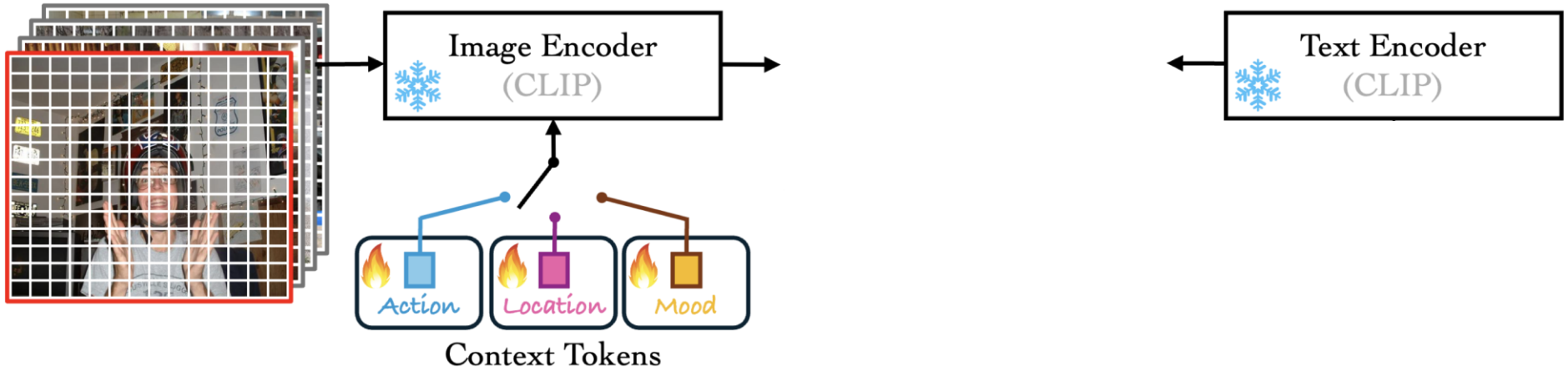
Source: https://datascience.stackexchange.com/questions/100015/orthogonal-decision-boundary-in-decision-tree

Our OAK Framework



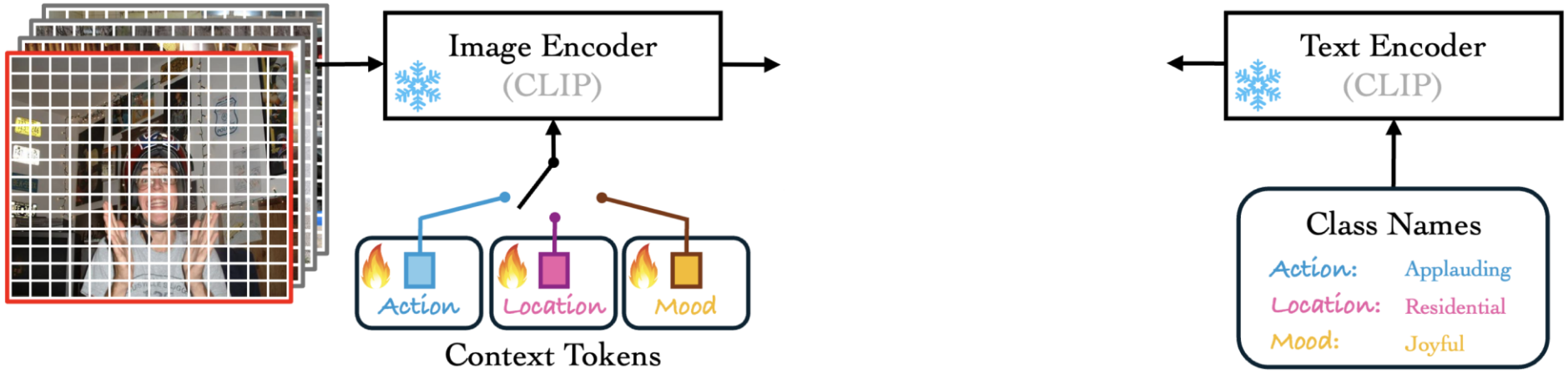
1. **Build on CLIP** to leverage its general knowledge

Our OAK Framework



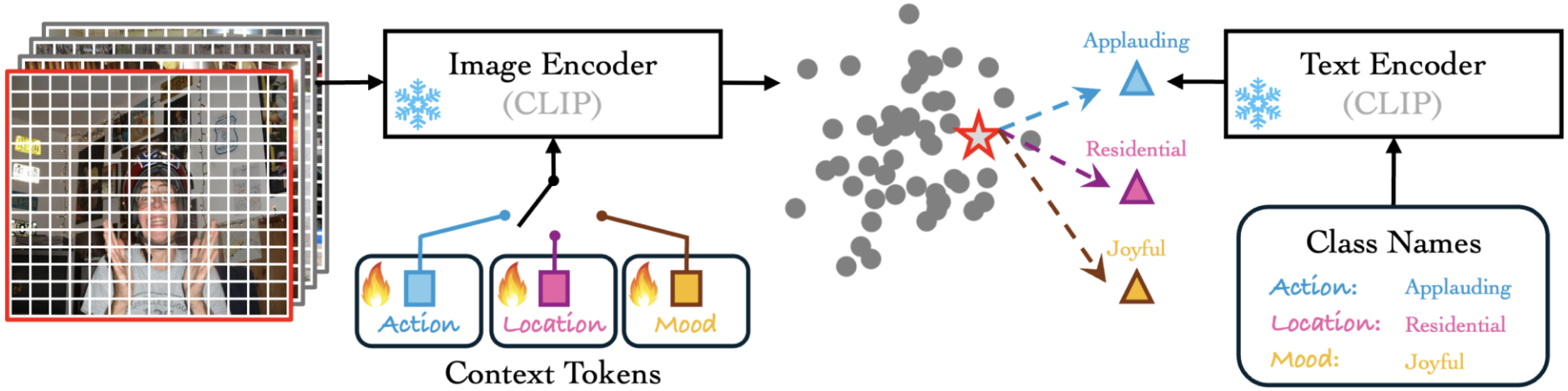
1. Build on CLIP to leverage its general knowledge
2. Tune only the **context token** to adapt visual features

Our OAK Framework



1. Build on CLIP to leverage its general knowledge
2. Tune only the **context token** to adapt visual features
3. Top-down guidance from **semantic cues**

