

Reasoning About Object Affordances in a Knowledge Base Representation

Yuke Zhu, Alireza Fathi, and Li Fei-Fei
ECCV14

Presented by Fereshteh Sadeghi
fsadeghi@cs.washington.edu

Affordance

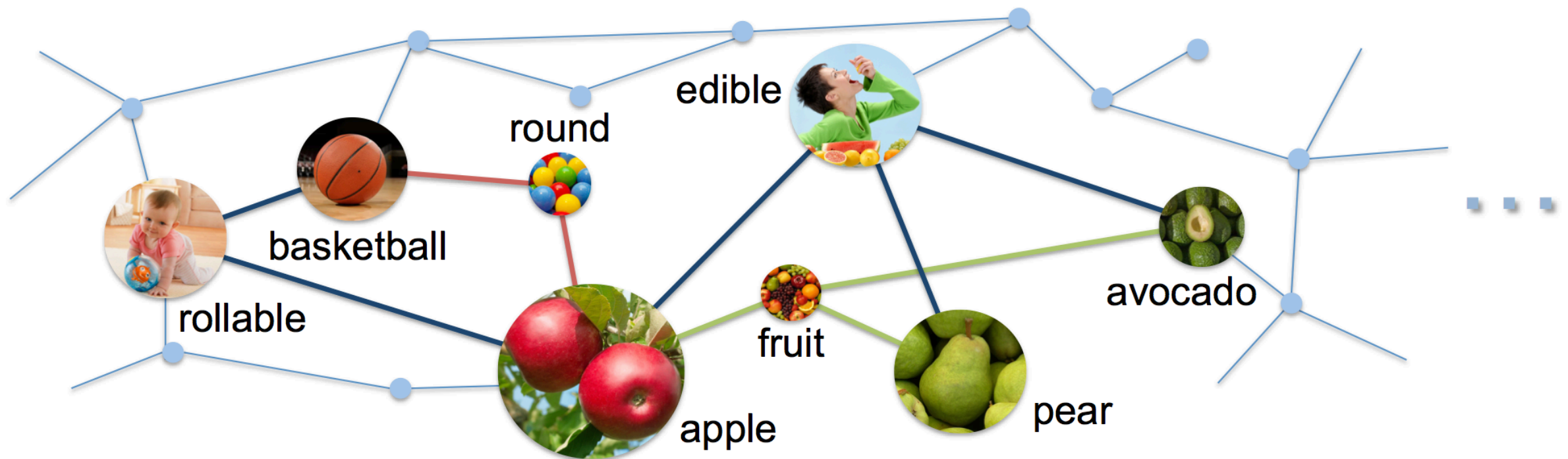
“Properties of an object [...] that determine what actions a human can perform on them”

--Gibson 1979

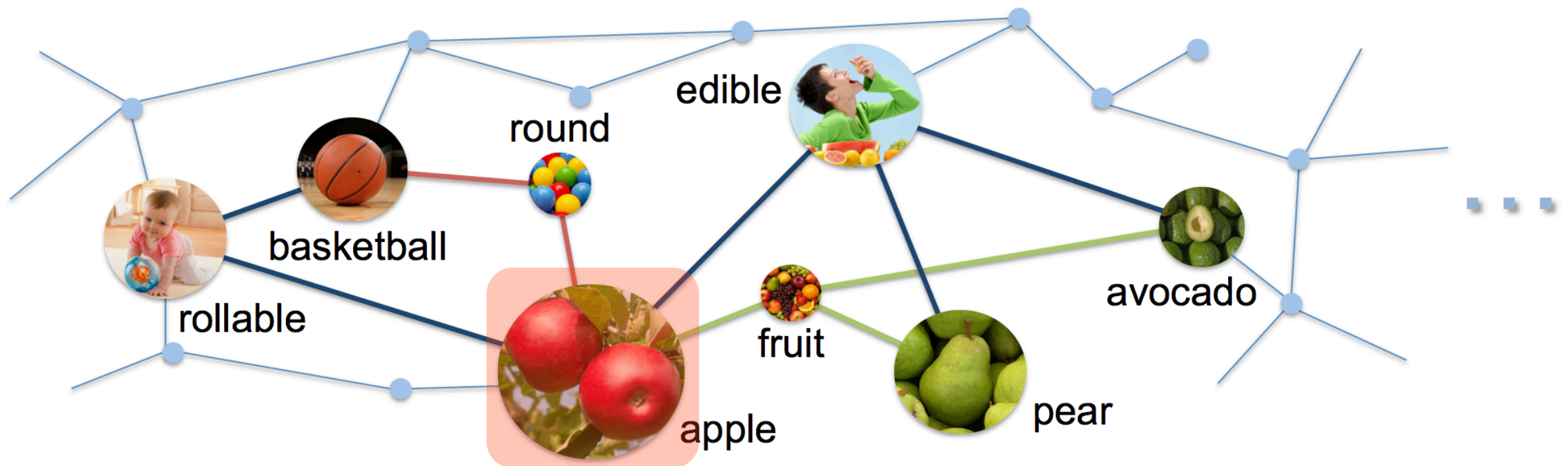
Affordance

- Combination of :
 - An affordance label (e.g. edible)
 - A human pose representation of the action (e.g. skeleton form)
 - Relative position of the object with respect to human pose (e.g. next to)

