



Faculty of Electronic and Computer Engineering

**WiFi MAC ADDRESS TAGGING ASSISTED FAST SURVEILLANCE
VIDEO RETRIEVAL SYSTEM**

Tan Kien Leong

Master of Science in Electronic Engineering

2020

**WiFi MAC ADDRESS TAGGING ASSISTED FAST SURVEILLANCE VIDEO
RETRIEVAL SYSTEM**

TAN KIEN LEONG

**A thesis submitted
in fulfillment of the requirements for the degree of Master of Science
in Electronic Engineering**

Faculty of Electronic and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020

DECLARATION

I declare that this thesis entitled “WiFi MAC Address Tagging Assisted Fast Surveillance Video Retrieval System” 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 :

Date :

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 Master of Science in Electronic Engineering.

Signature :

Supervisor Name :

Date :

DEDICATION

To my beloved father and mother

ABSTRACT

Conventional public safety surveillance video camera systems required 24/7 monitoring of security officers with video wall display installed in the control room. When a crime or incident is reported, all the recorded surveillance video streams nearby the incident area are played back simultaneously on video wall to help locate the target person. The security officers can fast forward the video playback to speed up the video search but it requires massive manpower if there are hundreds of video streams or multiple target persons required to be examined on the video wall. Even today with the Graphics Processing Unit (GPU) that is able to run the person search deep neural network model to automatic search for the target person from a large video database, it can take hours or even days to complete the search. This research aims to determine how to prioritize the surveillance camera video frames that need to be processed by the person search deep neural network model to reduce the time taken for getting the target person in the next camera (the cameras that may recorded the target person according to walkway topology). Thanks to the advancement in artificial intelligence, a person search deep neural network model trained to correctly match thousands of identical person can be used to automate the person search process. The person search matching process required the person in the image to be firstly detected before the matching can be carried out. Eight deep neural network based object detection models are re-trained on 55,272 labelled persons to determine the suitable object detection model that can be used to replace the person detection part of the person search model. As a result, applying Model 3 (Darkflow) for person detection is found to be able to provide reasonable speed/accuracy trade-off (0.62 mAP and 0.04s mean inference time). To further reduce the required time of automated person search without having to scale up the computing hardware, additional metadata (WiFi MAC address of smartphone) collected during the occurrence of the incident can be used to prioritize the retrieving of surveillance video frames for subsequent person search. Three ways of retrieving surveillance video are compared, in term of time taken for getting the target person, with a constructed testbed in UTeM. The developed WiFi sniffer enabled surveillance camera, with 3-stage WiFi frame inspection and the use of collected WiFi signal strength for filtering, is able to tag the collected WiFi MAC addresses to the surveillance video frames according to the time of the MAC address is sniffed. Using the formulated mathematical model, the proposed WiFi MAC address tagging assisted fast surveillance video retrieval method performs 9.6 times better in single person search and 6.2 times better in multiple persons search provided the WiFi MAC address of the target's smartphone is sniffed by the WiFi sniffer of the surveillance camera. Based on these results, the proposed fast video retrieval system with MAC address tagging is proven to take less time to get target person in the next camera as compared to video retrieval system without MAC address tagging. Further research is needed to identify how to prioritize the WiFi MAC address searching when multiple WiFi MAC addresses are sniffed.

