The 2nd Learning from Imperfect Data (LID) Workshop

Revisiting Class Activation Mapping for Learning from Imperfect Data

Wonho Bae*, Junhyug Noh*, Jinhwan Seo, and Gunhee Kim
Challenge Results

1st place
Track 3: Weakly Supervised Object Localization

2nd place
Track 1: Weakly Supervised Semantic Segmentation
Weakly-Supervised Object Localization
Class Activation Mapping (CAM)
Class Activation Mapping (CAM)

\[ \text{CNN} \rightarrow \text{F} \rightarrow \text{GAP} \rightarrow \text{p}_{\text{gap}} \rightarrow \text{CAM} \]

\[ w_{1,k} \ast F_1 + w_{2,k} \ast F_2 + w_{3,k} \ast F_3 + \cdots + w_{C,k} \ast F_C = M_k \]

\[ \text{resize} \rightarrow M'_k > \tau_{\text{loc}} \rightarrow B_k \rightarrow \text{localization result} \]

\( k: \text{monkey} \)
Class Activation Mapping (CAM) for Track 3

\[ \text{CNN} \rightarrow \text{F} \rightarrow \text{GAP} \rightarrow p_{gap} \]

\[ \sum w_{i,k} \cdot F_i = M_k \]

\[ M'_k \rightarrow \tau_{loc} \rightarrow B_k \rightarrow \text{localization result} \]
How to Grasp Whole Object Region?

[HaS] Singh, et al. ICCV 2017


[ACoL] Zhang, et al. CVPR 2018

[ADL] Choe, et al. CVPR 2019
Our Approach

• **Motivation**
  - Information to capture the whole area of the object already exists in feature maps

• **Problem**
  - Three modules (M1–M3) of CAM do not take phenomena (P1–P3) into account
  - It results in the localization being limited to small discriminative regions of an object

• **Solution**
  - Correctly utilize the information by simply modifying the three modules

![Diagram](image)
Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling (GAP) under P1

\[
p_{c}^{\text{gap}} = \frac{1}{H \times W} \sum_{(h,w)} F_{c}(h,w)
\]
Our Approach (1) Thresholded AveragePooling

- **Problem:** Global Average Pooling (GAP) under P1
Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling (**GAP**) under **P1**

Classification phase

Localization phase
Our Approach (1) Thresholded Average Pooling

- **Problem:** Global Average Pooling (GAP) under \( P1 \)

\[
p_{c}^{\text{gap}} = \frac{1}{H \times W} \sum_{(h,w)} F_c(h,w)
\]

- **Solution:** Thresholded Average Pooling (TAP)

\[
p_{c}^{\text{tap}} = \frac{\sum_{(h,w)} 1(F_c(h,w) > \tau_{\text{tap}}) F_c(h,w)}{\sum_{(h,w)} 1(F_c(h,w) > \tau_{\text{tap}})}
\]
Our Approach (2) Negative Weight Clamping

- Problem: Class Activation Maps (CAM) under P2

\[
M_k = \sum_{c=1}^{C} w_{c,k} \cdot F_c
\]

- M1: Global Average Pooling (GAP)
- M2: Class Activation Maps (CAM)
- M3: Thresholding

Phenomena observed in the feature map (F)
Our Approach (2) Negative Weight Clamping

- **Problem:** Class Activation Maps (CAM) under P2

  - Positive only
  - Negative only
  - Both
Our Approach (2) Negative Weight Clamping

- **Problem**: Class Activation Maps (CAM) under P2

IoA between the ground truth boxes and the CAMs

![Histogram of positive weights](chart-positive.png)

![Histogram of negative weights](chart-negative.png)
Our Approach (2) Negative Weight Clamping

- **Problem:** Class Activation Maps (CAM) under P2

\[
M_k = \sum_{c=1}^{C} w_{c,k} \cdot F_c
\]

- **Solution:** Negative Weight Clamping (NWC)

\[
M_k = \sum_{c=1}^{C} \mathbb{1}(w_{c,k} > 0) \cdot w_{c,k} \cdot F_c
\]
Our Approach (3) Percentile as a Thresholding Standard

- **Problem**: Maximum as a Standard (MaS) under P3

\[ \tau_{loc} = \theta_{loc} \cdot \max M'_k \]
Our Approach (3) Percentile as a Thresholding Standard

- **Problem:** Maximum as a Standard *(MaS)* under **P3**

| Num of channels (activation > \(\tau_{0.8}\)) | Result with CAM | CAM values (descending order) |
Our Approach (3) Percentile as a Thresholding Standard

- **Problem:** Maximum as a Standard (MaS) under P3

\[ \tau_{loc} = \theta_{loc} \cdot \max M'_k \]

- **Solution:** Percentile as a Standard (PaS)

\[ \tau_{loc} = \theta_{loc} \cdot \text{per}_{i}(M'_k) \]
Experimental Setting

• Backbone: ResNet50-SE
• Batch size: 210
• Input size: 384×384
• Random crop size: 336×336
• TAP threshold ($\tau_{tap}$): 0.05
• PaS percentile ($i$): 98
Results on Validation Set

- Results with different components

<table>
<thead>
<tr>
<th>Method</th>
<th>CRF</th>
<th>PaS</th>
<th>NWC</th>
<th>TAP</th>
<th>Peak IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>0.5254</td>
</tr>
<tr>
<td>+ Ours</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.5563</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.5881</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.6370</td>
</tr>
</tbody>
</table>

- To preserve the details of masks, we also applied a fully connected CRF.
- The performance gradually improves as each component is added.
## Leaderboard

- **Track 3: Weakly Supervised Object Localization**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Participant Team</th>
<th>Peak IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNUVL (Ours)</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>BJTU-Mepro-MIC</td>
<td>0.62</td>
</tr>
<tr>
<td>3</td>
<td>LEAP Group@PCA Lab</td>
<td>0.61</td>
</tr>
<tr>
<td>4</td>
<td>chohk (wsoi_aug)</td>
<td>0.53</td>
</tr>
<tr>
<td>5</td>
<td>TEN</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Qualitative Results

CAM + Ours

CAM + Ours

CAM + Ours

CAM + Ours
Expansion to Track 1

![Bar chart showing the ratio of classes in training and validation sets across different class counts.](chart.png)
Expansion to Track 1

Our target!
Class Activation Mapping (CAM) for Track 1

CNN → GAP → CAM

$\mathbf{F}$ → $\mathbf{p}_{gap}$ → $\mathbf{M}_k$

$w_{1,k} \mathbf{F}_1 + w_{2,k} \mathbf{F}_2 + w_{3,k} \mathbf{F}_3 + \cdots + w_{C,k} \mathbf{F}_C = \mathbf{M}_k$

$\mathbf{M}_k'$ > $\tau_{loc}$ → $\mathbf{B}_k$ → localization result

$k$: monkey

CAM (Class Activation Maps)
Leaderboard

- Track 1: Weakly Supervised Semantic Segmentation

<table>
<thead>
<tr>
<th>Rank</th>
<th>Participant Team</th>
<th>Mean IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cvl</td>
<td>45.18</td>
</tr>
<tr>
<td>2</td>
<td>SNUVL (Ours)</td>
<td>37.73</td>
</tr>
<tr>
<td>3</td>
<td>UCU &amp; SoftServe</td>
<td>37.34</td>
</tr>
<tr>
<td>4</td>
<td>IOOnlyHaveSevenDays</td>
<td>36.24</td>
</tr>
<tr>
<td>5</td>
<td>play-njupt</td>
<td>31.90</td>
</tr>
</tbody>
</table>
Thank You!