Pixel based Semantic Labelling and Image Parsing using Intelligent Learning Techniques

Thesis submitted in partial fulfillment

for the Award of Degree

DOCTOR OF PHILOSOPHY

by

Vishal Srivastava

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

(BANARAS HINDU UNIVERSITY),

VARANASI-221 005

Roll No: 17071022 **October 2020**

Certificate

It is certified that the work contained in this thesis titled ”Pixel based Semantic Labelling and Image Parsing using Intelligent Learning Techniques” by ”Vishal Srivastava” has been carried out under my supervision and that it has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.

Dr. Bhaskar Biswas

Associate Professor

Department of Computer Science and Engineering

Indian Institute of Technology (Banaras Hindu University),

Varanasi-221 005

DECLARATION BY THE CANDIDATE

I, Vishal Srivastava, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of Dr. Bhaskar Biswas from July-2017 to October , 2020, at the Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date :

Place :

Signature of the Student

(Vishal Srivastava)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Signature of Supervisor

(Dr. Bhaskar Biswas)

Signature of Head of Department

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis : **Pixel based Semantic Labelling and Image Parsing using Intelligent Learning Techniques**

Name of the Student : **Vishal Srivastava**

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the ”DOCTOR OF PHILOSOPHY”.

Date :

Place :

Signature of the Student

(Vishal Srivastava)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author’s personal use provided that the source and the Institute’s copyright notice are indicated.

Acknowledgements

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 take this opportunity to extend my sincere gratitude and appreciation to all those who made this Ph.D. thesis possible. First and foremost, I would like to express my sincere gratitude to my research guide, Dr. Bhaskar Biswas, for introducing me to this exciting field of Computer Science & Engineering and for his dedicated help, advice, inspiration, encouragement, and continuous support my Ph.D. His enthusiasm, integral view on research, and his mission to provide high-quality work have made a deep impression on me. During our course of interaction during the last three years, I have learned extensively from him, including how to raise new possibilities, regard an old question from a new perspective, approach a problem through systematic thinking, data-driven decision making, and exploiting serendipity. I owe him lots of gratitude for having me shown this way of research. My special words of thanks should also go to Prof. Rajeev Srivastava, Head, Computer Science & Engineering Department, for his kindness and valuable support for facilitating all the requirements. I express my heartfelt gratitude to faculty members and staff of the department. I especially thank the members of my Research Program Evaluation committee, Dr. Bhaskar Biswas, Dr. Tanima Dutta, Dr. Kalpana Choudhary and DPGC Convener Dr. Pratik Chattopadhyay for their valuable suggestions regarding the thesis. I have a special mention of thanks to my lab mates in Research Lab, Mr. Navin, Mr. Shivansh, and Ms. Sneha. I wish to extend my gratitude to all my seniors. Their advice, encouragements, and critics were always the source of inspiration. I would want to thank all my family members who continuously support me throughout my work. My parents did hard work to support the logistics for my studies and encouraged me every decision that I took. Finally, I am grateful to my almighty god shiv for blessings and for giving me the strength to persevere throughout the long and arduous journey.

- Vishal Srivastava

PREFACE

Semantic image labelling, also known as pixel-based classification, is a task of segmenting the objects within an image by pixel-level spectral similarity. Semantic image labelling has also referred to as image parsing, which is a process of decomposing the image in different regions and constructing a structured input. In this thesis, we summarise the four aspects of research in the semantic labelling, i.e., classical machine learning(ML), feature engineering, deep learning(DL), and relaxation labelling(RL).

The above mentioned four aspects of semantic labelling lead to realizing that it is not a separate domain but a natural step in moving from coarse to fine interpretation. The original procedure could have been derived from a classification scheme, which predicts the label for a complete input. This process is also known as image category classification in the literature. This process has various applications such as bio-metric image classification, classification of tumor in a different grade, classification of different classes of species, digit classification using digit databases, emotion detection using face databases, and category classification using features from CBIR systems. A vast set of image databases, such as MNIST, CIFAR, ORL, YALE, etc., are available to validate the category classification methods.