Knowledge structure

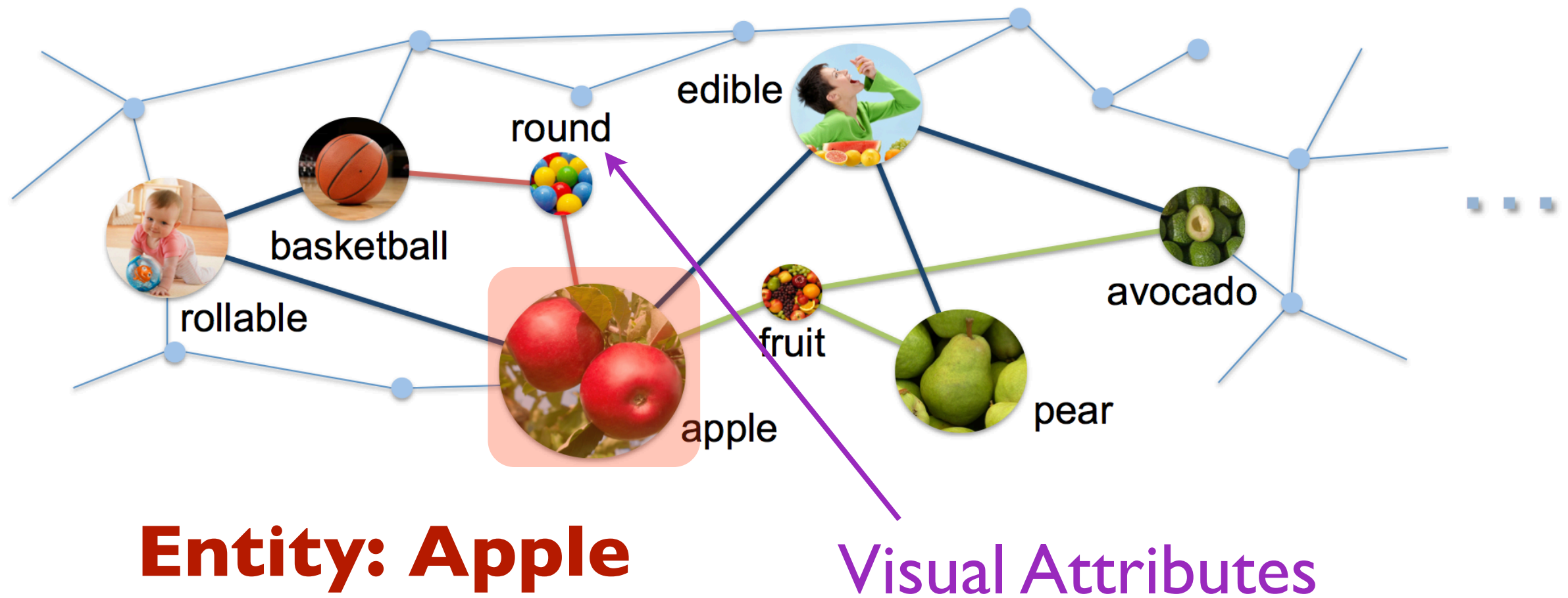


Knowledge structure

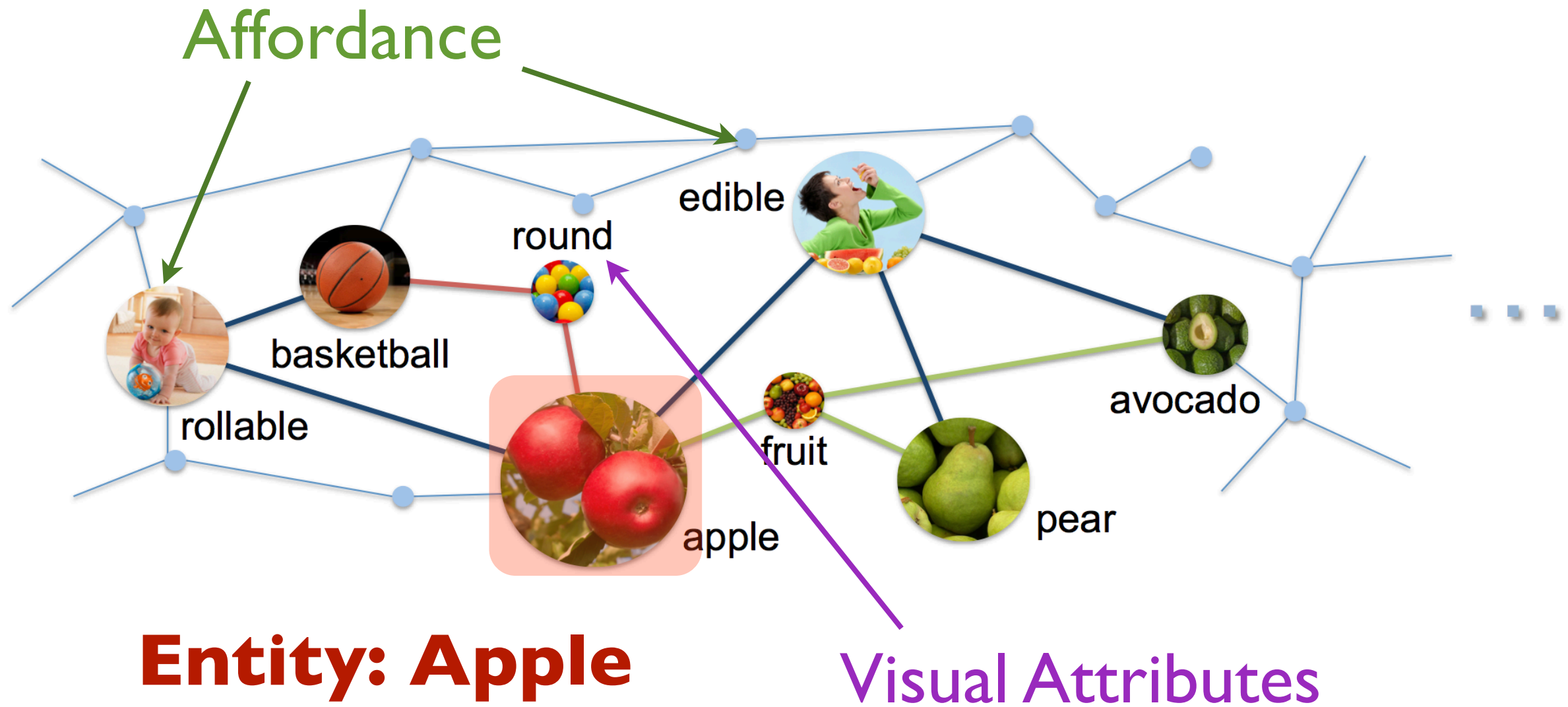


Entity: Apple

Knowledge structure



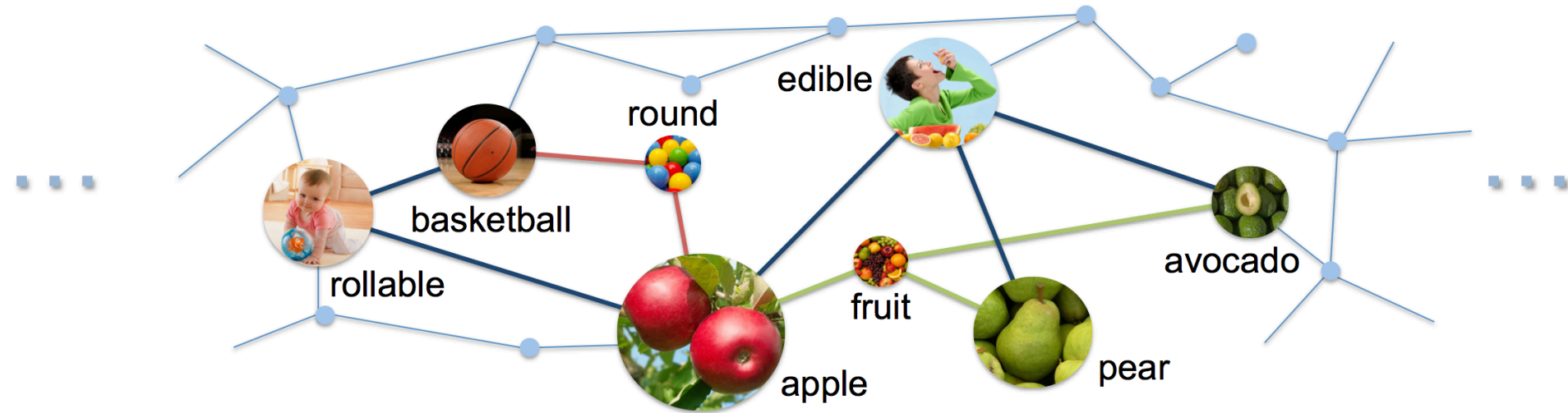
Knowledge structure



Main Goal

- Predict affordances of unseen objects
- Infer richer information beyond visual similarity
- Knowledge based approach for reasoning and answering various types of questions

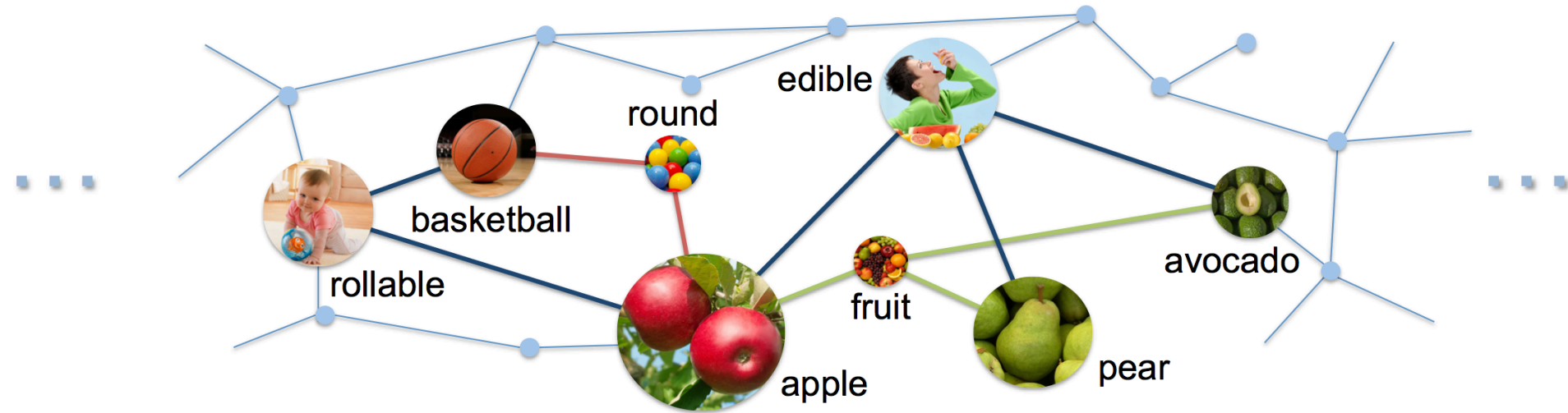
Knowledge Base (KB)



Nodes: Entities

Edges: General rules to characterize relations

Knowledge Base (KB)

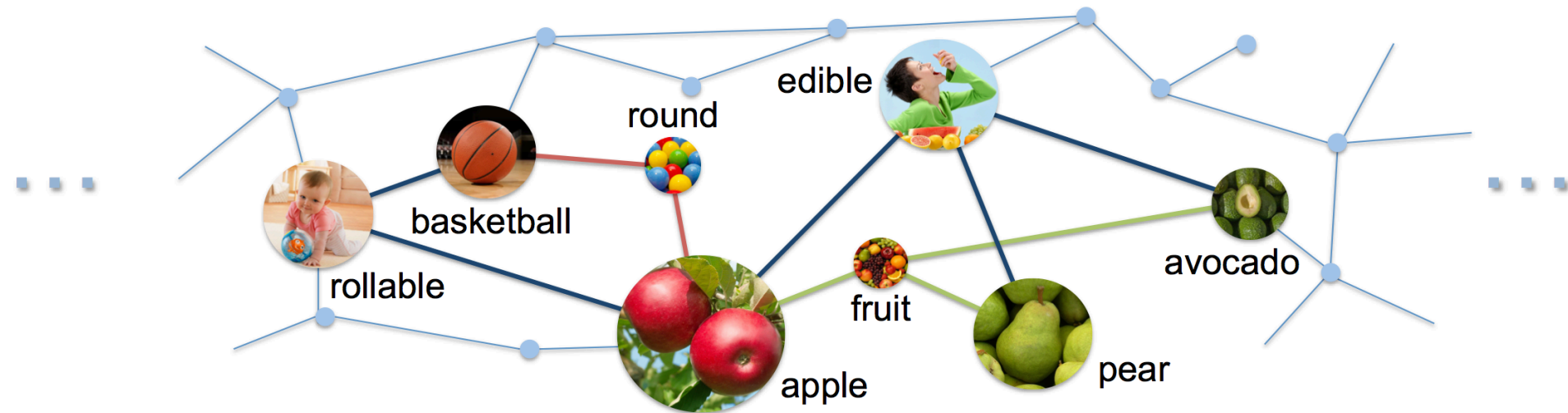


Nodes: Entities

- Visual attributes (e.g. round)
- Physical attributes (weight & size)
- Categorical attributes (e.g. apple)
- Affordance (e.g. edible)

