Progressive Cluster Purification for Unsupervised Feature Learning

Yifei Zhang[1,2*], Chang Liu[3*], Yu Zhou[1 †], Wei Wang[1,2], Weiping Wang[1] and Qixiang Ye[3 †]

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

[3]University of Chinese Academy of Sciences, Beijing, China

[*] Equal contribution, [†] Corresponding authors

Feature Learning

- Goal: Training CNNs with unlabeled dataset.
  - Feature representation of ‘immediate’ (convolutional) layers
  - Low dimension embedding feature representation

- Application:
  - Pre-trained feature as initialization
  - kNN classification, Image retrieval, clustering and so on

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Challenges

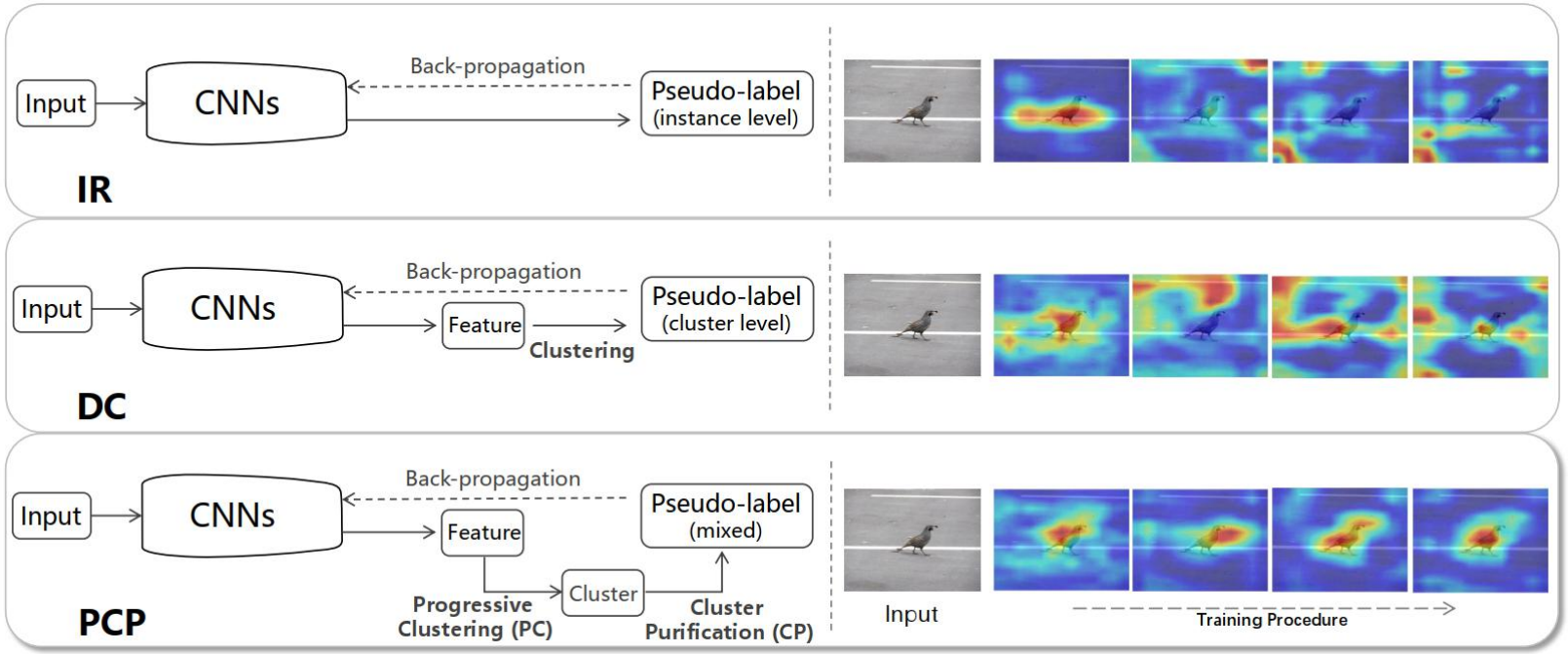
- How can we find **positive samples** for contrastive learning in unlabeled dataset?

