Multimodal Explanations by Predicting Counterfactuality in Videos (supplementary material)

1. Algorithm for finding maximum subpath in the 3D tensor

We show the algorithm to find the maximum subpath in the 3D tensor based on the dynamic programming proposed by [1] as below.

Input:
$M(u,t)$: the local discriminative scores; $\{u=(i,j)$: the 2D index of spatial coordinate$\}$ $\{t$: the frame in the video$\}$
Output:
$S(u,t)$: the accumulated scores of the best path leads to (u,t);
$P(u,t)$: the best path record for tracing back;
S^* : the accumulated score of the best path;
l^* : the ending location of the best path;

$S(u,1) = M(u,1), \forall u$;
$P(u,t) = null, \forall (u,t)$;
$S^* = -\infty$;
$l^* = null$;
for $i \leftarrow 2$ **to** n **do**
  foreach $u \in [1..w] \times [1..h]$ **do**
    $v_0 \leftarrow \arg\max_{v \in N(u)} S(v, i-1)$;
    if $S(v_0, i-1) > 0$ **then**
      $S(u,i) \leftarrow S(v_0, i-1) + M(u,i)$;
      $P(u,i) \leftarrow (v_0, i-1)$;
    else
      $S(u,i) \leftarrow M(u,i)$;
    end if
    if $S(u,i) > S^*$ **then**
      $S^* \leftarrow S(u,i)$;
      $l^* \leftarrow (u,i)$;
    end if
  end for
end for

Algorithm 1: Algorithm for finding maximum subpath in the 3D tensor [1]

