Northwestern | THE GRADUATE SCHOOL

Application for Admission

App Type **New Student**

Submitted Date **11-24-2018**

App ID# **78454995**

Intended Status **Full-time**

Entry Quarter **Fall 2019**

Prior TGS Applicant (Program)

Last Name **Sun**

First **Ping**

Middle

Gender Pronouns (US only)

Birthdate **05-11-1997**

Gender **Male**

Program **Computer Science: MS**

Specialization/Area of Interest **Artificial Intelligence and Machine Learning**

Cluster

Secondary PhD (MEAS Only)

MS Consideration (MEAS Only)

JD/PhD **No**

DPT/PhD **No**

Fee Waiver

US Vet/Active Forces

Ethnicity **Asian**

Hispanic **No**

Citizenship **CHINA**

Visa

Citizenship Status **International Student**

Country of Birth **CHINA**

Green Card #

Current Address
ROOM424, EAST1,
No.800 Dongchuaj Rd.,Minhang Dist., Shanghai
Shanghai, 200240
CHINA

Current Phone **+86 15821936985**

Cell Phone **+86 15821936985**

Email Address **spofnewage@sjtu.edu.cn**

Permanent Address
ROOM1403, 452 Qinyang Rd.
Chengguan Dist., Lanzhou
Lanzhou, 730030
CHINA

Permanent Phone **+86 15821936985**

Preferred Phone Number **Current Phone Number**

Previous Institution	From	To	Field of Study	Level	Degree	Date
Shanghai Jiaotong University	09-01-2015	06-30-2019	Computer Science and		International Undergraduate Degree	

Cumulative UG GPA	**3.38**	UG Junior/Senior Year GPA	**3.83**
Cumulative UG GPA - Unconverted	**3.4/4.3**	Max UG GPA Scale	**3.38/4.00**
Cumulative Grad GPA			
Cumulative Grad GPA - Unconverted		Max Grad GPA Scale	

Letters of Recommendation

1. **Guangtao Xue** **xue-gt@cs.sjtu.edu.cn**
2. **Liqing Zhang** **zhang-lq@cs.sjtu.edu.cn**
3. **Xiaoju Dong** **dong-xj@cs.sjtu.edu.cn**
4.
5.

Are you interested in studying with specific faculty members? (List names below)

1. First Name **OLIVER** Last Name **SCOSSAIRT**
2. First Name **KRISTIAN** Last Name **HAMMOND**
3. First Name **IAN** Last Name **HORSWILL**
4. First Name **HAOQI** Last Name **ZHANG**

Please indicate the highest level of education completed by your parent(s) or guardian(s) (the one or two people most responsible for raising you)

First individual's highest level of education completed: **Some college or university**

If other, please explain:

Second individual's highest level of education completed: **High school diploma or equivalent**

If other, please explain:

Language	Reading	Writing	Speaking

Self-Reported Test Scores

GRE Gen	09-14-2018	Verbal	154	65	Quant	170	96	A.W.	3.0	17					
GRE Sub						LSAT									
TOEFL	01-13-2018	Ovr	106	Read	29	List	27	Speak	22	Writ	28	IELTS		Ovr	
GMAT		Tot			Verb			Quant			A.W.			I.R.		
MCAT		Bioscience			Verbal			Physical Science							

Please list any honors you have been awarded

National Encouragement Scholarship 2016.

Have you applied for or been awarded an external fellowship?

Yes ⊗ No ○ If yes, please specify;

National Encouragement Scholarship 2016.

Please describe your plans for the future.

I aspire to pursue creativity both in the world of art and computer science, bringing unique and innovative technologies into the world. My career goal is to become a researcher in the field of computer vision, devising innovative new ways to improve the daily lives of ordinary people. Within five years, I plan to work as an algorithm engineer in the field of vision, interactive media.

Other Universities Applied (in preferred rank order)

1. School Drop Down **Carnegie Mellon**
2. School Drop Down **Georgia Institute of Tech**
3. School Drop Down **New York University**
4. School Drop Down **University of California-San Diego**
5. School "other" **University of California-Irvine**
6. School "other"
7. School "other"
8. School "other"

Academic misconduct? Yes ○ No ⊗ Convicted of crime? Yes ○ No ⊗

If answered yes, applicant is asked to upload explanation. If uploaded, explanation will be attached to end of application PDF.

Statement of Purpose

Ping Sun *CS Master Applicant*

My parents, like many people of their generation, are not so skilled in using various electronic devices and even afraid to download files from the Internet, saying that "they would bring viruses to our computers". My dad still tends to assume that as long as I'm in front of a screen, I must be playing games. In short, my parents have very little understanding of the vast practical applications of computer science, and they were baffled and disappointed when I selected it as my undergraduate major. When my dad asked me one day what on earth I was learning in computer science and technology, I gave his question serious consideration before answering. I believe that computer science should not be a mysterious, distant concept; it should be a welcome enhancement to our daily lives. After four years study in Shanghai Jiao Tong University pursuing a bachelor's degree in computer science and engineering, I do not want to answer my dad's question with a list of dry technical facts about pipelines in computer architecture or convolutional neural networks in deep learning. I prefer to explain that computer science aims to improve the world by taking the advantages of computers' fast calculation, optimal algorithms, and accurate predictions. I am currently applying for a master's degree in computer science because I have a strong desire to become a researcher, solving problems in people's everyday lives.

I have great enthusiasm for computer science, and I enjoy projects wherein I can develop my own original ideas. For instance, for the final project of the course of Thinking and Approach of Programming, most other students programmed games of gobang or chess as the TA suggested; at the time, however, I was independently studying game theory, so I decided instead to create a game based on the Prisoners' Dilemma. I had a very strong desire to discover the thoughts of people when they compete with others in a game. I designed an AI robot that could adjust his strategy according to my actions so that he could learn to behave smarter. As I made the AI robot smarter using different algorithms, it became much more difficult to improve his performance. I searched for information on how to win the game in an iterated Prisoners' Dilemma. This was the first project I completed which included some aspect of rudimentary artificial intelligence.

I once held a dream to become a unique UI designer, using coding skills to display a more artistic world of computer science. But I ended up feeling a front-end engineer so limited in the way they create. After seeing the fascinating results of A Neural Algorithm of Artistic Style, I chose computer vision as my main research direction. People can use apps with style-transferring functions to create great artwork without even knowing the underlying algorithms. The immense power of deep learning further strengthened my enthusiasm for studying in this field; I felt more determined than ever to achieve my ultimate goal of providing people with a better world using artificial intelligence and computer vision capabilities. I then began to study deep learning while I engaged in the course of Computer Graphics. While studying the project of "Convolutional Pose Machine", I learned the importance of datasets, the structure of VGG and many other networks, and the training-testing process; most importantly, I became aware of the power and magic of neural networks, which could be used to achieve seemingly impossible tasks. Through the CPM project—my first research project, as well as my first academic paper—I learned many valuable research methods.

I was deeply intrigued by the creative possibilities of Generative Adversarial Networks, and my research in this area proved GAN to be an excellent tool in helping me achieve many of my creative ideas. Together with two other students, we introduced a text attention-based approach to changing the clothes of people in images. In previous work in this field, the GAN structure commonly had multiple layers and greatly decreased the image information to a very low resolution. Based on the Fashion GAN, we added a CGAN to the existing structure in order to increase the resolution of results to the same size as the original image. Through this experience, I recognized the possibilities of computer vision and its potential to help me achieve my dreams.

In the summer of my junior year, I served as an intern in the lab at Versa-SJTU AI. The goals of the Versa company are similar to my own: giving people the chance to create works of art with the assistance of artificial intelligence. During my six-month internship, I worked in the algorithm department; working on the topic of segmentation, I read over 150 academic papers and came up with many creative approaches to improve the performance of both semantic and instance segmentation. I tried many network structures, combined several objective functions, and replaced dataset COCO with the more precise dataset MHP. In the end, I designed several algorithms to provide segmentation with both affinity and higher accuracy, outperforming the state-of-the-art methods and turned my work into a paper submitted to CVPR 2019 as the first author.

I was fortunate to continue my further studies in computer vision and pattern recognition in the lab of Professor Liqing Zhang. Under the supervision of Prof. Zhang, I developed better research habits and a broader understanding of machine learning. I joined the weekly forums in order to exchange ideas with others and gave a presentation about my thoughts on the book *Pattern Recognition and Machine Learning*. I worked with a fellow student and submitted a paper entitled "*Attribute Boosted Person Re-Identification via Decision Making and Attention*" to the CVPR 2019. My most recent project, Generative Image Inpainting with Contextual Attention, involves removing unimportant objects from images and filling in the content automatically with their surroundings.

I aspire to pursue creativity both in the world of art and computer science, bringing unique and innovative technologies into the world. My career goal is to become a researcher in the field of computer vision, devising innovative new ways to improve the daily lives of ordinary people. Honestly, I have applied for graduate programs in many institutions, but I suppose that NWU attracts me the most in many ways. I am so delighted that the department of EECS has several research groups including Cognitive Systems, Graphics & Interactive Media which are exactly consistent with my research experience. And I feel even happier to notice that each group has many outstanding professors sharing the same research interests with me. I have done similar researches as Prof. Oliver Cossairt in the field of computational imaging measuring the depth information. I also find the idea of Prof. Kristian Hammond 's research *Integrating Computer Science and AI into all aspects of life* consistent with my goal. In addition, I have the same career goal to bridge HCI, AI, and Crowd Computing and to further engage in the field of interactive art with AI as Prof. Haoqi Zhang and Prof. Ian Horswell do. I would really appreciate the opportunity to work with some of the professors in my graduate study. I firmly believe that your esteemed institution will prepare me to contribute to the field I love and finally achieve my goal.