In the previous classification methods, i.e., coarse grain, the classifiers predict the objects or provide a rank list in case of many objects. The next step is to localize and detect the objects within the image, which is a fine grain inference. The main aim of such inference is not only to provide classes but also some specific information such as the spatial location of classes, centroids, and bounding boxes. These kinds of fine-grain classification processes have performed on pixel-level, not image level. Therefore, they are computationally costly. Such inferences in various literature have denoted semantic labelling, pixel-based semantic classification/segmentation, or semantic image parsing. Consequently, we can summarise the semantic parsing as an image-based method to achieve fine-grain predictions. The goal is to make a dense prediction for the label of each pixel in such a way that each pixel is labelled with its class of enclosed regions. These kinds of fine-grained studies are the main interest of this thesis.

Another important aspect is to select the complex data-sets for semantic labelling. We have used low dimensional RGB images such as sift-flow, pascal-voc data-sets, and high dimensional hyperspectral(HD-HSI) data-sets to perform the semantic labelling based experiments. The facial expression based images in ORL, YALE-A, and B, COIL data-sets have been used for category prediction also.

We have discussed the classical machine learning methods, some feature engineering, and knowledge embedding techniques to develop accurate and efficient frameworks. The salient and CNN feature-based approaches have been adapted to achieve effective and robust features from raw data. Feature selection and dimension embedding have also played a pivotal role in proposed frameworks. Deep learning and CNN based approaches have been used to design the custom CNN architecture and exploited the significant results from the image. Some relaxation labelling-based methods have also detailed to improve the CNN and classical ML-based probabilistic outcomes significantly. The advantages and drawbacks of the proposed frameworks have discussed. The comparative analysis for benchmark image sets and evaluation matrices have also been performed. Finally, some encouraging future works have drawn out, and the conclusion has drawn for pixel-wise semantic scene labelling or image parsing.

Contents

Bibliography **175**

List of Figures

List of Tables

Abbreviations

DNN	**D**eep **N**eural **N**etwork
CNN	**C**onvolution **N**eural **N**etwork
ML	**M**achine **L**earning
RGB	**R**ed **G**reen **B**lue **I**mage
MLP	**M**ulti **L**ayer **P**erceptron
SVM	**S**upport **V**ector **M**achine
OA	**O**verall **A**ccuracy
AA	**A**verage **A**ccuracy
K	**K**appa **V**alue
SVD	**S**ingular value **D**ecomposition
PCA	**P**rincipal **C**omponent **A**nalysis
HD	**H**igh **D**imensional
RL	**R**elaxation **L**abelling
MRF	**M**arkov **R**andom **F**ield
CRF	**C**onditional **R**andom **F**ield
IOU	**I**ntersection **O**ver **U**nion
MLR	**M**ultinomial **R**egression
MO	**M**aximum **O**bject **S**ignal
GC	**G**raph **C**ut
EPS	**E**dge **P**reservation **S**cheme
RBF	**R**adial **B**asis **F**unction
MAP	**M**aximum **A** **P**osterior

paraKERNAL	**p**arametric **K**ernal
EFF	**E**dge **F**eature **F**usion
FC	**F**ully **C**onnected
LM	**L**arge scale **M**orphology
MFP	**M**ulti criteria **F**eature **P**ooling
SSIM	**S**tructural **S**imilarity
MI	**M**utual **I**nformation

Symbols

exp	exponential
γ	$\frac{spectral}{spatial}$
λ	eigenvectors or regularizers
α	regularizer
$\phi_{(k)}$	length k subspaces
λ_1, λ_2	multiregularizers
λ_d	step length for new dimensions
μ	regularizer
c_{min}	minimum correlation
σ	standard deviation
ϕ	relaxation prior
$\bigoplus_N^j$	Fusion operator on nearest neighbour and features
$\triangle$	gradient
η	Learning Rate