Northwestern | THE GRADUATE SCHOOL

Application for Admission

App Type **New Student**

Submitted Date **10-23-2018**

App ID# **78578700**

Intended Status **Full-time**

Entry Quarter **Fall 2019**

Prior TGS Applicant (Program)

Last Name **Yang**

First **Mingyan**

Middle

Gender Pronouns (US only)

Birthdate **02-04-1997**

Gender **Female**

Program **Computer Science: MS**

Specialization/Area of Interest **Graphics**

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
No.100 Pingleyuan
Chaoyang District
Beijing, 100124
CHINA

Permanent Address
#3-5-501 C District, Boyangjingyuan
Bailusi, Chaoyang District
Beijing, 100121
CHINA

Current Phone

Permanent Phone

Cell Phone **+8613693052017**

Preferred Phone Number **Cell Phone Number**

Email Address **yangmingyan9724@163.com**

Previous Institution	From	To	Field of Study	Level	Degree	Date
Xi'an Jiaotong University	08-16-2015	07-01-2019	Computer Science and		International Undergraduate Degree	

Cumulative UG GPA	**3.30**	UG Junior/Senior Year GPA	
Cumulative UG GPA - Unconverted	**82.43**	Max UG GPA Scale	**100**
Cumulative Grad GPA			
Cumulative Grad GPA - Unconverted		Max Grad GPA Scale	

Letters of Recommendation

1. **Jie Lin** **jielin@mail.xjtu.edu.cn**
2. **Yuanqi Su** **yuanqisu@mail.xjtu.edu.cn**
3. **Junbin Gao** **junbin.gao@sydney.edu.au**
4.
5.

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

1. First Name Last Name
2. First Name Last Name
3. First Name Last Name
4. First Name Last Name

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: **Graduate or professional degree**

If other, please explain:

Second individual's highest level of education completed: **Graduate or professional degree**

If other, please explain:

Language	Reading	Writing	Speaking

Self-Reported Test Scores

GRE Gen	05-13-2018	Verbal	156	73	Quant	170	96	A.W.	4.5	82				
GRE Sub								LSAT						
TOEFL	10-29-2017	Ovr	100	Read	25	List	26	Speak	22	Writ	27	IELTS		Ovr
GMAT		Tot			Verb			Quant			A.W.			I.R.
MCAT		Bioscience			Verbal			Physical Science						

Please list any honors you have been awarded

Awarded Siyuan Scholarship of Xi'an Jiaotong University successively in the academic years of 2016 and 2017
due to excellent academic performance 10/2016, 10/2017
Awarded the title of "Social Activity Activists" of Nanyang College of Xi'an Jiaotong University 10/2016

Have you applied for or been awarded an external fellowship?

Yes ○ No ⊗ If yes, please specify;

Please describe your plans for the future.

If I got an opportunity to finish the graduate degree in your university, my short term goal will be focusing on developing and researching as a programmer in a Fortune Global 500 company in Silicon Valley, which offers perfect professional trainings and working environment for its staff. After I gain certain experiences, I will develop my management ability. In the long run, I hope I will start my computer-related business that develops its own APPs, as my ultimate goal.

Other Universities Applied (in preferred rank order)

1. School Drop Down	5. School "other"
2. School Drop Down	6. School "other"
3. School Drop Down	7. School "other"
4. School Drop Down	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

-by Mingyan Yang

In the current age, informationalization has become the necessities in our daily life. The demand for computer talents in Computer Science (CS) remains strongly increasing in all trades and professions in the last two years, at present and in the near future. From my point of view, the prospect of computer science is promising and I believe pursuing further study in this field in the United States can pave me a road to a splendid and meaningful career achievement.

Nearly four years undergraduate study in CS at Xi'an Jiaotong University, one among the top universities in this field in China, equipped me comprehensive computer science knowledge and laid me solid theoretical foundation in computer science and technology. I believe I am ready for doing overseas study with fully support from my family. It also indeed helped me understand that the United States has a leading position in the academic field of Computer Vision, as a major of CS, which is really attracting me to study in the US.

In my childhood, I was so fascinated by computers that I chose it as my major when I was admitted into Xi'an Jiaotong University. As time went on, I became more interested in computer and acquired more knowledge about it. Gradually, I came to realize that computer science is never just about programming like C Language, C++. It includes many hardware lessons such as Analog Electronics, Digital Circuits and Computer Organization as well as theoretical lessons such as Discrete Mathematics and Algorithm. And with the popularization, computer science is interactive and closely related with many other subjects like economics, medicine, etc. and at the same time has extensive applications in these fields.

While trying my best to lay a solid theoretical foundation for my future study, I also seized all the opportunities to do internships and actively involved in every aspect of these fields by doing professional practice, in which I became more and more proficient in applying computer science knowledge to interpret and solve practical computer problems. During June and July 2018, I found an internship as a Java Engineer in Sichuan Hwadee Information Technology Co. Ltd, where I participated in a project that developed the law enforcement information system of urban and rural management. I took part in the whole process including project determination, demand analysis, designing, coding and test. At the database designing stage, in order to establish relation between tables, I integrated some functions of the system and set ID attributes which was easy to differentiate. At the coding stage, in consideration of a short web developing experience, I learnt it all by myself and made some innovations so that we were able to develop a satisfactory front-end page. At the developing stage, when the interaction of input data and database couldn’t be realized, I tried Ajax and settled it. Besides, in this highly valuable professional experience, I also formed and trained my self-learning, communicating and professional competence. Therefore, with a sophisticated understanding in these areas, I have been well positioned to appreciate the basic theories of computer science and their applications.

