# DYNAMIC KERNELS AND SEMANTIC REPRESENTATIONS FOR RECOGNITION OF VARYING SIZE SCENE IMAGES

*A THESIS*

*submitted by*

**SHIKHA GUPTA**

*for the award of the degree*

*of*

**DOCTOR OF PHILOSOPHY**


**SCHOOL OF COMPUTING AND ELECTRICAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY MANDI**

**July, 2020**

*To My Parents*

*Savita Gupta and Susheel Gupta*

*And My Husband*

*Chirag Singhal*

# Declaration

I hereby declare that the entire work embodied in this thesis is the result of investigations carried out by me in the **School of Computing and Electrical Engineering, Indian Institute of Technology Mandi**, under the supervision of **Dr. Dileep A.D.**, and that it has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on finding of other investigators.

Mandi, 175005
Date:

**Shikha Gupta**

# THESIS CERTIFICATE

This is to certify that the thesis titled **DYNAMIC KERNELS AND SEMANTIC REPRESENTATIONS FOR RECOGNITION OF VARYING SIZE SCENE IMAGES**, submitted by **Shikha Gupa**, to the Indian Institute of Technology, Mandi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by her under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Mandi, 175005
Date:

**Dr. Dileep A. D.**
(Ph.D Supervisor)

# Acknowledgments

Undertaking this Ph.D. has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

I would like to first say a very big thank you and my sincere gratitude to my advisor Dr. Dileep A. D., Associate Professor, School of Computing and Electrical Engineering, IIT Mandi, for all the counsel, scholarly inputs and consistent encouragement I received throughout the research work. I am very much influenced by his constant emphasis on correctness and clarity, patience, discipline, and punctuality. I am inspired by his meticulous attention to detail in correcting reports, papers, and thesis. His expertise and knowledge inspired me to convert my degree from M.S. to Ph.D. Without his timely guidance and constant feedback, this Ph.D. would not have been achievable.

Many thanks also to Dr. C Chandra Sekhar, Professor, Head of the Department, Computer Science and Engineering, IIT Madras, for his always ready to help behavior and immense knowledge. I would like to thanks Dr. Veena Thenkanidiyoor, Associate Professor, Computer Science and Engineering, NIT Goa, for her guidance and support in my research work. Our interactions gave me an opportunity to understand the problem better and work toward the solution. I am also thankful for the moral support extended by her to me during the time of need.

I am very grateful to Dr. Renu M. Rameshan, Assistant Professor, School of Computing and Electrical Engineering, IIT Mandi, A person with an amicable and positive disposition, for her all-time suggestions in personal and professional life. Her way of teaching always inspired me to do hard work. I have learned a lot by working as her teaching assistant. I am also grateful to her for the moral support extended by her to me during my difficult times.

I thank Dr. Timothy A. Gonsalves, Director, Computer Scientist and Professor, IIT

Mandi, for the academic support and the facilities provided to carry out the research work at the Institute. Faculty members of the institute have been very kind enough to extend their help at various phases of this research, whenever I approached them, and I do hereby acknowledge all of them. I apologize for not listing them all.

I am also thankful to the rest of my thesis committee, I thank Dr. Samar Agnihotri, Dr. Arnav Bhavsar, Dr. Shubhajit Roy Chowdhury, Dr. Manoj Thakur and Dr. Sarita Azad for their encouragement and insightful comments. I am also thankful to Dr. Hema Murthy and Dr. C Chandra from IIT Madras for taking core courses of machine learning through video conferencing NKN in starting semesters. These courses helped me a lot to build a basic foundation in the research work.

I thank my fellow lab mates in MANAS group for the stimulating discussions, for the sleepless nights when we worked together before deadlines, and for all the fun we had in the last five years. I thank my collaborators, Deepak kumar pradhan, Ashwin, Kishalaya De for their support, friendship and assistance throughout the years and their collaboration on some of the experiments. I thank my friends Abhijeet Sachdev, Krati Gupta, Krishan Sharma, Munender Varshney, Sai who supported me in every possible way to see the completion of this work.

