Gabriella
MEVA System
University of Central Florida (UCF)
Problem

Untrimmed security videos
  • Detect activities
    • Human and vehicles
    • Indoor and outdoor
  • Types
    • Single actors
    • Interaction between actors
    • Actor-object interactions
  • Spatio-temporal localization
    • Start/end
    • Spatial extent
Challenges

- Untrimmed nature
- Multiple activities
- Varying length of activities
- Multiple actors
- Interactions
  - Actor-object
  - Actor-actor
Challenges

- Untrimmed nature
- Multiple activities
- Varying length of activities
- Multiple actors
- Interactions
  - Actor-object
  - Actor-actor
- Multiple scales
  - Tiny activities
Challenges

- Untrimmed nature
- Multiple activities
- Varying length of activities
- Multiple actors
- Interactions
  - Actor-object
  - Actor-actor
- Multiple scales
  - Tiny activities
- Unbalanced dataset
Challenges

- Untrimmed nature
- Multiple activities
- Varying length of activities
- Multiple actors
- Interactions
  - Actor-object
  - Actor-actor
- Multiple scales
  - Tiny activities
- Unbalanced dataset
  - Sparse activities
Motivations

• Region proposal based approach [1, 2]
  • Scaling issue with videos
  • Multiple actors
    • How to pair?

• Object detection [3]
  • Time consuming
  • Multiple actors
    • How to pair?

Approach

• A two-stage process
  • Detect activity tubelets from long untrimmed videos
  • Recognize activities in the detected tubelets

• Encoder-decoder architecture
  • No region proposal

• Video level detection
  • No object detection

• RGB video
  • No optical flow
Architecture
Architecture
Architecture
Foreground scale imbalance

• Dice loss
  • Imbalanced pixel-wise classification
  • Focuses on ratio of foreground to background regions.
Patch-Dice loss

Generalized Dice Loss:

\[ \mathcal{L}_{GD} = 1 - \frac{2 \sum_{i=1}^{N} p_i \hat{p}_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} \hat{p}_i^2 + \epsilon} \]

Patch-Dice Loss:

\[ \mathcal{L}_{PDL} = \sum_{k=1}^{K} \left( 1 - \frac{2 \sum_{i=1}^{M} p_{ki} \hat{p}_{ki}}{\sum_{i=1}^{M} p_{ki}^2 + \sum_{i=1}^{M} \hat{p}_{ki}^2 + \epsilon} \right) \]

- N is the total number of pixels in the frame.
- K is the total number of patches.
- M is the total number of pixels in a patch.
Multiscale Patch-Dice Loss

\[ \mathcal{L}_{MPDL} = \lambda_1 \mathcal{L}_{PDL}(3 \times 3) + \lambda_2 \mathcal{L}_{PDL}(5 \times 5) + \lambda_3 \mathcal{L}_{PDL}(7 \times 7) \]

- (3 x 3), (5 x 5), (7 x 7) are the patch sizes.
Architecture
Sample activity tubelets
Sample activity tubelets
Architecture
Architecture
Architecture
Architecture
Architecture
Real-time system

- **Main process**
  - Manages the GPU threads
  - Perform task distribution.
Task Distribution

- Main process
- Initialize models
- Distributes videos into GPU threads
- Each GPU has a copy of all the models
Frame Reader

• Reads frame into memory
  • Without extraction

• Failsafe system
  • Two readers for codec problems.
    • FFMPEG reader
    • OpenCV reader
Pre-Processing

• Pre-processing is very slow on CPU
  • Bottleneck: resize operation
• Transferring frames to GPU
  • memory intensive
• **Solution:** 3-stage buffering mechanism for GPU
  • High resolution frame buffer
  • Low resolution frame buffer for localization network
  • Tube proposal buffer for action classifier
Buffering

HR Buffer

Resize

LR Buffer

Crop

Tube Coordinates

Localization Network

Tube Buffer

Action Classifier

GPU Thread

Frame Reader

Pre-Processing

Buffering

Pipeline-Models

Post-Processing
Evaluation

• Speed
  • ~100 frames per second

• Performance
  • Best performing system on the leaderboard
  • NIST will present
Sample detections

- *person_talks_to_person*
- *vehicle_turns_left*
- *vehicle_reverses*
Sample detections

vehicle_turns_right
Limitations

• Two stages
  • Need of connected components
  • Can we merge these two stages?

• Close by multiple instances
  • Single action tube
  • Instance separation
References

Acknowledgement

• Source of annotations
  • University of Maryland (UMD)
    • Professor Rama Chellappa and group
  • Carnegie Mellon University (CMU)
    • Professor Alexander Hauptmann and group
  • Kitware
  • IBM

• Evaluation
  • NIST – Jonathan Fiscus and team

• Funding - IARPA
Thank You!

Q & A