Efficient RGB-T Tracking via Cross-Modality Distillation

1. Attribute-based performance

We further analyze the attribute-based performance on RGBT234 [4] and LasHeR [6] dataset.

1.1. RGBT234 dataset

RGBT234 [4] is a large-scale RGB-T tracking dataset. It contains 12 challenge attribute labels, including no occlusion (NO), partial occlusion (PO), heavy occlusion (HO), low illumination (LI), low resolution (LR), thermal crossover (TC), deformation (DEF), fast motion (FM), scale variation (SV), motion blur (MB), camera moving (CM) and background clutter (BC).

As shown in Table 1, we analyze the attribute-based performance on RGBT234. For clarity, we only illustrate the four re-trained trackers and another four advanced trackers, i.e., JMMAC [11], M5L [7], CAT [5] and MANet++ [12]. From the results, we can see that our proposed method still performs well in most annotated attributions. Compared with such trackers based on MDNet (i.e., M5L [7], CAT [5], MANet++ [12], DAFNet [2] and FANet [14]), our method has remarkable improvements in case of PO, LI, DEF and SV. Compared with mfDiMP [10], which is based on DiMP [1] and employs two ResNet50 [3] for feature extraction, our algorithm achieves competitive performance but significantly reduce parameters.

1.2. LasHeR dataset

LasHeR [6] is currently the largest RGB-T tracking dataset. In addition to such challenges in RGBT234, LasHeR contains more challenges, including total occlusion (TO), hyaline occlusion (HO), high illumination (HI), abrupt illumination variation (AIV), similar appearance (SA), aspect ratio change (ARC), out-of-view (OV) and frame lost (FL).

As shown in Table 2, we further analyze the attribute-based performance on LasHeR. The results of our proposed method and some other state-of-the-art trackers, including MANet [8], DAPNet [13], DAFNet [2], MACNet [9], CAT [5], mfDiMP [10], FANet [14] and MANet++ [12], demonstrate that the our method performs the best under the most challenging conditions. First, in adverse lighting conditions, thermal crossover and low resolution, our method outperforms all other trackers. This demonstrates that the proposed method can enable such a compact model to fully explore the complementary information within multi-modal images. Second, our framework is robust to significant appearance changes, such as deformation, scale variation, camera moving and similar appearance. Finally, our model struggles to handle the out of view challenge and hyaline occlusion challenge. It may be due to the fact that student model can not learn effective information from the teacher model when targets are invisible.

References

[9] Hui Zhang, Lei Zhang, Li Zhuo, and Jing Zhang. Object tracking in RGB-T videos using modal-aware attention net-
Table 1. Attribute-based precision rate and success rate (PR/SR) scores obtained by using different trackers on RGBT234 dataset. The numbers with red colors indicate the best results.

<table>
<thead>
<tr>
<th>Trackers</th>
<th>DAPNet</th>
<th>S</th>
<th></th>
<th>mDiMP</th>
<th>Student-Distill</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CM</strong></td>
<td><strong>MB</strong></td>
<td><strong>OV</strong></td>
<td><strong>ALL</strong></td>
<td><strong>TO</strong></td>
<td><strong>DEF</strong></td>
</tr>
<tr>
<td><strong>PR/SR</strong></td>
<td>63.400.3</td>
<td>62.200.2</td>
<td>63.000.3</td>
<td>63.000.3</td>
<td>63.000.3</td>
</tr>
<tr>
<td><strong>Student-Distill</strong></td>
<td>63.400.3</td>
<td>62.200.2</td>
<td>63.000.3</td>
<td>63.000.3</td>
<td>63.000.3</td>
</tr>
</tbody>
</table>

Table 2. Attribute-based precision rate and success rate (PR/SR) scores obtained by using different trackers on LaSReR dataset. The numbers with red colors indicate the best results.

<table>
<thead>
<tr>
<th>Trackers</th>
<th>DAPNet</th>
<th>S</th>
<th></th>
<th>mDiMP</th>
<th>Student-Distill</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CM</strong></td>
<td><strong>MB</strong></td>
<td><strong>OV</strong></td>
<td><strong>ALL</strong></td>
<td><strong>TO</strong></td>
<td><strong>DEF</strong></td>
</tr>
<tr>
<td><strong>PR/SR</strong></td>
<td>63.400.3</td>
<td>62.200.2</td>
<td>63.000.3</td>
<td>63.000.3</td>
<td>63.000.3</td>
</tr>
<tr>
<td><strong>Student-Distill</strong></td>
<td>63.400.3</td>
<td>62.200.2</td>
<td>63.000.3</td>
<td>63.000.3</td>
<td>63.000.3</td>
</tr>
</tbody>
</table>