Edges: General rules to characterize relations

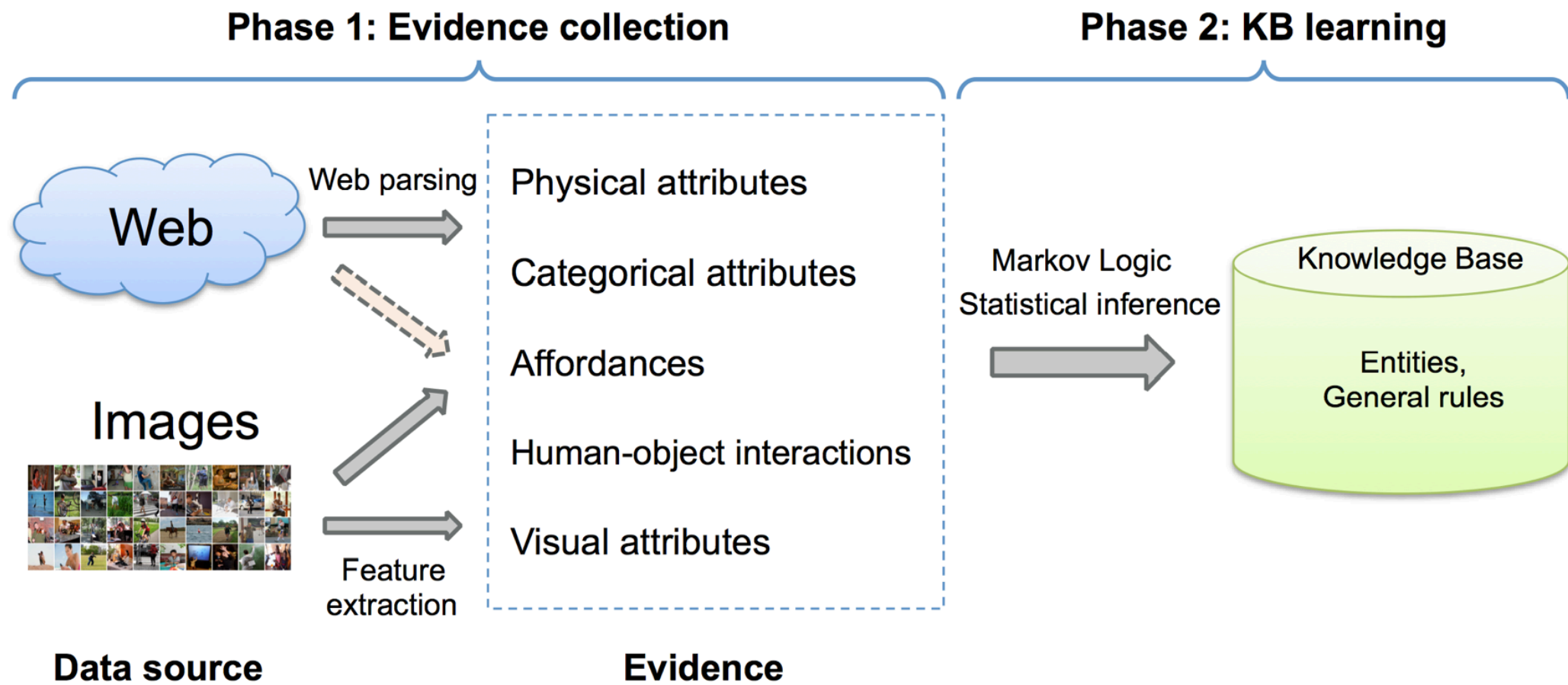
Knowledge Base (KB)



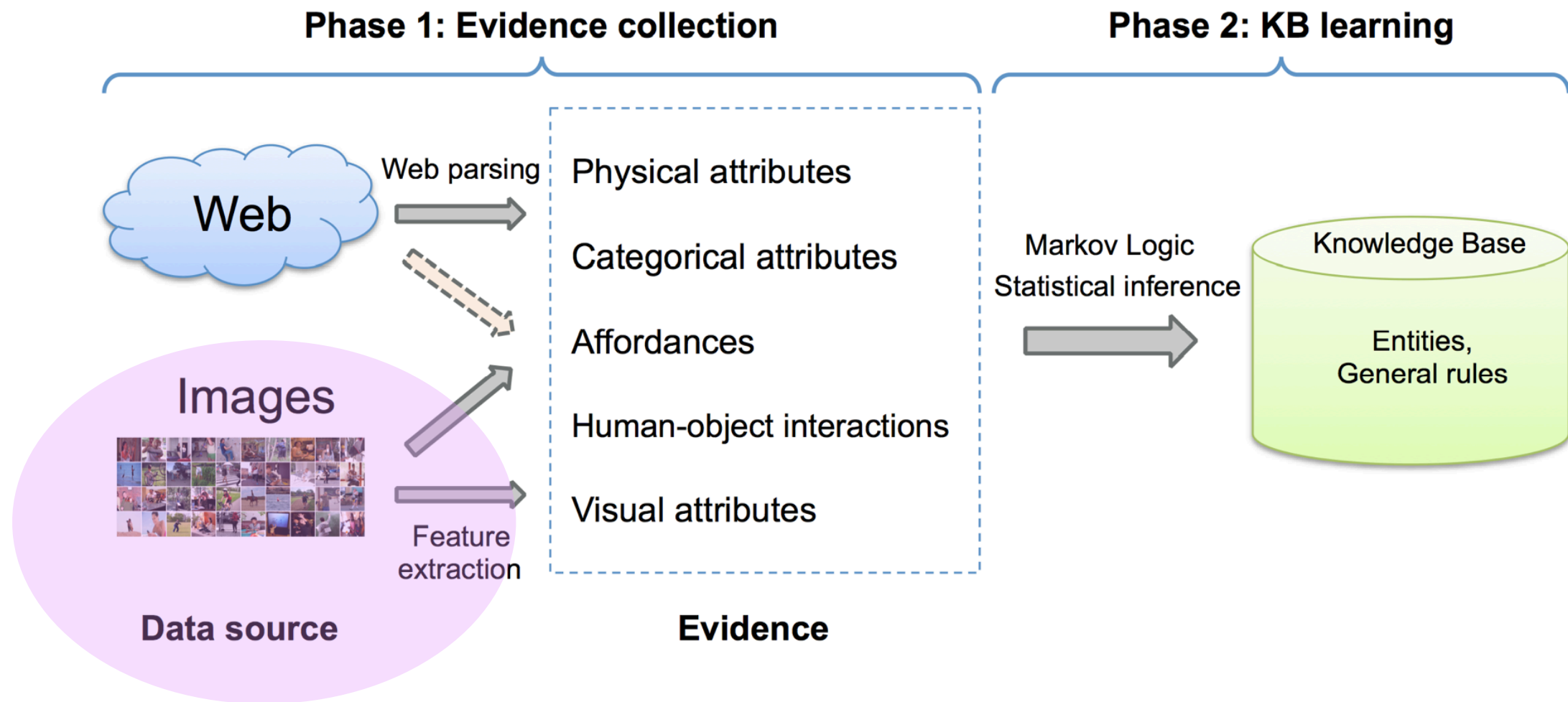
Nodes: Entities

Edges: General rules to characterize relations

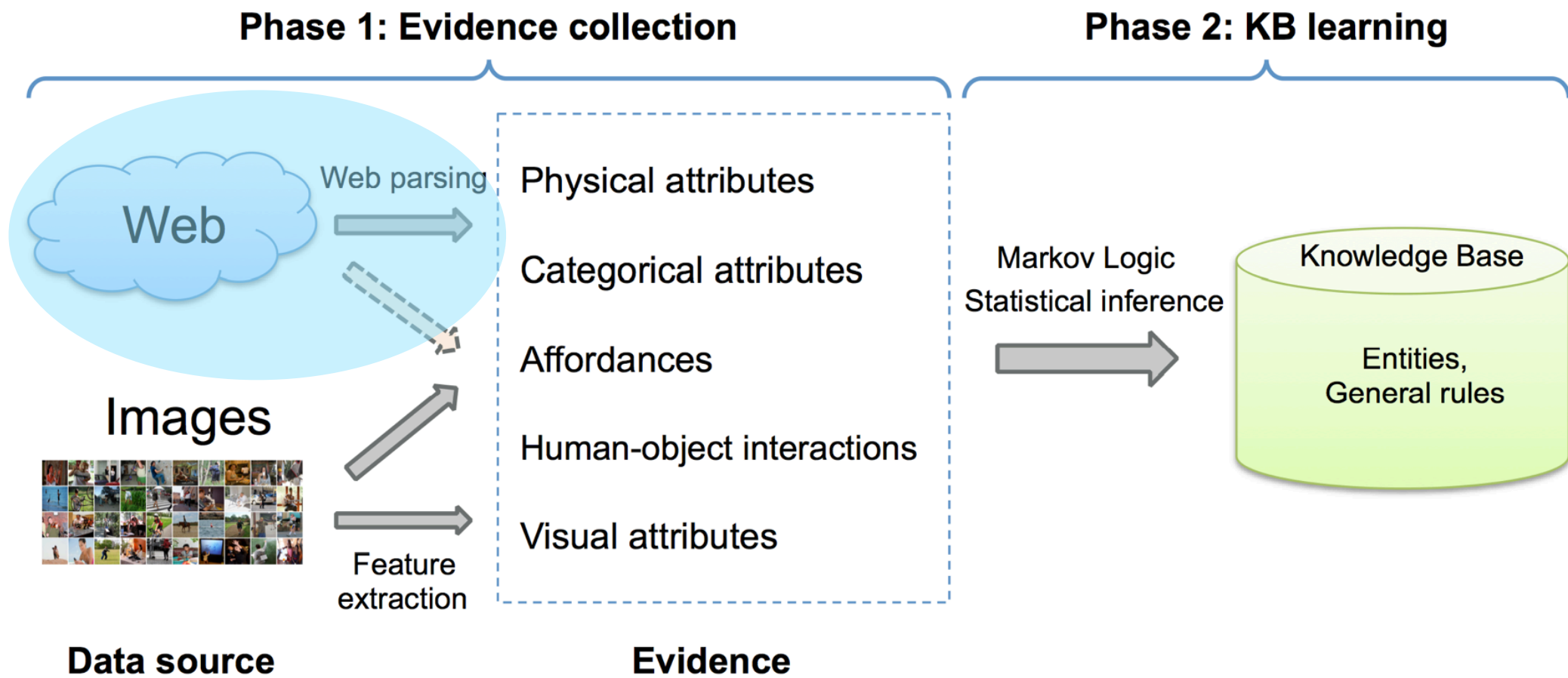
- attribute-attribute
- attribute-affordance
- human-object-interaction
 - attribute-pose, affordance-pose, attribute-location, affordance-location



