Unsupervised Embedding Learning via Invariant and Spreading Instance Feature

Mang Ye [1], Xu Zhang [2], Pong C. Yuen [1], Shih-Fu Chang [2]

[1] Hong Kong Baptist University [2] Columbia University

香港浸會大學 HONG KONG BAPTIST UNIVERSITY COLUMBIA UNIVERSITY

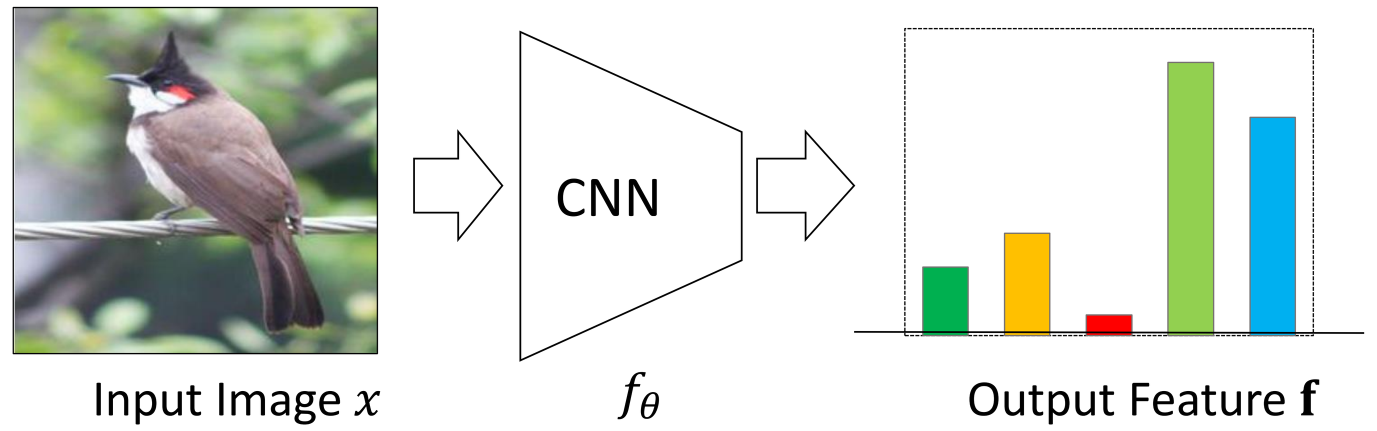
CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Code: https://github.com/mangye16/Unsupervised_Embedding_Learning

Introduction

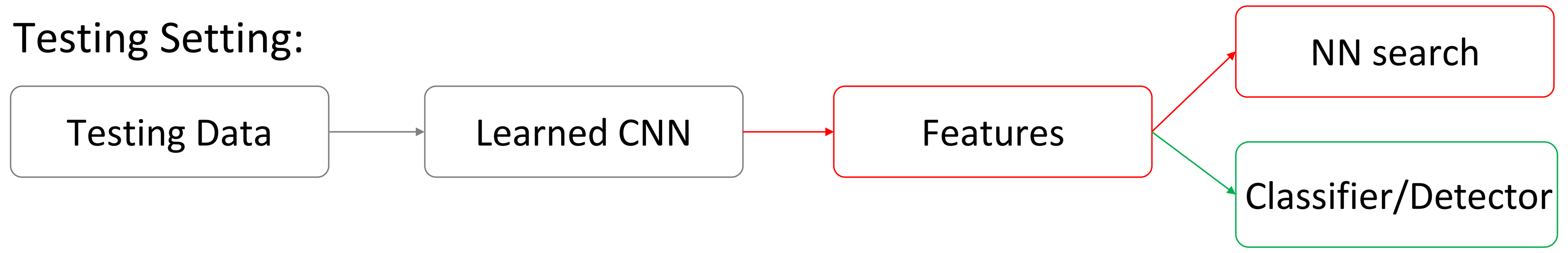
Unsupervised Embedding Learning:

