

CONTENTS

APPROVAL PAGE	
SELF DECLARATION AGAINST PLAGIARISM	
ABSTRACT	i
ACKNOWLEDGMENTS	ii
CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
LIST OF ABBREVIATION	ix
I INTRODUCTION	1
1.1 Background	1
1.2 Problem Identification	3
1.3 Objectives	4
1.4 Hypotheses	5
1.5 Research Methodology	5
1.6 Problem Limitations	5
II BASIC CONCEPTS	7
2.1 Fine-Grained Visual Classification	7
2.2 Vision Transformer	8
2.3 Internal Ensemble Learning Transformer	10
2.3.1 Multi-Head Voting Module	11
2.3.2 Dynamic Selection Module	12
2.3.3 Cross-Layer Refinement Module	13
2.4 Batch-based Dynamic Mask	15
III SYSTEM MODEL AND THE PROPOSED DESIGN	16
3.1 System Model	16
3.2 Dataset	17

3.3	Proposed Method	18
3.3.1	Kernel Modifications in the MHV Module	18
3.3.2	Method of Applying the BDMM Algorithm	20
3.3.2.1	First Masking Method	21
3.3.2.2	Second Masking Method	22
3.4	Performance Parameter	23
3.4.1	Top-1 accuracy	23
3.4.2	Precision	24
3.4.3	Recall	24
3.4.4	F1 Score	24
IV PERFORMANCE EVALUATIONS		25
4.1	Implementation Details	25
4.2	Scenario Performance	25
4.3	Impact of Convolution Kernel Types on Model Performance	26
4.3.1	Result on CUB-200-2011	27
4.3.2	Result on Stanford Dogs	27
4.4	Impact of Masking Methods on Model Performance	28
4.4.1	Result on CUB-200-2011	28
4.4.2	Result on Stanford Dogs	29
4.5	Impact of Combined Methodologies on Model Performance	29
4.5.1	Result on CUB-200-2011	30
4.5.2	Result on Stanford Dogs	31
4.6	Visualizations	32
4.7	Comparison with the State-Of-The-Art	33
4.7.1	Result on CUB-200-2011	33
4.7.2	Results on Stanford Dogs	34
V CONCLUSIONS AND FUTURE WORKS		36
5.1	Conclusions	36
5.2	Future Works	36
REFERENCES		38