“Uncertainty-aware Score Distribution Learning for Action Quality Assessment”
Supplementary Material

Table 1. Comparisons of different strategies to segment on the Diving action of AQA-7 [3].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp. Corr.</td>
<td>0.7642</td>
<td>0.8099</td>
<td>0.7928</td>
</tr>
</tbody>
</table>

1. Different Strategies for Segmenting Videos

During our experiments, we explored three strategies to divide the videos into multiple segments. Table 1 presents the results on the Diving action of the AQA-7 dataset [3], where the length of each video clip is 103 frames. Specifically, “6-seg” denotes the scheme used in [4], which first normalized the video into 96 frames, and then divided them into 6 segments where each clip contained 16 frames\(^1\). In fact, as pointed in the recent work [2], 10 is more proper for the number of segments. Based on this, we further studied two schemes. The first used \([0, 10, 20, 30, 40, 50, 60, 70, 80, 87]\) as the indices of beginning frames for the ten segments (denoted as “10-seg-s1”). Since \(103/10 = 10.3\), we set the stride to be 10 in most cases. And the last beginning index was set to be 87 due to the length of the video is 103 frames. The second scheme utilized \([0, 9, 19, 29, 38, 48, 58, 67, 77, 87]\) as the indices of beginning frames (denoted as “10-seg-s2”). As shown from the results, “10-seg-s1” achieves best result among the three. Hence, we applied this scheme to the other actions in AQA-7 and MTL-AQA datasets in our paper.

2. Visualization of Temporal Evolution

In Section 4.3, we present a visualization result on the Gym_Vault action [3]. Here we further display two visualization results on the Diving action [3] in Figure 1. As it illustrates, the stage that the player enters the water plays

\(^1\)We used I3D model [1] as the backbone. It took 16 frames as inputs.

Figure 1. Evolution of the score distributions in temporal domain. The \(x, y, z\) axis represents clip number, score and probability predicted for the certain score. The two samples are selected from the Diving action in AQA-7 [3].
a prominent part for action quality assessment. For example, in the bottom instance, the player causes a large splash from the 8th segment to the 10th segment. Hence, for these segments, the distributions reach the peak at the low-level scores.

References