- ❑ Learn a similarity-preserving embedding on a set of unlabeled images,
  - ✓ Similarity between learned features is consistent with the visual similarity



f_θ is the deep network parameters, $\mathbf{f}$ is the feature descriptor of input image x

Testing Setting:



Doesn't Require similarity consistency

Main Idea

Instance-wise Feature Learning

- ✓ **Augmentation Invariant**:
  - ❖ Use random data augmentation to generate the positive sample for each image instance
- ✓ **Instance Spread-out**:
  - ❖ Separate each instance from all the other sampled instances

Highlights

Data augmentation invariant and instance spread-out feature:

- ❑ Optimize the embedding directly on the real-time instance features with softmax function
- ❑ Achieve much faster learning speed and better accuracy than existing methods
- ❑ Perform well on both seen and unseen testing categories

Proposed Method

Exemplar CNN:

- ❑ Considers each instance as one category and learns a multi-class classifier for all samples.
  - ✓ Too many weights, weights and samples are not aligned.

NCE:

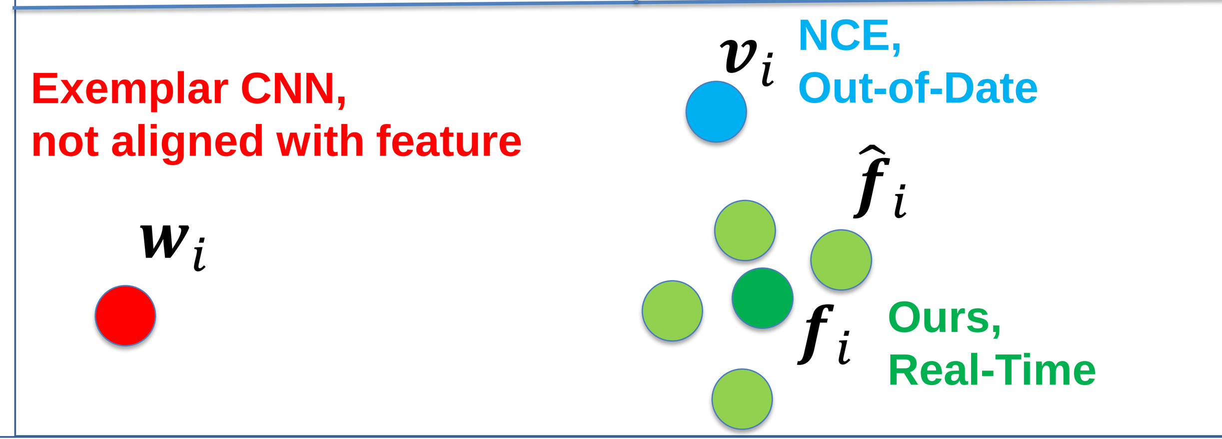
- ❑ Uses feature from last epoch as weights.
  - ✓ Stored features are out-of-date.

Exemplar CNN:

$$P(i|\mathbf{x}_j)=\frac{\exp(\mathbf{w}_i^T\mathbf{f}_j)}{\sum_{k=1}^{n}\exp(\mathbf{w}_k^T\mathbf{f}_j)}$$

NCE:

$$P(i|\mathbf{x}_j)=\frac{\exp(\mathbf{v}_i^T\mathbf{f}_j/\tau)}{\sum_{k=1}^{n}\exp(\mathbf{v}_k^T\mathbf{f}_j/\tau)}$$



Softmax Embedding on 'Real' Instance Feature

- ❑ We consider a small random sampled batch $\{\mathbf{x}_1,\ldots,\mathbf{x}_m\}$ rather than all instances
  - ➢ The augmented sample $\hat{\mathbf{x}}_i$ should be classified into instance i

