Deep Clustering by Semantic Contrastive Learning

Jiabo Huang
jiabo.huang@qmul.ac.uk
Queen Mary University of London, London, E1 4NS, UK

Shaogang Gong
s.gong@qmul.ac.uk

1 Deep clustering of Unlabelled Images

- Unlabelled representations: $f_0$
- Representations: $f_\Phi$
- Clusters: $k$

Semantic clusters:
Can be described by human-understandable words or phrases, e.g. class labels

One-to-one mappings
Elephant Python Fish

2 Related works & Limitations

(a) Sample-specific
(b) Neighbourhood
(c) Joint Clustering

(a) Insensitive to potential class memberships
(b) Restricted to subtle intra-class variations
(c) Instance contrast v.s. clustering - contradictory

3 Overview

(a) Start from randomly initial representations and decision boundaries
(b) Pulling each instance away from only its pseudo-negative samples of other clusters
(c) Simultaneously learning decision boundaries according to the distance-based cluster structure

4 Cross-cluster instance discrimination

- Semantic Memory $\mathcal{M} = \{M_1, M_2, ..., M_C\}$
  Maintaining $C$ independent memory banks with each corresponding to a target cluster
- Per-instance contrastive set $Q_i$
  For sample $I_i$ with a pseudo label $y_i$, its contrastive set $Q_i$ is composed of its pseudo negative samples
  $Q_i = \{k | k \in M_j \forall j \in [1, C] \text{ and } j \neq y_i\}$
- Instance discrimination objective
  To identify samples’ feature $q_i$ and their perturbed copies $k_i$ from their contrastive set
  $L_{ID}(I_i) = -\log\left(\frac{\exp(\cos(q_i, k_i)/\tau)}{\sum_{k \in Q_i \cup \{k_i\}} \exp(\cos(q_i, k)/\tau)}\right)$

5 Online cluster discrimination

- Distance-based cluster structure
  Taking samples in each semantic memory bank $M_i$ as the anchors of the corresponding cluster
  $\tilde{p}_{i,j} = \frac{\sum_{k \in M_j} \exp(\cos(q_i, k)/\tau)}{\sum_{k \in M_j} \exp(\cos(q_i, k)/\tau)}$
- Decision boundaries
  Modelling decision boundaries by an FC layer $p_i = \text{Softmax}(W^T q_i + B) \in \mathcal{R}^C$
- Cluster discrimination objective
  Updating boundaries to yield consistent $p_i$ with $\tilde{p}_i$
  $L_{CD}(I_i) = \sum_{j=1}^{C} -\tilde{p}_{i,j} \log p_{i,j}$

6 Experiments

Deep clustering

<table>
<thead>
<tr>
<th>Method</th>
<th>CIFAR10</th>
<th>CIFAR100</th>
<th>STLL10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCCM</td>
<td>0.327</td>
<td>0.383</td>
<td>0.108</td>
</tr>
<tr>
<td>PICA</td>
<td>0.337</td>
<td>0.352</td>
<td>0.098</td>
</tr>
<tr>
<td>GAT</td>
<td>0.281</td>
<td>0.322</td>
<td>-</td>
</tr>
<tr>
<td>CC+</td>
<td>0.429</td>
<td>0.429</td>
<td>0.140</td>
</tr>
<tr>
<td>GRLC+</td>
<td>0.425</td>
<td>0.484</td>
<td>-</td>
</tr>
<tr>
<td>GCC+</td>
<td>0.472</td>
<td>0.526</td>
<td>0.138</td>
</tr>
<tr>
<td>SCL+</td>
<td>0.482</td>
<td>0.763</td>
<td>0.172</td>
</tr>
</tbody>
</table>

IN*: ImageNet; Method*: w/ contrastive learning

Representation Learning

- MoCo: 0.528
- PAD: 0.526
- DeepCluster: 0.374
- SCL: 0.813

Feature visualisation

Visual case examples

Confident vs Unconfident