Deep Adaptive Image Clustering

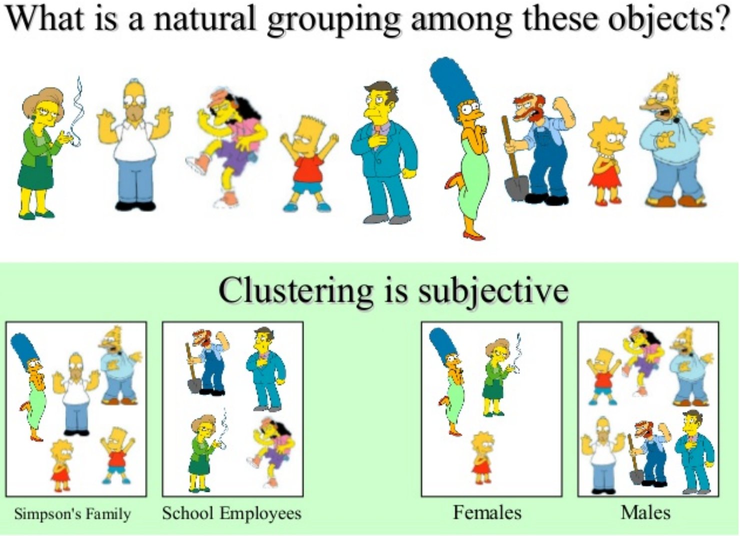
Jianlong Chang[1,2], Lingfeng Wang[1], Gaofeng Meng[1], Shiming Xiang[1,2], Chunhong Pan[1]