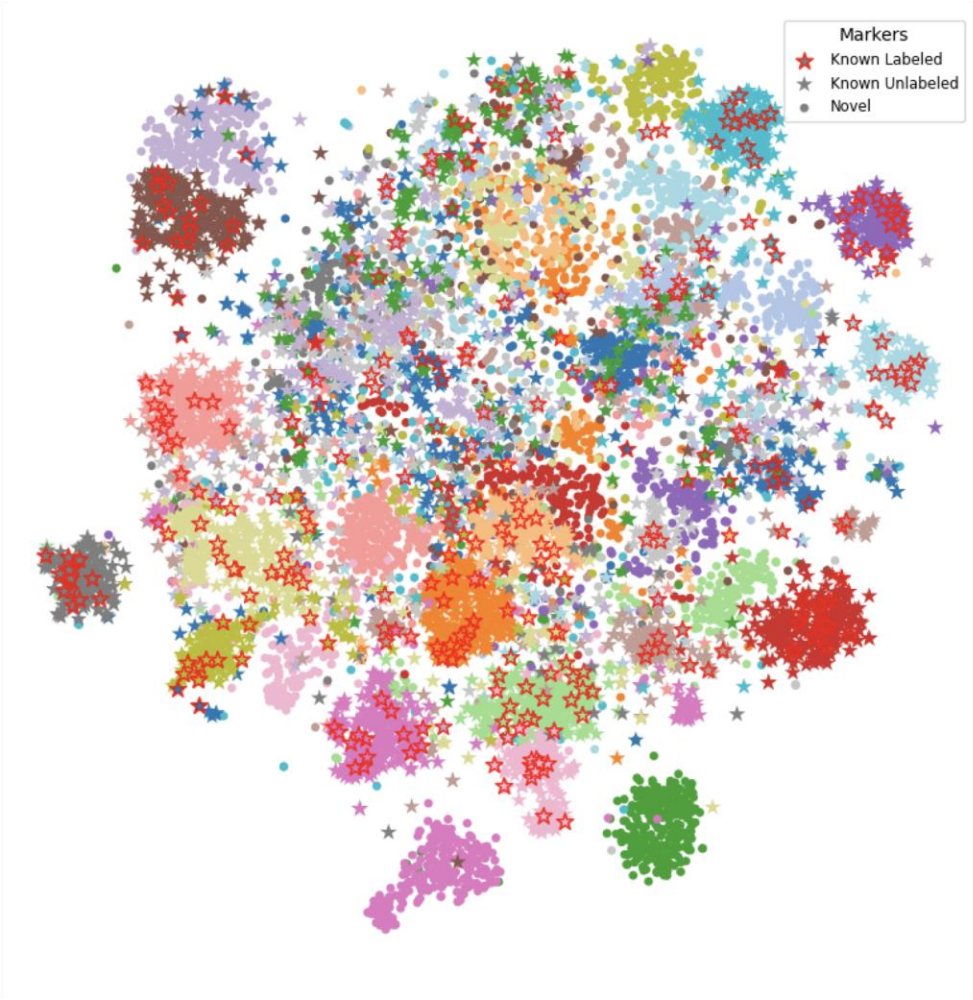
Our OAK Framework



1. Build on CLIP to leverage its general knowledge
2. Tune only the context token to adapt visual features
3. Top-down guidance from semantic cues
4. Bottom-up clustering of **visual cues** → novel clusters

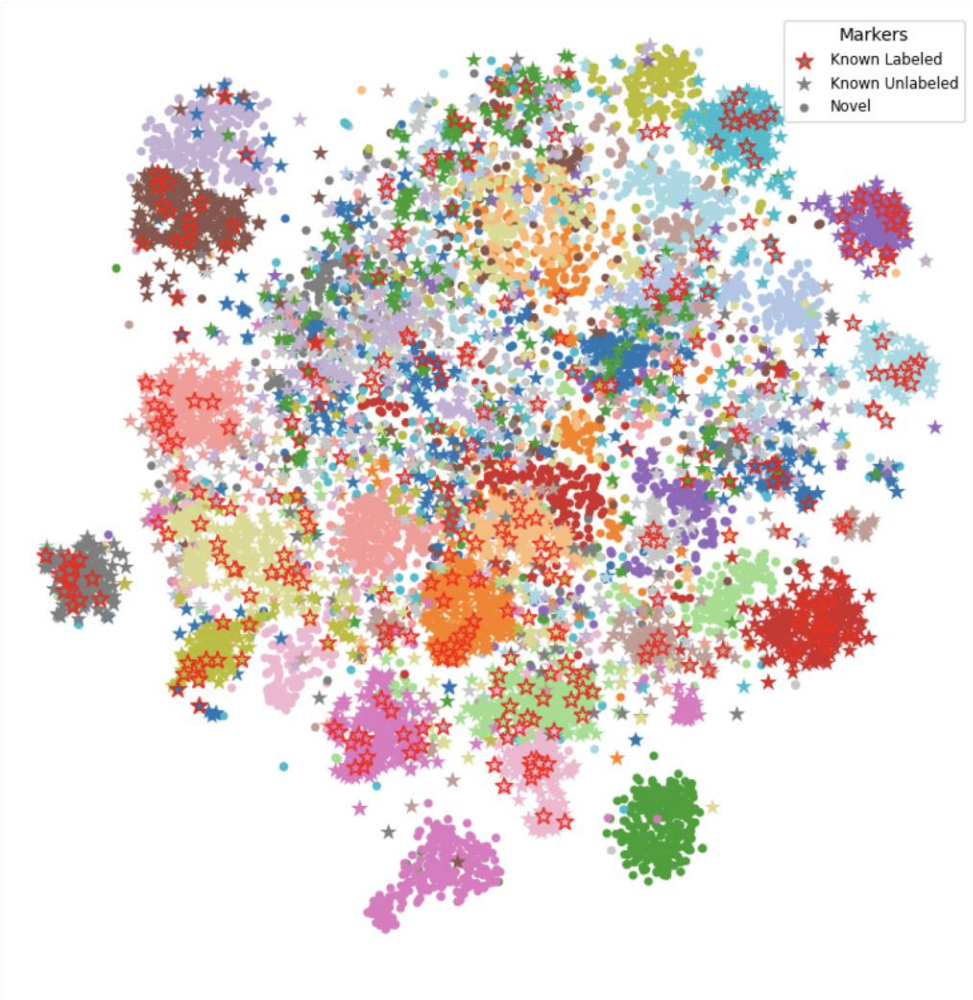
OAK **Contextualizes** Visual Features

Before (CLIP)

