IN DEFENSE OF PSEUDO-LABELING: AN UNCERTAINTY-AWARE PSEUDO-LABEL SELECTION FRAMEWORK FOR SEMI-SUPERVISED LEARNING
Semi-Supervised Learning

- Supervised learning relies on large labeled datasets
- Constructing large labeled datasets
  - expensive
  - time-consuming
- SSL leverages
  - a small amount of labeled data
  - a large amount of unlabeled data concurrently
- One of the fundamental problems in machine learning
Dominant SSL Approaches

- **Consistency Regularization** [1, 2, 3]
  - obtain perturbation/augmentation invariant output distribution
  - rely on domain-specific heavy data augmentation
  - limited applicability on domains which do not have a rich set of augmentations

2. *Interpolation consistency training for semi-supervised learning*; Verma et al.; IJCAI 2019
Dominant SSL Approaches

- Pseudo-Labeling [1]
  - generate pseudo-labels for unlabeled samples
  - does not require domain-specific data augmentation
  - performs poorly in comparison to consistency regularization

1. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks; Lee; ICML Workshop 2013
Objective

Bridge the performance gap between

Pseudo-Labeling

and

Consistency Regularization
Fundamental Issues with Pseudo-Labeling

● Training with small labeled set
  ○ leads to erroneous pseudo-label generation
  ○ many incorrect pseudo-labels leads to noisy training

● Incorrect pseudo-labels must be discarded
  ○ use high-confidence pseudo-labels for training
Pseudo-Label Selection

- Using highly confident pseudo-labels is insufficient
  - neural networks suffer from poor calibration [1]
  - many incorrect pseudo-labels are still selected

- Another interpretation of calibration
  - a notion of network’s overall prediction uncertainty [2]

- We have empirically analyzed
  - the relationship between the calibration error and
  - individual output prediction uncertainties

1. On Calibration of Modern Neural Networks; Guo et al.; ICML 2017
2. Simple and scalable predictive uncertainty estimation using deep ensembles; Lakshminarayanan et al.; Neurips 2017
Pseudo-Label Selection

- Based on our observations
  - we select a subset of generated pseudo-labels with the following two criteria
    - the confidence of a prediction has to be high
    - the network has to be certain about the output prediction
- We call this method Uncertainty-Aware Pseudo-Label Selection (UPS)
Pseudo-Label Selection

- Networks can be confident and certain about a class *not* being present
- We can use this information and select negative pseudo-labels
- This gives us two benefits:
  - negative learning for single label classification
  - allows for multi-label classification
UPS Method

Training

Labeled (+Unlabeled) Data

\(f_0\)

\(\mathcal{L}_{CE} (+\mathcal{L}_{NCE})\)

Pseudo-Label Selection

\(p_{c} > \tau \) and \(u(p_{c}) < \xi\)

Pseudo-Label Generation

Unlabeled Data

\(f_0\)

\(p_{c} \geq \gamma\)

Pseudo-Labels

<table>
<thead>
<tr>
<th>(p_{c})</th>
<th>(u(p_{c}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.01</td>
</tr>
<tr>
<td>0.83</td>
<td>0.05</td>
</tr>
<tr>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>0.87</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Car

Bird

Ship
Results (CIFAR-10 and CIFAR-100)

Error rate (%) on the CIFAR-10 and CIFAR-100 test set:

