Self-Supervised Learning in Gender Classification using full-body images extracted from CCTV footage

Gowshalini Rajalingam

219335N

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2023

Self-Supervised Learning in Gender Classification using full-body images extracted from CCTV footage

Gowshalini Rajalingam

219335N

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2023

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Thanuja Ambegoda

Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

I would like to express my deep and sincere gratitude to Dr. Thanuja Ambegoda for guiding me to initiate, finding a great research topic to conduct the research, and for continuing support and encouragement. His supervision greatly helped me in setting goals and engaging in the research.

I would like to express my greatest gratitude to the Department of Computer Science and Engineering, the University of Moratuwa for providing the support to overcome this effort. Last but not least, my heartfelt gratitude goes to my parents and friends who supported me throughout this effort.

ABSTRACT

Gender classification is regarded as one of the vital components of security systems, recommendation systems, data access authentication and surveillance. Facial features and supervised learning remain the predominant metrics to classify genders currently. But facial feature driven approach would falter in case of incomplete or unavailable details especially when analyzing masked faces or CCTV footage and supervised learning driven approach becomes tedious and time-consuming provided large volume of labelled data. Therefore, the need of analyzing full-body images is established instead of the sole focus of facial features driven analysis as well as the less dependency on supervised learning. The proposed approach establishes the implementation of convolutional neural network (CNN) based on self-supervised learning classification algorithm that needs fewer volumes of labelled data for fine-tuning. dBOT classifier, a state-of-the-art self-supervised image classification model, is used to perform transfer learning and the subsequential fine-tuning to facilitate the training on low-quality images. The proposed model on evaluation significantly outperforms SSL based methods for small, unclear full-body gender image classification techniques applied on CCTV footage extracts.

Keywords: CNN, dBOT, Gender-Classification, CCTV

TABLE OF CONTENTS

Declaration i

Acknowledgement ii

Abstract iii

Table of Contents iv

List of Figures vi

List of Tables viii

List of Abbreviations ix

List of Appendices xi

Chapter 1 1

Introduction 1

1.1 Introduction 1

1.1.1 Gender Classification 1

1.1.2 Self-Supervised Learning 4

1.2 Research problem 10

1.3 Research objective 12

1.3.1 Main objective 12

1.3.2 Sub objectives 12

1.3.3 Technical objectives 13

1.4 Research gap identified 14

Chapter 2 15

Statement of the work 15

2.1 Background information and overview of previous work 15

2.1.1 Overview 15

2.1.2 SSL in computer vision 17
2.1.3 Gender Classification 20
Chapter 3 24
Research Methodology 24
3.1 Dataset 24
3.2 Algorithm 25
3.3 Methodology 28
Chapter 4 33
Results and discussion 33
Chapter 5 38
Conclusion 38
References 39
Appendices 48
Appendix A 48

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Facial Feature difference between male and female	3
Figure 1.2	Men vs women body shape	3
Figure 1.3	Jigsaw puzzles (3x3). Nine patches are sent into a DCNN (Deep Convolutional Neural Network), which then forecasts the positions of the patches.	5
Figure 1.4	Representation of an image inpainting ruse. The model is used to recreate the missing portion of the image (right) using the rest of the image, given the corrupted image (left).	5
Figure 1.5	The representation of the algorithm used in [12]	6
Figure 1.6	image examples from Goodfellow et al 2014	6
Figure 1.7	image examples from Karras et al. 2019	7
Figure 1.8	GAN Framework. Source [8]	8
Figure 1.9	In contrastive learning the main idea is to bring the original and positive images together while bringing the original and negative images away.	8
Figure 1.10	Contrastive learning pipeline for self-supervised training	9

Figure 1.11	IMDB-WIKI, Adience Benchmark Dataset, FG-NET Aging Database, UTKFace Dataset, and Morph Album 2 Dataset[23]	10
Figure 2.1	Standard SSL framework.	16
Figure 3.1	The market1501 dataset	25
Figure 3.2	Transfer classification accuracy on various datasets for dBOT classification in [2]	25
Figure 3.3	The architecture of Masked Image Modeling	27
Figure 3.4	The architecture of dBOT algorithm	27
Figure 3.5	Grayscale, segmented and background removed images	28
Figure 3.6	Methodology Diagram	31
Figure 4.1	Downstream accuracy vs dataset size	36

LIST OF TABLES

Table	Description	Page
Table 4.1	Accuracy score for gender classification	33
Table 4.2	The downstream accuracy with the different dataset sizes	35
Table 4.3	Comparison of self-supervised learning models with our model	37

LIST OF ABBREVIATIONS

Abbreviation	Description
CCTV	closed-circuit television
AI	Artificial Intelligence
CNN	Convolutional Neural Network
dBOT	knowledge Distillation with BOotstrapped Teachers
GCP	Google Cloud Platform
CPU	Central Processing Unit
SSL	Self Supervised Learning
CNN	Convolutional Neural Network
CPC	Contrastive Predictive Coding
GAN	Generative Adversarial Networks
NLP	Natural Language Processing
CL	Contrastive Learning
DCNN	Deep Convolutional Neural Network
SEER	SElf-supERvised
CV	Computer Vision
CFN	Context-Free Network
CPC	Contrastive Predictive Coding
CBAM	Convolution Block Attention Module
AUC	Area Under the Curve
SVM	Support Vector Machine

DCT	Discrete Cosine Transform
LBP	Local Binary Pattern
GDF	Geometrical Distance Feature
VFL	Variational Feature Learning
PHVW	Pyramid Histogram of Visual Words
PHOG	Pyramid Histogram of Gradients
SIFT	Scale-invariant Feature Transform
PCA	Principal Component Analysis
HOG	Histogram of Oriented Gradients
mAP	mean Average Precision
MIM	Masked Image Modeling

LIST OF APPENDICES

Appendix	Description	Page
Appendix - A	Literature review summary for gender classification	44