I am very much indebted to the many countless contributors to the "Open Source" programming community for sharing the numerous tools and systems, which were helpful in numerous ways.

I would also like to thank Dr. SN Jha, Principal Sports Officer, IIT Mandi, for teaching me many sports, especially table tennis, providing wisdom and life values. It helped me to stay focused and active during hard times. I would like to mention hostel wardens, caretaker, Medical unit doctors, staff and guards, I am grateful to them and their families for their friendship and the warmth they extended to me during my time at the IIT Mandi. I would like to mention Ramlila aunty, Vandana, Punit for giving me so much love and family-like support. Teaching Vandana and Punit, always gave peace to my mind and seeing them successful make me feel unconditionally happy.

I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this dream come true. Thank you for blessing me with

this life and all there is in it, for their immense belief in me and my commitment towards my work. I am especially grateful to my mother, for her consistent faith and unconditional love. Shekhar and Tushar, best brothers, I could have ever asked for, blessed for all their help, suggestions, warmth they surround me with all the time. I am also thankful to my in-laws whose support was the key in the completion of my research work, their loving and caring nature always supported me and my work.

Sanchi, my daughter, my blessing, my pride, my joy. At the time of thesis writing, she was in my womb, she gave me enough motivation, for keep the work going and writing the thesis.

And last, but certainly not the least, Chirag Singhal, my husband, thank you from the depths of my being for the endless support, love and time you have given to me. He gives me happiness by simply being who he is. Thank you for always motivating me, believing in me and visiting me at IIT Mandi, over the many weekends. My husband deserves my eternal gratefulness, as always, he patiently listened to my research problems and always suggest in the right direction. I learn a lot from you. Thank you for being on my side.

Above all, I owe it all to Almighty God for granting me the wisdom, health, and strength to undertake this research task and enabling me to its completion.

*Shikha Gupta*

# Abstract

This thesis addresses the task of scene image classification. The real-world scene images are of different sizes and constituent with complex semantic concepts. Typical techniques for scene recognition resize the images to a fix standard size and then extract the features. However, this leads to a significant loss of information as the size of images varies significantly in the range of $10^4$ to $10^6$ pixels. This thesis addresses this issue by considering varying size images in their true resolution to avoid the loss of information due to resizing. True size images results in varying size feature representation of scene images. To build the support vector machine (SVM) based classification model for such representation, two novel dynamic kernels are proposed. Since a scene is composed of complex semantic concepts, obtaining a concept-based representation is quite challenging. This thesis further addresses this issue by the proposed framework for the generation of semantic concept-based representation of varying size scene images.

This thesis proposes two dynamic kernels, namely, spatial probabilistic sequence kernel (SPSK) and deep spatial pyramid match kernel (DSPMK) for the classification of varying length patterns of scene images. Dynamic kernels are the similarity functions that take two varying sizes of the input and compute the similarity score. In SPSK images are represented by sets of low-level local feature vectors. SPSK incorporates spatial configuration of local feature vectors in the computation of probabilistic sequence kernel. Low-level features used in the computation of SPSK are the local descriptors and failed to capture the complex geometric structure of scene images. For better feature representation, low-level features are replaced by a learned convolutional neural network (CNN) based features. The main challenge is the usage of CNN requires to bring different sized input images to a fixed predefined size either by reducing, enlarging or cropping. To handle this, the proposed work

provides a mechanism to pass the images to CNN in their original resolution. This results in varying size sets of deep activation maps as image representation. To build the SVM-based classifier for such representation, a DSPMK as the novel dynamic kernel is proposed. DSPMK operates over sets of activation maps on different pyramid levels. At each level, activation maps are divided into fix number of spatial regions and the final similarity score between two examples is obtained by computing the weighted combination of intermediate matching scores.