In all of my research experiences, I cherish two projects most. Through the research procedures, I learnt Python all by myself and became more and more adept at it. Above all, I identified one developing direction for my future academic and professional careers, which run as follows: computer vision. The most significant project was *Generation of SMPL 3D Human Body Model Attitude and Recognition in Real Scene*, which I conducted from April 2018 till present in New Computer Research Institute, Xi'an Jiaotong University. From this valuable research experience, I got to know the wide applications of computer vision in my favorite movie industry. And interest is always the best teacher. I was deeply attracted by computer vision and decided to learn more about it in my graduate study. The other meaningful research project I conducted in in Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology of Beijing University of Technology was *Locality Preserving Projection via Deep Neural Network*. While widening my knowledge in deep neural network and sharpening my programming skills in Python through this project, I also greatly improved my English literature review ability. Also an academic paper as the main outcomes of this project has been submitted to "ACM Transactions on Knowledge Discovery from Data" and is currently under review. Deep neural network, combined with computer vision, will play significant roles in AI research, development and application in the future. Doing researches in this field certainly helps me understand how to apply theories to solve real-world problems.

Nevertheless, the reason I included computer vision in my future academic and professional plans was not only because I have a keen interest in it myself, but also because it will change people's life greatly. E.g., in autonomous car researches, computer vision is used to identify the image and make the right decision. Additionally, computer vision is also used in security, police investigation and special effect of the film.

On this note, it is obvious that the US will be the ideal destination for my further studies. I would like to seek the opportunity to have my graduate study in Northwestern University, compared with other schools and institutes. The courses that you offer particularly drew my interest, not only because I am interested in this area but more importantly, your courses represent world class quality. Meanwhile, your strong academic atmosphere, advanced teaching facilities and equipment as well as the distinguished faculties are also the important reasons for me to choose your prestigious university.

If I got an opportunity to finish the graduate degree in your university, my short term goal will be focusing on developing and researching as a programmer in a Fortune Global 500 company in Silicon Valley, which offers perfect professional trainings and working environment for its staff. After I gain certain experiences, I will develop my management ability. In the long run, I hope I will start my computer-related business that develops its own APPs, as my ultimate goal.

I am looking forward to the day that I can use my knowledge and skills to bring more benefits to our society and make people enjoy the convenience of computer.

西安交通大学
XI'AN JIAOTONG UNIVERSITY

本科生成绩单
Transcript of Undergraduate Student

姓名: 杨明妍
Name: Yang Mingyan
学院: 电子与信息工程学院
School: The School of Electronic and Information Engineering

入学年月：2015年08月 Date of Enrollment: August 2015	学号：2150500147 Student No.: 2150500147	出生日期：1997年02月04日 Date of Birth: February 04,1997	专业：计算机科学与技术 Major: Computer Science And Technology
毕业年月：年月 Date of Graduation:	学制：4年制 Length of Schooling: Four Years	性别：女 Gender: Female	专业班：计算机55

绩点：3.24 GPA: 3.24
学分成绩：82.43 Average Score: 82.43

课程 Course	学分 Credit	成绩 Score	绩点 GPA
第一学年（2015-2016）第一学期 1st Academic Year（Semester 1）（2015-2016）			
体育1 Sports 1	0.5	98	4.3
军训 Military Training	1	95	4.3
国防教育 National Defence Education	1	83	3.3
大学英语IV College English IV	2	81	3.3
大学英语四级(CET4) College English Test(CET4)		561	
心理学基础 Basics of Psychology	2	93	4.0
思想道德修养与法律基础 Moral and Legal Education	3	94	4.0
文物鉴赏 Appreciation of Cultural Relics	2	85	3.7
程序设计基础 Programming Fundamentals	3	87	3.7
线性代数与解析几何II Linear Algebra and Analytic Geometry II	3.5	86	3.7
计算机科学技术导论 Introduction of Computer Science and Technology	1	81	3.3
高等数学I1 Advanced Mathematics I 1	6.5	74	2.3
第一学年（2015-2016）第二学期 1st Academic Year（Semester 2）（2015-2016）			
中国近现代史纲要 Outline of Modern Chinese History	2	63	1.3
体育2 Sports 2	0.5	100	4.3
大学物理II1 University Physics II1	4	68	2.0
大学物理实验I1 University physics experiments I 1	1	87	3.7
大学生职业发展与规划 Career Planning	2	82	3.3
大学英语六级(CET6) College English Test (CET 6)		569	
材料与人类文明 Materials and Human Civilization	2	92	4.0
毛泽东思想和中国特色社会主义理论体系概论1 Mao Zedong Thought and Introduction to the theoretical system of socialism with Chinese characteristics1	2	91	4.0
离散数学A Discrete Mathematical Structures	4	78	3.0
通用学术英语 English for General Academic Purpose	2	84	3.3
面向对象程序设计 Object-oriented Programing	2.5	76	2.7
高等数学I2 Advanced Mathematics I 2	6.5	82	3.3
第一学年（2015-2016）第三学期 1st Academic Year（Semester 3）（2015-2016）			
健康与疾病 Health and Diseases	2	86	3.7
第二学年（2016-2017）第一学期 2nd Academic Year（Semester 1）（2016-2017）			
“数据结构与程序设计”专题实验 "Data structure and program design"Subject experiment	1	B +	3.3
体育3 Sports 3	0.5	99	4.3
公共演讲 Public Speaking	2	84	3.3
大学物理II2 University Physics II2	4	80	3.0
大学物理实验I2 University physics experiments I 2	1	85	3.7
戏剧鉴赏 Drama Appreciation	2	89	3.7
数据结构与算法A Data Structure and Algorithm (A)	3.5	86	3.7
模拟电子技术基础 Analog electronics	3.5	86	3.7
毛泽东思想和中国特色社会主义理论体系概论2 Mao Zedong Thought and Introduction to the theoretical system of socialism with Chinese characteristics2	2	71	2.0
电子技术实验1 Electronics Experiment1	0.75	优秀 A	4.0
电路 Circuits	4.5	71	2.0
第二学年（2016-2017）第二学期 2nd Academic Year（Semester 2）（2016-2017）			
优化方法基础 Fundamental of optimization	2	87	3.7
体育4 Sports 4	0.5	94	4.0
工程制图III Engineering Drawing III	2	91	4.0
形式语言与编译 Formal Language and Compiler	3.5	70	2.0
数字逻辑电路 Digital Logic Circuit	3.5	71	2.0
数理逻辑 Mathematical Logic	2	82	3.3
概率统计与随机过程 Probability Theory and Stochastic Process	4	79	3.0
汇编语言 Assembly Language	2.5	80	3.0
生命科学基础I essentials of life scienceI	3	79	3.0
电子技术实验2 Electronics Experiment2	0.75	A -	3.7
电工实习 Electrician Practice	1	A -	3.7
编译器设计专题实验 Lab Course of Compiler Design	1	90	4.0
英语辩论 Debate in English	2	83	3.3
金工实习 Metalworking Practice	2	A	4.0
第二学年（2016-2017）第三学期 2nd Academic Year（Semester 3）（2016-2017）			
专业实习Ⅰ Specialized practice I	1	优秀 A	4.0
测控实习 Measurement and Control Practice	1	A +	4.3
第三学年（2017-2018）第一学期 3rd Academic Year（Semester 1）（2017-2018）			
24式太极拳 24-Style Tai Ji Quan	0	70	2.0
操作系统原理I The Principle of Operating System	3	87	3.7
操作系统设计专题实验 The special topic experiment of Operating System	1	94	4.0
算法分析与设计 The AnalysisandDesign of Algorithms	2	76	2.7
计算机图形学 Computer Graphics	2.5	86	3.7