上海交通大学 RECORDS FOR UNDERGRADUATE

School of Electronic Information and Electrical Engineering **STUID：515030910451** **CLASS：F1503018**

MAJOR：Computer Science and Technology NAME：Sun Ping

ACADEMIC YEAR：2015-2016 SEMESTER ONE

CODE	COURSES	CREDIT	SCORE
CS902	Thinking and Approach of Programming	3	80
EI901	Science and Technology Innovation (Part 1)	2	87
EN025	University English I	3	74
MA077	Linear Algebra	3	83
MA115	Discrete Mathematics	2	86
MA118	Mathematical Analysis I	6	80
PE001	Physical Education I	1	82
SP209	3S Technology:Remote Sensing, Global Positioning Systems and Geography Information Systems	1	80
TH000	Cultivation of Ethics and Fundamentals of Law	3	89
TH004	Military Theory	1	86
TH009	Circumstance and Policy	1	A-

ACADEMIC YEAR：2015-2016 SEMESTER TWO

CODE	COURSES	CREDIT	SCORE
AR901	Architecture Appreciation	2	85
CS048	C++ Programming	3	85
EN026	University English II	3	71
HI054	The Tokyo Trial	1	71
MA043	Mathematical Analysis II	6	69
MA119	Probability and Statistics	3	66
PE002	Physical Education II	1	88
PH001	Physics I	4	91
PH028	Physics Lab. I	1	85
PU109	International Relations in UK/US TV Series	2	84
PU936	Nationalism and Ethnic Politics	3	86
SO936	Career Development and Planning	2	93
TH009	Circumstance and Policy	1	A
TH010	Military Training	3	P
TH021	Modern Chinese History	2	78
XP000	General Education Practice	2	P

ACADEMIC YEAR：2016-2017 SEMESTER ONE

CODE	COURSES	CREDIT	SCORE
CS144	Principles and Practice of Computer Algorithms	3	83
CS358	Data Structure	4	80
EI228	Science and Technology Innovation (Part 2－B)	2	95
EI235	Synthesis of Circuits Systems	3	97
EI236	Synthesis of Circuits Systems Lab.	2	92
EN027	University English III	3	76
MA097	Mathematical Methods in Physics	3	72
PE003	Physical Education III	1	85
PH002	Physics II	4	87
PH029	Physics Lab. II	1	82
TH007	Basic Theory of Marxism	3	82
TH009	Circumstance and Policy	1	B

ACADEMIC YEAR：2016-2017 SEMESTER TWO

CODE	COURSES	CREDIT	SCORE
CS145	Experiments in Computer Organization	2	80
CS214	Algorithm and Complexity	3	75
CS307	Operating Systems	3	77
CS356	Project Workshop of Operating System	2	65
CS359	Computer System Architecture	3	80
CS499	Mathematical Foundations of Computer Science	3	71
EE204	Embedded System and Microcomputer Principles	4	86
EN028	University English IV	3	80
ME122	Manufacturing Practice B	2	A-
PE004	Physical Education IV	1	82
TH009	Circumstance and Policy	1	A
TH012	Introduction to Mao Zedong's Thoughts and Theoretical System of Socialism with Chinese Characteristics	6	78

ACADEMIC YEAR：2017-2018 SEMESTER ONE

CODE	COURSES	CREDIT	SCORE
CS337	Computer Graphics	3	91
CS339	Computer Network	3	89
CS410	Artificial Intelligence	3	77
CS438	Internet Information Extraction	3	80
EI312	Science and Technology Innovation (Part 3-C)	2	95
IN901	Information Literacy and Practice	2	91
LI901	Literature And Life	2	88

ACADEMIC YEAR：2017-2018 SEMESTER TWO

CODE	COURSES	CREDIT	SCORE
BI022	Life Science Development History	2	97
CS238	Virtual Reality and Augmented Display	3	94
CS239	Data Visualization and Visual Analytics	3	92
CS245	Introduction to Data Science	3	86
DR002	Seal Cutting	2	93
EI327	Science and Technology Innovation (Part 4-I)	2	96
MU902	Music Theory	2	90

NOTE1-MARK"△"Means the Course Failed NOTE2-The sheet should be stamped to be official

Registrar: Registrar's Office,Shanghai Jiao Tong University http://jwc.sjtu.edu.cn 2018/9/21

上海交通大学教务处 学生成绩证明专用章

说　　明

学年

每学年开始于九月，结束于次年八月。2011年（含）起，每学年包括两个长学期和一个暑期学期，长学期有16周的规定课程，短学期有4周的规定课程；2011年之前，每学年包括两个学期，每学期有18周规定课程。

考核与记分方式

考核根据课程类别分为考试和考查两类，其中考试课程的记分方式为百分制或等级制，考查课程的记分方式为合格/不合格（Pass/Failure）两级制。详细注释如下：

1、从2005届毕业生起，我校成绩记录不再使用五级记分制（优秀、良好、中、及格、不及格），已计入的成绩参照附表进行折算；2004届（含2004届）以前学生成绩仍按原记分方式执行，同时由学校出具的中英文成绩证明中成绩折算方法也不做调整，具体参照附表；

2、考查课程不计入平均积点，但计入总学分，考查课程总学分达不到培养计划要求不能毕业；

3、自03届毕业生起部分课程为双语或英语授课（双语或英语授课课程不另标注），03届以前所有课程除英语、日语等语言类课程外均采用中文授课；02届以前（含02届）毕业生如果英语从二级开始修读，对应英文成绩单英语提高一级；

4、学时、学分与GPA：2011年（含）起，16学时=1学分；2011年之前，18学时=1学分；GPA=Σ（学分×积点）/ Σ学分，教务处不受理GPA公证。

Explanatory Notes

Academic Year

The academic year of the university begins in September and ends in August of the following year.From the year of 2011, it includes two long semesters and one summer semester, each long semester has sixteen weeks of scheduled classes,summer semester has four weeks of scheduled classes;Before the year of 2011, it includes two semesters,each semester has eighteen weeks of scheduled classes.

Score-Transformation Rules For Undergraduate Courses of SJTU

1. For students graduated in 2005 or after, some courses are graded by the " Pass/Failure " grading system,and others are graded by the hundred-mark system. The Chinese five-level score system(优秀excellent.良好good.中fair.及格pass.不及格 failure)is no longer in force.The transformation rules are illustrated in the attached chart. For students graduated in 2004 or before, the transformation rules are unchanged.

2. The grade point average does not include the courses graded by the " Pass/Failure ", but the credits of these courses are added to the total credits. For graduation, students need to accumulate the required credits,as specified for each program.

3. For students graduated in 2003 or after, some courses are taught bilingually or in English; for students graduated in 2002 or before, all the courses were taught in Chinese language only, except for language courses such as English courses, Japanese courses, and so on; For students graduated in 2002 or before, the score of English is improved one level if he or she studies English from Band 2.

4. From the year of 2011, one credit is designated for one lecture hour per week for 16 weeks; Before the year of 2011, one credit is designated for one lecture hour per week for 18 weeks; GPA =Σ(course credit×point)/Σcourse credit, Academic Affairs Division does not verify students′ GPA.

附表/Attached Chart

新记分制(2005届起)New Grading System(For Students Graduated in year 2005 or After)						旧记分制 Old Grading System(For Students Graduated in year 2004 or Before)				
考查考核			考试考核			五（四）级考核			百分考核	
中文计分	对应英文 百分制	对应英文 等级制	百分制计分	对应英文等级	积点	中文记分	对应英文 百分制	对应英文 等级制	中文计分	对应英文等级制
合格	Pass	Pass	[95,100]	A+	4.3	优+、优、优-	Excellent	A	[85,100]	A
			[90,95)	A	4.0					
			[85,90)	A-	3.7					
			[80,85)	B+	3.3	良+、良、良-	Good	B	[75,85)	B
			[75,80)	B	3.0					
			[70,75)	B-	2.7	中+、中、及格 C+、C、C-			[70,75)	C
			[67,70)	C+	2.3		Fair	C	[60,70)	
			[65,67)	C	2.0					
			[62,65)	C-	1.7					
			[60,62)	D	1.0					
不合格	Failure	Failure	<60	F	0	不及格	Failure	D	<60	D

缓考（DF）:Deferred Final Examination

Ping Sun

spofnewage@sjtu.edu.cn
(+86) 158 2193 6985

EDUCATION

Shanghai Jiao Tong University (SJTU) **Shanghai, China**

Bachelor of Engineering, Computer Science and Engineering *Sept. 2015 - Jun. 2019*

GPA: **3.83/4.00** in the 3rd year
National Encouragement Scholarship 2016 (20%)

Core Coursework: Data Structure, C++ Programming, Algorithms, Computer Network, Operating Systems, Artificial Intelligence

PROJECTS

Automatic Tracking and Service System for Parking Lots

Course Project *Oct. 2017 – Dec. 2017*

- Provided convenient services to car owners so that they could view positions and conditions of cars through mobile phones
- Used motion detection algorithms to dynamically capture the tracks of multiple cars in video monitor
- Recognized vehicle license plates based on pattern recognition and optical character recognition
- Compared with the traditional way of using electronic tags and signals, I provided precise positioning in complex indoor environment under different illumination conditions using image processing technology
- Used short-term keys to ensure privacy of users and security of the server system

Realtime Multi-Person Pose Estimation

Research Project *Oct. 2017 – Dec. 2017*

- Detected multi-person body posture using a bottom-up 2-stage CNN approach
- Used VGG-19 to get feature maps of body parts and link parts using the Part Affinity Field algorithm in both Keras and Pytorch
- Optimized the algorithm of limb connection to reduce processing time
- Achieved real-time test and worked on video format

Feature Recombination Based on Spatially-aware Fashion Concept

Research Project *Apr. 2018 – Jun. 2018*

- Generated images and changed only the clothes of person in original images according to the input text description
- Added a CGAN to the work of Fashion GAN to get results with high resolution
- Generated segmentation map and added one new label of hair style to the existing labels

INTERNSHIP

Versa AI Lab Intern **Shanghai, China**

Algorithm Group, Computer Vision *Jul. 2018 – Dec. 2018*

- Made great improvement over the existing work of segmentation
- Combined the depth information of objects with RGB image to provide affinity information to original segmentation and deploy this new function on iPhone X
- Designed networks and added two loss functions to combine the work of Mask-RCNN and Deeplabv3
- Increased the accuracy (IoU) of instance segmentation to 89%, which outperformed results of Mask-RCNN on MSCOCO on three labels of human, cat and dog
- Worked on the project of Generative Image Inpainting to fill blanks in images with background information
- Improved the temporal consistency of video semantic segmentation using optical flow

EXPERIENCE

- **China Vis Data Challenge 2018**

Analysis and Visualization of potential risks in an Internet company *May. 2018*

1) Used keywords extraction to find the organization of the subject company from email content
2) Analyzed data of employee behaviors with SVM, logistic regression and other machine learning methods
3) Built Spring framework to process data from back end to front end
4) Built a system to detect potential risks in the company and visualize the whole system using JavaScript and D3

- **Outstanding members of Campus Youth League**
- **Research in the lab of Prof. Liqing Zhang in the senior year**

COMPETENCE AND SKILLS

- Coding Languages: C/C++, Python, SQL, JavaScript.
- Technologies/Environment: Linux, OpenCV, TensorFlow, Pytorch
- TOEFL: 106 (speaking 23), GRE 154+170 AW4.0

PUBLICATIONS

[1] Attribute Boosted Person Re-Identification via Decision Making and Attention (CVPR 2019 submitted)

[2] High Accuracy Instance Segmentation with Edge Loss (CVPR 2019 submitted)

Attribute Boosted Person Re-Identification via Decision Making and Attention

Anonymous CVPR submission

Paper ID ****

Abstract

Person re-identification aims to match pedestrian images across various scenes in video surveillance. The existing works usually compute similarities between two images without semantic explanation. In contrast, in this paper, we propose an interpretable approach to identify persons by learning "what" and "where". Specifically, we integrate attribute information into a decision making process, in which we predict whether two images belong to the same person by selecting and verifying their key attributes sequentially and adaptively. In this process, we learn "what", i.e., which are the key attributes to verify two images, by using our proposed attribute selection module, and learn "where", i.e., which local information is related to the key attributes, by using attention modules. Experimental results demonstrate that our proposed method achieves the state-of-the-art results on three benchmark datasets. We also provide in-depth qualitative analyses to interpret our decision making process.

