Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

R&D Project Proposal

Survey on Video-based Anticipation for Anomaly Detection

Anna Rose Johny

Supervised by

Prof. Dr. Paul G. Plöger
M.Sc. Santosh Thoduka

May 2020

Abstract

The aim of this Research and Development is to conduct a comprehensive survey on the various learning methods used for anticipation and also a study of how these methods can be applied for the task of anomaly detection. The learning approaches can be supervised, semi-supervised or unsupervised learning. This survey covers all anticipation methods and focuses on their application for task of anomaly detection.

1 Introduction

This section describes the motivation behind doing this project, problem statement, project scope, related work and expected results.

1.1 Motivation

Anticipation is a process of predicting the future based on some inputs such as video clips. The prediction can be trajectory prediction, early action recognition, future frame prediction, and so on. Anticipation can be applied in many fields such as self-driving cars, human-robot interaction and anomaly detection. Consider the example of self-driving cars [2] operating in real-world traffic. In this case it is mandatory to evaluate the actions which are happening around in order to make a collision free travel. Here the anticipation can be used for predicting changing a lane, taking turns, pedestrians information and so on [4]. It is a challenging task, since the environments in which cars and autonomous robots operate are dynamic and the possible events are diverse. Here future positions of the pedestrians and other vehicles are predicted using Dynamic-Spatial-Attention (DSA) Recurrent Neural Network (RNN) [4]. A vision based sensor is provided with the data obtained by using this deep learning approach for anticipating the accidents. By providing the information to the vision based sensor, the self-driving ability of the cars can be improved [4]. Main challenge in the estimation of the exact position of the pedestrians comes with the difficulty in collecting the image features required for tracking [6]. The solution for this can be directly analysing the data observed from the dashboard camera. The anomalies in this case can be pedestrians crossing roads. In this approach two models are evaluated. One for detecting the motion of the pedestrians for video prediction and other for finding the action that can be performed by these pedestrians.

Another application is in the case of human-robot interaction [5], in which robots operate along with humans in the same field. This task requires high amounts of safety measures to be fulfilled. This anticipation is a kind of action prediction in which the robots will be able to observe the actions of the humans for understanding their behavior and predicting what they will do next. Consider the case

of patrolbot [13] in which Generative Adversarial Networks (GAN) are used for detecting anomalies. Firstly the robot is trained with images obtained from the objects seen previously. During the testing phase, the new objects are detected which were not part of the trained data and are considered as anomalies. In this case it can be a person walking through a corridor or an object placed on the floor can be anomaly. The nominal conditions are anticipated and if we see new things that does not match our anticipation is an indication of anomaly. This prediction helps the autonomous robot in operating with safety conditions by knowing the future prediction earlier.

The Figure 1 [15] depicts the task of anomaly detection in video frames. Figure depicted here shows the input in the form of video, predicted frames and their ground truth for normal and abnormal events [15]. The figure demonstrates a walking area in which frames are predicted more accurately for normal events, while for abnormal events such as two men fighting and bicycle intrusion, the predictions are blurred. This is an example of anticipation used for anomaly detection.