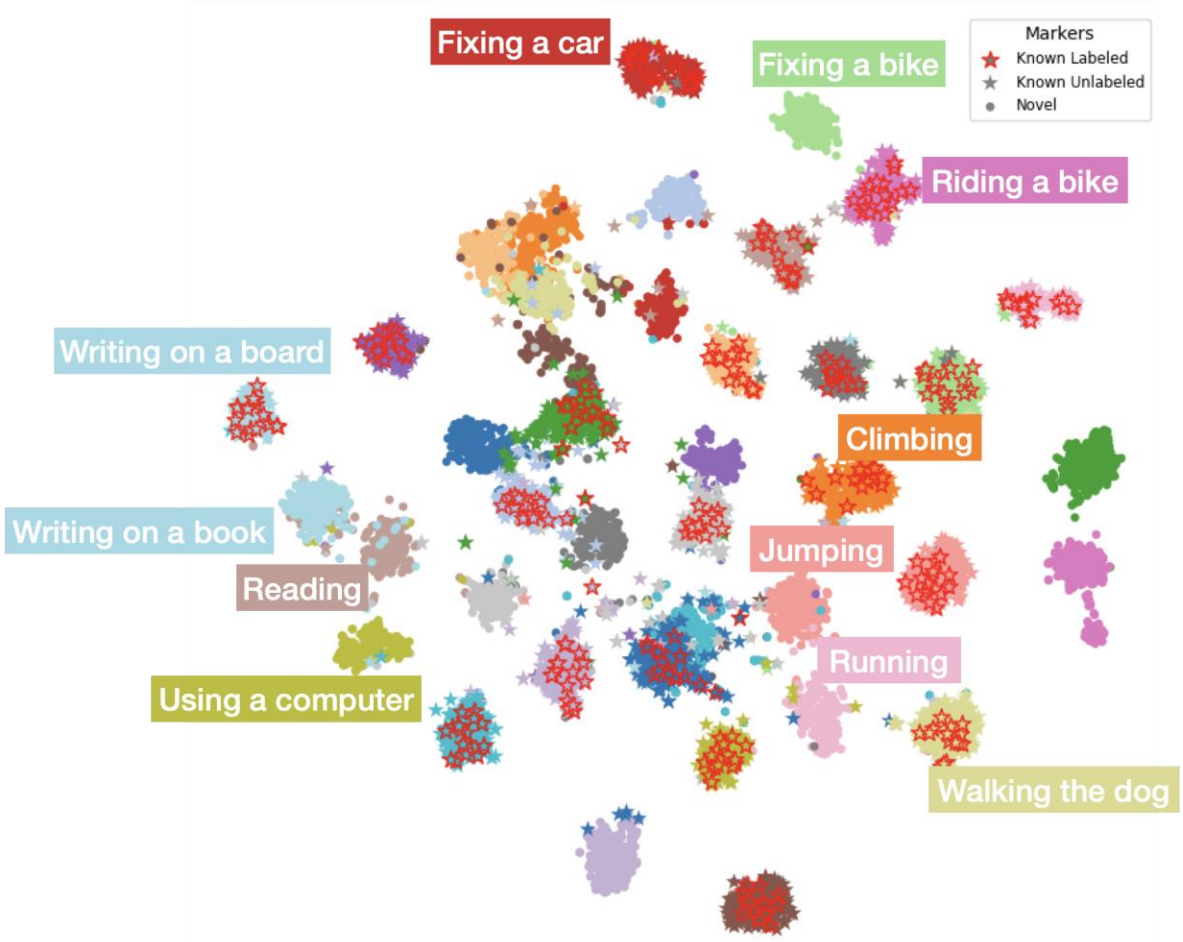


OAK **Contextualizes** Visual Features

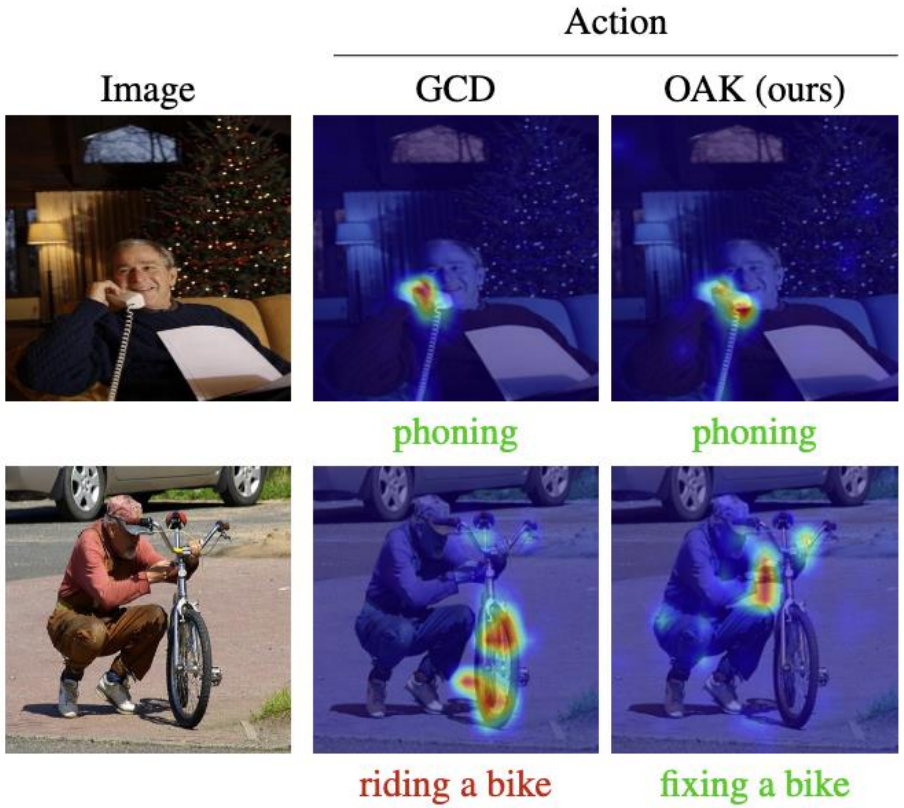
Before (CLIP)



After (OAK)