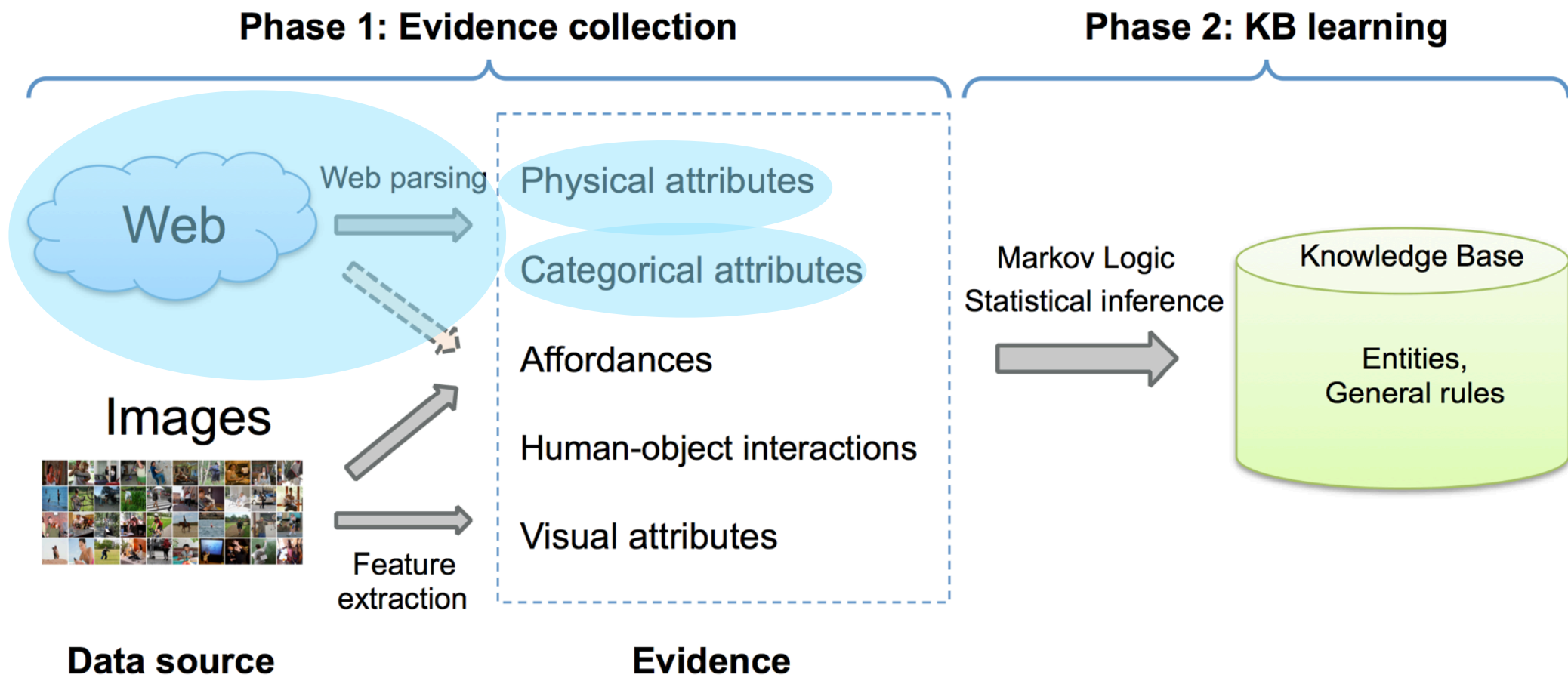
System overview



- Seed KB with 40 objects & actions (Stanford 40 dataset)
- 14 affordance (Stanford 40 dataset)
- 100 images for each object (ImageNet)



- WordNet: hypernym hierarchy
- Freebase: animal synopsis
- Amazon & eBay: physical attributes (weight & size)



- WordNet: hypernym hierarchy

- Freebase: animal synopsis

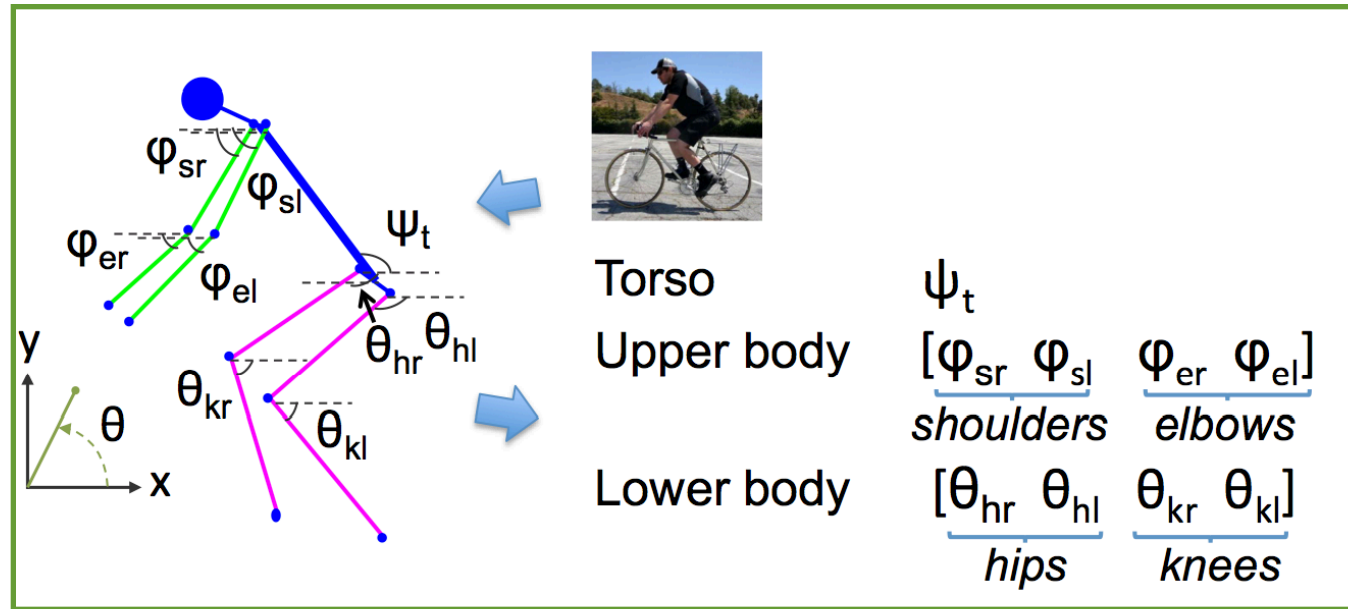
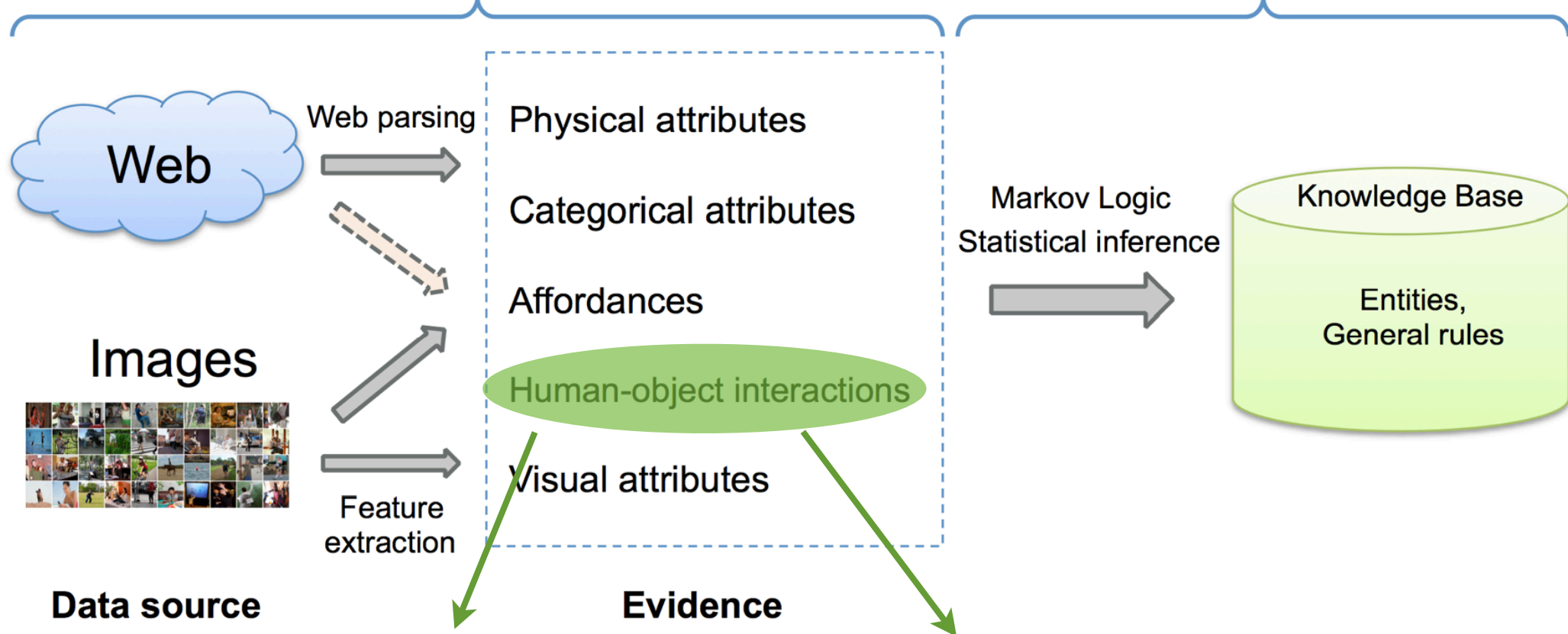
- Amazon & eBay: Physical attributes (weight & size)

