REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
26-04-2023	Final Report	1-May-2021 - 15-Jun-2022

4. TITLE AND SUBTITLE
Final Report: Dynamic Scene Graphs for Extracting Activity-based Intelligence

5a. CONTRACT NUMBER
W911NF-21-1-0236

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Rochester Institute of Technology
141 Lomb Memorial Drive

Rochester, NY 14623 -5603

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
78672-MI.9

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Yu Kong
					19b. TELEPHONE NUMBER
					517-355-1855

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 27-Apr-2023

Agency Code: 21XD

Proposal Number: 78672MI
Agreement Number: W911NF-21-1-0236

INVESTIGATOR(S):

Name: Qi Yu
Email: qi.yu@rit.edu
Phone Number: 5854756929
Principal: N

Name: Yu Kong
Email: yukong@msu.edu
Phone Number: 5173551855
Principal: Y

Organization: **Rochester Institute of Technology**
Address: 141 Lomb Memorial Drive, Rochester, NY 146235603
Country: USA
DUNS Number: 002223642
EIN:

Report Date: 15-Sep-2022
Date Received: 26-Apr-2023
Final Report for Period Beginning 01-May-2021 and Ending 15-Jun-2022
Title: Dynamic Scene Graphs for Extracting Activity-based Intelligence
Begin Performance Period: 01-May-2021
End Performance Period: 15-Jun-2022
Report Term: 0-Other
Submitted By: Yu Kong
Email: yukong@msu.edu
Phone: (517) 355-1855
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees:
STEM Participants:

Major Goals: Artificial intelligence and machine learning research have made significant contributions that exploit machine intelligence to support Activity-based Intelligence (ABI) extraction in many military operations. However, ABI extraction in open-world scenarios poses a variety of technical difficulties that have not been fully addressed. For example, as the battlefield is a complex and highly dynamic environment, novel objects and events will frequently appear and it is necessary to understand these novelties. In addition, while good accuracy can often be achieved using state-of-the-art deep learning-based approaches, their poor explainability makes the extracted information hard to be used by the human-decision team. Furthermore, it is also challenging to incorporate human knowledge to further enhance the performance of machine learning approaches. This pioneering research project will build a comprehensive framework to support ABI extraction by collectively addressing the following specific technical challenges:

1) The proposed framework will be able to analyze large-scale data of heterogeneous types as data from different sources may provide complementary information that is instrumental to improve the overall accuracy of ABI.
2) It is important to recognize and detect objects in open-set scenarios as it is very likely to observe novel objects in practical scenarios. Novel objects may raise an alarm to the human decision team as the team have no knowledge about the objects.
3) Spatial interactions of elements in the event data as well as temporal dynamics of the event need to be modeled in order to better capture complex relationships for extracting ABI. Such relationships will enable anomaly event detection before it happens.
4) The proposed models for extracting ABI should be adaptive to a new environment with limited training data as the battlefield scenarios are dynamically changing.

The primary objective of this project is to explore a dynamic learning framework that effectively extracts ABI in various highly complicated military operations. We plan to develop Dynamic Scene Graphs (DSGs) over large-scale multimodal time series data for representation learning. The new representations will enable learning from complex and dynamic environments, where a variety of vision tasks can be achieved including open-set object classification and detection, event detection, question-answering, and dense captioning. The proposed framework will achieve the following three properties:

(i) Combining the strength of multimodal data for learning effective representations.
(ii) Enabling open-set object classification and detection by learning the boundary of data distributions. This allows the proposed framework to understand novel objects in open-world scenarios.
(iii) Effective event detection and model adaptation from limited samples. The proposed event detection model is practical as it can be efficiently adapted to novel events using few training samples. All of these properties enable the extraction of ABI to support military operations.

Accomplishments: Research activity I: Open-set Classification (ICCV 2021)
We have developed an Open-set Classification approach. In military operations, it is very likely for a system to observe samples from classes not seen during training. Our approach can handle that and label these classes as "unknown". To enable the model to know the unknowns, our method formulates it as an uncertainty estimation problem by leveraging evidential deep learning (EDL). EDL utilizes deep neural networks to predict a Dirichlet distribution of class probabilities, which can be regarded as an evidence collection process. The learned evidence is informative to quantify the predictive uncertainty of diverse classes so that unknown classes would incur high uncertainty, i.e., the model knows the unknown. Furthermore, to overcome the potential over-fitting risk of EDL in a closed set, we propose a novel model calibration method to regularize the evidential learning process. Besides, to mitigate the static bias problem for the data classes, we propose a plug-and-play module to debias the learned representation through contrastive learning. Experimental results show that our method boosts the performance of existing powerful models, while still maintains a high performance in traditional closed set recognition setting.