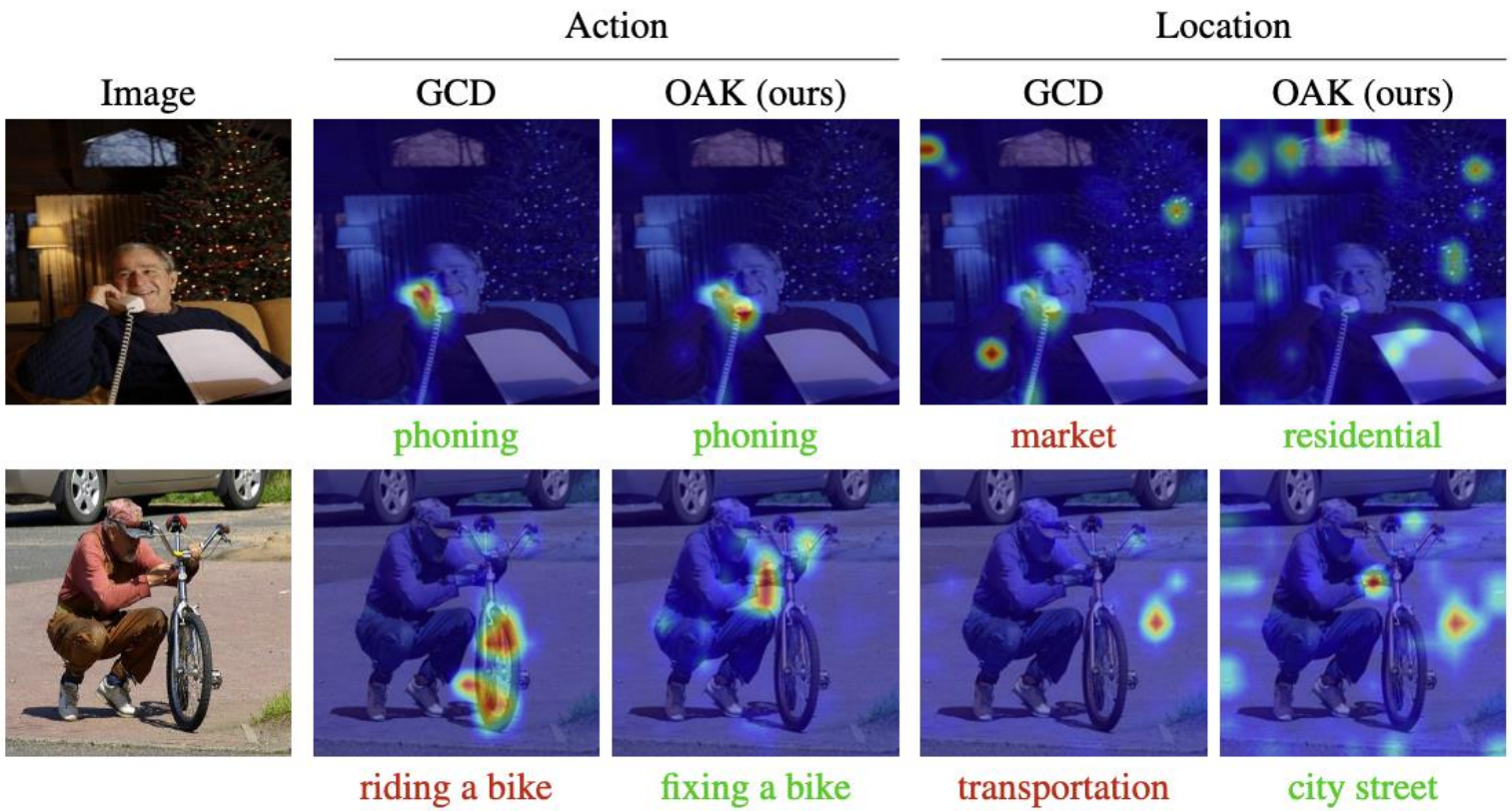


OAK Contextualizes **Visual Attention**



Focus on
hands

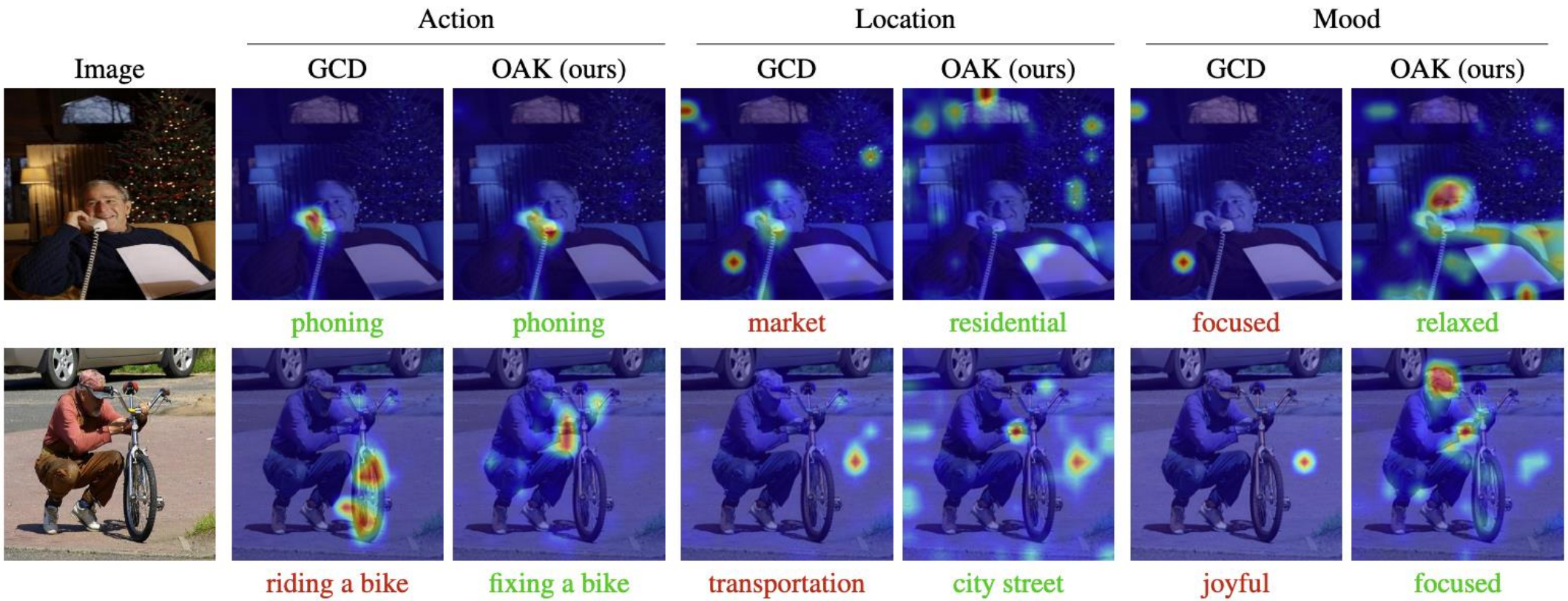
OAK Contextualizes **Visual Attention**



Focus on hands

Focus on **background**

OAK Contextualizes **Visual Attention**



Semantic & Visual Cues are Both Necessary

Semantic only

Method	Known				Novel				Overall			
	Act	Loc	Mood	Omni[a]	Act	Loc	Mood	Omni	Act	Loc	Mood	Omni
CLIP-ZS	**96.2**	60.4	**87.4**	75.0	-	-	-	-	-	-	-	-
CLIP-ZS + LLM vocab	93.5	58.3	75.9	75.0	38.6	34.2	35.4	0.0	65.2	47.5	55.0	43.0
CLIP-ZS + GT vocab	93.6	59.6	78.8	75.0	80.3	59.7	65.8	29.4	86.7	59.7	72.1	38.3
SS-KMeans	67.0	56.3	43.2	0.0	53.9	71.1	78.2	35.3	60.2	62.9	61.3	47.7
GCD	89.4	75.4	64.3	75.0	67.8	80.8	40.6	0.0	78.3	77.8	52.1	52.3
OAK (ours)	88.9	**83.9**	68.8	0.0	**85.1**	**88.4**	**87.4**	**47.1**	**86.9**	**85.9**	**78.4**	**70.3**

[a] With only 8 images overlapping across all contexts, text is toned down to light gray due to unreliable results.