chalk, pen, bottle, frisbee, toothbrush, can, handset, mobile phone, hand saw, food turner, fishing pole, umbrella, camera, cleaver, pitcher, carving knife, dustcloth, teapot, laptop, axe, dish, microscope, power saw, violin, guitar, telescope, mop, television, vacuum cleaner, desktop computer, small boat, car tire, chair, wheelbarrow, dog, bicycle, sofa, shopping cart, automobile engine, horse

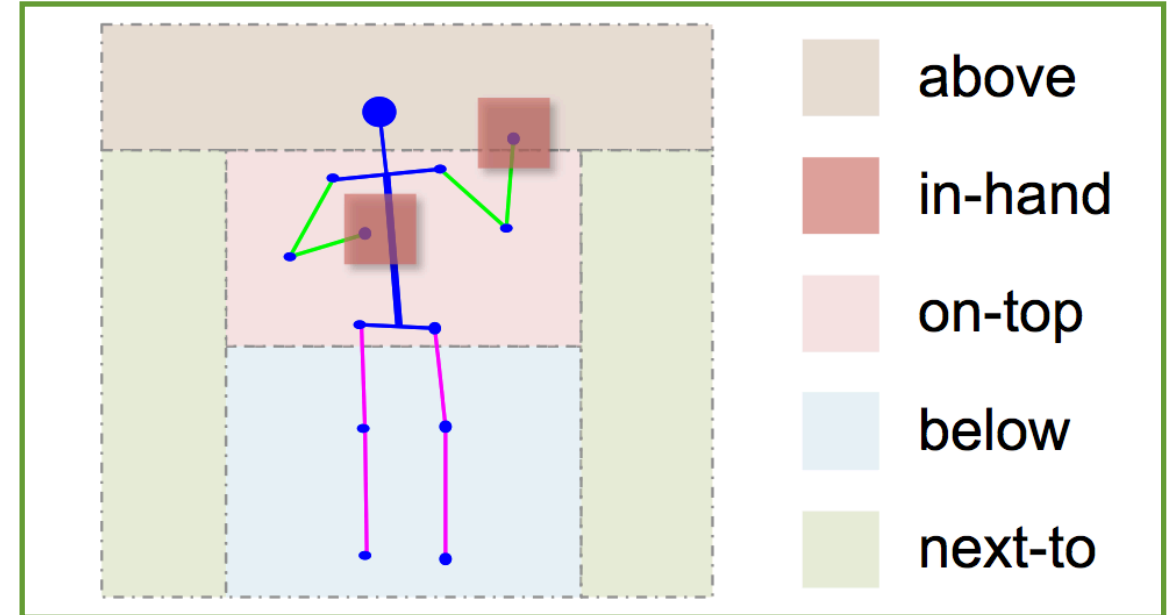
■ <1kg ■ 1~10kg ■ 10~100kg ■ >100 kg

Phase 1: Evidence collection

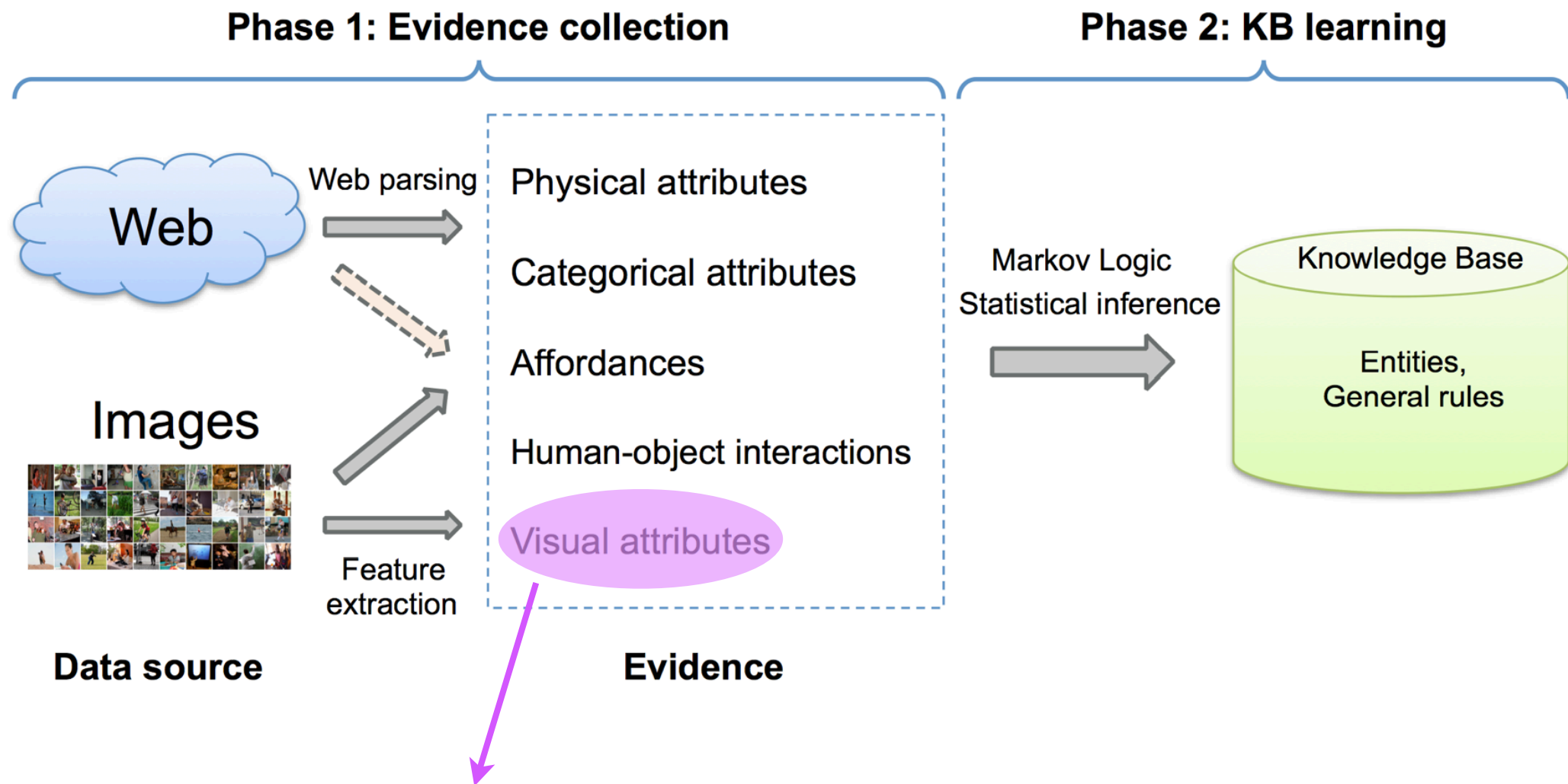
Phase 2: KB learning



Pose descriptor



Relative Locations



33 pre-trained visual attribute classifier
 Describe shape, material & parts of objects

Learning KB using Markov Logic

Schema

$\text{hasAffordance}(\text{object}, \text{affordance})$
 $\text{isA}(\text{object}, \text{category})$
 $\text{hasVisualAttribute}(\text{object}, \text{attribute})$
 $\text{hasWeight}(\text{object}, \text{weight})$
 $\text{hasSize}(\text{object}, \text{size})$
 $\text{locate}(\text{object}, \text{location})$
 $\text{torso}(\text{object}, \text{torso_id})$
 $\text{upperBody}(\text{object}, \text{ubody_id})$
 $\text{lowerBody}(\text{object}, \text{lbody_id})$

