Complex Pairwise Activity Analysis via Instance Level Evolution Reasoning

Sudipta Paul[1], Carlos Torres[2], Shivkumar Chandrasekaran[3], Amit K. Roy-Chowdhury[1]

[1]University of California, Riverside
[2]TwoSixLabs, LLC
[3]Mayachitra, Inc

Objective

- ❑ Video activity analysis
- ❑ Object level activity detection
- ❑ **Pairwise** and **unary** activity analysis

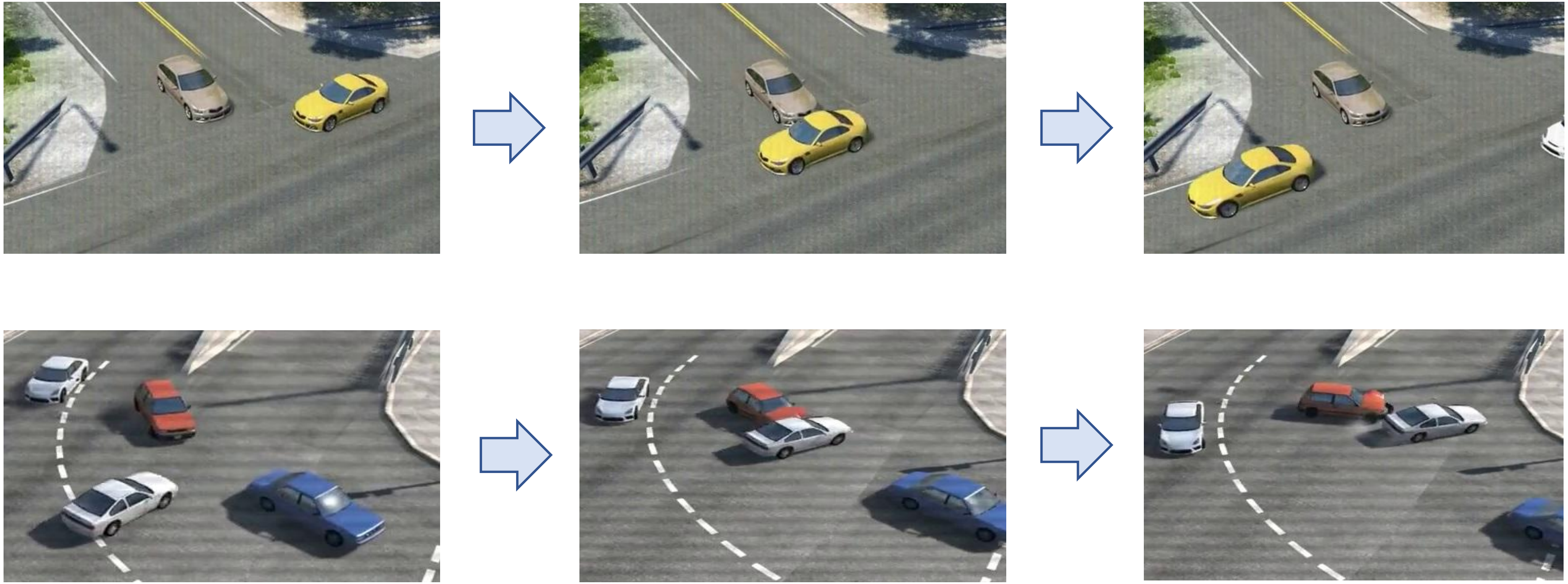
Understanding Activity Involving Multiple Objects