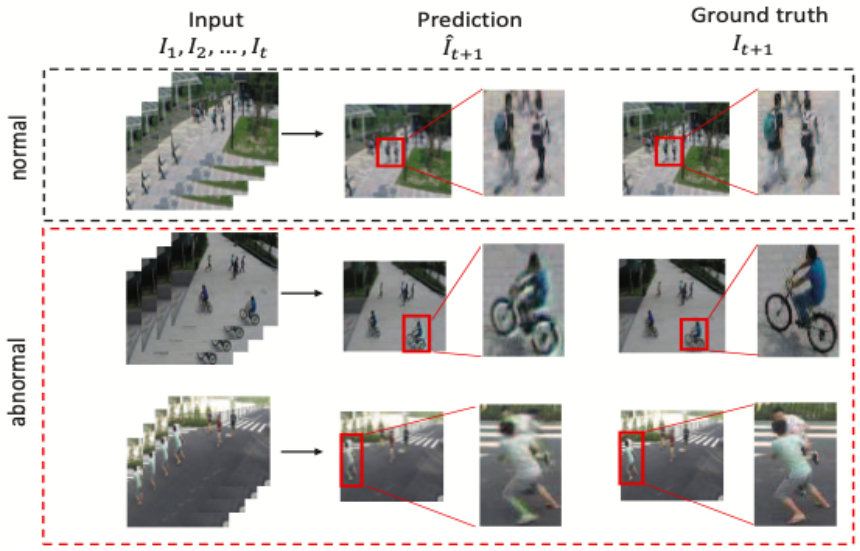

Figure 1: Anticipation used for anomaly detection

1.2 Problem Statement

The main intention of this R&D is to conduct a comprehensive survey on anticipation methods for anomaly detection. The first task is to identify the learning approaches such as supervised, semi-supervised and unsupervised learning based on the application for anticipation. The main objective is to identify how these approaches can be applied for detecting anomalies by providing different datasets as inputs. The task of anomaly detection is a problem because of the difficulty in identifying the unexpected behaviors which are identified as a change from normal behavior. Anomaly detection is important in the area of robotics because most of these systems require higher levels of autonomy. Identifying and correcting these anomalies are furthermore difficult. Another problem comes in choosing the datasets based on the application. All these problems are evaluated in this survey and the best approach is proposed finally based on the observations made.

1.3 Project Scope

This project shares some commonalities with the project *Comparative Evaluation of Generative Models for Future Frame Prediction*. In particular, both projects have a specific focus on the application of prediction or anticipation to video anomaly detection. In order to clarify the differences between the two projects, we define the scope of our project and the inputs and outputs which will be considered.

This project aims to cover anticipation in autonomous systems. As described earlier, anticipation can be expressed in the form of a video prediction [3], activities [1], human motion [7], driving maneuvers [10], etc. As illustrated in Figure 2, the types of inputs include video, human skeleton poses, various miscellaneous sensors such as GPS, robot dynamics, trajectories, etc. The anticipated outputs, whether video, motion, or activity labels, can be generated by both classical machine learning algorithms or deep learning algorithms.

In contrast, future frame prediction is focussed *only* on the prediction of future video frames with the input comprising of at least video frames, but potentially other sensor data as well.

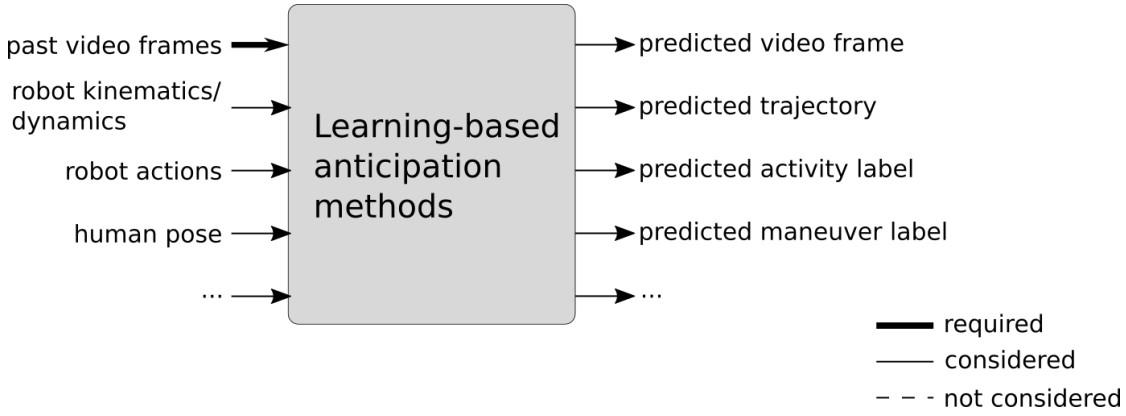

Figure 2: Scope of the project in terms of inputs and outputs considered

1.4 Related Work

This section describes the existing learning approaches with their challenges. Human action anticipation which predicts future actions have great importance in real-world applications such as self-driving and video surveillance systems. One of the approaches for action anticipation uses RNN models [2]. In this approach the early prediction is based on previous prediction inputs. This is an iterative approach and there are chances for error accumulation. This can affect the process of anticipation for long-term actions. Another approach uses anticipating future action in one-shot fashion [17]. This method firstly evaluates the temporal features and then provides initial anticipation. The final prediction is made based on a time-conditioned skip connection [17]. This method is more accurate and efficient compared to the previous approach. This approach is also suitable for long-term and short-term action anticipation. The prediction can also be applied for detecting anomalies on trajectories. One of the approaches uses a structured one-class support vector machine [6] for predicting the goals from trajectories and also detecting anomalies based on these predictions. This is intention-based anomaly detection for obtaining behaviors from trajectories.