ABSTRAK

Sistem kamera pengawasan konvensional memerlukan pemantauan oleh pegawai keselamatan sepanjang hari di dalam bilik kawalan. Apabila jenayah atau insiden dilaporkan, semua rakaman video pengawasan yang dirakam berhampiran kawasan kejadian akan diulang semula secara serentak untuk mencari sasaran orang. Pegawai keselamatan boleh mempercepatkan masa ulangan video untuk mengenalpasti orang yang perlu disasarkan tetapi ia memerlukan tenaga kerja yang banyak jika terdapat ratusan rakaman video atau beberapa sasaran yang perlu dipantau. Walaupun Graphics Processing Unit (GPU) dapat mengaplikasi model rangkaian neural untuk melaksanakan proses pencarian orang secara automatik, tetapi ia masih mengambil masa yang panjang untuk proses pencarian orang jika terdapat ratusan video yang perlu diproses. Kajian ini bertujuan untuk menentukan bagaimana mengutamakan rangka video yang perlu diproses oleh model rangkaian neural bagi mengurangkan masa yang diperlu untuk mencari orang sasaran dalam kamera yang berikut (kamera yang mungkin dikunjungi oleh sasaran berdasarkan topologi laluan). Seiring dengan pencapaian dan kemajuan dalam kecerdasan buatan, model rangkaian neural mampu melaksanakan proses pencarian orang secara automatik. Proses pepadanan orang memerlukan imej orang dikesan terlebih dahulu. Lapan model pengesanan objek berasaskan rangkaian neural yang mendalam dilatih semula pada 55,272 orang berlabel untuk menentukan model pengesanan objek yang sesuai seterusnya menggantikan bahagian pengesanan model carian tersebut. Keputusannya, Model 3 (Darkflow) diaplikasikan untuk pengesanan orang mampu memberikan keseimbangan antara kelajuan dan ketepatan dalam melaksanakan tugas (0.62 mAP dan 0.04s purata masa inferensi). Metadata tambahan (alamat MAC WiFi telefon pintar) yang dikumpulkan semasa kejadian boleh digunakan untuk mengutamakan pencarian rangka video yang berkaitan untuk pemprosesan imej berikutnya. Melalui cara ini, masa yang diperlukan untuk pencarian orang dapat dikurangkan tanpa perlu menaikkan taraf perkakasan komputer yang tersedia ada. Tiga cara telah dibandingkan dari segi masa untuk mendapatkan sasaran telah dijalankan pada kawasan ujian yang bertempat di UTeM. Snifer WiFi yang terletak dalam kamera pengawasan mempunyai fungsi tiga peringkat pemeriksaan rangka WiFi dan penggunaan kekuatan isyarat WiFi yang dikumpul untuk penapisan rangka WiFi. Ia digunakan untuk menanda alamat MAC WiFi yang telah dikumpul ke rangka video mengikut masa alamat MAC yang dikesan. Dengan penggunaan model matematik yang dicadangkan, kaedah pencarian video yang dibantu dengan alamat MAC WiFi dapat memberikan 9.6 kali lebih pantas dalam carian orang tunggal dan 6.2 kali lebih pantas dalam mencari orang untuk jumlah yang banyak, dengan andaian alamat MAC WiFi yang dimiliki oleh sasaran orang dapat dikesan oleh snifer WiFi. Kesimpulannya, kaedah pencarian video yang dibantu dengan alamat MAC WiFi terbukti mengambil masa yang lebih pendek berbanding dengan kaedah pencarian video yang tidak dibantu dengan alamat MAC WiFi. Kajian lanjutan diperlukan untuk mengenal pasti bagaimana untuk mengutamakan alamat MAC WiFi apabila terdapat banyak alamat MAC WiFi yang dikumpul.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my supervisor, Associate Professor Dr. Lim Kim Chuan, for the continuous support of my master study and related research, for his patience, motivation, and immense knowledge.

I would also like to thank my co-supervisor, Associate Professor Dr. Soo Yew Guan, for his guidance in WebApp video surveillance dashboard interface development.

I would also like to thank Collaborative Research in Engineering, Science and Technology (CREST) and Recogine Technology Sdn Bhd for awarded me the Graduate Research Assistant Scholarship Program (GRASP) and provide me the experiment equipment used to build the testbed.

