



One-shot, Zero-shot and Open-set Recognition of Social Media Analysis

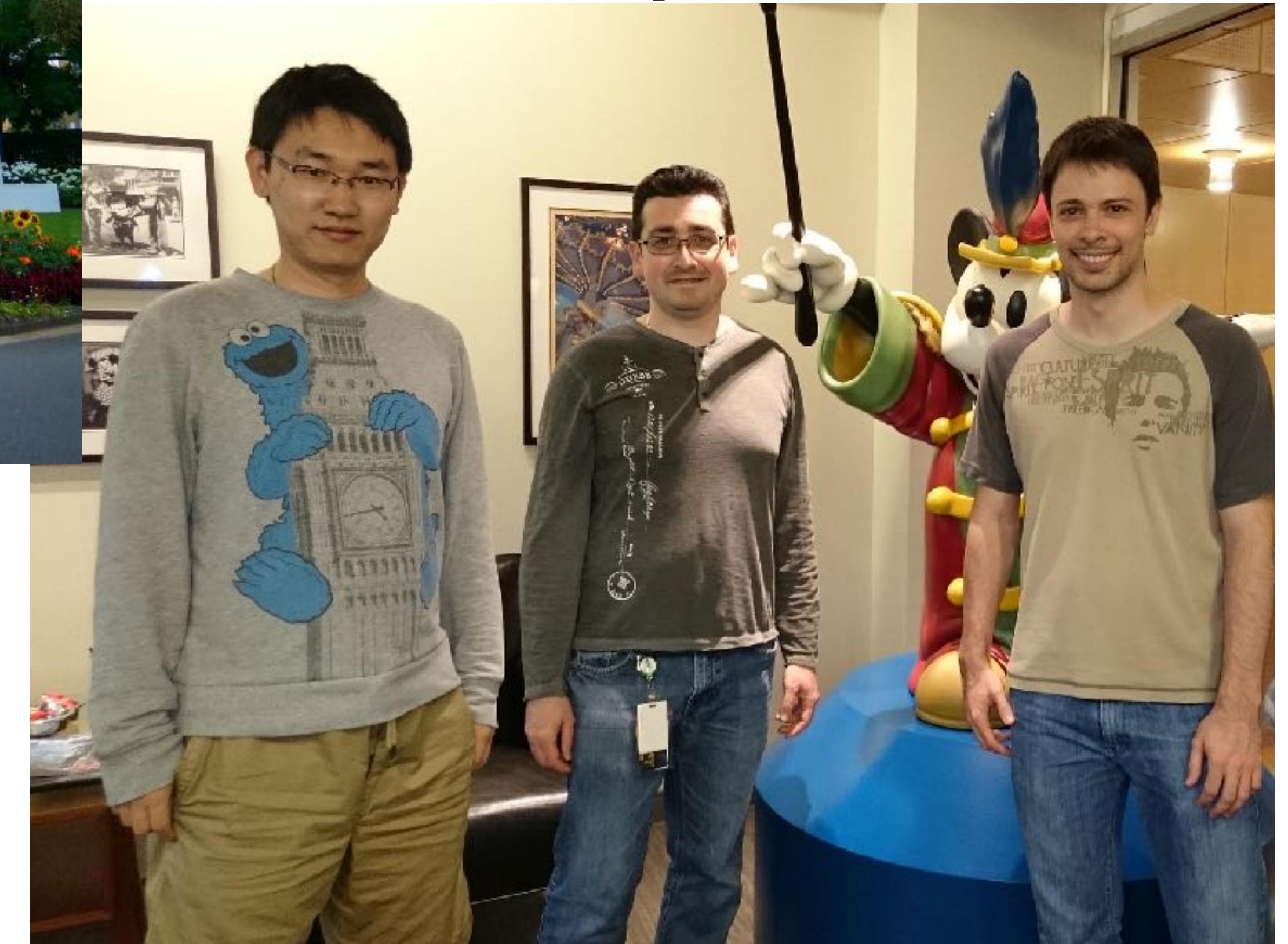
付彦伟
复旦大学大数据学院



About me



Disney Research



Content

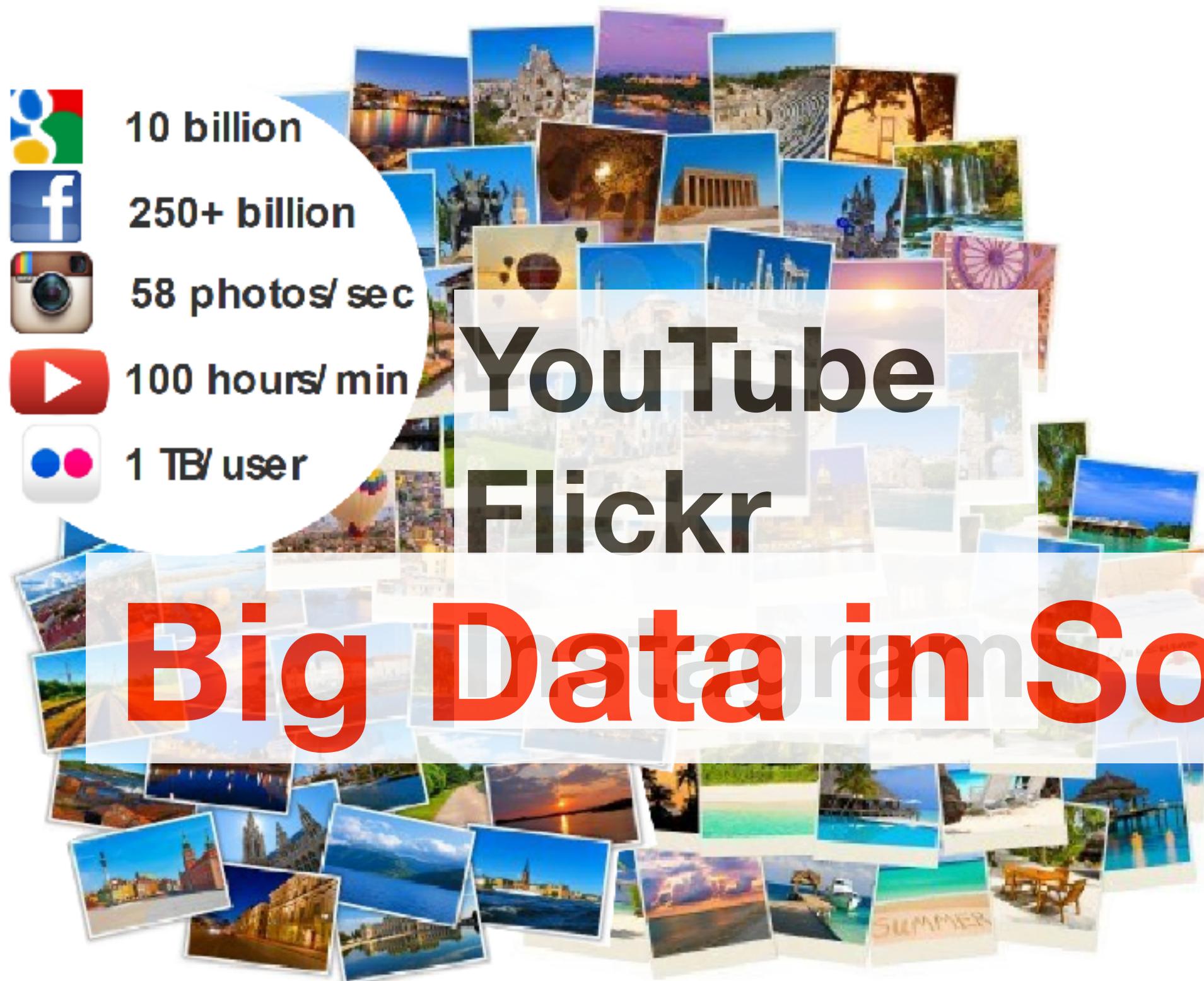
1. Overview
2. Definition
3. Embedding
4. More



Content

1, Overview,
of our works on
Social Media Analysis





YouTube

vimeo

Dailymotion

twitch ...



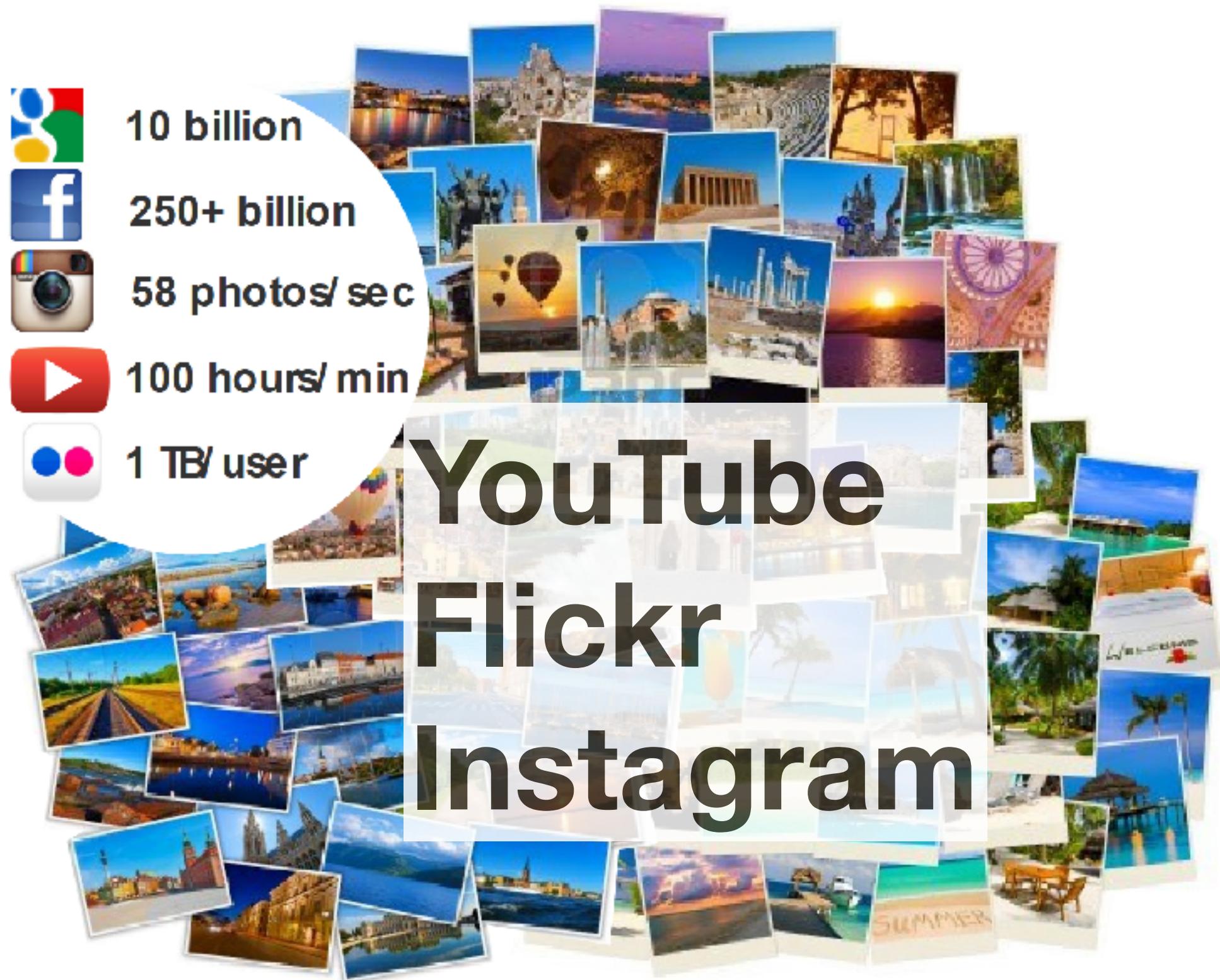
10 billion

250+ billion

58 photos/sec

100 hours/min

1 TB/user



YouTube
Flickr
Instagram



Hockey

Cello Performance



Football



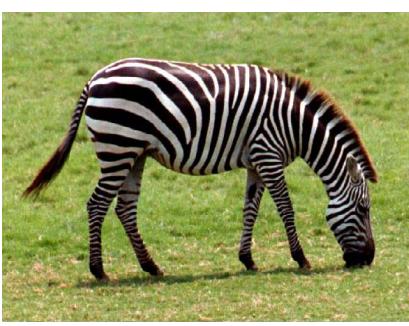
Skiing



Running



Parade



Sadness

Fear

Anger

Joy

Disgust

Image Classes

Action

Activity

Emotion



1. **Yanwei Fu**, Timothy M. Hospedales, Tao Xiang and Shaogang Gong. “*Learning Multi-modal Latent Attributes*” **IEEE TPAMI 2014**;
2. **Yanwei Fu**, Timothy M. Hospedales, Tao Xiang and Shaogang Gong. “*Transductive Multi-view Zero-Shot Learning*” **IEEE TPAMI, 2015**;
3. **Yanwei Fu**, Timothy M. Hospedales, Tao Xiang and Shaogang Gong. “*Attribute Learning for Understanding Unstructured Social Activity*”, **ECCV 2012**;
4. **Yanwei Fu**, Timothy M. Hospedales, Zhenyong Fu, Tao Xiang and Shaogang Gong. “*Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*” **ECCV 2014**;
5. **Yanwei Fu**, Leonid Sigal. “*Semi-supervised Vocabulary-informed learning*”, **CVPR 2016**;
6. ZuXuan Wu, **Yanwei Fu**, Yu-gang Jiang, Leonid Sigal, *Harnessing Object and Scene Semantics for Large-Scale Video Understanding*, **CVPR 2016**;
7. *Heterogeneous Knowledge Transfer in Video Emotion Recognition, Attribution and Summarization*, Baohan Xu, **Yanwei Fu**, Yu-Gang Jiang, Boyang Li and Leonid Sigal. **IEEE Transactions on Affective Computing 2017**.

Crowdsourcing Ranking on Internet



Our Story Start from a movie – The Social Network



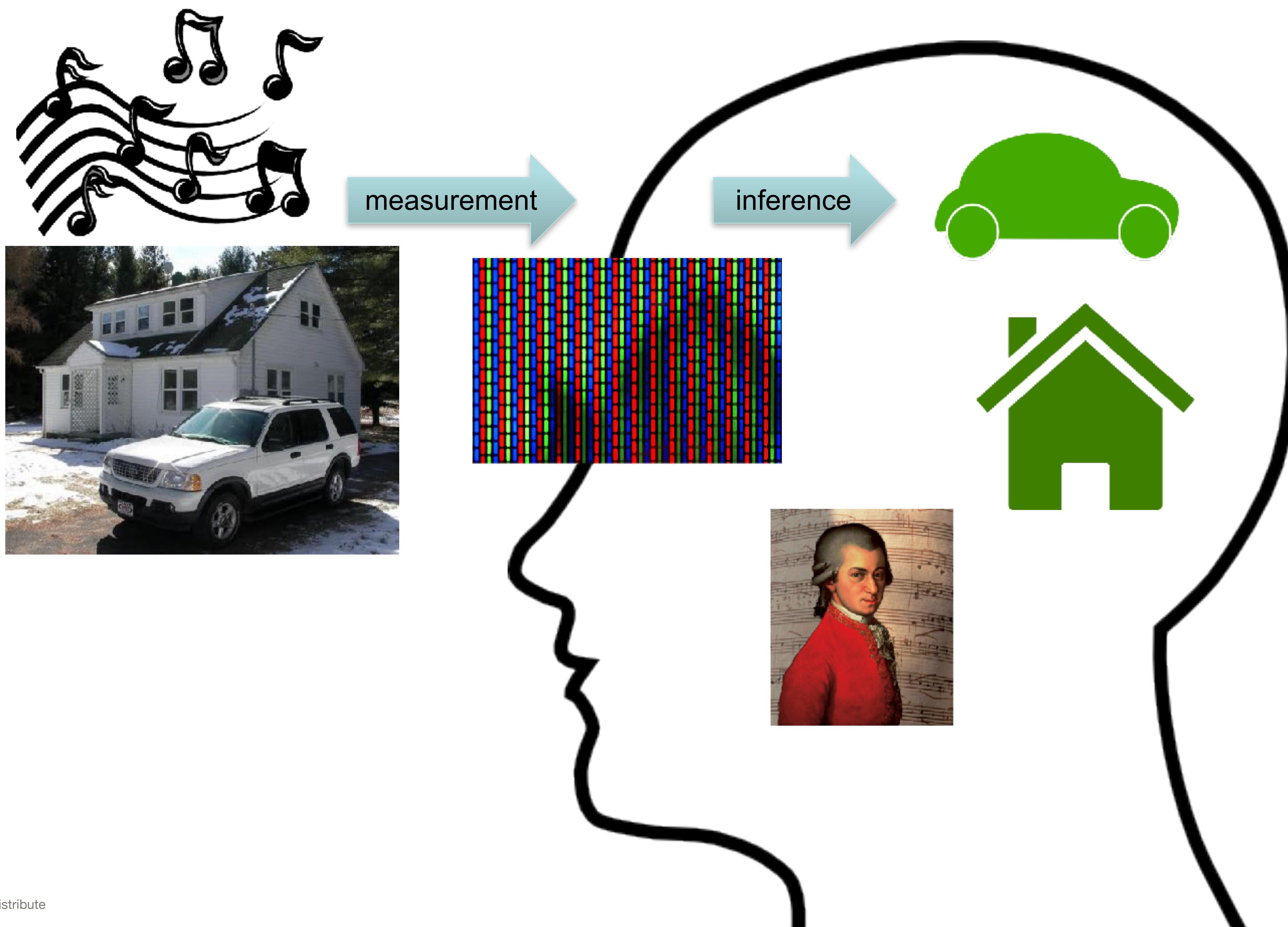
1. **Yanwei Fu**, Timothy M. Hospedales, Tao Xiang, Jiechao Xiong, Shaogang Gong, Yizhou Wang and Yuan Yao. “*Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels*” **IEEE TPAMI 2016**;
2. Yu-ting Qiang, **Yanwei Fu**, Yanwen Guo, Zhi-hua Zhou and Leonid Sigal. “*Learning to Generate Posters of Scientific Papers*”, **AAAI 2016**;
3. **Yanwei Fu**, Timothy M. Hospedales, Tao Xiang and Shaogang Gong and Yuan Yao. “*Interestingness Prediction by Robust Learning to Rank*” **ECCV 2014**;
4. **Yanwei Fu**, Yanwen Guo, Yanshu Zhu, Feng Liu, Chuanming Song and Zhi-Hua Zhou. *Multi-view Video Summarization*, **IEEE TMM 2010**;
5. **Yanwei Fu**, De-an Huang, Leonid Sigal, Robust Classification by Pre-conditioned LASSO and Transductive Diffusion Component Analysis, <http://arxiv.org/abs/1511.06340>.