Deep learning approaches such as Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) can be used for detecting the objects and anticipating the future frames. One of the approaches uses the Dynamic-Spatial-Attention RNN [4] model for anticipating accidents. In this approach model is able to anticipate

accidents with 56.14% precision and predictions are made 2s before the accident occurs [4]. This method fails for very complex traffic data sets. One of the methods used for long-term action anticipation [2] uses Convolutional Neural Networks (CNN) and RNN. This approach is suitable for providing accurate results for different datasets with videos having different lengths. But this method consumes more time and also requires more parameters for predicting long sequences.
For predicting the realistic pixel values in future frames, the model should possess the capability of capturing the motion changes and pixel-wise appearance, in order to give the previous frames as the inputs for producing new frames. The most of existing state-of-the-art-approaches in unsupervised learning uses the Generative Neural Networks(GANs) [12] for the video prediction, in which the RGB pixels are synthesized directly leading to blurry predictions. For reducing these effects, a Dual GAN architecture can be used [14], which consist of a future frame generator and future flow generator. This model is good in predictive learning, but this model needs to be improved for handling the real-world videos with complex motions involving sets. Another approach for video anomaly detection is to use Spatial Temporal AutoEncoder (STAE) [17]. This approach fits well for some applications and fails for complex scenarios.

1.5 Expected results

- To identify different approaches for anticipation tasks.
- Classification of the types of anticipation approaches used for anomaly detection.
- Selecting best approaches for task of anomaly detection..

2 Project Plan

2.1 Work Packages

WP1 : Literature survey on anticipation methods

- Finding different papers related to anticipation.
- Categorizing them into different types of anticipation such as trajectory planning, action recognition, future frame prediction.

WP2 : Selection of anticipation methods for the task of anomaly detection

- Identifying the papers for anomaly detection.
- Performing a search on how the existing approaches can be modified for different datasets.

WP3 : Project Report

- Documentation of the literature survey.
- Categorization of results.
- Final report.

