



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

**ARTIFICIAL INTELLIGENCE SYNTHESIZED
FACE SWAPPING DETECTION MODEL USING
UNIFIED DATA SETS**



GONG DAFENG
اونيور سيني ليكسيكس مليسيا ملاك
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DOCTOR OF PHILOSOPHY

2023



Faculty of Information and Communication Technology

**ARTIFICIAL INTELLIGENCE SYNTHESIZED FACE SWAPPING
DETECTION MODEL USING UNIFIED DATA SETS**



Gong Dafeng

Doctor of Philosophy

2023

**ARTIFICIAL INTELLIGENCE SYNTHESIZED FACE SWAPPING DETECTION
MODEL USING UNIFIED DATA SETS**

GONG DAFENG

**A thesis submitted
in fulfillment of the requirements for the degree of Doctor of Philosophy**



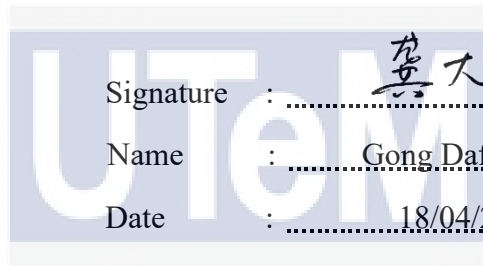
Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I declare that this thesis entitled “Artificial Intelligence Synthesized Face Swapping Detection Model Using Unified Data Sets” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



Signature :

Name : Gong Dafeng


Date : 18/04/2023

اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

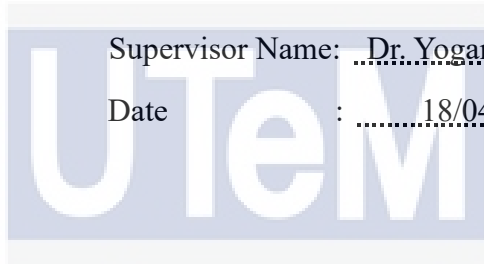
APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : 

Supervisor Name: Dr. Yogan Jaya Kumar

Date : 18/04/2023



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEDICATION

To my beloved father, mother, wife, sisters, brothers, and daughters



ABSTRACT

Today's image generation technology can generate high-quality face images, and it is not easy to recognize the authenticity of the generated images through human eyes. Due to the rise of image generation technology based on deep learning, software related to image generation is used widely, including some popular face-swapping software. If misused, it will directly affect forensics and security-related industries. As an essential branch of computer security, image forensics technology also needs to be improved with the development of image forgery technology. This study aims to improve deepfake detection, a face-swapping forgery, by absorbing the advantages of deep learning technologies. In order to solve the problem of poor detection performance on cross data sets, this study generates unified data sets from multiple sources using spatial enhancement technology to obtain approximately four million images, 36 times the size of the original data set, and was proved effective with traditional feature methods. Taking the advantages of ResNet and Inception networks, DeepfakeNet architecture composed of 32 parallel branches and 20 network layers is proposed as the deepfake detection model with FLOPs of 2.05×10^9 and parameters of 10.87×10^6 . To further improve the proposed DeepfakeNet model, a univariate method is used to obtain the ideal model values of hyperparameters, including batch size, epochs, dropout, learning rate, and sample ratio. Accuracy of 98.69%, loss value of 3.42% and AUC of 0.96 are achieved. The evidence of this study shows that the proposed DeepfakeNet has significantly improved over the mainstream methods in terms of loss value, accuracy, AUC, FLOPs, and parameters.