1. Introduction

Given a query image, the goal of person re-identification (re-id) is to identify the images belonging to the same person from the gallery images captured by non-overlapping surveillance cameras. The previous re-id works either learn discriminative features which are invariant across different camera views and illumination conditions [37, 12, 44], or learn the distance metric which can preserve the ranking order of training samples [2, 18, 38]. With the rapid development of deep learning [11, 34, 7], deep learning techniques have been used for the re-id task and achieved excellent performance [15, 19, 5, 32, 17, 20, 42, 16, 29, 1], which can learn discriminative features and good metric at the same time in an end-to-end system. Despite the competitive performance of deep learning based methods for the re-id task, the majority of them can hardly provide semantic explanation for *"what"*: which key attributes indicate the essential difference between two pedestrian images, and *"where"*: which local information forms the primary evidence of verifying two pedestrian images.

In this paper, we propose an interpretable person re-id method by formulating the re-id task as attribute-based decision making process. In particular, given two pedestrian images, at each time step, we select one attribute (*e.g.*, gender) and judge whether these two images have the same attribute value (*e.g.*, male) *w.r.t.* the selected attribute, which is illustrated in Figure 1. Intuitively, if two pedestrian images belong to the same person (*i.e.*, ID), they are supposed to have the same attribute value *w.r.t.* most attributes. Otherwise, if two pedestrian images belong to different persons (*i.e.*, IDs), they are supposed to have different attribute values *w.r.t.* at least one attribute. In our problem, we use ID-level attribute, which means that all the images belonging to the same ID should have the same attribute value. In contrast, image-level attribute means that different images from the same person could have different attribute values, which is ill-suited for our method. Moreover, image-level attribute annotations are not easy to acquire due to the high annotation cost. Note that we only use the annotated attribute labels of training IDs as auxiliary information in the training stage, but do not require attribute labels for test IDs in the testing stage, considering that test IDs usually do not have attribute annotation in the real-world applications.

One problem of using ID-level attribute is the attribute annotation noise. On one hand, the image quality issues such as low resolution, motion blur, and cluttered background often induce annotation errors of human annotators. On the other hand, for different images from the same ID, some attributes may vary along with time (*e.g.*, vanishing attributes due to the occlusion caused by camera view change), resulting in the inconsistency of ID-level attribute on the image level. The abovementioned attribute annotation noise could hinder the performance of re-id methods based on ID-level attribute.

Fortunately, in our re-id framework, the annotation noise of ID-level attribute can be mitigated to some extent by dropping the noisy attributes that may harm the verification. In particular, we propose a novel Attribute Selection Module (ASM). In our ASM, we use reinforcement learning based on our designed reward mechanism to select the key attributes, which are more robust to low image quality and

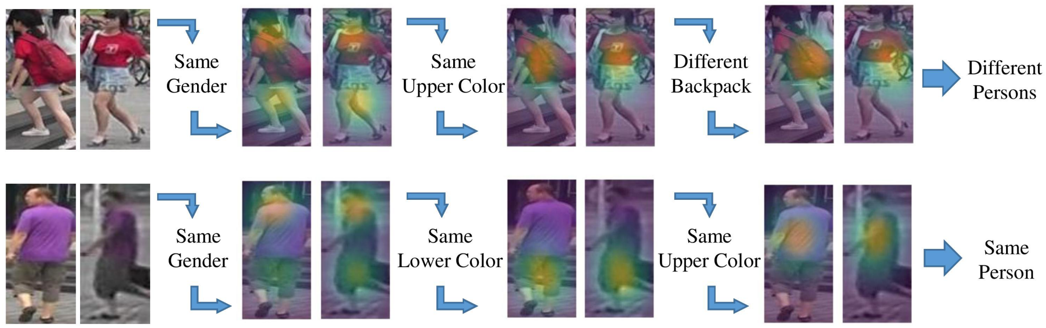

Figure 1: An illustration of learning "what" (*i.e.*, the key attributes) and 'where" (*i.e.*, attribute-related local information) in the person re-identification task, in which our method can automatically select the key attributes for verification in a sequential manner (best viewed in color).

more consistent across different images from the same ID. In this way, we learn "what": which are the key attributes to judge whether two images belong to the same person. Moreover, we employ spatial and channel attention modules to learn "where": how much attention should be paid to certain regions or channels for each key attribute. The main contributions of our work are listed as follows:

- We are the first to formulate the person re-id task as an attribute-based decision making process, to make this task more interpretable. This procedure will help automatically choose some key attributes for verification and determine the number of attributes for comparison.
- Technically, we propose an Attribute Selection Module to learn "what", *i.e.*, the key attributes for verification. Besides, we further utilize attribute-specific attention modules to learn "where", *i.e.*, attribute-related regions or channels.
- We show that our method achieves the state-of-the-art performance on three popular re-id datasets. Compared with the baselines using attribute information, our method achieves as high as 8% improvement based on rank-1 accuracy, which demonstrates the advantage of method to leverage attribute information.

2. Related Works

In this section, we conduct a comprehensive review of the existing methods for the person re-id task.

Re-Id Based on Deep Learning: There are mainly two research lines for the re-id task: representation learning and metric learning: 1) representation learning tends to learn a discriminative ID-level representation [39, 31]; 2) metric learning optimizes the distance metric using verification loss [40] or triplet loss [27]. Currently, the research trend is combining representation learning and metric learning [36, 4, 5, 17, 42, 29] in one system.

Re-Id Based on Attribute: The attribute can be categorized into ID-level attribute and image-level attribute, as discussed in Section 1. For image-level attribute, some previous works use the attribute information to select tuples for metric learning [30, 10] or mine the rules for verifying persons [24]. There are more works using ID-level attribute. Specifically, Lin *et al.* [21] proposed to combine person identification and ID-level attribute recognition while Schumann *et al.* [28] proposed to use the attribute information as auxiliary information to train a network with triplet loss. These methods treat different attributes equally without selecting the key attributes, and thus suffer from attribute annotation noise. In contrast, our proposed method can select key attributes in a sequential manner, which can suppress the adverse effect of noisy attributes.

Re-Id Based on Reinforcement Learning: Reinforcement Learning (RL) trains an agent to exploit the experience during exploration and simultaneously maximize the cumulative reward. Recently, RL has been successfully applied to many computer vision tasks [3, 23, 9, 13, 41]. However, there are relatively few works using RL for person Re-ID. For instance, In [13], Lan *et al.* proposed a method to refine the bounding box of persons to improve re-id task performance by RL. In another work [41], RL is applied to learn the adequate number of image pairs to verify in the multi-shot re-id task. Distinguished from the above two methods, our method formulates person re-id as an attribute-based decision making process under the reinforcement learning framework, which has not been studied before.

Re-Id Based on Attention Models: The attention mechanism [25] can guide the network towards the discriminative local information. Some previous attention based approaches for the re-id task employ hard attention [13, 14, 43, 29] or soft attentions [17, 22] to locate discriminative regions or channels. Nevertheless, the regions or channels located by their methods are not associated with semantic information, and thus not interpretable in the semantic level. Instead, with the aid of attribute information, we select key attributes se-

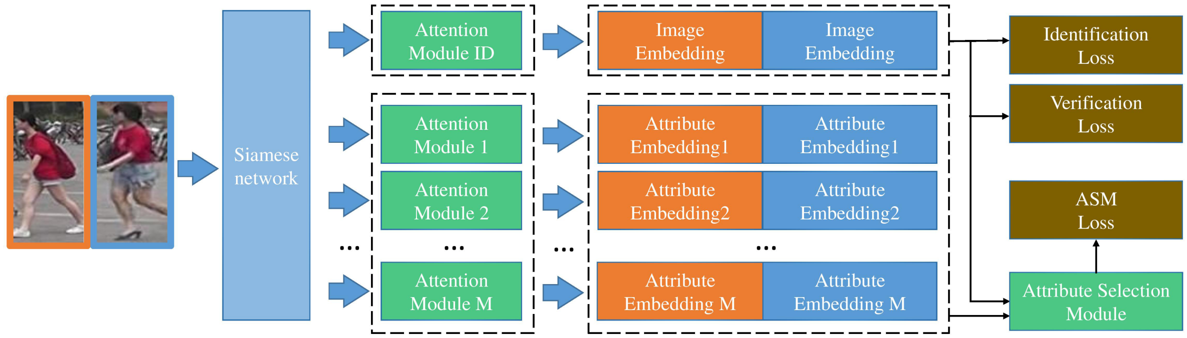

Figure 2: The pipeline of our proposed method (best viewed in color). Given a pair of pedestrian images, we use Siamese network to generate the feature maps for two images. Then, we use attention module for the entire image and M attributes to generate $M+1$ attention heatmaps, and apply the attention maps to the feature maps of two images, yielding an image embedding and M attribute embeddings for each image. Next, image embeddings are used to calculate the identification loss and verification loss, while both image embeddings and attribute embeddings are used to train our Attribute Selection Module (ASM) based on the ASM loss.

quentially and identify the attribute-related local information, leading to an interpretable decision making process.

3. Our Proposed Method

In the person re-id task, given a dataset composed of K images (each image contains a detected bounding box enclosing a pedestrian) from N different persons captured by different surveillance cameras, we split the dataset into training set and test set according to person ID, so that training IDs and test IDs have no overlap. Each training ID is associated with a suite of ID-level attribute labels. In the testing stage, the test set is divided into a set of query images and a set of gallery images. Then, given a query image, the goal is to rank all the gallery images in light of the probability that it belongs to the same person as the query image. Formally, we assume the dataset contains an image set $X = \{x_1, x_2, \ldots, x_K\}$ with ID labels $Y = \{y_1, y_2, \ldots, y_K\}$ where $y_i \in \{1, 2, \ldots, N\}$. Each training image x_i from training IDs has the ID-level attribute labels $z_i = (z_i^1, \ldots, z_i^M) \in \mathcal{N}^M$, in which M is the number of attributes.