西安交通大学成绩单专用章 (2)

西安交通大學
XI'AN JIAOTONG UNIVERSITY

本科生成绩单
Transcript of Undergraduate Student

姓名: 杨明妍
Name: Yang Mingyan
学院: 电子与信息工程学院
School: The School of Electronic and Information Engineering

入学年月：2015年08月 Date of Enrollment: August 2015	学号：2150500147 Student No.: 2150500147	出生日期：1997年02月04日 Date of Birth: February 04,1997	专业：计算机科学与技术 Major: Computer Science And Technology
毕业年月：年月 Date of Graduation:	学制：4年制 Length of Schooling: Four Years	性别：女 Gender: Female	专业班：计算机55

绩点：3.24
GPA: 3.24
学分成绩：82.43
Average Score: 82.43

课程 Course	学分 Credit	成绩 Score	绩点 GPA
计算机组成 Computer Organization	4	78	3.0
计算机组成与结构专题实验 Computer Organization and Architecture Laboratory	1	85	3.7
计算机网络原理 Principles of Computer Networks	3	79	3.0
马克思主义基本原理 basic principle of Marxism	3	85	3.7
第三学年（2017-2018）第二学期 3rd Academic Year（Semester 2）（2017-2018）			
数据库系统 Database System	3	81	3.3
计算机系统综合设计实验 Computer Systems Design Experiments	1	A	4.0
计算机网络专题实验 Special Experiments on Computer Networks	1	92	4.0
计算机视觉与模式识别 Computer Vision and Pattern Recognition	2.5	91	4.0
软件定义网络 Software Defined Networking	2	86	3.7
软件工程 Software Engineering	2.5	93	4.0

百分制 Centesimal Grade	95~100	90~94	85~89	81~84	78~80
等级 Grades	优+(A+)	优(A)	优-(A-)	良+(B+)	良(B)
绩点 GPA	4.3	4.0	3.7	3.3	3.0

75~77	72~74	68~71	64~67	60~63	0~59
良-(B-)	中+(C+)	中(C)	中-(C-)	及格(D)	不及格(F)
2.7	2.3	2.0	1.7	1.3	0

GPA=∑课程学分X绩点/∑课程学分
(采用二等级制记载的课程成绩不参与GPA计算)
GPA=∑creditX grade/∑credit
(Course marks recorded by two-tier system are not aclculated by GPA)

咨询电话(Tel)：+86-29-82666406

注册员：周湘灏
Registrar: Zhou Xianghao

打印时间:2018年9月4日
Issued Date:September 4,2018

西安交通大学成绩单专用章 (2)
盖章 Seal:

西安交通大學 教务处
Administrative Department for Undergraduate Education
Add: Room 1303 Main Teaching Building, No. 28 Xianning West Road, Xi'an, Shaanxi Province, China, 710049
Tel:+86-29-82665804 Fax:+86-29-82668301 E-mail:xawei@mail.xjtu.edu.cn

证 明

(Certificate)

杨明妍，性别：女，出生年月：1997年02月04日。2015年08月进入西安交通大学学习，现就读于电子与信息工程学院计算机科学与技术专业（本科四年制）。学号：2150500147。

特此证明。

Yang Mingyan， female, born on Feb 04, 1997， was enrolled in Xi''an Jiaotong University in Aug , 2015. Now she is studying in The School of Electronic and Information Engineering, majoring in Computer Science And Technology. The length of schooling is four years and student NO. is 2150500147.