MODEL PENGESANAN PERTUKARAN MUKA MELALUI KECERDASAN BUATAN MENGGUNAKAN SET DATA BERSATU

ABSTRAK

Teknologi penjaan imej hari ini boleh menjana imej muka berkualiti tinggi, dan bukan mudah untuk mengenali ketulenan imej yang dijana melalui mata manusia. Disebabkan oleh peningkatan teknologi penjaan imej berdasarkan pembelajaran mendalam, perisian yang berkaitan dengan penjaan imej digunakan secara meluas, termasuk beberapa perisian pertukaran muka yang popular. Jika disalahgunakan, ia akan menjejaskan industri forensik dan berkaitan keselamatan secara langsung. Sebagai cabang penting dalam keselamatan komputer, teknologi forensik imej juga perlu dipertingkatkan dengan pembangunan teknologi pemalsuan imej. Kajian ini bertujuan untuk meningkatkan pengesanan deepfake, pemalsuan pertukaran muka, dengan menyerap kelebihan teknologi pembelajaran mendalam. Untuk menyelesaikan masalah prestasi pengesanan yang lemah pada set data silang, kajian ini menjana set data bersatu daripada pelbagai sumber menggunakan teknologi peningkatan spatial untuk mendapatkan kira-kira empat juta imej, 36 kali ganda saiz set data asal, dan telah terbukti berkesan dengan kaedah ciri tradisional. Mengambil kelebihan rangkaian ResNet dan Inception, seni bina DeepfakeNet yang terdiri daripada 32 cawangan selari dan 20 lapisan rangkaian dicadangkan sebagai model pengesanan deepfake dengan FLOP 2.05×10^9 dan parameter 10.87×10^6 . Untuk menambah baik lagi model DeepfakeNet yang dicadangkan, satu univariate kaedah digunakan untuk mendapatkan nilai model ideal hiperparameter, termasuk saiz kelompok, zaman, keciciran, kadar pembelajaran, dan nisbah sampel. Ketepatan 98.69%, nilai kehilangan 3.42% dan AUC sebanyak 0.96 dicapai. Bukti kajian ini menunjukkan bahawa DeepfakeNet yang dicadangkan telah bertambah baik dengan ketara berbanding kaedah arus perdana dari segi nilai kehilangan, ketepatan, AUC, FLOP dan parameter.

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to my supervisors, Dr. Yogan Jaya Kumar and Professor Ts. Dr. Goh Ong Sing, for their continuous support and availability in the development of my research study, especially because both of them have never been too busy to keep an eye on my progress in spite of their numerous obligations. They have greatly helped me in a lot of ways throughout this study. The most important lesson I learnt from them is that whatever you work on, work with dedication and sincerity. Again, I owe them my deepest thanks.

In addition, I am grateful to my colleagues at my research lab, especially those in a deep learning research group, for the useful discussions, comments, and knowledge sharing. Thanks also to my friends here at UTeM, who have in one way or another, helped, encouraged, and motivated me during the progression of my study.

I am also grateful to all staff of Faculty of Information and Communication Technology and other departments, UTeM, for their kind cooperation during my study and stay here.

Last but not least, I like to thank my parents, wife, sisters, brothers and daughters for their support, guidance, love and prayers. Thank you for all the encouragement you have given me. Once again, thank you.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

TABLE OF CONTENTS

	PAGE
DECLARATION	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF APPENDICES	xiii
LIST OF ABBREVIATIONS	xiv
LIST OF SYMBOLS	xvi
LIST OF PUBLICATIONS	xvii
CHAPTER	
1. INTRODUCTION	1
1.1 Introduction	1
1.2 Problem background	5
1.3 Research questions	10
1.4 Objectives of study	11
1.5 Research scope	12
1.6 Significance of study	12
1.7 Thesis organization	13
2. LITERATURE REVIEW	15
2.1 Introduction	15
2.2 Deepfake methods and applications	17
2.2.1 GANs based on the improvement of loss hunction	21
2.2.2 GANs based on the improvement of the model	23
2.2.3 Applications of GANs and deepfakes	25
2.3 Deepfake detection technology	30
2.3.1 Deepfake detection	31
2.3.2 Multimedia forensic	36
2.3.3 Anti-Counterfeiting	39
2.3.4 Convolutional neural networks	39
2.3.5 Data driven method	41
2.4 Model interpretability technology	45
2.5 Concepts of convolution neural network	47