I would also like to thank my colleagues and lab technician in Research Lab 3, Faculty of Electronics and Computer Engineering of UTeM who have provided assistance when I was doing my research.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF APPENDICES	xi
LIST OF ABBREVIATIONS	xii
LIST OF PUBLICATIONS	xiv
CHAPTER	
1. INTRODUCTION	1
1.1 Background	1
1.2 Problem statement	4
1.3 Research question	6
1.4 Hypothesis	6
1.5 Objectives	7
1.6 Research scope	8
1.7 Contribution of research	8
2. LITERATURE REVIEW	9
2.1 Metadata in the surveillance video	9
2.2 Related works on person search by using deep learning	11
2.2.1 Person detection model	13
2.2.1.1 Region proposal based	14
2.2.1.2 Regression based	23
2.2.1.3 TensorFlow object detection API	29
2.2.2 Person re-identification model	29
2.2.3 Person search model	33
2.2.4 Person search datasets	37
2.3 IEEE 802.11 standard	39
2.4 MAC frame format	40
2.5 Related works on WiFi sniffer	41
2.6 Summary	42
3. METHODOLOGY	44
3.1 Experiment setup for convolutional object detector in person detection with Tensorflow	45
3.1.1 Training strategy for transfer learning	46
3.1.2 Benchmarking hardware and system	47
3.1.3 Determine the accuracy of person detection with mean average precision (mAP)	47
3.2 Design of WiFi sniffer for WiFi MAC address tagging	49
3.2.1 WiFi sniffer setup	50
3.2.2 Smartphones setting for different WiFi connection	55

status	
3.2.3 Applying RSSI thresholding for WiFi MAC address tagging during the recording of surveillance camera video frame	56
3.2.4 Channel hopping time calculation	60
3.3 Testbed for person search	63
3.4 Methods used in single and multiple persons search experiments	64
3.4.1 Single person search experiment	67
3.4.2 Multiple persons search experiment	68
3.4.3 Evaluation metric	70
3.5 Summary	75
4. RESULT AND DISCUSSION	77
4.1 Performance analysis for convolutional object detector in person detection with Tensorflow	77
4.1.1 Number of training epoch versus mAP	77
4.1.2 Test-time performance	82
4.1.3 Calculate number of FLOPs	83
4.1.4 Person detection accuracy (mAP) versus mean inference time	84
4.1.5 Mean inference time versus number of Conv2D layer	85
4.2 WiFi MAC address sniffing	87
4.2.1 Analysis on broadcast frequency of WiFi frame types under different scenario	87
4.2.2 Comparison of RSSI values at different distance from the WiFi sniffer	89
4.2.3 The effect of RSSI thresholding on the probability of misidentify the distance between the sniffed WiFi MAC address and WiFi sniffer	92
4.3 Performance analysis for target search in video database with MAC address filtering	96
4.3.1 Performance analysis of proposed MAC address filtering technique for single person search	96
4.3.2 Performance analysis of proposed MAC address filtering technique for multiple person search	105
4.4 Summary	113
5. CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK	118
5.1 Conclusion	118
5.2 Recommendation for future work	122
5.2.1 Re-train the person search deep neural network model with model 3 (Darkflow)	122
5.2.2 Speeding up the deep learning model inference process	123
5.2.3 Intelligent on the edge	124
5.2.4 Prioritize the WiFi MAC address searching	124
REFERENCES	126
APPENDICES	138

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Two types of object detection methods	14
2.2	Summary of region proposal based person detection model	22
2.3	Summary of regression-based person detection model	28
2.4	Summary of person search model	37
2.5	Overall data statistics for CUHK-SYSU and PRW dataset	38
2.6	Statistics of the CUHK-SYSU and PRW dataset with respect to training and testing splits	38
2.7	Summary of spectrum properties of IEEE 802.11a/b/g/n	40
3.1	New model name for the 8 object detection models	45
3.2	Summary of training strategy	47
3.3	WiFi frames sniffed by WiFi sniffer	53
4.1	mAP at 45 th epoch of 8 object detection models	79
4.2	Inference result of the 8 models on the selected test images in CUHK-SYSU person search dataset	80
4.3	Number of Conv2D layer and mean inference time of 8 object detection models	86
4.4	Time taken by method 2 and method 3 for getting the target in the next camera in single and multiple persons search experiment (F : Total surveillance video frames to be processed)	117

