ICCV 2019

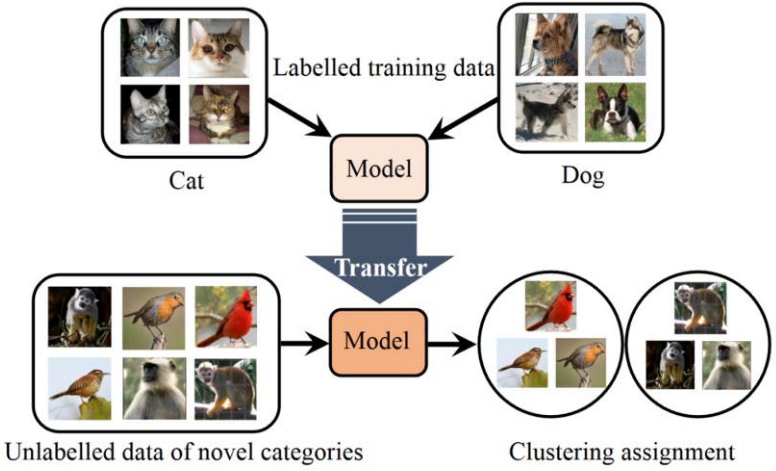
Learning to Discover Novel Visual Categories via Deep Transfer Clustering

Kai Han, Andrea Vedaldi, Andrew Zisserman

- **Problem/Objective**
  - Discover novel categories
- **Contribution/Key Idea**
  - Deep Transfer Clustering

전유진

- Task



$$D^l = \{(x_i^l, y_i^l), i = 1, \ldots, N\} \quad y_i^l \in \{1, \ldots, L\}$$

$$D^u = \{x_i^u, i = 1, \ldots, M\} \quad y_i^u \in \{1, \ldots, K\}$$

1. Deep embedded clustering

: representation learning (f_θ) & clustering ($U = \{\mu_k, k = 1, \ldots, K\}$)

$$z = f_\theta(x) \in \mathbb{R}^d$$

$$p(k|i) \propto \left(1 + \frac{\|z_i - \mu_k\|^2}{\alpha}\right)^{-\frac{\alpha+1}{2}}. \quad (1) \qquad i \in \{1, \ldots, N\} \ k \in \{1, \ldots, K\}$$

$$p(i,k) = p(k|i)/N \qquad (p(i) = 1/N)$$

결론, $q(i,k) = q(k|i)/N$ 와 $p(i,k) = p(k|i)/N$ 간의 KL divergence를 minimize.

$$E(q) = KL(q||p) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} q(k|i) \log \frac{q(k|i)}{p(k|i)}. \quad (2)$$

$$q(k|i) \propto p(k|i) \cdot p(i|k)$$

$$q(k|i) \propto \frac{p(k|i)^2}{\sum_{i=1}^{N} p(k|i)}. \quad (3)$$

전유진

Learning to Discover Novel Visual Categories via Deep Transfer Clustering

ICCV 2019

2. Transferring knowledge from known categories

Labeled set에서 pre-trained된 image representation f_θ

3. Bottleneck

$$\mathcal{Z}^u = \{z_i = f_\theta(x_i^u), i = 1, \ldots, M\} \qquad z_i \in \mathbb{R}^d$$

차원 축소 Layer : $\hat{z}_i = Az_i + b \qquad A \in \mathbb{R}^{K \times d}$

Figure 2. Representation visualization on CIFAR-10. Left: t-SNE projection on our learned features of unlabelled data (colored with GT labels); Middle: failure cases of clustering horses as dogs; Right: failure cases of clustering trucks as ships.

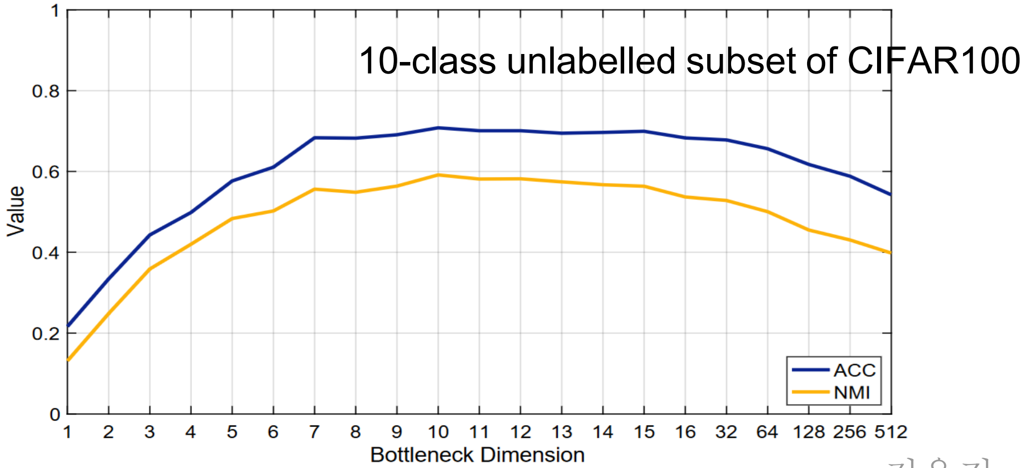


Figure 3. ACC and NMI w.r.t. different bottleneck dimensions.