Content

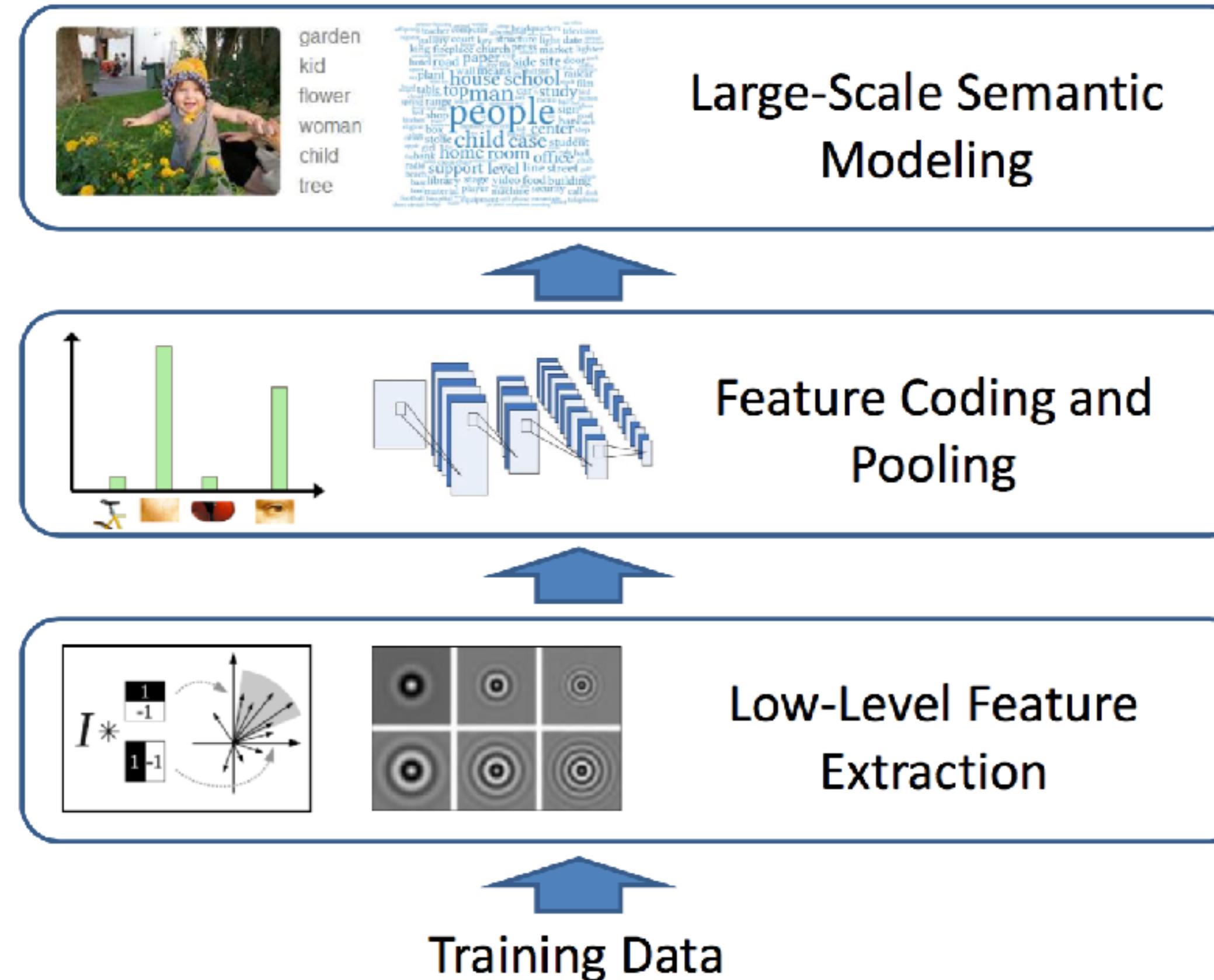
2, Definition,
of one-shot, zero-shot, and
open-set recognition



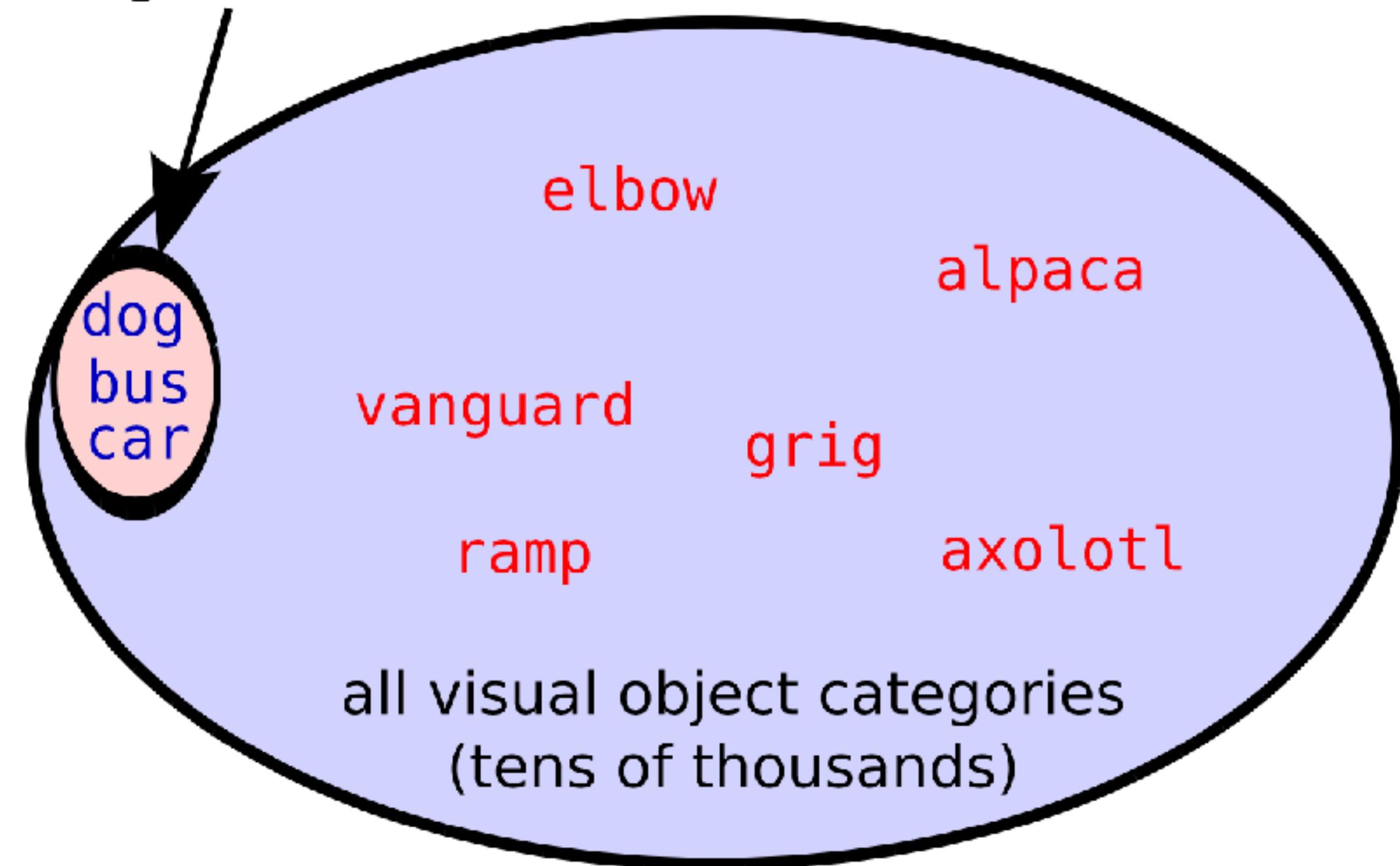
Recognition in a nutshell



Supervised Recognition



visual object categories
for which we have training
images (hundreds)