[1] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

[2] University of Chinese Academy of Sciences

Introduction

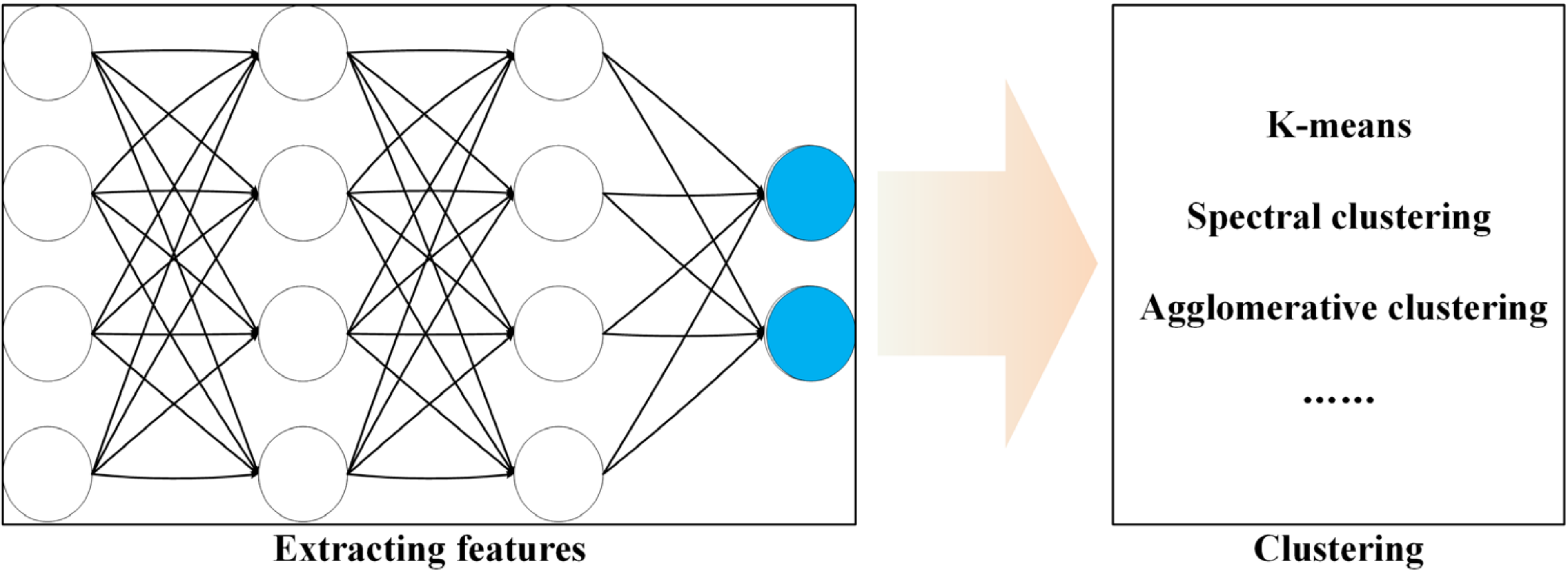
Image search
Image retrieval



Find potential customers
Consumer behavior research

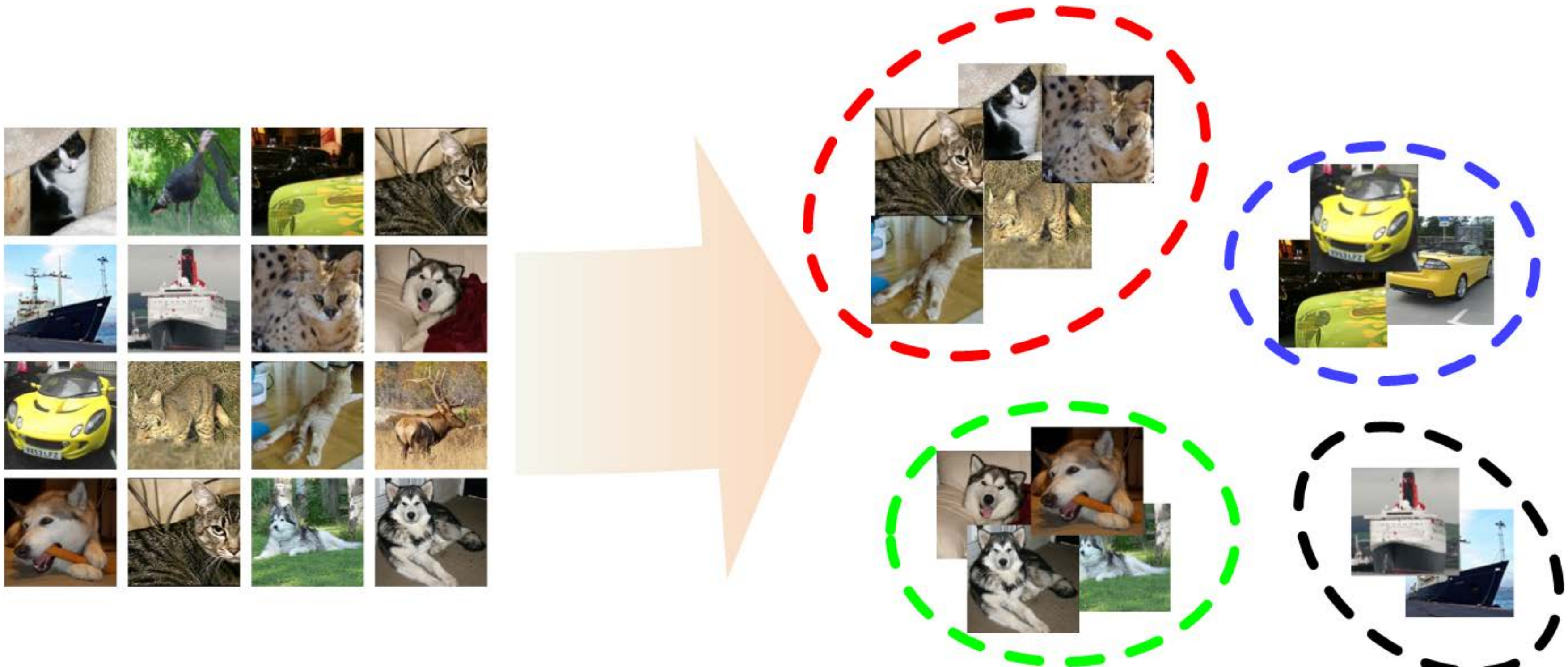
Related work

- **Multi-stage**
  - Extracting features(HoG, etc.) or learning features
  - Clustering by using the features
  - The learned features are **fixed**, the representations can not be further improved to obtain better performance.