西安交通大学教务处

2018年09月04日

学籍证明专用章 (2)

Educational Administration Office of

Xi''an Jiaotong University

Sep 04, 2018

Mingyan Yang

yangmingyan9724@163.com | +8613693052017

#3-5-501 C District, Boyangjingyuan, Bailusi, Chaoyang District, Beijing 100121, P.R.China

EDUCATION BACKGROUND

Xi'an Jiaotong University 08/2015-07/2019

- **Major:** Computer Science and Technology
- **Degree:** Bachelor of Engineering in Computer Science and Technology Expected in July 2019

PUBLICATION

- Tianhang Long, Junbin Gao, **Mingyan Yang**, Yongli Hu, Baocai Yin, "Locality Preserving Projection via Deep Neural Network", ACM Transactions on Knowledge Discovery from Data, Under review

RESEARCH EXPERIENCES

Undergraduate Research Assistant 04/2018-present

New Computer Research Institute, Xi'an Jiaotong University

Program: Generation of SMPL 3D Human Body Model Attitude and Recognition in Real Scene

Supervisor: Prof. Yuanqi Su

✧ **Responsibilities:**

- Used SMPL package to generate tens of thousands of sequence images samples of human body under normal walking conditions;
- Generated a video of human walking with image samples;
- Improved the SMPL model and converted a wide range of BVH file materials into different human body postures of the SMPL model;
- Generated videos in .obj file format; applied the generated videos to recognize different postures of human body in real scene through machine learning knowledge.

✧ **Procedures:**

- Applied the two variables m.pose[:] and m.betas[:] to control the changes of attitude and stature(such as tall, low, fat and thin) of different parts of a 3D SMPL model;
- Obtained the dimension of these two variables and figured out the corresponding characteristics of different body parts or postures of each dimension;
- Gained the result that m.pose[:] was applied to control rotation directions and angles of different joints of human body and m.betas[:] was used to control the length and thinness of different parts of human body.

✧ **Innovation:**

- Selected a BVH file of walking action that is very close to the walking status of human being and had some of its joint nodes corresponded to the 23 joint nodes of the SMPL model, found the rotation angle of each joint and recorded the corresponding relation into an excel file;
- Wrote a program to read the nodes of corresponding BVH file and SMPL model, saved it as a txt format file, then read txt file to generate an image sequence of walking postures and connected the image sequence into a video;
- Improved the SMPL model, read the .pkl file in original SMPL package and added a root node to the original 23 joint nodes and formed a SMPL model based on numpy;
- Filtered the joint points of human body defined in the BVH file of multiple human posture movements, selected the joint points that could correspond to the SMPL model and named them uniformly;
- Wrote a program to read the joint points and rotation angles of BVH file and directly formed the corresponding joint points and rotation angles in the SMPL model, obtaining a more general conversion method from extensive BVH file materials to SMPL model;
- Exported the generated action model in in .obj file format, which could observe the generated human motion model all round.

✧ **Achievement:**

- Self-learned Python and sharpened my programming skills; understood the wide applications of computer vision in animation, film&TV industry and inspired my great interest in computer vision.

Undergraduate Research Assistant 01/2018-08/2018

Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing University of Technology

Program: Locality Preserving Projection via Deep Neural Network

Supervisor: Prof. Junbin Gao

- Implemented a proposed novel nonlinear locality preserving projection method via deep neural network, termed as DNLPP, which replaces the linear projection with an appropriate deep neural network;

- Found the proposed method is more discriminative than the conventional LPP, benefiting from the nonlinearity of neural networks and its powerful representation capability;
- Implemented a proposed iterative optimization algorithm to solve the new model;
- Did extensive experiments on several public datasets and found that the proposed method is overall superior to the state-of-art dimensionality reduction methods;
- Completed the academic paper and submitted it to "ACM Transactions on Knowledge Discovery from Data" and it is currently under review.

INTERNSHIP EXPERIENCES

Sichuan Hwadee Information Technology Co. Ltd 06/2018-07/2018

Position: *Java Engineer*

- Participated in a project that developed a law enforcement information system of urban and rural management;
- Analyzed the required functions of the project, confirmed the requirement degree level of different functions and drew use-case diagram and activity diagram;
- Designed the database, draw class diagram and sequence diagram;
- Involved in the web development; completed the front-end and back-end development and tested the system.

ACADEMIC REPORTS

Camera calibration and library height measurement 06/2018

- Applied MATLAB's own camera Calibrator to calibrate;
- Selected the library and flagpole in the pictures after gaining the internal parameters of the camera;
- Calculated the library height with cross-ratio calculations and obtained the depth information to generate point cloud.

COMPETITION, EXTRACURRICULAR&VOLUNTEER EXPERIENCES

***Participator*, "Haier Maker Cup" Creativity Design Competition** 09/2016

Project: Intelligent Integrated Storage Space of Future Kitchen

- Designed an integrated storage space in a cylindrical shape, with a double-door structure, equipped with a temperature sensor, humidity sensor and pressure sensor inside that can help people take foods conveniently, store foods to avoid them go mouldy and intelligently remind you food intake;
- Equipped with a corresponding APP to implement information interaction and real-time control, having the chip transmitted the data received by the sensor to the cloud in real time through the WiFi module and APP acquired the corresponding data of the user in the cloud;
- Performed data analysis and set timing reminding intelligently based on the data processing result.

Class commissary in charge of psychology 09/2015-present

- Organized psychological trainings and provided psychological counseling for students;
- Observed students' psychological status; kept timely and effective communication with students that have abnormal behaviors and helped them release psychological stress.

***Swim coach*, Swimming Team of Nanyang College of Xi'an Jiaotong University** 09/2015-07/2016

- Taught team members how to swim and improved their swimming skills.

Active participator in various volunteer activities

- Participated in various volunteer activities including volunteering in Model UN, subway station, Xi'an International Horticultural Expo Garden, School Open Day as well as volunteer for the maintenance of STEI, etc.