General Rules

Attribute-attribute relations
 Attribute-affordance relations
 Human-object-interaction relations

Examples

$\text{isA}(x, \text{Vehicle}) \Rightarrow \text{isA}(x, \text{Animal})$
 $\text{hasVisualAttribute}(x, \text{Furry}) \Rightarrow \text{hasAffordance}(x, \text{Feed})$
 $\text{hasWeight}(x, \text{W4}) \Rightarrow \text{hasAffordance}(x, \text{SitOn})$
 $\text{hasAffordance}(x, \text{Ride}) \wedge \text{locate}(x, \text{Below})$
 $\text{isA}(x, \text{Animal}) \wedge \text{locate}(x, \text{Below})$
 $\text{hasAffordance}(x, \text{Push}) \wedge \text{torso}(x, \text{T1})$
 $\text{isA}(x, \text{Vehicle}) \wedge \text{upperBody}(x, \text{U3})$

- Markov Logic Network (MLN)
- Unify MRF with first-order logic

$$P(X = \mathbf{x}) = \frac{1}{Z} \exp \left(\sum_{i=1}^n w_i f_i(x_{\{i\}}) \right)$$

Possible worlds

feature function

$$f_i(x_{\{i\}}) = 1 \text{ if } F_i(x_{\{i\}}) \text{ is true}$$

Learning KB using Markov Logic

Schema

hasAffordance(*object*, *affordance*)
isA(*object*, *category*)
hasVisualAttribute(*object*, *attribute*)
hasWeight(*object*, *weight*)
hasSize(*object*, *size*)
locate(*object*, *location*)
torso(*object*, *torso_id*)
upperBody(*object*, *ubody_id*)
lowerBody(*object*, *lbody_id*)


General Rules

Attribute-attribute relations
Attribute-affordance relations
Human-object-interaction relations

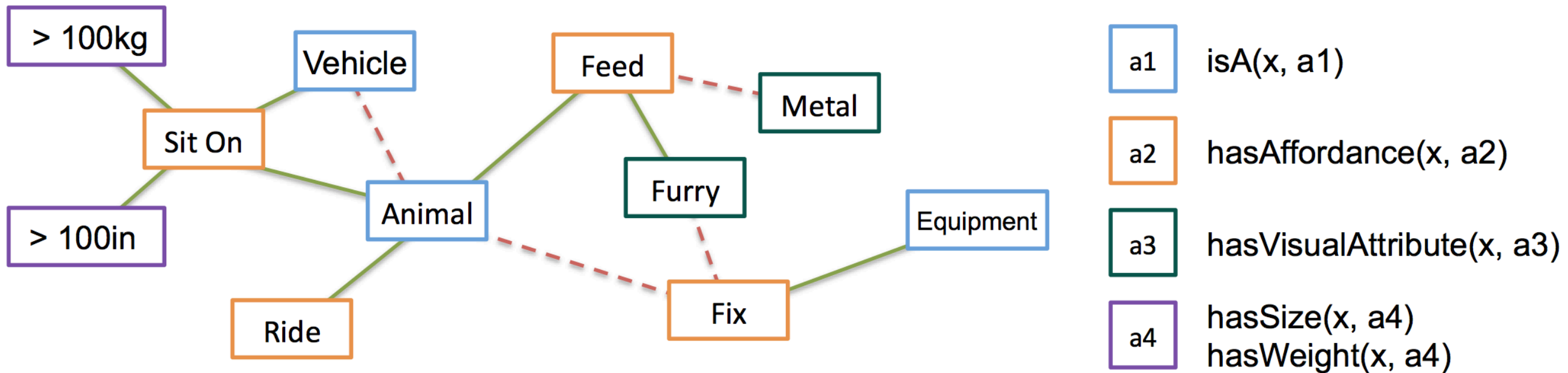
Examples

isA(x, Vehicle) \Rightarrow isA(x, Animal)
hasVisualAttribute(x, Furry) \Rightarrow hasAffordance(x, Feed)
hasWeight(x, W4) \Rightarrow hasAffordance(x, SitOn)
hasAffordance(x, Ride) \wedge locate(x, Below)
isA(x, Animal) \wedge locate(x, Below)
hasAffordance(x, Push) \wedge torso(x, T1)
isA(x, Vehicle) \wedge upperBody(x, U3)

- Markov Logic Network (MLN)
- Unify MRF with first-order logic

$$P(X = x) = \frac{1}{Z} \exp \left(\sum_{i=1}^n w_i f_i(x_{\{i\}}) \right)$$


Likelihood of the formulae being true
L-BFGS Optimization



- 0.8232 hasVisualAttribute(x, Saddle) \Rightarrow hasAffordance(x, SitOn)
- 0.7467 hasVisualAttribute(x, Pedal) \Rightarrow hasAffordance(x, Lift)
- 0.7155 hasVisualAttribute(x, Screen) \Rightarrow hasAffordance(x, Fix)
- 0.7012 hasVisualAttribute(x, Head) \Rightarrow hasAffordance(x, Feed)
- 0.6540 hasVisualAttribute(x, Furry) \Rightarrow hasAffordance(x, Feed)

(a) Top positive attributes (Visual)

- 1.0682 hasVisualAttribute(x, Metal) \Rightarrow hasAffordance(x, Feed)
- 1.0433 hasVisualAttribute(x, Shiny) \Rightarrow hasAffordance(x, Feed)
- 1.0115 hasVisualAttribute(x, Boxy_3D) \Rightarrow hasAffordance(x, Feed)
- 0.8317 hasVisualAttribute(x, Wheel) \Rightarrow hasAffordance(x, Feed)
- 0.7987 hasVisualAttribute(x, Text) \Rightarrow hasAffordance(x, Feed)

(b) Top negative attributes (Visual)

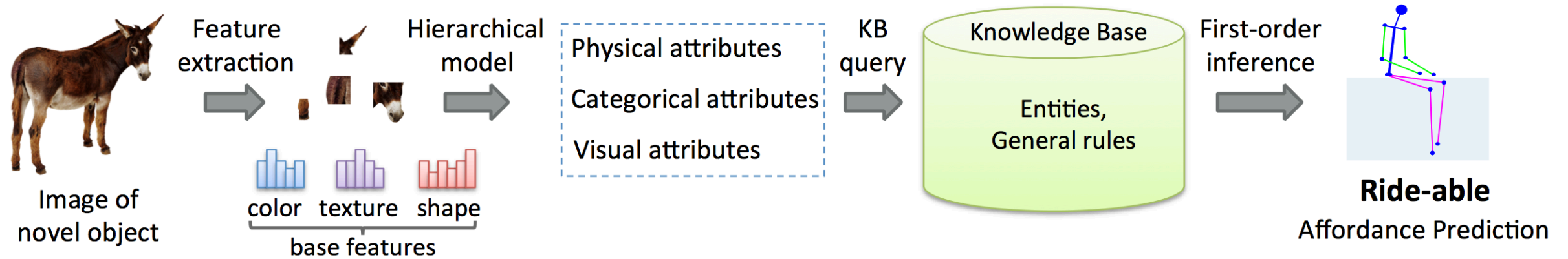
- 5.4734 isA(x, Animal) \Rightarrow hasAffordance(x, Feed)
- 3.3196 isA(x, Vehicle) \Rightarrow hasAffordance(x, Ride)
- 3.2436 isA(x, Vehicle) \Rightarrow hasAffordance(x, Row)
- 2.7976 isA(x, Container) \Rightarrow hasAffordance(x, PourFrom)
- 2.6208 isA(x, Animal) \Rightarrow hasAffordance(x, SitOn)

