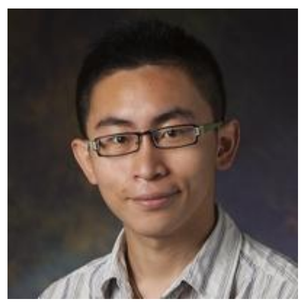
Unsupervised Learning of Object Representations from Geodesic Space Clustering of Disparate Views


[1]Dong Li


[2]Wei-Chih Hung


[3]Jia-Bin Huang


[1]Shengjin Wang


[3]Narendra Ahuja


[2]Ming-Hsuan Yang

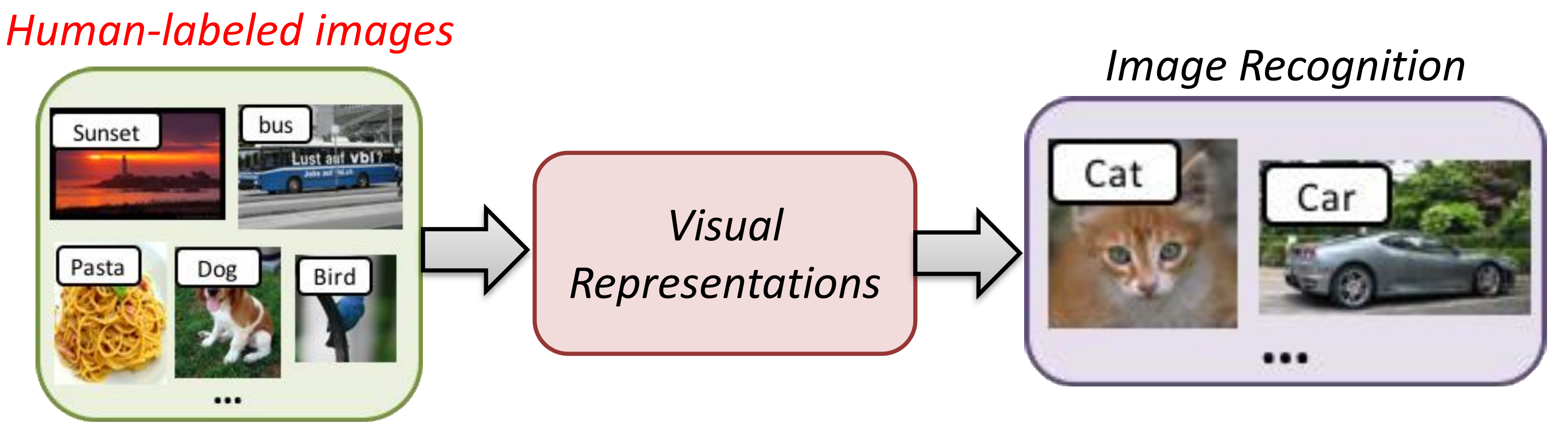
[1]Tsinghua University [2]University of California, Merced [3]University of Illinois, Urbana-Champaign

http://bit.ly/feature-learning-eccv2016

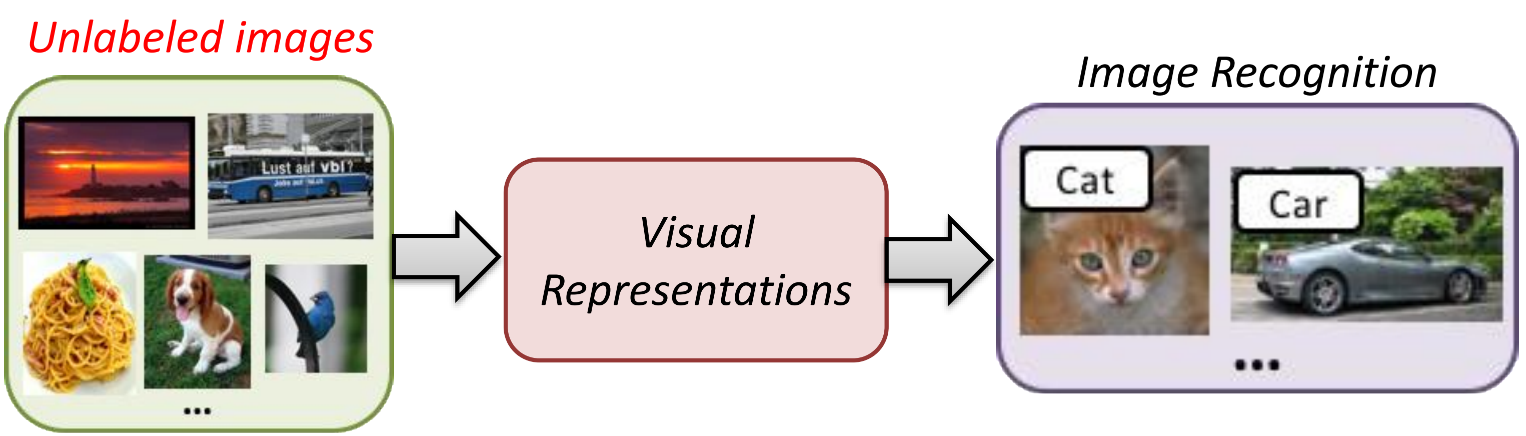
To be presented at ECCV 2016

Representation Learning

- Supervised learning: Expensive annotations & Poor scalability



- **Goal**: Visual representation learning with a large, unlabeled image collection