HONORS AND AWARDS

- Awarded Siyuan Scholarship of Xi'an Jiaotong University successively in the academic years of 2016 and 2017 due to excellent academic performance 10/2016, 10/2017
- Awarded the title of "Social Activity Activists" of Nanyang College of Xi'an Jiaotong University 10/2016

PROFESSIONAL SKILLS

- **Computer skills:** Proficient in using C language(3 yrs) and C++(2 yrs), MATLAB(2 yrs), Java(1 yr), and Python(1 yr)
- **Design Skills:** AutoCAD(1 yr), Photoshop(1 yr)
- **GRE:** 326(V:156+Q:170+AW:4.5)
- **Hobbies:** Swimming, Travelling, Reading, Film, and Music

Locality Preserving Projection via Deep Neural Network

Tianhang Long, Beijing University of Technology
Junbin Gao, The University of Sydney
Mingyan Yang, Xi'an Jiaotong University
Yongli Hu, Beijing University of Technology
Baocai Yin, Dalian University of Technology

Dimensionality reduction is an essential problem in data mining and machine learning fields. Locality Preserving Projection (LPP) is a well-known dimensionality reduction method which can preserve the neighborhood graph structure of data, and has achieved promising performance. However linear projection makes it difficult to analyze complex data with nonlinear structure. In order to deal with this issue, this paper proposes a novel nonlinear locality preserving projection method via deep neural network, termed as DNLPP, which replaces the linear projection with an appropriate deep neural network. Benefiting from the nonlinearity of neural networks and its powerful representation capability, the proposed method is more discriminative than the conventional LPP. In order to solve the new model, we propose an iterative optimization algorithm. Extensive experiments on several public datasets illustrate that the proposed method is overall superior to the other state-of-art dimensionality reduction methods.

CCS Concepts:•**Computing methodologies** → *Recognition and understanding; Dimensionality reduction; Learning latent representations;*

Additional Key Words and Phrases: Dimension Reduction, Locality Preserving Projection, Deep Neural Network, Subspace Clustering

ACM Reference Format:
Tianhang Long, Junbin Gao, Mingyan Yang, Yongli Hu and Baocai Yin, 2018. Locality Preserving Projection via Deep Neural Network. *ACM Trans. Knowl. Discov. Data.* 1, 1, Article 1 (June 2018), 18 pages.
DOI: 0000001.0000001

1. INTRODUCTION

In many practical problems, acquired data often have high dimensionality, such as image and video data. High dimensional data not only increase computational complexity and memory requirement of algorithms, but also affect algorithm performance in real applications. Many researches have confirmed that high dimensional data generally lie in or close to a low dimensional space or manifold. A dimension reduction algorithm aims to effectively and properly reveal such an embedded low-dimensional space from high dimensional data by removing redundant information and noise. Do-

Tianhang Long and Yongli Hu are with Beijing Advanced Innovation Center for Future Internet Technology, Beijing Municipal Key Lab of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China. E-mail: long54482000@163.com, huyongli@bjut.edu.cn
Junbin Gao is with the Discipline of Business Analytics, The University of Sydney Business School, The University of Sydney, NSW 2006, Australia. E-mail: junbin.gao@sydney.edu.au
Mingyan Yang is with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: yangmingyan9724@126.com
Baocai Yin is with the College of Computer Science and Technology, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology , Dalian 116024, China. E-mail: ybc@dlut.edu.cn

DOI: 0000001.0000001

ing so can improve algorithm performance in a wide range of applications, including but not limited to face recognition [He et al. 2005b; Xie et al. 2016], feature extraction [Luo et al. 2016; Wang and Gao 2016] and so on.

In the past few decades, great progress in the research on dimensionality reduction algorithms and models have been achieved. The most widely used dimension reduction methods include the Principal Component Analysis (PCA) [Jolliffe 1986], Canonical Correlation Analysis (CCA) [Sun et al. 2010], Independent Component Analysis (ICA) [Comon 1994] and Locality Preserving Projections (LPP) [He et al. 2005b].

PCA aims to learn a linear transformation by maximizing the variance of projected data or equivalently minimizing the reconstruction error when recovering the data from their low dimensional counterpart. CCA performs dimensionality reduction by finding a set of canonical variables (linear combinations of primitive variables) with the largest correlation coefficient, as the correlation between the canonical variables is used to reflect the correlation of the original variables. ICA is a method to find hidden factors or components from multivariable (multidimensional) statistics. Different from PCA, ICA aims to find a linear transformation that maximizes the independence of each component. Because ICA believes that all components are of equally important and maximizing the independence of each component can reveal some hidden factors. LPP focuses on seeking the low dimensional linear projection while preserving the locality structure of data. This has been realized via preserving the data neighborhood graph structure under the graph Laplacian framework.

These methods have been used in machine learning and computer vision tasks widely. In addition to the face recognition [He et al. 2005b; Cai et al. 2006; Xie et al. 2016] and feature extraction [Ohbuchi et al. 2008; Luo et al. 2016; Wang and Gao 2016] mentioned before, they are also used in dimensionality reduction [He et al. 2005a; Lee and Verleysen 2007; Van der Maaten et al. 2009].

All these methods assume that the mapping from higher dimensional data space to the lower dimensional counterpart is linear, represented by a matrix projection mathematically. However, the real data is always high dimensional and complicated, these methods can not get good performance on it.

To deal with high dimensional and complicated data, many nonlinear dimension reduction methods have been proposed. The typical one is the kernel-based method, including KPCA [Hoffmann 2007] and KLDA [Huang et al. 2004]. Like the most kernel methods, a kernel-based dimension reduction method usually implements a nonlinear mapping by mapping data onto a high dimensional feature where a linear mapping is applied. This enables the nonlinear relationship of data in low dimensions to be represented by linear relations of the high dimensional features. Although the kernel-based methods are successful, they suffer from difficulties in choosing suitable kernel functions. It is well known that the performance of a particular kernel-based method depends on the chosen kernel function. How to choose an appropriate kernel function for a particular learning task is still an open question [Hofmann et al. 2008].