<u>Requirement:</u>

- ❑ Understanding spatial interaction among objects

Understanding Activity Involving Multiple Objects

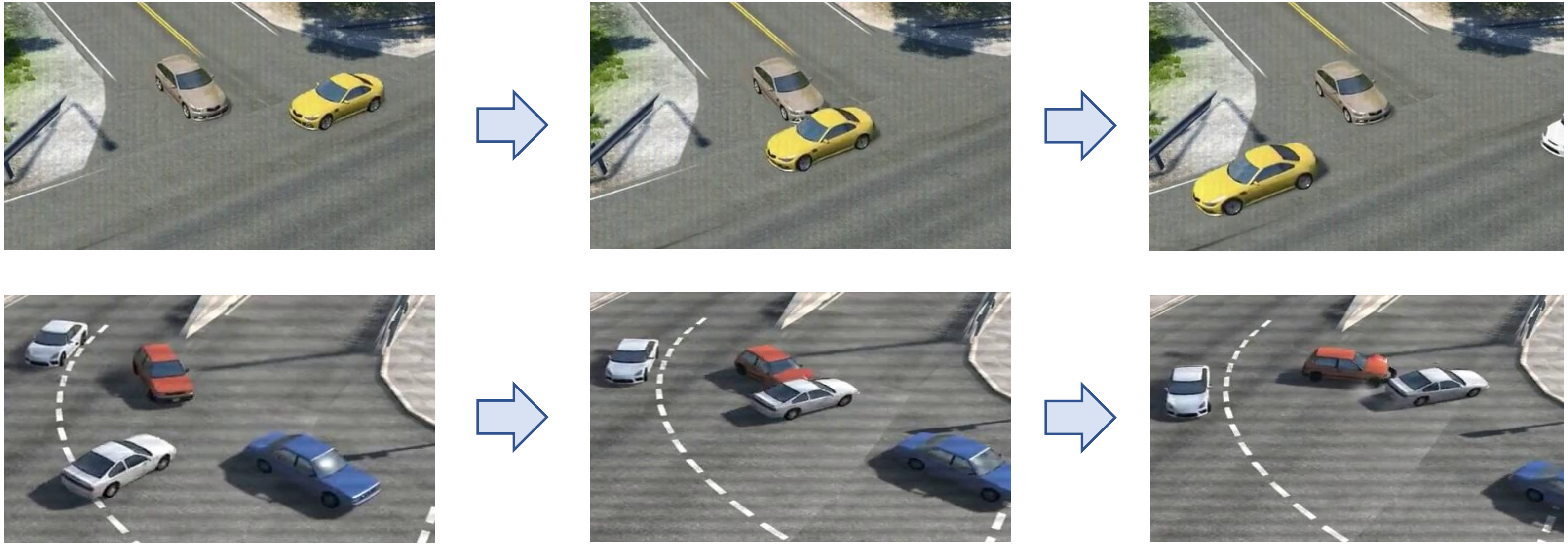
- However, spatial interaction is not enough



Understanding Activity Involving Multiple Objects

<u>Requirements:</u>

- ❑ Reason on spatial object interaction
- ❑ Reason on short-term and long-term evolution of the dynamics of objects with time



Application

Detection of multiple objects involvement in an activity can be useful for:

- ❑ Surveillance and urban planning applications
- ❑ Detection of objects involved in a collision is important for an autonomous driving application

Contribution

- ❑ Proposed a novel method that reasons about the relation of the interacting objects and captures the temporal dynamics of objects to boost activity recognition performance

- ❑ Addressed a novel problem of activity recognition for pairs of objects and introduce a new dataset **CarBump**, containing synthetic videos of car collision events with pairwise activity annotation

- ❑ Empirically showed the activity recognition performance on the CarBump dataset