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Example of automated person detection (CUHK-SYSU person search dataset, 2016)	2
1.2	(a) The persons of input image are automatic labelled with a unique ID (b) The result of person re-identification (CUHK-SYSU person search dataset, 2016)	3
1.3	Path taken by person A	7
2.1	Example of person detection	11
2.2	Example of person re-identification	12
2.3	Example of person search	12
2.4	(a) Sliding window is used to find objects in an image (b) Different sizes and aspect ratios of sliding windows are used for different objects or same kind of object	14
2.5	The flowchart of R-CNN (Girshick et. al., 2014)	15
2.6	The flowchart of SPP-net for object detection (He et. al., 2015)	16
2.7	The flowchart of Fast R-CNN (Girshick, 2015)	17
2.8	The flowchart of Faster R-CNN (Ren et. al., 2015)	18
2.9	The output of Region Proposal Network (RPN) in Faster R-CNN (Ren et. al., 2015)	18
2.10	The flowchart of R-FCN (Li et. al., 2016)	20
2.11	The flowchart of Mask R-CNN (He et. al., 2017)	21
2.12	The flowchart of AttentionNet (Yoo et. al., 2015)	24
2.13	The flowchart of G-CNN (Najibi et. al., 2016)	25
2.14	(a) Apply multi-scale regular grid over the input image (b) Adjust the bounding boxes iteratively in the grid (c) Each iteration will move the bounding boxes toward the objects and classify the objects	25
2.15	The network architecture of YOLO (Redmon et. al., 2016)	26
2.16	(a) The input image is divided into a 13×13 grid (b) Each grid cell predicts B boundary boxes and each box has one confidence score (c) Non-maximum suppression is applied to remove overlapping bounding box	27
2.17	The architecture of SSD (Liu et. al., 2016)	27
2.18	The architecture of Siamese neural network (Bromley et. al., 1993)	30
2.19	The architecture of DML (Yi et. al., 2014)	31
2.20	The architecture of FPNN (Li et. al., 2014)	31
2.21	An improved Siamese model with cross-input neighbourhood differences (Ahmed et. al., 2015)	32
2.22	The Siamese model with RNN (McLaughlin et. al., 2016)	33
2.23	The architecture of RSS person search (Xiao et. al., 2016)	34

2.24	The architecture of OIM person search (Xiao et. al., 2017)	34
2.25	The architecture of IAN (Xiao et. al., 2017)	35
2.26	The architecture of NPSM (Liu et. al., 2017)	36
2.27	The architecture of I-Net (He et. al., 2018)	36
2.28	The 13 channels in 2.4GHz frequency band	40
2.29	General MAC frame format	41
3.1	Example image with TP, FP and FN (CUHK-SYSU person search dataset, 2016)	49
3.2	The monitor mode is not one of the supported interface modes	50
3.3	Nexmon firmware patch installation flow	51
3.4	Execute command <i>iw list</i> to check the monitor mode is supported after the installation of Nexmon firmware patch	52
3.5	WiFi sniffer workflow	54
3.6	(a) Person A is out of camera viewing distance but the smartphone's MAC address is sniffed (b) Only person A is within the camera viewing distance but the smartphone's MAC address of person A and other persons (B, C, and D) are sniffed	57
3.7	(a) The sniffed WiFi MAC address is within camera viewing distance but misidentify as outside camera viewing distance (b) The sniffed WiFi MAC address is outside camera viewing distance but misidentify as within camera viewing distance	59
3.8	Channel hopping flowchart	60
3.9	Mean channel switching time	62
3.10	Location of the three cameras (C1, C2 and C3)	63
3.11	Overall system design of proposed fast video retrieval system for person search	64
3.12	Path taken by person A	67
3.13	Timeline of person A appears in different cameras cameras (surveillance video is played back at time T_0 and paused at T_1)	68
3.14	Path taken by person A and person B	69
3.15	Timeline of person A and person B appearing in different cameras (surveillance video is played back at time T_0 and paused at T_1)	70
3.16	Single person search level	72
3.17	Multiple person search level (a) Person A and person B are moving to different direction (b) Person A person B are moving toward the same direction	73
3.18	Application of F_{search}^k and F_n^k in method 2 and method 3 for single person search	74
3.19	Application of F_{search}^k and F_n^k in method 2 and method 3 for multiple person search	75
4.1	mAP versus number of epochs for the 8 object detection models	79
4.2	Inference time for 2000 images (600 × 600 pixels) for the 8 CNN base object detection models implemented with TensorFlow framework on Nvidia GTX 1080 GPU	83
4.3	Mean inference time (Nvidia GTX 1080) and number of FLOPs for the 6 models	84
4.4	mAP at 45 th epoch versus Mean inference time (Nvidia GTX 1080)	85
4.5	Mean inference time (Nvidia GTX 1080) of 8 object detection models	86