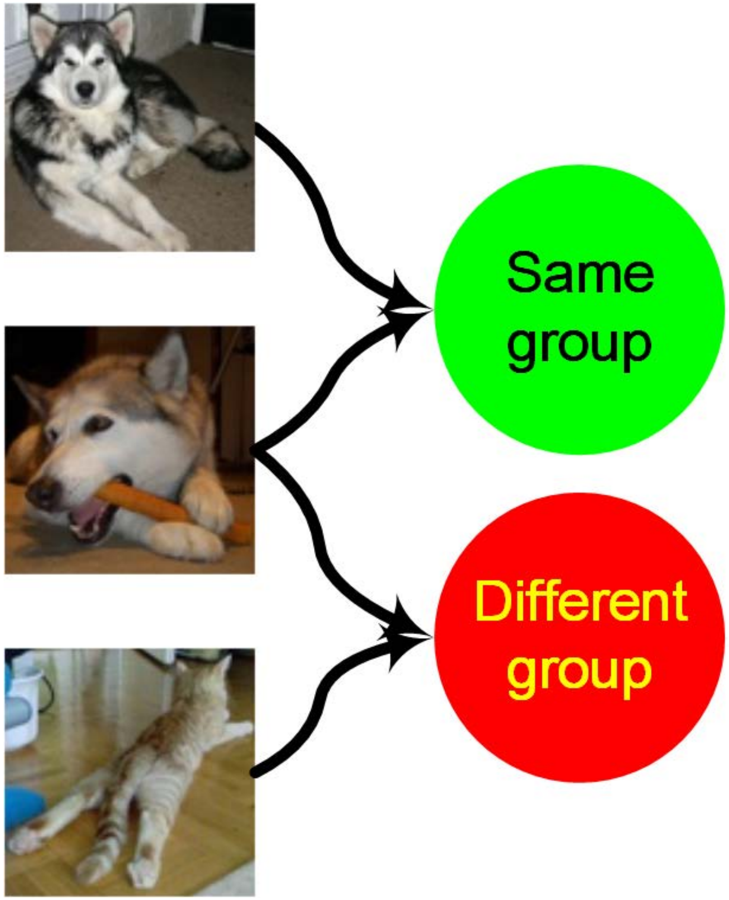
Definition

- **Clustering** is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. --from Wikipedia



DAC: Motivation

- **From the definition**
  - For two data points
    - **Same group**
    - **Different group**
  - Binary pairwise classification



DAC: Motivation

- **DAC model**

$$\min_{\mathbf{w}} \mathbf{E}(\mathbf{w}) = \sum_{i,j} L(r_{ij}, g(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}))$$

r_{ij} : the unknown binary variable (1:same cluster; 0:differernt cluster).

$g(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w})$: the estimated similarity.

- **Problems**
  - The clusters are unacquirable by only accessing to $g(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w})$
  - r_{ij} is unknown in clustering.

DAC: Label features

- **Clustering constraint**

$$g(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}) = f(\mathbf{x}_i; \mathbf{w}) \cdot f(\mathbf{x}_j; \mathbf{w}) = \mathbf{l}_i \cdot \mathbf{l}_j,$$

$$\boxed{\forall\, i,\ \| \mathbf{l}_i \|_2 = 1,\ \text{and}\ l_{ih} \geq 0,\ h = 1, \cdots, k,}$$

  - *k* is the predefined number of clusters.
  - *g(.,.)* represents the cosine distance.
  - *f* is a **CNN** model in our method.

- **DAC model**

$$\min_{\mathbf{w}} \mathbf{E}(\mathbf{w}) = \sum_{i,j} L(r_{ij}, \mathbf{l}_i \cdot \mathbf{l}_j), \tag{5}$$

$$\text{s.t.}\ \forall\, i,\ \| \mathbf{l}_i \|_2 = 1,\ \text{and}\ l_{ih} \geq 0,\ h = 1, \cdots, k.$$

DAC: Label features

- **We have**

THEOREM 1. *If the optimal value of Eq.* ($\boxed{5}$) *is attained, for* $\forall\ i,\ j,\ \mathbf{l}_i \in \mathbb{E}^k,\ \mathbf{l}_i \neq \mathbf{l}_j \Leftrightarrow r_{ij} = 0$ *and* $\mathbf{l}_i = \mathbf{l}_j \Leftrightarrow r_{ij} = 1$.

E^k : the standard basis of the k-dimensional Euclidean space

- Label features are ***k* diverse one-hot vectors** ideally.
- $\mathbf{l}_i \neq \mathbf{l}_j \Leftrightarrow r_{ij} = 0$ and $\mathbf{l}_i = \mathbf{l}_j \Leftrightarrow r_{ij} = 1$.
- Clustering based on the learned label features.

DAC: Similarity estimation

- **Selecting similar/dissimilar samples**

$$r_{ij} := \begin{cases} 1, & \text{if } \mathbf{l}_i \cdot \mathbf{l}_j \geq u(\lambda), \\ 0, & \text{if } \mathbf{l}_i \cdot \mathbf{l}_j < l(\lambda), \\ \text{None}, & \text{otherwise}, \end{cases} \quad i, \ j = 1, \cdots, n,$$

- **Curriculum learning (Self-paced Learning)**
  - $u(\lambda)$ is gradually decreased.
  - $l(\lambda)$ is gradually increased.
  - $u(\lambda) = l(\lambda)$: all the samples are used for training.