전유진

4. Temporal ensembling

$$P^t(k|i) = \beta \cdot P^{t-1}(k|i) + (1-\beta) \cdot p^t(k|i), \qquad (4)$$

$$\tilde{p}^t(k|i) = \frac{1}{1-\beta^t} \cdot P^t(k|i). \qquad (5)$$

$$\tilde{q}^t(k|i)$$

$$q(k|i) \propto \frac{p(k|i)^2}{\sum_{i=1}^{N} p(k|i)}. \qquad (3)$$

$$E(q) = KL(q||p) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} q(k|i) \log \frac{q(k|i)}{p(k|i)}. \qquad (2)$$

5. Consistency

$$E(q) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} q(k|i) \log \frac{q(k|i)}{p(k|i)} + \omega(t) \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K} \|p(k|i) - p'(k|i)\|^2 \qquad (6)$$

6. How to predict # of unlabeled classes

$$D^l = \{(x_i^l, y_i^l), i = 1, \ldots, N\} \quad y_i^l \in \{1, \ldots, L\}$$

$$D^u = \{x_i^u, i = 1, \ldots, M\} \quad y_i^u \in \{1, \ldots, K\}$$

결론, $D_r^l \cup D^u$ 에서 전체 카테고리 수 C 변화시키며 constrained (semi-supervised) k-means clustering 여러번 실행하고, 각 C값에 대해 cluster quality 측정하여 최적의 클래스 개수 결정함.

$L - L_r$ $(\underline{D^l \setminus D_r^l})$: supervised feature representation learning

train set

L $(\underline{D^l})$

labeled set

(e.g., $L_r^a : L_r^v = 4 : 1$)

L_r $(\underline{D_r^l})$

probe set

L_r^a $(\underline{D_{ra}^l})$: gt label을 따라 cluster 되도록

anchor probe set

L_r^v $(\underline{D_{rv}^l})$: 추가적인 unlabeled data로 간주

validation probe set

전유진

- Experiment

Known scenario : OmniGlot, ImageNet, CIFAR-10, CIFAR-100, SVHN

Table 1. Visual category discovery (known number of categories).

	CIFAR-10		CIFAR-100		SVHN		OmniGlot		ImageNet	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
k-means [21]	65.5%	0.422	66.2%	0.555	42.6%	0.182	77.2%	0.888	71.9%	0.713
DTC-Baseline	74.9%	0.572	72.1%	0.630	57.6%	0.348	87.9%	0.933	**78.3%**	0.790
DTC-Π	**87.5%**	**0.735**	70.6%	0.605	**60.9%**	**0.419**	**89.0%**	**0.949**	76.7%	0.767
DTC-TE	82.8%	0.661	**72.8%**	**0.634**	55.8%	0.353	87.8%	0.931	78.2%	**0.791**
DTC-TEP	75.2%	0.591	72.5%	0.632	55.4%	0.329	87.8%	0.932	**78.3%**	**0.791**

Table 2. Results on OmniGlot and ImageNet with known number of categories.

	OmniGlot		ImageNet	
Method	ACC	NMI	ACC	NMI
k-means [21]	21.7%	0.353	71.9%	0.713
LPNMF [4]	22.2%	0.372	43.0%	0.526
LSC [8]	23.6%	0.376	73.3%	0.733
KCL [15]	82.4%	0.889	73.8%	0.750
MCL [16]	83.3%	0.897	74.4%	0.762
Centroid Networks [17]	86.6%	-	-	-
DTC	**89.0%**	**0.949**	**78.3%**	**0.791**

Table 3. Comparison with KCL and MCL on CIFAR-10/CIFAR-100/SVHN.

	CIFAR-10		CIFAR-100		SVHN	
	ACC	NMI	ACC	NMI	ACC	NMI
KCL [15]	66.5%	0.438	27.4%	0.151	21.4%	0.001
MCL [16]	64.2%	0.398	32.7%	0.202	38.6%	0.138
DTC	**87.5%**	**0.735**	**72.8%**	**0.634**	**60.9%**	**0.419**

전유진

- Experiment

Unknown scenario : OmniGlot, ImageNet, CIFAR-100

Table 4. Category number estimation results.