	versus number of Conv2D layer	
4.6	Number of sniffed probe request in five minutes for scenario 1	87
4.7	Number of sniffed packets for Huawei P10 Android 8.0.0 in 5 minutes duration	88
4.8	Number of sniffed packets for Samsung Galaxy Note 4 Android 6.0.1 in 5 minutes durations	89
4.9	The RSSI distribution of both smartphones at (a) 1.5 meters, (b) 3 meters, (c) 5 meters, and (d) 10 meters over 5 minutes duration	91
4.10	(a) The RSSI threshold is set to -59 dBm if 90% of the MAC addresses at 5 meters are required to be tagged to surveillance camera video frames (b) Applying threshold value of -59dBm to the collected RSSI values of WiFi MAC address at distance of 10 meters resulting in 43% of the MAC addresses wrongly tagged as 5 meters (c) & (d) The probability of misidentify the smartphone is outside camera viewing distance is very low (less than 3%) and can be ignored	93
4.11	Difference between the probability of misidentify the smartphone is outside or within camera viewing distance. The minimum point is located at -56 dBm	95
4.12	The probability of misidentify the smartphone is outside or within camera viewing distance are almost the same when the RSSI threshold is -56dBm, (a) $P_{wrongdiscard}$ of 16% and (b) $P_{wrongtag}$ of 18.3%	95
4.13	Person A appears on camera C2 with WiFi MAC address tagged (circled with red circle) in the single person search experiment	98
4.14	The result of MAC address filtering technique (circled with red dash circle) and the result of person search on camera C2 (person A in green bounding box) in the single person search experiment	99
4.15	The result of person search on camera C3 (person A in green bounding box) in the single person search experiment	99
4.16	Comparison of video duration that need to be processed by person search deep neural network model for method 2 and method 3 in the single person search experiment. Person A re-identification is firstly performed on camera C2 (red segment T_1 to T_2) and camera C3 (red segment T_3 to T_4)	100
4.17	Average $k\%$ of video surveillance frames tagged with WiFi MAC address, before and after discarded the MAC addresses of stationary WiFi devices nearby camera C2 for a single person search	102
4.18	Average $k\%$ of video surveillance frames tagged with WiFi MAC address of different cameras for a single person search	103
4.19	Person A and person B appear on camera C2 with WiFi MAC addresses tagged (circled with red circle) in the multiple persons search experiment	107
4.20	The result of MAC address filtering technique (circled with red dash circle) and the result of person search on camera C2 (Person A and Person B in green bounding box) in the multiple persons search experiment	107
4.21	The result of person search on camera C3 (Person A in green bounding box) in the multiple persons search experiment	108
4.22	The result of person search on camera C1 (Person B in green bounding box) in the multiple persons search experiment	108
4.23	Comparison of video duration that needs to be processed by person search deep neural network model for method 2 and method 3 in the multiple persons search experiment. Person A re-identification is first performed on	109

camera C2 (red segment T_1 to T_2) and camera C3 (red segment T_5 to T_6).
 Person B re-identification is first performed on camera C2 (red segment T_1 to T_2) and camera C1 (red segment T_3 to T_4)

4.24	Average $k\%$ of video surveillance frames tagged with WiFi MAC address, before and after discarded the MAC addresses of stationary WiFi devices nearby camera C2 for multiple persons search	110
4.25	Average $k\%$ of video surveillance frames tagged with WiFi MAC address of different cameras for multiple persons search	110
4.26	Average $k\%$ of video surveillance frames tagged with WiFi MAC address of different cameras with MAC address of stationary WiFi devices removed for single and multiple person search	116
5.1	Components of OpenVINO toolkit	123
5.2	Components of TensorRT	123
5.3	WiFi MAC addresses with its corresponding probability. The WiFi MAC address with highest probability will be given priority to be processed by the proposed MAC address filtering technique	125