DAC: Model

- **DAC model**

Learn label feature **Select samples**

$$\min_{\mathbf{w},\lambda} \mathbf{E}(\mathbf{w},\lambda) = \boxed{\sum_{i,j} v_{ij} L(r_{ij}, \mathbf{l}_i \cdot \mathbf{l}_j)} + \boxed{u(\lambda) - l(\lambda)},$$

$$\text{s.t. } l(\lambda) \leq u(\lambda),$$

$$v_{ij} \in \{0,1\},\ i,\ j = 1, \cdots, n,$$

$$\forall\, i,\ \| \mathbf{l}_i \|_2 = 1, \text{ and } l_{ih} \geq 0,\ h = 1, \cdots, k,$$

$$r_{ij} := \begin{cases} 1, & \text{if } \mathbf{l}_i \cdot \mathbf{l}_j \geq u(\lambda), \\ 0, & \text{if } \mathbf{l}_i \cdot \mathbf{l}_j < l(\lambda), \\ \text{None}, & \text{otherwise}, \end{cases} \quad i,\ j = 1, \cdots, n,$$

where $\mathbf{v}$ is an indicator coefficient, *i.e.*,

$$v_{ij} := \begin{cases} 1, & \text{if } r_{ij} \in \{0,1\}, \\ 0, & \text{otherwise}, \end{cases} \quad i,\ j = 1, \cdots, n,$$

DAC: Algorithm

- **Clustering constraint**
  - A restraint layer is devised in CNN to learn label features

$$L_h^{out} := \exp^{L_h^{in} - \max_h(L_h^{in})},\ h = 1, \cdots, k,$$

$$L_h^{out} := \frac{L_h^{out}}{\| \mathbf{L}^{out} \|_2},\ h = 1, \cdots, k,$$

- **Alternating iterative optimization**

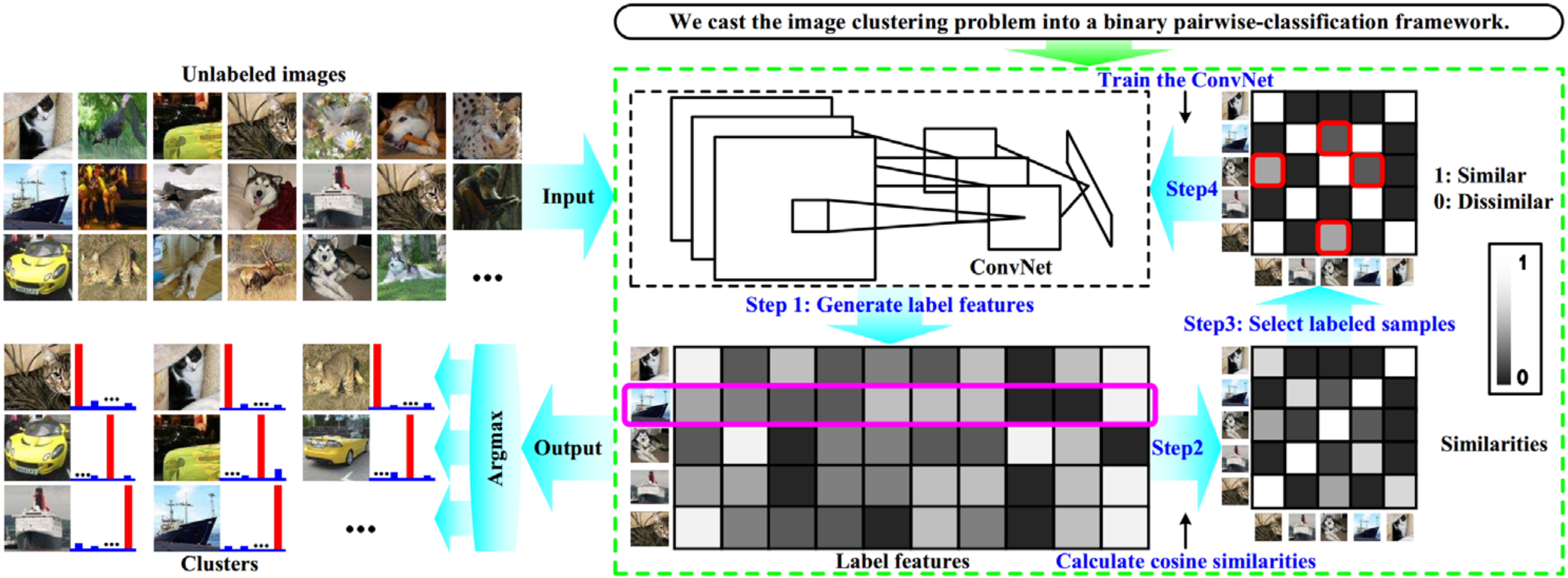
$$\text{fixing}: \lambda \Rightarrow \min_{\mathbf{w}} \mathbf{E}(\mathbf{w}) = \sum_{i,j} v_{ij} L(r_{ij}, f(\mathbf{x}_i; \mathbf{w}) \cdot f(\mathbf{x}_j; \mathbf{w}))$$

$$\text{fixing}: w \Rightarrow \min_{\lambda} \mathbf{E}(\lambda) = u(\lambda) - l(\lambda)$$

- **Clustering**

$$c_i := \arg\max_h (l_{ih}),\ h = 1, \cdots, k,$$

DAC: Flowchart



- Step 1 generates the label features of the samples by using a CNN.
- Step 2 calculates the cosine similarities based on the label features.
- Step 3 selects training samples according to the cosine similarities.
- Step 4 trains the CNN the binary pairwise-classification model.
- Iterate step 1 to step 4 until all the samples are considered.