(c) Top positive rules

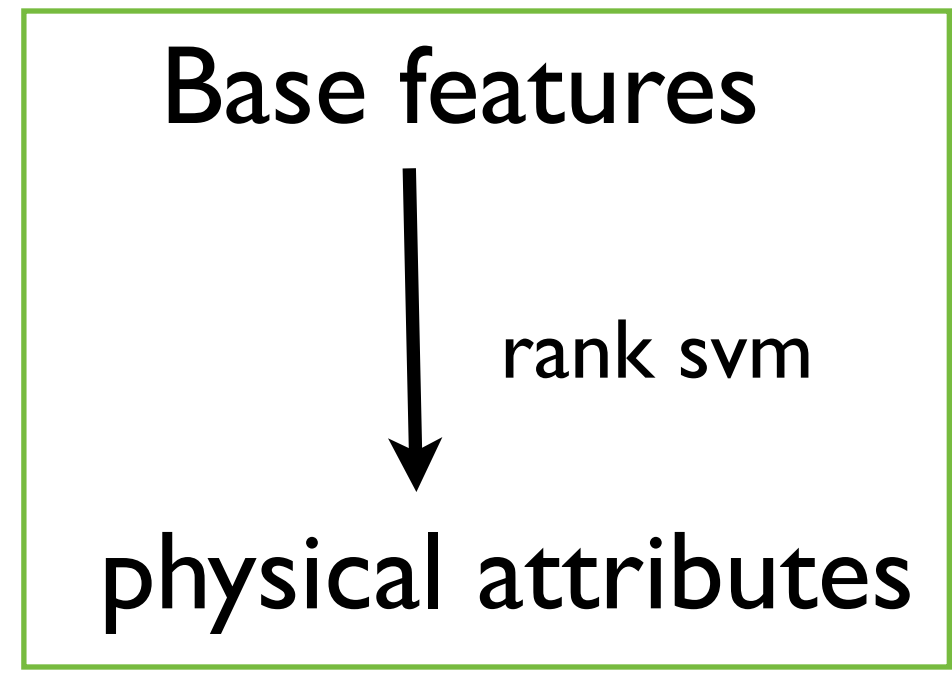
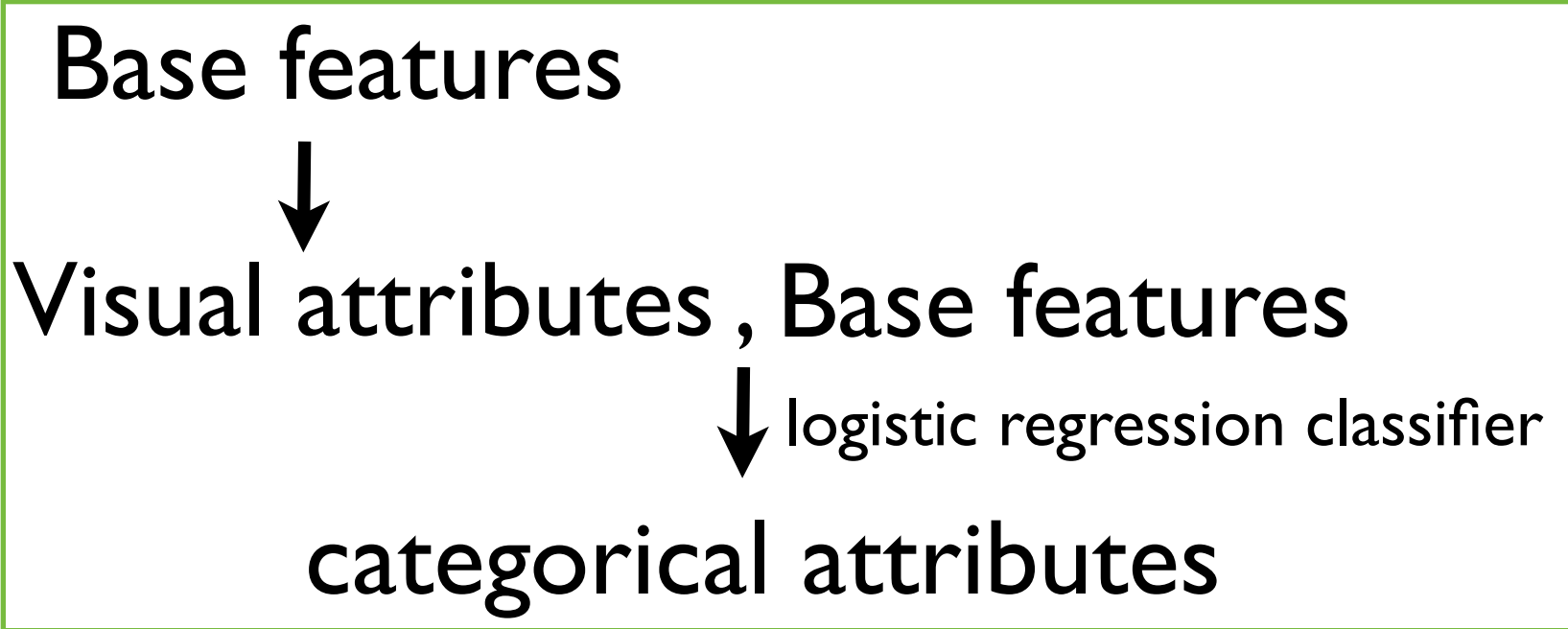
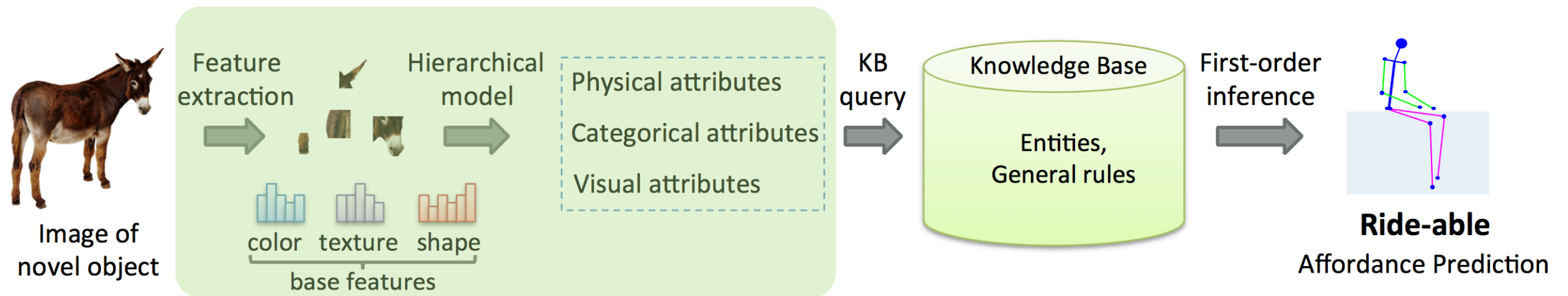
- 3.8636 isA(x, Animal) \Rightarrow hasAffordance(x, Fix)
- 2.2209 isA(x, Seat) \Rightarrow hasAffordance(x, Push)
- 1.8066 isA(x, Vehicle) \Rightarrow hasAffordance(x, Lift)
- 1.7254 isA(x, Instrumentality) \Rightarrow hasAffordance(x, Feed)
- 1.3258 isA(x, Instrumentality) \Rightarrow hasAffordance(x, Fix)

(d) Top negative rules

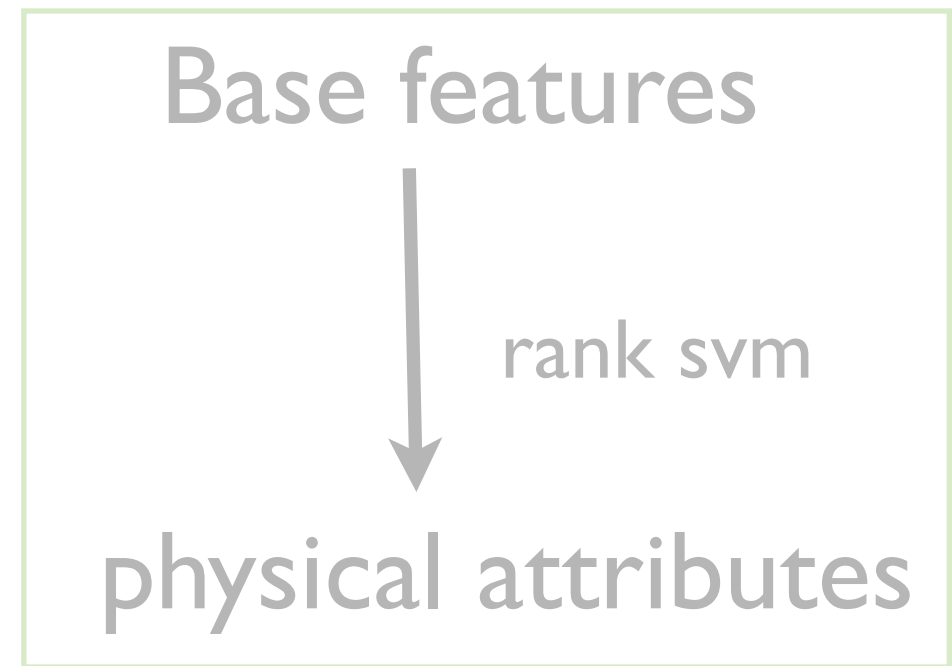
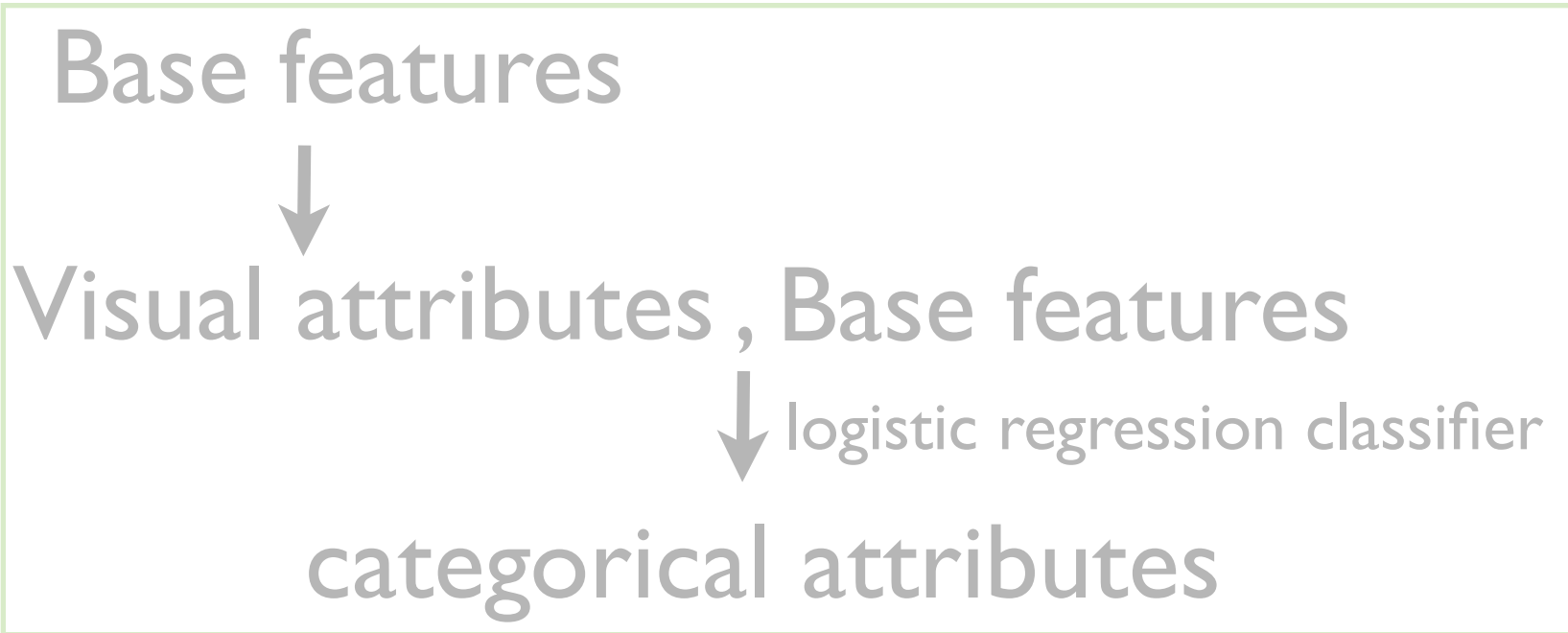
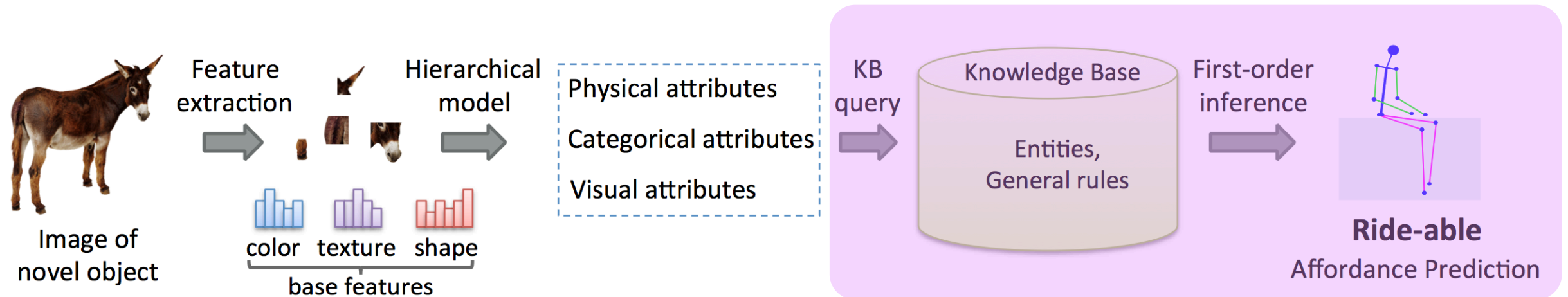
zero-shot affordance prediction



