

Half-Yearly Progress Report for Jan-May 2019

Data Sheet for Ph.D Scholars

Name: Sonam Gupta

Registration No: CS18D005

Department: Computer Science and Engineering

Date of Joining: 16th July 2018

Date of Upgradation (if any):

Specialization / Stream: Computer Vision

Area of Research work: Future Frame Prediction in Video

Category of Admission: HTRA

Guide: Prof. Sukhendu Das

Co-Guide(s):

Date of DC meetings:

Description	Event	Date
1 st DC meeting	Comprehensive Viva	Attempt 1 : August 2019 Attempt 2 :
2 nd DC meeting	Research Proposal Seminar (1 st Seminar)	
3 rd DC meeting	Mid-Term Review DC meeting (3-3.5 years from the date of joining)	
4 th DC meeting	Research Colloquium (2 st Seminar)	
Six monthly DC meeting	After five years from the date of registration, upto maximum period of the programme	SMD 1: SMD 2: SMD 3: SMD 4:
Final DC meeting	Synopsis at Dean AR Office	

Details of Course Work

S.No	Course No	Course Title	Sem/Year	Credit	Grade
Core Courses					
1.	CS6015	Linear Algebra and Random Processes	1	12	S
2.	CS5691	Pattern Recognition and Machine Learning	1	15	B
Elective Courses					
1.	CS5020	Nonlinear Optimization: Theory and Algorithms	1	12	A
2.	CS5800	Advanced Data Structures & Algorithms	1	12	A
3.	CS6350	Computer Vision	2	12	S
4.	CS7015	Deep Learning	2	12	A
5.	CS6870	Digital Video Processing	2	12	B
Compulsory Courses / Optional Courses					
1.	ID6020	Introduction to Research (Institute Module)	1	0	Pass
2.	CS6021	Introduction to Research	1	0	Pass

Signature of Scholar

Signature of Guide

Signature of Co-Guide 1

Signature of Co-Guide 2

Contents

i) Title of Research Work

Probabilistic Video Frame prediction for Human Interaction Sequences

ii) Problem Definition / Research Objectives

Much work in video prediction tasks have been focused on generation of a deterministic future. The output of such models generally represents either a deterministic or the average of the possible futures, and is also blurred. Also, the future prediction is done only upto a few frames.

During my PhD research work, I will try to design new Deep Learning methods, built on top of existing models to address the above challenges in video prediction containing interaction of human with humans and or objects. In particular, I will work on the following aspects of the problem:

1. Stochastic Multiple frame prediction from a given set of frames.
2. Design of efficient objective function, integrating probabilistic criteria in feature space for the task of learning on a heterogenous deep CNN model.
3. Exploring appropriate evaluation metric for evaluating the results of multiple future predictions.
4. Creation of a tailormade Dataset for training and evaluating the stochastic model.

iii) Summary of Work Done before Review (From Date of Admission till now)

During the first semester, I worked mostly upon improving/gaining knowledge in foundation courses that will help in research. I took Linear Algebra and Random Processes, Non-Linear Optimization courses which in which I learnt about Matrices and their properties, Eigen values and Eigen Vectors, Singular Value Decomposition, Convex Sets, Convex Functions, Primal Dual Formulations. These mathematical tools will enable me to understand and formulate the mathematical explanations better. As Machine Learning and Deep Learning are prevalent in Computer Vision community, I took Pattern Recognition and Machine Learning course wherein I have learnt about Regression, Gaussian Mixture Models, Hidden Markov Models, Support Vector Machine through assignments and class lectures. Also, to brush up the Data Structures and Algorithms concepts, I took Advanced Data Structure and Algorithms course. The course covered advanced topics like Max Flow algorithms along with other Graph Algorithms. Many problems in domain of Computer Vision can be mapped to graphs. I believe that the fundamentals built by learning these courses will help me to better understand the literature in the area. With these four courses, I completed the coursework requirement of PhD.

In second semester I have primarily focussed upon building domain knowledge through courses and literature. I have taken three courses namely Deep Learning, Computer Vision and Digital Video Processing. These courses have taught me the concepts that will help me to understand the state of art methods in the literature.

iv) Work Done during Review (Even Semester 2019)

During the second semester, Computer Vision course exposed me to different problems that exists in literature like Depth Estimation, Segmentation, Object Detection and Recognition. I also learnt some of the possible ways in which these problems can be solved (partially or fully) using classical (Non-deep) methods. The course also covered fundamentals of Computer Vision and various handcrafted features like edges, corners, SIFT, SURF which have been a great success in the past and are still used successfully in many applications. As part of the course, I did a term project, in which I worked on the problem of finding the Depth from a monocular RGB Image. For this project, after following the literature, I have used an existing Deep Learning method based on

Transfer learning [1]. The methods followed an Encoder-Decoder based architecture, where both Encoder and Decoder were CNN based models. I also got to explore different loss functions for the same architecture.

The Deep learning course covered topics like Multi Layer Feed Forward Neural Networks, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long short term memory (LSTM), AutoEncoders. I implemented these concepts for tasks like Image classification, Function Approximation, Image Annotation, Image Captioning.

Digital Video Processing course, covered the seminal papers for Background Subtraction, Edge Detection, Corner Detection, Markov Random Fields based methods for Segmentation, Stereo Matching, Shape based matching etc. This course introduced me to various challenges like texture, illumination, scale, rotation that may come when proposing an algorithm.

Along with the courses, I also explored the current papers in some of the top conferences as CVPR, NIPS, ECCV. The problem of prediction of future frames caught my attention. At present, I am carrying out a literature survey [2, 3, 4, 5] for the same. Based on few papers that I have read during the semester, I have also applied for Google PhD Fellowship, 2019 the results for which will be declared in July.

v) Issues affecting Research Progress, if any

None

vi) Future Plans, with proposed timeline

For the next semester, I plan to do the following:

1. Carry out Literature survey for Future Frame Video Prediction.
2. Attempt Comprehensive Exam in August.
3. Define a concrete Research Problem
4. Propose a solution to the defined problem and submit to a conference if results obtained are up to the mark.
5. Take the course “Optimization methods for Computer Vision Applications” to learn more optimization techniques.

vii) Visible Research Output:

(a) Paper(s) Published in Journals

None

(b) Full Paper(s) Published in Conference Proceedings

None

(c) Seminars/Workshops/Conferences/Exchange programs attended and Papers Presented

- Attended 11th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2018).
- I will be attending the Computer Vision workshop at IIIT Hyderabad from July 1, 2019 to July 7, 2019. The summer school curriculum roughly consists of a series of lectures and demo/lab sessions designed to work in tandem. Experts talks will be delivered where they will share the view works with the attendees. Special focus will be given to recent advances in the area.

(d) Awards/Honours, if any

None

viii) References:

- [1] I. Alhashim, P. Wonka, “High Quality Monocular Depth Estimation via Transfer Learning by”, in arXiv preprint 1812.11941, 2019.
- [2] H. Cai and C. Bai, “Deep Video Generation, Prediction and Completion of Human Action Sequences,” in Proceedings of the European Conference on Computer Vision, 2018.
- [3] J. Xu, B. Ni and X. Yang, “Video Prediction via Selective Sampling,” in Advances in Neural Information Processing Systems, 2018.
- [4] P. Bhattacharjee and S. Das, “Predicting Video Frames using Feature Based Locally Guided Objectives,” in Asian Conference on Computer Vision, 2018.
- [5] P. Bhattacharjee and S. Das, “Temporal Coherency based Criteria for Predicting Video Frames using Deep Multi-stage Generative,” in Advances in Neural Information Processing Systems, 2017.