3.1. System Configuration

The pipeline of our proposed method is depicted in Figure 2. In particular, given a pair of images (x_a, x_b), we adopt a Siamese network structure [6] (the backbone is ResNet50 [7] pretrained on ImageNet [11]), which passes a pair of images through two parallel CNN models with shared model parameters. Then, we learn an individual attention module (see Section 3.3 for the details) for the entire human body and each attribute, leading to in total $M+1$ attention modules. For each image, we apply its $M+1$ attention heatmaps produced by attention modules to its output of Siamese network, yielding M *attribute embeddings* $(e^1, e^2, \ldots, e^M)$ for all M attributes and one *image embedding* e^{id} for the entire human body. Based on the image embedding, we classify each image into different IDs and verify whether a pair of images belong to the same ID based on the identification loss and verification loss respectively. Besides, the image embedding together with M attribute embeddings is fed into our proposed Attribute Selection Module (ASM) to select key attributes sequentially, which is trained based on the ASM loss.

To this end, we are to optimize an objective function with three types of losses:

- L_{ide} (Identification Loss): We aim to classify input images into different IDs based on their image embeddings. The classifier is implemented as a fully-connected (FC) layer with the output size as N, which is trained using softmax classification loss (*i.e.*, identification loss) with ground-truth ID labels.

- L_{ver} (Verification Loss): We tend to verify whether a pair of images belong to the same ID, which is transformed to a binary classification problem. Technically, given a pair of images, the absolute difference of their image embeddings is calculated, followed by a FC layer with the output size as 2, which is trained using binary classification loss (*i.e.*, verification loss) with ground-truth pairwise ID labels.

- L_{asm} (ASM Loss): We use our proposed Attribute Selection Module (ASM) to select key attributes sequentially and adaptively. Our ASM is trained based on the ASM loss, which will be detailed in Section 3.2.

To this end, the aggregated loss function can be written as

$$L = \lambda_i(L_{ide} + L_{ver}) + \lambda_a L_{asm}, \tag{1}$$

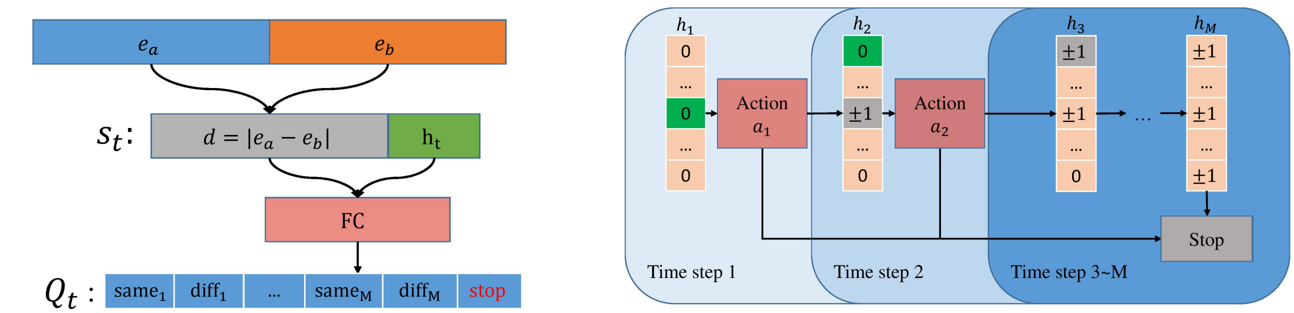

Figure 3: An illustration of the proposed Attention Selection Module. The left subfigure shows the structure of Q-network with state $s_t = (d, h_t)$ and Q-values for each type of action. The right subfigure represents the process of updating h_t after taking a sequence of actions in an episode.

in which λ_a and λ_i are trade-off parameters. L_{ide} and L_{ver} are commonly used for the re-id task [31, 40] while L_{asm} is the key component contributing to an interpretable decision making process, so we will fully describe our Attribute Selection Module (ASM) in the following section.

3.2. Attribute Selection Module (ASM)

We use Attribute Selection Module (ASM) to interpret “what": which key attributes indicate the essential difference between two pedestrian images. The goal of ASM is to verify two pedestrian images through verifying an adequate number of selected attributes sequentially. Note that one major contribution of this paper is the automatic attribute selection and adaptive decision making for person verification. In the remainder of this paper, we refer to the whole process of verifying two images as an episode. At each time step in an episode, we choose an attribute and judge whether the two images have the same attribute value *w.r.t.* the chosen attribute. In this way, we cast the person re-id task as a decision making process under the reinforcement learning framework, in which an agent needs to interact with the dynamic environment by exploring the unknown and exploiting the experience. The policy for an agent to take the optimal *actions* based on different *states* is learnt by maximizing the cumulative *rewards* in each episode. In our problem, *state*, *action*, and *reward* are defined as follows.

State: The state describes the observation of the agent in the environment. By using s_t to represent the state at the time step t, the state $s_t = (d, h_t)$ consists of d and h_t, as shown in the left subfigure in Figure 3. To be exact, d is the embedding difference between two input images. Recall that after an image pair (x_a, x_b) passes through the Siamese network and $M+1$ attention modules, M attribute embeddings $(e^1, e^2, \ldots, e^M)$ and one image embedding e^{id} will be generated for each image (see Section 3.3 for the details of attention modules). We flatten and concatenate these $M+1$ embeddings for each image as $e_a = (e_a^{id}, e_a^1, \ldots, e_a^M)$ and $e_b = (e_b^{id}, e_b^1, \ldots, e_b^M)$, and then compute their absolute difference $d = |e_a - e_b|$. $h_t \in \mathcal{Z}^M$ is used to record the action history of the current episode at the time step t. When $t = 1$, $h_t = \mathbf{0}$ with all entries being 0. When $t > 1$, the i-th entry will be updated accordingly if we select the i-th attribute for verification, which will be detailed next.

Action: In our task, the agent needs to know which key attribute to choose and whether the two input images have the same attribute value *w.r.t.* the chosen attribute at each time step in an episode. The agent also needs to know when to terminate the current episode. Therefore, the agent has in total $2M+1$ types of actions including predicting “same" or “different" for one of the M attributes and the “stop" action. Formally, we use a_t to denote the action at the time step t in an episode. Given two images x_a and x_b with their ID labels y_a and y_b, at the time step t, the agent selects the i-th attribute and predict whether x_a and x_b have the “same” attribute value (*i.e.*, $a_t = (i, 1)$) or “different” attribute values (*i.e.*, $a_t = (i, -1)$) *w.r.t.* the i-th attribute, or chooses “stop” (*i.e.*, $a_t = (0, 0)$) to terminate the current episode and drop the remaining attributes. Note that the agent is not allowed to choose the attribute which has been selected before in the same episode, in order to avoid the duplicated selection of key attributes. In an extreme case, at the time step $M+1$, which indicates that the agent has selected all M attributes, the agent is enforced to choose “stop". When $t > 1$ and $a_t \neq (0, 0)$, h_t in state s_t will be updated as follows: $h_{t+1}^i = 1$ (*resp.*, $h_{t+1}^i = -1$) if $a_t = (i, 1)$ (*resp.*, $a_t = (i, -1)$). The process of updating h_t in an episode is illustrated in the right subfigure in Figure 3.

Reward: The reward is defined as the benefit or penalty that the agents will receive after taking specific action in the current state. In an episode of verifying two images, we expect them to have the same attribute value *w.r.t.* most attributes when they belong to the same ID, while to have different attribute values *w.r.t.* at least one attribute when they belong to different IDs. When designing the reward function, on one hand, we assume that the rewards of the sequence of attribute verification should follow the above rules until the end of an episode; on the other hand, we hope that the attribute verification results through an episode are

CVPR #****

Algorithm 1 Training based on the ASM loss.

Input: Training images with their ID labels and the attributes of training IDs.
Output: Model parameters θ

Initialize θ_0.
for $i \leftarrow 1, 2, \ldots, I$ **do**
  Sample a batch of image pairs $\mathcal{P}$ and initialize the transition memory $\mathcal{D} = \emptyset$.
  for (x_a, x_b) in $\mathcal{P}$ **do**
    Generate transitions $\{(s_t, a_t, r_t, s_{t+1})\}$ within one episode.
    $\mathcal{D} = \mathcal{D} \cup \{(s_t, a_t, r_t, s_{t+1})\}$.
  end for
  for (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$ **do**
    if $a_t = (0, 0)$ **then** $Q^*(s_t, a_t) \leftarrow r_t$.
    else $Q^*(s_t, a_t) \leftarrow r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta_{i-1})$.
    end if
  end for
  Update θ_i with gradient $\nabla_{\theta_i} \mathbb{E}_{s_t, a_t}[(Q(s_t, a_t; \theta_i) - Q^*(s_t, a_t))^2]$.
end for
return θ_I

correct, that is, $a_t = (i, 1)$ when $z_a^i = z_b^i$ and $a_t = (i, -1)$ when $z_a^i \neq z_b^i$, in which z_a^i (*resp.*, z_b^i) is the ground-truth attribute label *w.r.t* the i-th attribute for x_a (*resp.*, x_b).

Then, our reward function r_t after taking the action a_t can be described as follows:

1. When $a_t = (i, \pm 1)$ for $i = 1, \ldots, M$:
   (a) $r_t = +1$: If $a_t = (i, 1)$ (*resp.*, $a_t = (i, -1)$) and $z_a^i = z_b^i$ (*resp.*, $z_a^i \neq z_b^i$);
   (b) $r_t = -1$: If $a_t = (i, 1)$ (*resp.*, $a_t = (i, -1)$) and $z_a^i \neq z_b^i$ (*resp.*, $z_a^i = z_b^i$).
2. When $a_t = (0, 0)$:
   (a) $r_t = +1$: If $y_a = y_b$ and $\forall j \;\, h_t^j \neq -1$;
   (b) $r_t = -1$: If $y_a = y_b$ and $\exists j \;\, h_t^j = -1$;
   (c) $r_t = +1$: If $y_a \neq y_b$ and $\exists j \;\, h_t^j = -1$;
   (d) $r_t = -1$: If $y_a \neq y_b$ and $\forall j \;\, h_t^j \neq -1$.

Deep Q-Learning: After introducing the state s_t, action a_t, and reward r_t, we use deep Q-learning [26] to train the agent to seek for the optimal policy. We use $Q(s_t, a_t)$ to represent the Q value at the time step t, which estimates the discounted cumulative reward from the current time step to the end of the episode. $Q(s_t, a_t)$ can be calculated based on the recursive formulation $Q(s_t, a_t) = r_t + \gamma \max_{a'} Q(s_{t+1}, a')$. We use a simple network with one FC layer to estimate the Q-value $Q(s_t, a_t)$ given s_t and a_t, which is referred to as Q-network. Instead of feeding s_t and each type of action a_t into Q-learning network, we use a single forward pass to compute Q-values for all actions, as shown in the left subfigure in Figure 3. Since there are totally $2M + 1$ types of actions, the output size of our Q-network is $2M + 1$.