Semantic & Visual Cues are Both Necessary

Semantic only

Visual only

Method	Known				Novel				Overall			
	Act	Loc	Mood	Omni[a]	Act	Loc	Mood	Omni	Act	Loc	Mood	Omni
CLIP-ZS	**96.2**	60.4	**87.4**	75.0	-	-	-	-	-	-	-	-
CLIP-ZS + LLM vocab	93.5	58.3	75.9	75.0	38.6	34.2	35.4	0.0	65.2	47.5	55.0	43.0
CLIP-ZS + GT vocab	93.6	59.6	78.8	75.0	80.3	59.7	65.8	29.4	86.7	59.7	72.1	38.3
SS-KMeans	67.0	56.3	43.2	0.0	53.9	71.1	78.2	35.3	60.2	62.9	61.3	47.7
GCD	89.4	75.4	64.3	75.0	67.8	80.8	40.6	0.0	78.3	77.8	52.1	52.3
OAK (ours)	88.9	**83.9**	68.8	0.0	**85.1**	**88.4**	**87.4**	**47.1**	**86.9**	**85.9**	**78.4**	**70.3**

[a] With only 8 images overlapping across all contexts, text is toned down to light gray due to unreliable results.

Semantic & Visual Cues are Both Necessary

	Known				Novel				Overall			
Method	Act	Loc	Mood	Omni[a]	Act	Loc	Mood	Omni	Act	Loc	Mood	Omni
Semantic only												
CLIP-ZS	**96.2**	60.4	**87.4**	75.0	-	-	-	-	-	-	-	-
CLIP-ZS + LLM vocab	93.5	58.3	75.9	75.0	38.6	34.2	35.4	0.0	65.2	47.5	55.0	43.0
CLIP-ZS + GT vocab	93.6	59.6	78.8	75.0	80.3	59.7	65.8	29.4	86.7	59.7	72.1	38.3
Visual only												
SS-KMeans	67.0	56.3	43.2	0.0	53.9	71.1	78.2	35.3	60.2	62.9	61.3	47.7
GCD	89.4	75.4	64.3	75.0	67.8	80.8	40.6	0.0	78.3	77.8	52.1	52.3
Use both												
OAK (ours)	88.9	**83.9**	68.8	0.0	**85.1**	**88.4**	**87.4**	**47.1**	**86.9**	**85.9**	**78.4**	**70.3**

[a] With only 8 images overlapping across all contexts, text is toned down to light gray due to unreliable results.

OAK Can **Name** the Discovered Clusters

a) Stanford	Action		Location		Mood	
GT Label	cooking	reading	restaurant or dining area	urban area or city street	focused	relaxed
Prediction	preparing a meal	reading a book	commercial kitchen	suburban street	explorative	admiring

b) Clevr-4	Texture		Color		Shape		Count	
GT Label	zigzag	circles	pink	cyan	star	diamond	4	9
Prediction	wavy lines	checkerboard	pink	turquoise	star shape	diamond shape	4	17

OAK Can **Name** the Discovered Clusters

a) Stanford	Action		Location		Mood	
GT Label	cooking	reading	restaurant or dining area	urban area or city street	focused	relaxed
Prediction	preparing a meal	reading a book	commercial kitchen	suburban street	explorative	admiring

b) Clevr-4	Texture		Color		Shape		Count	
GT Label	zigzag	circles	pink	cyan	star	diamond	4	9
Prediction	wavy lines	checkerboard	pink	turquoise	star shape	diamond shape	4	17

jumping named as *dancing*



pouring liquid named as *carrying a box*



Part 3. **Future Research & Teaching Plan**

Two-track Research & Education

AI Core

AI + Business

Two-track Research & Education

AI Core

Current IME dept:
Focus on fundamentals
(math, stats, code)

+ My expertise:
Latest AI technology

AI + Business

Two-track Research & Education

AI Core	AI + Business
Current IME dept: Focus on fundamentals (math, stats, code)	**Current IME dept:** Focus on classic sectors (e.g., manufacturing)
+ My expertise: Latest AI technology	**+ My expertise:** New business insights