<table>
<thead>
<tr>
<th>Method</th>
<th>CIFAR-10 1000 labels</th>
<th>CIFAR-10 4000 labels</th>
<th>CIFAR-100 4000 labels</th>
<th>CIFAR-100 10000 labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPS†</td>
<td>8.18 ± 0.15</td>
<td>6.39 ± 0.02</td>
<td>40.77 ± 0.10</td>
<td>32.00 ± 0.49</td>
</tr>
<tr>
<td>MixMatch</td>
<td>-</td>
<td>6.84</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R2-D2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>32.87 ± 0.51</td>
</tr>
<tr>
<td>DualStudent</td>
<td>14.17 ± 0.38</td>
<td>8.89 ± 0.09</td>
<td>-</td>
<td>32.77 ± 0.24</td>
</tr>
<tr>
<td>ICT</td>
<td>15.48 ± 0.78</td>
<td>7.29 ± 0.02</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MT + DeepLP</td>
<td>16.93 ± 0.70</td>
<td>10.61 ± 0.28</td>
<td>43.73 ± 0.20</td>
<td>35.92 ± 0.47</td>
</tr>
<tr>
<td>MT</td>
<td>19.04 ± 0.51</td>
<td>11.41 ± 0.25</td>
<td>45.36 ± 0.49</td>
<td>36.08 ± 0.51</td>
</tr>
<tr>
<td>TSSDL†</td>
<td>21.13 ± 1.17</td>
<td>10.90 ± 0.23</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DeepLP†</td>
<td>22.02 ± 0.88</td>
<td>12.69 ± 0.29</td>
<td>46.20 ± 0.76</td>
<td>38.43 ± 1.88</td>
</tr>
</tbody>
</table>
## Results (UCF-101 and Pascal VOC2007)

### Accuracy (%) on the UCF-101 test set:

<table>
<thead>
<tr>
<th>Method</th>
<th>20% labeled</th>
<th>50% labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>33.5</td>
<td>45.6</td>
</tr>
<tr>
<td>MT*</td>
<td>36.3</td>
<td>45.8</td>
</tr>
<tr>
<td>PL*</td>
<td>37.0</td>
<td>47.5</td>
</tr>
<tr>
<td>S4L*</td>
<td>37.7</td>
<td>47.9</td>
</tr>
<tr>
<td>UPS</td>
<td><strong>39.4</strong></td>
<td><strong>50.2</strong></td>
</tr>
</tbody>
</table>

### mAP scores on the Pascal VOC2007 test set:

<table>
<thead>
<tr>
<th>Method</th>
<th>10% labeled</th>
<th>20% labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>18.36 ± 0.65</td>
<td>28.84 ± 1.68</td>
</tr>
<tr>
<td>PL</td>
<td>27.44 ± 0.55</td>
<td>34.84 ± 1.88</td>
</tr>
<tr>
<td>MixMatch</td>
<td>29.57 ± 0.78</td>
<td>37.02 ± 0.97</td>
</tr>
<tr>
<td>MT</td>
<td>32.55 ± 1.48</td>
<td>39.62 ± 1.66</td>
</tr>
<tr>
<td>UPS</td>
<td><strong>34.22 ± 0.79</strong></td>
<td><strong>40.34 ± 0.08</strong></td>
</tr>
</tbody>
</table>
Analysis (PL Accuracy)

UPS achieves higher pseudo-label accuracy while selecting similar number of pseudo-labels.
Analysis (Compatibility and Robustness)

UPS is compatible with most uncertainty estimation methods

<table>
<thead>
<tr>
<th>Method</th>
<th>1000 labels</th>
<th>4000 labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC-Dropout</td>
<td>8.14</td>
<td>6.36</td>
</tr>
<tr>
<td>MC-SpatialDropout</td>
<td>8.28</td>
<td>6.60</td>
</tr>
<tr>
<td>MC-DropBlock</td>
<td>9.76</td>
<td>7.50</td>
</tr>
<tr>
<td>DataAug</td>
<td>8.28</td>
<td>6.72</td>
</tr>
</tbody>
</table>

UPS is robust to pseudo-label selection hyperparameters
Conclusion

- We have introduced UPS,
  - a simple and efficient framework for effective pseudo-labeling based SSL
- UPS competes with SOTA consistency regularization based methods
  - without inherently relying on strong data augmentation
- We are first to propose negative pseudo-labeling for SSL
- UPS is versatile:
  - it is domain agnostic
  - can be easily used for multi-label classification
Thank You