2.5.1	Convolutional neural network	48
2.5.2	Convolution layer	48
2.5.3	Pooling layer	49
2.5.4	Fully connected layers	51
2.5.5	Activation function	53
2.5.6	Loss function	58
2.5.7	Neural network optimization algorithm	60
2.6	Mainstream convolutional neural network	62
2.6.1	VGG19	63
2.6.2	GoogLeNet	64
2.6.3	XceptionNet	66
2.6.4	ResNet101	67
2.6.5	ResNeXt50	69
2.7	Deepfake data sets	70
2.7.1	FaceForensics++ (FF++)	71
2.7.2	FaceForensics (FF)	72
2.7.3	Deepfake-TIMIT	72
2.7.4	DFDC preview data set	72
2.7.5	DFDC	73
2.7.6	Deepfake detection (DFD)	73
2.7.7	MesoNet data	73
2.7.8	Celeb-DF	74
2.7.9	UADFV	74
2.8	Discussion	75
2.9	Summary	78
3.	RESEARCH METHODOLOGY	79
3.1	Introduction	79
3.2	Research design	79
3.3	Research operational framework	81
3.3.1	Phase 1: preliminary research literature	83
3.3.2	Phase 2: build a unified and enhanced data set	83
3.3.3	Phase 3: propose DeepfakeNet architecture	90
3.3.4	Phase 4: obtain optimized hyperparameter values	92
3.3.5	Phase 5: report writing	97
3.4	Experimental evaluation	97
3.5	Selected benchmark methods for comparison	102
3.6	Summary	104
4.	DATA UNIFICATION AND ENHANCEMENT	105
4.1	Introduction	105
4.2	Building a unified data set	107
4.2.1	Extract video frames	107
4.2.2	Convert different images into the same specification	111

4.3	Data enhancement	112
4.3.1	Rotating	112
4.3.2	Flipping	113
4.3.3	Scaling	114
4.3.4	Clipping	115
4.3.5	Changing Colors	115
4.4	Data validation and analysis using traditional feature methods	118
4.4.1	Introduction of traditional feature methods	118
4.4.2	Selection of traditional feature methods	126
4.4.3	Experiment and result analysis	128
4.5	Summary	132
5.	DEEPPFAKENET	133
5.1	Introduction	133
5.2	Deepfake detection algorithm	136
5.2.1	ResNet principle	137
5.2.2	Inception principle	142
5.2.3	DeepfakeNet principle	149
5.3	Designing DeepfakeNet	152
5.3.1	Conv1 layer	152
5.3.2	Conv2 layer	155
5.3.3	Conv3 layer	156
5.3.4	Conv4 layer	158
5.3.5	Conv5 layer	160
5.3.6	Softmax layer	161
5.4	Algorithm implementation	161
5.5	Experimental analysis	163
5.6	Summary	166
6.	HYPERPARAMETER VALUES OPTIMIZATION	167
6.1	Introduction	167
6.2	Optimization methods	167
6.3	Hyperparameter values optimization	169
6.4	Experimental result	175
6.5	Summary	183
7.	CONCLUSION AND FUTURE WORK	184
7.1	Introduction	184
7.2	Proposed methods	185
7.2.1	Generation method of a unified data set	185
7.2.2	A new deepfake detection method	186
7.2.3	A method of training hyperparameter values and model	187
7.3	Contributions of the study	187
7.4	Future work	190

7.5	Summary	192
	REFERENCES	193
	APPENDICES	220



LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Classification of GANs derived models (Wu Shaoqian, 2019)	21
2.2	Typical deepfake tools	29
2.3	Error rate of CIFAR-10 (Khalid et al., 2020)	56
2.4	Error rate of CIFAR-100 (Khalid et al., 2020)	57
2.5	Error rate of NDSB (Khalid et al., 2020)	57
2.6	Comparison of Common Data Sets	75
3.1	Kaggle JSON file data format	88
3.2	Analysis of three data sources	88
3.3	Population, intervention, comparison and outcome (PICO)	103
4.1	Composition of Data Sources	116
4.2	Approaches of enhancing data	117
4.3	Total data after enhancement	118
4.4	Predicted results	131
5.1	DeepfakeNet with a $32 \times 4d$ template	151
5.2	Preset random hyperparameter values	164
5.3	Other experiment results	165
6.1	Accuracy of experimental results of different sample ratio	170
6.2	Accuracy of experimental results of different learning rate	171
6.3	Accuracy of experimental results of different dropout	172
6.4	Accuracy of experimental results of different batch size	174
6.5	Accuracy of experimental results of different epochs	174
6.6	Preset hyperparameter values	175
6.7	Accuracy comparison of experimental results	181