DAC: Experiments

- **Datasets (5 image datasets)**

Table 1. The image datasets used in our experiments.

Dataset	Images	Clusters	Image size
MNIST [16]	70000	10	28×28
CIFAR-10 [14]	60000	10	$32 \times 32 \times 3$
CIFAR-100 [14]	60000	20	$32 \times 32 \times 3$
STL-10 [5]	13000	10	$96 \times 96 \times 3$
ImageNet-10 [7]	13000	10	$96 \times 96 \times 3$
ImageNet-Dog [7]	19500	15	$96 \times 96 \times 3$

DAC: Experiments

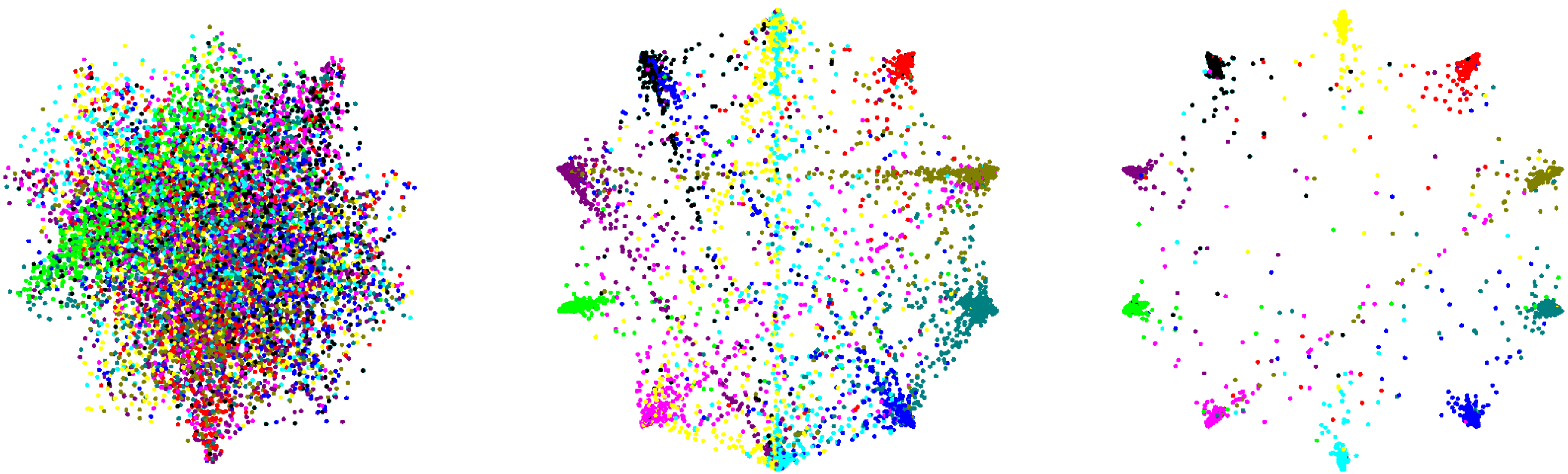
- **Compared methods (13 approaches)**

Table 2. The clustering results of various methods on six datasets. The best three results are highlighted in **bold**. DAC* represents that all the samples are considered for training in each iteration.