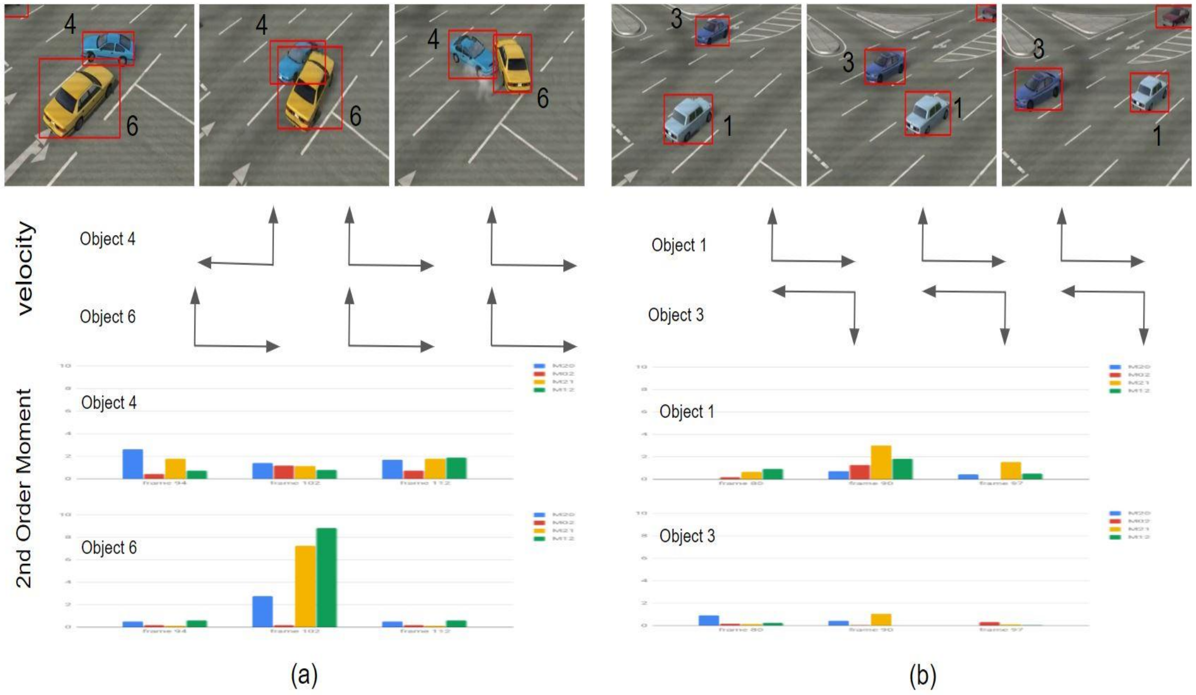
Methodology

- ❑ Feature Extraction
- ❑ Temporal Reasoning
- ❑ Spatial Reasoning

Feature Extraction

For each of the detected object, we compute:

- ❑ Centroid,
- ❑ Bounding box information,
- ❑ Change of speed,
- ❑ Second order moment of area