On the other front, many researchers design nonlinear dimensionality reduction methods by either explicitly or implicitly exploring the manifold structure hidden in the training data. In manifold learning research, many nonlinear dimension reduction algorithms or models have been proposed, including isometric feature mapping (ISOMAP) [Tenenbaum et al. 2000], locally linear embedding (LLE) [Roweis and Saul 2000], Laplacian Eigenmaps (LE) [Mikhail and Partha 2009] and (FEDRA) [Magdalinos et al. 2011]. ISOMAP, LLE and LE explore manifold information by constructing the adjacency matrix with local information, then the intrinsic low dimensional representation is obtained by keeping the similar local manifold structure. In FEDRA, one follows the landmark-based manifold structure to embed data objects in a low-dimensional projection space.

Although a manifold learning method may reveal the low dimensional manifold structure embedded in high-dimensional data, it may suffer from the so-called out-of-sample problem [Trosset and Priebe 2008]. The problem comes from the fact that there is no concrete mapping from high dimensional space to low dimensional space in many manifold learning algorithms. Additionally, some manifold learning methods may also be limited by the specific assumption of the nonlinear properties of the data. For example, LLE implicitly assumes that the data is locally linear and LPP incorporates the local manifold structural information or neighbor information of the data, but this hypothesis may not be always correct in practice. LPP can be regarded as the linear approximation of the LE, offering the manifold learning ability and preserving the local structure in the transformed domain.

In order to make more flexible nonlinear dimensionality reduction methods, we turn our attention to deep learning techniques. The primary reason is that deep learning has the power to represent any nonlinear mapping and has achieved huge success in numerous applications, especially in supervised learning, e.g. image classification [Krizhevsky et al. 2012], metric learning [Hu et al. 2014], image super-resolution [Wang et al. 2015], and data embedding [Chang et al. 2015]. The other reason is that to combine traditional algorithms with deep learning is an interesting field of research and may improve the performance of traditional algorithms. We have observed that few recently works have successfully used deep learning in learning tasks. such as Structural Deep Network Embedding (SDNE) [Wang et al. 2016], Projective Low-rank Subspace Clustering (PLrSC) [Li et al. 2017] and the Cascade Subspace Clustering (CSC) [Peng et al. 2017]. SDNE presents a new semi-supervised deep learning model, which combines the advantages of first-order proximity and second-order proximity to represent the global structure properties and local structural properties of the network. PLrSC and CSC learn the low-rank or compact representations by using neural network instead of the conventional transformation methods and make great improvement in clustering performance.

Although deep learning has the potential of improving the performance of unsupervised learning methods, to the best of our knowledge, there are few works in deep learning for unsupervised data dimensionality reduction. The most related works are the Auto-Encoder (AE) based dimensionality reduction methods [Ma et al. 2014; Tian et al. 2014; Peng et al. 2016]. In [Ma et al. 2014; Peng et al. 2016], a certain clustering criterion, such as the Davies-Bouldin validity index [Ma et al. 2014] or the Sparse Subspace Clustering criterion [Peng et al. 2016], is added into the AE objective of learning. However, in [Peng et al. 2016], the sparse prior is pre-learned before the deep learning procedure, so it is a two stage procedure. In [Tian et al. 2014], a deep neural network (DNN) was used to learn the similarity of data graph, by training a series of AEs layer by layer. In the procedure, a dimensionality reduction can be automatically implemented by using a hidden layer with a smaller number of neurons.

In this paper, we propose a novel nonlinear dimensionality reduction model combining LPP and deep neural network models, named as NLPP, in which deep neural network is employed as the nonlinear mapping which not only maps the samples to a lower dimensional space, but also preserves local neighbor structure simultaneously. The key idea in our approach is to introduce a oblique manifold constraint [Absil et al. 2008] on the deep networks, resulting in a new optimization problem on manifolds.

The rest of the paper is organized as follows. In Section 2, we review the necessary knowledge for classic Locality Preserving Projection (LPP). Section 3 proposes the nonlinear locality preserving projection via the deep neural network model, and the detailed optimization algorithm is given. In Section 4, the performance of the proposed model is evaluated on several public datasets. Finally, the conclusions and future works are summarized in Section 5.

2. BACKGROUND THEORY

In this section, we review the Locality Preserving Projection (LPP) and Deep Neural Networks to pave the way for introducing the proposed method.

2.1. Locality Preserving Projection (LPP)

For the purpose of extending LPP to deep learning architecture, we first give a quick review of related concepts of the LPP. The goal of LPP is to find low-dimensional embedding of high dimensional data such that the local neighbor structure of high dimension data can be preserved in the low dimension space. The LPP implements this goal via building a linear projective mapping which preserves the neighborhood structure of the data, encoded in the so-called Laplacian graph. Thus, the LPP falls in the category of linear dimensionality reduction algorithm.

Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N] \in \mathbb{R}^{P\times N}$ be N training samples, where $\mathbf{x}_i \in \mathbb{R}^P$ is the i-th sample with dimension P. LPP seeks for a projection direction $\mathbf{a} \in \mathbb{R}^P$ such that the projected value $y_i = \mathbf{a}^T\mathbf{x}_i$ ($i = 1, ..., N$) fulfills the following condition,

$$\min_{\mathbf{a}} \sum_{i,j=1}^{N} (y_i - y_j)^2 \cdot s_{ij} = \sum_{i,j=1}^{N} (\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j)^2 \cdot s_{ij} = \mathbf{a}^T\mathbf{X}(\mathbf{D}-\mathbf{S})\mathbf{X}^T\mathbf{a}. \tag{1}$$

where matrix $\mathbf{S} = [s_{ij}] \in \mathbb{R}^{N\times N}$ is a symmetric weight matrix whose elements s_{ij} measures the similarity or affinity of samples $\mathbf{x}_i$ and $\mathbf{x}_j$. To avoid a trivial solution for problem (1), it is a common practice to impose a constraint such as

$$\mathbf{y}\mathbf{D}\mathbf{y}^T = \mathbf{a}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a} = 1, \tag{2}$$

where $\mathbf{y} = [y_1, ..., y_N]$ and $\mathbf{D} = \text{diag}[d_{ii}]$ is a diagonal matrix whose elements are column sum of matrix $\mathbf{S}$ with $d_{ii} = \sum_{j=1}^{N} s_{ij}$.

