

Appendix

Mengnan Liu¹, Le Wang^{1*}, Sanping Zhou¹, Kun Xia¹, Qi Wu¹, Qilin Zhang², and Gang Hua³

¹ National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

² DeepNight.ai

³ Multimodal Experiences Research Lab, Dolby Laboratories

1 Additional Experiments

Tab. 1 shows the results with respect to different boundary weighting parameter on THUMOS'14. And we can observe that when $\lambda = 0.5$, the model performs better. We also conduct experiments to analyze the hyper-parameters of each loss item on the THUMOS14 in Tab. 3. We compared SMBD with LACP in terms of the boundaries obtained during training in Fig. 1.

We calculate the distance between the predicted boundaries and the ground truth (GT) for each video, and then count the number of videos. The boundaries obtained by SMBD have a smaller discrepancies with GT, and the accuracy of *distance* $\in [0, 5]$ improves 6.5%.

Table 1: The boundary weighting parameter λ analysis on THUMOS'14. AVG[0.1 : 0.5] and AVG[0.3 : 0.7] represent the mean average precision (mAP) at thresholds [0.1 : 0.1 : 0.5] and [0.3 : 0.1 : 0.7], respectively.

parameter	mAP@IoU(%)					AVG	AVG
	0.3	0.4	0.5	0.6	0.7	[0.1 : 0.5]	[0.3 : 0.7]
$\lambda = 0.1$	64.9	56.6	46.5	35.7	21.6	63.1	45.1
$\lambda = 0.2$	66.0	58.1	46.9	36.0	21.5	64.2	45.7
$\lambda = 0.3$	66.5	58.3	46.8	35.4	21.2	64.5	45.6
$\lambda = 0.4$	64.9	56.7	46.6	35.4	21.7	63.2	45.0
$\lambda = 0.5$	66.0	57.9	47.0	36.0	22.0	64.2	45.7
$\lambda = 0.6$	65.6	57.8	46.5	35.4	21.0	63.8	45.3
$\lambda = 0.7$	64.8	56.4	46.3	35.2	21.3	63.0	44.8
$\lambda = 0.8$	66.5	58.5	46.3	34.6	20.6	64.6	45.3
$\lambda = 0.9$	66.2	58.3	46.3	35.5	21.1	64.3	45.5

*Corresponding author

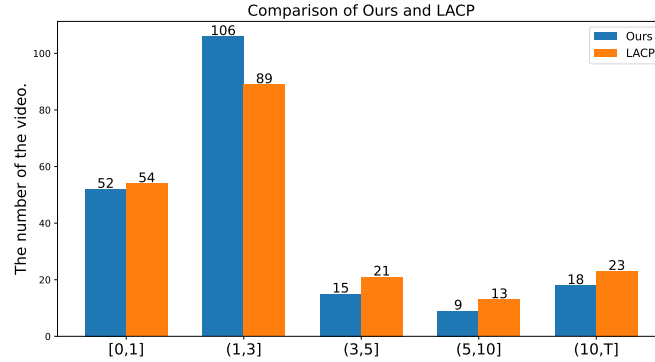


Fig. 1: Comparison with the LACP on boundary accuracy during training. The abscissa represents the intervals of the distance between the predicted boundaries and the ground truth. The vertical axis represents the number of videos.

Table 2: The loss coefficients analysis on THUMOS’14 at different values. AVG[0.1 : 0.5] and AVG[0.3 : 0.7] represent the mean average precision (mAP) at thresholds [0.1 : 0.1 : 0.5] and [0.3 : 0.1 : 0.7], respectively.

parameter	mAP@IoU(%)					AVG	AVG
	0.3	0.4	0.5	0.6	0.7	[0.1 : 0.5]	[0.3 : 0.7]
$\gamma_1 = 0.5$	65.7	57.6	46.4	35.6	21.2	63.7	45.3
$\gamma_1 = 0.7$	66.0	57.9	47.0	36.0	22.0	64.2	45.7
$\gamma_1 = 1.0$	65.7	57.8	46.2	35.3	21.0	63.8	45.2
$\gamma_2 = 0.75$	64.9	56.9	45.9	35.4	21.4	63.1	44.9
$\gamma_2 = 1.0$	66.0	57.9	47.0	36.0	22.0	64.2	45.7
$\gamma_2 = 1.25$	64.8	56.9	46.6	35.5	21.4	63.2	45.1

Table 3: We conduct an experiment with the state-of-the-art method HR-Pro for comparison on THUMOS14 in Tab.

	mAP@IoU(%)								AVG
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	[0.1:0.7]	
SF-Net	68.3	62.3	53.2	40.7	29.3	18.4	9.6	40.3	
Ju <i>et al.</i>	72.8	64.9	58.3	46.4	34.5	21.8	11.9	44.3	
LACP	75.7	71.4	64.6	56.5	45.3	34.5	21.8	52.8	
CRRC-Net	77.8	73.5	67.1	57.9	46.6	33.7	19.8	53.8	
HR-Pro	85.6	81.6	74.3	64.3	52.2	39.8	24.8	60.3	
SF-Net+SMBD	72.3	70.0	59.1	46.4	33.4	20.7	10.5	44.6	
LACP+SMBD	77.7	72.5	66.0	57.9	47.0	36.0	22.0	54.2	
HR-Pro+SMBD	86.1	82.1	75.2	65.1	53.5	41.2	26.3	61.4	