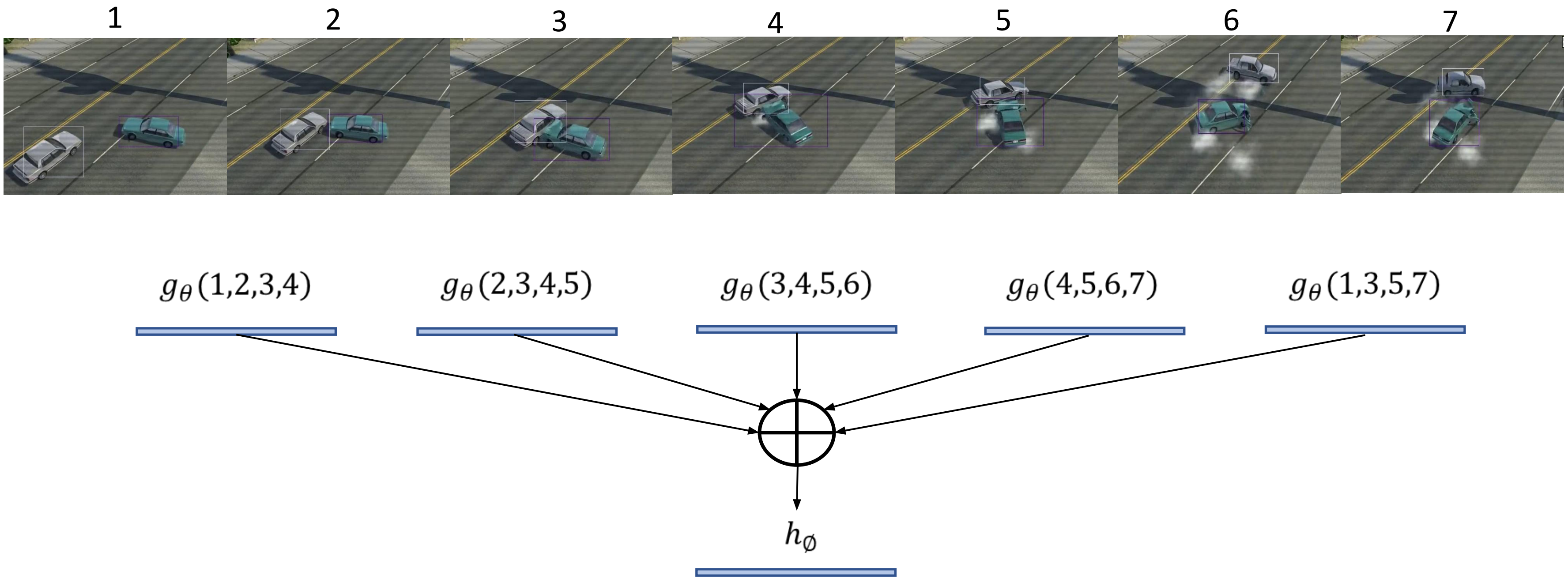


Methodology

Temporal Reasoning

- ❑ Composite function of two neural networks
- ❑ For each object, samples temporally ordered features from four frames
- ❑ From a set of seven frames, sample five combinations

$$TR(\mathcal{O}) = h_{\phi_t}\Big(\sum_{i<j<k<l} g_{\theta_t}(\boldsymbol{x}_i^o, \boldsymbol{x}_j^o, \boldsymbol{x}_k^o, \boldsymbol{x}_l^o)\Big).$$



Methodology

Spatial Reasoning

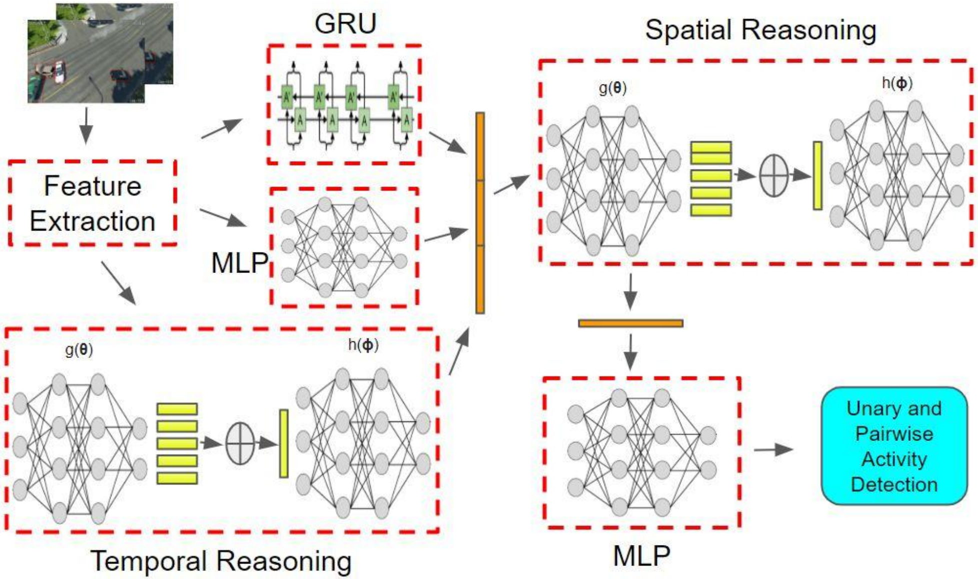
- Also a composite function of two neural networks

- Proven to be effective for relational reasoning*

- Inputs are the combined feature representation of each object, extracted using a MLP, a bidirectional GRU and temporal reasoning unit

- Outputs are object representation with spatial relational encoding

$$SR(\mathcal{F}) = h_{\phi_s}(\sum_{i,j} g_{\theta_s}(\boldsymbol{x}_i^T, \boldsymbol{x}_j^T))$$



*A simple neural network module for relational reasoning, NIPS 2017

Dataset

- Introduced CarBump Dataset
- Contains 141 synthetic videos of car collision
- Each video clip is at least 5 second long
- Annotation includes detected object information, tubelets and activity information for each frame.
- Activity annotation includes individual activity and pairwise activity
- Object IDs are consistent with COCO-2014

