Learning Where to Focus for Efficient Video Object Detection

Zhengkai Jiang[1,2], Yu Liu[3], Ceyuan Yang[3], Jihao Liu[3], Peng Gao[3], Qian Zhang[4], Shiming Xiang[1,2], and Chunhong Pan[1,2]

[1] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences
[3] The Chinese University of Hong Kong
[4] Horizon Robotics
{zhengkai.jiang, smxiang}@nlpr.ia.ac.cn

Abstract. We conduct extensive ablation studies for different settings of sampling locations initialization. For Uniform Initialization, experiments for different neighbours on the feature grid are conducted to study the influence of neighbourhood size on the final performance. And for Gaussian Initialization, the different number of points settings are studied. Note that the sampling locations are fixed during training.

Different Neighbors for Uniform Initialization. In this section, we study the influence of neighbourhood sizes. As Table 1 shows, larger the neighbourhood size is, more runtime will be. When the neighbourhood size equals 4, the better tradeoff is obtained.

Table 1. Comparison about different neighbors on the final performance.

The size of neighbor	1	2	3	4	5	6
runtime(FPS)	25.0	23.8	22.6	21.7	20.6	18.9
mAP(%)	74.3	75.0	75.3	75.5	75.8	75.6

Different Number of Points for Gaussian Initialization. In this section, we study the influence of the number of points. As Table 2 shows, more points are, more runtime will be. When the number of points equals 40, the better tradeoff is obtained.

Table 2. Comparison about different neighbors on the final performance.

Points	10	20	30	40	50	100
Runtime(FPS)	25.0	24.0	23.4	23.0	22.6	21.1
mAP(%)	74.3	75.0	75.3	75.5	75.8	75.6

Table 3. Comparison on different keyframe intervals between our method and flow-warping. mAP(%) is used in default

Keyframe Interval	4	6	8	10	12	14	16
Flow-Warping	76.5	76.3	76.2	76.1	75.9	75.6	75.1
Ours	**77.6**	**77.4**	**77.4**	**77.2**	**77.3**	**76.5**	**76.1**

Fig. 1. Visualization comparisons between our method and R-FCN image detector, DFF video object detector.

Keyframe Interval. We conduct an ablation experiment to study the influence of the testing keyframe interval. The results can be seen in Table 3. Our method could consistently outperform optical flow-warping based method at different keyframe intervals.