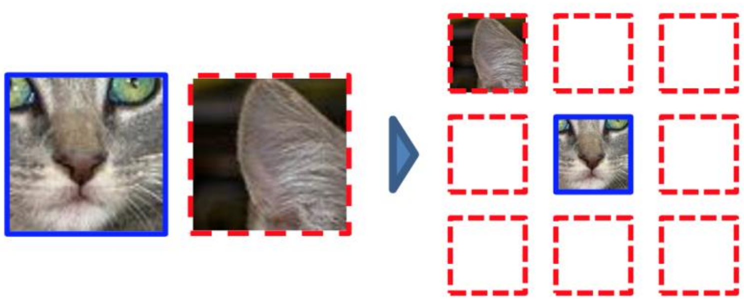


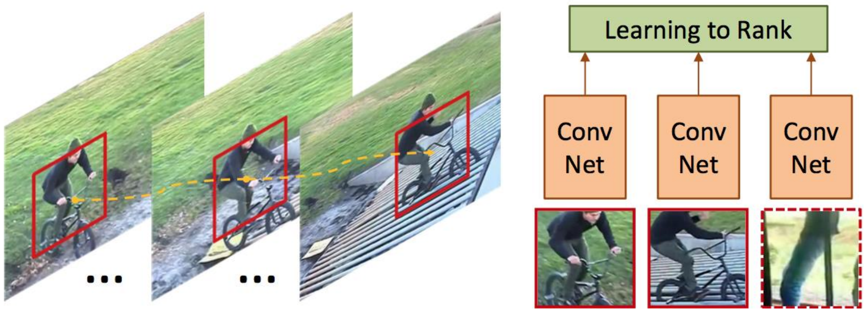
Prior Work - Representation Learning

- Class labels *[Krizhevsky et al. NIPS'12]*
- Web resources *[Chen and Gupta ICCV'15; Joulin et al. ECCV'16]*
- Ego-motion *[Agrawal et al. ICCV'15; Jayaraman et al. ICCV'15]*
- Context *[Doersch et al. ICCV'15]*
- Tracking *[Wang and Gupta ICCV'15]*