Results

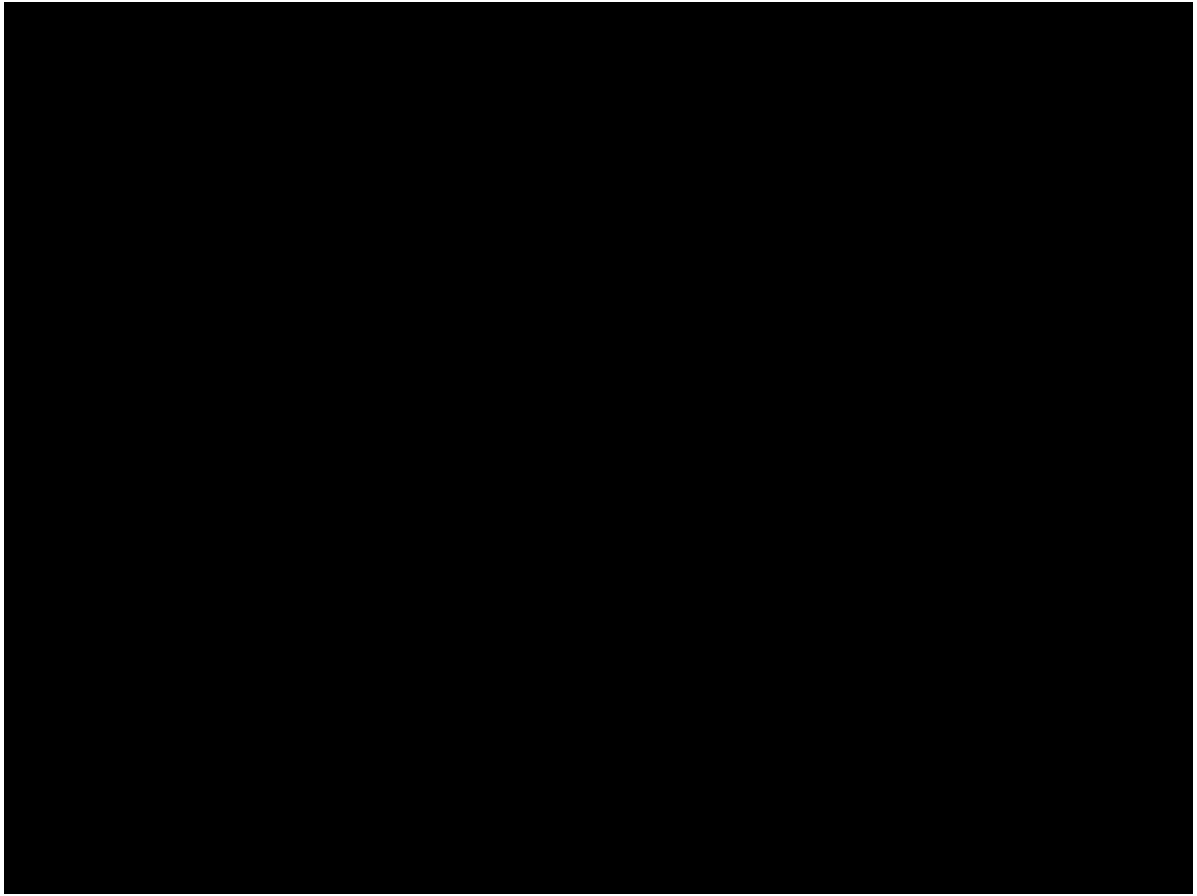
Approach	Unary (mAP)	Pairwise + Unary (mAP)
RGB	0.30	0.007
RGB + FLOW	0.18	0.058
Proposed	0.45	0.42

RGB: used pooled I3D feature from RGB stream to classify activity

RGB+FLOW: used pooled I3D feature from RGB stream and Flow stream to classify activity

- Significant performance improvement for pairwise activity detection

Examples



Conclusion

- Instance level spatial and temporal reasoning to boost activity recognition performance.
- It enables the system to distinguish between insufficient appearance variance to detect collision events.
- Introduced **CarBump** dataset will promote complex activity analysis using relational reasoning.

Future Work

- Extensions to applications in natural videos
- Extensions to multiple activities and multi-object involvement analysis
- Utilization of scene context information to optimize the spatio-temporal search space and improve performance

Thank You!

For any query contact:

Sudipta Paul
spaul007@ucr.edu

Carlos Torres
carlitos408@gmail.com