Unsupervised Visual Representation Learning by Graph-Based Consistent Constraints

Dong Li[1], Wei-Chih Hung[2], Jia-Bin Huang[3], Shengjin Wang[1], Narendra Ahuja[3], Ming-Hsuan Yang[2]
[1]Tsinghua University, [2]University of California, Merced, [3]University of Illinois, Urbana-Champaign

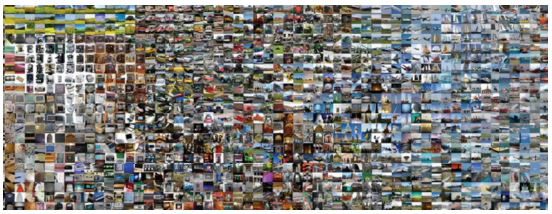
http://bit.ly/feature-learning-eccv2016

ECCV'16 EUROPEAN CONFERENCE ON COMPUTER VISION

Introduction

Supervised learning: Expensive annotations & poor scalability

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



Prior work:

Context [Doersch et al. ICCV'15]

Tracking [Wang and Gupta ICCV'15]

instance-level training data **within** the same image/video

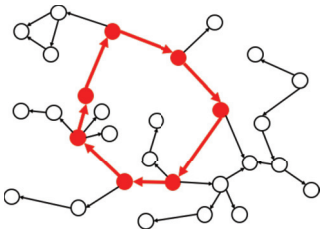
Contributions

- Generating semantically similar and dissimilar image pairs from a large, unlabeled image collection via an unsupervised constraint mining approach.
- Learning visual representations in an unsupervised manner via binary classification based on Siamese network.
- Improving the classification performance by applying the mined constraints for both unsupervised feature learning and semi-supervised learning.

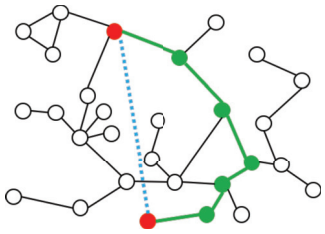
Unsupervised Visual Representation Learning

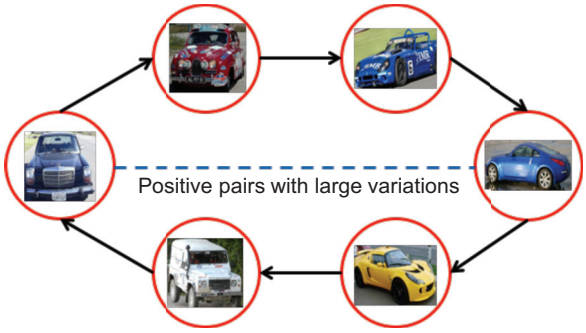
Mining: Generate **category-level** training samples **across** different images

Cycle consistency → Positive mining

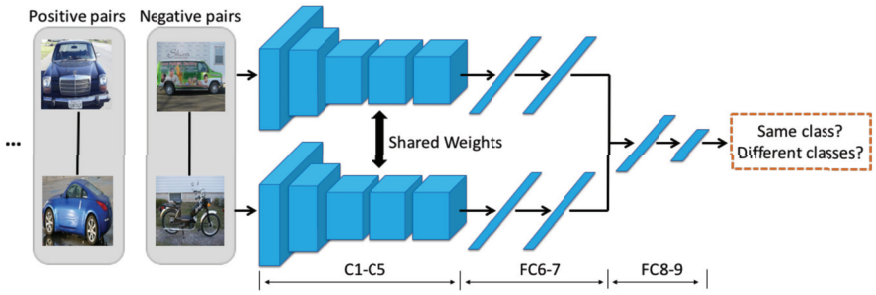


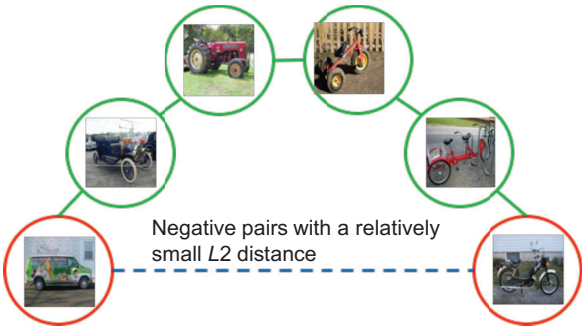
Geodesic distance → Negative mining





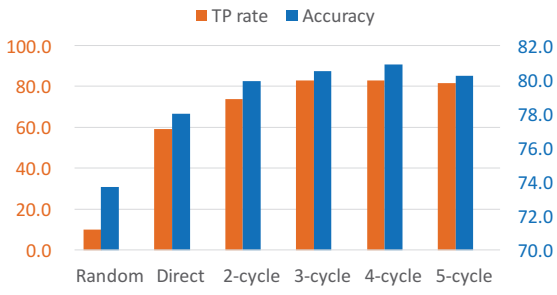
Training: Learn visual representations for **binary** classification





Experiments

Positive mining (CIFAR10)



Accurate positive pairs & Better CNN representations

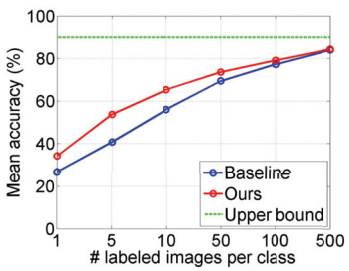
Image classification (VOC'07)

Methods	Supervision	mAP
Agrawal ICCV'15	Ego-motion	52.9
Doersch ICCV'15	Context	55.3
Wang ICCV'15	Tracking triplet	58.4
SIFT+FV	Matching pair	46.0
Ours	Matching pair	56.5
Krizhevsky NIPS'12	Class labels	69.5

Competitive performance with the state-of-the-arts

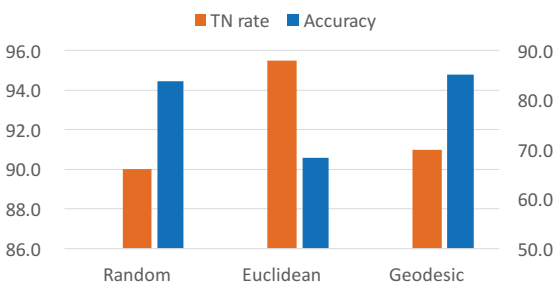
Significant improvement over hand-crafted features

Semi-supervised learning (CIFAR10)



Boosted performance over directly fine-tuning

Negative mining (CIFAR10)



Euclidean → *Easy samples*

Geodesic → *Hard samples*

Image search (ImageNet)

Comparable to supervised learned representations