Specific Objectives: 1) performing open set classification using novel deep evidential learning; 2) efficient uncertainty evaluation for understanding the unknowns; 3) effectively mitigating over-confident predictions using evidential uncertainty calibration; 4) mitigating static bias problem using contrastive evidential debiasing.

Research activity II: Event Detection (CVPR 2022)
We have developed a video detection approach that can detects visual events in streaming videos. This is important for military scenarios where anomaly events, actions, and activities should be detected. A key challenge in this work is that we do not have access to the future, and have to solely rely on history frames to make predictions. Therefore, it is important to accentuate parts of the history that are more informative to the prediction of the current frame. We present GateHUB, Gated History Unit with Background Suppression, that comprises a novel position-guided gated cross-attention mechanism to enhance or suppress parts of the history as per how informative they are for current frame prediction. GateHUB further proposes a Future-augmented History component to make history features more informative. GateHUB also introduces a background suppression objective to further mitigate false positive background frames that closely resemble the action frames. Extensive experiments demonstrates that GateHUB significantly outperforms all existing methods and is also more efficient than the existing best work.

Specific objectives: 1) enhancing or suppressing parts of video history using a novel position-guided gated cross-attention; 2) enhancing history encoding using Future-augmented History module, which extracts features for a history frame using its subsequent observed frames; 3) mitigating the false positive prediction of background frames using a background suppression objective.

Research activity III: Event/Accident Anticipation (ICCV 2021)
We have developed an approach, DRIVE, for detecting an anomaly event at an early stage before it happens. This is critical in military operations because we are not only detecting these anomaly events, but we can also anticipate them. The DRIVE model simultaneously learns the policies of accident anticipation and fixation prediction based on a deep reinforcement learning (DRL) algorithm. At each time step, the agent takes actions to predict the occurrence probability of a future accident, as well as the fixation point indicating where drivers will look in the next time step. Our environment model dynamically provides the observation state by considering both the bottom-up and top-down visual attention, which is recurrently modulated by the actions from the previous time step. We develop a novel dense anticipation reward to encourage early and accurate prediction, as well as a sparse fixation reward to enable visual explanation. Moreover, to effectively train the DRIVE model on real-world datasets, substantial improvements are made based on the DRL algorithm SAC. Our method is demonstrated to be effective on the DADA-2000 dataset, and can be easily extended to the DAD dataset without fixation annotations.

Specific objectives: 1) anticipating anomaly events based on deep reinforcement learning (DRL); 2) visually explainable by explicitly simulating human visual attention within the DRL framework; 3) performing effective in training using the proposed dense anticipation reward and sparse fixation reward.

Research activity IV: Open-set Localization (CVPR 2022)
We, for the first time, step toward the Open Set Temporal Action Localization (OpenTAL) problem. OpenTAL provides a more practical means to localize anomaly events and can be potentially used in anomaly event detection in the military domain. We propose a general framework OpenTAL based on Evidential Deep Learning (EDL). Specifically, the OpenTAL consists of uncertainty-aware action classification, actionness prediction, and temporal location regression. With the proposed importance-balanced EDL method, classification uncertainty is learned by collecting categorical evidence majorly from important samples. To distinguish the unknown actions from background video frames, the actionness is learned by the positive-unlabeled learning. The classification uncertainty is further calibrated by leveraging the guidance from the temporal localization quality. The OpenTAL is general to enable existing TAL models for open set scenarios, and experimental results on THUMOS14 and ActivityNet1.3 benchmarks show the effectiveness of our method.

Specific objectives: 1) first attempt solving this challenging open-set temporal localization problem; 2) quantifying classification uncertainty using evidential deep learning; 3) selecting the top negative samples from the mixture of samples using positive-unlabeled learning; 4) calibrating uncertainty using the temporal Intersection-over-Union as the localization quality.