To capture the constituent concept information of scene images, a scene image is represented in semantic concept space by the posterior probabilities of concepts present in it and such representation is known as semantic multinomial (SMN) representation. SMN representation requires concept annotated dataset with concept specific features for concept modeling which are infeasible to generate manually due to large size of database. The proposed research work focused on building the concept models via pseudo-concepts in the absence of true concept annotated data. For the sets of local feature vector representation of scene images, clusters of local feature vectors of all the database images are proposed as cues to the pseud-concepts. Further, the pseudo-concept models are built using the proposed dynamic kernel-based SVM framework. Disadvantages of the low-level feature-based SMN representation include, concept models are built using features from the complete image instead of concept specific features, and handcrafted features used for building the concept models are local descriptors, moreover, it do not capture much of the semantic information. To overcome these limitations, a novel deep CNN-based SMN representation is proposed that uses the deeper convolutional layers filter responses of pre-trained CNNs as cues to pseudo-concepts. Convolutional layer filters are considered as concepts detector, but ground truth information of filters (i.e., which filter is learning what concept) is not known during the training process of CNNs. Hence, the true-concept identity of a particular filter from its activation is not inferred. However, activation maps responses can be visualized using different visualization techniques. The non-significant pseudo-concepts are removed using the proposed filter specific threshold-based approach and similar pseudo-concepts are grouped using subspace modeling. Pseudo-concept models are built using linear kernel-based SVM to generate novel SMN representation. The proposed procedure for building pseudo-concept

models is weakly supervised as an image may contain multiple pseudo-concepts.

Further to improve the pseudo-concept modeling, the proposed thesis work focuses on the semantic analysis of filter responses of true size images. A strategy is proposed in which filter responses of true resolution images act as cues for pseudo-concepts in the absence of true concepts labeled data. Procedure to select prominent pseudo-concepts and group the similar one is proposed. In the end, pseudo-concept models are built using proposed modified DSPMK-based framework to generate SMN representation of varying sized images. Potential of the proposed approaches are evaluated using standard scenes recognition datasets such as MIT8 scene, Vogel Schiele, MIT67 indoor and SUN397.

**Keywords:** *Scene images, varying size scene image recognition, varying length patterns, set of low-level local feature vectors, set of varying size activation maps, dynamic kernel, spatial probabilistic sequence kernel, deep spatial pyramid match kernel, support vector machine, semantic multinomial representation, pseudo-concepts, pseudo-concept selection, pseudo-concepts grouping, subspace analysis, kernel clustering.*

# Contents

**Acknowledgment** i

**Abstract** v

**List of Tables** xi

**List of Figures** xv

**List of Algorithms** xviii

**Abbreviations** xxi

**1 Introduction** 2

1.1 Classification of varying size patterns of scene images . . . . . 4

1.2 Classification using semantic concept based scene image representations . . . . . 5

1.3 Objective and scope of the work . . . . . 7

1.4 Contributions of the thesis . . . . . 9

1.5 Organization of the thesis . . . . . 10

**2 Related Works on Scene Image Classification** 14

2.1 Scene image representation . . . . . 15

2.1.1 Low-level local feature based representations . . . . . 17

2.1.2 Learning-based feature representations . . . . . 20

2.1.3 Semantic concept-based representation . . . . . 21

2.2 Approaches to scene image classification . . . . . 24

2.2.1 GMM-based Bayes classifier for scene image classification . . . . . 24

2.2.2 SVM-based classifier for scene image classification . . . . . 26

2.3 Dynamic kernels for varying size scene image classification . . . . . 27
2.3.1 Improved Fisher kernel . . . . . 28
2.3.2 GMM supervector kernel . . . . . 29
2.3.3 GMM-UBM mean interval kernel . . . . . 30
2.3.4 Intermediate matching kernel . . . . . 32
2.3.5 Histogram intersection kernel . . . . . 34
2.3.6 Pyramid match kernel . . . . . 34
2.3.7 Spatial pyramid match kernel . . . . . 35
2.4 Summary . . . . . 38