By using θ to denote the model parameters of Q-network and the modules preceding Q-network, we adopt ϵ-greedy learning [33] to update θ in the training stage. In each training iteration, we randomly sample a batch of image pairs and create a sequence of transitions for each image pair. Specifically, in the i-th training iteration, with the Q-network parameters θ_{i-1} and the initial state s_1, we select actions according to ϵ-greedy policy [33] and obtain the corresponding reward until one episode terminates, yielding a sequence of transitions $\{(s_t, a_t, r_t, s_{t+1})\}$. For each transition (s_t, a_t, r_t, s_{t+1}), we can estimate the optimal Q-value $Q^*(s_t, a_t)$: $Q^*(s_t, a_t) = r_t$ if $a_t = (0, 0)$ and otherwise $Q^*(s_t, a_t) = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta_{i-1})$. We update Q-network parameters θ_i to enforce $Q(s_t, a_t; \theta_i)$ to be close to the optimal Q-value $Q^*(s_t, a_t)$, leading to our ASM loss:

$$L_{asm} = \mathbb{E}_{s_t, a_t}[(Q(s_t, a_t; \theta_i) - Q^*(s_t, a_t))^2]. \qquad (2)$$

The details of training ASM are summarized in Algorithm 1, in which γ is set as 0.9, ϵ shrinks from 1 to 0.1 linearly in the first 20 epochs, and the learning rate is set as 0.0001.

In the testing stage, how to use ASM for evaluation is non-trivial. Since ASM selects different sets of attributes for different input image pairs, the generated sequence of Q-values corresponding to different sets of attributes cannot be directly used to measure the similarity between two input images in a consistent manner. Thus, instead of using ASM, we adopt the traditional evaluation strategy only using image embeddings. Following most previous methods [15, 40, 21, 28, 46], given an image pair (x_a, x_b), we obtain their normalized image embeddings e_a^{id} and e_b^{id}, and calculate their Euclidean distance (equivalent to cosine distance in the normalized case), which measures the probability that x_a and x_b are from the same person. Although we do not use ASM for evaluation, ASM can help learn better image embeddings by using auxiliary attribute information in the training stage (see Section 4.4) and also provide semantic explanation when verifying test image pairs in the testing stage (see Section 4.5).

Stabilizing ASM training: In practice, we observe that joint training of ASM and attention modules is unstable. Therefore, we pretrain the attribute related attention modules to stabilize subsequent ASM training. In the pretraining stage, we replace ASM with Attribute Classification and Verification (ACV) module for each attribute. Specifically, for attribute classification, similar to L_{ide} in (1), we perform multi-class classification for each attribute. For attribute verification, similar to L_{ver} in (1), we perform binary classification based on the absolute difference between two at-

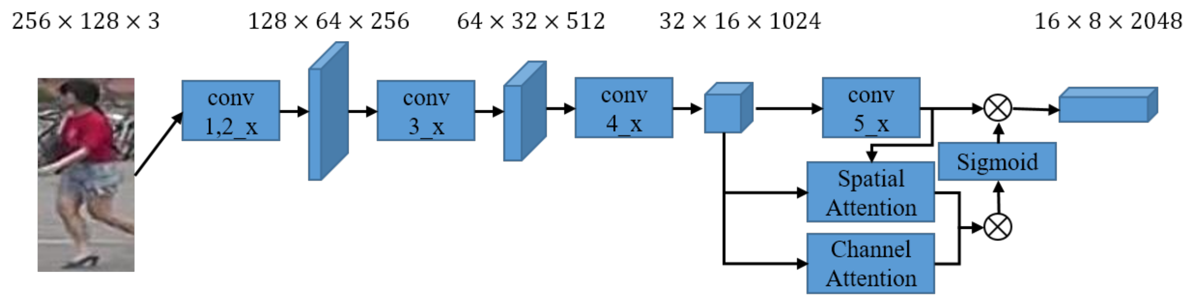

Figure 4: The illustration of our attention module composed of spatial attention and channel attention. The conv 1~5_x represent five stages in ResNet101 [7] and the details of each stage can be found in [7]. We apply our attention modules between the conv x_4 and the conv x_5 stage.

tribute embeddings. We refer to the collected losses of ACV module as L_{acv}. The pretraining process based on L_{acv} can provide better initialization and reduce the total training time significantly.

3.3. Attention Module

In the person re-id task, to interpret "where", *i.e.*, which local information forms the main evidence of verifying two pedestrian images, we learn an attention module to focus on the entire human body and M attention modules to focus on attribute-related local information for M attributes, as shown in Figure 2. Each attention module can be further decomposed into spatial attention and channel attention, as illustrated in Figure 4. Since we use ResNet50 with conv 1~5_x stages as the backbone Siamese network, our attention modules are applied between the conv 4_x stage and the conv 5_x stage.

Spatial Attention: For spatial attention, we learn different weights for each location on the feature map, paying attention to specific regions. We adopt the attention model structure in [35] and the details are omitted here.

Channel Attention: For channel attention, we learn different weights for different channels of the feature map, paying attention to specific channels. We follow the attention model structure in [8] and the details are omitted here.

Combining Spatial Attention and Channel Attention: We combine spatial attention and channel attention in the following way. We resize the output of both spatial attention and channel attention to the same size as the output of conv 5_x stage, and element-wise multiple two outputs, followed by sigmoid activation layer for the normalized scale in $[0, 1]$. Then, we element-wise multiply the combined attention map and the output of conv 5_x stage, producing image embedding or attribute embedding. The idea of combining spatial attention and channel attention is not new [17]. However, unlike [17], we replace the spatial attention module in [17] with that in [35] to improve the performance. Moreover, our work is the first to apply attribute-related attention model for the re-id task.

4. Experiments

In this section, we compare our method with our special cases and state-of-the-art baselines on three benchmark datasets together with in-depth qualitative analyses.

4.1. Datasets

We evaluate our method and baselines on three datasets: Market-1501 [45], DukeMCMT-ReID [21], and CUHK03 [15].

Market-1501: Market-1501[45] dataset contains 751 training IDs and 750 test IDs, with annotated ID-level attribute labels for 30 attributes provided in [21]. We merge 9 attributes for colors of lower-body clothing, 8 attributes for colors of upper-body, and 4 attributes for ages, leading to in total 12 attributes.

DukeMCMT-ReID: The DukeMTMC-reID[47] dataset consists of an equal number of 702 IDs for both training and testing, with annotated ID-level attribute labels for 23 attributes provided in [21]. Similar to the Market-1501 dataset, we merge 8 attributes for colors of lower-body clothing and 7 attributes for colors of upper-body, resulting in totally 10 attributes.

CUHK03: Following [48], we split all IDs into 767 training IDs and 707 test IDs. Note that CUHK03[15] dataset does not have attribute annotations, so we use pseudo attribute labels instead. Specifically, we use the ACV module pretrained on the Market-1501 dataset to generate attribute probabilities for each image. Then, we sum up the attribute probabilities of all images from each ID, and select the attribute labels corresponding to the highest probabilities as the pseudo attribute labels for that ID. Finally, we rely on pseudo ID-level attribute labels to train our model.

4.2. Training Strategies

Recall that we have three losses in (1) with λ_i weighing the ID-related losses and λ_a weighing the attribute-related loss. In terms of different training strategies, besides jointly optimizing all losses, we can focus on ID-related losses or attribute-related loss first. Since we use auxiliary attribute information to help learn better image embeddings, we are prone to focus on attribute-related loss first instead of ID-related losses. Thus, we have the following two strategies:

- Joint Learning (JL): jointly optimizing all losses with fixed λ_i and λ_a.
- Step-by-Step (SS): in the early stage, only optimize attribute-related loss by setting $\lambda_i = 0$ and $\lambda_a = 1$; in the late stage, only optimize ID-related losses by setting $\lambda_i = 1$ and $\lambda_a = 0$.

In fact, that is a soft version of SS, that is, we gradually transfer our focus from attribute-related loss to ID-related losses, which is referred to as Incremental learning (IL):

- Incremental learning (IL): in the early stage, increase λ_i from 0 to 1 and decrease λ_a from 1 to 0; in the late stage, keep $\lambda_i = 1$ and $\lambda_a = 0$.

We evaluate JL, SS, IL using rank-1 accuracy on the Market-1501 dataset and find that SS is the best strategy (see Supplementary for the details). After identifying the best training strategy, we incorporate the pretraining process mentioned in Section 3.2 and train our entire system in the following three phases:

Phase 1: pretraining based on L_{acv} for 30 epochs.

Phase 2: training based on L_{asm} by setting $\lambda_i = 0$ and $\lambda_a = 1$ in (1) for 40 epochs.

Phase 3: training based on L_{ide} and L_{ver} by setting $\lambda_i = 1$ and $\lambda_a = 0$ in (1) for 70 epochs.

4.3. Ablation Studies

To evaluate our proposed ASM and used attention modules, we conduct extensive ablation studies. We evaluate 10 special cases including our complete method on Market-1501 and list the rank-1 accuracy results in Table 1. These special cases can be grouped into four quadrants based on whether using ID labels (with ID or w/o ID) or whether using Attention Modules (with AM or w/o AM), as shown in Table 1. In each quadrant, we create special cases by manipulating Phase 2 or Phase 3 in the training procedure (see Section 4.2). In particular, in the quadrant "with ID with AM", we evaluate three special cases: 1) no change; 2) replace L_{asm} with L_{acv} in Phase 2; 3) skip Phase 2 without using ASM. In the quadrant "w/o ID with AM", we evaluate two special cases: 1) skip Phase 3 without using ID information; 2) replace L_{asm} with L_{acv} in Phase 2 and skip Phase 3 without using ID information. Similarly, the quadrant "with ID w/o AM" (*resp.*, "w/o ID w/o AM") is almost the same as the quadrant "with ID with AM" (*resp.*, "w/o ID with AM") except that we remove the attention modules. For the special

	with ID			w/o ID	
	ASM	ACV	w/o ASM	ASM	ACV
w/o AM	89.3	84.1	79.0	56.1	47.3
with AM	**92.1**	87.7	82.6	56.7	47.5

Table 1: Rank-1 accuracy(%) of our special cases on the Market-1501 dataset. The best result is denoted in boldface.

cases without using ID information, evaluation is performed based on the concatenation of all the attribute embeddings instead of image embeddings.

According to Table 1, when ID information is available, attention modules can achieve 3~4% improvement and ASM can outperform ACV by 4~5%, which indicates the benefit of attention modules and our proposed ASM. Equipped with both attention modules and ASM, our method obtains 13.1% improvement (92.1 *v.s.* 79.0), which shows that it is helpful to learn "what" and "where" by using ASM and attention modules. When ID information is unavailable, the performances drop sharply. However, attention modules still achieve slight improvement and ASM still outperforms ACV by a large margin, which again indicates the benefit of our proposed ASM.