Define $\mathbf{L} = \mathbf{D} - \mathbf{S}$, which is called the graph Laplacian matrix. Thus the final LPP model becomes

$$\begin{aligned} &\min_{\mathbf{a}} \sum_{i,j}^{N} \frac{1}{2}(\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j)^2 s_{ij} = \mathbf{a}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}, \\ &\text{s.t. } \mathbf{a}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a} = 1. \end{aligned} \tag{3}$$

A possible choice for the element s_{ij} of similarity matrix $\mathbf{S}$ is through the so-called Gaussian heat kernel as defined as follows:

$$s_{ij} = \begin{cases} \exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right\}, & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ in } K\text{- neighborhood}; \\ 0, & \text{otherwise}. \end{cases} \tag{4}$$

With a given $\mathbf{S}$, minimizing the objective function (1) of LPP is to ensure if $\mathbf{x}_i$ and $\mathbf{x}_j$ are similar to each other, then the projected values $y_i = \mathbf{a}^T\mathbf{x}_i$ and $y_j = \mathbf{a}^T\mathbf{x}_j$ are close to each other.

The optimization (3) of the LPP model can be easily solved by the following generalized eigenvalue problem,

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a}. \tag{5}$$

2.2. Deep Neural Network

Deep neural networks give a way of defining a complex, non-linear form of hypothesis $f(\mathbf{x}, \Theta)$ where Θ collects all the weights and biases parameters in the deep network. In a typical deep structure, the hypothesis is formed in the following forward process:

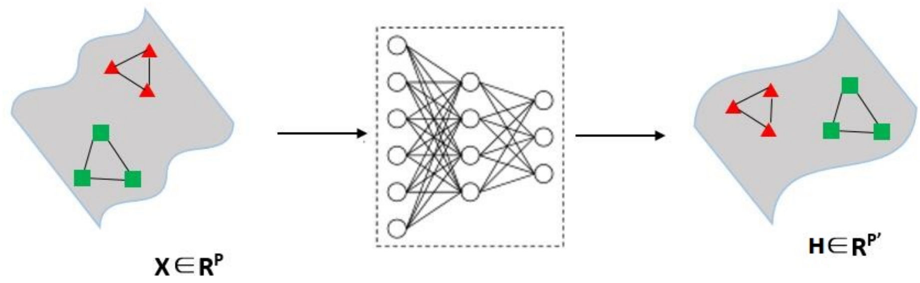


Fig. 1: The illustration of our NLPP: Taking a given data set $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ as inputs, with the constraint of (13), the outputs $\mathbf{H} = \{\mathbf{h}_1, \cdots, \mathbf{h}_N\}$ produced by neural network to be regarded as the low-dimension representation of the input data.

Let $\mathbf{h}_i^{(0)} = \mathbf{x}_i \in \mathbb{R}^{p_0}$ (where $p_0 = P$ for convenience, $i = 1, ..., N$) denote the i-th input sample and

$$\mathbf{h}_i^{(m)} = g(\mathbf{W}^{(m-1)}\mathbf{h}_i^{(m-1)} + \mathbf{b}^{(m-1)}) \in \mathbb{R}^{p_m};\ m = 1, 2, ..., M,$$

be the activations of the m-th layer and $f(\mathbf{x}_i, \Theta) = \mathbf{h}_i^{(M)}$ is regarded as the activation over the last layer for the i-th input sample and $\Theta = \{\mathbf{W}^{(0)}, \mathbf{b}^{(0)}, \mathbf{W}^{(1)}, \mathbf{b}^{(1)}, ..., \mathbf{W}^{(M-1)}, \mathbf{b}^{(M-1)}\}$ contains all the network weight and bias parameters. Clearly we have $\mathbf{W}^{(m-1)} \in \mathbb{R}^{p_m \times p_{m-1}}$ and $\mathbf{b}^{(m-1)} \in \mathbb{R}^{p_m}$, which are associated with the m-th layer, respectively. The dimension $p = p_M$ of the output layer will serves as the data dimension after dimensionality reduction. We are interested in the outputs $\{\mathbf{h}_i^{(M)}\}$. Further, we denote by

$$\mathbf{H}^{(m)} = [\mathbf{h}_1^{(m)}, \mathbf{h}_2^{(m)}, ..., \mathbf{h}_N^{(m)}] \in \mathbb{R}^{p_m \times N}. \tag{6}$$

For our convenience, we write $\mathbf{H} = \mathbf{H}^{(M)}$ and $p = p_M$.

In a supervised learning setting, one constructs an appropriate objective function for training the network by matching the outputs $\mathbf{h}_i$ with the given targets $\mathbf{y}_i$ under a task dependent loss or criterion. For example, the classical neural networks regression aims to minimize the overall loss, aligned with a regulariser, defined as

$$\begin{aligned}\mathcal{J} =& \frac{1}{\mathbf{N}}\sum_i \frac{1}{2}\|f(\mathbf{x}_i, \Theta) - \mathbf{y}_i\|_2^2 + \frac{\lambda}{2}r(\Theta)\\ =& \frac{1}{\mathbf{N}}\sum_i \frac{1}{2}\|\mathbf{h}_i - \mathbf{y}_i\|_2^2 + \frac{\lambda}{2}r(\Theta) = \frac{1}{2N}\|\mathbf{H} - \mathbf{Y}\|_F^2 + \frac{\lambda}{2}r(\Theta).\end{aligned} \tag{7}$$