$$P(i|v) = \frac{exp(v_i^T v/\tau)}{\sum_{j=1}^{n} exp(v_j^T v/\tau)}$$

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Motivation

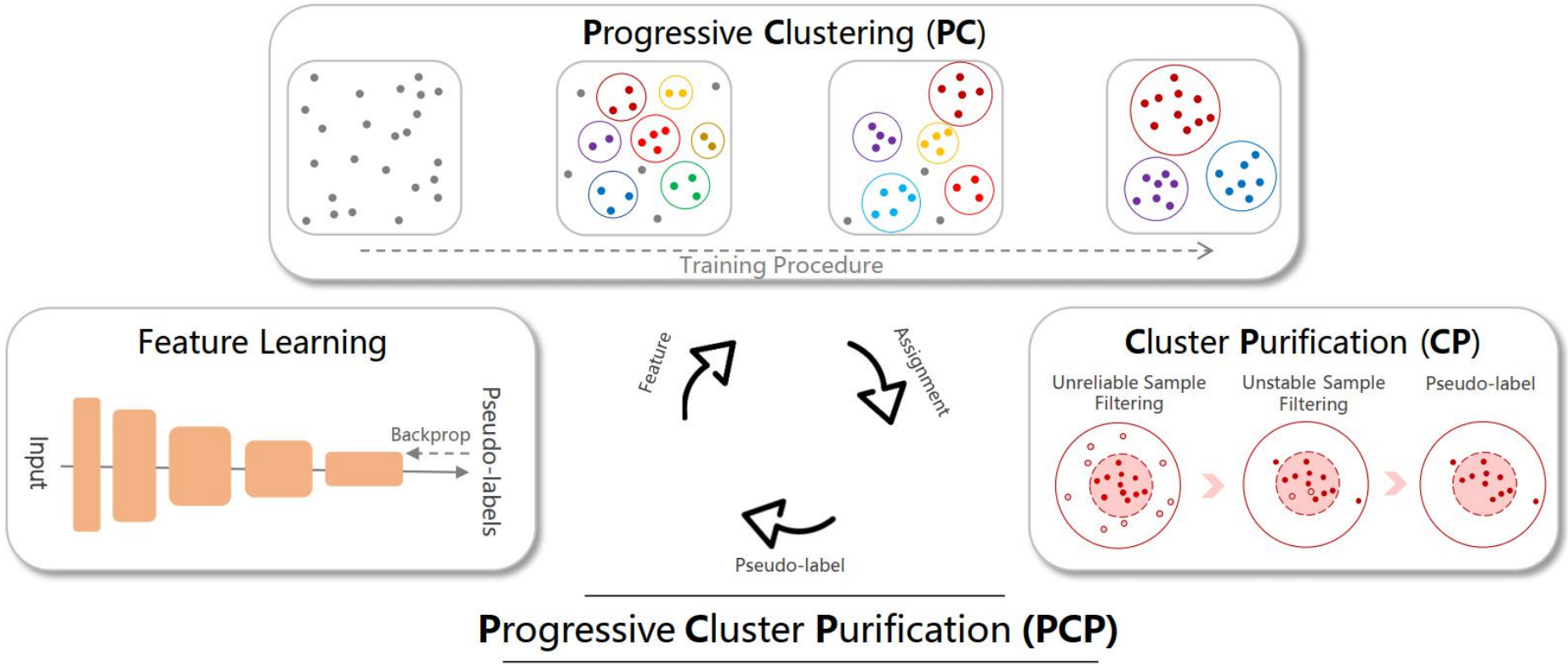
- **Instance based methods** ignore inter-image information;
- **Clustering based methods** exist class-inconsistent samples (noise);
- We need more **reliable positive samples**.



Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Overview

PCP consists three components, including Progressive Clustering, Cluster Purification and Feature Learning.



Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Progressive Clustering

- IR can learn class discriminative feature representation with solely **instance-level supervision**.
- We go one more step to infer that deep models can extract the underlying class information under different grain-level supervision **from instance-wise to class-wise**.
- As the total of samples N is tremendous, we implement a **linear declining strategy** on its logarithm to decide **the number of clusters**.

$$lg(N_t) = (1 - \frac{t}{T})lg(N)$$

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Cluster Purification

- To further make the cluster more reliable and stable to optimize the feature learning under supervision of **clear pseudo-labels** with **less noise**, we design a Cluster Purification mechanism which consists of unreliable sample filtering and unstable sample filtering .

- Based on the observation that samples **near cluster centroids** share higher apparent similarity, thus they are more likely to belong to the same class.

- We discard the samples far away from the centroids and temporarily regard each one as a distinct class in the subsequent learning procedure.

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Cluster Purification

- Easily distinguished samples are likely to be consistently assigned to the same cluster at **different iterations** of clustering.

- Inspired by this, we propose a **voting function** which utilizes the **previous clustering results** to quantitatively estimate the class consistency of samples.

$$V(v_i(t), v_{i_c}(t)) = \sum_{k=0}^{n} \alpha^k \cdot \delta(C(v_i(t-k)), C(v_{i_c}(t-k)))$$

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Feature Learning

- The probability of an input x being recognized as i-th example is shown in left.
- We develop our **instance-wise supervision** loss and **cluster-wise supervision** loss.

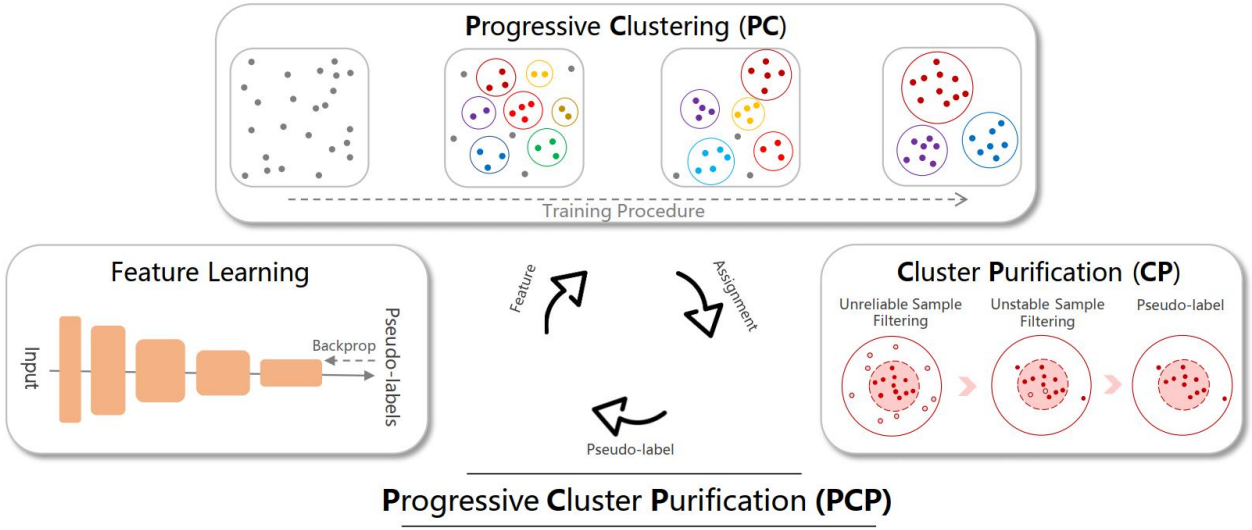
$$P(i|v) = \frac{exp(v_i^T v/\tau)}{\sum_{j=1}^{n} exp(v_j^T v/\tau)} \longrightarrow \begin{cases} L_{instance}^t = -\sum_{i \in \mathcal{N}^s(t)} log(P(i|v_i(t))) \\ L_{cluster}^t = -\sum_{c=1}^{N_t} \sum_{i,j \in \mathcal{S}_c^s(t)} log(P(i|v_j(t))) \end{cases} \longrightarrow L_{pcp}^t = L_{instance}^t + L_{cluster}^t$$

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Overview

Pipeline of PCP

Progressive Clustering → Cluster Purification
→ Feature Learning → Progressive Clustering
→ ……



Algorithm 1 Progressive Cluster Purification.