4.4. Comparison with State-of-the-art Methods

We compare our method with state-of-the-art baselines on three benchmark datasets using Rank-1 accuracy (R1) and mAP as evaluation metrics. The baselines can be divided into two groups: 1) APR [21] and ACRN [28] using ID-level attribute information, which is close to our method; 2) SVDN [32], CSA [49], JLML [16], GAN [47], and HAN [17] without using attribute information. Recall that the CUHK03 dataset is not accompanied by attribute annotation, so we generate pseudo attribute labels for our method and baselines [21, 28] as described in Section 4.1. Besides, for fair comparison, we evaluate all methods with the same input size as 256×128.

The results of different methods are reported in Table 2. It can be seen that the baselines [21, 28] using attribute information are generally worse than those without using attribute information, which is likely to be caused by the attribute annotation noise (see Section 1) and their unsophisticated model design. However, with the aid of attribute information, our method outperforms all the other baselines and achieves the state-of-the-art results on all three datasets, which demonstrates the superiority of our method for the person re-id task by using ASM and attention modules to help learn better image embeddings. Especially when compared with the baselines [21, 28] using attribute information, our method achieves about 8~9% improvement on mAP and about 12~13% improvement on R1. The results indicate that our method can take full advantage of the attribute information effectively, that is, our ASM can select key attributes

Datasets	Measure	SVDN [32]	CSA [49]	JLML [16]	GAN [47]	HAN [17]	APR [21]	ACRN [28]	Ours
Market-1501	R1(%)	82.3	87.2	85.1	84.0	90.6	84.3	83.6	**92.1**
	mAP(%)	62.1	68.7	65.5	66.1	75.5	64.7	62.6	**76.0**
DukeMTMC-reID	R1(%)	76.7	76.5	73.3	67.7	78.0	70.7	72.6	**80.9**
	mAP(%)	56.8	58.4	56.4	47.1	62.3	51.9	52.0	**64.3**
CUHK03	R1(%)	41.5	45.1	42.5	43.7	50.0	40.5	41.7	**52.5**
	mAP(%)	37.3	41.1	38.0	39.1	48.7	36.8	37.4	**51.1**

Table 2: Results of our proposed method and state-of-the-art baselines on three datasets. The best results are denoted in boldface.

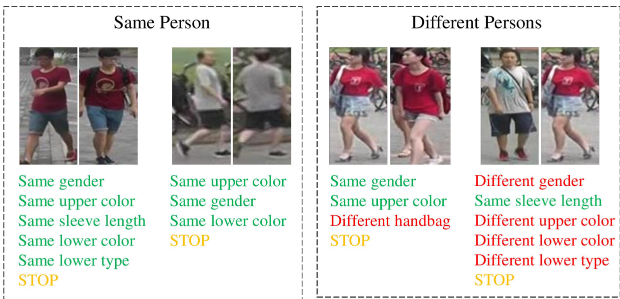

Figure 5: Four example episodes showing the decision making process of Attribute Selection Module (ASM) by selecting and verifying key attributes sequentially.

Figure 6: The attention heatmaps generated by spatial attention modules for the entire human body and three most important attributes (*i.e.*, gender, upper color, and backpack).

and drop noisy attributes while the attention modules can help focus on attribute-related local information.

4.5. Qualitative Analyses

We reformulate the person re-id problem as an attribute-based decision making process, which makes the re-id task more interpretable. Next, we first provide several example episodes of verifying image pairs, and then investigate the most important attributes.

Example Episodes: Although we do not use ASM for evaluation in the testing stage, ASM can still interpret the decision making process of test image pairs to help our understanding. In Figure 5, we show four example episodes of test image pairs generated by our ASM. In each example episode, the performed actions are listed in the execution order with different colors indicating different types of actions. We can observe that for the image pairs from the same person, two images are predicted to have the same attribute value *w.r.t.* all selected attributes. For the image pairs from different persons, two images are predicted to have different attribute values *w.r.t.* at least one selected attribute. The example episodes show the effectiveness of our ASM, which can help learn better image embeddings by exploiting attribute information, contributing to the superior performance of our method in Table 2.

Attribute Importance: We claimed that our ASM is able to select key attributes and drop noisy attributes. To prove this point, we show the most important attributes identified by our ASM. Assuming that the frequency of the agent selecting one attribute implies the significance of this attribute, we calculate the selection frequency of each attribute in the training stage on the Market-1501 dataset. According to the calculated frequency, Gender, the color of the upper clothing, and carrying backpack are three most important attributes while age, hat, and lower type are three least important attributes. We show the spatial attention heatmaps (smoothed by the Gaussian filter) generated by our attention modules for the entire human body and three most important attributes in Figure 6. Intuitively, the important attributes are less affected by low image quality because the attribute-related regions are very large, and more consistent across different images from the same person (*e.g.*, the upper body is less likely to be occluded than the lower body when the pedestrian moves). Therefore, these important attributes are more resilient to the attribute annotation noise mentioned in Section 1.

5. Conclusion

In this paper, we have proposed an interpretable approach for person re-identification (re-id) by formulating the re-id task as an attribute-based decision making process, which can utilize auxiliary attribute information effectively. Comprehensive experiments on three benchmark datasets have demonstrated the superiority of our method.

References

[1] S. Bai, X. Bai, and Q. Tian. Scalable person re-identification on supervised smoothed manifold. In *CVPR*, 2017. 1

[2] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a Mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 2005. 1

[3] J. C. Caicedo and S. Lazebnik. Active object localization with deep reinforcement learning. In *ICCV*, 2015. 2

[4] W. Chen, X. Chen, J. Zhang, and K. Huang. A multi-task deep network for person re-identification. In *AAAI*, 2017. 2

[5] Y. Chen, N. Wang, and Z. Zhang. Darkrank: Accelerating deep metric learning via cross sample similarities transfer. In *AAAI*, 2018. 1, 2

[6] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *CVPR*, 2005. 3

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1, 3, 6

[8] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *CVPR*, 2018. 6

[9] C. Huang, S. Lucey, and D. Ramanan. Learning policies for adaptive tracking with deep feature cascades. In *ICCV*, 2017. 2

[10] S. Khamis, C. Kuo, V. K. Singh, V. D. Shet, and L. S. Davis. Joint learning for attribute-consistent person re-identification. In *ECCV*, 2014. 2

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 1, 3

[12] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *Transactions on Pattern Analysis and Machine Intelligence*, 2013. 1

[13] X. Lan, H. Wang, S. Gong, and X. Zhu. Identity alignment by noisy pixel removal. In *BMVC*, 2017. 2

[14] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017. 2

[15] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 1, 5, 6

[16] W. Li, X. Zhu, and S. Gong. Person re-identification by deep joint learning of multi-loss classification. In *IJCAI*, 2017. 1, 7, 8

[17] W. Li, X. Zhu, and S. Gong. Harmonious attention network for person re-identification. In *CVPR*, 2018. 1, 2, 6, 7, 8

[18] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013. 1

[19] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2015. 1

[20] J. Lin, L. Ren, J. Lu, J. Feng, and J. Zhou. Consistent-aware deep learning for person re-identification in a camera network. In *CVPR*, 2017. 1

[21] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang. Improving person re-identification by attribute and identity learning. *CoRR*, 2017. 2, 5, 6, 7, 8

[22] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang. Hydraplus-net: Attentive deep features for pedestrian analysis. In *ICCV*, 2017. 2

[23] S. Mathe, A. Pirinen, and C. Sminchisescu. Reinforcement learning for visual object detection. In *CVPR*, 2016. 2

[24] T. Matsukawa and E. Suzuki. Person re-identification using CNN features learned from combination of attributes. In *ICPR*, 2016. 2

[25] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu. Recurrent models of visual attention. In *NIPS*, 2014. 2

[26] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. A. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 2015. 5

[27] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015. 2

[28] A. Schumann and R. Stiefelhagen. Person re-identification by deep learning attribute-complementary information. In *CVPR workshop*, 2017. 2, 5, 7, 8

[29] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose-driven deep convolutional model for person re-identification. In *ICCV*, 2017. 1, 2

[30] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian. Deep attributes driven multi-camera person re-identification. In *ECCV*, 2016. 2

[31] Y. Sun, Y. Chen, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. In *NIPS*, 2014. 2, 4

[32] Y. Sun, L. Zheng, W. Deng, and S. Wang. Svdnet for pedestrian retrieval. In *ICCV*, 2017. 1, 7, 8

[33] R. S. Sutton and A. G. Barto. Reinforcement learning: An introduction. *Transactions on Neural Networks*, 1998. 5

[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015. 1

[35] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In *CVPR*, 2017. 6

[36] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang. Joint learning of single-image and cross-image representations for person re-identification. In *CVPR*, 2016. 2

[37] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. H. Tu. Shape and appearance context modeling. In *ICCV*, 2007. 1

[38] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 2009. 1

[39] T. Xiao, H. Li, W. Ouyang, and X. Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*, 2016. 2

[40] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Deep metric learning for person re-identification. In *ICPR*, 2014. 2, 4, 5

[41] J. Zhang, N. Wang, and L. Zhang. Multi-shot pedestrian re-identification via sequential decision making. In *CVPR*, 2018. 2

[42] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, 2017. 1, 2

[43] L. Zhao, X. Li, Y. Zhuang, and J. Wang. Deeply-learned part-aligned representations for person re-identification. In *ICCV*, 2017. 2

[44] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014. 1

[45] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 6

[46] Z. Zheng, L. Zheng, and Y. Yang. A discriminatively learned cnn embedding for person reidentification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2017. 5

[47] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In *ICCV*, 2017. 6, 7, 8

[48] Z. Zhong, L. Zheng, D. Cao, and S. Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, 2017. 6

[49] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang. Camera style adaptation for person re-identification. In *CVPR*, 2018. 7, 8

View Scores

ETS Registration Number: 0000000032319970

Test	Test Date	Reading	Listening	Speaking	Writing	Total
TOEFL iBT	Sat Jan 13 08:54:13 EST 2018	29	27	22	28	106

[How to interpret scores]