Research activity V: model adaptation (AAAI 2022)
We have developed a dynamic meta-learning model that allows a model to quickly adapt to new and sparse observations. In particular, we leverage recommender systems as a platform to evaluate the effectiveness of the proposed model with a focus on providing accurate recommendation to users with limited recent interactions. The key challenge lies in how to effectively capture their historical preference encoded through past interactions with the system while making adaptions according to the most recent but very limited (i.e., few-shot) interactions. The proposed model consists of two distinct modules: a meta-learning module and a recurrent module. The former aims to capture time-specific latent factors through limited interaction data by leveraging the shared knowledge learned from other users. The latter aims to capture time-evolving latent factors by nesting a recurrent neural network, and it can be jointly optimized with the meta-learning module through the model-agnostic meta-learning approach. Finally, we seamlessly integrate the two modules by merging the time-specific and time-evolving factors to form the user representation. This user representation further interacts with an item embedding (which is also optimized during model training) to provide the final recommendations.

Specific objectives: (1) detecting novel phenomena (e.g., new event or change of taste) from limited observations through meta-learning, (2) capturing the temporal evolution of a model by encoding important patterns from historical data, and (3) integrating a model that captures historical patterns and current observations for quick adaptation.

Training Opportunities: During this reporting period, PI Kong has been mentoring three PhD students, each of which is working on a different aspect of the project. These students are in different stages of their PhD studies, varying from second year to fourth year. Their research focuses cover all the major areas of this project. These topics will also be a major part of these PhD students' dissertation research. I expect one student will receive her PhD degree in the next 18 months, under the full or partial support of our ARO project.

The PI holds weekly one-on-one or small group meetings with each PhD student. The PhD students also attend the regular project meetings with the PI and Co-PIs that are held biweekly. In addition, the PhD students regularly attend and present at a machine learning special interest group organized by the PI.

Results Dissemination: Major results from this reporting period have been disseminated to the communities of interest through publications in top-tier computer vision and machine learning venues including, CVPR 2022, AAAI 2022, ICCV 2021. So far, seven papers have been published in the current reporting cycle and details are provided in the Accomplished section of the report. In addition to the publications, both oral and poster presentations were delivered at these venues. Four other papers are currently under submission.

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

RPPR Final Report
as of 27-Apr-2023

PARTICIPANTS:

Participant Type: PD/PI
Participant: Yu Kong
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Qi Yu
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Wentao Bao
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Junwen Chen
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Conference on Computer Vision (ICCV)
Date Received: 24-Aug-2021 Conference Date: 11-Oct-2021 Date Published:
Conference Location: Virtual
Paper Title: Evidential Deep Learning for Open Set Action Recognition
Authors: Wentao Bao, Qi Yu, Yu Kong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Conference on Computer Vision (ICCV)
Date Received: 24-Aug-2021 Conference Date: 11-Oct-2021 Date Published:
Conference Location: Virtual
Paper Title: DRIVE: Deep Reinforced Accident Anticipation with Visual Explanation
Authors: Wentao Bao, Qi Yu, and Yu Kong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Conference on Computer Vision (ICCV)
Date Received: 24-Aug-2021 Conference Date: 11-Oct-2021 Date Published:
Conference Location: Virtual
Paper Title: Explainable Video Entailment with Grounded Visual Evidence
Authors: Junwen Chen and Yu Kong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)
Date Received: 22-Jun-2022 Conference Date: 21-Jun-2022 Date Published:
Conference Location: New Orleans, Louisiana
Paper Title: OpenTAL: Towards Open Set Temporal Action Localization
Authors: Wentao Bao, Qi Yu, Yu Kong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)
Date Received: 22-Jun-2022 Conference Date: 21-Jun-2022 Date Published:
Conference Location: New Orleans, Louisiana
Paper Title: GateHUB: Gated History Unit with Background Suppression for Online Action Detection
Authors: Junwen Chen, Gaurav Mittal, Ye Yu, Yu Kong, Mei Chen
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The 32nd British Machine Vision Conference
Date Received: 22-Jun-2022 Conference Date: 22-Nov-2021 Date Published:
Conference Location: virtual
Paper Title: Gradient Frequency Modulation for Visually Explaining Video Understanding Models
Authors: Xinmiao Lin, Wentao Bao, Matthew Wright, Yu Kong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 36th AAAI Conference on Artificial Intelligence (AAAI)
Date Received: 22-Jun-2022 Conference Date: 22-Feb-2022 Date Published:
Conference Location: virtual
Paper Title: A Dynamic Meta-Learning Model for Time-Sensitive Cold-Start Recommendations
Authors: Krishna Prasad Neupane, Ervine Zheng, Yu Kong, Qi Yu
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Yu Kong
Signature Date: 4/26/23 7:53PM