Two-track Research & Education

AI Core

Deep experience

Google

Meta

NVIDIA

UNIVERSITY OF MICHIGAN

NAVER

kakao

AI + Business

Two-track Research & Education

AI Core

Deep experience

Google

Meta

NVIDIA

UNIVERSITY OF MICHIGAN

NAVER

kakao

AI + Business

Wide experience

MSSA
Management Strategy Student Association

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

PEKING UNIVERSITY 1898

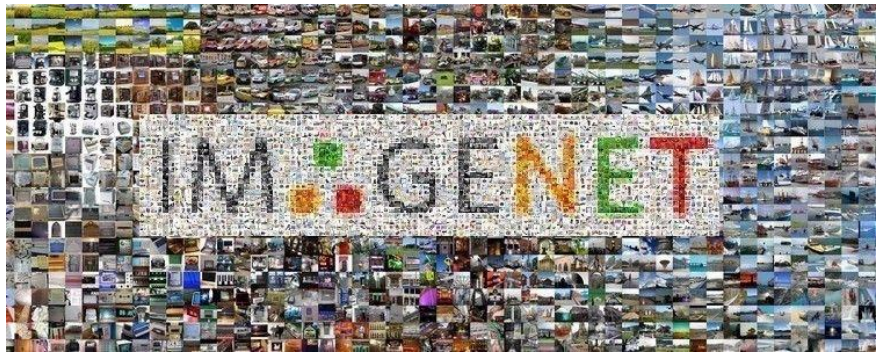
BCG

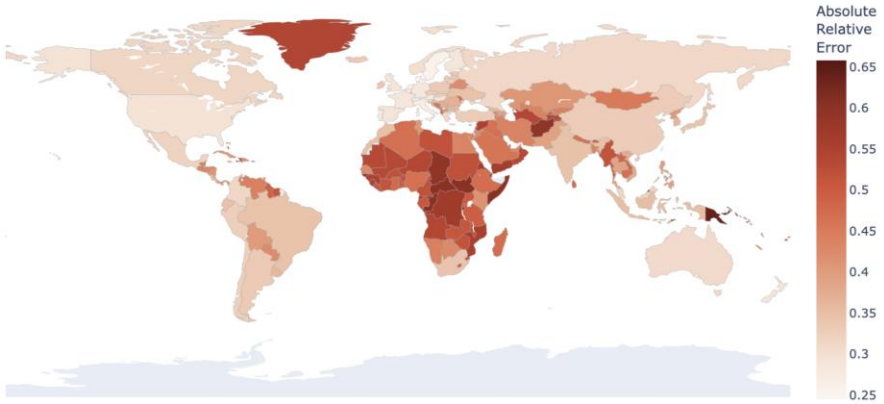
FAST TRACK ASIA

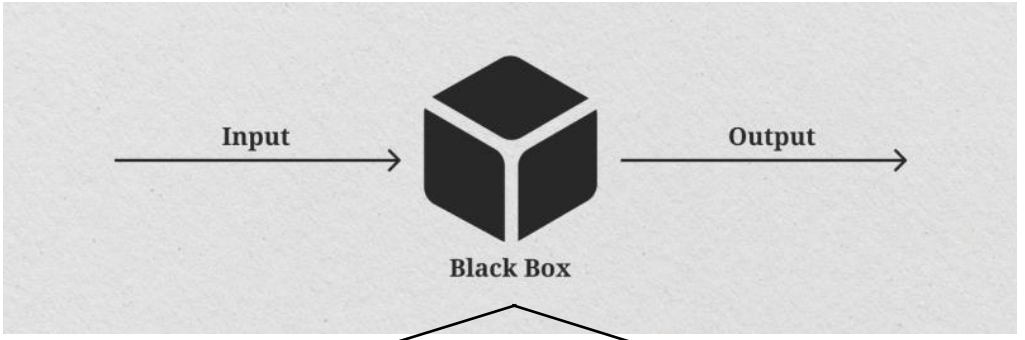
NRI

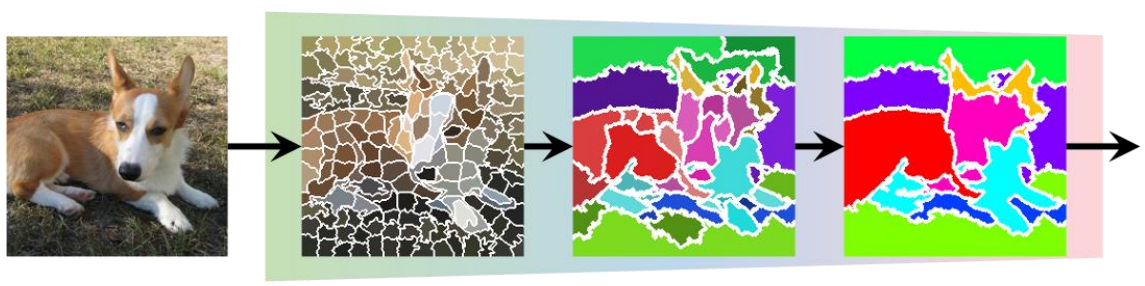
GE x McKINSEY
LEADERSHIP WORKSHOP

AI Core - Extend Scalable Priors









Statistical priors in data manifolds

Structural priors in data instances

AI Core - Ongoing Projects



Generative models



Vision-language



3D vision



Reasoning



Robotics

AI Core - Community Service

CVPR'25 Pixel-level Vision Foundation Model (PixFoundation) Workshop



Antonio Torralba
(MIT)



Cordelia Schmid
(INRIA)



Hengshuang Zhao
(HKU)



Ishan Misra
(Meta)

ICCV'25 Structural Priors for Vision (SP4V) Workshop



Danfei Xu
(Georgia Tech)



Joao Carreira
(DeepMind)



Jiajun Wu
(Stanford)



Kristen Grauman
(UT Austin)



Saining Xie
(NYU)

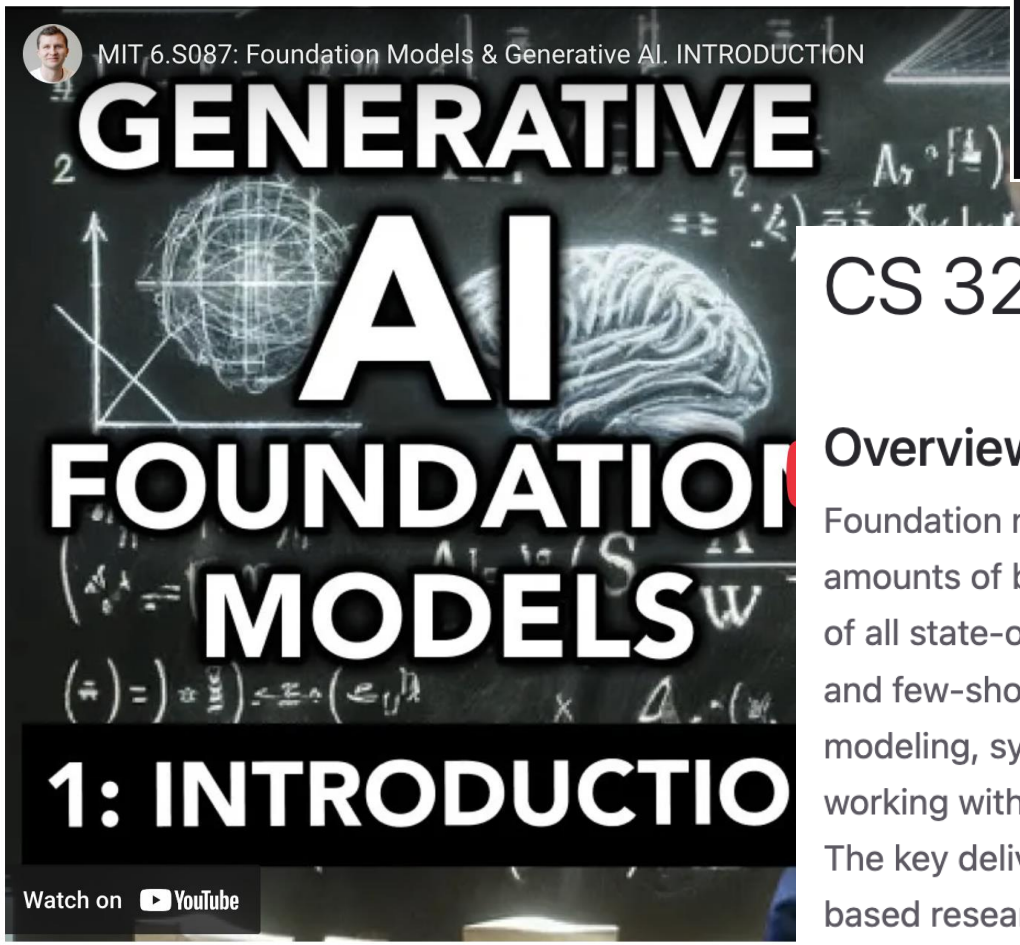


Vincent Sitzmann
(MIT)

AI Core - Potential Lectures

MIT FUTURE OF AI

6.S087: Foundation Models & Generative AI





16-892
Seminar on Multimodal Foundation Models
Fall 2024

CS 324 - Advances in Foundation Models

Overview

Foundation models (FMs) are models (e.g., DALL-E, GPT-3, Stable Diffusion) that are trained on large amounts of broad data and are adaptable to a wide range of downstream tasks. They form the basis of all state-of-the-art systems across a wide range of tasks and have shown impressive generative and few-shot learning abilities. In this course, students will learn the fundamentals about the modeling, systems and ethical aspects of foundation models, as well as gain hands-on experience working with them. In addition, the course will feature speakers from industry working on these FMs. The key deliverable will be a quarter-long project through which students will design their own FM-based research project or application, targeting a problem they care about.