One-shot Learning

Problem Definition

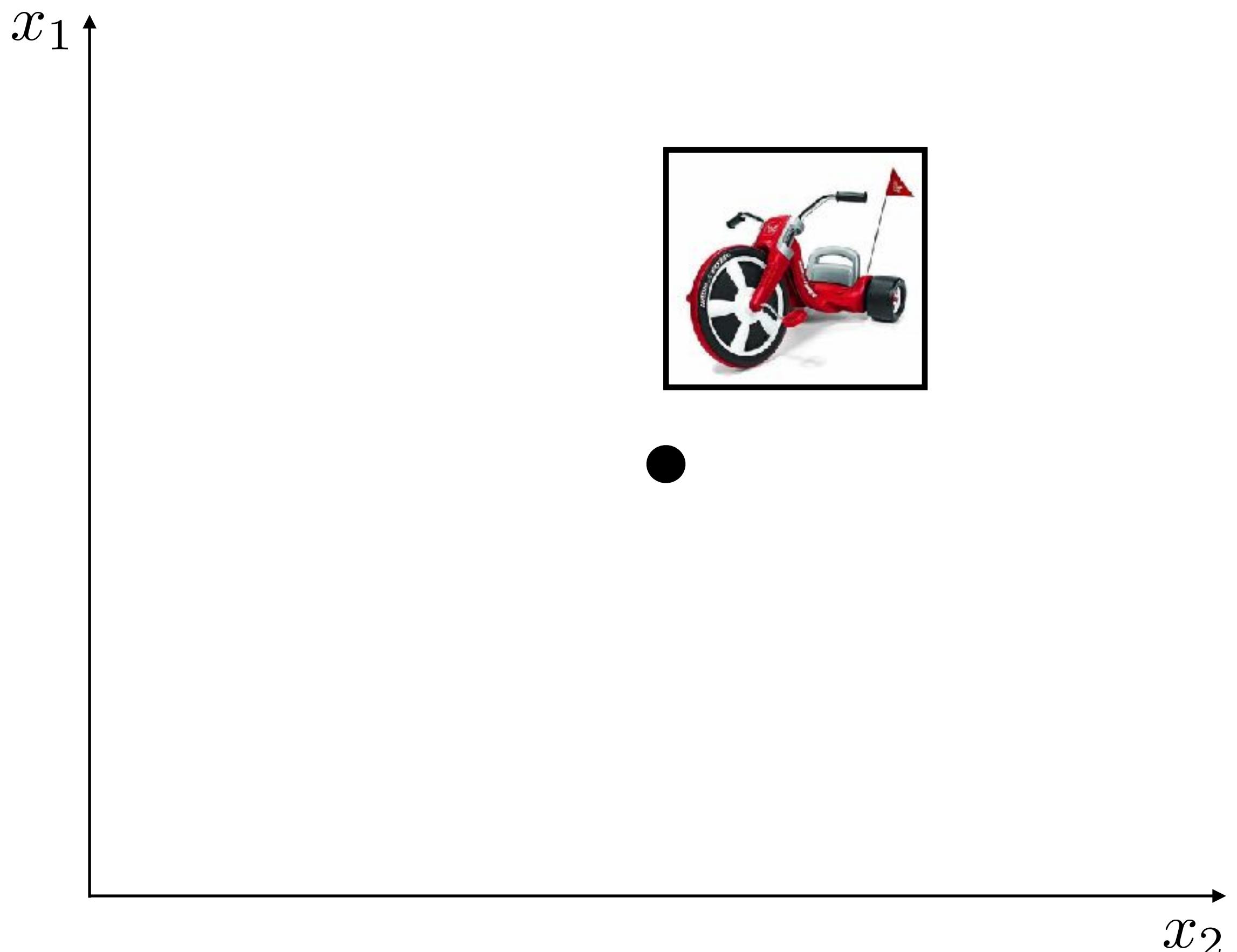


airplane

car

unicycle

tricycle



One-shot Learning

Learning

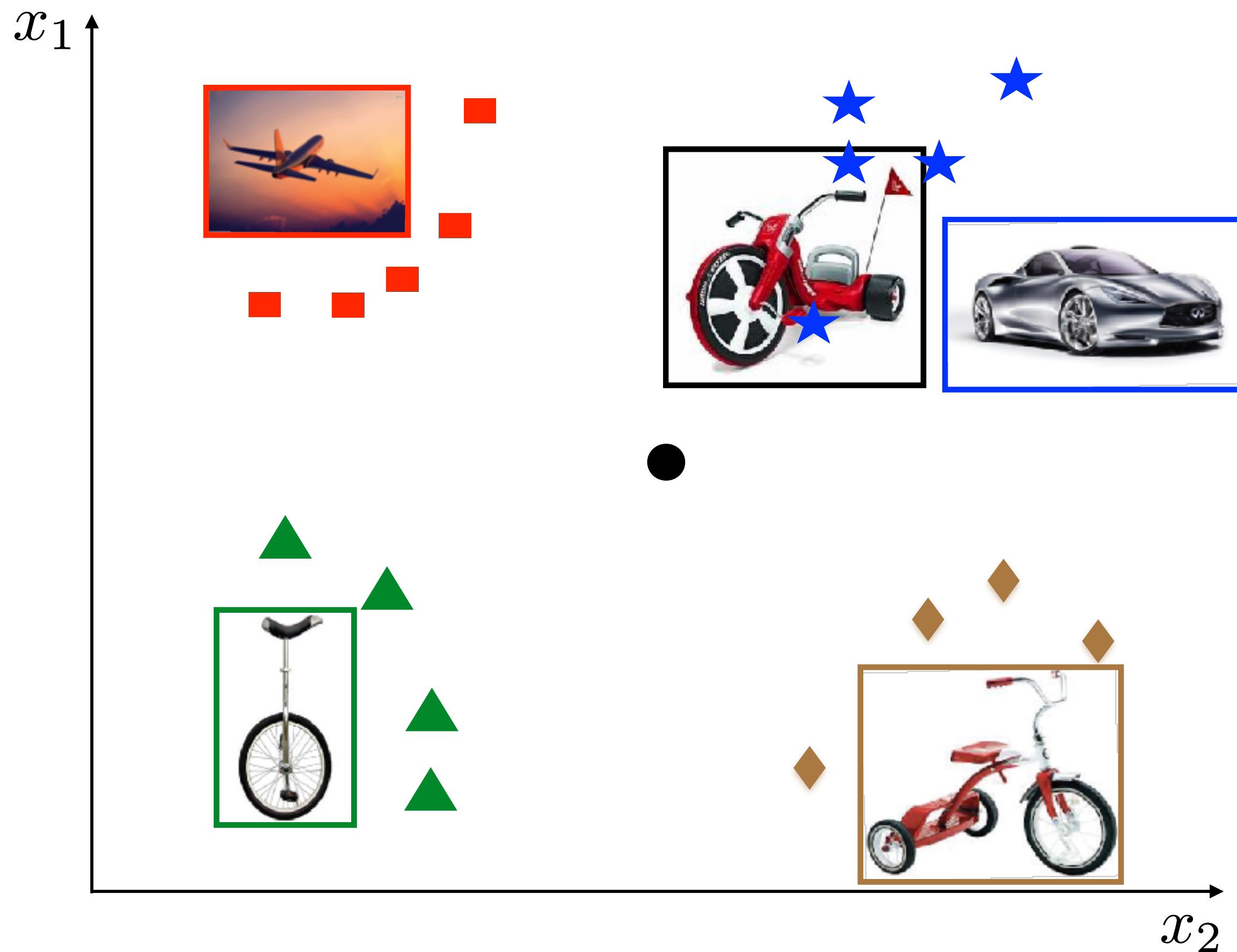


airplane

car

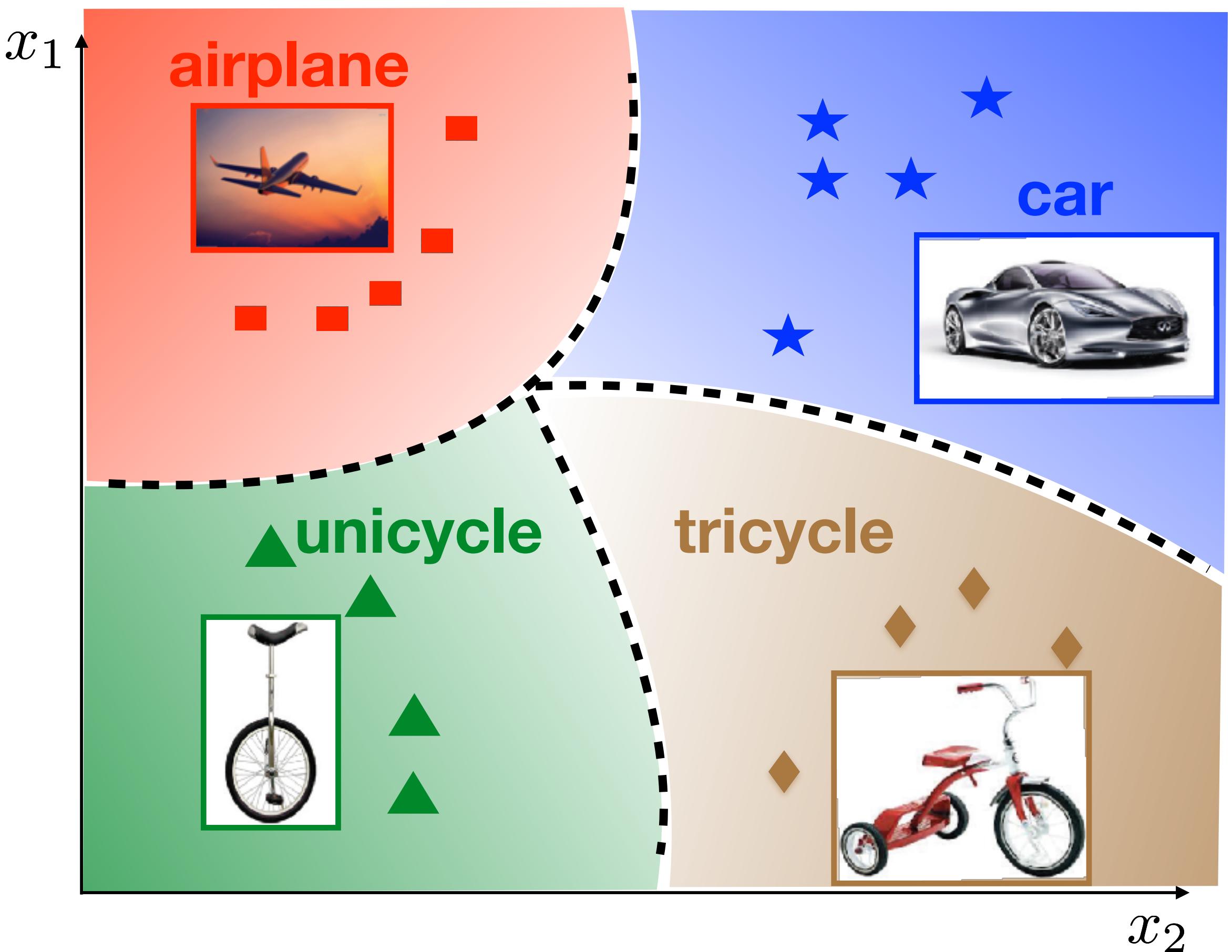
unicycle

tricycle



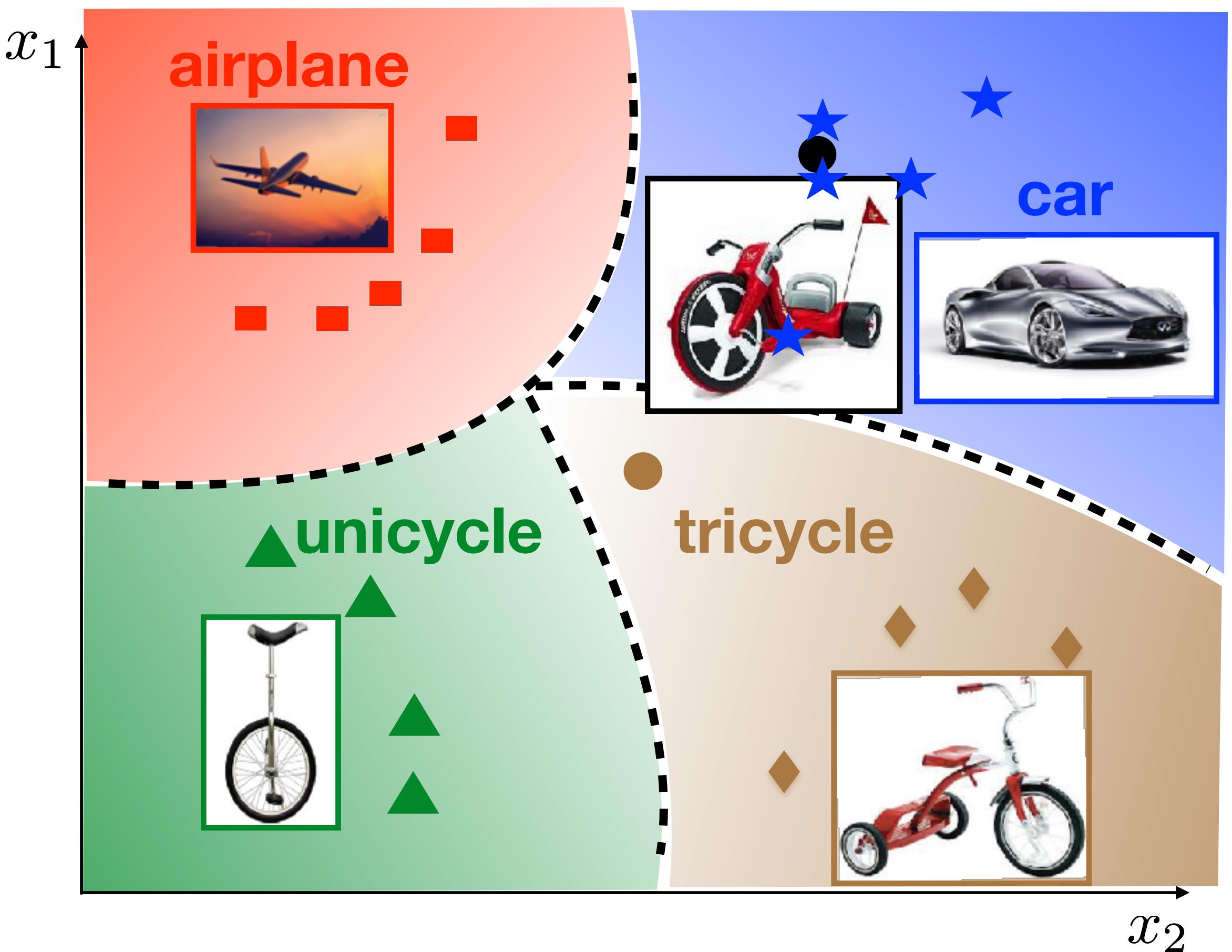
Supervised Learning

Learning



Supervised Learning

Inference



Zero-shot Learning

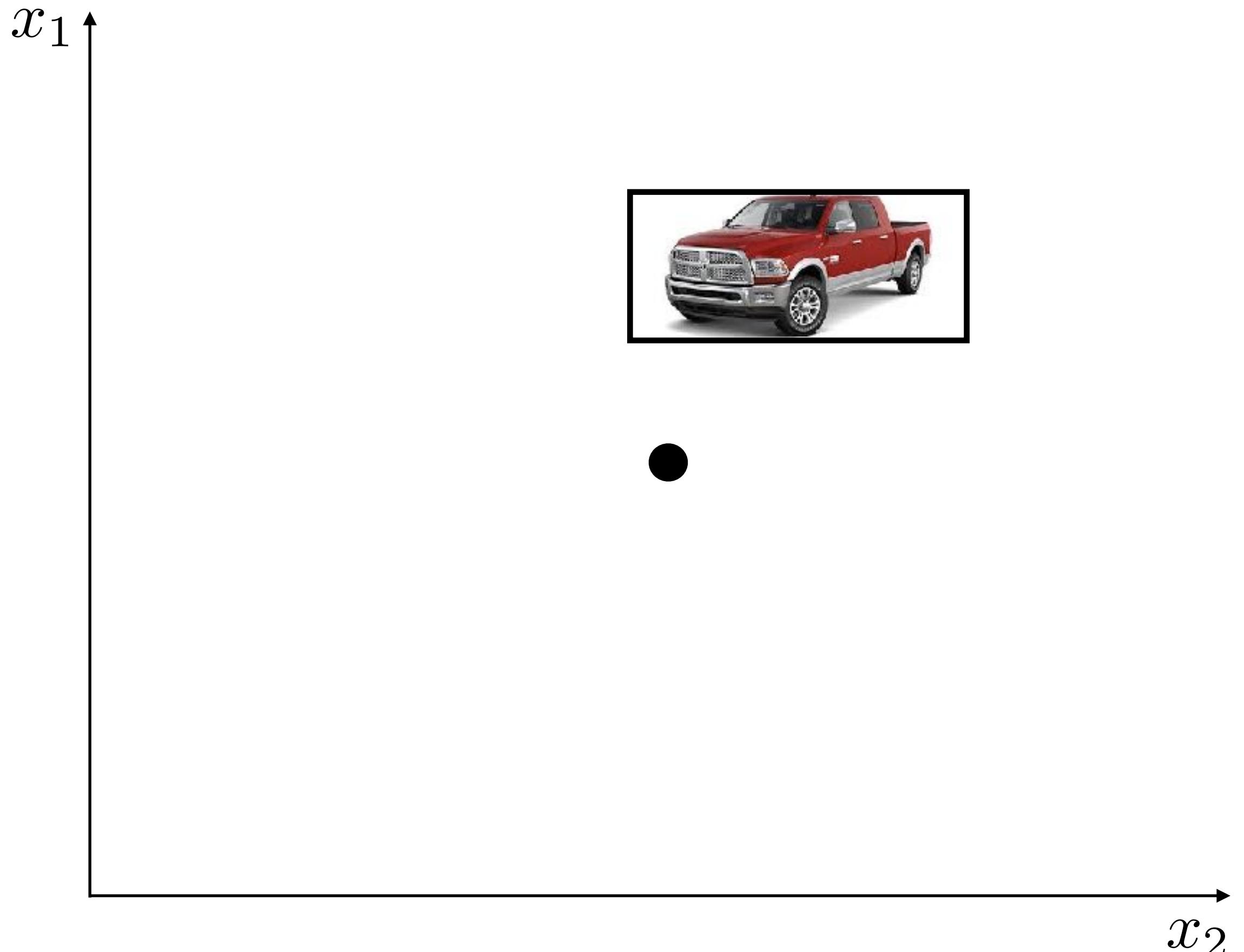
Problem Definition



We do not have any visually labeled instances of what these look like

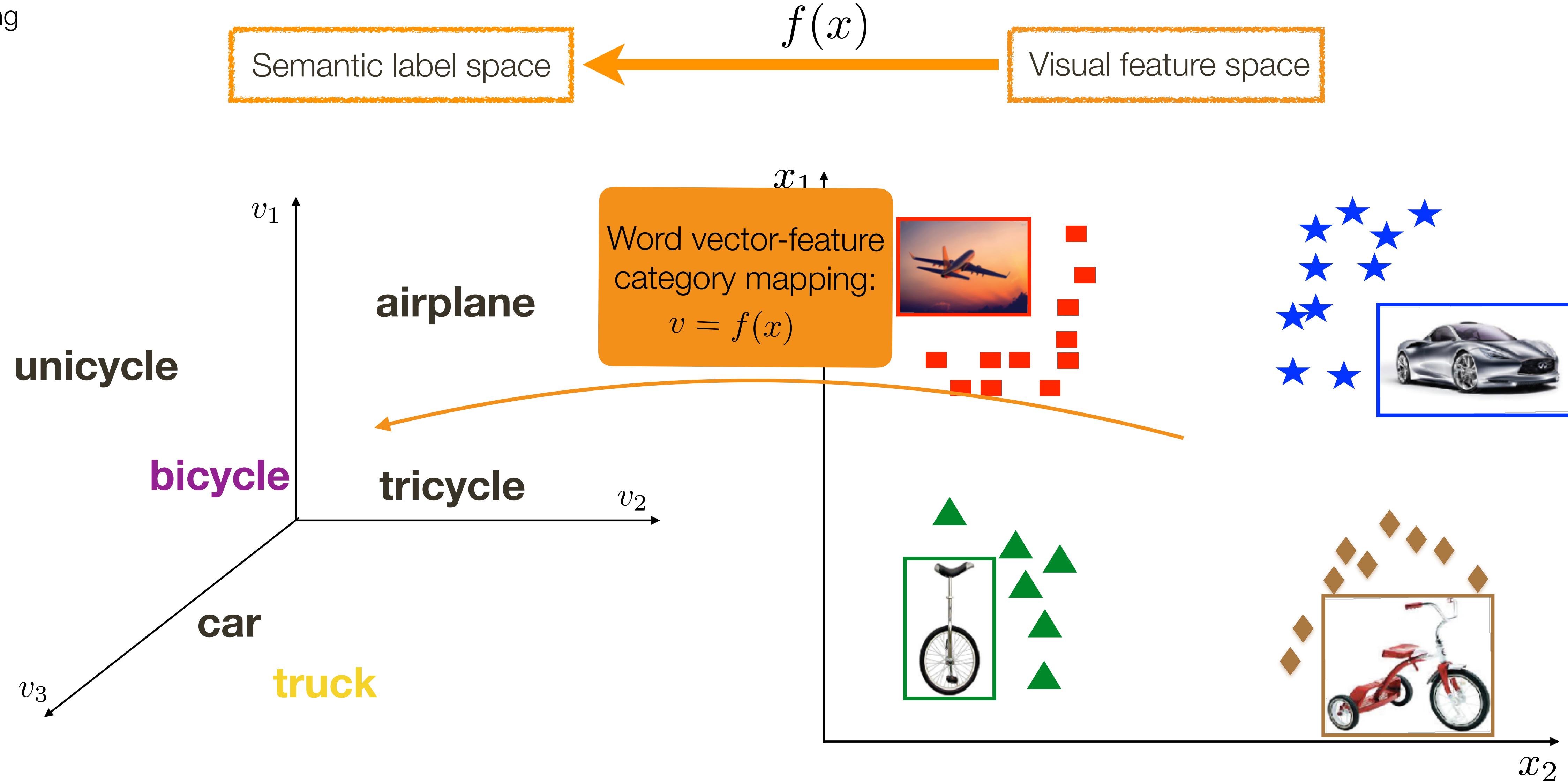
bicycle

truck



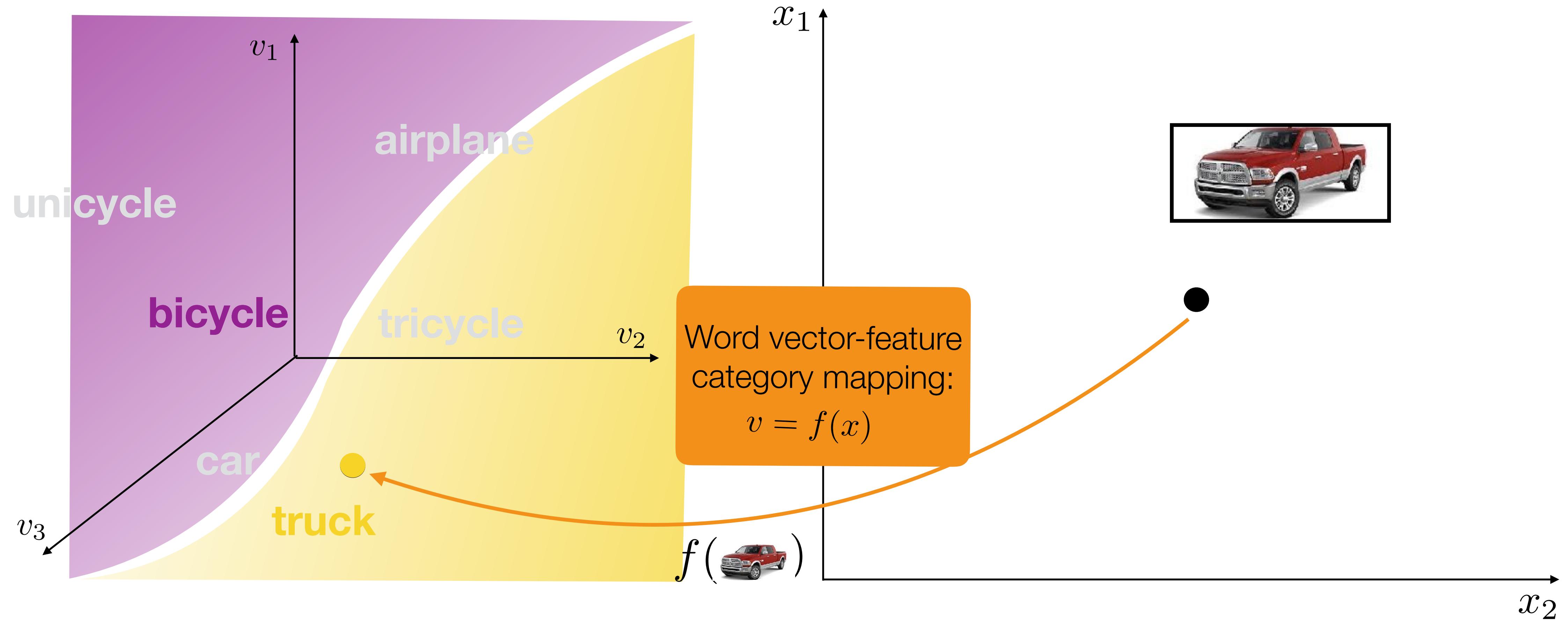
Zero-shot Learning

Learning



Zero-shot Learning

Inference



Key Question: How do we define semantic space?



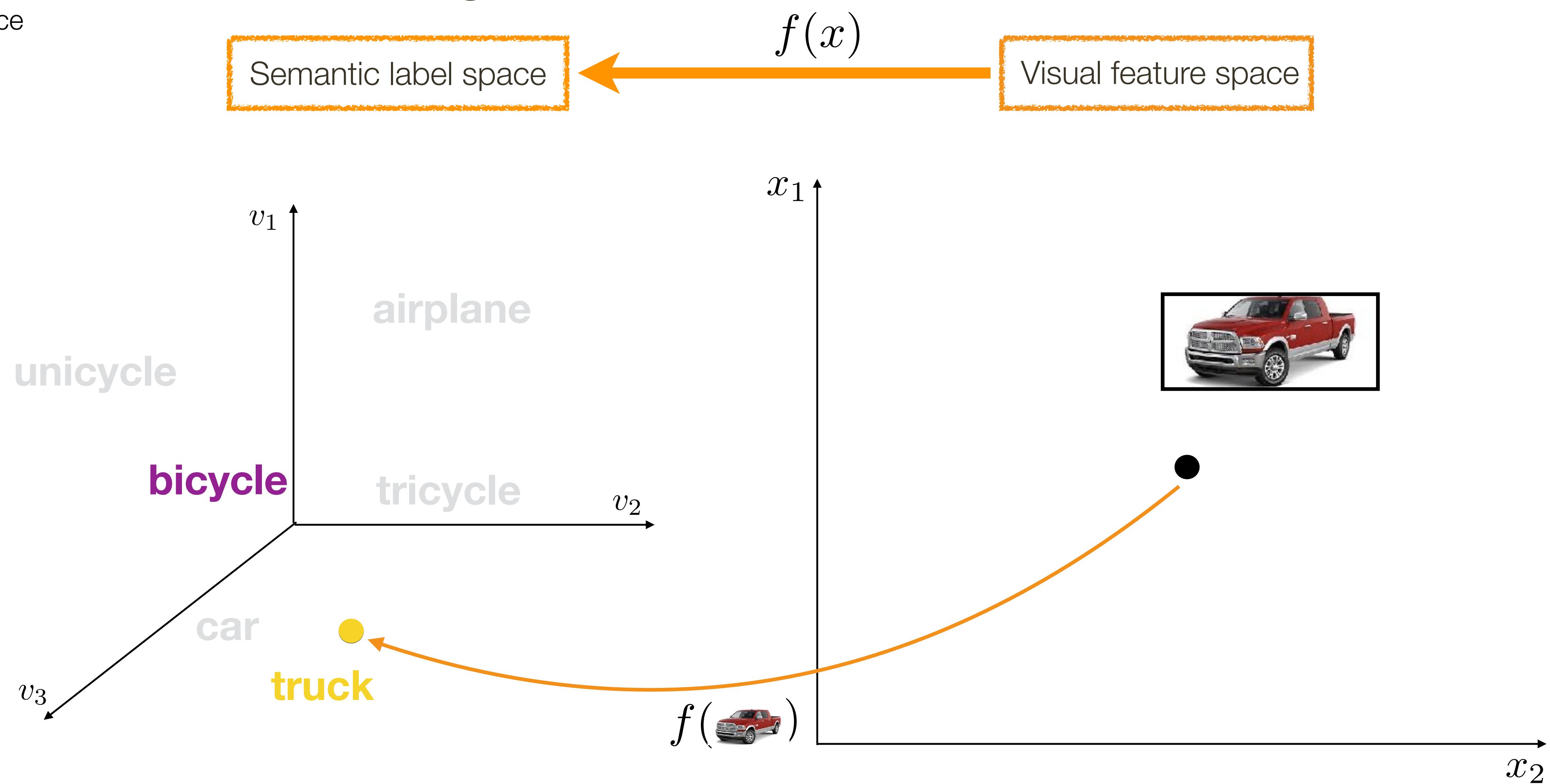
Semantic Label Vector Spaces

Spaces	Type	Advantages	Disadvantages
Semantic Attributes	Supervised	Good interpretability of each dimension: airplane := fixed_wing, propelled, has_pilot	Manual annotation Limited vocabulary
Semantic Word Vectors (e.g. word2vec)	Unsupervised	Good vector representation for millions of vocabulary $v(\text{Berlin}) - v(\text{Germany}) = v(\text{Paris}) - v(\text{France})$	Limited interpretability of each dimension