Abstract

In this report, we describe our work progress within the 14 months of the project (05/2021-07/2022). During that period, our main work was to study Task 2 (open-set object classification and detection) and Task 3 (event detection and model adaptation). Our research shows that uncertainty estimation can be useful in open-set recognition. We have proposed two approaches for open-set recognition in the video domain based on the idea of uncertainty estimation. We have also found that modeling history frames and background frames are useful in event detection. In addition, we have found that integrating historical patterns and current observations in temporal data is the key for quick model adaptation. Our findings have been published on top-tier venues including CVPR, ICCV, and AAAI.

Objectives

1. Develop approaches for open-set classification and detection.
2. Develop approaches for event/activity detection.
3. Develop approaches for learning event models from few training data.

Findings

Objective 1: We have developed an open-set recognition approach [1] and an open-set temporal localization approach [2]. Both approaches provide us with required knowledge for developing open-set object detection approaches. Our **findings** are:

1. Uncertainty estimation can be a great tool for open-set recognition problem, and can be potentially extended to open-set object detection. We have developed an evidential neural network [1] that naturally captures model's predictive uncertainty. Our results show that our approach can overcome the over-confidence issue of the Softmax in open-set recognition. Results can be found in **Table 1**.
2. Static bias needs to be addressed in open-set recognition problem because it heavily hurts the performance. Therefore, we have proposed a new module to mitigate static bias in [1]. This module can further improve the performance by 6% using Open maF1 metric (HMDB-51 is used as the unknown). Results can be found in **Table 2** (row 2 vs row 4).
3. Our approach [2] significantly outperforms the baselines by large margins on all metrics for the open-set temporal localization problem, while still keeping comparable closed set temporal localization performance (less than 1% mAP decrease) when THUMOS14 or ActivityNet dataset is used as the unknown. Results can be found in **Table 3**.
4. In localization/detection problem, positive and negative samples are highly imbalanced. Also, actionness assessment is important for localization as well. Therefore, we have proposed an importance-balanced evidential learning approach [2]. Comprehensive ablation studies demonstrate the effectiveness of these components. Results can be found in **Table 4**.

Table 1: Comparison with state-of-the-art methods. Models are trained on the closed set UCF-101 and tested on two different open sets where the samples of unknown class are from HMDB-51 and MiT-v2, respectively. For Open maF1 scores, both the mean and standard deviation of 10 random trials of unknown class selection are reported. Closed set accuracy is for reference only.

Models	OSAR Methods	UCF-101 [54] + HMDB-51 [31]		UCF-101 [54] + MiT-v2 [39]		Closed Set Accuracy (%)
		Open maF1 (%)	Open Set AUC (%)	Open maF1 (%)	Open Set AUC (%)	**(For reference only)**
I3D [8]	OpenMax [5]	67.85 ± 0.12	74.34	66.22 ± 0.16	77.76	56.60
	MC Dropout	71.13 ± 0.15	75.07	68.11 ± 0.20	79.14	94.11
	BNN SVI [27]	71.57 ± 0.17	74.66	68.65 ± 0.21	79.50	93.89
	SoftMax	73.19 ± 0.17	75.68	68.84 ± 0.23	79.94	94.11
	RPL [10]	71.48 ± 0.15	75.20	68.11 ± 0.20	79.16	94.26
	DEAR (ours)	**77.24** ± 0.18	**77.08**	**69.98** ± 0.23	**81.54**	93.89
TSM [35]	OpenMax [5]	74.17 ± 0.17	77.07	**71.81** ± 0.20	83.05	65.48
	MC Dropout	71.52 ± 0.18	73.85	65.32 ± 0.25	78.35	95.06
	BNN SVI [27]	69.11 ± 0.16	73.42	64.28 ± 0.23	77.39	94.71
	SoftMax	78.27 ± 0.20	77.99	71.68 ± 0.27	82.38	95.03
	RPL [10]	69.34 ± 0.17	73.62	63.92 ± 0.25	77.28	95.59
	DEAR (ours)	**84.69** ± 0.20	**78.65**	70.15 ± 0.30	**83.92**	94.48
SlowFast [14]	OpenMax [5]	73.57 ± 0.10	78.76	72.48 ± 0.12	80.62	62.09
	MC Dropout	70.55 ± 0.14	75.41	67.53 ± 0.17	78.49	96.75
	BNN SVI [27]	69.19 ± 0.13	74.78	65.22 ± 0.21	77.39	96.43
	SoftMax	78.04 ± 0.16	79.16	74.42 ± 0.22	82.88	96.70
	RPL [10]	68.32 ± 0.13	74.23	66.33 ± 0.17	77.42	96.93
	DEAR (ours)	**85.48** ± 0.19	**82.94**	**77.28** ± 0.26	**86.99**	96.48
TPN [61]	OpenMax [5]	65.27 ± 0.09	74.12	64.80 ± 0.10	76.26	53.24
	MC Dropout	68.45 ± 0.12	74.13	65.77 ± 0.17	77.76	95.43
	BNN SVI [27]	63.81 ± 0.11	72.68	61.40 ± 0.15	75.32	94.61
	SoftMax	76.23 ± 0.14	77.97	70.82 ± 0.21	81.35	95.51
	RPL [10]	70.31 ± 0.13	75.32	66.21 ± 0.21	78.21	95.48
	DEAR (ours)	**81.79** ± 0.15	**79.23**	**71.18** ± 0.23	**81.80**	96.30