2.2 Milestones

M1 : Literature search

M2 : Classification of data

M3 : Report submission

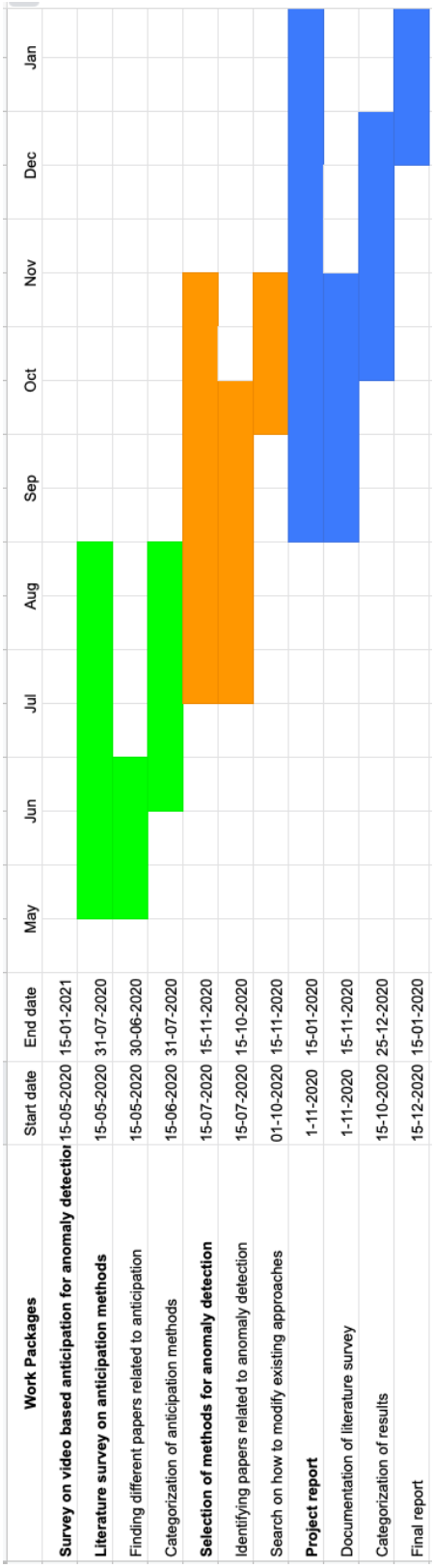

Figure 3: Work packages with timeline

2.3 Deliverables

1. Minimum Viable
    - Literature study on Anticipation methods.
    - Comparison of the above methods.
2. Expected
    - Comprehensive survey of the anticipation methods for anomaly detection.
    - Creating a list of most promising approach for anomaly detection.
3. Maximum
    - Implementation of any one of the approach in a test dataset.

References

[1] Yazan Abu Farha and Juergen Gall. Uncertainty-aware anticipation of activities. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[2] Yazan Abu Farha, Alexander Richard, and Juergen Gall. When will you do what? - anticipating temporal occurrences of activities. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. URL `http://openaccess.thecvf.com/content_cvpr_2018/html/Abu_Farha_When_Will_You_CVPR_2018_paper.html`.

[3] L. Castrejon, N. Ballas, and A. Courville. Improved conditional vrnns for video prediction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7607–7616, 2019.

[4] Fu-Hsiang Chan, Yu-Ting Chen, Yu Xiang, and Min Sun. Anticipating accidents in dashcam videos. In *Asian Conference on Computer Vision*, pages 136–153. Springer, 2016. URL `https://yuxng.github.io/chan_accv16.pdf`.

[5] Enric Galceran, Alexander G Cunningham, Ryan M Eustice, and Edwin Olson. Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction. In *Robotics: Science and Systems*, volume 1, 2015. URL `https://link.springer.com/article/10.1007/s10514-017-9619-z`.

[6] Pratik Gujjar and Richard Vaughan. Classifying pedestrian actions in advance using predicted video of urban driving scenes. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 2097–2103. IEEE, 2019. URL `http://autonomy.cs.sfu.ca/doc/gujjar_icra19.pdf`.

[7] Alejandro Hernandez, Jurgen Gall, and Francesc Moreno-Noguer. Human motion prediction via spatio-temporal inpainting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7134–7143, 2019.

[8] C. Huang and B. Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE International Conference on Human-*

Robot Interaction (HRI), pages 83–90, 2016. URL https://ieeexplore.ieee.org/document/7451737.

[9] Fan Hung, Xu Xie, Andrew Fuchs, Michael Walton, Siyuan Qi, Yixin Zhu, Doug Lange, and Song-Chun Zhu. Intention-based behavioral anomaly detection. 2019.

[10] Ashesh Jain, Avi Singh, Hema S Koppula, Shane Soh, and Ashutosh Saxena. Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3118–3125. IEEE, 2016.

[11] Qiuhong Ke, Mario Fritz, and Bernt Schiele. Time-conditioned action anticipation in one shot. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9925–9934, 2019.

[12] B Ravi Kiran, Dilip Mathew Thomas, and Ranjith Parakkal. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging*, 4(2):36, 2018.

[13] W. Lawson, E. Bekele, and K. Sullivan. Finding anomalies with generative adversarial networks for a patrolbot. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. URL https://ieeexplore.ieee.org/abstract/document/8014802?section=abstract.

[14] Xiaodan Liang, Lisa Lee, Wei Dai, and Eric P. Xing. Dual motion gan for future-flow embedded video prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. URL https://arxiv.org/abs/1708.00284.

[15] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6536–6545, 2018.

[16] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015. URL https://arxiv.org/abs/1511.05440.

[17] Yiru Zhao, Bing Deng, Chen Shen, Yao Liu, Hongtao Lu, and Xian-Sheng Hua. Spatio-temporal autoencoder for video anomaly detection. In *Proceedings of the 25th ACM International Conference on Multimedia*, page 1933–1941. Association for Computing Machinery, 2017. ISBN 9781450349062. doi: 10.1145/3123266.3123451. URL https://www.researchgate.net/publication/320543620_Spatio-Temporal_AutoEncoder_for_Video_Anomaly_Detection.