Input: An imagery dataset X without labels;
Output: CNN model f_θ with parameters θ ;

Preset embedding feature dimension D, training epochs T, cluster number N_{t_0} for stopping declining;
for epoch $t = 1$ **to** T **do**
  Get the number of clusters $N_t = max(N_t, N_{t_0})$ during the process of **PC**, Eq.(1);
  Obtain D-dimensional feature space $\mathcal{V}(t)$ by CNN model, $v(t) = f_{\theta_t}(x)$;
  Implement k-means clustering algorithm to get $\cup_c \mathcal{S}_c(t)$ with N_t clusters;
  for each cluster $c = 1$ **to** N_t **do**
    Split $\mathcal{S}_c(t)$ into class consistent set $\mathcal{S}_c^r(t)$ and noise set $\mathcal{N}_c^r(t)$ by $\mathbf{CP}_r$;
    Update class consistent set as $\mathcal{S}_c^s(t)$ and noise set as $\mathcal{N}_c^s(t)$ by $\mathbf{CP}_s$, Eq.(2);
  Calculate objective loss L_{pcp}^t (Eq.(6)) according to the union of set, $\cup_c \mathcal{S}_c^s(t)$ and $\cup_c \mathcal{N}_c^s(t)$;
  Feature learning by gradient back-propagation and updating model weights;
return f_θ.

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Experiments

- Experiment Setting

➢ Dataset

Cifar10, Cifar100, ImageNet100, CUB200, PASCAL VOC

➢ Evaluation Metrics

kNN, Linear Classification, Detection

Experiments

Components Analysis

TABLE I: Effects of the components in our approach with kNN classification accuracy.

DC [2]	PC	CP_r	CP_s	Acc
√	-	-	-	73.6
√	√	-	-	76.9
√	√	√	-	78.9
√	√	√	√	81.6

TABLE II: Evaluation of CP_r and CP_s under different filtering ratio.

γ	0	0.3	0.5	0.7	0.9	0.95	0.99
CP_r	76.9	77.7	78.9	79.8	81.4	81.6	81.7
CP_r+CP_s	80.3	81.5	81.6	81.3	81.1	80.9	80.6

TABLE III: Performance under different cluster numbers. * denotes with warm-up.

N_c	10k	5k	3k	1k	100	10	5
DC [2]	82.9	83.0	82.1	80.6	73.6	62.0	56.9
PCP	81.7	81.6	81.8	82.3	81.6	82.0	81.8
PCP*	84.1	84.4	84.3	84.7	83.9	83.1	83.2

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Experiments

Curriculum Learning

TABLE IV: Comparison of AND and PCP (w or w/o warm-up) under different training rounds (kNN accuracy). * denotes with warm-up. N denotes neighborhood size.

Round	0	1	2	3	4
AND [14] (N = 1)	80.0	83.9	85.0	85.9	86.3
AND [14] (N = 5)	80.0	83.8	85.0	85.6	85.8
AND [14] (N = 10)	80.0	83.8	84.9	85.1	85.1
AND [14] (N = 20)	80.0	83.3	84.3	84.6	84.6
PCP (Ours)	82.3	84.8	85.3	86.0	86.0
PCP* (Ours)	**84.7**	**86.4**	**86.7**	**87.0**	**87.3**