Table 2: Ablation studies of our approach. HMDB-51 is used as the unknown. Best results are shown in bold.

$\mathcal{L}_{EUC}$	CED	Joint Train	Open maF1 (%)	OS-AUC (%)
✗	✗	✓	74.95 ± 0.18	77.12
✓	✗	✓	75.88 ± 0.16	77.49
✓	✓	✗	81.18 ± 0.15	79.02
✓	✓	✓	**81.79 ± 0.15**	**79.23**

Table 3. Open-set temporal localization results (%). Models trained on the THUMOS14 closed set are tested on the open sets by including the unknown classes from THUMOS14 and ActivityNet1.3, respectively. The mAP is provided as the reference of the TAL results on THUMOS14 closed set.

Methods	THUMOS14 as the Unknown				ActivityNet1.3 as the Unknown				mAP
	FAR@95 (↓)	AUROC	AUPR	OSDR	FAR@95 (↓)	AUROC	AUPR	OSDR	
SoftMax	85.58	54.70	31.85	23.40	85.05	56.97	53.54	27.63	55.81
OpenMax [6]	90.34	53.26	33.17	13.66	91.36	51.24	54.88	15.73	36.36
EDL [4]	81.42	64.05	40.05	36.26	84.01	62.82	53.97	38.56	52.24
OpenTAL	**70.96**	**78.33**	**58.62**	**42.91**	**63.11**	**82.97**	**80.41**	**50.49**	55.02

Table 4. Ablation Results (%). The proposed EDL re-weighting method (MIB), the actionness prediction (ACT), and the IoUC loss are individually ablated from the open-set temporal localization.

Variants	MIB	ACT	IoUC	FAR@95 (↓)	AUROC	AUPR	OSDR
(1)		✓	✓	77.20	76.41	56.65	12.10
(2)	✓		✓	82.85	58.12	31.80	37.89
(3)	✓	✓		79.64	62.73	37.86	39.39
OpenTAL	✓	✓	✓	**70.96**	**78.33**	**58.62**	**42.91**

Publications

[1] Wentao Bao, Qi Yu, and **Yu Kong**. Evidential Deep Learning for Open Set Action Recognition. ICCV 2021

[2] Wentao Bao, Qi Yu, and **Yu Kong**. OpenTAL: Towards Open Set Temporal Action Localization. CVPR 2022

Objective 2: We have developed an event detection approach [3] and an anomaly anticipation approach [4]. Both approaches provide us with required knowledge for developing event and detection approaches. Our **findings** are:

1. Cross-attention model is powerful in assessing the contributions of history frames. We have further extended this model in [3] by adding a gating scheme. Also, background frames needs to be suppressed in event detection. We have proposed a novel approach that combines the two capabilities. Our approach shows an improved performance of about 3% in mAP over state-of-the-art methods on THUMOS'14 dataset. Results are shown in **Table 5**.
2. Our approach [3] does not sacrifice speed for accuracy. Our overall inference speed is almost the same as state-of-the-art method LSTR, but our mAP is higher than LSTR. Results are shown in **Table 6**.
3. The proposed Gate History Unit in our approach [3] helps detect events more accurately. Please refer to **Figure 1** for visualization.
4. Deep reinforcement learning (DRL) is a nice model for anomaly event anticipation. We have proposed a novel approach [4] with dense anticipation reward and sparse fixation reward, which has shown effective results in our experiment. Results are shown in **Table 7**.
5. Model's attention is important when anticipating future anomaly event. Therefore, we have proposed a *dynamic* bottom-up and top-down visual attention fusion (DAF) in [4]. This module shows significant performance improvement, and also enable visual explanation. Results in **Table 8** demonstrates the effectiveness of DAF over the static attention fusion (SAF).