Context: instances ***within*** the same image

Tracking: instances ***within*** the same video

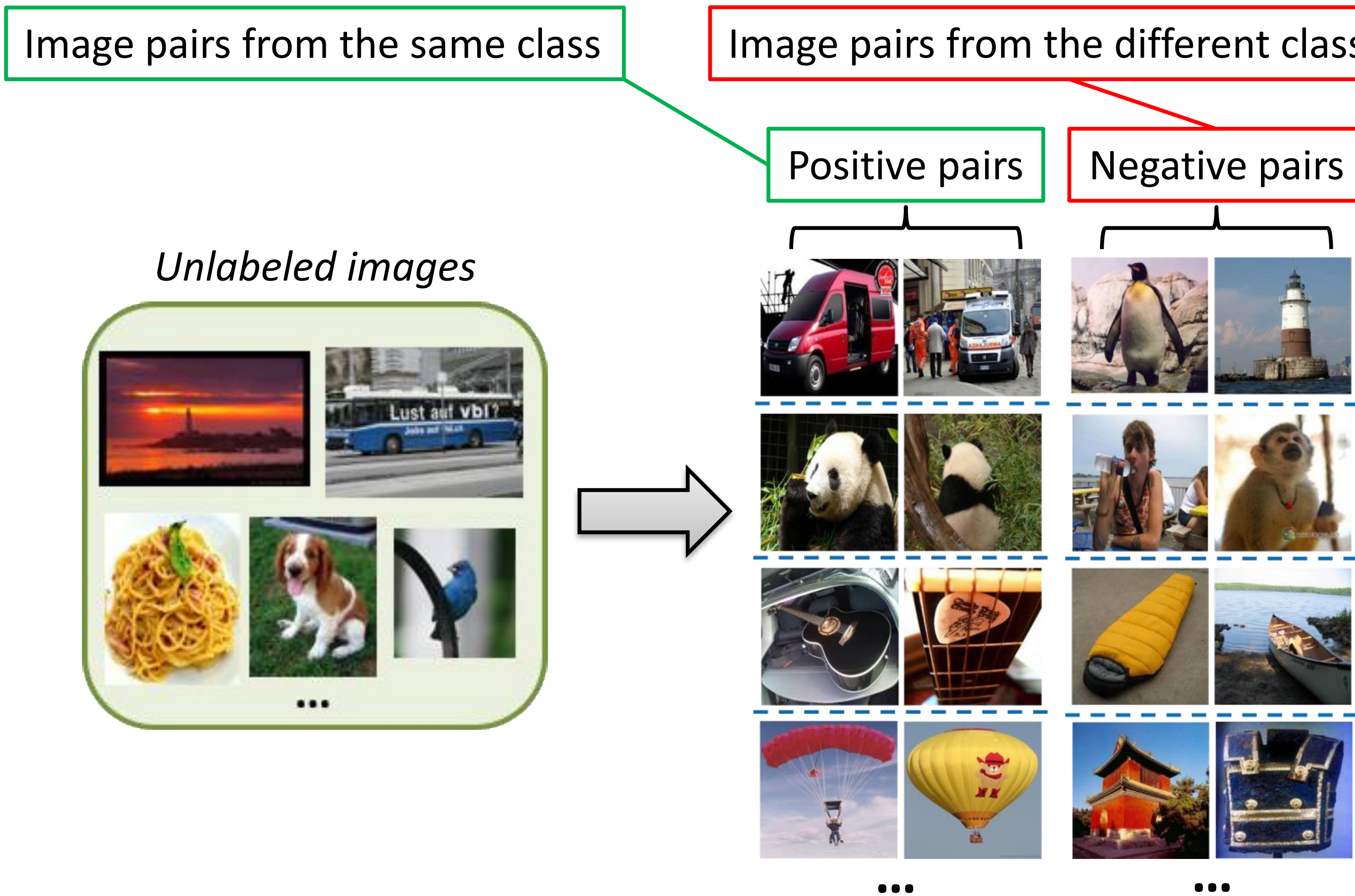




instance-level training data

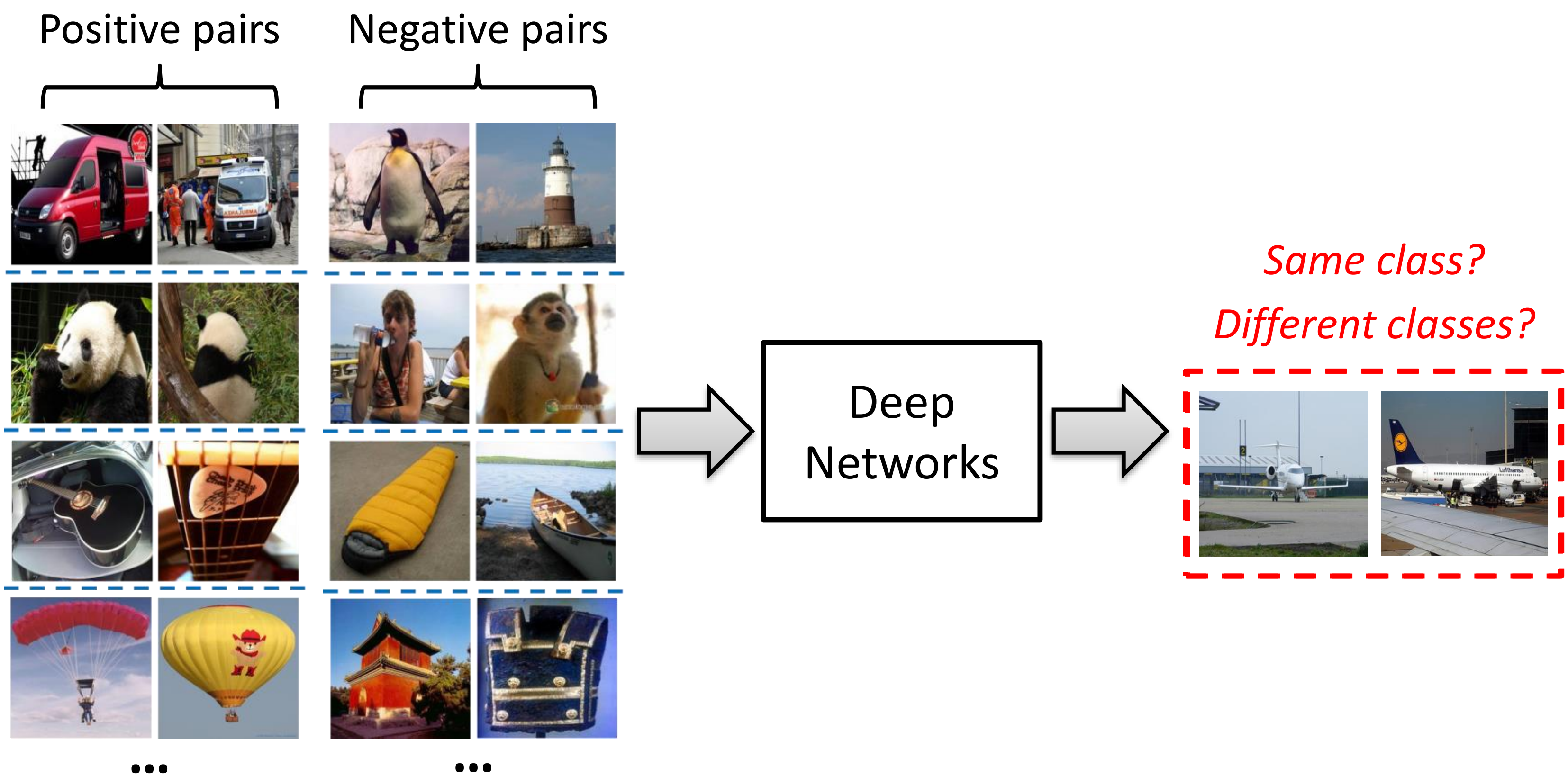
Main Idea - Mining