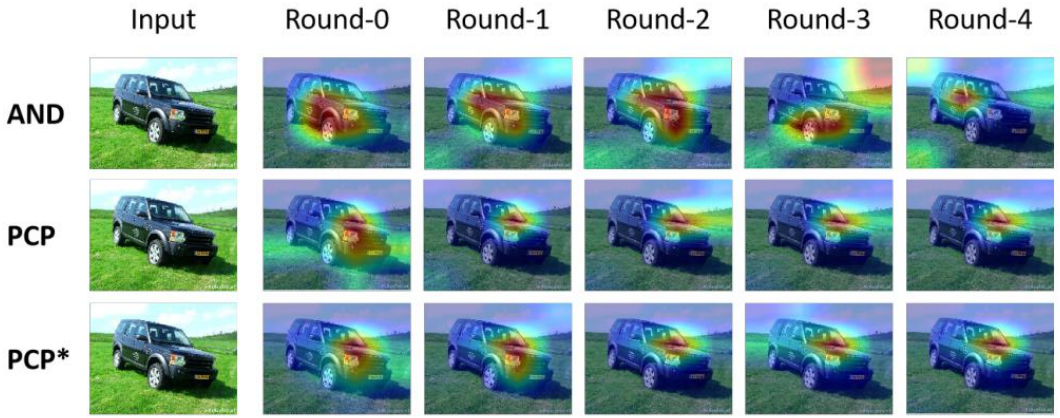


Fig. 4: Visualization of features extracted by AND and PCP, which clearly shows that the features learned by PCP steadily focus on objects. * denotes with warm up.

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Experiments

Comparison

Image Classification

TABLE V: Comparison of classification accuracy (kNN) on CIFAR10. F-S denotes fully supervised. * denotes the performance produced by our implementation. + denotes 5 rounds training. The compared results of IR/IS/AND are directly transcribed from their references.

Model	Random	F-S*	DC* [2]	IR [28]	IS [30]
Acc	32.1	93.1	80.6	80.8	83.6
Model	Random	AND [14]	PCP	AND$^+$	**PCP$^+$**
Acc	32.1	84.2	84.7	86.3	**87.3**

TABLE VI: Evaluation on CIFAR10 and CIFAR100 with AlexNet by performing linear classifier on the features from $conv5$, and kNN from FC. F-S denotes fully supervised. * denotes our rerunning. AND in our implementation has two rounds (one-off) while PCP with one round.

Classifier	Weighted kNN (FC)		Linear Classifier ($conv5$)	
Dataset	CIFAR10	CIFAR100	CIFAR10	CIFAR100
DC* [2]	70.3	27.4	77.1	44.0
IR* [28]	68.1	39.6	76.6	49.5
IS* [30]	76.4	46.3	78.7	51.2
AND* [14]	76.1	44.2	79.2	52.8
PCP (Ours)	**77.1**	**48.4**	**79.9**	**53.0**
F-S	91.9	69.7	91.8	71.0

TABLE VII: Comparison with MoCo [11] on ImageNet subset IN-100 with AlexNet by performing linear classifier(LC) on the features from $conv5$, and kNN from FC.

Classifier	kNN (FC)	LC ($conv5$)
MoCo [11]	50.1	55.4
PCP(Ours)	50.7	56.9

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Experiments

Comparison

Fine-grained dataset and initialization for **Object Detection**

TABLE VIII: Comparison of fine-grained classification performance. * denotes our rerunning. PCP is implemented with one round.

Model	Random*	IR [28]	DC* [2]	AND [14]	**PCP (Ours)**
Acc	2.6	11.6	13.1	14.1	**16.9**

TABLE IX: Comparison of object detection performance. AND in our implementation has two rounds (one-off) while PCP with one round.

Model	Random	DC [2]	IR [28]	AND [14]	**PCP (Ours)**	F-S
mAP	0.5	27.8	30.6	36.9	**39.8**	46.1