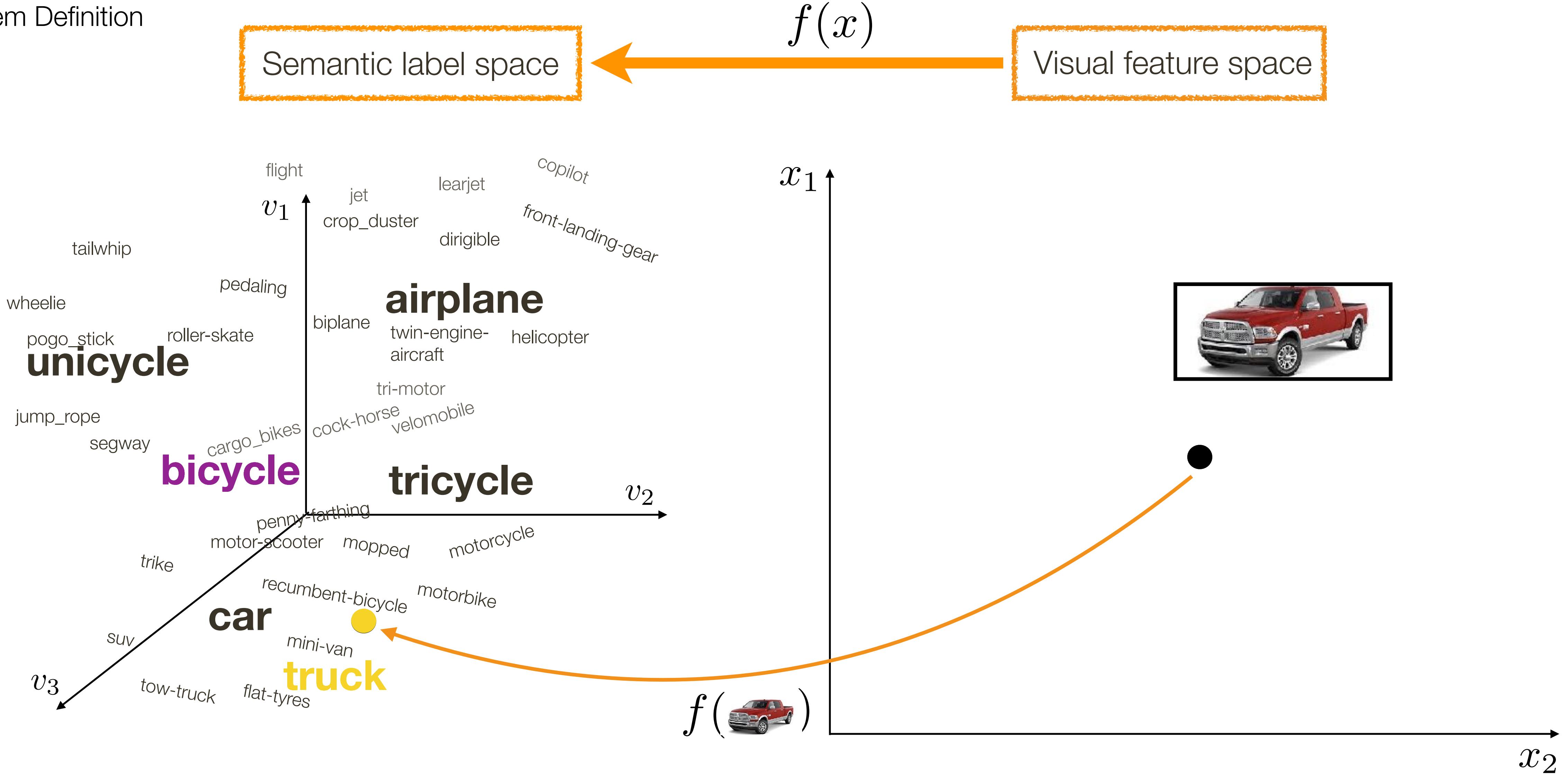
Zero-shot Learning

Inference



Open-set Recognition

Problem Definition



Supervised Learning:

Pros: Very good quantitative performance

Cons: Relatively small vocabulary (~1,000 classes)
Requires **manual labeling** of all the data

One-shot (N-shot, Few-shot) Learning

Pros: Only require few instances per class

Cons: Relatively small vocabulary (~1,000 classes)
Requires **manual labeling** of all the data

Zero-shot Learning:

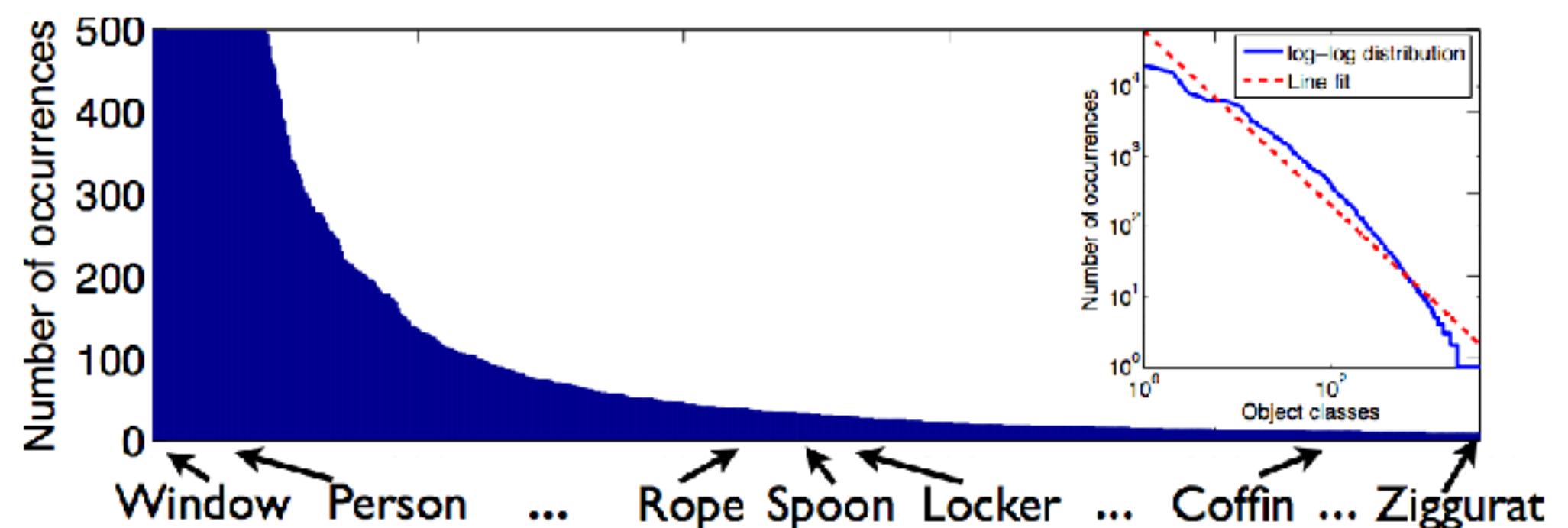
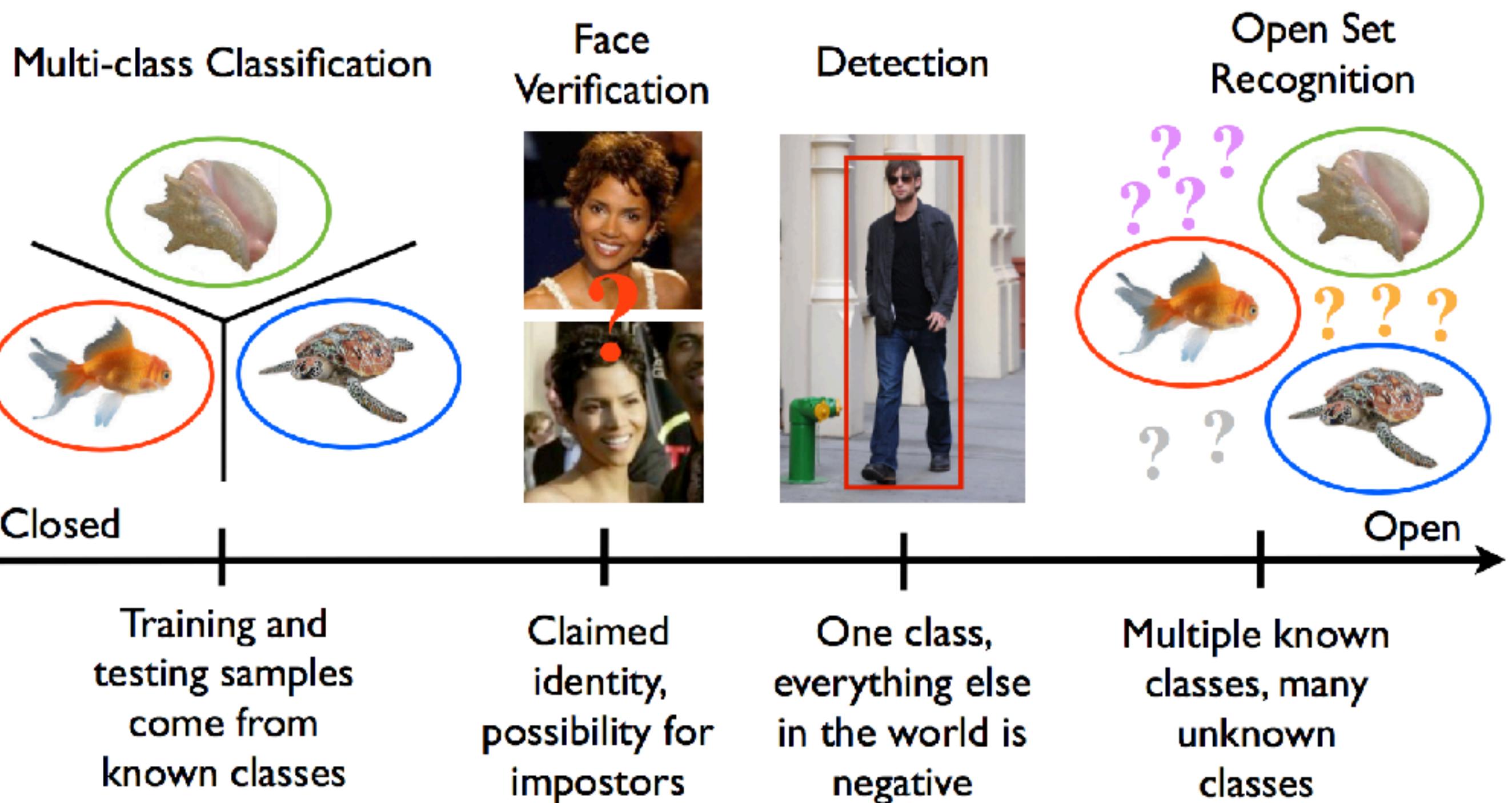
Pros: Does not require instance labeling for target classes

Cons: Typically limited to recognition with target classes only
Relatively **small vocabulary** (~50-10K classes typically)

Open-set Learning

Pros: Does not require instance labeling for target classes

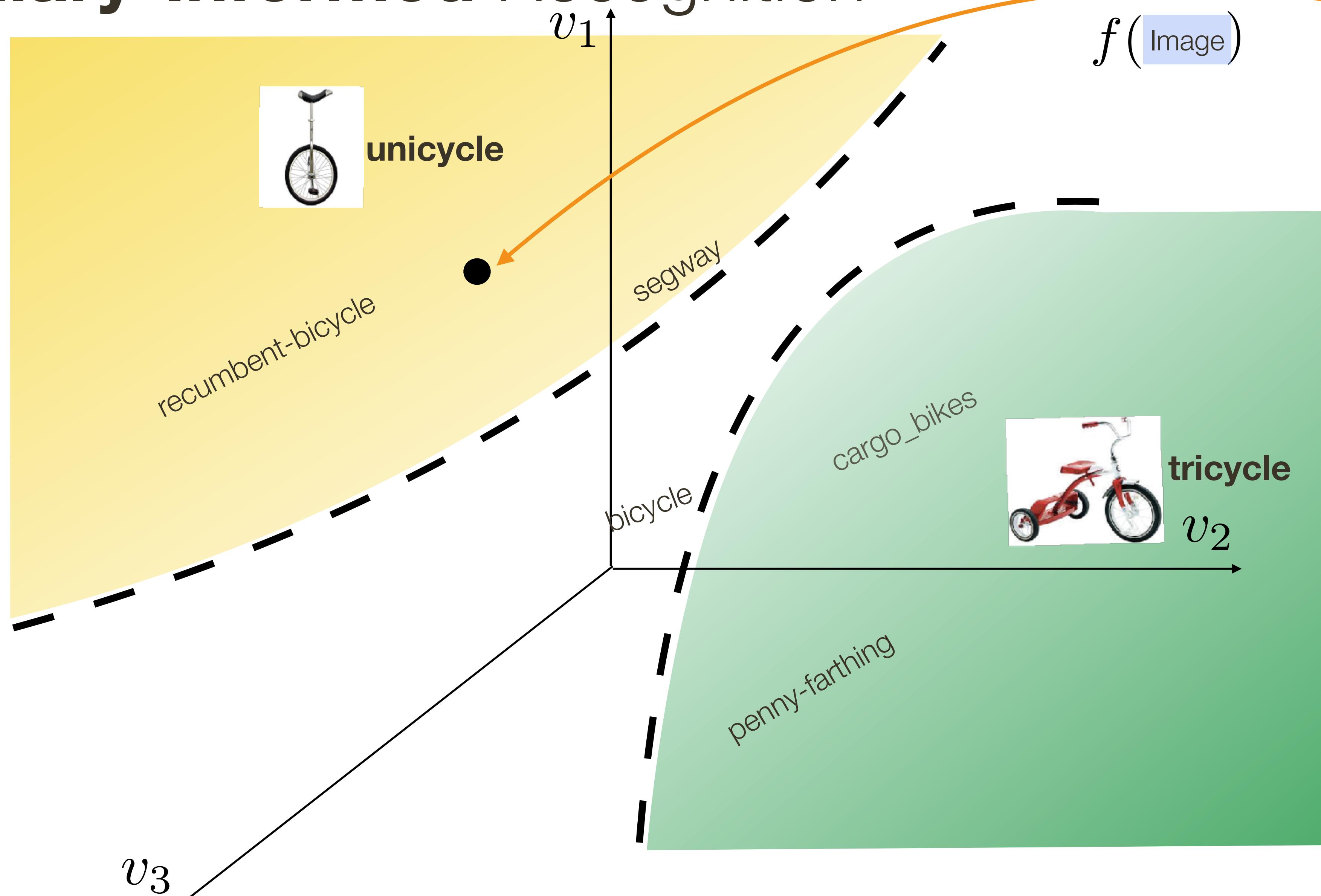
Large vocabulary (e.g. 300K classes)



(a) The number of examples by object class in SUN dataset

Long-tailed Distribution of object categories in the world

Vocabulary-Informed Recognition

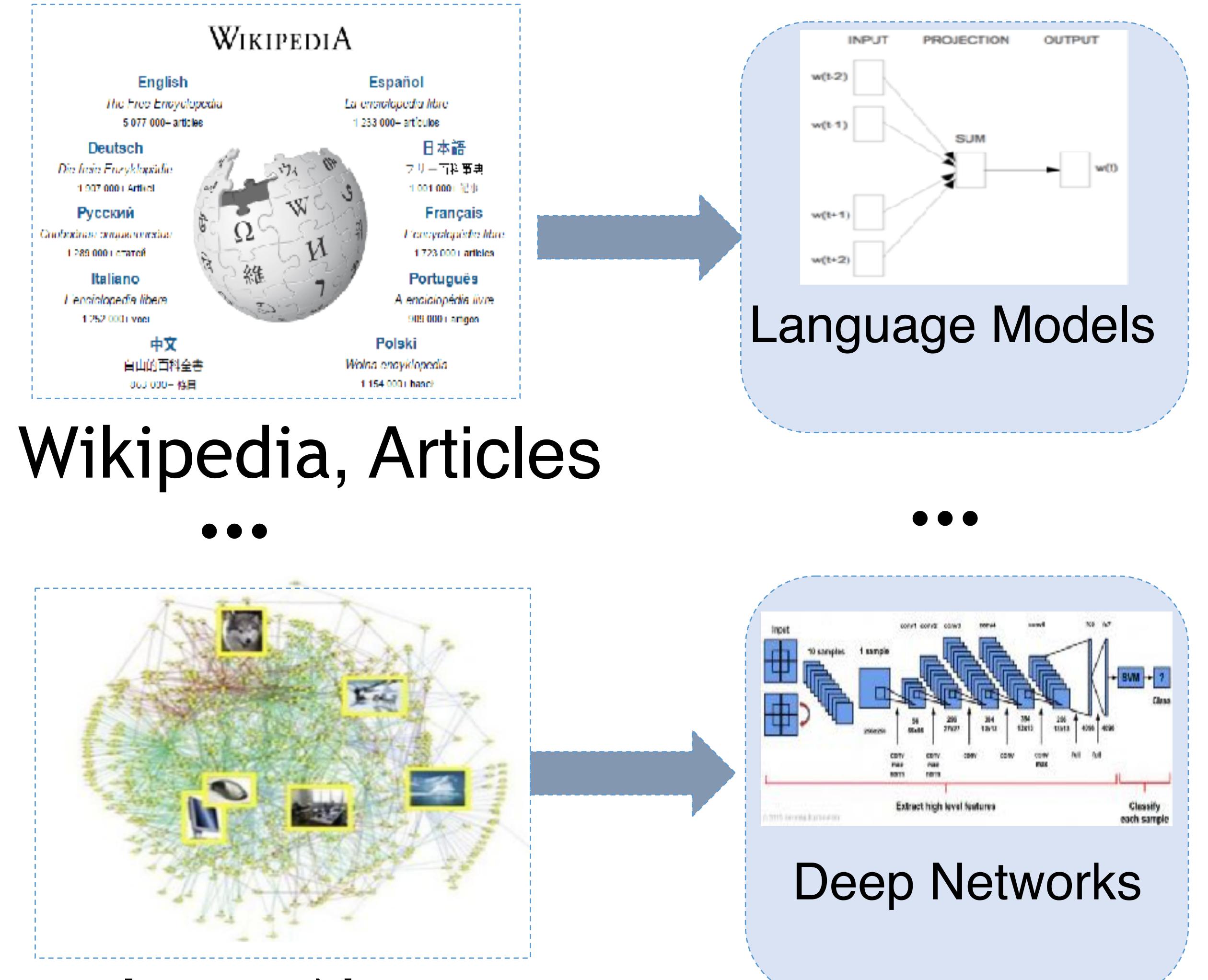


Content

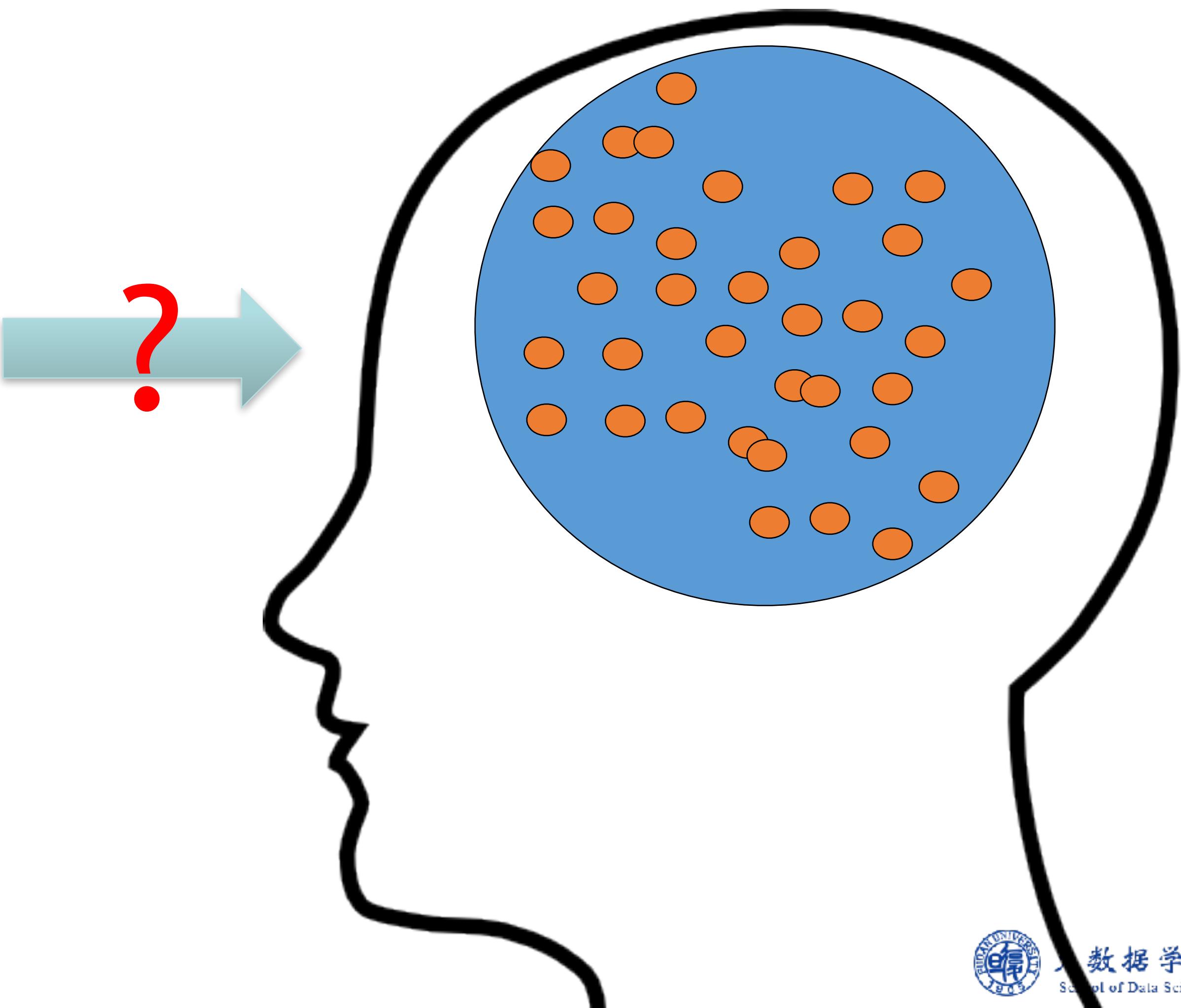
3. Embedding
in Multi-view or Multi-modal



