A Self-supervised GAN for Unsupervised Few-shot Object Recognition

Khoi Nguyen  Sinisa Todorovic

Paper ID: 970
Problem Statement

Training

Unlabeled images

Testing

A few labeled support images

Images from Imagenet dataset: http://image-net.org
Challenges

Our setting differs from standard few-shot classification which requires a large set of labeled examples for the episodic training.

Key challenges:

• Learn a deep image representation on unlabeled data such that it would generalize well to unseen classes in testing

• Estimate an accurate similarity between query and support images
Applications

• In domains with very few labeled images per class:
  • Medical images
  • Unsupervised video object segmentation
  • Online tracking

Prior Work

• Generate pseudo labels for training set and apply episodic training as in standard few-shot classification:
  • Image clustering (Hsu et al., ICLR 2019)
  • Data augmentation where each training example is a unique class (Khodadadeh et al., NeurIPS 2019)

➤ Limitations of pseudo labels:
  • Are noisy and not reliable for learning a distance based representation
  • Do not respect the semantic similarity relationship between images
Proposed Unsupervised Training

Vanilla GAN

Real image \( x \sim \text{Data} \)

Generated image \( x' = G(z') \)

Sampled latent code \( z' \)

Discriminator

CNN

Discriminating head \( D_{r/f} \)

Encoding head \( D_z \)

"real"

"fake"

Contribution 1: Reconstruction loss between \( z' \) and \( \hat{z}' \)

Contribution 2: Triplet loss
Image Masking and Triplet Loss

Triplet loss \((z, z^+, z^-)\)
Few-shot Classification Testing

Support images
- Elephant
- Cat
- Boat
- Dog
- Airplane

Support representations

Query image

Query representation

Cosine similarities
- 0.4
- 0.9
- 0.1
- 0.6
- 0.2

$\text{cos}(\cdot, \cdot)$
Experiments

• Datasets:
  • *Mini-Imagenet*: 100 classes: train: 64, validation: 16, and test: 20 classes
  • *Tiered-Imagenet*: 608 classes grouped into 34 high-level categories: train: 20, validation: 6, and test: 8 categories (minimize semantic similarity between the splits)

• Evaluation metrics:
  • Average $N$-way $K$-shot classification accuracy with 95% confidence interval over 1000 episodes where $N$ is the number of classes and $K$ is the number of labeled examples for each class
## Few-shot Classification Results

<table>
<thead>
<tr>
<th>Unsupervised Methods</th>
<th>Mini-Imagenet, 5-way</th>
<th>Tiered-Imagenet, 5-way</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-shot</td>
<td>5-shot</td>
</tr>
<tr>
<td>SN-GAN (Miyato et al., ICLR 2018)</td>
<td>34.84 ± 0.68</td>
<td>44.73 ± 0.67</td>
</tr>
<tr>
<td>AutoEncoder (Vincent et al., JMLR 2010)</td>
<td>28.69 ± 0.38</td>
<td>34.73 ± 0.63</td>
</tr>
<tr>
<td>Rotation (Gidaris et al., ICLR 2018)</td>
<td>35.54 ± 0.47</td>
<td>45.93 ± 0.62</td>
</tr>
<tr>
<td>BiGAN kNN (Donahue et al., ICLR 2017)</td>
<td>25.56 ± 1.08</td>
<td>31.10 ± 0.63</td>
</tr>
<tr>
<td>AAL-ProtoNets (Antonious et al., Arxiv 2019)</td>
<td>37.67 ± 0.39</td>
<td>40.29 ± 0.68</td>
</tr>
<tr>
<td>CACTUs-ProtoNets (Hsu et al., ICLR 2019)</td>
<td>39.18 ± 0.71</td>
<td>53.36 ± 0.70</td>
</tr>
<tr>
<td><strong>Our GdBT2</strong></td>
<td><strong>48.28 ± 0.77</strong></td>
<td><strong>66.06 ± 0.70</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fully-supervised Methods</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ProtoNets (Snell et al., NeurIPS 2017)</td>
<td>46.56 ± 0.76</td>
</tr>
</tbody>
</table>
Conclusions

• We have addressed unsupervised few-shot classification
• Our approach extends the vanilla GAN to integrate adversarial learning with two new strategies for self-supervised learning:
  • Latent code reconstruction loss
  • Image masking with triplet loss
• Outperform SOTAs with significant margin (9% on Mini-Imagenet)
• Beat a fully-supervised few-shot classification approach (ProtoNet)!