Dataset	MNIST [16]			CIFAR-10 [14]			CIFAR-100 [14]			STL-10 [5]			ImageNet-10 [7]			ImageNet-Dog [7]		
Metric	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC
K-means [32]	0.4997	0.3652	0.5723	0.0871	0.0487	0.2289	0.0839	0.0280	0.1297	0.1245	0.0608	0.1920	0.1186	0.0571	0.2409	0.0548	0.0204	0.1054
SC [40]	0.6626	0.5214	0.6958	0.1028	0.0853	0.2467	0.0901	0.0218	0.1360	0.0978	0.0479	0.1588	0.1511	0.0757	0.2740	0.0383	0.0133	0.1111
AC [9]	0.6094	0.4807	0.6953	0.1046	0.0646	0.2275	0.0979	0.0344	0.1378	0.2386	0.1402	0.3322	0.1383	0.0674	0.2420	0.0368	0.0207	0.1385
NMF [3]	0.6082	0.4298	0.5447	0.0814	0.0338	0.1895	0.0791	0.0263	0.1175	0.0962	0.0458	0.1804	0.1316	0.0652	0.2302	0.0442	0.0155	0.1184
AE [1]	0.7257	0.6139	0.8123	0.2393	0.1689	0.3135	0.1004	0.0476	0.1645	0.2496	0.1610	0.3030	0.2099	0.1516	0.3170	0.1039	0.0728	0.1851
SAE [18]	0.7565	0.6393	0.8271	0.2468	0.1555	0.2973	0.1090	0.0436	0.1567	0.2520	0.1605	0.3203	0.2122	0.1740	0.3254	0.1129	0.0729	0.1830
DAE [30]	0.7563	0.6467	0.8316	0.2506	0.1627	0.2971	0.1105	0.0460	0.1505	0.2242	0.1519	0.3022	0.2064	0.1376	0.3044	0.1043	0.0779	0.1903
DeCNN [39]	0.7577	0.6691	0.8179	0.2395	0.1736	0.2820	0.0923	0.0378	0.1327	0.2267	0.1621	0.2988	0.1856	0.1421	0.3130	0.0983	0.0732	0.1747
SWWAE [41]	0.7360	0.6518	0.8251	0.2330	0.1638	0.2840	0.1034	0.0391	0.1472	0.1962	0.1358	0.2704	0.1761	0.1603	0.3238	0.0936	0.0760	0.1585
AEVB [13]	0.7364	0.7129	0.8317	0.2451	0.1674	0.2908	0.1079	0.0403	0.1517	0.2004	0.1464	0.2815	0.1934	0.1683	0.3344	0.1074	0.0786	0.1788
GAN [21]	0.7637	0.7360	0.8279	**0.2646**	**0.1757**	**0.3152**	0.1200	0.0453	0.1510	0.2100	0.1390	0.2984	0.2250	0.1571	0.3459	0.1213	0.0776	0.1738
JULE [36]	**0.9130**	**0.9270**	**0.9640**	0.1923	0.1377	0.2715	0.1026	0.0327	0.1367	0.1815	0.1643	0.2769	0.1752	0.1382	0.3004	0.0537	0.0284	0.1377
DEC [35]	0.7716	0.7414	0.8430	0.2568	0.1607	0.3010	**0.1358**	**0.0495**	**0.1852**	**0.2760**	**0.1861**	**0.3590**	**0.2819**	**0.2031**	**0.3809**	**0.1216**	**0.0788**	**0.1949**
DAC*	**0.9246**	**0.9406**	**0.9660**	**0.3793**	**0.2802**	**0.4982**	**0.1623**	**0.0776**	**0.2189**	**0.3474**	**0.2351**	**0.4337**	**0.3693**	**0.2837**	**0.5026**	**0.1815**	**0.0953**	**0.2455**
DAC	**0.9351**	**0.9486**	**0.9775**	**0.3959**	**0.3059**	**0.5218**	**0.1852**	**0.0876**	**0.2375**	**0.3656**	**0.2565**	**0.4699**	**0.3944**	**0.3019**	**0.5272**	**0.2185**	**0.1105**	**0.2748**

