Supplementary Material: Spatiotemporal Deformable Scene Graphs for Complex Activity Detection

Salman Khan
salmankhan@brookes.ac.uk
Fabio Cuzzolin
fabio.cuzzolin@brookes.ac.uk

Visual Artificial Intelligence Laboratory
Oxford Brookes University
Oxford, UK

1 Additional Details

1.1 Overview

A general overview of our complex activity framework is visualized in Figure 1.

1.2 Tube Detection

The detail of AMTNet tube generation is illustrated in Figure 2.

1.3 Algorithm

The complete process of our framework is described in Algorithm 1.

2 Datasets Details

Due to the lack of space here we provided the detail of both our augmented datasets with their statistics.

2.1 ROAD

The detail of each class and its statistics with complete description is given in Table 1.

2.2 SARAS-ESAD

The detail of each phase and its statistics with complete description is given in Table 2.

© 2021. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

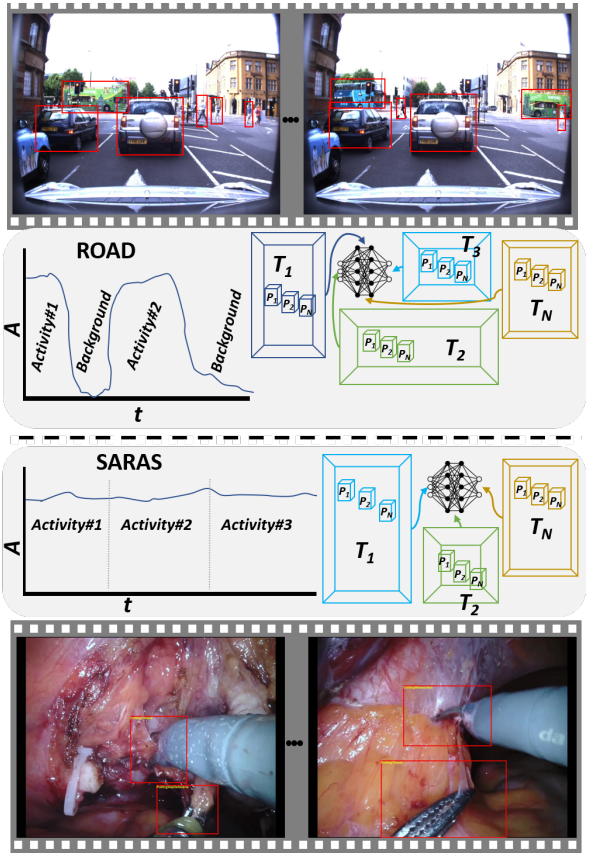

Figure 1: Our complex activity recognition framework workflow for two videos from the ROAD (autonomous driving) and SARAS (surgical robotics) datasets. The overall concept of detecting complex activities via a part-based deformation mechanism is the same in both cases. However, in ROAD background frames exist which do not contain any complex road activity. In contrast, the SARAS complex activities are actually the phases of a surgical procedure, which are contiguous without the need for a background label.

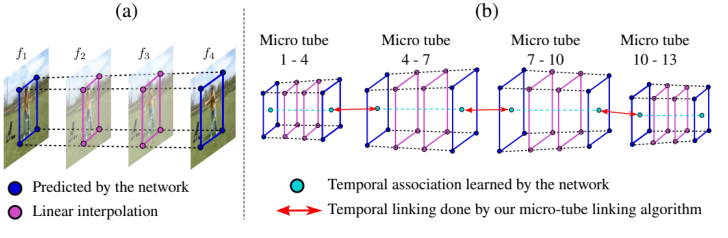

Figure 2: (a) AMTNet generates micro-action tubes by predicting bounding boxes for the start and the end frame of a snippet. Predictions for a fixed number of intermediate frames are constructed by bilinear interpolation. (b) Complete action tubes are created by temporally linking micro action tubes by dynamic programming.

Algorithm 1: Complex Activity Detection

Input: Input Video V divided into N number of Snippets S consists of fixed M number of frames $F^{1,2,3,\ldots,M}$
Initialization: Load pretrained action tube detector ATD on desire dataset, import our 3D deformable RoI pooling $DRoI$ and sparsity Sp modules ;
for $i \in 1,2,3,\ldots, N$ **do**
  Extract features and detect action tubes ${B_A}^{M\times K\times 5}, {X_B}^{M\times 64\times 300\times 300} \leftarrow ATD(S_i)$;
  ${X_{DRoI}}^{K\times 64\times M\times 7\times 7} \leftarrow DRoI({B_A}^{M\times K\times 5}, {X_B}^{M\times 64\times 300\times 300})$;
  $graphs^k \leftarrow Construct_Graph({X_{DRoI}}^{K\times 64\times M\times 7\times 7})$;
  ${X_G}^{k\times 2048} \leftarrow G(graphs^k)$;
  $CA_i \leftarrow Softmax({X_G}^{k\times 2048})$;
  if $CA_{i-1} == CA_i$ **then**
    Same activity;
  else
    if $CA_{i-1} == CA_{i+1}$ **then**
      same activity;
    else
      end of activity;
      start new activity;
    end
  end
end
Output: Start and end frames with activity labels

No	Complex activity	Train snippets	Test snippets	Scenario/Description
1	Negotiating intersection	523	414	Autonomous vehicle (AV) stops at intersection/junction, waits for other vehicles to pass, then resumes.
2	Negotiating pedestrian crossing	381	25	AV Stops at pedestrian crossing signal, waits for pedestrian(s) and cycles to cross, then resumes.
3	Waiting in a queue	570	447	AV stops and waits in a queue for a signal (either pedestrian crossing or junction traffic light).
4	Merging into vehicle lane	68	3	AV stops and waits for bus/car to merge into the vehicle lane from the opposite side of the road.
5	Sudden appearance	3	3	A pedestrian/cycle/vehicle suddenly appears in front of the vehicle, coming from a side street.
6	Walking in the middle of road	13	19	A pedestrian walking in the middle of the road in front of the AV, causing the AV to slow down or stop.

Table 1: List of ROAD complex activities with number of train and test snippets splits and brief description.

No	Complex activity/ Phase #	Train snippets	Test snippets	Scenario/Description
1	Phase#1 (Bladder mobilization)	1808	999	Tools only used for mobilization
2	Phase#2 (AFP Dissection)	6263	4455	Anytime only Anterior Prostatic Fat (AFP) is removed
3	Phase#3 (Bladder neck transection)	8353	2419	Whenever the scissor cut the neck, until all the neck is transected
4	Phase#4 (NVP dissection)	361	46	Located just under the seminal vescicles.
5	Phase#5 (Exposure of seminal vescicles)	2858	800	White bags surrounding the bladder neck. Comprises fat removal.
6	Phase#6 (Urethral division)	2553	1685	Start even when removing fat on the urethra.
7	Phase#7 (Prostate liberation)	1092	2903	After urethral division. It is not AFP. Stops only when completely free.
8	Phase#8 (Vesicourethral anastomosis)	4430	4690	Every time there is string and needle in the scene. Stops when urethra is linked to bladder

Table 2: SARAS video phases with number of train and test snippets and a brief description.