Mine ***category-level*** training samples ***across*** different images



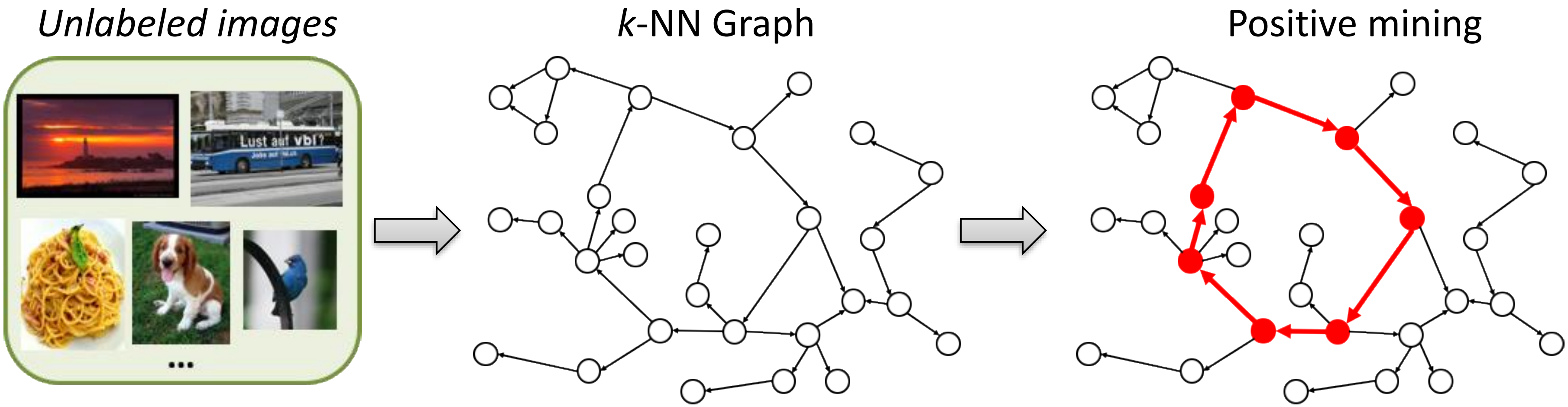
Main Idea - Training

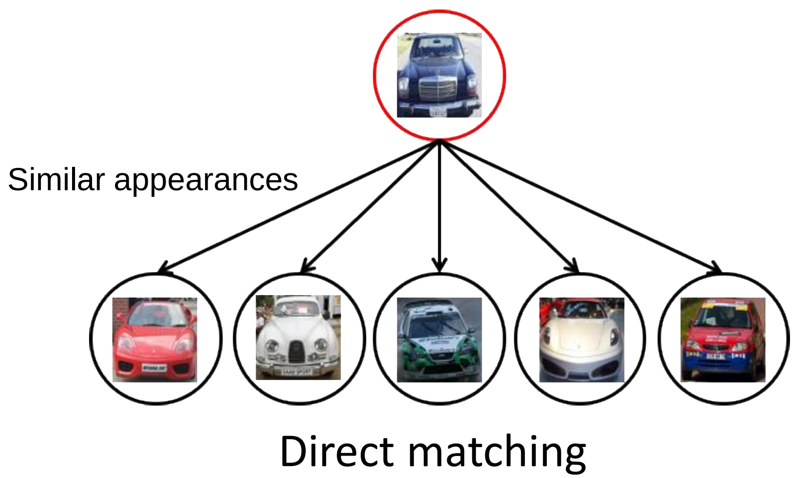
Learn visual representations for ***binary*** classification

