TURN TAP: Temporal Unit Regression Network for Temporal Action Proposals Supplementary Materials

Jiyang Gao[1*]   Zhenheng Yang[1*]   Chen Sun[2]   Kan Chen[1]   Ram Nevatia[1]

[1]University of Southern California   [2]Google Research

{jiyangga, zhenheny, kanchen, nevatia}@usc.edu,   chensun@google.com

1. Seven methods of proposal generation

In the paper, 7 proposals were used in the correlation analysis between temporal action proposal performance and temporal action localization performance. The 7 proposals are: (1) Random: randomly assigning a starting time and an ending time, subject to that ending time is larger, (2) sliding windows: {16,32,64,128,256,512} frames, 75% overlap; (3) sliding windows: {32,64,128,256} frames, 75% overlap; (4) sliding windows: {64,128} frames, 75% overlap; (5) S-CNN proposals: length between 16 to 512 frames, (6) S-CNN proposals: length between 32 to 256 frames, (7) S-CNN proposals: length between 64 to 128 frames.

2. Frame sampling method

We sample frames in a unit for feature extraction. The different methods of frame sampling for different features are shown in Figure 1.
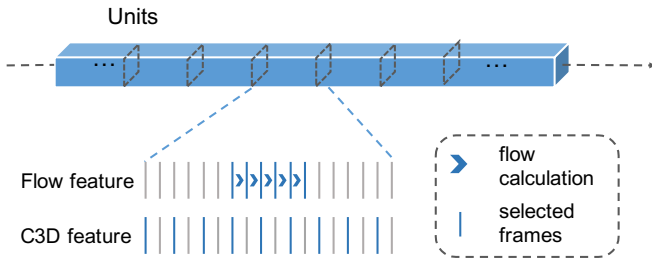

Figure 1: Illustration of frame sampling

3. Details of computation time

Table 1: Detailed runtime analysis

	Feature Extraction (ms)	Classification & Regression (ms)	Total Time (ms)	FPS
TURN-FL-16	7.5	0.2	7.7	129.4
TURN-C3D-16	0.9	0.2	1.1	880.8

Runtime analysis of testing TURN on THUMOS-14 is shown in Table 1. The first column "Feature Extraction" includes the time of both optical flow extraction and forward pass of BN-inception network. We divide the runtime of processing all testing videos by the total number of frames to get processing time per frame, and cacluate FPS accordingly. The runtime is evaluated on a single Nvidia Titan X Pascal GPU.

4. Pooling method comparison

Two different pooling methods are compared by evaluating AR-F performance on THUMOS-14 dataset. The results in Table 2 show that two pooling methods give similar TAP performance.

Table 2: AR-F performance comparison of different pooling methods on THUMOS-14

Frequency	0.1	0.5	1.0	2.0	5.0
Average pooling	0.132	0.312	0.424	0.543	0.647
Max pooling	0.129	0.309	0.420	0.548	0.649