Experiments

Class conceptualization

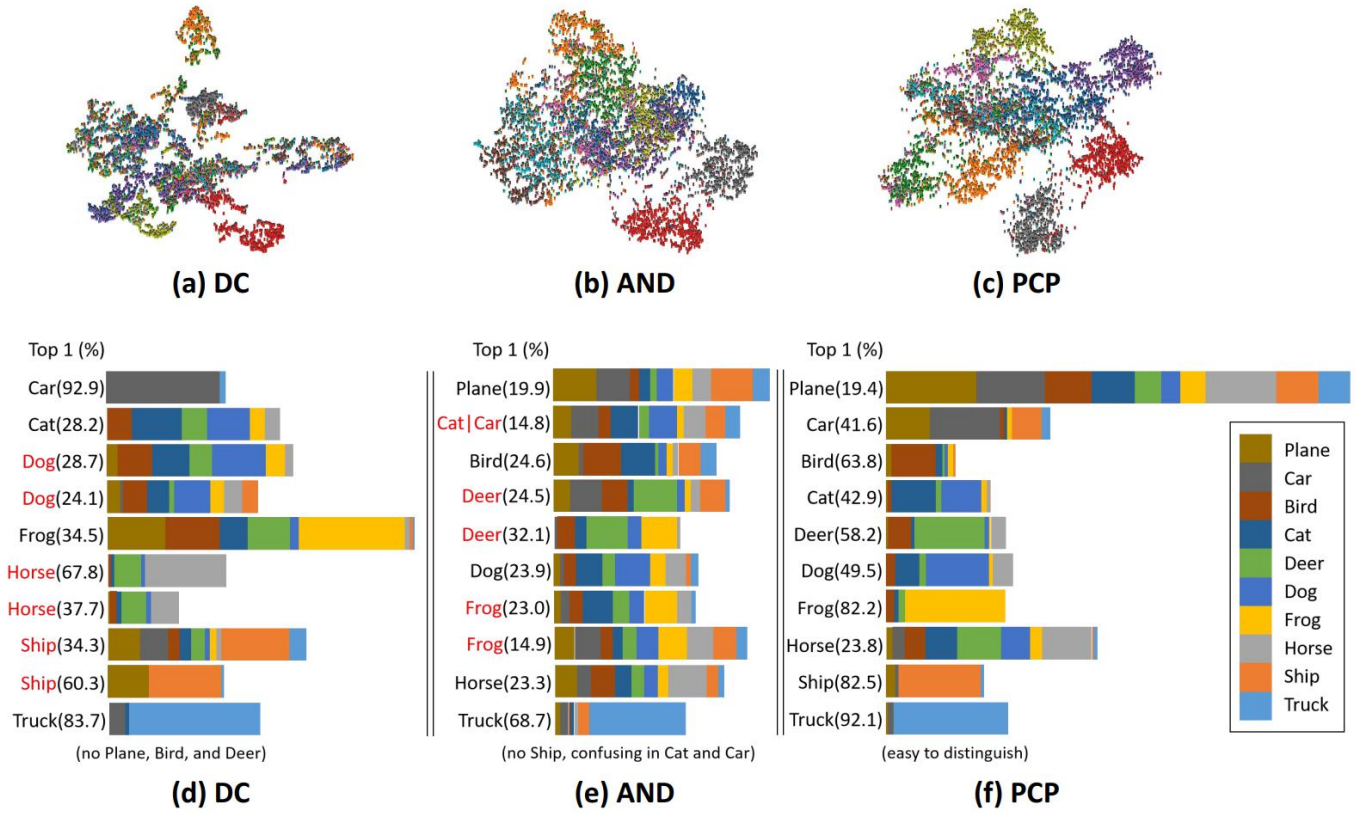


Fig. 5: Visualization of 2-dimensional t-SNE distributions of the feature space (a-c) and its class statistics under k-means (k=10) clustering results (d-f) by DC , AND and PCP. (Best viewed in color)

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Conclusion

To alleviate the impact of noise samples, we designed

- The Progressive Clustering (PC) strategy to gradually expand the cluster size consistently with the growth of the model representation capability.
- The Cluster Purification (CP) mechanism to reduce unreliable and unstable noise samples in each cluster to a significant extent.
- With warmup training strategy, PCP avoids network focusing on low-level feature for early clustering.

Extensive experiments on classification and detection benchmarks demonstrated that the proposed PCP approach has improved the classical clustering method and provided a fresh insight into the unsupervised learning problem.

Y. Zhang, C. Liu, Y. Zhou, W. Wang, W. Wang and Q. Ye, Progressive Cluster Purification for Unsupervised Feature Learning, 2020.

Thank you !