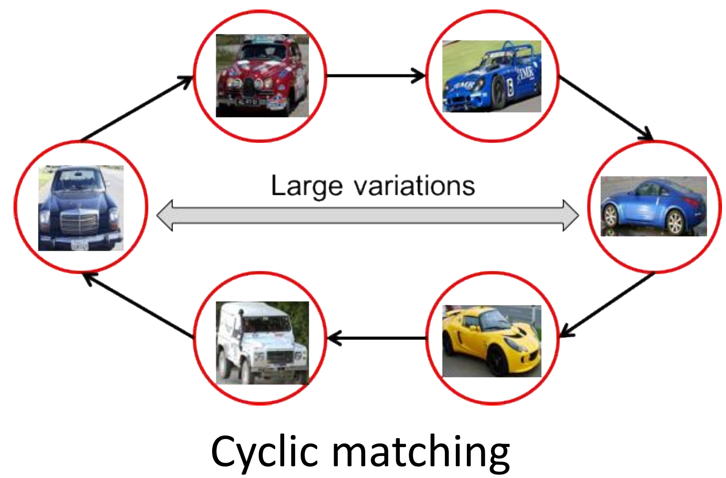


Positive Mining

Cycle consistency: positive pairs with large appearance variations

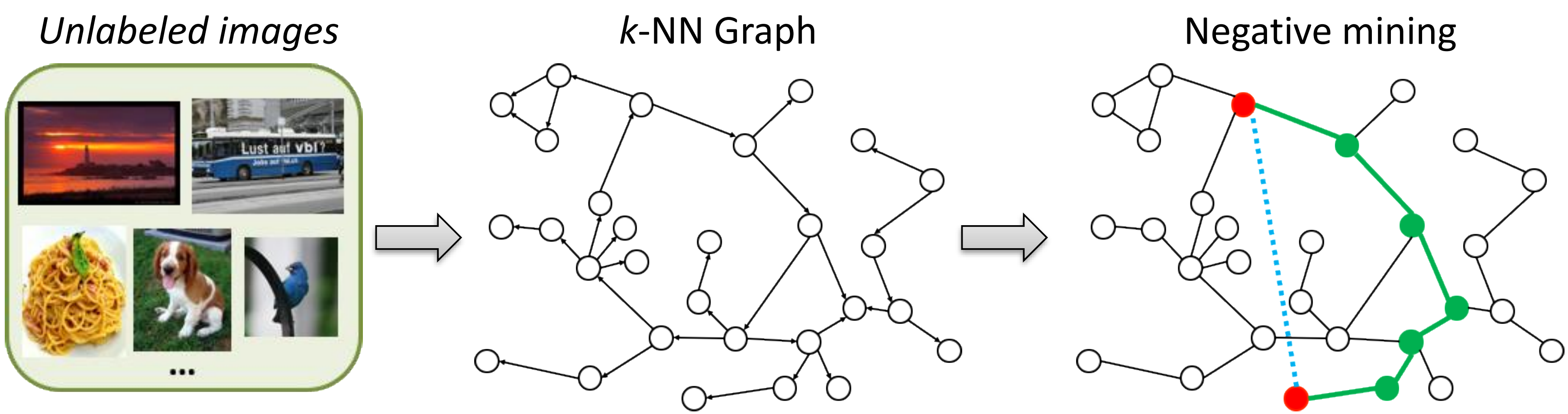


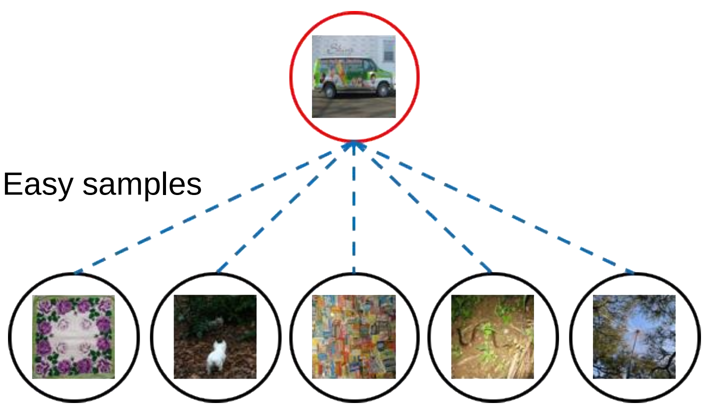




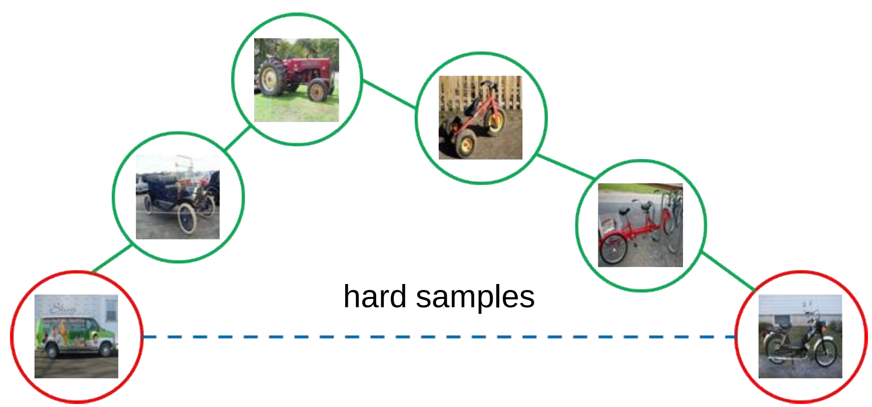
Negative Mining

Geodesic distance: hard negative pairs with a relatively small *L*2 distance





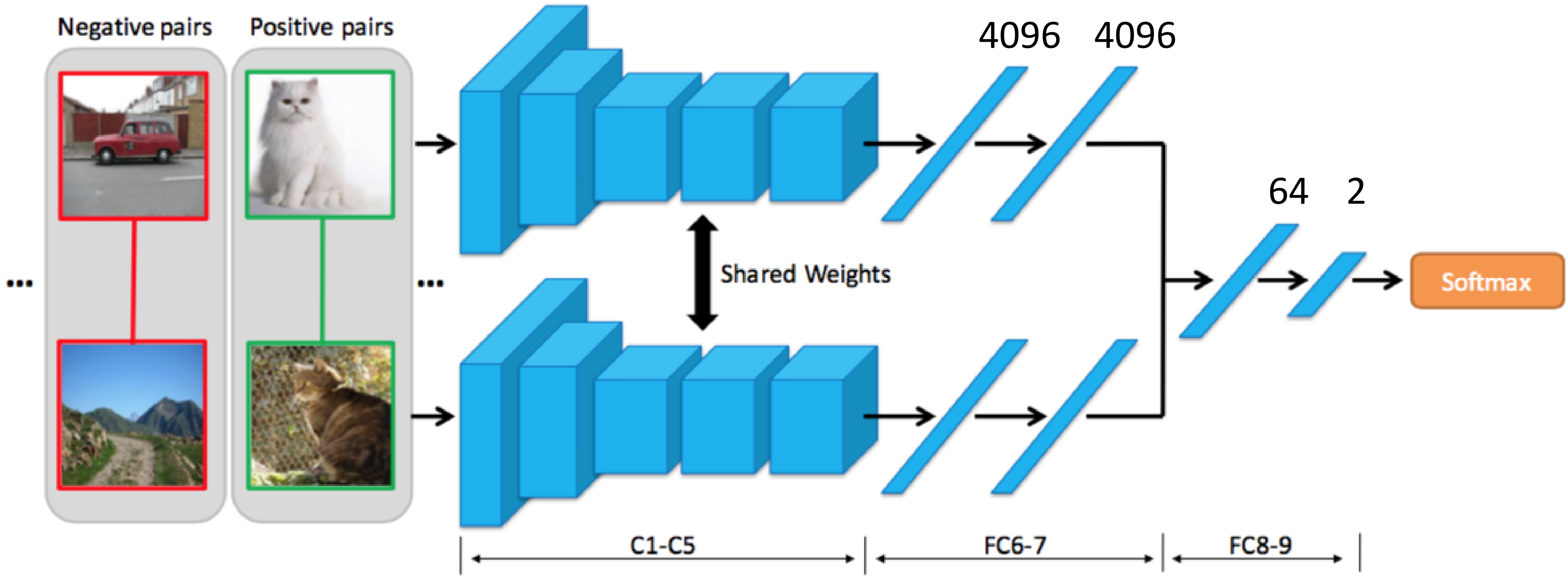
Large *L*2 distance



Large geodesic distance

Pair-wise Training

Siamese network for binary pair classification