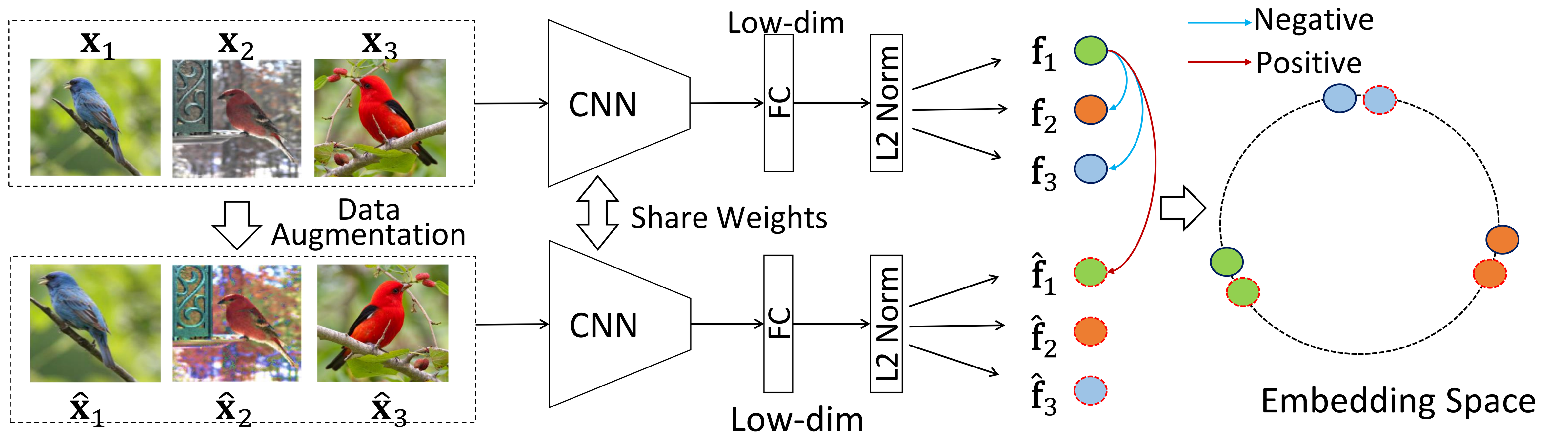
$$P(i|\hat{\mathbf{x}}_i)=\frac{\exp(\mathbf{f}_i^T\hat{\mathbf{f}}_i/\tau)}{\sum_{k=1}^{m}\exp(\mathbf{f}_k^T\hat{\mathbf{f}}_i/\tau)} \quad \uparrow \text{ Maximize}$$

  - ➢ The negative sample $\mathbf{x}_j$ should not be classified into instance i

$$P(i|\mathbf{x}_j)=\frac{\exp(\mathbf{f}_i^T\mathbf{f}_j/\tau)}{\sum_{k=1}^{m}\exp(\mathbf{f}_k^T\mathbf{f}_j/\tau)},\ j\neq i \quad \downarrow \text{ Minimize}$$

Siamese Network Training

- ❑ **Branch 1**: n randomly selected anchor instances
- ❑ **Branch 2**: Random data augmentation of the n anchor instances

