Self-supervised Difference Detection for Refinement CRF and Seed Interpolation

Wataru Shimoda  Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

Objective

Weakly supervised segmentation
- Use only image-level annotation and generate segmentation masks

CRF is a good refinement method but it often degrades the results due to unstable unary terms in weakly supervised segmentation
- Our motivation is to use CRF results as not teacher but adviser
- In our situation, we suppose advisers give us noisy information

Motivation

To utilize information from advisers
- We consider that many people first determine whether it is against their principles, and utilize opinions of other advisers for problems that are difficult to judge
- We model this scheme by difference detection task

Overview

Overview of Difference Detection Network (DD-Net)
- Training
Train a difference detection model using difference regions between raw segmentation masks generated by PSA[1] and its CRF results
- Inference
Integrate a pair of mask using DD-net outputs

Difference detection network

Definition of difference detection task
- Input: Pair of mask
\[(m_{\text{before}}, m_{\text{after}})\]
- Output: Difference region
\[M_{\text{before after}}\]

Definition of difference detection network (DD-net)
- Input: One of the pair of mask \((m_{\text{before}}, m_{\text{after}})\) and feature maps \((e(x), e'(x))\)
- Output: Probability map
\[p_{\text{before after}} \in \mathbb{R}^{N \times W} \]

Evaluation of input masks for integrations

Modeling the decision process about whether to use the adviser’s opinion
- Decision by own principles
  Make low confidence on pixels predicted by large differences between the input feature maps and segmentation masks
  - In both of decisions, high value (highlighted by white) indicates low confidence
- Decision taking into consideration the opinions of other advisers
  Make low confidence on pixels predicted by large differences between the input feature maps and segmentation masks

Static region refinement

Loss for difference detection in PSA and CRF
\[L_{\text{change}} = \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{after}}(u) \]
Loss for segmentation network for obtaining good representation from backbone network
\[L_{\text{seg-main}} = \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{before}}(u) + \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{after}}(u) \]
Final loss
\[L_{\text{static}} = L_{\text{seg}} + L_{\text{change}} \]

Dynamic region refinement

Losses for difference detection
\[L_{\text{diff-crf}} = \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{after}}(u) \]
\[L_{\text{diff-seg}} = \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{before}}(u) \]
Losses for segmentation network
\[L_{\text{seg-main}} = \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{before}}(u) + \frac{1}{m} \sum_{u \in \Omega} m_{u} \text{e}_{\text{after}}(u) \]
Final loss
\[L_{\text{dynamic}} = L_{\text{diff-crf}} + L_{\text{diff-seg}} + L_{\text{seg-main}} + L_{\text{seg-rag}} \]

Experiments

- Dataset: Pascal VOC 2012 dataset
- Evaluation metric: mean IoU

Comparison with WSS methods w/ additional supervision.

<table>
<thead>
<tr>
<th>Method</th>
<th>Val Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA</td>
<td>48.0</td>
<td>49.0</td>
</tr>
<tr>
<td>PSA-CRF refinement</td>
<td>48.0</td>
<td>49.0</td>
</tr>
<tr>
<td>Static region refinement</td>
<td>53.4</td>
<td>61.4</td>
</tr>
</tbody>
</table>

Comparison with WSS methods w/o additional supervision.

<table>
<thead>
<tr>
<th>Method</th>
<th>Val Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM-CRF (CVPR2018)</td>
<td>38.2</td>
<td>35.6</td>
</tr>
<tr>
<td>OCSNet (CVPR18)</td>
<td>48.1</td>
<td>45.1</td>
</tr>
<tr>
<td>DFPS (ICCV2018)</td>
<td>46.8</td>
<td>46.8</td>
</tr>
<tr>
<td>BEL-CRF(2018)</td>
<td>50.7</td>
<td>51.7</td>
</tr>
<tr>
<td>FSR (CVPR2018)</td>
<td>57.1</td>
<td>56.8</td>
</tr>
<tr>
<td>CRF16-SSD (CVPR17)</td>
<td>57.6</td>
<td>57.7</td>
</tr>
<tr>
<td>Dynamic region refinement</td>
<td>61.4</td>
<td>65.4</td>
</tr>
</tbody>
</table>

References