Controlled Experiments (CIFAR-10)

- Evaluation on positive mining

	Random sampling	Direct matching	2-cycle	3-cycle	4-cycle	5-cycle
TP rate	10.0	59.0	73.8	82.9	83.0	81.7
Accuracy	73.7	78.0	79.9	80.5	80.9	80.2

Accurate positive pairs & Better CNN representations

- Evaluation on negative mining

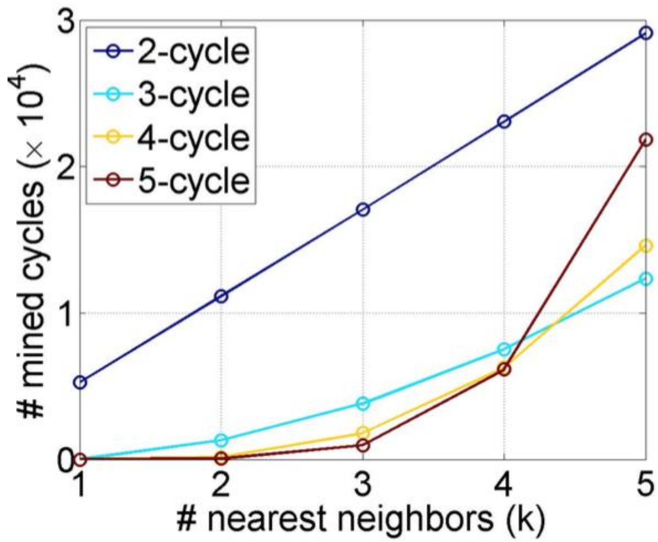
	Random sampling	Original distance	Geodesic distance
TN rate	90.0	95.5	91.0
Accuracy	83.8	68.3	85.2

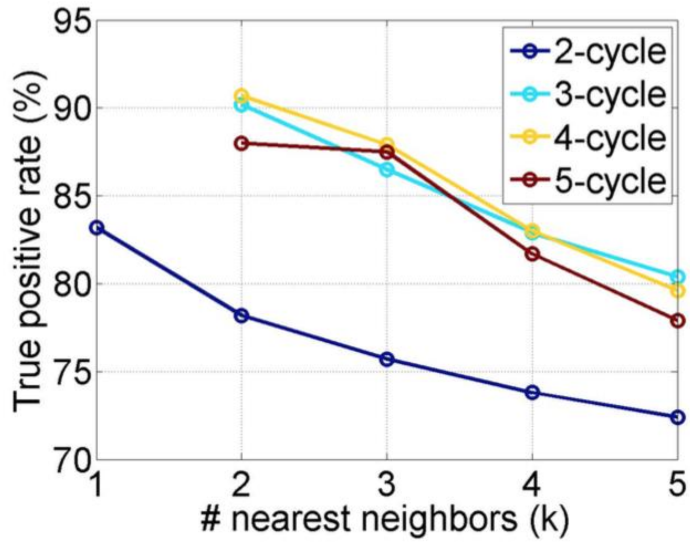
Easy samples

Hard samples

Controlled Experiments (CIFAR-10)