DAC: Experiments

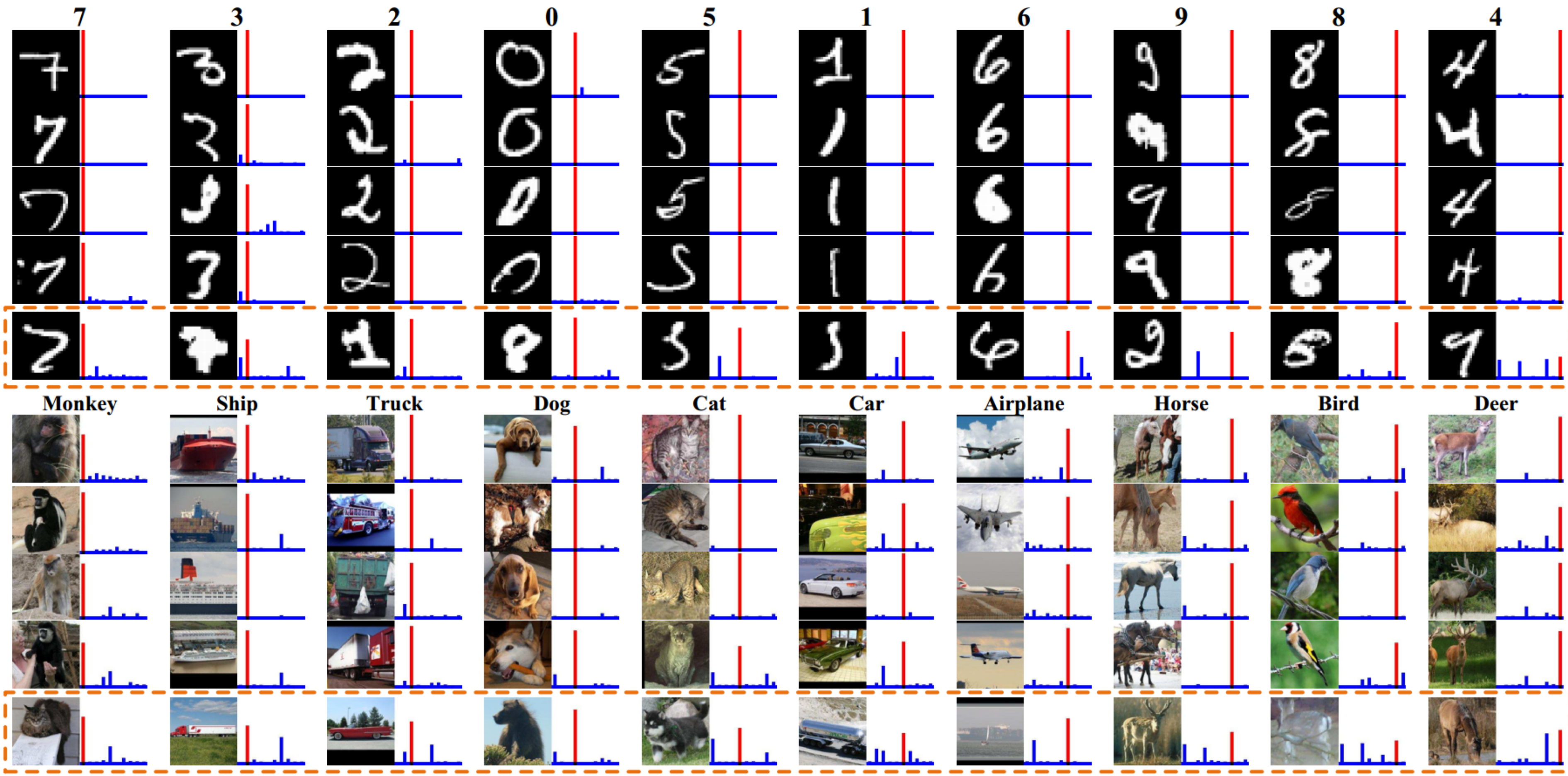
- **MNIST (10 classes)**



Initial stage Intermediate stage Final stage

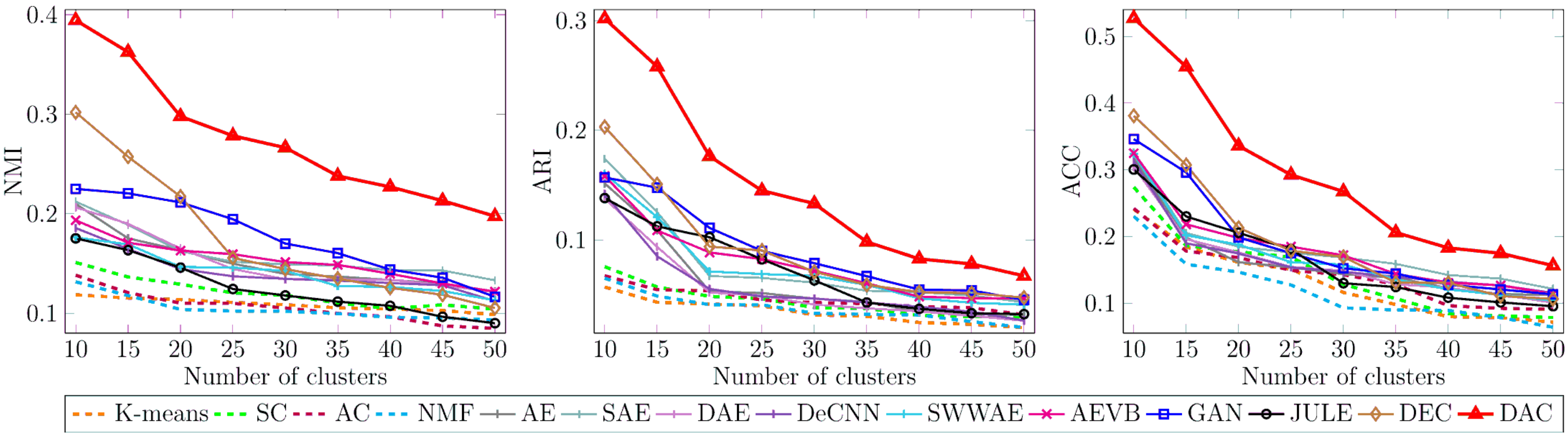
DAC: Experiments

- The label features learned by DAC
- MNIST/STL-10 datasets



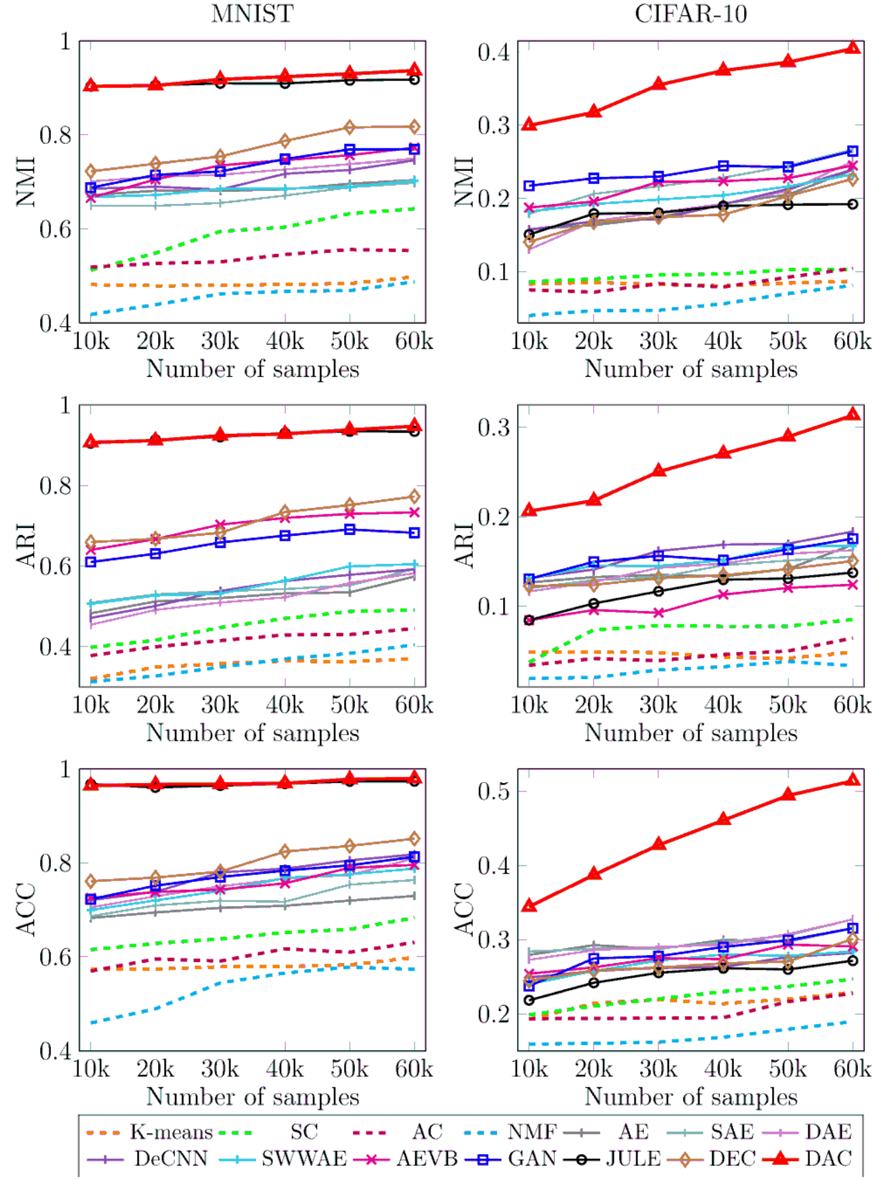
DAC: Experiments

- Various Number of Clusters on ImageNet (1300 images per clusters)



DAC: Experiments

- Various Number of Clusters on MNIST and CIFAR-10
  - The superiority of DAC holds with the various number of samples.



Conclusions

- A single-stage method for clustering images
- A binary constrained pairwise-classification model
- Features are one-hot vectors (effective and efficient)
- Relationships between data points is important

Future work

THEOREM 1. *If the optimal value of Eq. (5) is attained, for* $\forall\ i,\ j,\ \mathbf{l}_i \in \mathbb{E}^k,\ \mathbf{l}_i \neq \mathbf{l}_j \Leftrightarrow r_{ij} = 0$ *and* $\mathbf{l}_i = \mathbf{l}_j \Leftrightarrow r_{ij} = 1.$

- **Clustering constraint**

$$g(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}) = f(\mathbf{x}_i; \mathbf{w}) \cdot f(\mathbf{x}_j; \mathbf{w}) = \mathbf{l}_i \cdot \mathbf{l}_j,$$

$$\boxed{\forall\ i,\ \| \mathbf{l}_i \|_2 = 1,\ \text{and}\ l_{ih} \geq 0,\ h = 1, \cdots, k,}$$

  - Are there any other constraints?
  - Which one is the best constraint?
  - Why?

Thank you for your attention!

Any questions?