

# Appendix

## 1. Additional Experiments

Tab.1 presents a comparative analysis of the number of queries constructed for each category. And we can observe that when  $q = 3$ , the model performs best. We also conduct experiments to analyze the balanced weight of Transportation Cost Map on the THUMOS14 in Tab.2, and the model performs best when  $\beta = 0.05$ . We present the experiment of the experiment of concatenation compares with addition in Tab.3 and the setting of coefficient of Eq.9 in Tab.4

Table 1. Comparison of different number of queries constructed for each category.

Parameter	mAP@IoU(%)					AVG [0.3:0.7]
	0.3	0.4	0.5	0.6	0.7	
$q = 5$	71.7	62.6	53.1	39.4	26.6	50.7
$q = 4$	72.9	63.5	53.7	40.4	26.8	51.5
$q = 3$	73.1	64.2	54.3	41.3	27.4	52.1
$q = 2$	72.8	64.3	53.8	40.7	27.2	51.8
$q = 1$	72.3	63.9	51.8	40.9	26.7	51.1

Table 2. Comparison of different balanced weight of Transportation Cost Map on the THUMOS14.

Parameter	mAP@IoU(%)					AVG [0.3:0.7]
	0.3	0.4	0.5	0.6	0.7	
$\beta = 0.5$	69.8	60.2	49.3	38.2	24.1	48.3
$\beta = 0.1$	71.3	61.8	51.9	39.6	25.8	50.1
$\beta = 0.05$	73.1	64.2	54.3	41.3	27.4	52.1
$\beta = 0.01$	71.9	61.9	52.3	40.5	26.2	50.6

Table 3. The experiment of concatenation compares with addition.

Fusion Method	mAP@IoU(%)					AVG [0.3:0.7]
	0.3	0.4	0.5	0.6	0.7	
Addition	73.6	63.9	52.9	41.1	25.2	51.3
Concatenation	73.1	64.4	54.3	43.1	27.4	52.1

Table 4. The setting of coefficient of Eq.9

Coefficient	mAP@IoU(%)					AVG [0.3:0.7]
	0.3	0.4	0.5	0.6	0.7	
$\lambda_{detr} = 1, \lambda_{act} = 0.05$	71.9	63.9	53.0	39.0	23.5	52.3
$\lambda_{detr} = 1, \lambda_{act} = 0.1$	71.2	62.3	51.1	38.1	23.7	49.3
$\lambda_{detr} = 1, \lambda_{act} = 0.5$	73.1	66.4	54.3	43.1	27.4	52.1
$\lambda_{detr} = 1, \lambda_{act} = 1$	72.0	61.5	50.5	38.6	24.9	49.5

Fig.1 illustrates the query construction process for the 20 categories in the THUMOS14 dataset. In our method, three queries are generated for each category, corresponding to three cluster centers. With a total of 20 categories in the complete THUMOS14 dataset,  $20 \times 3$  static queries are constructed and used as prior information input to the decoder.

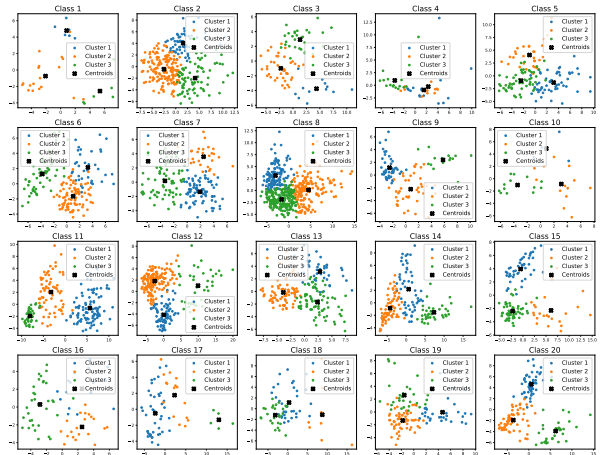


Figure 1. Visualization of the clustering results of all categories from the THUMOS14 dataset. Each category is divided into three clusters, resulting in three static queries per category.