Multi-view/Multi-modal Embedding

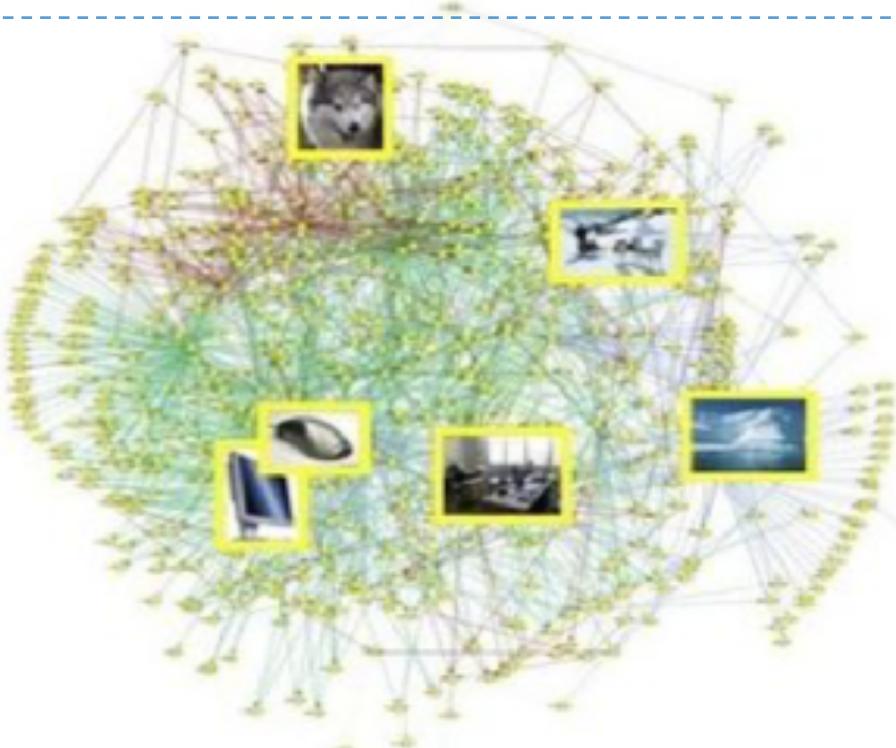


A unified embedding space



Wikipedia, Articles

...



ImageNet

Embedding

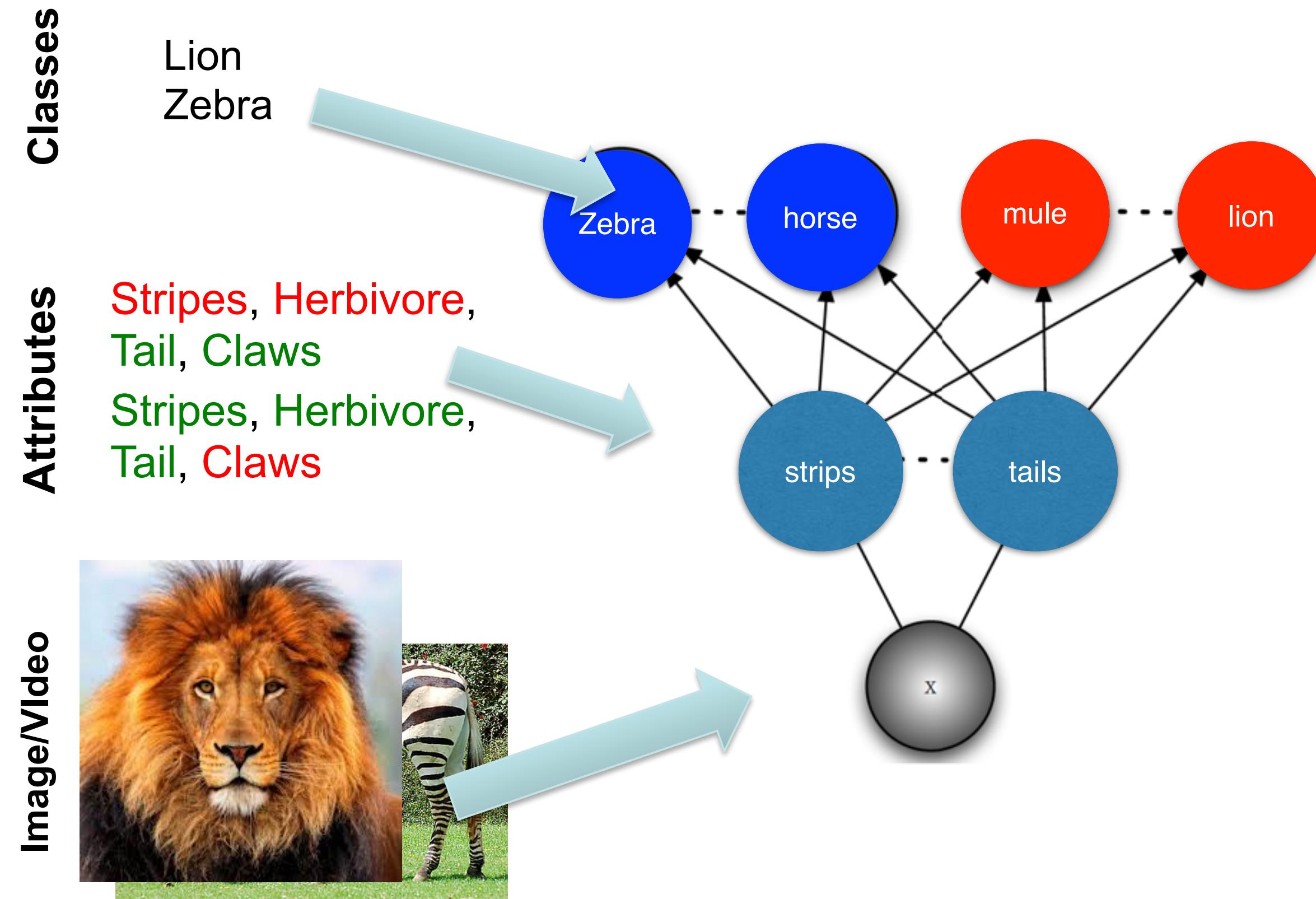
*Learning Multi-modal
Latent Attributes*

Fu et al. *Attribute Learning for Understanding Unstructured Social Activity*”, ECCV 2014
Fu et al. *Learning Multi-modal Latent Attributes*, IEEE TPAMI 2014



Multi-modal Embedding

Existing Attribute Learning Pipeline

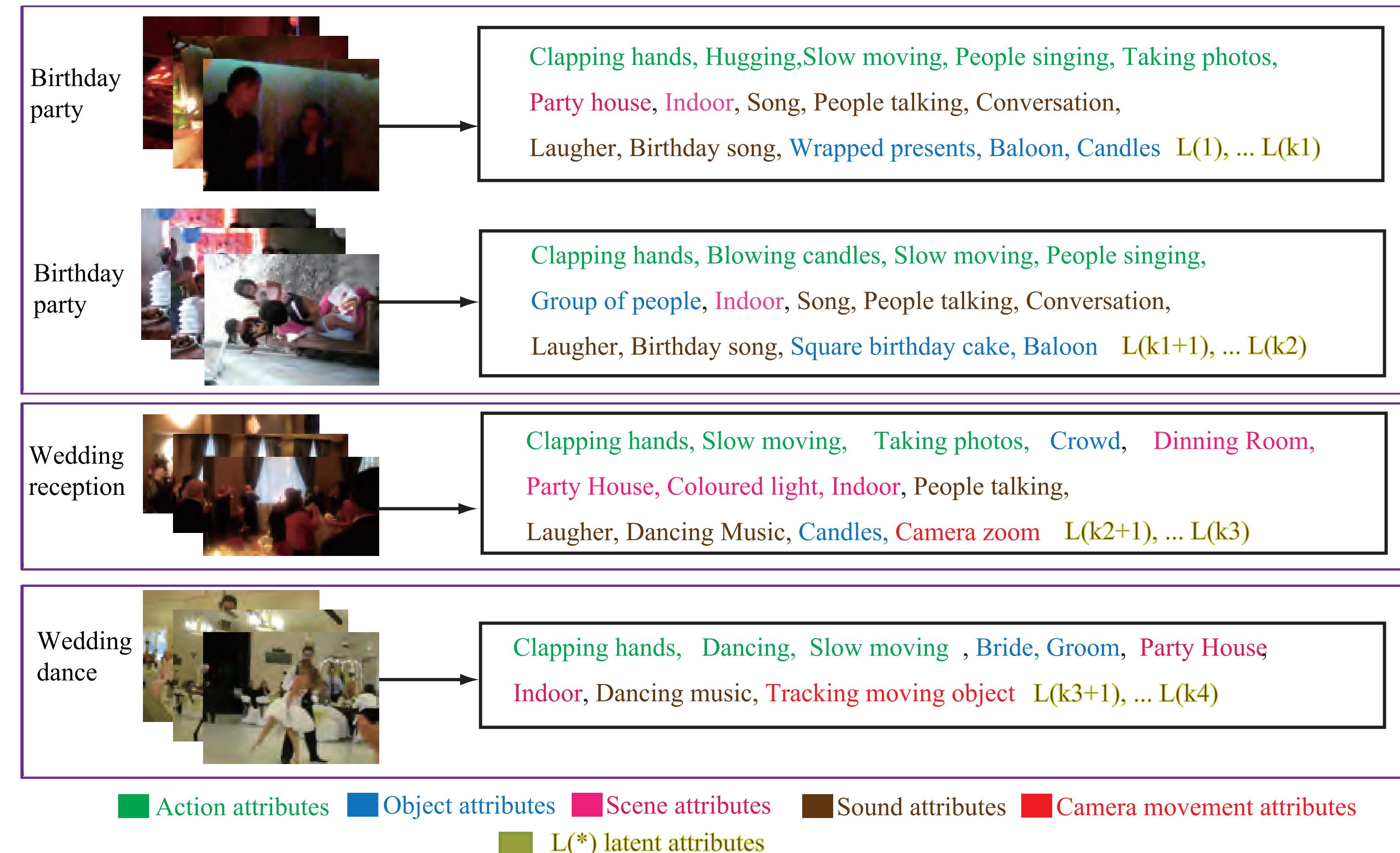


Challenges of attribute annotations

- Incomplete user-defined attributes;
- Sparse user-defined attributes;
- Ambiguous user-defined attributes;

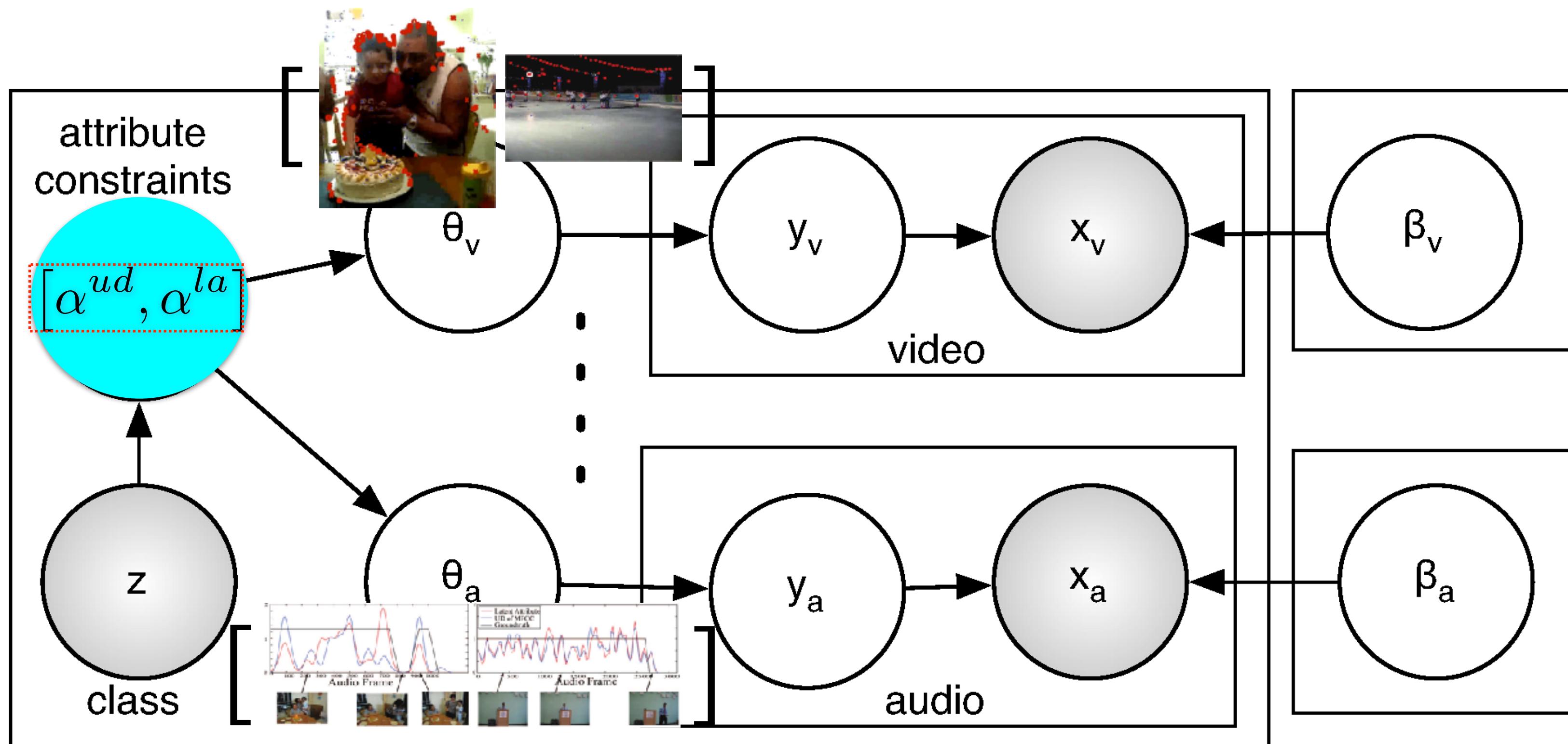
Fu et al. *Attribute Learning for Understanding Unstructured Social Activity*", ECCV 2014
Fu et al. *Learning Multi-modal Latent Attributes*, IEEE TPAMI 2014

Multi-modal Embedding



Fu et al. *Attribute Learning for Understanding Unstructured Social Activity*", ECCV 2014
Fu et al. *Learning Multi-modal Latent Attributes*, IEEE TPAMI 2014

Multi-modal Latent Attribute Topic Model(M2LATM)



$$\theta_m \sim Dir(\alpha) \quad y_m \sim Multi(\theta_m) \quad m \in \{v, a, \dots\}$$

Fu et al. *Attribute Learning for Understanding Unstructured Social Activity*", ECCV 2014
Fu et al. *Learning Multi-modal Latent Attributes*, IEEE TPAMI 2014

Experimental Results

Dataset & Settings:

USAA dataset (8 video classes, e.g. birthday party, graduation party.);

Animal with Attributes (AwA) dataset (40 auxiliary cls; 10 testing cls);

Comparisons

Direct: KNN/SVM of features to classes;

DAP: Direct Attribute Prediction [Lampert et al. CVPR 2009];

SVM-UD: a SVM generalization of DAP;

SCA: Topic models in [Wang et al CVPR 2009];

ST: Synthetic Transfer in [Yu et al ECCV 2010];



Multi-task Learning

	Direct	SVM-DAP	SCA	M2LATM
100I, A/69	66.0	65.7	44.0	65.6
10I, A/69	26.8	40.2	32.2	40.6
10I, R/7	26.8	26.4	25.6	38.3
10I, N/0	26.8	-	17.3	40.4
10I, T/7	26.8	32.4	26.0	38.3
10I, B/7	26.8	18.2	26.0	38.9

Supervised classification performance for USAA. 8 classes, chance = 12.5%. Row labels are I: number of training instances per class, A: all attributes, R: random subset of attributes, N: no attributes, T: top attributes, B: bottom attributes.

