PSEUDO-LABEL NOISE SUPPRESSION TECHNIQUES FOR SEMI-SUPERVISED SEMANTIC SEGMENTATION

INTRODUCTION

- Semi-supervised learning (SSL) reduce the need for large labelled datasets by incorporating unlabelled data into the training. This is particularly interesting for semantic segmentation, where labelling data is very costly and time-consuming.
- Current SSL approaches use an unsupervised trained model to generate predictions for unlabelled images, called pseudo-labels, which are subsequently used for training a new model from scratch.
- Since the predictions do not come from an error-free neural network, they are naturally full of errors. However, training with partially incorrect labels often reduce the final model performance. Thus, it is crucial to manage errors/noise of pseudo-labels wisely.
- In this work, we use three mechanisms to control pseudo-label noise and errors:
  1. We construct a base framework by mixing images with cow-patterns to reduce the negative impact of wrong pseudo-labels.
  2. We propose a simple and effective loss weighting scheme for pseudo-labels defined by the feedback of the model trained on these pseudo-labels.
  3. We study the common practice to ignore pseudo-labels with low confidence during training and empirically analyse the influence and effect of pseudo-labels with different confidence ranges and the contribution of pseudo-label filtering to the achievable performance gains.

METHOD

- We use the iterative pseudo-label refinement method and first train an initial segmentation model on the labelled dataset.
- The model is used as teacher $F_T$ to generate pseudo-labels for unlabelled images.
- A new student model $F_S$ is trained on the labelled and unlabeled dataset, while latter used the pseudo-labels from $F_T$.
- Cycle process where $F_S$ becomes the $F_T$ at next iteration.
- To prevent the model from memorizing wrong pseudo-labels, we utilize image mixing as perturbation when training with pseudo-labels [1].

ABSTRACTION


Ablation Study

Table: Ablation Study on three different datasets: Cityscapes [2] (C), Mapillary [3] (M) and PASCAL Voc 2012 [4] (P), with 15, 100 and 183 labelled images respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>ST</th>
<th>STCM</th>
<th>PLF</th>
<th>PLW</th>
<th>PLW_SCE</th>
<th>iteration</th>
<th>$\mathcal{C}(%)$</th>
<th>$\mathcal{M}(%)$</th>
<th>$\mathcal{P}(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.0</td>
<td>50.7</td>
<td>57.0</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>52.1</td>
<td>50.2</td>
<td>56.2</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56.2</td>
<td>53.9</td>
<td>66.3</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60.3</td>
<td>56.1</td>
<td>68.1</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60.9</td>
<td>56.8</td>
<td>67.8</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.6</td>
<td>57.0</td>
<td>68.9</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62.4</td>
<td>57.8</td>
<td>69.5</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.0</td>
<td>58.0</td>
<td>69.3</td>
</tr>
<tr>
<td>DL+</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66.5</td>
<td>60.1</td>
<td>-</td>
</tr>
</tbody>
</table>

Pseudo Label Filtering

- Remove pseudo-labels with low confidence from training and assume that a low confidence correlates with wrong pseudo-labels.
- A higher confidence is equated with a higher certainty that the pseudo-label is correct.

Pseudo Label Weighting

- Observation: The confidence of correct label predictions grow faster than the confidences of wrong label predictions over the course of model training.
- We propose a weighting scheme with the aim to assign smaller weights to possibly wrong pseudo-labels and therefore reduce the contribution of noisy pseudo-labels at training.
- Further adaption of the Symmetric Cross Entropy Loss as more robust loss function to outliers.
- With $p_i = F_k(y_i)$ as prediction of the model on the perturbed image, $y_i^k = \arg\max_k (F_k(x_i))$ as pseudo-labels and $a, \beta$ as balancing coefficients the loss is defined as:
  $$l^i = w^i \cdot \left( a CE(p_i, y_i^k) + \beta CE(y_i^k, p_i) \right)$$
- The dynamic weights is defined by the softmax output of $F_k$:
  $$w^i = \frac{P(y_i^k)}{P(p_i)}$$

EXPERIMENTS

- Cityscapes Dataset [2]: 2975 training and 500 validation images.
- We use different fraction of ratios between labelled and unlabelled data from the training set.

HUMAN POSE ESTIMATION

- Simple Unet architecture with ResNet18 backbone used.
- Metric: Percentage of correct keypoints (PKC) with 0.2 ratio.
- We observe similar findings as in semantic segmentation.

References

[2] Cordts et al., The cityscapes dataset for semantic urban scene understanding, 2016
[3] Neuhäusl et al., The reprisal video dataset for semantic understanding of street scenes, 2017
[4] Everingham et al., The pascal visual object classes (voc) challenge, 2010
[5] Sam Johnson and Mark Everingham, Distilled pose and nonlinear appearance models for human pose estimation, 2010