zero-shot affordance prediction



zero-shot affordance prediction



➔ First-order inference to predict affordances

zero-shot affordance prediction

Method	L1-LR [9]	χ^2 -SVM [25]	Ours
base features (BF)	0.7858	-	-
visual attributes (VA)	0.7525	0.7533	0.7432
categorical & physical (CP)	0.7919	0.7924	0.8234
combined (VA+CP)	0.8006	0.7985	0.8409

- ✓ KB models complex general rules
- ✓ Classifiers fail to take correlations into account

Estimating human pose

\mathcal{T} Pose: (torso, lower body, upper body)

↘ cluster centroids

$$\textit{hamming}(\mathcal{T}) = \min_{\mathcal{T}' \in P_o \cup \hat{P}_o} \sum_{i=1..3} \mathbb{1}(\mathcal{T}_i = \mathcal{T}'_i)$$

↓
Set of ground-truth poses of the canonical affordance of the object

Method	nearest neighbor	attributes	affordances	attributes+affordances
Distance	0.928	1.027	0.630	0.527

Question Answering

Question	Evidence	Query	Top Answers
What do animals look like?	isA(N1, Animal)	hasVisualAttribute(N1, x)	hasVisualAttribute(N1, Leather) hasVisualAttribute(N1, Head) hasVisualAttribute(N1, Tail) hasVisualAttribute(N1, Furry)
I saw something shiny and metallic. What is it?	hasVisualAttribute(N1, Shiny) hasVisualAttribute(N1, Metal)	isA(N1, x)	isA(N1, Instrumentality) isA(N1, Device) isA(N1, Container) isA(N1, Computer)
Here is a vehicle and it's quite heavy. What can I do with it?	isA(N1, Vehicle) hasWeight(N1, W4) (> 100 kg)	hasAffordance(N1, x)	hasAffordance(N1, Ride) hasAffordance(N1, Row) hasAffordance(N1, SitOn) hasAffordance(N1, Fix)
Tell me how heavy and large a wooden musical instrument is.	isA(N1, Musical_instrument) hasVisualAttribute(N1, Wood)	hasWeight(N1, x) hasSize(N1, x)	hasSize(N1, D2) (10-100 in) hasWeight(N1, W2) (1-10 kg)

MLN infers the probability or the most likely state of each query from the evidence

why KB

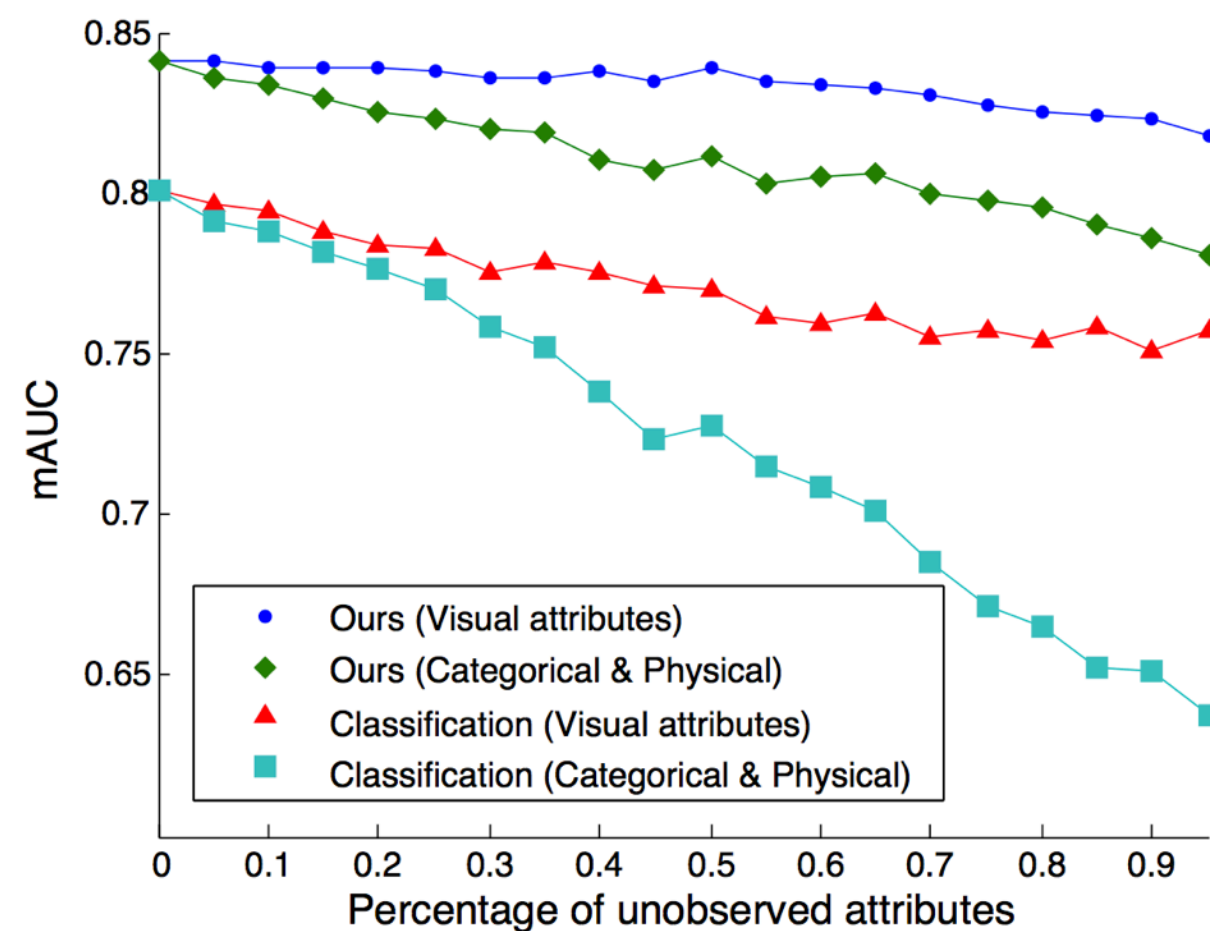


Fig. 12: **Performance variations against partial observation.** The x -axis denotes the percentage of unobserved evidence. The y -axis denotes the performance (mAUC). The top two curves correspond to our method. The bottom two are the classification-based method. In comparison, the knowledge base representation is more robust against partial observation.