**3 Spatial Probabilistic Sequence Kernel for Sets of Local Feature Vectors 42**
3.1 Probabilistic sequence kernel for sets of local feature vectors . . . . . 43
3.2 Motivation for spatial PSK . . . . . 45
3.3 Spatial probabilistic sequence kernel . . . . . 46
3.3.1 Pooling techniques for probabilistic alignment vectors . . . . . 48
3.3.2 Classification using class-specific SPSK based SVM classifier . . . . . 50
3.4 Databases and features used in the studies . . . . . 52
3.5 Experimental studies on scene image classification task . . . . . 54
3.6 Summary . . . . . 57

**4 Deep Spatial Pyramid Match Kernel for Varying Size Sets of Activation Maps 59**
4.1 Motivation for using true varying size images as input to CNNs . . . . . 62
4.2 Image representation using varying size sets of activation maps . . . . . 64
4.3 Deep spatial pyramid match kernel . . . . . 65
4.4 Experimental Studies . . . . . 71
4.4.1 Datasets . . . . . 71
4.4.2 Experimental details . . . . . 71
4.4.3 Results on scene image classification . . . . . 73
4.5 Summary . . . . . 77

**5 Low-level Local Features based Semantic Multinomial Representation 78**
5.1 Semantic multinomial representation . . . . . 81
5.2 Pseudo-concept modeling . . . . . 83

5.2.1 Motivation for pseudo-concepts . . . 83
5.2.2 Generation of data for pseudo-concept classes . . . 84
5.2.3 GMM-based approach to build pseudo-concept models . . . 85
5.2.4 SVM-based approach to build pseudo-concept models . . . 86
5.3 Generation of SMN representation using pseudo-concept SVMs . . . 87
5.4 Experimental studies on scene classification using SMN representation . . . 89
5.4.1 Experimental details . . . 89
5.4.2 Results on scene image classification . . . 92
5.5 Illustration of pseudo-concepts . . . 96
5.6 Summary . . . 96

**6 Deep CNN-based Semantic Multinomial Representation for Scene Images 100**
6.1 Deep SMN representation generation framework . . . 104
6.1.1 Selection of pseudo-concepts and generation of pseudo-concepts class specific data . . . 107
6.1.2 Grouping of similar pseudo-concepts using subspace modeling . . . 110
6.1.3 Building weakly supervised pseudo-concepts model . . . 112
6.1.4 Deep SMN representation generation of scene images . . . 112
6.1.5 Classification framework . . . 113
6.2 Experimental Studies . . . 114
6.2.1 Experimental Details . . . 114
6.2.2 Results on scene image classification . . . 115
6.2.3 Comparison of proposed approach to state-of-the-art approaches . . . 117
6.2.4 Analysis of classification accuracy vs. number of pseudo-concepts . . . 120
6.2.5 Visualization of pseudo-concepts . . . 122
6.2.6 Visualizing SMN representation using t-SNE . . . 123
6.3 Summary . . . 125

**7 Modified DSPMK-based Pseudo-concept Modeling using Varying Size Activation Maps 128**
7.1 Motivation for using true varying size images for semantic analysis . . . 131
7.2 Framework for varying size scene recognition with deep SMN representation . . . 132
7.2.1 Selection of pseudo-concepts using varying size activation maps . . . 134
7.2.2 Grouping of similar pseudo-concepts using dynamic kernel-based clustering 137

7.2.3 Pseudo-concept modeling using modified DSPMK-based SVM . . . . . 139
7.2.4 Deep SMN representation generation for varying size scene images . . . . 141
7.3 Experimental studies . . . . . 142
7.3.1 Experimental details . . . . . 142
7.3.2 Experimental results and analysis . . . . . 143
7.3.3 Analysis of effective number of pseudo-concepts . . . . . 147
7.3.4 Visualization of pseudo-concept vs. true concept . . . . . 149
7.4 Summary . . . . . 150

**8 Summary and Future Work** **152**
8.1 Summary of the work . . . . . 152
8.2 Directions for further work . . . . . 155

**A Spectral Clustering** **158**
A.1 Spectral clustering algorithm . . . . . 158

**References** **160**

**List of Publications** **174**