Transfer Learning – ZSL

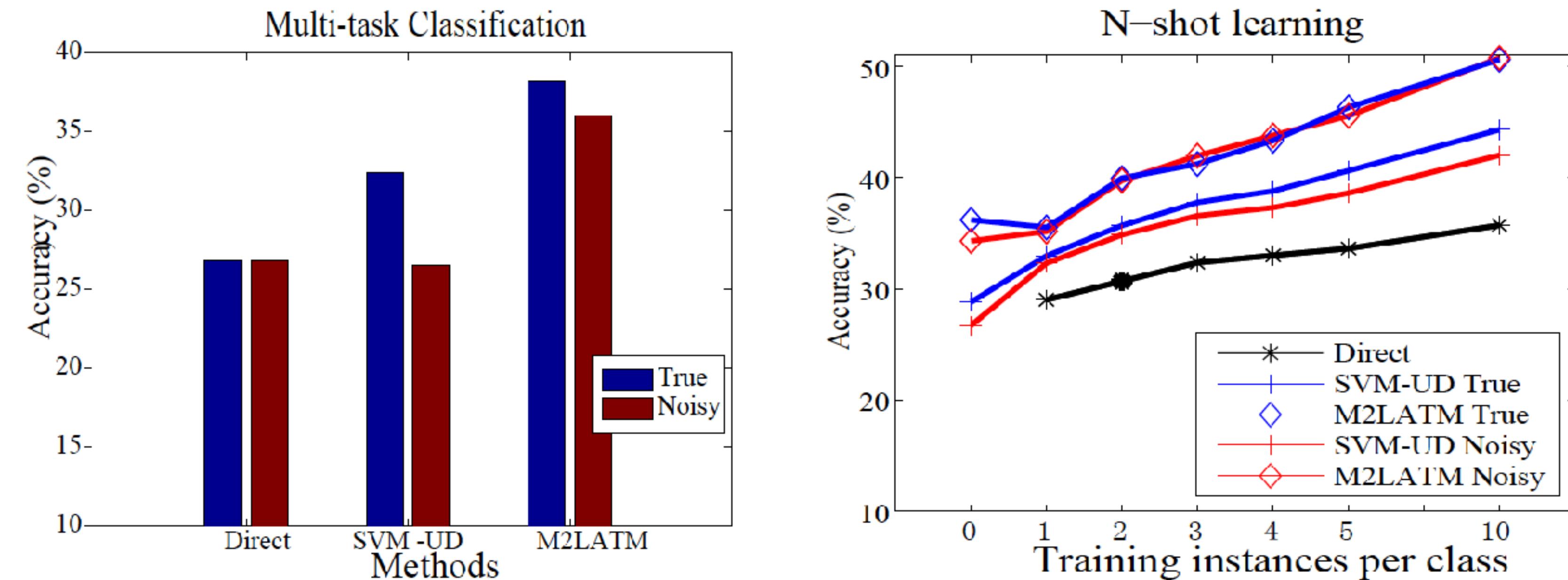
USAA dataset: 4 auxiliary classes; 4 testing classes

	SVM-UD	ST[34]	M2LATM
R/7	27.1	18.1	33.8
O/15	31.3	36.9	39.4
R/34	36.7	30.9	39.2
A/69	33.2	31.0	41.9

AwA dataset: 40 auxiliary classes; 10 testing classes;

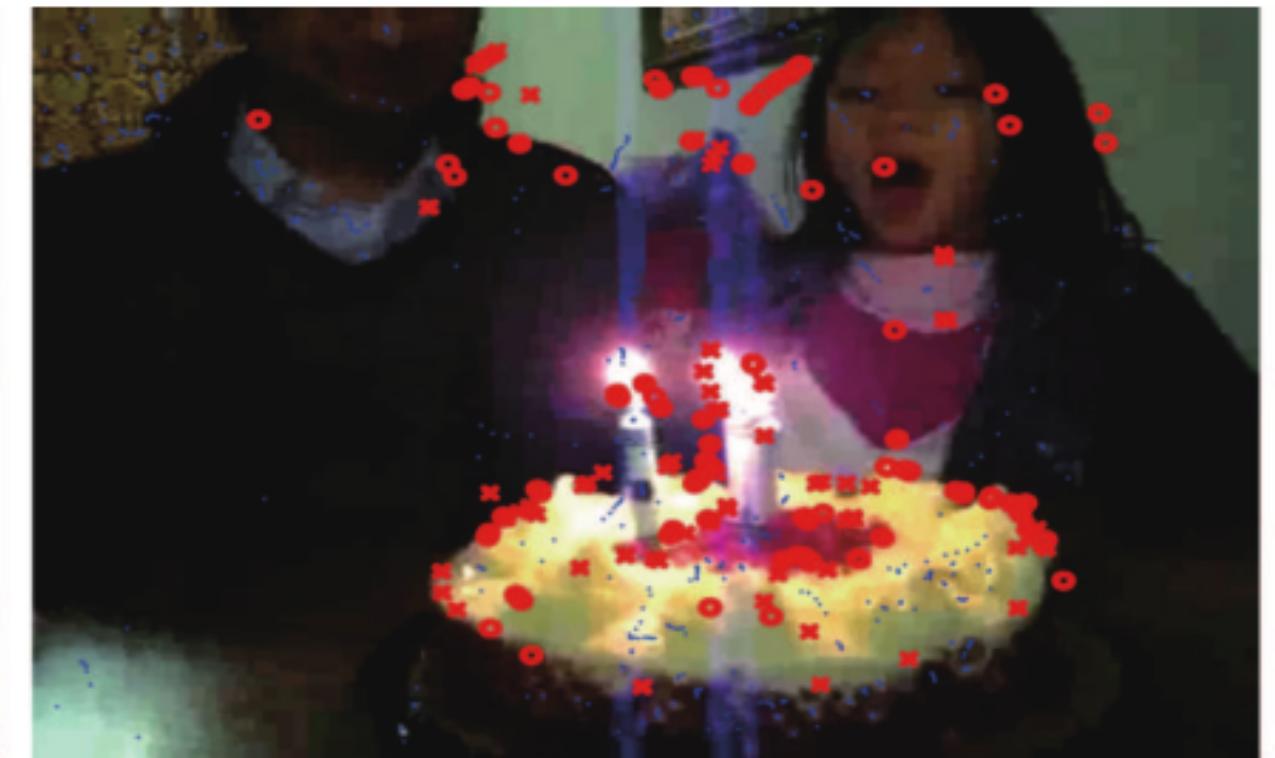
[39]/[26]/[17]	No Attrib Prior		Attrib Prior	
	DAP	M2LATM	DAP	M2LATM
R/9	-	26.3	26.9	27.8
R/42	-	34.4	38.2	36.0
A/85	33.0/33.0/32.7	37.0	39.2	39.2

Robustness to Label Noise



To simulate label noise, we randomly flipped 50% of attribute annotations on 50% of the training videos (so 25% wrong annotations).

Qualitative Results



Red illustrates interest-points from UD or latent attributes;
Blue illustrates interest-points not related to attributes concerned.

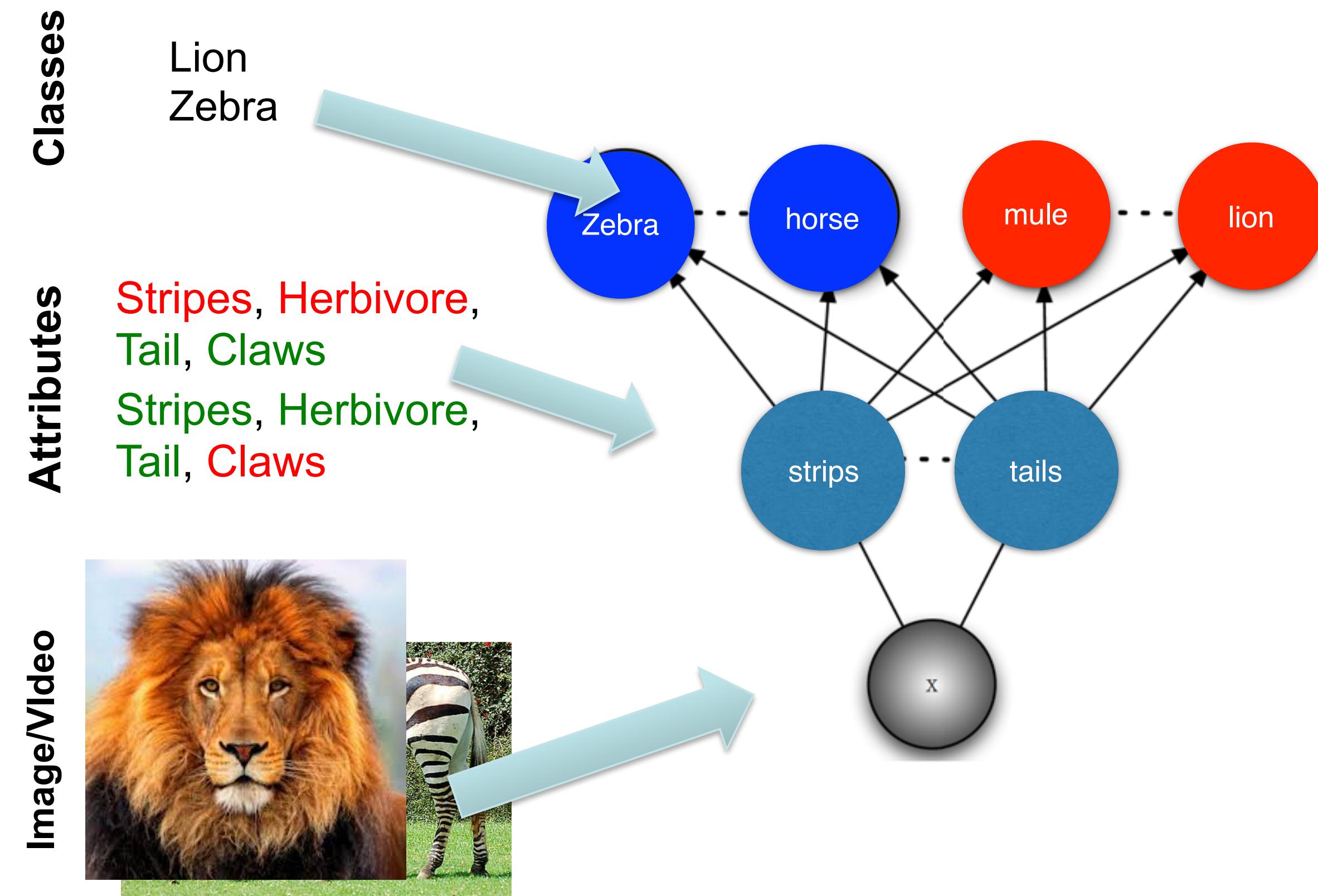
Embedding

Multi-view Embedding

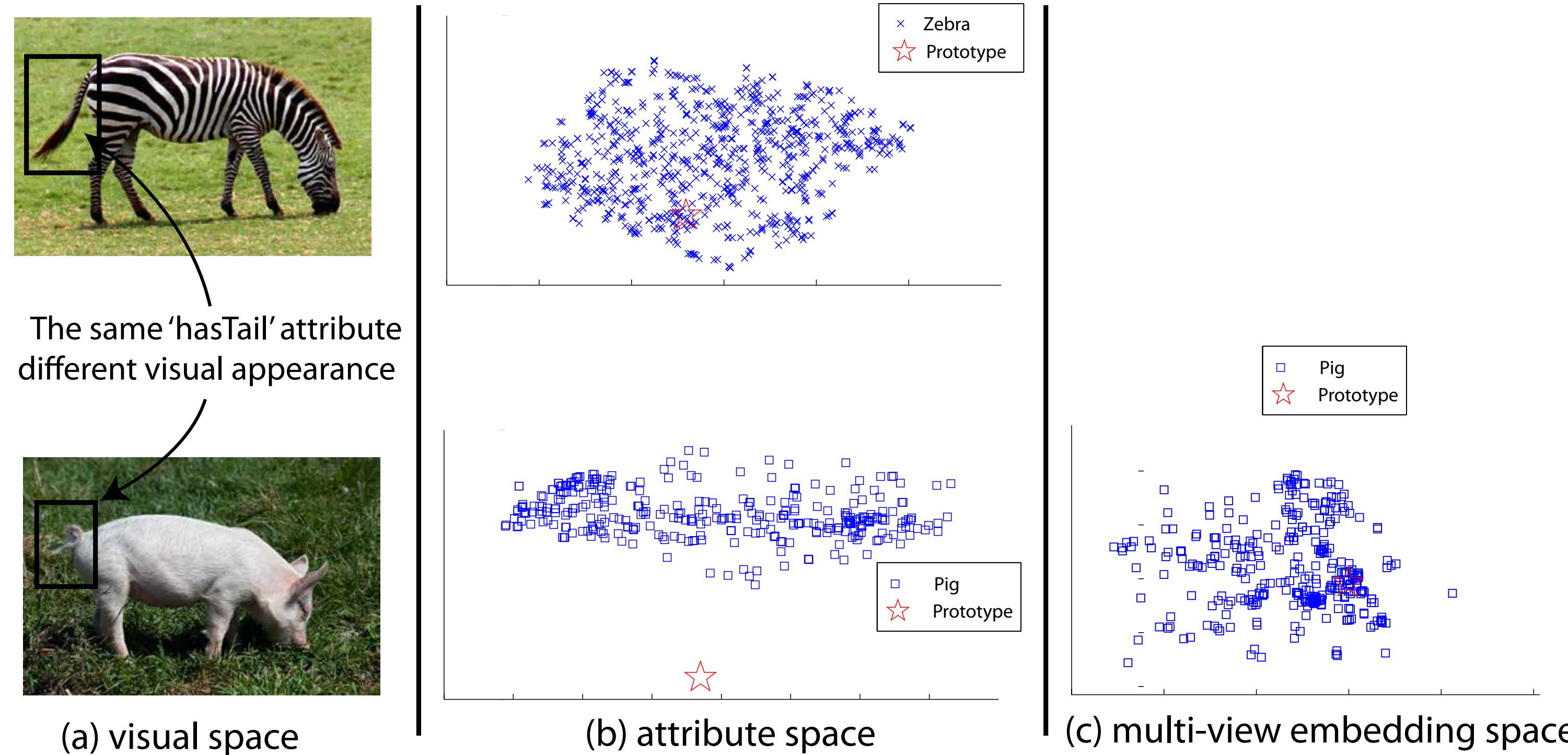


Multi-view Embedding

Existing Attribute Learning Pipeline



Projection Domain Shift Problem



- Zebra (one of auxiliary classes) and Pig (one of target classes) share 'hasTail' attributes and visual appearance differs greatly;
- Projection of low-level features to attributes learned from auxiliary data is shifted when used in target data.

Fu et al. *Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*, ECCV 2014
Fu et al. *Transductive Multi-View Zero-Shot Learning*, IEEE TPAMI 2015

Prototype Sparsity and Multiple Representation Embedding

Prototype sparsity problem,

Each class only has a single attribute prototype which is insufficient to fully represent what that class looks like;

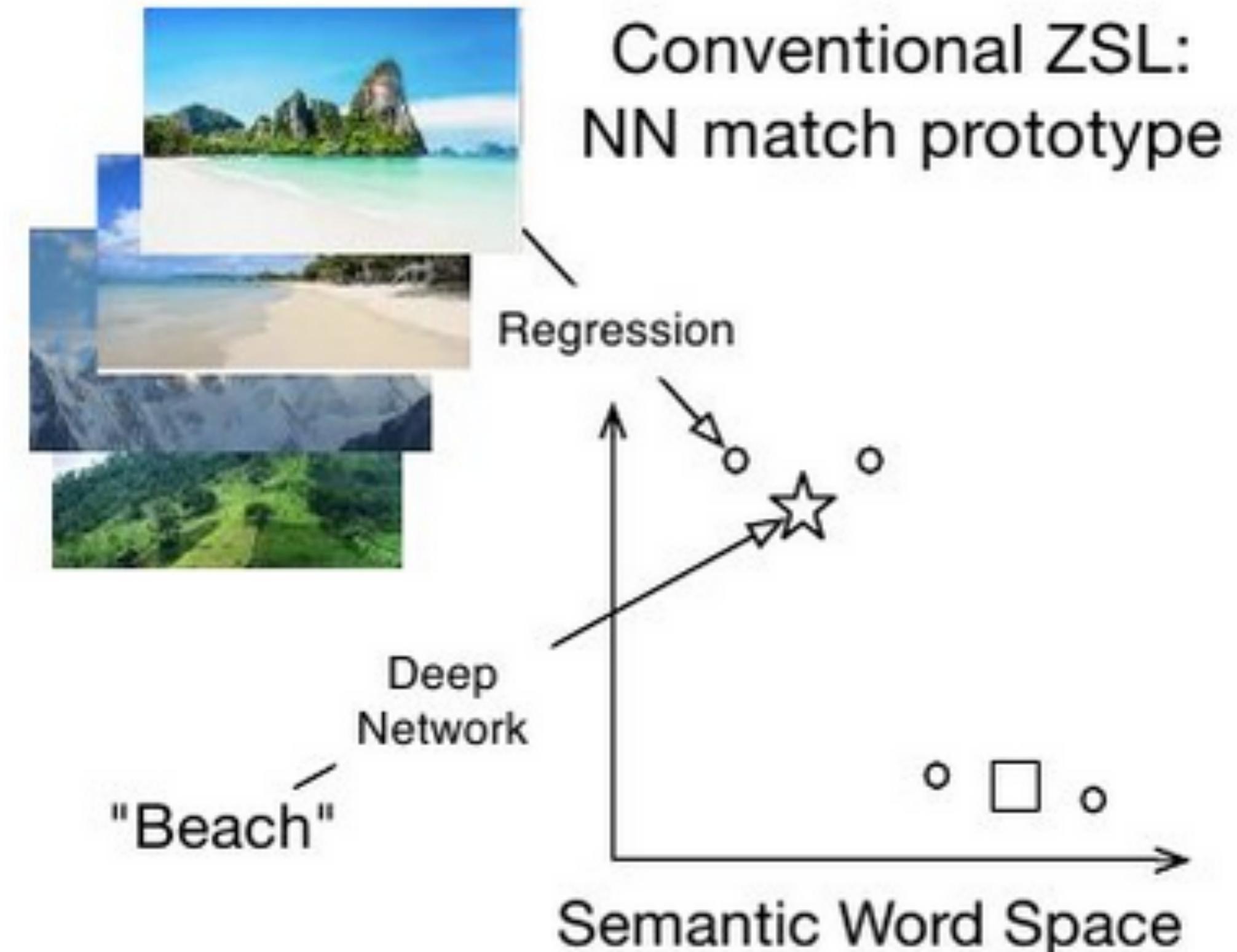
Semantic representations (e.g. word vectors) have manifold structure.