Reading Skills	Level	Your Performance
Reading	**High(22-30)**	Test takers who receive a score at the **HIGH** level, as you did, typically understand academic texts in English that require a wide range of reading abilities regardless of the difficulty of the texts. Test takers who score at the **HIGH** level, typically • have a very good command of academic vocabulary and grammatical structure; • can understand and connect information, make appropriate inferences, and synthesize ideas, even when the text is conceptually dense and the language is complex; • can recognize the expository organization of a text and the role that specific information serves within the larger text, even when the text is conceptually dense; and • can abstract major ideas from a text, even when the text is conceptually dense and contains complex language.
Listening Skills	**Level**	**Your Performance**
Listening	**High(22-30)**	Test takers who receive a score at the **HIGH level**, as you did, typically understand conversations and lectures in English that present a wide range of listening demands. These demands can include difficult vocabulary (uncommon terms, or colloquial or figurative language), complex grammatical structures, abstract or complex

		ideas, and/or making sense of unexpected or seemingly contradictory information. When listening to lectures and conversations like these, test takers at the **HIGH** level typically can • understand main ideas and important details, whether they are stated or implied; • distinguish more important ideas from less important ones; • understand how information is being used (for example, to provide evidence for a claim or describe a step in a complex process); • recognize how pieces of information are connected (for example, in a cause-and-effect relationship); • understand many different ways that speakers use language for purposes other than to give information (for example, to emphasize a point, express agreement or disagreement, or convey intentions indirectly); and • synthesize information, even when it is not presented in sequence, and make correct inferences on the basis of that information.
Speaking Skills	**Level**	**Your Performance**
Speaking about familiar topics	**Fair(2.5 - 3.0)**	Your responses indicate you are able to speak in English about your personal experiences and opinions in a mostly clear and coherent manner. Your speech is mostly clear with only occasional errors. Grammar and vocabulary are somewhat limited and include some errors. At times, the limitations prevent you from elaborating fully on your ideas, but they do not seriously interfere with overall communication.
Speaking about campus situations	**Fair(2.5 - 3.0)**	Your responses demonstrate an ability to speak in English about reading material and experiences typically encountered by university students. You are able to convey relevant information about conversations, newspaper articles, and campus bulletins; however, some details are missing or inaccurate. Limitations of grammar, vocabulary, and pronunciation at times cause difficulty for the listener. However, they do not seriously interfere with overall communication.

Speaking about academic course content	**Fair(2.5 - 3.0)**	Your responses demonstrate that you are able to speak in English about academic reading and lecture material, with only minor communication problems. For the most part, your speech is clear and easy to understand. However, some problems with pronunciation and intonation may occasionally cause difficulty for the listener. Your use of grammar and vocabulary is adequate to talk about the topics, but some ideas are not fully developed or are inaccurate.
Writing Skills	**Level**	**Your Performance**
Writing based on reading and listening	**Good(4.0 - 5.0)**	You responded well to the task, relating the lecture to the reading. Weaknesses, if you have any, might have to do with • slight imprecision in your summary of some of the main points and/or • use of English that is occasionally ungrammatical or unclear.
Writing based on knowledge and experience	**Good(4.0 - 5.0)**	You responded with a well-organized and developed essay. Weaknesses, if you have any, might have to do with • use of English that is occasionally ungrammatical, unclear, or unidiomatic and/or • elaboration of ideas or connection of ideas that could have been stronger.

PING SUN

Address: ROOM424, EAST1, NO.800,DONGCHUAN, RD., MINHANG DIST., SHANGHAI, SHANGHAI, Shanghai, 200240 China

Email: 919374094@qq.com
Phone: 86-15821936985
Date of Birth: May 11, 1997
Social Security Number (Last Four Digits):
Gender: Male
Intended Graduate Major: Computer Science (0402)

Most Recent Test Date: July 5, 2018

Registration Number: 3232243
Print Date: September 12, 2018

Your Scores for the General Test Taken on July 5, 2018

Verbal Reasoning	Quantitative Reasoning	Analytical Writing
Your Scaled Score: 153	Your Scaled Score: 169	Your Score: 4.0
130 – 170	130 – 170	0 – 6
61st Percentile	96th Percentile	59th Percentile

Your Test Score History

General Test Scores

	Verbal Reasoning		Quantitative Reasoning		Analytical Writing	
Test Date	Scaled Score	Percentile	Scaled Score	Percentile	Score	Percentile
July 5, 2018	153	61	169	96	4.0	59

Subject Test Scores

You do not have reportable test scores at this time.

Your Score Recipient(s)

Undergraduate Institution

Report Date	Institution (Code)	Department (Code)	Test Title	Test Date
July 18, 2018	UM-SJTU Joint Institute (2147)	ANY DEPARTMENT NOT LISTED (5199)	General Test	July 5, 2018

Designated Score Recipient(s)

Report Date	Score Recipient (Code)	Department (Code)	Test Title	Test Date

PING SUN

Most Recent Test Date: July 5, 2018

Date of Birth: May 11, 1997

Registration Number: 3232243
Print Date: September 12, 2018

About Your *GRE*® Score Report

Score Reporting Policies

With the *ScoreSelect*® option, you can decide which test scores to send to the institutions you designate. There are three options to choose from:

- Most Recent option – Send your scores from your most recent test administration
- All option – Send your scores from all administrations in the last five years
- Any option – Send your scores from one OR as many test administrations in the last five years (this option is not available on test day when you select up to four FREE score reports)

Scores for a test administration must be reported in their entirety. Institutions will receive score reports that show only the scores that you selected to send to them. There will be no special indication if you have taken additional GRE tests. See the *GRE® Information Bulletin* for details. The policies and procedures explained in the Bulletin for the current testing year supersede previous policies and procedures in previous bulletins.

Scores will be sent to designated score recipients approximately 10-15 days after a computer-delivered test and 5 weeks after a paper-delivered test. If your scores are not available for any reason, you will see "Not Available" in Your Test Score History.

GRE test scores are reportable according to the following policies:

- For tests taken prior to July 1, 2016, scores are reportable for five (5) years following the testing year in which you tested (July 1 – June 30). For example, scores for a test taken on May 15, 2015, are reportable through June 30, 2020. GRE scores earned prior to August 2011 are no longer reportable.
- For tests taken on or after July 1, 2016, scores are reportable for five (5) years following your test date. For example, scores for a test taken on July 3, 2016, are reportable through July 2, 2021.

Note: Score recipients will only receive scores from test administrations that you have selected to send to them.

Percentile Rank (% Below)

A percentile rank for a test score indicates the percentage of test takers who took that test and received a lower score. Regardless of when the reported scores were earned, the percentile ranks for General Test and Subject Test scores are based on the scores of all test takers who tested within the most recent three-year period.

Retaking a GRE Test

You can take the *GRE*® General Test *once every 21 days*, up to *five times* within any continuous rolling 12-month period (365 days). This applies even if you canceled your scores on a test taken previously. You can take the paper-delivered GRE General Test and *GRE*® Subject Tests as often as they are offered.

Note: This policy will be enforced even if a violation is not immediately identified (e.g., inconsistent registration information) and test scores have been reported. In such cases, the invalid scores will be canceled and score recipients will be notified of the cancellation. Test fees will be forfeited.

For More Information

For information about interpreting your scores, see *Interpreting Your GRE Scores* at [www.ets.org/gre/understand.](www.ets.org/gre/understand)

For detailed information about your performance on the Verbal Reasoning and Quantitative Reasoning sections of the computer-delivered GRE General Test, access the free GRE Diagnostic Service from your ETS account. This service includes a description of the types of questions you answered right and wrong, the difficulty level of each question, and the time spent on each question. This service is available approximately 15 days after your test administration and for six months following your test administration.

If you have any questions concerning your score report, email GRE Services at **gre-info@ets.org** or call 1-609-771-7670 or 1-866-473-4373 (toll free for test takers in the U.S., U.S. Territories and Canada) between 8 a.m. and 7:45 p.m. (New York Time).

GRE

TEST TAKER SCORE REPORT

Note: This report is not valid for transmission of scores to an institution.

PING SUN

Most Recent Test Date: September 14, 2018

Address: ROOM424, EAST1, NO.800,DONGCHUAN, RD., MINHANG DIST., SHANGHAI, SHANGHAI, Shanghai, 200240 China

Registration Number: 3748436
Print Date: October 1, 2018

Email: 919374094@qq.com
Phone: 86-15821936985
Date of Birth: May 11, 1997
Social Security Number (Last Four Digits):
Gender: Male
Intended Graduate Major: Computer Science (0402)

Your Scores for the General Test Taken on September 14, 2018

Verbal Reasoning	Quantitative Reasoning	Analytical Writing
Your Scaled Score: 154	Your Scaled Score: 170	Your Score: 3.0
130 – 170	130 – 170	0 – 6
65th Percentile	96th Percentile	17th Percentile

Your Test Score History

General Test Scores

	Verbal Reasoning		Quantitative Reasoning		Analytical Writing	
Test Date	Scaled Score	Percentile	Scaled Score	Percentile	Score	Percentile
September 14, 2018	154	65	170	96	3.0	17
July 5, 2018	153	61	169	96	4.0	59

Subject Test Scores

You do not have reportable test scores at this time.

Your Score Recipient(s)

Undergraduate Institution

Report Date	Institution (Code)	Department (Code)	Test Title	Test Date
July 18, 2018	UM-SJTU Joint Institute (2147)	ANY DEPARTMENT NOT LISTED (5199)	General Test	July 5, 2018

Designated Score Recipient(s)

Report Date	Score Recipient (Code)	Department (Code)	Test Title	Test Date

PING SUN

Most Recent Test Date: September 14, 2018

Date of Birth: May 11, 1997

Registration Number: 3748436
Print Date: October 1, 2018

About Your *GRE*® Score Report

Score Reporting Policies

With the *ScoreSelect*® option, you can decide which test scores to send to the institutions you designate. There are three options to choose from:

- Most Recent option – Send your scores from your most recent test administration
- All option – Send your scores from all administrations in the last five years
- Any option – Send your scores from one OR as many test administrations in the last five years (this option is not available on test day when you select up to four FREE score reports)

Scores for a test administration must be reported in their entirety. Institutions will receive score reports that show only the scores that you selected to send to them. There will be no special indication if you have taken additional GRE tests. See the *GRE® Information Bulletin* for details. The policies and procedures explained in the Bulletin for the current testing year supersede previous policies and procedures in previous bulletins.

Scores will be sent to designated score recipients approximately 10-15 days after a computer-delivered test and 5 weeks after a paper-delivered test. If your scores are not available for any reason, you will see “Not Available” in Your Test Score History.

GRE test scores are reportable according to the following policies:

- For tests taken prior to July 1, 2016, scores are reportable for five (5) years following the testing year in which you tested (July 1 – June 30). For example, scores for a test taken on May 15, 2015, are reportable through June 30, 2020. GRE scores earned prior to August 2011 are no longer reportable.
- For tests taken on or after July 1, 2016, scores are reportable for five (5) years following your test date. For example, scores for a test taken on July 3, 2016, are reportable through July 2, 2021.

Note: Score recipients will only receive scores from test administrations that you have selected to send to them.

Percentile Rank (% Below)

A percentile rank for a test score indicates the percentage of test takers who took that test and received a lower score. Regardless of when the reported scores were earned, the percentile ranks for General Test and Subject Test scores are based on the scores of all test takers who tested within the most recent three-year period.