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	RSSI distribution of two smartphones at different distances	138

LIST OF ABBREVIATIONS

AI	- Artificial Intelligence
AP	- Average Precision
API	- Application Programming Interface
cm	- centimetre
CNN	- Convolutional Neural Network
Conv2D	- Two-dimensional convolution
CUDA	- Compute Unified Device Architecture
cuDNN	- CUDA Deep Neural Network
dBm	- decibels (dB) with reference to one milliwatt
DS	- Distribution System
FLOPs	- Floating point operations
FN	- False Negative
FP	- False Positive
fps	- frame per second
GB	- Gigabyte
GHz	- Gigahertz
GPU	- Graphics Processing Unit
IANA	- Internet Assigned Numbers Authority
ID	- Identity
IEEE	- Institute of Electrical and Electronics Engineers

IoU	- Intersection over Union
IP	- Internet Protocol
LTS	- Long Term Support
m	- Meter
MAC	- Media Access Control
mAP	- Mean Average Precision
Max	- Maximum
Min	- Minimum
mm	- millimetre
no.	- number
OpenCV	- Open source Computer Vision
OS	- Operating System
OUI	- Organizational Unique Identifier
PIL	- Python Imaging Library
PVC	- Polymerizing Vinyl Chloride
RAM	- Random Access Memory
RSSI	- Receive Signal Strength Indication
s	- seconds
SSID	- Service Set Identifier
TP	- True Positive
WEP	- Wired Equivalent Privacy
WPA	- WiFi Protected Access
WPA 2	- WiFi Protected Access II

LIST OF PUBLICATIONS

Journal

1. Tan, K.L. and Lim, K.C., 2019. Fast Surveillance Video Indexing & Retrieval with WiFi MAC Address Tagging, *Indonesian Journal of Electrical Engineering and Computer Science*, 16(1), pp. 473-481.

Technical Report

1. Tan, K.L., Lim, K.C. and Tan, X.Y., Alvin, S., 2018. MAC Address tagged Fast Video Retrieval System across Multiple Cameras for Person Re-identification, *Recognize Technology Sdn Bhd*.

CHAPTER 1

INTRODUCTION

The WiFi MAC address tagging assisted fast surveillance video retrieval system with deep learning based person search will be firstly explained at the beginning of this chapter. The problem statement, research question, hypothesis, objectives and research scope of this research are presented subsequently. The contribution of this research is presented at the end of this sectionp.

1.1 Background

Surveillance camera systems have been widely used in public places in recent years for the purpose of public safety. Traditional type surveillance camera systems are monitored by security officers 24 hours per day. When an incident has taken place, all the recorded surveillance video streams nearby the incident area are playback simultaneously to help locate the target person. The security officer usually fast forward the video playback to speed up the video search.

Deep learning is one of the greatest achievements in the field of computer vision over the last decade. It outperforms the handcrafted methods in image classification (Krizhevsky et. al., 2012), object detection (Ren et. al., 2015), object segmentation (He et. al., 2017), and others. An automatic approach to search for a target person from the surveillance camera video database is by applying deep neural network model. A person search deep neural network model consists of two parts; the first part is person detector while the second part is person re-identification. The person detector is used to automate locate the persons inside an

image. Red bounding boxes are drawn on the image to indicate where the persons are located (see Figure 1.1). The person feature inside the bounding box is extracted and passed to person re-identification model. It compares the similarity of person features of the target person with the person features of the persons appeared in recorded surveillance video frames. For example, given an input image which contains five target persons, the target persons are automatically labelled with a box and a unique ID as shown in Figure 1.2(a). The person re-identification model will search through the surveillance camera video database to find the images that contain either one of the target persons. An example of the result of person re-identification is as shown in Figure 1.2(b). Same ID will be assigned to the person detected in the recorded surveillance camera video frame if that person is detected and recognized as the same person in the input image. The percentage value displayed beside the ID is the percentage of similarity (in between the range of 0%-100%).