2. The influence of the complexity of classification model on the negative class accuracy

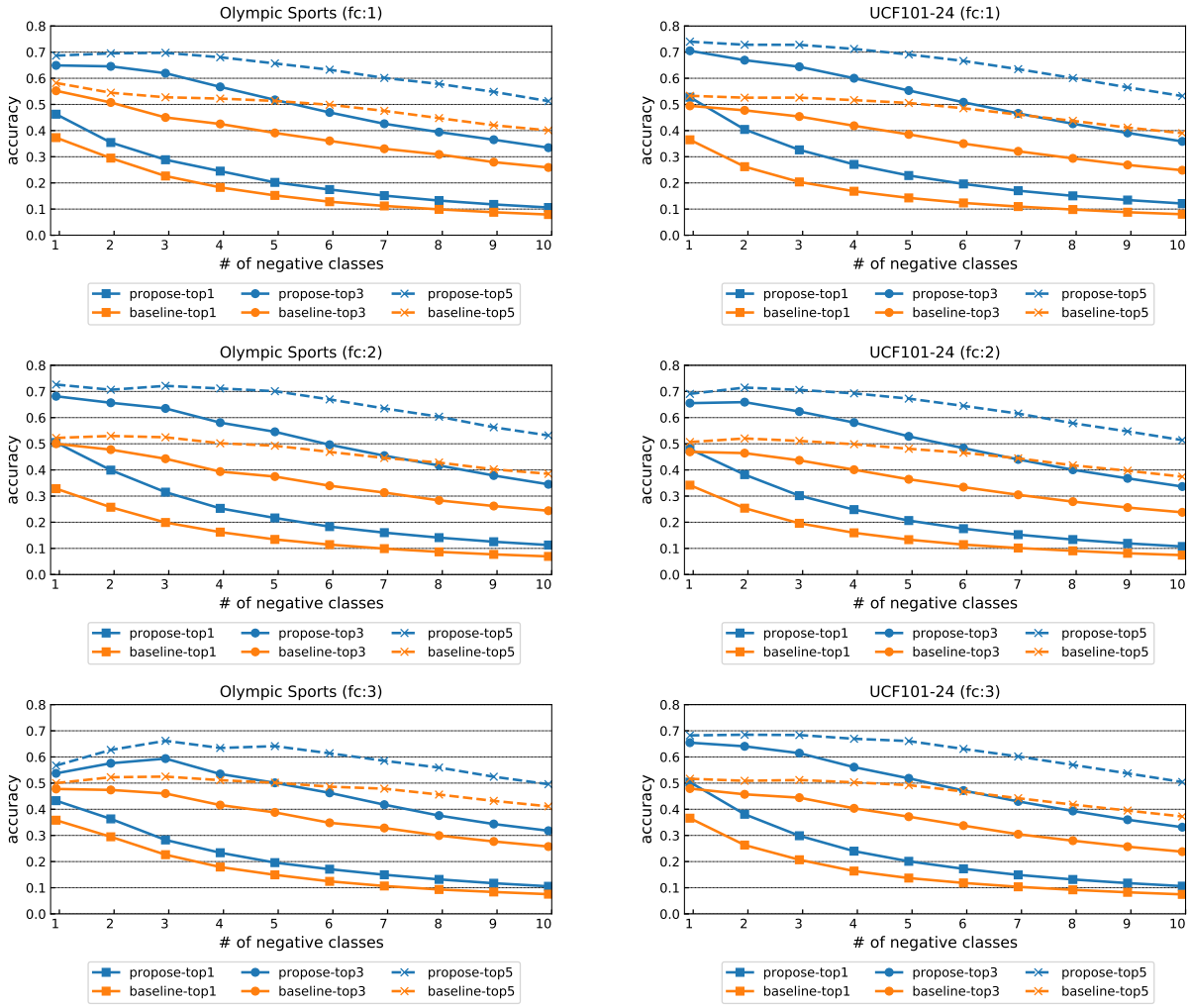


Figure 1. The negative class accuracy on Olympic Sports dataset (left) and UCF101-24 dataset (right). Each row corresponds to the number of fully-connected layer of the classification module. y-axis indicates the mean accuracy and x-axis means the number of the negative classes used for averaging whose prediction value is maximum.

3. List of attributes

3.1. Olympic Sports dataset

'Run', 'Slow run', 'Fast run', 'Indoor', 'outdoor', 'Ball', 'small ball', 'big ball', 'Jump', 'Small local Jump', 'Local jump up', 'Jump Forward', 'Track', 'Bend', 'StandUp', 'Lift something', 'Raise Arms', 'One Arm Open', 'Turn Around', 'Throw Up', 'Throw away', 'water', 'Down Motion in Air', 'Up Motion in Air', 'Up Down Motion Local', 'Somersault in Air', 'With Pole', 'Two hand holding pole', 'One hand holding pole', 'Spring Platform', 'Motion in the air', 'one arm swing', 'Crouch', 'Two Arms Open', 'Two Arms Swing overhead', 'Turn around with two arms open', 'Run in Air', 'Big Step', 'Open Arm Lift', 'With Pat'

3.2. UCF101-24 dataset

'Body Motion is Flipping', 'Body Motion is Walking', 'Body Motion is Running', 'Body Motion is Riding', 'Body Motion is Up down', 'Body Motion is Pulling', 'Body Motion is Lifting', 'Body Motion is Pushing', 'Body Motion is Diving', 'Body Motion is Jumping Up', 'Body Motion is Jumping Forward', 'Body Motion is Jumping Over Obstacle', 'Body Motion is Spinning', 'Body Motion is Climbing Up', 'Body Motion is Horizontal', 'Body Motion is Vertical Up', 'Body Motion is Vertical Down', 'Body Motion is Bending', 'Object is Ball Like', 'Object is Big Ball Like', 'Object is Stick Like', 'Object is Rope Like', 'Object is Sharp', 'Object is Circular', 'Object is Cylinderical', 'Object is Musical Instrument', 'Object is Portable Musical Instrument', 'Object is Animal', 'Object is Boat Like', 'Posture is Sitting', 'Posture is Sitting In Front Of Table Like Object', 'Posture is Standing', 'Posture is Lying', 'Posture is Handstand', 'Body Parts Used is Head', 'Body Parts Used is Hands', 'Body Parts Used is Arms', 'Body Parts Used is Legs', 'Body Parts Used is Foot'

4. Output Examples

SalsaSpin not SoccerJuggling because Body_Motion is Pulling

RopeClimbing not SalsaSpin because Body_Motion is Vertical_Up

Biking not SkateBoarding because Posture is Sitting

IceDancing not Skiing because Body_Motion is Spinning

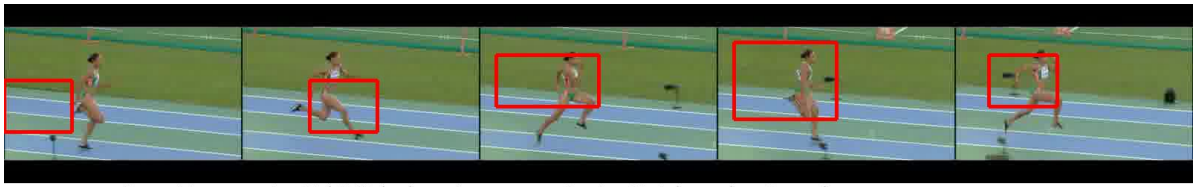
Skijet not Surfing because Body_Parts_Used is Arms