Retaking a GRE Test

You can take the *GRE*® General Test *once every 21 days*, up to *five times* within any continuous rolling 12-month period (365 days). This applies even if you canceled your scores on a test taken previously. You can take the paper-delivered GRE General Test and *GRE*® Subject Tests as often as they are offered.

Note: This policy will be enforced even if a violation is not immediately identified (e.g., inconsistent registration information) and test scores have been reported. In such cases, the invalid scores will be canceled and score recipients will be notified of the cancellation. Test fees will be forfeited.

For More Information

For information about interpreting your scores, see *Interpreting Your GRE Scores* at www.ets.org/gre/understand.

For detailed information about your performance on the Verbal Reasoning and Quantitative Reasoning sections of the computer-delivered GRE General Test, access the free GRE Diagnostic Service from your ETS account. This service includes a description of the types of questions you answered right and wrong, the difficulty level of each question, and the time spent on each question. This service is available approximately 15 days after your test administration and for six months following your test administration.

If you have any questions concerning your score report, email GRE Services at **gre-info@ets.org** or call 1-609-771-7670 or 1-866-473-4373 (toll free for test takers in the U.S., U.S. Territories and Canada) between 8 a.m. and 7:45 p.m. (New York Time).

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name:	**Ping Sun**
Program:	**Computer Science: MS**
Applicant Waived Rights*:	**This applicant has waived the right to view their recommendation.**
Recommender Name:	**Guangtao Xue**
Organization Name:	**Shanghai Jiao Tong University**
Title:	**Professor, PhD Advisor, Deputy Dean**
E-mail Address:	**xue-gt@cs.sjtu.edu.cn**
Telephone Number:	**+86[0]21-34205411**
Relationship to Applicant:	**Course instructor**
Certification (Date):	**10-19-2018**

*"Public Law 93-380, Educational Amendments Act of 1974, grants students the right to have access to letters of recommendation in their placement files. By selecting the "Waive access" option you are waiving access to these letters."

上海交通大學
Shanghai Jiao Tong University

Dear Admissions Committee,

I am writing in enthusiastic support of Ping Sun who is applying for your PROGRAM. I have known Ping since he was a student in my course in Computer Networks last fall. He is a young man with both genuinely creative ideas and scientific research abilities who has impressed me with his extraordinary project in my class.

Mr. Sun focused his project on solving problems around our lives. One of his most impressive projects "Automatic Tracking and Service System for Parking Lots" stood out immediately among students who attended my course in Computer Networks. His innovation of combining methods from computer visions on which he has long researched with knowledge in computer networks finally made a great success.

With careful preparation before the project, he communicated with me about his own creative idea and actively found support from the campus security office. As the leader of his group with another three students, he kept a regular meeting every week and assigned tasks to each member according to their individual abilities and experiences. Ping mainly focused his own part on the recognition of texts on vehicle plates and tracking the positions of vehicles with the technology of computer visions and other machine learning methods. During the whole process, Ping actively reported me the progression of his project and asked for advice to improve the project. With great effort, he invented new approaches to solve the hard problem of indoor positioning in various conditions replacing the traditional usage of the electronic signal in vehicle positioning, which has been proved to be both reliable and energy-saving. This project eventually facilitated the campus parking lot and helped improve the efficiency of management. This original invention was promising and has a potential to be applied in a larger market with further effort.

In addition to his innovative thinking and excellent insight in solving problems in reality with computer technology, he also showed great interpersonal skills. With strong coordination, communication and organization skills, the group he led was exceptional united and each one of them had a chance to fully contribute with their own abilities. Ping made each member complete a satisfying work with the reasonable amount with no exception. The final presentation he gave was very well prepared with fluent speaking and compelling content including both video and live demo.

Ping Sun is an impressive innovator with remarkable interpersonal skills, practical enough to think and solve problems using cross-disciplinary methods from multiple angles. Motivated by his own interests and enthusiasm, this extraordinary student is distinctive from many other test-oriented students and is very potential to make more innovations in the future academic study and research work. I highly recommend him to you without reservation.

Sincerely,

Guangtao Xue（薛广涛） Guangtao Xue
Ph.D., Professor
Deputy Dean, the School of Electronic, Information and Electrical Engineering (SEIEE)
Shanghai Jiao Tong University

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name: **Ping Sun**

Program: **Computer Science: MS**

Applicant Waived Rights*: **This applicant has waived the right to view their recommendation.**

Recommender Name: **Liqing Zhang**

Organization Name: **Shanghai Jiao Tong University**

Title: **Professor, Vice Dean, Dept. of Computer Science &**

E-mail Address: **zhang-lq@cs.sjtu.edu.cn**

Telephone Number: **+86-21-3420-4423**

Relationship to Applicant: **Intern Supervisor**

Certification (Date): **11-23-2018**

*"Public Law 93-380, Educational Amendments Act of 1974, grants students the right to have access to letters of recommendation in their placement files. By selecting the "Waive access" option you are waiving access to these letters."

Shanghai Jiao Tong University

Liqing Zhang
Professor, Vice Dean
Dept. of Computer Science & Eng.
Shanghai Jiao Tong University
Shanghai, 200240, China
Email: zhang-lq@cs.sjtu.edu.cn
Tel: +86-21-34204423

Dear Members of the Search Committee,

It is my pleasure to recommend Ping Sun for the graduate program in your department. Over the last year, Ping worked in my laboratory and showed his excellent research level and learning abilities.

I have known him for one year. Ping mentioned his previous researches with me including human pose estimation with CNN and Image Synthesis with GAN. Focusing on the tasks of Computer Visions, Mr. Sun dedicated himself to the projects wholeheartedly.

In the last summer, Ping did his intern in the Versa-SJTU AI joint lab and achieved some excellent research results there. He first was assigned the task of detecting affiliation in the image, with the purpose of figuring out the relationship between human and their accessories. Ping was actively involved in this project and tried varieties of approaches including adding depth information and salient objects detection to the existing RGB image information. Ping carried out experiments on various platforms under different conditions. Particularly, he took advantage of the new function of iPhone8 which provides a selfie with depth information and successfully made his algorithm work on iPhone8, resulting in high performance in semantic segmentation of affiliation.

Ping is innovative in research. In order to improve the performance of instance segmentation using Mask R-CNN, he read over 15 related papers and made many attempts. After several tries and error, he designed a new network combining ideas of both Mask R-CNN and Deeplabv3, resulting in high performance algorithm for instance segmentation. He is now working on how to solve the problem of body occlusion using the knowledge of human pose estimation.

He is a hard-working student. He reaches his dream step by step with a firm belief. He is a man with his own original opinions and thoughts. Before he entered my laboratory, he had focused his research on how to generate pictures according to the predefined concept. He worked with another student in my lab and finished the project Image Synthesis with Generative Adversarial Networks, which helps to change clothes of human in an image.

Ping Sun completed his senior year in my lab. He is one of the most hard-working students and skilled researchers. Since the beginning when he was invited to conduct his internship in Versa-SJTU AI lab, he worked the longest time of all the students and devoted himself with great passion to various projects as many as possible. The internship experience witnessed his rapid growth from an undistinguished beginner overcoming difficulties and making exciting progress. During his study in

Liqing Zhang
Professor, Vice Dean
Dept. of Computer Science & Eng.
Shanghai Jiao Tong University
Shanghai, 200240, China
Email: zhang-lq@cs.sjtu.edu.cn
Tel: +86-21-34204423

my lab, he has shown excellent adaptability, quick learning abilities and a great curiosity about new knowledge.

In short, I highly recommend Ping Sun to the graduate program of your department. With his ability and experience, I am confident that he will be successful in his future academic career. I would appreciate your favorable consideration of his application.

Liqing Zhang Liqing Zhang

Professor

Vice Dean of Computer Science Department
Shanghai Jiao Tong University

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name: **Ping Sun**

Program: **Computer Science: MS**

Applicant Waived Rights*: **This applicant has waived the right to view their recommendation.**

Recommender Name: **Xiaoju Dong**

Organization Name: **Shanghai Jiao Tong University**

Title: **Associate Professor**

E-mail Address: **dong-xj@cs.sjtu.edu.cn**

Telephone Number: **+86[0]21-34205060**

Relationship to Applicant: **Course Instructor**

Certification (Date): **10-26-2018**

*"Public Law 93-380, Educational Amendments Act of 1974, grants students the right to have access to letters of recommendation in their placement files. By selecting the "Waive access" option you are waiving access to these letters."

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

800 DONGCHUAN ROAD, SHANGHAI 200240, P. R. CHINA

To whom it may concern,

I'm very glad to learn that Ping Sun is applying for your graduate program. As his teacher, of the course Data Visualization and Visual Analytics, I deem it a great pleasure to recommend Ping, a very excellent student in the class.

Ping showed excellent performance in his studies. Although he had already acquired some data analysis foundation and JavaScript skills before he took up my class, he was still very modest and willing to take advice. Ping discussed a lot of questions through email with my teaching assistant and tried his best to gain knowledge and fill his learning gaps. Ping proved his rigorous and meticulous learning attitude through his homework. In the end, Ping ranked the top five for this course.

As in previous years, the best students in my class will team up for a national competition called China Vis Data Challenge. Ping teamed up with another three top students in the class to participate in the competition this year. In about two weeks, the four students worked together and contributed their strength without reservation. Ping bravely shouldered the heavy tasks to build the backend of the entire system, which was very unfamiliar to each member. Ping had to start from a beginner, but he showed very quick learning abilities and successfully built a framework within three days. After that, Ping immediately threw himself into the task of making front-end web pages helping his teammates to improve the answers. Sparing no efforts, Ping confidently completed five interactive web pages with cool visual effects and compelling information including various charts and maps. Finally, the team finished their paper and answers in analyzing the organization structure and some potential risks of an Internet company.

I consider that Ping Sun is an extraordinary student, not only because he works hard and keeps an inquisitive attitude, but he also proves himself a brave and reliable man with a strong sense of social responsibility. One day, I was teaching the course in the front of class. A boy was secretly taking pictures of a girl, which infringed the girl's privacy. Ping noticed the boy's indecent behavior. He calmly took photos of the boy's behavior and went up to catch the boy.

The course of Data Visualization and Visual Analytics requires students to learn new contents continuously and also need rich experience of data analysis. Ping Sun left a very deep impression on me in this course. He is a young man with an outstanding future. Therefore, I give him my highest recommendation for your consideration.

If I can be of any further assistance, or provide you with any further information, please do not hesitate to contact me.

Sincerely yours,
Dr. Xiaoju Dong, Associate Professor of Computer Science Xiaoju DONG
Department of CS, Shanghai Jiao Tong University, Shanghai, China
Email: xjdong@sjtu.edu.cn