SELF-SUPERVISED LEARNING FOR FEW-SHOT IMAGE CLASSIFICATION

Da Chen, Yuefeng Chen, Yuhong Li, Feng Mao, Yuan He, Hui Xue

chen.cd, yuefeng.chenyf, daniel.lyh, maofeng.mf, heyuan.hy, hui.xueh@alibaba-inc.com

Motivation

- Difficulties
  - Limited labelled data for training
  - Novel classes during test comparing to training
  - Highly relied on the quality of pretrained on the backbone
- A robust few-shot learning method will benefit
- Tasks that have very limited data for training such as Medical Images
- Who cannot afford for very expensive annotation

Overall Architecture of Proposed Method

Our Contribution

- A simple framework that applies self-supervised learning to learn a very deep backbone for few-shot learning classification task
- State-of-the-art results in
  - 5-way 1-shot & 5-way 5-shot in MiniImageNet dataset[1]
  - 5-way 1-shot & 5-way 5-shot in CUB dataset[2]
  - Cross-domain few-shot learning task[3]
- Code is available at https://github.com/phecy/SSL-FEW-SHOT

Few-shot learning Pipeline in details

General Pipeline

- General pipeline for most of existing methods with good performance
- Pre-train the backbone on training set.
- Meta-learning based training with pretrained backbone on training set.
- Test the performance of the solution by training with limited data(1-shot or 5-shot) with novel classes(5-way) in test set and testing on query samples in these novel classes.

Self-supervised learning stage

The core is to maximize mutual information between global features and local features from two views \((x_k, x_m)\) of the same image. The NCE loss is defined as:

\[
\mathcal{L}_{NCE}(f(x_k), f(x_m)) = \frac{1}{N} \sum_{k=1}^{N} \log \left( \frac{\exp(\mathcal{D}(f(x_k)))}{\sum_{m \neq k} \exp(\mathcal{D}(f(x_m)))} \right)
\]

Meta-learning stage

The representation of class k is represented by the control of embedding features of training samples and can be obtained as:

\[
c_k = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} f(s)
\]

Datasets

- MiniImageNet[1]: 60,000 images from 100 classes, 64 classes for training, 16 classes for validation, 20 classes for the test.
- CUB[2]: 11788 images from 200 classes, 100 classes for training, 50 classes for validation, and 50 classes for the test.

Comparison to the state of the art:

<table>
<thead>
<tr>
<th>Method</th>
<th>Embedding Net</th>
<th>Few-shot 1-way</th>
<th>Few-shot 5-way</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAML[11]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>WANDA[12]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>Alpago[13]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>MBIL[14]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>MAE[15]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>FeRAM[16]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>SuperMAML[17]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>PanAF[18]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>MetaSAN[19]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>CBF[20]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>VA-MAE[21]</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>FEW-SHOT</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
<tr>
<td>SELF-SUPERVISED LEARNING FOR FEW-SHOT IMAGE CLASSIFICATION</td>
<td>4-Corn</td>
<td>4-Way</td>
<td>4-Way</td>
</tr>
</tbody>
</table>

Our method obtains remarkable improvement.

Acknowledgement

The authors are supported by Alibaba Group

Reference