Embedding multiple representations,

Distributed word vectors (word2vec, GloVec);

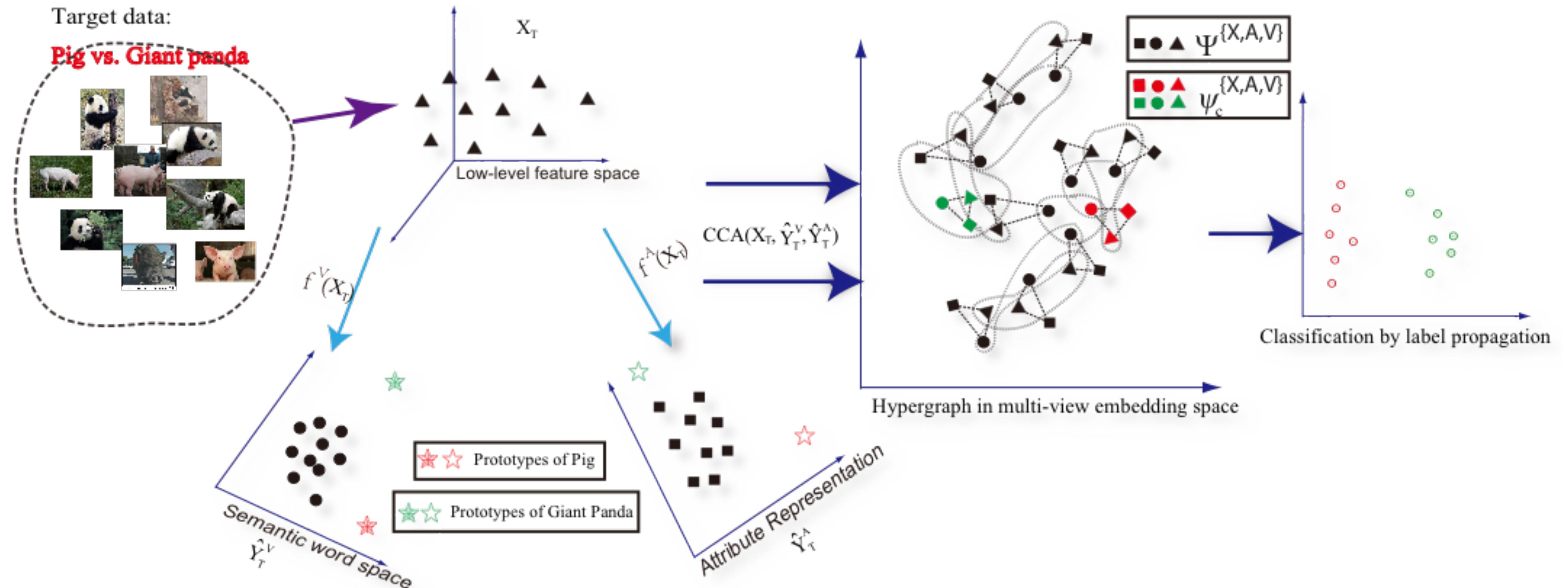
Non-distributed word vectors (tf-idf); and WordNet;

Each may contain complementary information of different classes.



Fu et al. *Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*, ECCV 2014
Fu et al. *Transductive Multi-View Zero-Shot Learning*, IEEE TPAMI 2015

Tranductive Multi-view CCA Embedding



Fu et al. *Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*, ECCV 2014
Fu et al. *Transductive Multi-View Zero-Shot Learning*, IEEE TPAMI 2015

Experimental Results

Approach	AwA (\mathcal{H} [27])	AwA (\mathcal{O})	AwA (\mathcal{O}, \mathcal{D})	USAA	CUB (\mathcal{O})	CUB (\mathcal{F})
DAP	40.5([27])/41.4([28])/38.4*	51.0*	57.1*	33.2([15])/35.2*	26.2*	9.1*
IAP	27.8([27])/42.2([28])	—	—	—	—	—
M2LATM [15]***	41.3	—	—	41.9	—	—
ALE/HLE/AHLE [1]	37.4/39.0/43.5	—	—	—	—	18.0*
Mo/Ma/O/D [38]	27.0/23.6/33.0/35.7	—	—	—	—	—
PST [36]***	42.7	54.1*	62.9*	36.2*	38.3*	13.2*
[54]	48.3**	—	—	—	—	—
TMV-BLP [14]***	47.7	69.9	77.8	48.2	45.2	16.3
TMV-HLP***	49.0	73.5	80.5	50.4	47.9	19.5

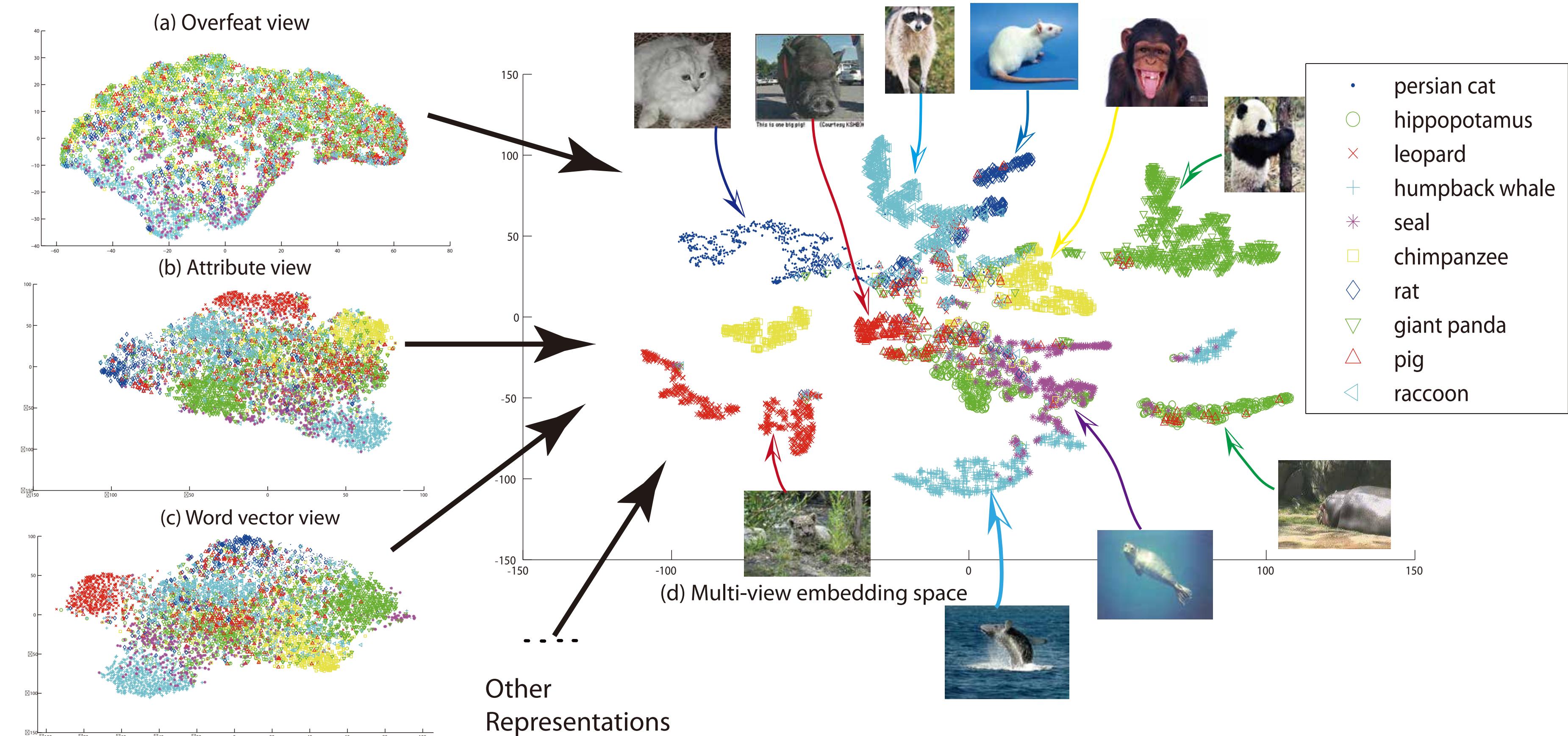
Comparison with State-of-the-Art of ZSL on AwA, USAA and CUB

1. \mathcal{H} , \mathcal{O} , \mathcal{D} and \mathcal{F} represent hand-crafted, OverFeat, DeCAF, and Fisher Vector.
2. *: our implementation; **: requires additional human annotations; ***: requires unlabelled data, i.e. a transductive setting.
3. ‘—’: no result reported;

Fu et al. *Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*, ECCV 2014
Fu et al. *Transductive Multi-View Zero-Shot Learning*, IEEE TPAMI 2015



Visualization of AwA Dataset



t-SNE Visualisation of (a) OverFeat view, (b) attribute view, (c) word vector view, (d) transition probability of pairwise nodes by our framework. The unlabelled target classes are much more separable in (d).

Fu et al. *Transductive Multi-view Embedding for Zero-Shot Recognition and Annotation*, ECCV 2014
Fu et al. *Transductive Multi-View Zero-Shot Learning*, IEEE TPAMI 2015

Embedding

*Pairwise Graph
Embedding*



Subjective Visual Properties

Richer semantic representations;

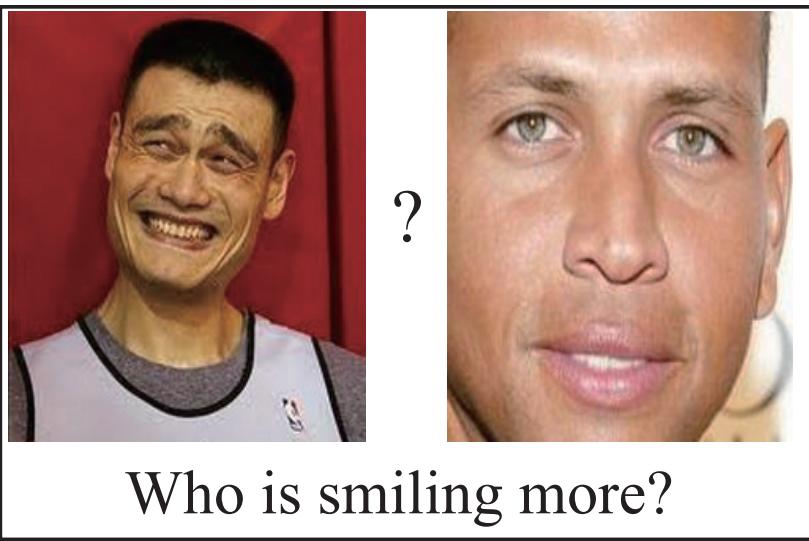
Less ambiguity: *who is smiling more?* Vs.
how much is smiling?

Better for transfer learning & active learning;

Better for interactive image retrieval;

Subjective visual properties,

Image/video interestingness & aesthetics;
Image memorability & image/video quality.



Who is more beautiful?

Fu et al. *Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels*", IEEE TPAMI 2016;
Fu et al. *Interestingness Prediction by Robust Learning to Rank*, ECCV 2014

Crowdsourced Paired Comparisons

Advantages:

1. Cheap and easier for annotators (AMT);
2. We can annotate large-scale datasets;

Problems:

Outliers:

- Ambiguous or Malicious/Lazy annotators;
- Cultural&Psychological factors for inconsistent comparisons;

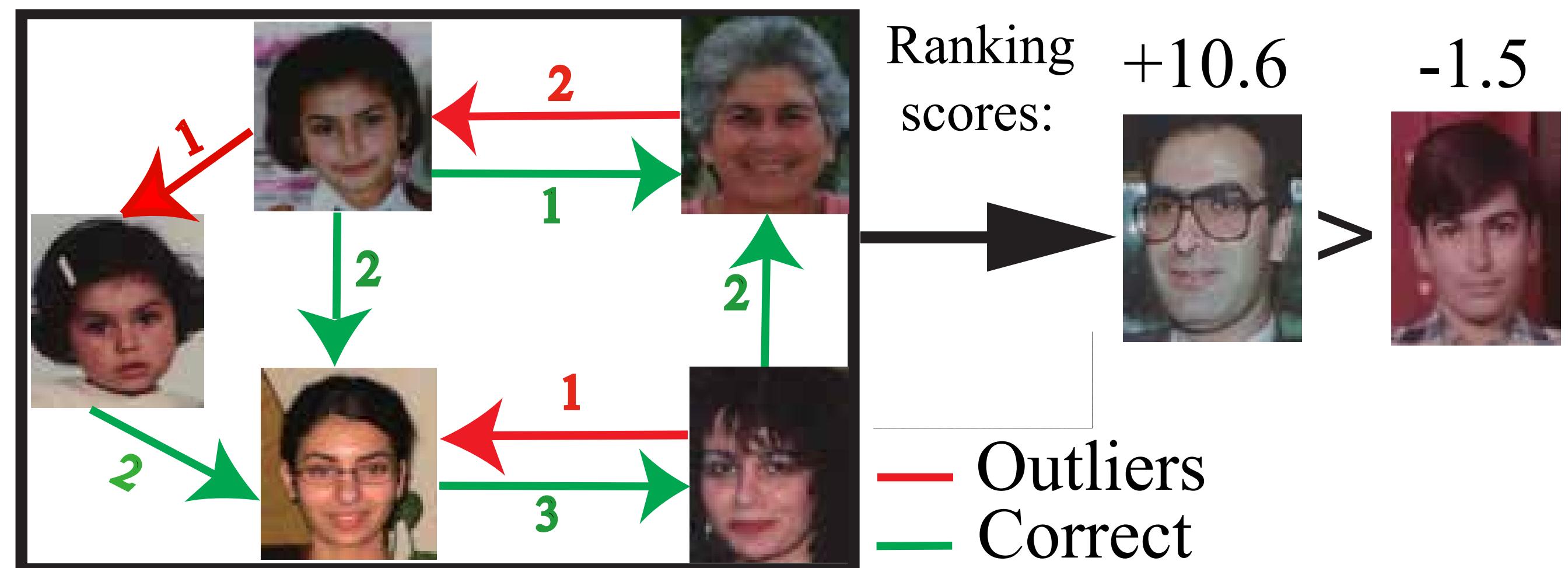
Sparsity:

- $O(n^2)$ for n instances;
- Large-scale dataset even with lots of comparisons, resulting in sparse data.

Fu et al. *Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels*”, IEEE TPAMI 2016;
Fu et al. *Interestingness Prediction by Robust Learning to Rank*, ECCV 2014



Overview — Robust Learning to Rank



Synthetic experiment: Human Age Prediction from Face Images

Pairs data are naturally directed graphs $G = (V, E)$ $E = \{i \rightarrow j | Y_{ij} > 0\}$

Low-level feature matrix $\Phi = [\phi_i^T]_{i=1}^I \in R^{I \times d}$; $V = \{i\}_{i=1}^I$

The edge $i \rightarrow j$ exists if $Y_{ij} > 0$.

Fu et al. *Robust Subjective Visual Property Prediction from Crowdsourced Pairwise Labels*, IEEE TPAMI 2016;
Fu et al. *Interestingness Prediction by Robust Learning to Rank*, ECCV 2014

Robust Learning to Rank (URLR) Framework

Each edge $i \rightarrow j \in E$, Y_{ij} is modelled as

$$Y_{ij} = \beta^T \phi_i - \beta^T \phi_j + \gamma_{ij}$$

The edges of the whole dataset are

$$Y = C\Phi\beta + \Gamma$$

where C is incidence matrix. γ_{ij} is also known as incidental parameter [*_1].

We minimize the discrepancy between annotation and prediction $C\Phi\beta + \Gamma$ as well as keeping sparse outliers Γ ,

$$\begin{aligned} & \min_{\beta, \Gamma} \frac{1}{2} \|Y - C\Phi\beta - \Gamma\|_2^2 + \lambda \|\Gamma\|_1 \\ & := \sum_{i \rightarrow j \in E} \left[\frac{1}{2} (Y_{ij} - \gamma_{ij} - \beta^T \phi_i - \beta^T \phi_j)^2 + \lambda |\gamma_{ij}| \right] \end{aligned}$$

[*_1] **Jianqing Fan**, Runlong Tang and Xiaofeng Shi, Partial Consistency with Sparse Incidental Parameters, arXiv:1210.6950, 2012.



Robust Learning to Rank (URLR) Framework (cont.)

To solve it, we rewrite the cost function as Lagaragain,

$$L(\beta, \Gamma) = \frac{1}{2} \|Y - X\beta - \Gamma\|_2^2 + \lambda \|\Gamma\|_1$$

where $X = C\Phi$, $X = U\Sigma A^T$ and $U = [U_1; U_2]$

We have

$$\begin{aligned}\hat{\beta} &= (X^T X)^{\dagger} X^T (Y - \Gamma) \\ \hat{\Gamma} &= \arg \min_{\Gamma} \|U_2^T Y - U_2^T \Gamma\| + \lambda \|\Gamma\|_1\end{aligned}$$

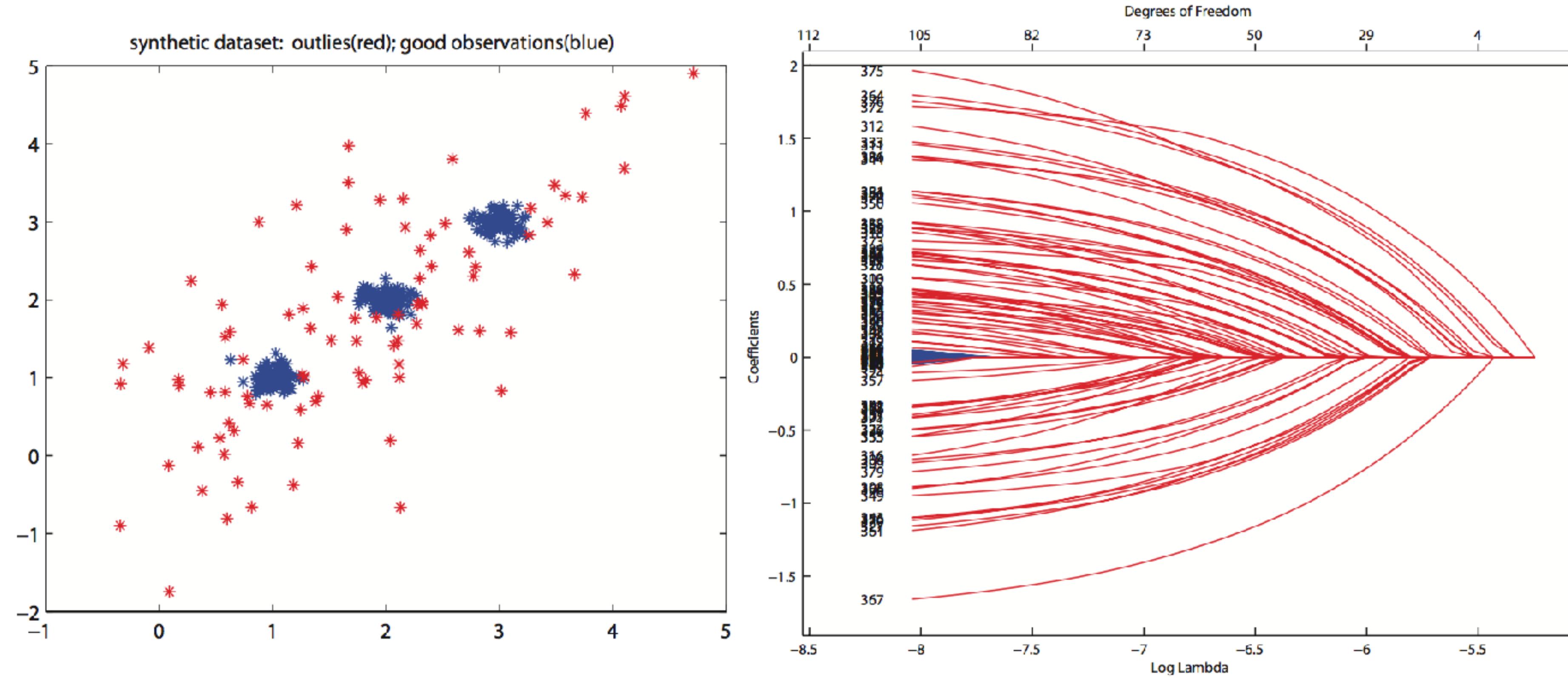
We solve Γ checking regularisation path,

It's one type of Preconditioned Lasso![*2].