AI + Business - Potential Lectures

Harvard Business School Online

Courses For Organizations Insights More Info



AI Essentials for Business

Artificial Intelligence in Business: Creating Value with Machine Learning

Leverage new technologies to build value for your organization.

Competing in the Age of AI

Discover how artificial intelligence is transforming today's businesses and how you can leverage it to grow your company into an AI-first leader.

AI + Business - Potential Collaborations

APGC-Lab

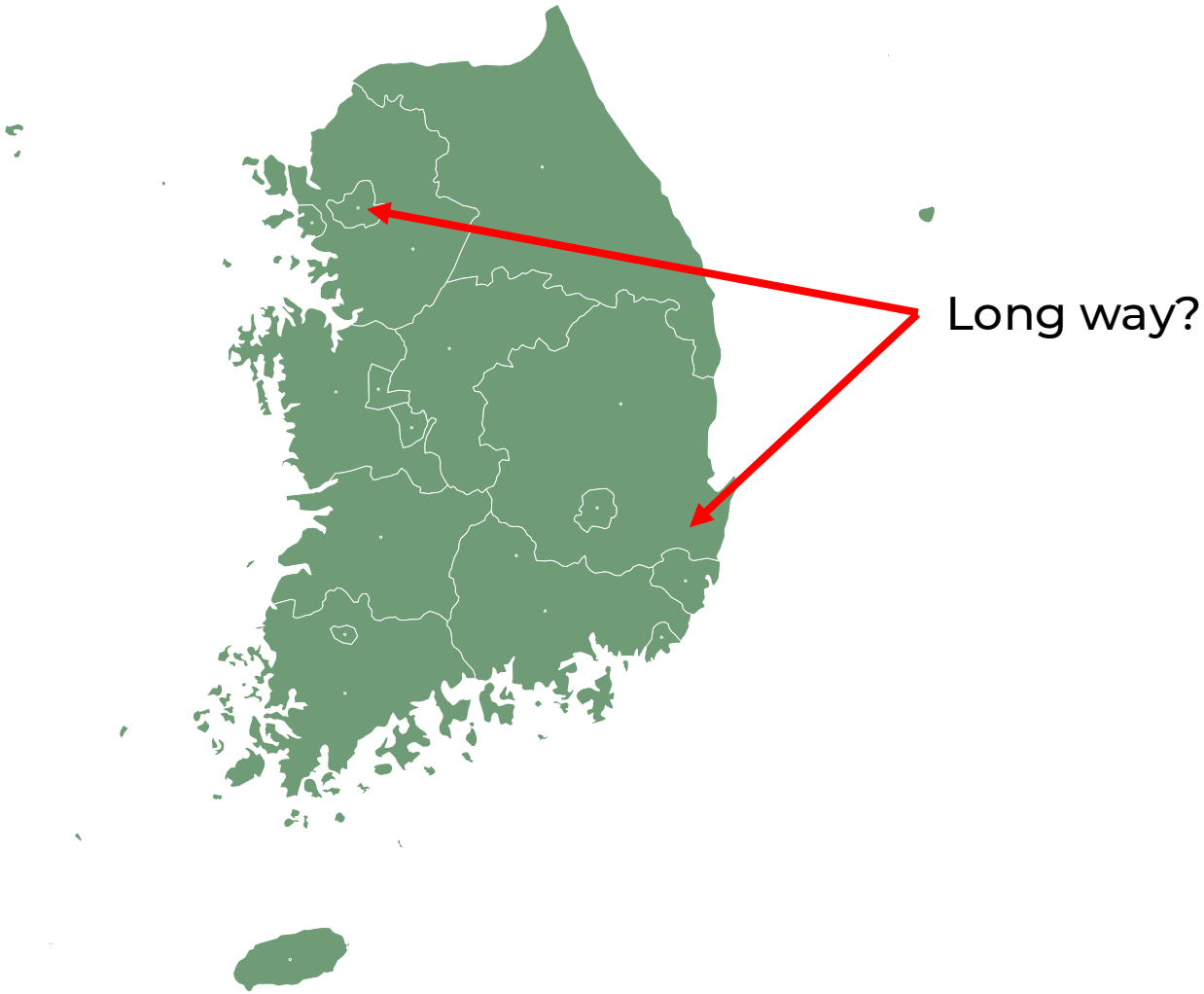
산업경영공학과



컴퓨터공학과