Table 5. Online action detection results on THUMOS'14 comparing GateHUB with SoTA methods on mAP (%) when the RGB-based features are extracted from either TSN or TimeSformer. Optical flow-based features are extracted from TSN in all settings.

Method	Feature Backbone		THUMOS14
	RGB	Optical Flow	mAP (%)
FATS [26]	TSN	TSN	59.0
IDN [15]			60.3
TRN [50]			62.1
PKD [54]			64.5
OadTR [48]			65.2
WOAD [20]			67.1
LSTR [51]			69.5
GateHUB (Ours)			**70.7**
TRN [50]	TimeSformer	TSN	68.5
OadTR [48]			65.5
LSTR [51]			69.6
GateHUB (Ours)			**72.5**

Table 6. Efficiency comparison of GateHUB using RGB and optical flow features and our optical flow-free version with existing methods. GateHUB using RGB and optical flow has the least parameter count compared to existing methods, and higher accuracy and lower GFLOPs than the existing best method. Moreover, our flow-free version attains higher or close accuracy compared to existing methods that require RGB and optical flow features at 2.8× faster inference speed.

Method	Model		Inference Speed (FPS)					mAP(%)
	Parameter Count	GFLOPs	Optical Flow Computation	RGB Feature Extraction	Flow Feature Extraction	Model	Overall	
TRN [52]	402.9M	1.46	8.1	70.5	14.6	123.3	8.1	62.1
OadTR [48]	75.8M	2.54	8.1	70.5	14.6	110.0	8.1	65.2
LSTR [51](Flow-free)	54.2M	6.33	-	22.7	-	99.2	22.7	63.5
LSTR [51]	58.0M	7.53	8.1	70.5	14.6	91.6	8.1	69.5
Ours (Flow-free)	41.8M	3.47	-	22.7	-	83.3	**22.7**	66.5
Ours	45.2M	6.98	8.1	70.5	14.6	71.2	8.1	**70.7**

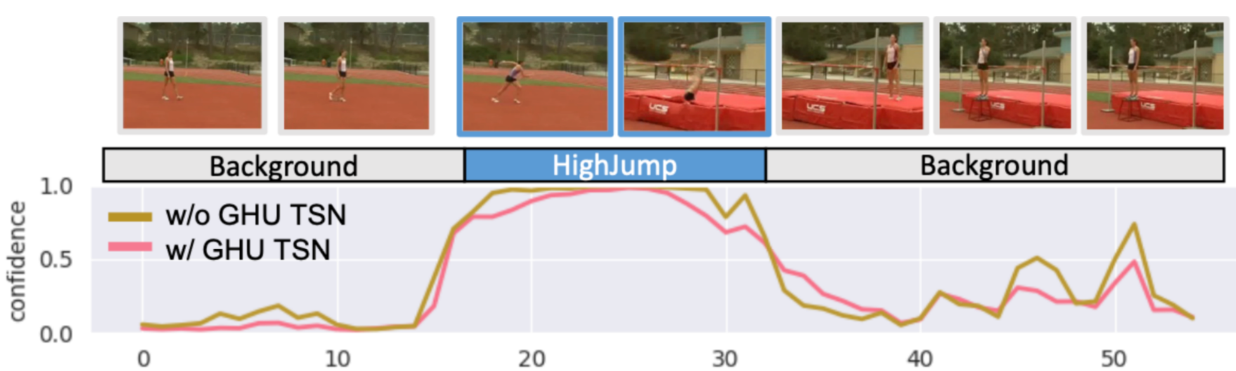

Figure 1. Visualization of our approach's online prediction. The curves indicate the predicted confidence of the ground-truth class (High Jump) using TSN backbone with and without GHU.

Table 7: Comparison with state-of-the-art methods. Best results are marked with bold fonts. AUC and TTA evaluate the correctness and earliness of accident anticipation, respectively.