Data	GT	Ours	Error
OmniGlot	20-47	22-51	4.60
$\text{ImageNet}_{\text{A, B, C}}$	{30, 30, 30}	{34, 31, 32}	2.33
CIFAR-100	10	11	1

Table 5. Results on OmniGlot and ImageNet with unknown number of categories.

	OmniGlot		ImageNet	
Method	ACC	NMI	ACC	NMI
k-means [21]	18.9%	0.464	34.5%	0.671
LPNMF [4]	16.3%	0.498	21.8%	0.500
LSC [8]	18.0%	0.500	33.5%	0.655
KCL [15]	78.1%	0.874	65.2%	0.715
MCL [16]	80.2%	0.893	71.5%	0.765
DTC	**87.0%**	**0.945**	**77.6%**	**0.786**

Table 6. KCL and MCL with our category number estimation.

	OmniGlot		ImageNet	
	ACC	NMI	ACC	NMI
KCL [15]	78.1%	0.874	65.2%	0.715
KCL [15] w/ our k	80.3%	0.875	71.4%	0.740
MCL [16]	80.2%	0.893	71.5%	0.765
MCL [16] w/ our k	80.5%	0.879	72.9%	0.752
DTC	**87.0%**	**0.945**	**77.6%**	**0.786**

[Learning to Discover Novel Visual Ca](#)

ICCV 2019

- **Visualization**

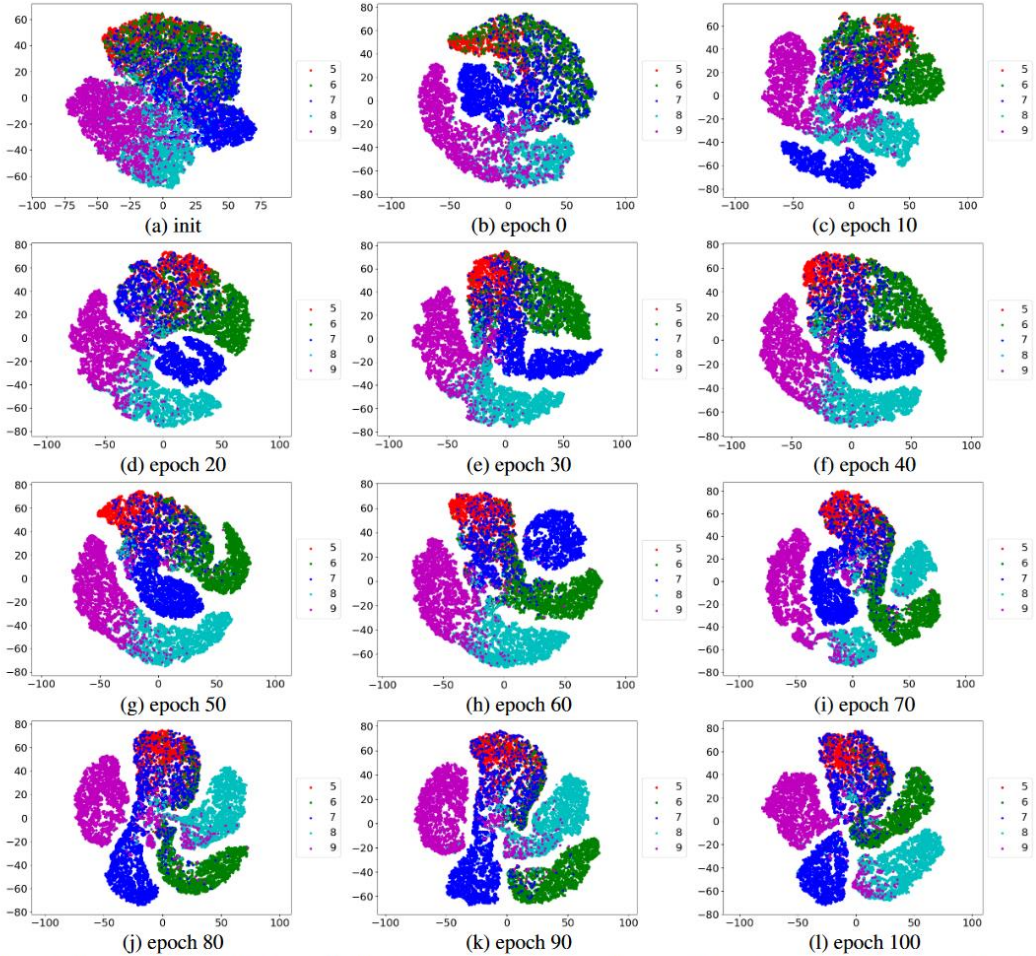


Figure 5. Representation visualization of unlabelled data (i.e., dog, frog, horse, ship, truck) by our deep transfer clustering model. 'init' means the feature obtained with the feature extractor trained on the labelled data (i.e., airplane, automobile, bird, cat, deer). Data points are colored according to their ground-truth labels.

전유진