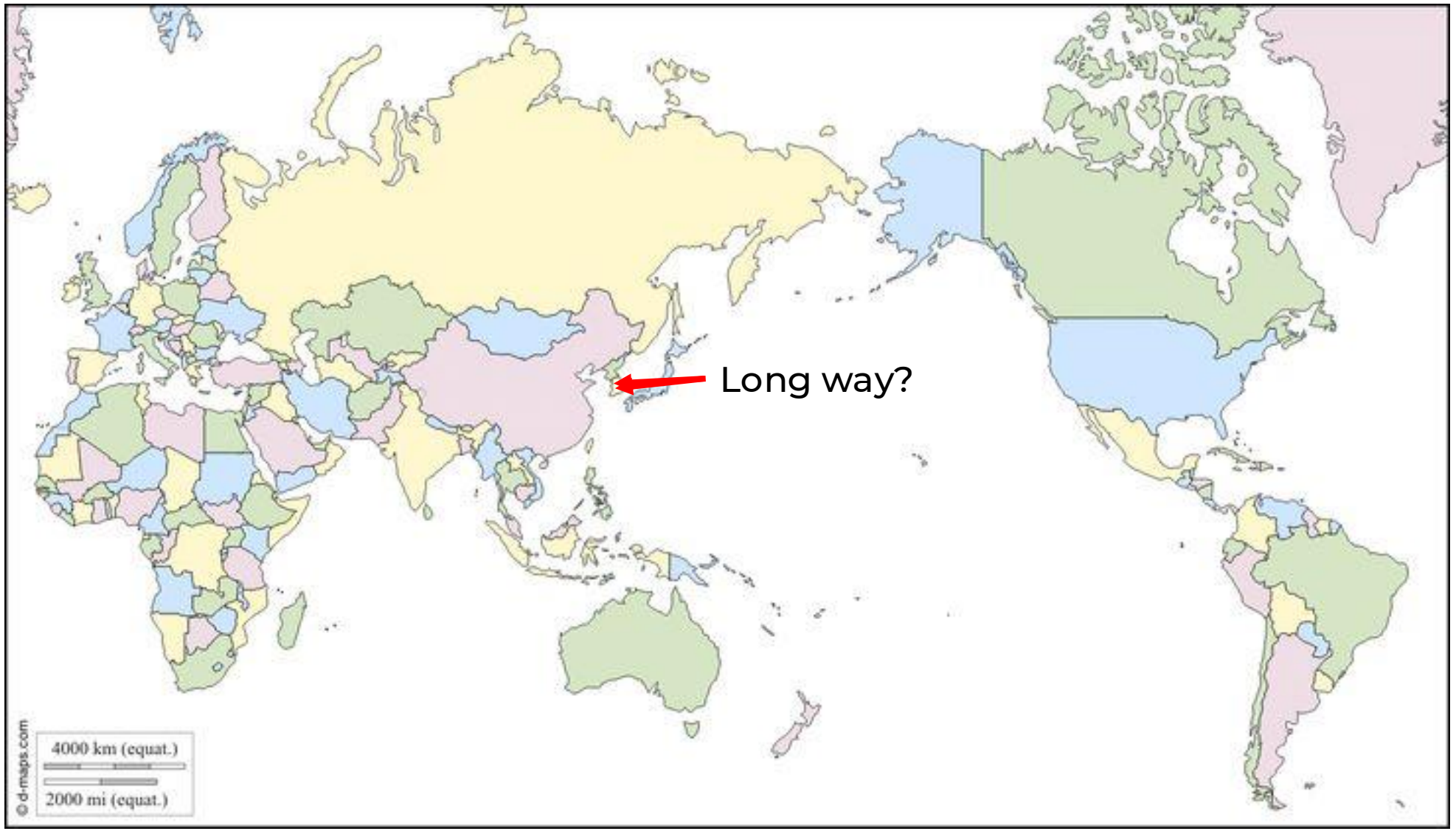


Global POSTECH



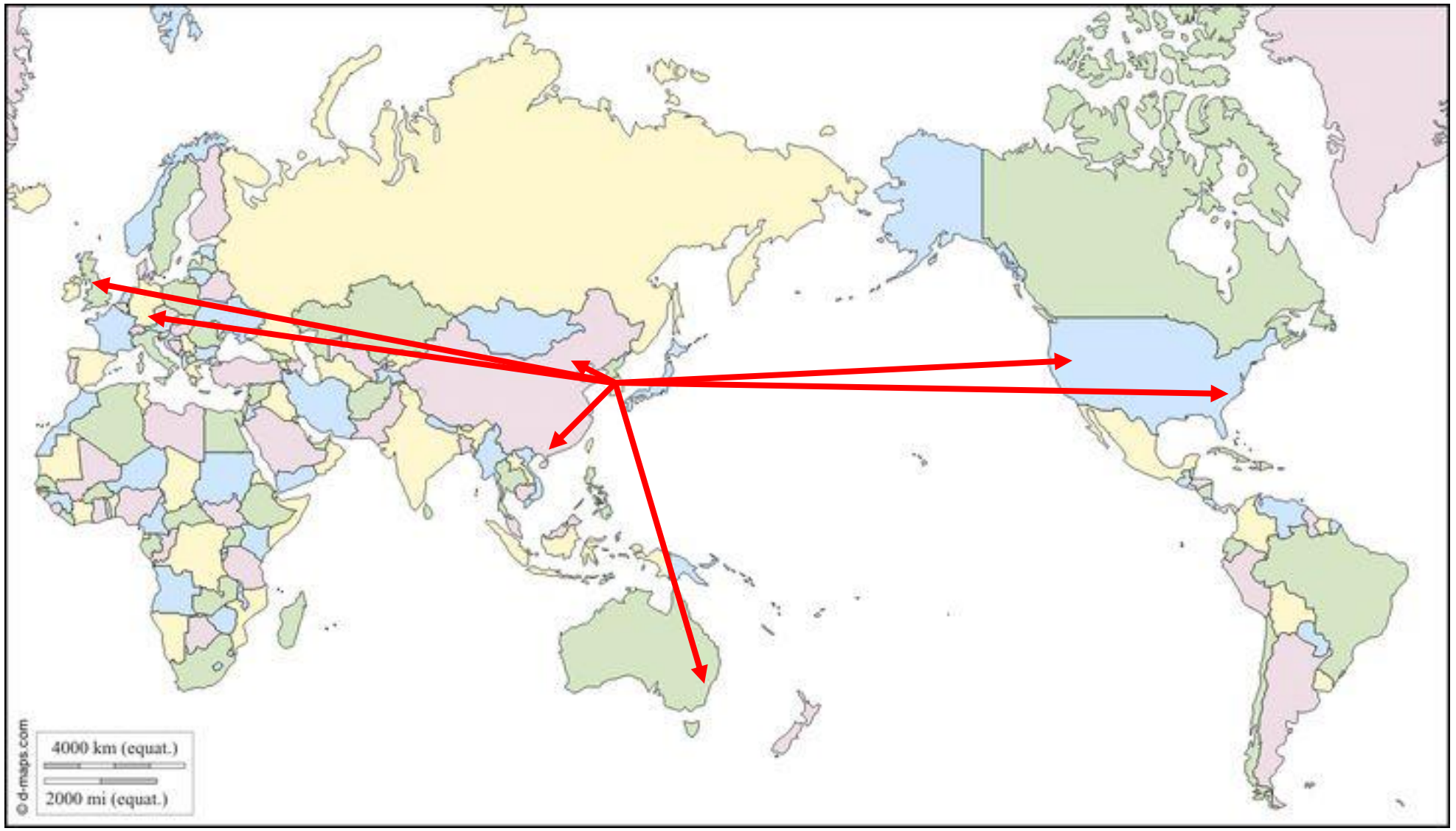
Source: https://uk.pinterest.com/pin/world-pacific-ocean-centered-free-map-free-blank-map-free-outline-map-free--493988652871838315/

Global POSTECH



Source: https://uk.pinterest.com/pin/world-pacific-ocean-centered-free-map-free-blank-map-free-outline-map-free--493988652871838315/

I hope I can contribute to Global POSTECH



Source: https://uk.pinterest.com/pin/world-pacific-ocean-centered-free-map-free-blank-map-free-outline-map-free--493988652871838315/