SkateBoarding not Surfing because Body_Parts_Used is Legs

Biking not Basketball because Posture is Sitting

TennisSwing not SoccerJuggling because Object is Circular

Fencing not Basketball because Object is Sharp



LongJump not CliffDiving because Body_Motion is Running

LongJump not SkateBoarding because Body_Motion is Running

Biking not PoleVault because Body_Parts_Used is Legs

SkateBoarding not HorseRiding because Body_Motion is Bending

5. Dataset collection

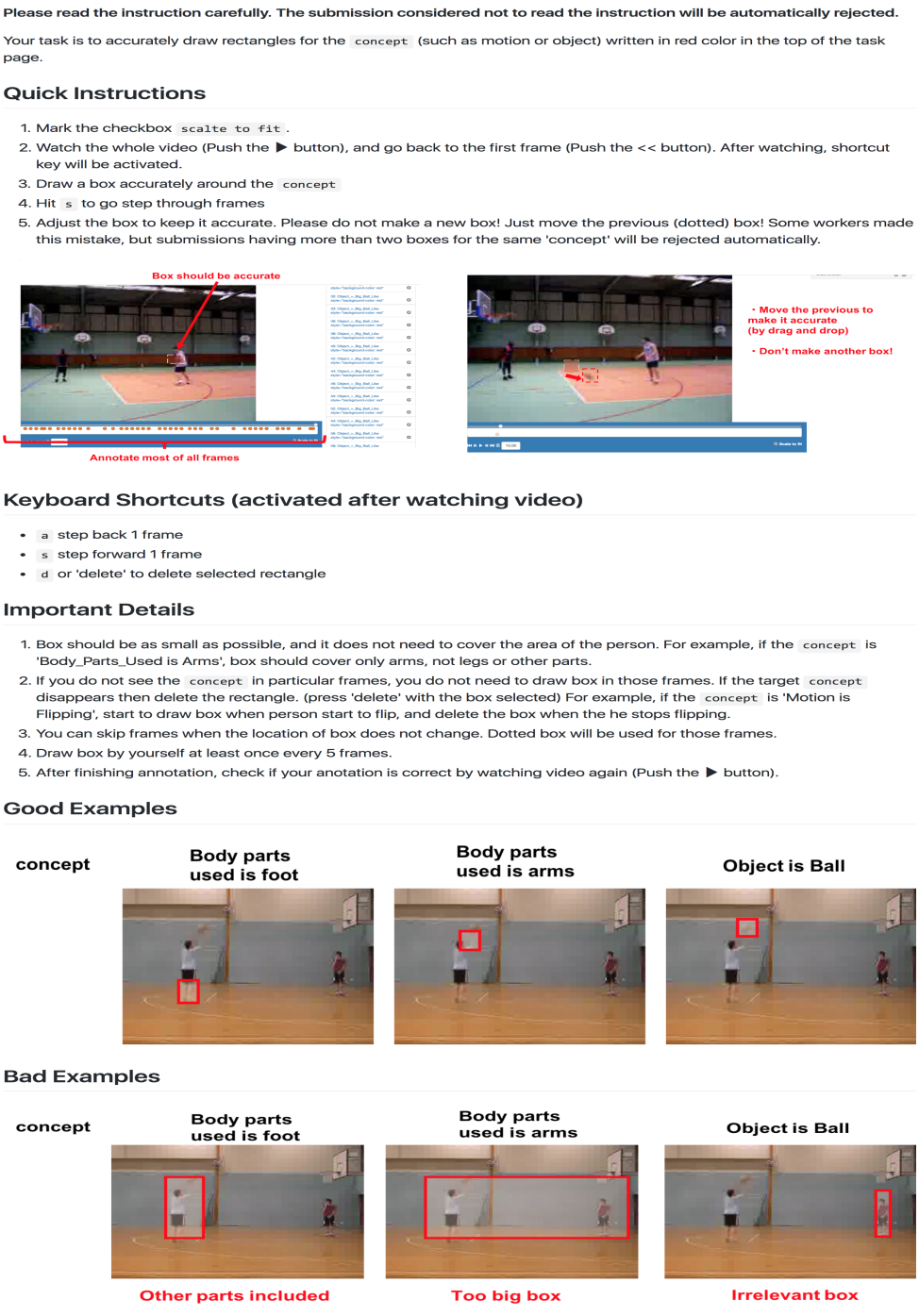


Figure 2. Screen shot of the instruction for collecting bounding box annotation on AWS.