Figure 1.1: Example of automated person detection (CUHK-SYSU person search dataset,

2016)

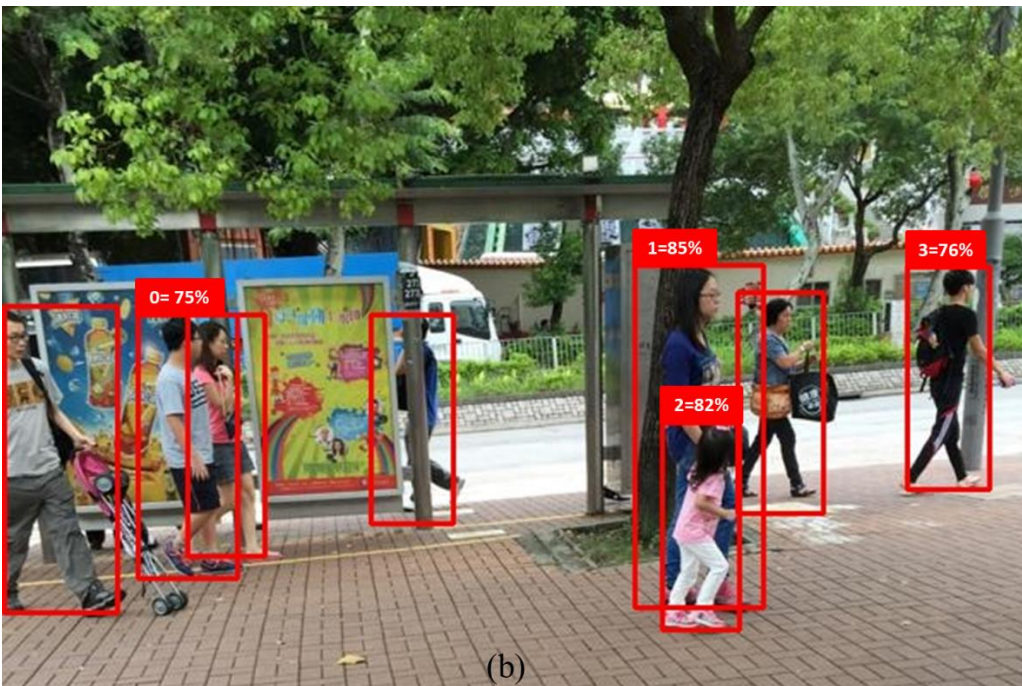
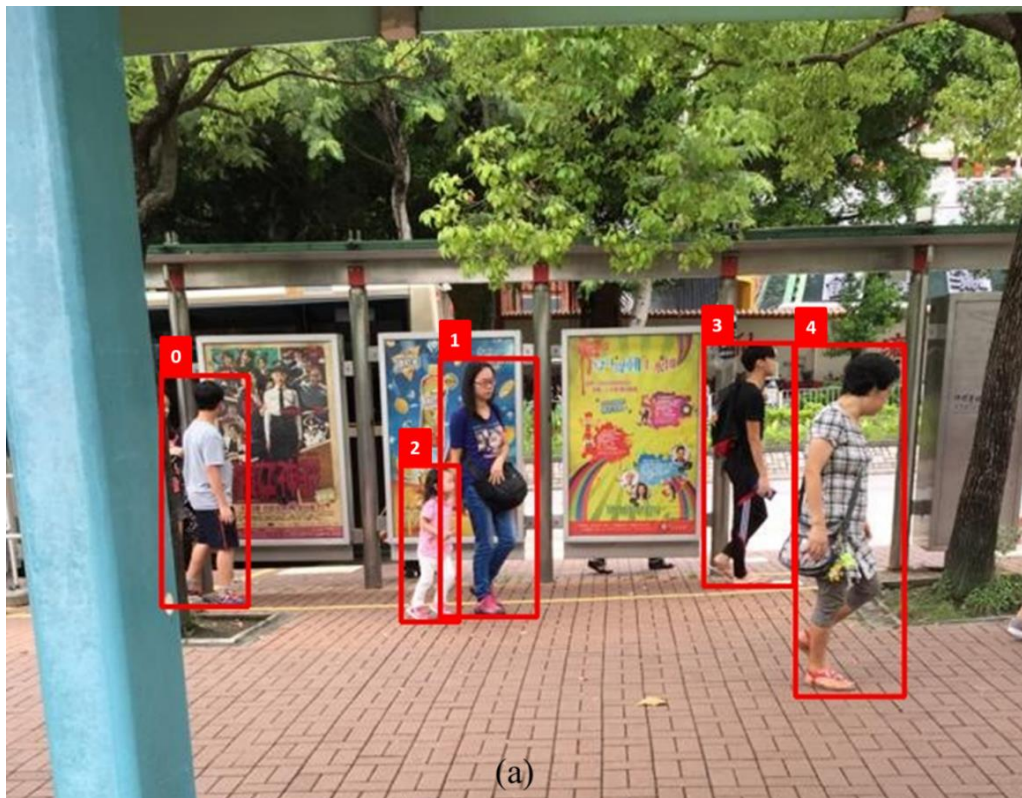