Methods	DADA-2000 [11]		DAD [3]	
	AUC (%)	TTA (s)	AUC (%)	TTA (s)
DSA-RNN [3]	47.19	3.095	71.57	1.169
AdaLEA [40]	55.05	**3.890**	58.06	2.228
UString [2]	60.19	3.849	65.96	0.915
DRIVE (ours)	**72.27**	3.657	**93.82**	**2.781**

Table 8: Evaluation of visual attention and accident anticipation on DADA-2000. Best results are shown in bold. The parameters (Params) represent the values of ρ for SAF and m for DAF.

Params	Methods	AUC	SIM	CC	KLD (↓)
0.5	SAF	0.645	0.188	0.322	2.679
	DAF	**0.659**	**0.192**	**0.331**	**2.654**
0.8	SAF	0.691	0.144	0.190	3.087
	DAF	**0.726**	**0.158**	**0.226**	**2.986**
1.0	SAF	0.632	0.080	0.079	12.948
	DAF	**0.679**	**0.112**	**0.143**	**7.836**

Publications

[3] Junwen Chen, Gaurav Mittal, Ye Yu, **Yu Kong**, and Mei Chen. GateHUB: Gated History Unit with Background Suppression for Online Action Detection. CVPR 2022

[4] Wentao Bao, Qi Yu, and **Yu Kong**. DRIVE: Deep Reinforced Accident Anticipation with Visual Explanation. ICCV 2021

Objective 3: We have developed an event detection approach [5]. Both approaches provide us with required knowledge for developing few-shot learning approaches. Our **findings** are:

1. Integrating historical patterns and current observations in temporal data is the key for quick model adaptation. Therefore, we have proposed a recurrent module to capture time-evolving latent factors for few-shot learning purposes.
2. Capturing limited interaction data from various data sources is useful for learning feature representation for few-shot learning. Inspired by that, we have proposed a meta-learning module to leverage shared knowledge among the sources.
3. Our novel approach combines the two components and outperforms state-of-the-art methods for few-shot learning in recommendation systems on three datasets, including Netflix, Last.fm, and MovieLens-1M. Results are shown in **Table 9**.

Table 9. Recommendation results.

Category	Model	Netflix		Last.fm		MovieLens-1M	
		RMSE	**NDCG**	**RMSE**	**NDCG**	**RMSE**	**NDCG**
MF	SVD++	0.9797±0.03	0.2915	1.7829±0.08	0.2882	1.0825±0.04	0.3023
Dynamic	timeSVD++	0.9538±0.06	0.3115	1.6912±0.11	0.2962	1.0483±0.03	0.3224
	CKF	0.9337±0.04	0.3130	1.5316±0.32	0.3018	1.0652±0.04	0.3151
	DPF	N/A	N/A	1.5227±0.43	0.3085	N/A	N/A
Deep Learning	Wide and Deep	0.9904±0.04	0.2864	1.7253±0.22	0.2727	1.1364±0.06	0.2932
	DeepFM	0.9811±0.03	0.2930	1.6815±0.21	0.2971	1.1723±0.05	0.2882
	DIEN	1.0345±0.04	0.2832	1.9225±0.26	0.2714	1.1872±0.14	0.2843
Graph	GC-MC	1.0760±0.03	0.2901	N/A	N/A	1.1704±0.08	0.2913
	NGCF	1.0321±0.03	0.3026	1.5612±0.23	0.2896	1.1216±0.05	0.3103
Sequential	Caser	1.0124±0.03	0.3101	1.5824±0.31	0.2931	1.1339±0.08	0.3012
	SASRec	N/A	0.3246	N/A	0.3103	N/A	0.3238
Meta-Learning	MeLU	0.9213±0.05	0.3232	1.2580±0.28	0.3122	1.0685±0.08	0.3214
	ML-ICS	0.9332±0.04	0.3173	1.2408±0.24	0.3142	1.0845±0.06	0.3244
Proposed	**Ours**	**0.8925±0.03**	**0.3472**	**1.2203±0.16**	**0.3385**	**0.9945±0.08**	**0.3351**

Publications

[5] Krishna Prasad Neupane, Ervine Zheng, **Yu Kong**, Qi Yu. A Dynamic Meta-Learning Model for Time-Sensitive Cold-Start Recommendations. AAAI 2022