[*2] Fabian L. Wauthier, Nebojsa Jojic and Michael I. Jordan, A Comparative Framework for Preconditioned Lasso Algorithms, NIPS 2013.



Checking Regularisation Path



Red lines & red points indicate outliers;
Blue lines & blue points are inliers. Figures from [*3].

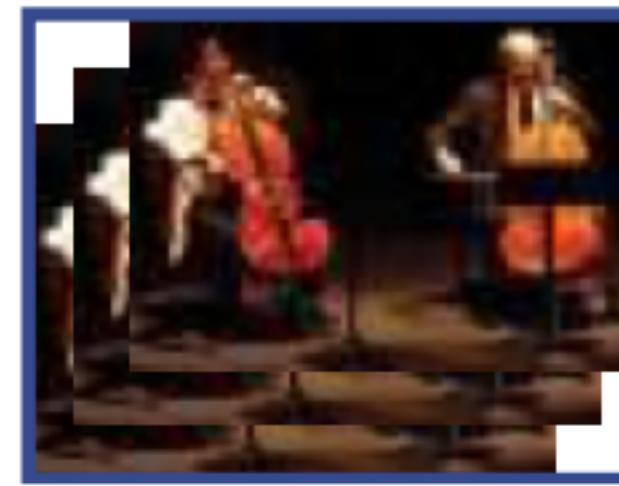
[*3] **Yanwei Fu**, De-An Huang, Leonid Sigal, Robust Classification by Pre-conditioned LASSO and Transductive Diffusion Component Analysis, submitted to ICLR 2016: <http://arxiv.org/abs/1511.06340>

Content

More
of our works



Video Understanding with Objects/Scenes



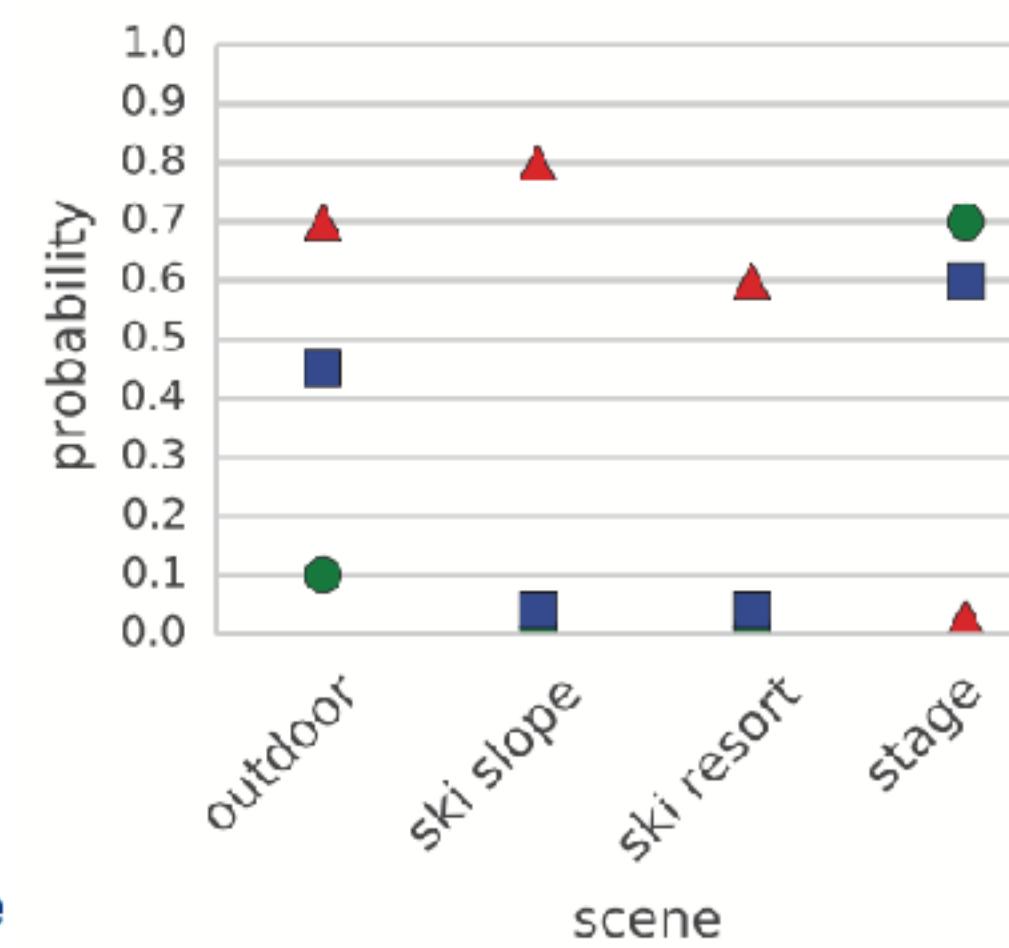
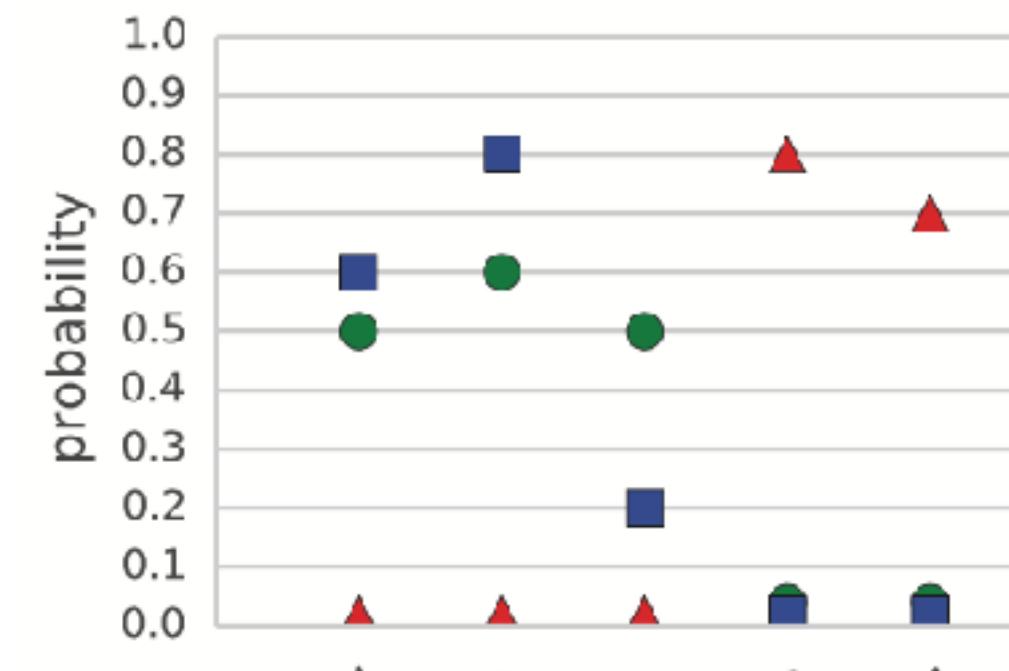
Cello Performance



Skiing



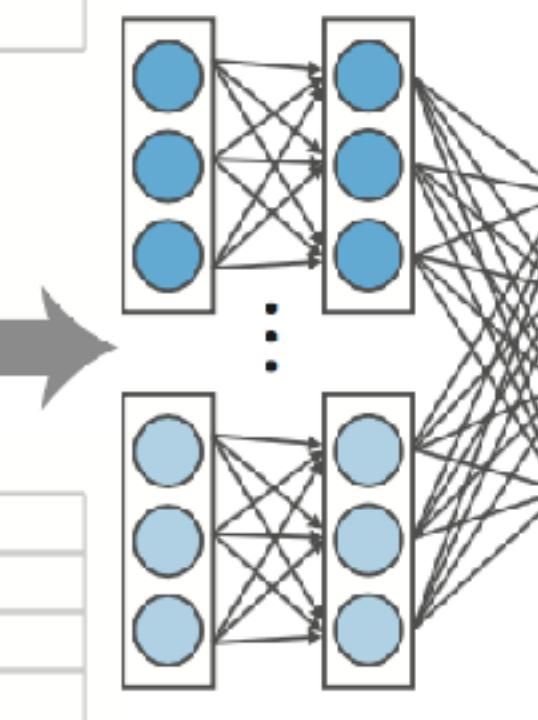
Symphony Performance



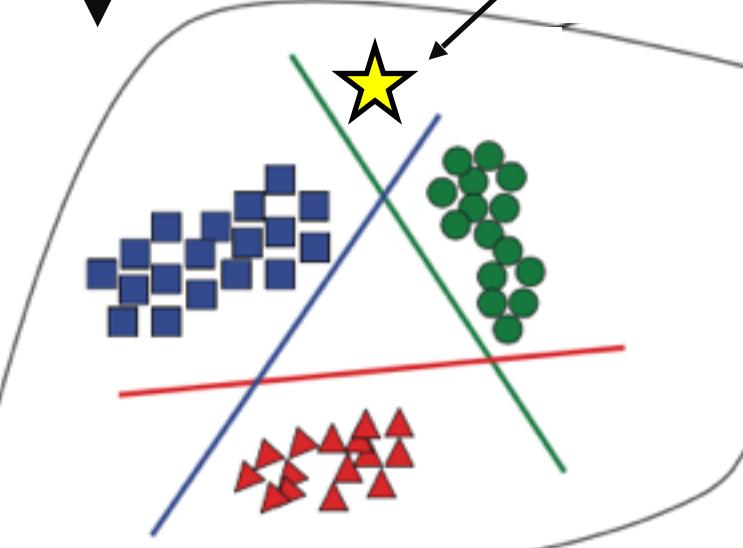
Mining Relationships



Zero-Shot Learning



Flute Performance



Classification

Learning to Generate Posters of Scientific Papers

FACE SPOOFING DETECTION THROUGH PARTIAL LEAST SQUARES AND LOW-LEVEL DESCRIPTORS

William Robson Schwartz, Anderson Rocha, Helio Pedrinil
Institute of Computing, University of Campinas

Introduction

- Problem: 2-D image-based facial verification or recognition system can be spoofed with no difficulty (a person displays a photo of an authorized subject either printed on a piece paper).
- Idea: anti-spoofing solution based on a holistic representation of the face region through a robust set of low-level feature descriptors, exploiting spatial and temporal information.
- Advantages: PLS allows to use multiple features and avoids the necessity of choosing before-hand a smaller set of features that may not be suitable for the problem.

EXPERIMENTAL RESULTS

Print-Attack Dataset

Dataset: 200 real-access and 200 printed-photo attack videos [1].

PARTIAL LEAST SQUARES

- PLS deals with a large number of variables and a small number of examples.
- Data matrix X and response matrix $Y_{N \times N} = TP_f + E$.
- Practical Solution: NIPALS algorithm iterative approach to calculate PLS factors.
- PLS weights the feature descriptors and estimates the location of the most discriminative regions.

NUAA Dataset

Dataset: 1743 live images and 1740 non-live images for training, 3362 live and 5761 non-live images for testing [4].

Setup: faces are detected and images are scaled to 64 x 64 pixels.

Comparison: Tan et al. [4] achieved AUC of 0.95.

ANTI-SPOOFING PROPOSED SOLUTION

- A video sample is divided into m parts, feature extraction is applied for every k -th frame. The resulting descriptors are concatenated to compute the feature vector.
- PLS is employed to obtain the latent feature space, in which higher weights are attributed to feature descriptors extracted from regions containing discriminatory characteristics between the two classes.
- The test procedure evaluates if a novel sample belongs either to the live or non-live class. When a sample video is presented to the system, the face is detected and the frames are cropped and rescaled.

[1] <http://www.idiap.ch/datasets/printattack>
[2] W. H. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In IEEE ICCV, pages 2403–2009.
[3] W. R. Schwartz, R. C. da Silva, and H. Pedrinil. A Novel Feature Descriptor Based on the Shearlet Transform. In IEEE ICIP, 2011.
[4] X. Tan, Y. Li, J. Liu, and L. Jiaqin. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In ECCV, pages 504–517, 2010.

(a) Designed by novice

FACE SPOOFING DETECTION THROUGH PARTIAL LEAST SQUARES AND LOW-LEVEL DESCRIPTORS

William Robson Schwartz, Anderson Rocha, Helio Pedrinil
Institute of Computing, University of Campinas

Introduction

- Problem: 2-D image-based facial verification or recognition system can be spoofed with no difficulty (a person displays a photo of an authorized subject either printed on a piece paper).
- Idea: anti-spoofing solution based on a holistic representation of the face region through a robust set of low-level feature descriptors, exploiting spatial and temporal information.
- Advantages: PLS allows to use multiple features and avoids the necessity of choosing before-hand a smaller set of features that may not be suitable for the problem.

Anti-Spoofing Proposed Solution

- A video sample is divided into m parts, feature extraction is applied for every k -th frame. The resulting descriptors are concatenated to compute the feature vector.
- PLS is employed to obtain the latent feature space, in which higher weights are attributed to feature descriptors extracted from regions containing discriminatory characteristics between the two classes.
- The test procedure evaluates if a novel sample belongs either to the live or non-live class. When a sample video is presented to the system, the face is detected and the frames are cropped and rescaled.

Partial Least Squares

- PLS deals with a large number of variables and a small number of examples.
- Data matrix X and response matrix $Y_{N \times N} = TP_f + E$, $Y_{N \times N} = UQ^T + F$.
- Practical Solution: NIPALS algorithm iterative approach to calculate PLS factors.
- PLS weights the feature descriptors and estimates the location of the most discriminative regions.

Experimental Results

Print-Attack Dataset

Dataset: 200 real-access and 200 printed-photo attack videos [1].

Setup: face detection, rescale to 110 x 40 pixels, 10 frames are sampled for feature extraction (HOG, intensity, color frequency (CF) [2], histogram of shearlet coefficients (HSC) [3], GLCM).

Classifier evaluation: SVM type C with linear kernel achieved EER of 10.

NUAA Dataset

Dataset: 1743 live images and 1740 non-live images for training, 3362 live and 5761 non-live images for testing [4].

Setup: faces are detected and images are scaled to 64 x 64 pixels.

Comparison: Tan et al. [4] achieved AUC of 0.95.

Experimental Results

Print-Attack Dataset

Dataset: 200 real-access and 200 printed-photo attack videos [1].

Setup: face detection, rescale to 110 x 40 pixels, 10 frames are sampled for feature extraction (HOG, intensity, color frequency (CF) [2], histogram of shearlet coefficients (HSC) [3], GLCM).

Classifier evaluation: SVM type C with linear kernel achieved EER of 10%.

NUAA Dataset

Dataset: 1743 live images and 1740 non-live images for training, 3362 live and 5761 non-live images for testing [4].

Setup: faces are detected and images are scaled to 64 x 64 pixels.

Comparison: Tan et al. [4] achieved AUC of 0.95.

Feature combination

Name	# Descriptors	EER (%)
HOG	326,800	11.57
Intensity	154,000	8.50
CF	27,160	5.67
GLCM	159,380	5.67
HSC	581,120	4.33
Combination	1,094,600	1.67

Comparisons

Team	FAR (%)	FRR (%)
IDMAP	0.09	0.09
UDIULU	0.09	0.09
AMLAB	0.08	1.25
CASIA	0.08	0.08
SIAU	0.08	21.25
Our results	1.23	0.08

Feature combination

[1] <http://www.idiap.ch/datasets/printattack>
[2] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In IEEE ICCV, pages 2403–2009.
[3] W. R. Schwartz, R. C. da Silva, and H. Pedrinil. A Novel Feature Descriptor Based on the Shearlet Transform. In IEEE ICIP, 2011.
[4] X. Tan, Y. Li, J. Liu, and L. Jiaqin. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In ECCV, pages 504–517, 2010.

(b) Our result

FACE SPOOFING DETECTION THROUGH PARTIAL LEAST SQUARES AND LOW-LEVEL DESCRIPTORS

William Robson Schwartz, Anderson Rocha, Helio Pedrinil
Institute of Computing, University of Campinas

Introduction

Problem: 2-D image-based facial verification or recognition system can be spoofed with no difficulty (a person displays a photo of an authorized subject either printed on a piece paper).

Idea: anti-spoofing solution based on a holistic representation of the face region through a robust set of low-level feature descriptors, exploiting spatial and temporal information.

Advantages: PLS allows to use multiple features and avoids the necessity of choosing before-hand a smaller set of features that may not be suitable for the problem.

Anti-Spoofing Proposed Solution

A video sample is divided into m parts, feature extraction is applied for every k -th frame. The resulting descriptors are concatenated to compose the feature vector.

PLS is employed to obtain the latent feature space, in which higher weights are attributed to feature descriptors extracted from regions containing discriminatory characteristics between the two classes.

The test procedure evaluates if a novel sample belongs either to the live or non-live class. When a sample video is presented to the system, the face is detected and the frames are cropped and rescaled.

Partial Least Squares

PLS deals with a large number of variables and a small number of examples.

Data matrix X and response matrix $Y_{N \times N} = TP_f + E$, $Y_{N \times N} = UQ^T + F$.

Practical Solution: NIPALS algorithm iterative approach to calculate PLS factors.

PLS weights the feature descriptors and estimates the location of the most discriminative regions.

Experimental Results

Print-Attack Dataset

Dataset: 200 real-access and 200 printed-photo attack videos [1].

Setup: face detection, rescale to 110 x 40 pixels, 10 frames are sampled for feature extraction (HOG, intensity, color frequency (CF) [2], histogram of shearlet coefficients (HSC) [3], GLCM).

Classifier evaluation: SVM type C with linear kernel achieved EER of 10%.

NUAA Dataset

Dataset: 1743 live images and 1740 non-live images for training, 3362 live and 5761 non-live images for testing [4].

Setup: faces are detected and images are scaled to 64 x 64 pixels.

Comparison: Tan et al. [4] achieved AUC of 0.95.

Feature combination

Name	# Descriptors	EER (%)
HOG	4,020	52.20
Intensity	6,981	11.80
CF	1,246	11.40
GLCM	3,552	9.60
Combination	22,952	8.20

Comparisons

Team	FAR (%)	FRR (%)
IDMAP	0.09	0.09
UDIULU	0.09	0.09
AMLAB	0.08	1.25
CASIA	0.08	0.08
SIAU	0.08	21.25
Our results	1.23	0.08

Feature combination

[1] <http://www.idiap.ch/datasets/printattack>
[2] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In IEEE ICCV, pages 2403–2009.
[3] W. R. Schwartz, R. C. da Silva, and H. Pedrinil. A Novel Feature Descriptor Based on the Shearlet Transform. In IEEE ICIP, 2011.
[4] X. Tan, Y. Li, J. Liu, and L. Jiaqin. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In ECCV, pages 504–517, 2010.

(c) Original poster

Qiang et al. “Learning to Generate Posters of Scientific Papers”, AAAI 2016;
 Qiang et al. “Learning to Generate Posters of Scientific Papers by Probabilistic Graphical Model”, submitted to a journal;



Thanks!

yanwei.fu@iclouds.com

