TALL: Temporal Activity Localization via Language Query (Supplemental Material)

Jiyang Gao[1] Chen Sun[2] Zhenheng Yang[1] Ram Nevatia[1]
[1]University of Southern California [2]Google Research
{jiyangga, zhenheny, nevatia}@usc.edu, chensun@google.com

To directly compare our the temporal regression method with previous state-of-the-art methods on traditional action detection task, we did additional experiments on THUMOS-14.

Since THUMOS is a classification task with limited number of action classes, we removed the cross-modal part and trained the localization network with classification loss (cross-entropy loss) and regression loss. We trained a model on the validation set (train set only contains trimmed videos which are not suitable for localization task) and tested it on the test set. The regression model contains 20*2 outputs, corresponding to the 20 categories in the dataset, α is set to 2.0 and 10.0 for non-parameterized and parameterized regression respectively. For each category, we use NMS to eliminate redundant detections in every video, the NMS threshold is set to (tIoU - delta), where tIoU = 0.5 and delta=0.2. We report mAP at tIoU=0.5. For training sample generation, we use the same procedure as SCNN [24], we set the high IoU threshold as 0.5 (SCNN used 0.7) and low IoU threshold as 0.1 (SCNN used 0.3) for generating training samples. Note that, our method and SCNN both use C3D features.

Table 1. Temporal action localization experiments on THUMOS-14

	SCNN	cls	reg-p	reg-np	reg-np (p+d)
mAP	19.0	16.3	18.9	19.8	20.5

As shown, "cls" for only using classification loss, "reg-p" for classification loss+parameterized regression loss, "reg-np" for classification loss+ non-parameterized regression loss. For "cls", "reg-p","reg-np", we use the proposals generated by SCNN (from their github codes) as input, so that we can fairly compare the effect of classification loss, localization loss (used in SCNN) and temporal regression loss. "reg-np (p+d)" means that we apply temporal regression on both proposal generation and action detection.

Our method (reg-np) outperforms SCNN. Comparing with "cls" and "reg-np", we can see the improvement by the temporal regression. By applying temporal regression on proposal generation, we can see a further improvement from 19.8 to 20.5.