Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen
International Center for
Information Technology

R&D Project Proposal

Survey on Learning methods for video based anticipation

Anna Rose Johny

Supervised by

M.Sc. Santhosh Thoduka
Prof. Dr. Paul G Ploeger

May 2020

Abstract

This Research and Development project finds the different learning methods for future frame prediction with sequences of input frames given as input. This is a challenging task because video prediction requires large amounts of video data and high-dimensionality of video frames. This chalslenge can be addressed using various unsupervised learning techniques such as Variational Autoencoders(VAE), Generative Adversarial Networks(GAN), Variational Recurrent Neural Networks(VRNNs). This video prediction can be used for trajectory planning, future frame prediction, object recognition, human pose estimation, and activity prediction.

1 Introduction

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

1.1 Motivation

Machine learning algorithms mainly focus on improving the ability to learn without being programmed fully. This learning starts by providing some input data, which can be frame, patterns and then finding decisions of how to act based on these previous inputs. There are mainly three types of machine learning algorithms, supervised learning, unsupervised learning and semi-supervised learning. In supervised learning methods, training data is given to a machine(trains the model for some inputs), which creates a model and then checks the model for accuracy by providing the test data. In this approach, the observations made in the past are given to the new inputs in order to predict future frames. The main disadvantage of using supervised learning is that the models trained on one dataset cannot classify the objects that are not classified on the same dataset. If the input provided to the training the system is not classified or unlabeled, then unsupervised learning methods are used. These unsupervised learning methods are used in density estimation, clustering, anomaly detection,dimensionality reduction. Semi-supervised learning techniques use small amounts of labeled data and large amounts of unlabeled data with greater accuracy in learning. This Research and development project comparison of various learning approaches used for video prediction. After completing this project, it is able to identify the various approaches by which video prediction can be made.

1.2 Problem Statement

The main intention of this RD is to provide a literature survey on the unsupervised learning methods that can be applied for video prediction. By this survey, it is able to provide a clear idea about the methods by which video prediction can be performed. Unsupervised learning techniques have difficulties due to the usage of unlabeled data for prediction tasks, which is a challenging task. The main challenge associated with video prediction is that it requires large amounts of data, and also high dimensionality of video frames. The naturalistic and accurate video predictions are a problem, because the existing approaches predict the future frames for only a few frames. Another issue associated with video anticipation is blurry predictions. In this literature study, a comparison of the approaches that can be used for video prediction is performed and finally a best model that fits the video prediction is proposed.

1.3 Related Work

This section describes the existing learning approaches with their challenges.
Unsupervised learning approaches 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 uses the Generative Neural Networks(GANs)[1] 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[2], 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 Spatio Temporal AutoEncoder(STAE)[3], which utilizes the 3D convolutional networks for extracting the features from the video frames in spatial and temporal dimensions and further helps in reconstruction and future frame prediction. This approach fits well for some applications and fails for complex scenarios.
In this RD best approaches for the video prediction using various reinforcement learning methods will be identified.

1.4 Expected results

- To identify various learning methods that can be used for video prediction and also various video prediction methods.
- Classify them according to the field of application along with their advantages

and disadvantages.

- Finally finding a best approach that can be used for video prediction.

1.5 Work Packages

WP1 Literature Search

- Study of various learning methods.
- Study of various video prediction techniques.
- Comparison of the existing approaches.

WP2 Classification of approaches

- Classifying different approaches obtained before according to the different application fields.
- Formulating a general model.

WP3 Project Report

- Documentation of the literature survey.
- Documentation of comparison results
- Initial draft of the report.
- Final report.

1.6 Milestones

M1 Literature search

M2 Classification of data

M3 Formulating a general model

M4 Report submission

References

[1] 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.

[2] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015.

[3] 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*, MM '17, page 1933–1941, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349062. doi: 10.1145/3123266.3123451. URL https://doi.org/10.1145/3123266.3123451.