- Parameter analysis





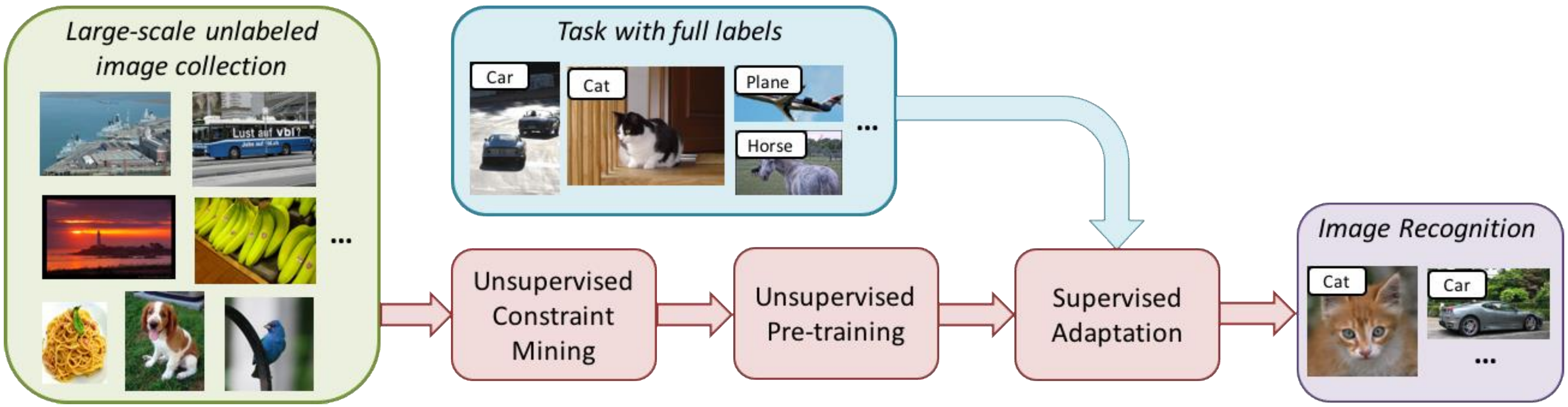
- Effect of different features

Features	LBP	HOG	SIFT+FV	Pre-trained CNN
Accuracy	76.7	80.7	80.9	81.6

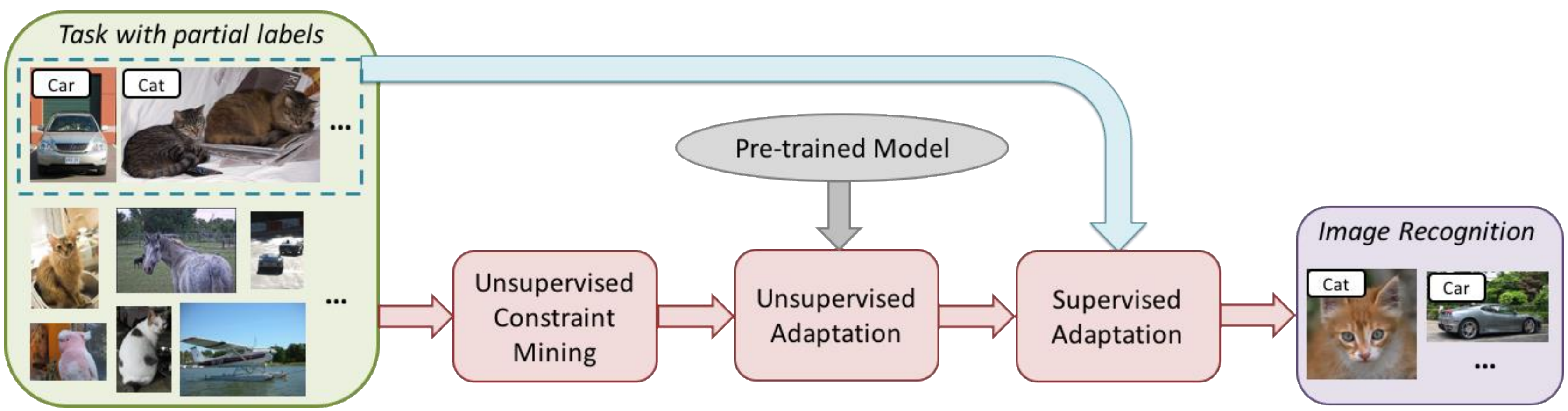
Cycle consistency works well on different hand-crafted features.

Applications

I. Unsupervised feature learning

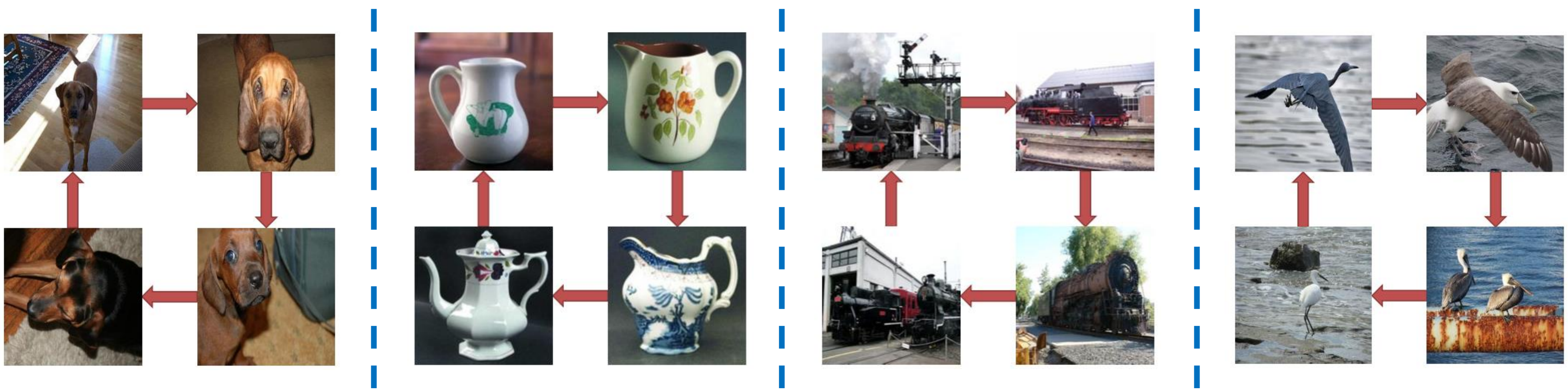


II. Semi-supervised learning



Unsupervised Feature Learning

- Implementation details
  - Dataset: ImageNet 2012 without any labels (~1.3M images)
  - Base features: SIFT+FV
  - Mining results: ~1M positives and ~13M negatives
- Cycle detection results