6. Dataset Examples


Body_Motion is Vertical_Down
Body_Parts_Used is Hands
Body_Motion is Vertical_Down
Object is Ball_Like
Posture is Standing
Object is Stick_Like
Posture is Standing
Body_Parts_Used is Hands
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Motion is Running
Body_Motion is Running
Body_Motion is Jumping_Forward
Object is Animal
Body_Motion is Bending
Posture is Sitting
Body_Motion is Diving
Jeromy Cummings 3.80m
Body_Parts_Used is Foot
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Parts_Used is Arms
Body_Parts_Used is Hands
Body_Motion is Spinning
Object is Boat_Like
Body_Parts_Used is Hands
Body_Parts_Used is Legs
Body_Motion is Jumping_Forward
Body_Parts_Used is Foot
Body_Motion is Riding
Body_Parts_Used is Foot
Body_Motion is Riding
Body_Motion is Vertical_Down
Body_Parts_Used is Foot
Body_Motion is Flipping
Object is Animal
Body_Parts_Used is Arms
Object is Stick_Like
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Motion is Bending


Body_Motion is Vertical_Down
Body_Parts_Used is Hands
Body_Motion is Vertical_Down
Object is Ball_Like
Posture is Standing
Object is Stick_Like
Posture is Standing
Body_Parts_Used is Hands
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Motion is Running
Body_Motion is Running
Body_Motion is Jumping_Forward
Object is Animal
Body_Motion is Bending
Posture is Sitting
Body_Motion is Diving
Jeromy Cummings 3.80m
Body_Parts_Used is Foot
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Parts_Used is Arms
Body_Parts_Used is Hands
Body_Motion is Spinning
Object is Boat_Like
Body_Parts_Used is Hands
Body_Parts_Used is Legs
Body_Motion is Jumping_Forward
Body_Parts_Used is Foot
Body_Motion is Riding
Body_Parts_Used is Foot
Body_Motion is Riding
Body_Motion is Vertical_Down
Body_Parts_Used is Foot
Body_Motion is Flipping
Object is Animal
Body_Parts_Used is Arms
Object is Stick_Like
Body_Parts_Used is Arms
Body_Parts_Used is Foot
Body_Motion is Bending


Body_Motion is Climbing_Up
Body_Motion is Flipping
Body_Parts_Used is Hands
Object is Stick_Like
Body_Parts_Used is Hands
Body_Parts_Used is Arms
Body_Parts_Used is Legs
Body_Parts_Used is Legs
Posture is Standing
Object is Stick_Like
Posture is Standing
Body_Motion is Horizontal
Body_Motion is Vertical_Up
Body_Motion is Horizontal
Body_Parts_Used is Legs
Body_Motion is Horizontal
Body_Parts_Used is Hands
Body_Motion is Climbing_Up
Body_Parts_Used is Legs
Body_Parts_Used is Arms
Object is Big_Ball_Like
Posture is Standing
Body_Motion is Flipping
Body_Motion is Jumping_Forwar
Posture is Standing
Object is Stick_Like
Object is Circular
Object is Rope_Like
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Parts_Used is Foot
Body_Motion is Diving
Body_Motion is Vertical_Up
Body_Motion is Horizontal
Object is Ball_Like
Body_Motion is Jumping_Forwar
Body_Parts_Used is Hands
Body_Parts_Used is Foot
Body_Parts_Used is Legs
Body_Parts_Used is Hands
Posture is Standing
Body_Motion is Bending

Object is Stick_Like

Body_Parts_Used is Hands

Body_Parts_Used is Foot

Body_Motion is Horizontal

Body_Parts_Used is Foot

Body_Parts_Used is Arms

Body_Parts_Used is Foot

Body_Parts_Used is Hands

Body_Motion is Jumping_Up

Object is Boat_Like

Posture is Standing

Body_Parts_Used is Foot

Body_Parts_Used is Hands

Body_Parts_Used is Legs

Object is Ball_Like

Body_Parts_Used is Legs

Body_Motion is Bending

Object is Ball_Like

Body_Motion is Vertical_Down

Body_Motion is Bending

Body_Motion is Jumping_Forward

Body_Motion is Vertical_Down

Body_Motion is Vertical_Down

Object is Ball_Like

Body_Motion is Vertical_Down

Object is Big_Ball_Like

Body_Motion is Spinning

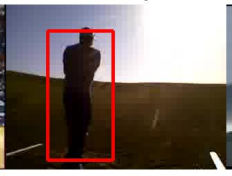
Posture is Standing

Posture is Sitting

Body_Parts_Used is Legs

Body_Motion is Jumping_Up

Posture is Standing



Body_Parts_Used is Hands

Body_Parts_Used is Hands

Body_Parts_Used is Legs

Body_Motion is Jumping_Up

Body_Parts_Used is Arms

Posture is Standing

Body_Motion is Running

Body_Parts_Used is Hands

Body_Motion is Bending

Posture is Standing

References

[1] D. Tran, J. Yuan, and D. Forsyth. Video event detection: From subvolume localization to spatio-temporal path search. 2014. 1