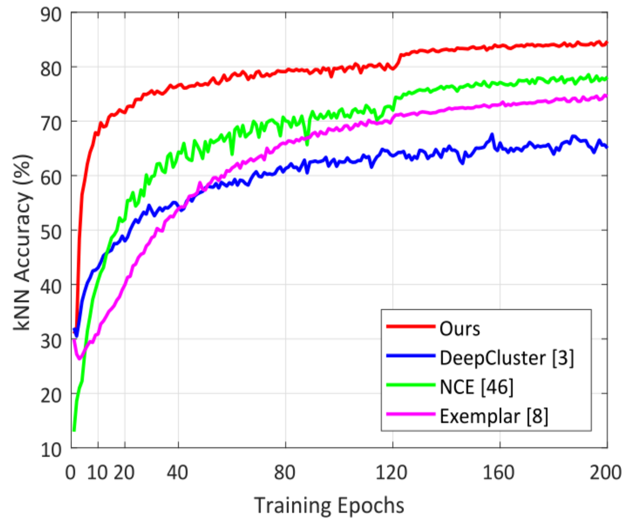


Experimental Results

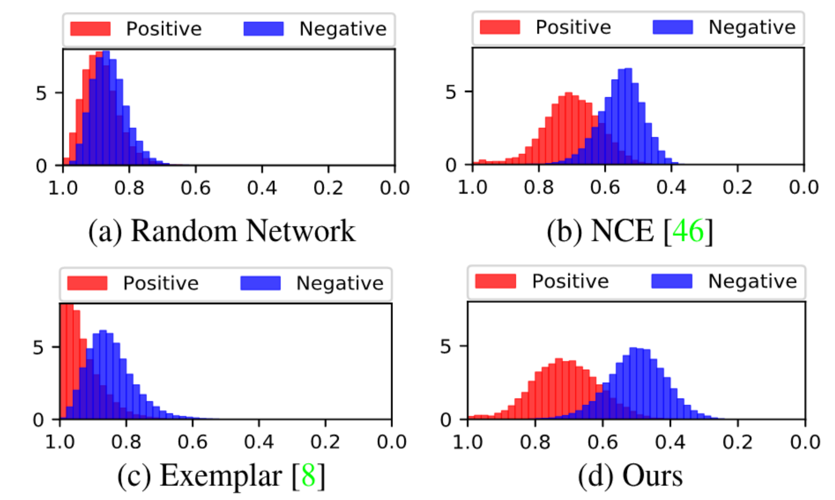
Seen Testing Category

CIFAR-10 dataset

- ❑ Training and testing set share the same categories



(a) Learning Speed



(b) Cosine similarity distributions

Unseen Testing Category

- ❑ Training and testing set do not have the same categories

Methods	R@1	R@2	R@4	R@8	NMI
Initial (FC)	39.2	52.1	66.1	78.2	51.4
Supervised Learning					
Lifted	43.6	56.6	68.6	79.6	56.5
Clustering	48.2	61.4	71.8	81.9	59.2
Triplet+	45.9	57.7	69.6	79.8	58.1
Smart+	49.8	62.3	74.1	83.3	59.9
Unsupervised Learning					
Cyclic	40.8	52.8	65.1	76.0	52.6
Exemplar	38.2	50.3	62.8	75.0	45.0
NCE	39.2	51.4	63.7	75.8	45.1
DeepCluster	42.9	54.1	65.6	76.2	53.0
MOM	45.3	57.8	68.6	78.4	55.0
Ours	**46.2**	**59.0**	**70.1**	**80.2**	**55.4**

(a) CUB200-2011 dataset

Methods	R@1	R@10	R@100	NMI
Initial (FC)	40.8	56.7	72.1	38.3
Exemplar	45.0	60.3	75.2	85.0
NCE	46.6	62.3	76.8	85.8
DeepCluster	34.6	52.6	66.8	82.8
MOM	43.3	57.2	73.2	84.4
Ours	**48.9**	**64.0**	**78.0**	**86.0**

(b) Stanford Product dataset

Methods	R@1	R@2	R@4	R@8	NMI
Initial (FC)	35.1	47.4	60.0	72.0	38.3
Exemplar	36.5	48.1	59.2	71.0	35.4
NCE	37.5	48.7	59.8	71.5	35.6
DeepCluster	32.6	43.8	57.0	69.5	38.5
MOM	35.5	48.2	60.6	72.4	**38.6**
Ours	**41.3**	**52.3**	**63.6**	**74.9**	35.8

(c) Car196 dataset

Acknowledgement:

This work is partially supported by Research Grants Council (RGC/HKBU12200518), Hong Kong. This work is partially supported by the United States Air Force Research Laboratory (AFRL) and the Defense Advanced Research Projects Agency (DARPA) under Contract No. FA8750-16-C-0166. Any opinions, findings and conclusions or recommendations expressed in this material are solely the responsibility of the authors and does not necessarily represent the official views of AFRL, DARPA, or the U.S. Government.