Positive pairs with large appearance variations

Qualitative evaluation - Search

Random

Supervised

Unsupervised

Random

Supervised

Unsupervised

Comparable to supervised learned representations

Qualitative evaluation - Search

Random

Supervised

Unsupervised

Random

Supervised

Unsupervised

Comparable to supervised learned representations

Quantitative Evaluation - Classification

Comparisons of image classification performance on VOC 2007

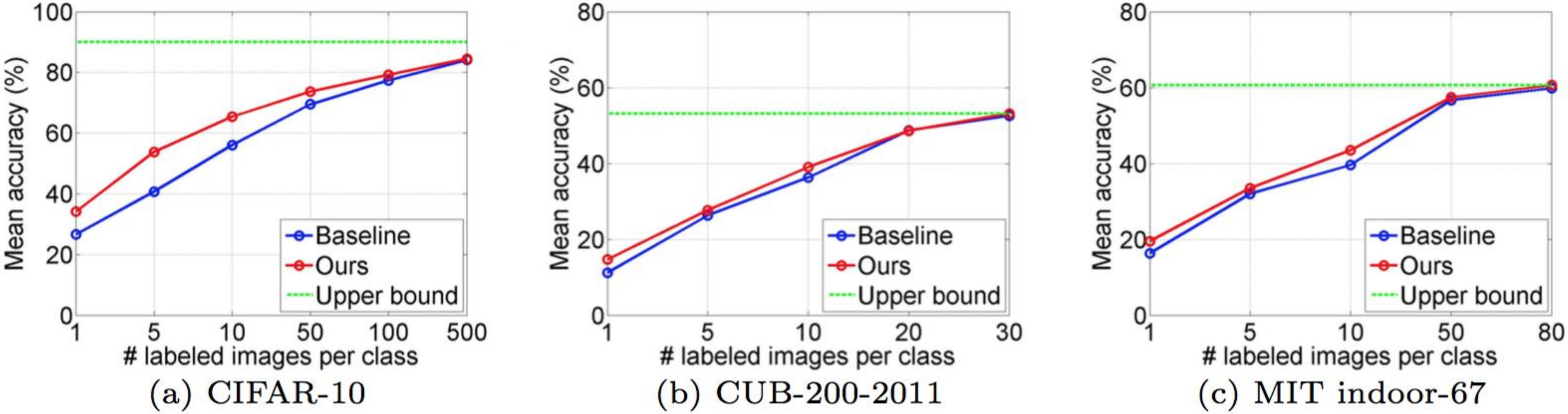
Methods	Supervision	Classification
Agrawal et al. ICCV'15	Ego-motion	52.9
Doersch et al. ICCV'15	Context	55.3
Wang et al. ICCV'15	Tracking triplet	58.4
Ours (SIFT+FV)	Matching pair	46.0
Ours (Learned features)	Matching pair	56.5
Krizhevsky et al. NIPS'12	Class labels	69.5

Competitive performance with the state-of-the-art

Significant improvement over hand-crafted features

Semi-supervised Learning

Classification results on three vision datasets



True positive rate on three vision datasets

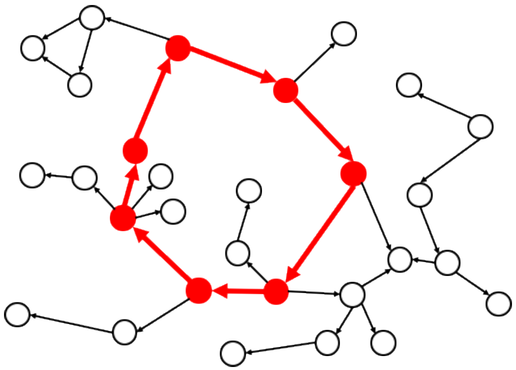
	CIFAR-10	CUB-200-2011	MIT indoor-67
Random sampling	10.0	0.5	1.5
4-cycle	83.0	55.8	65.8

Accurate positive pairs despite small inter-class differences

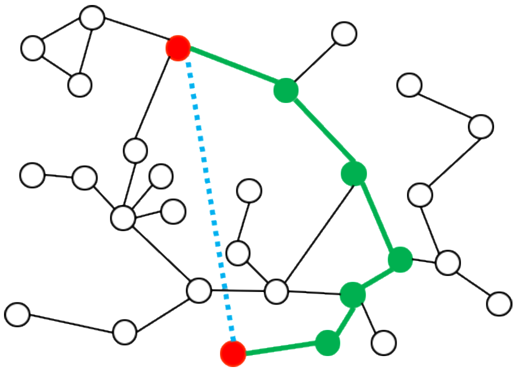
Conclusions

- **Unsupervised constraint mining**
  - Positive mining: positive pairs with large appearance variations
  - Negative mining: hard negative pairs with a relatively small *L*2 distance

Cycle consistency → Positive mining



Geodesic distance → Negative mining



- **Unsupervised feature leaning**
  - Image search: comparable to supervised learned representations
  - Image classification: competitive with the state-of-the-arts
- **Semi-supervised leaning**
  - Image classification: boosted performance over directly fine-tuning