6.8	AUC value comparison of experimental results	181
6.9	Comparison of Number of FLOPs and Params	182

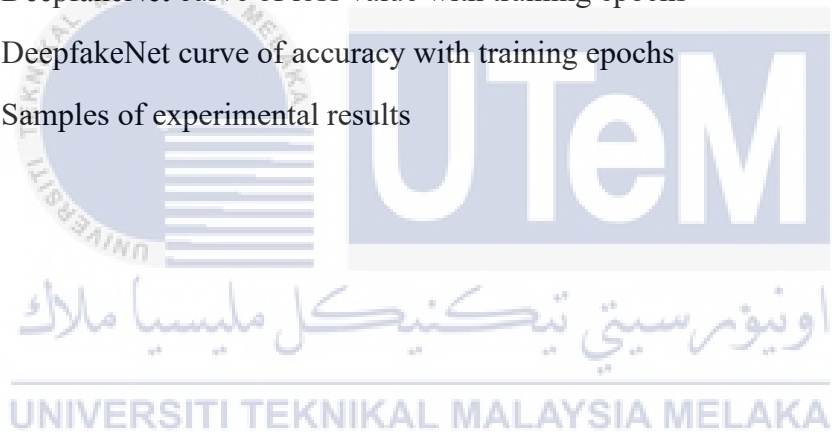


LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	Top: fake videos, bottom: true videos	16
2.2	Number of papers on GANs published since 2014	18
2.3	Architecture of GAN (Goodfellow et al., 2014)	19
2.4	Images generated by ProGAN (Karras et al., 2017)	25
2.5	Images generated by StarGAN (Choi et al., 2018)	26
2.6	Architecture of ESRGAN (Wang et al., 2018)	27
2.7	DeepNude App (Yuanxiaosc, 2019)	28
2.8	Number of papers deepfake detection published since 2014	31
2.9	Process of LRCN (Li et al., 2018)	32
2.10	Detecting deepfake eyes and teeth (Matern et al., 2019)	33
2.11	Architecture of Capsule-Forensics (Nguyen et al., 2019a)	34
2.12	Architecture of detecting forensics (Nhu-tai et al., 2018)	36
2.13	Pipeline of detecting the forgery face (Rössler et al., 2019)	37
2.14	Block scheme for detecting deepfake (Cozzolino et al., 2019)	39
2.15	Computer generated face (CGFace) (Dang et al., 2018)	40
2.16	Schematic diagram of convolution algorithm	49
2.17	2*2 diagram of maximum pooling	50
2.18	Schematic diagram of fully connected network	51
2.19	Schematic diagram of full connection layer feature recognition	52
2.20	Sigmoid function (Liu et al., 2019)	54
2.21	Tanh function (Liu et al., 2019)	55
2.22	ReLU function (Liu et al., 2019)	56
2.23	VGG network structure parameters (Simonyan, 2014)	64

2.24	GoogLeNet network structure parameters (Szegedy et al., 2014)	65
2.25	XceptionNet model (Chollet, 2017)	67
2.26	ResNet network structure parameters (He et al., 2016)	69
2.27	Operation adopted by ResNeXt network (Xie et al., 2017)	70
3.1	Operational research framework	82
3.2	Overview of research study	82
3.3	Data unified and enhanced process	89
3.4	Network structure of face-swapping detection	91
4.1	Process of data unification and enhancement	107
4.2	Process of converting a video into frames	108
4.3	Rotating an image	113
4.4	Flipping an image	113
4.5	Scaling an image	114
4.6	Clipping an image	115
4.7	Changing colors of an image	116
4.8	JPEG image processing flow (Pete, 2018)	120
4.9	Image ELA features	122
4.10	Model network structure	128
5.1	ResNet network block structure (He et al., 2016)	137
5.2	Two shortcut connection methods (He et al., 2016)	138
5.3	Increasing network layers leads to greater errors (He et al., 2016)	139
5.4	Two structures of residual block (He et al., 2016)	140
5.5	Inception module (Szegedy et al., 2014)	144
5.6	Inception module with dimensionality reduction (Szegedy et al., 2014)	144
5.7	A block with cardinality value of 32 (Xie et al., 2017)	149
5.8	DeepfakeNet architecture	150
5.9	Bottleneck structure of conv2 Layer	156
5.10	Bottleneck-0 structure of conv3 layer	157
5.11	Bottleneck-1 structure of conv3 layer	158

5.12	Bottleneck-0 structure of conv4 layer	159
5.13	Bottleneck-1 structure of conv4 layer	160
5.14	Bottleneck structure of conv5 layer	161
5.15	DeepfakeNet code flow chart	162
5.16	DeepfakeNet curve of loss value with training epochs	164
5.17	DeepfakeNet curve of accuracy value with training epochs	164
6.1	VGG19 curve of loss value and accuracy with training epochs	175
6.2	GoogLeNet curve of loss value and accuracy with training epochs	176
6.3	XceptionNet curve of loss value and accuracy with training epochs	176
6.4	ResNet101 curve of loss value and accuracy with training epochs	177
6.5	ResNeXt50 curve of loss value and accuracy with training epochs	177
6.6	DeepfakeNet curve of loss value with training epochs	178
6.7	DeepfakeNet curve of accuracy with training epochs	178
6.8	Samples of experimental results	180



LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	FFmpeg parameter description	220
B	The main code for DeepfakeNet, using Pytorch	224



LIST OF ABBREVIATIONS

AI	-	Artificial Intelligence
BGD	-	Batch Gradient Descent
BN	-	Batch Normalization
CFFN	-	Common Fake Feature Network
CNN	-	Convolutional Neural Networks
DCT	-	Discrete Cosine Transform
DFAE	-	Deepfake Autoencoder
DFD	-	Deepfake Detection
DFDC	-	Deepfake Detection Challenge
DL	-	Deep Learning
DWT	-	Discrete Wavelet Transform
ELA	-	Error Level Analysis
EXIF	-	Exchangeable Image File Format
FC	-	Full Convolution
FCN	-	Full Convolution Neural Network
FF	-	FaceForensics
FFmpeg	-	Fast Forward Moving Picture Expert Group
FLOPs	-	Floating Point Operations
FN	-	False Negative
FP	-	False Positive
GAN	-	General Adverse Network
GPU	-	Graphic Processing Unit
IPM	-	Integral Probability Metrics
LAE	-	Locality-aware AutoEncoder

LBP	-	Local Binary Pattern
LR	-	Learning Rate
LSTM	-	Long Short-Term Memory
MBGD	-	Mini-Batch Gradient Descent
ML	-	Machine Learning
MLP	-	Multilayer Perceptron
PDR	-	Predictive, Descriptive, Relevant
PRNU	-	Photo Response Non-Uniformity
RAM	-	Random Access Memory
RF	-	Random Forest
RGB	-	Red-Green-Blue
RNN	-	Recurrent Neural Network
SGD	-	Stochastic Gradient Descent
SVD	-	Singular Value Decomposition
SVM	-	Support Vector Machines
TN	-	True Negative
TP	-	True Positive
VDM	-	Variable Divergence Minimization
VGG	-	Visual Geometry Group

LIST OF SYMBOLS

- Tanh* - Tanh activation function
ReLU - ReLU activation function



LIST OF PUBLICATIONS

1. Gong, D., Sing, G. O., Kumar, Y. J., Ye, Z. and W, Chi., 2020. Deepfake Forensics, an AI-synthesized Detection with Deep Convolutional Generative Adversarial Networks. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), pp. 2861–2870.
2. Gong, D., Sing, G. O., Kumar, Y. J., Ye, Z. and W, Chi., 2021. DeepfakeNet, an Efficient Deepfake Detection Method. *International Journal of Advanced Computer Science and Applications*, 12(6), pp. 201–207. (SCOPUS indexed)
3. Gong, D., Kumar, Y. J., Sing, G. O., Choo, Y. H., Ye, Z. and W, Chi., 2022. An Improved Deepfake Detection Method Based on CNNs. *Journal of Theoretical and Applied Information Technology*, 100(18), pp. 5684-5691. (SCOPUS indexed)
4. Jin Z., Liu L., Gong D., Li L., 2021, Target Recognition of Industrial Robots Using Machine Vision in 5G Environment. *Frontiers in Neurorobotics*, 15, pp. 1-9. (SCOPUS indexed, SCI indexed)
5. Gong D., Cai S., Chi W., 2022. Research on the feedback teaching of AI customized development. *Scientific Journal of Intelligent Systems Research*, 4(7), pp. 700-708.
6. Chi W., Choo, Y. H., Sing, G. O., Gong D., 2022. Lung Disease Diagnosis based on Transfer Learning. *Journal of Artificial Intelligence Practice*. 5(1), pp. 98-104.
7. Gong, D., Kumar, Y. J., Sing, G. O., Ye, Z. and W, Chi., 2021. Data Enhancement Technology on Deepfake Dataset. in *3rd International Conference on Intelligent and*

Interactive Computing 2021, pp. 1-2.

8. Ye, Z., Kumar, Y. J., Sing, G. O. and Gong, D., 2021. Research on Upgrading of Traditional Industry Driven by Intelligent Manufacturing. in *3rd International Conference on Intelligent and Interactive Computing 2021*, pp. 85.

