Learning Quality-aware Dynamic Memory for Video Object Segmentation

Yong Liu[1*], Ran Yu[1], Fei Yin[1], Xinyuan Zhao[2], Wei Zhao[2], Weihao Xia[3], and Yujiu Yang[1†]

[1] Tsinghua Shenzhen International Graduate School, Tsinghua University
[2] Huawei Technologies
[3] University College London
{liu-yong20,yu-r19}@mails.tsinghua.edu.cn, yang.yujiu@sz.tsinghua.edu.cn

1 Analysis on Long Videos.

To further prove the rationality and effectiveness of our dynamic memory updating strategy, we show the qualitative results on long videos in Fig. 1. In general, the most common practice for updating memory in practical applications is to retain the most recent memory frames. However, this approach is difficult to deal with object appearance changes or scene changes. As shown in Fig. 1, only retaining the most recent memory frames may cause the memory bank losing perception of the target object but our dynamic updating strategy allows for superior segmentation effect.

Fig. 1. Qualitative results of our proposed dynamic memory updating strategy. (a) is the results of retaining the most recent memory frames and (b) is applying our updating strategy. The memory frame storage limit is 25 frames.

Fig. 2. The bad case of QAM.

2 Failure Cases

The failure cases of QAM are the extremely difficult scenarios, *e.g.*, lots of similar objects overlapping each other, in which almost every frame is poorly segmented. Take the Fig. 2 as example, in this case, the target sheep are mixed with other background sheep. Along with the movement of the flock, it is difficult for the algorithm to correctly identify the target sheep as well as the outline of the sheep. Although we use relative quality scores in this scenario, QAM is less helpful.

3 Memory Interval

In order to avoid excessive memory redundancy in the same scene, we still choose to trigger the storage of every N frames, but the frame must meet our proposed principles. Otherwise, it will be deferred to the next frame to trigger. As Table 1 shows, it is not the smaller the memory interval, the better the segmentation effect. The reason is that although the smaller memory interval means more reference frames, excessive redundancy in the same scene affects the matching process.

Table 1. Experiment results of different memory interval on DAVIS2017 val set.

Memory Interval	$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$
3	82.3	88.1	85.2
5	**82.7**	**88.6**	**85.7**
7	81.6	87.8	84.7