Figure 1.2: (a) The persons of input image are automatic labelled with a unique ID (b) The result of person re-identification (CUHK-SYSU person search dataset, 2016)

Nowadays, people carry their smartphone with WiFi turned on wherever they go in order for their smartphone to automatically connect to Internet when there is WiFi service available. When the WiFi of the smartphone is turned on, the smartphone will broadcast management frames known as probe request to discover all nearby access point. The probe request contains WiFi media access control (MAC) address which is a unique identifier assigned to the network interface card of the smartphone. The length of WiFi MAC address is six bytes and separated by colons. The MAC address is generated by using OUI (Organizationally Unique Identifier) number provided by IANA (Internet Assigned Numbers Authority). The MAC address is never encrypted even though the WiFi devices are connected to a WiFi network with security encryption enabled (Wired Equivalent Privacy (WEP), WiFi Protected Access (WPA), and WiFi Protected Access II (WPA2)). Besides probe request, there are other WiFi frames broadcast by smartphone that contain WiFi MAC address for example: disassociation, authentication, deauthentication, and so on. Since each smartphone has a unique WiFi MAC address, it can be used to track a person who carries a smartphone with WiFi turned on. Research work has been carried out using WiFi MAC address for WiFi tracking (Petre et. al., 2017) (Julien, 2015) (Xu et. al., 2013) (Musa et.al., 2012).

1.2 Problem statement

With rapid urbanization, IP cameras are almost everywhere in our daily life such as pedestrian walkways, road junctions, schools, markets and public transport which provide a greater level of public safety. However, these IP cameras generate large volume of video data across time. This becomes an issue when there is an incident that happens for example a lost child or crime suspect, and the surveillance camera video frames that contain the target

person need to be retrieved in the shortest time to minimize further damage. Every second counts to avoid the situation from becoming more serious.

In the old days, human beings needed extensive time and intensive man-power to retrieve surveillance camera video frames that contained target person from a large video database. When an incident occurred, all the recorded surveillance video streams nearby the incident area were playback simultaneously to help locate the target person. The security officers would fast forward the video playback to speed up the video search but it required massive manpower if there were hundreds of video streams that needed to be monitored. The situation became more challenging to the security officers if there were multiple targets being tracked at the same time. Hence, an automatic approach to search for target person by applying deep learning model is applicable.

The speed and accuracy of a person search deep neural network model is mainly affected by the person detection part. A good person detector can provide more useful person features for the person re-identification model to perform person re-identification process (useful person features provide better accuracy in person re-identification). A deeper neural network model can extract more useful person features (higher accuracy) but at the same time it will consume more processing time (slower speed) (Huang et. al. 2016). Therefore, an appropriate person detector which provides reasonable speed and accuracy trade off need to be selected for person search deep neural network model.

Even today with the Graphics Processing Unit (GPU) that is able to run the person search deep neural network model to automatic search for the target person from a large video database, it can take hours or even days to complete the search. The video processing time can be reduced by using multiple desktop GPUs or a more powerful server grade GPU to run the person search deep neural network model, but this requires more investment on GPU and it is not cost-effective to allocate the GPU for every available camera. This has