Then training a network involves minimizing the objective function $\mathcal{J}$ with respect to all the network parameters Θ. The classic backpropogation (BP) can be used to calculate all the parameter derivatives which are needed in an optimization algorithm such as gradient descent (or stochastic gradient descent). The BP starts calculating $\frac{\partial \mathcal{J}}{\partial \mathbf{h}_i}$ and passes $\frac{\partial \mathcal{J}}{\partial \mathbf{h}_i}$ backwards layer by layer to calculate all the parameters derivatives $\frac{\partial \mathcal{J}}{\partial \mathbf{W}^{(m-1)}}$ and $\frac{\partial \mathcal{J}}{\partial \mathbf{b}^{(m-1)}}$. Finally one iteration of gradient descent step updates the parameters $\mathbf{W}^{(m-1)}, \mathbf{b}^{(m-1)}$ as follows:

$$\mathbf{W}^{(m-1)} = \mathbf{W}^{(m-1)} - \alpha \left(\frac{\partial \mathcal{J}}{\partial \mathbf{W}^{(m-1)}} + \lambda \frac{\partial r}{\partial \mathbf{W}^{(m-1)}} \right) \tag{8}$$

$$\mathbf{b}^{(m-1)} = \mathbf{b}^{(m-1)} - \alpha \frac{\partial \mathcal{J}}{\partial \mathbf{b}^{(m-1)}}. \tag{9}$$

where α is an appropriate learning rate.

3. THE NONLINEAR LOCALITY PRESERVING PROJECTIONS MODEL

The proposed Nonlinear Locality Preserving Projections (NLPP) model builds on a deep neural network to nonlinearly map high dimensional data to possibly low dimensional form for which the locality of the original data should be preserved. Fig. 1 illustrates the proposed NLPP model.

3.1. The NLPP Based on Deep Neural Network

As mentioned in subsection 2.2, the outputs produced by a deep network are presented in $\mathbf{H}$ as defined in (6). The deep network has actually defines a nonlinear mapping from the high dimensional data to possibly low dimensional, i.e., $f\colon \mathbf{x}_i \in \mathbb{R}^P \to \mathbf{h}_i = f(\mathbf{x}_i, \Theta) \in \mathbb{R}^p$ for the given training dataset $\mathbf{X}$ of N samples. The proposed NLPP will seek for the nonlinear mapping $f(\mathbf{x}, \Theta)$ by minimizing the LPP locality preserving criterion. Thus the mapping should not only fit the training data but also preserve locality property of the training data.

In the typical supervised learning setting in applying neural networks, the objective function is defined as the overall sum of the loss of each individual data pattern, e.g. see (7). In the LPP setting, the similarity $\mathbf{S}$ can be regarded as the target information, hence the LPP objective offers a collective loss function over the entire dataset, which can be similarly defined as

$$\min_{\mathbf{\Theta}} \sum_{i,j} \frac{1}{2} \|f(\mathbf{x}_i, \Theta) - f(\mathbf{x}_j, \Theta)\|_2^2 \cdot s_{ij} = \sum_{i,j} \frac{1}{2} \|\mathbf{h}_i - \mathbf{h}_j\|_2^2 \cdot s_{ij}, \tag{10}$$

where $\mathbf{\Theta} = \{\mathbf{W}^{(0)}, \mathbf{b}^{(0)}, \cdots, \mathbf{W}^{(M-1)}, \mathbf{b}^{(M-1)}\}$ denotes the neural network parameters of all the layers.

Different from the traditional neural network, in the LPP case, the objective function is a weighted summation as defined in (10), in which each term is calculated for a pair of data, resulting in N^2 terms. This may increase the complexity of Back Propagation (BP) algorithm in which derivative with respect to $\mathbf{W}$ and $\mathbf{b}$ must be backwards passed on. This gives arise a tremendous increase in computation when the amount of data is larger. We can also note that, due to the paired terms in (10), it is not easy to employ batch-sized stochastic training procedure. However when we take into account the special structure of the objective function, we manage to propose an efficient BP algorithm next.

As usual, let us assume $\mathbf{S} = [s_{ij}]_{i,j=1}^N$ is a symmetric similarity matrix in LPP, and define the Laplacian by $\mathbf{L} = \mathbf{D} - \mathbf{S}$ (where $\mathbf{D}$ is the diagonal matrix with the sum of each row of S as the components), then it is well known that the objective (10) can be written in the matrix form as follows,

$$\min_{\Theta} \frac{1}{2} \mathbf{tr}(\mathbf{H}\mathbf{L}\mathbf{H}^T). \tag{11}$$

Obviously, optimization problem (11) may achieve a trivial solution, which will give a non-meaningful dimensionality reduction. To avoid this problem, certain constraint regularizations must be imposed. Some widely-used regularizations method constraint network parameters, and are independent of training samples, including weight decay, DropOut [Srivastava et al. 2014], and DropConnect [Wan et al. 2013]. Intuitively, weight decay abates overfitting by reducing the magnitude of the weights and the features. DropOut and DropConnect can be regarded as low-cost methods for training deep neural networks. Their effectiveness as network regularizers can be quantified by analyzing the Rademacher complexity, which provides an upper bound for the generalization error [Bartlett 1998; Wan et al. 2013]. These methods only change the network structure between layers to avoid non-meaningful solutions, but not encourage the network to produce desirable features. Along this line, borrowing the ideas from [Ma et al. 2014; Peng et al. 2016], we can simply add a decoder networks top-up our network so that an AE criterion can be added into (11). However, this will further increase the computational complexity, and do not play an important role in producing meaningful features.

To focus on the structures of both the input data and the output features, we tend to encourage the network to produce structurally meaningful features.

The trivial solution of (11) may come from the unbounded scale of $\mathbf{H}$. In fact, the standard LPP has the similar issue which is resolved by constraining the reduced features to a given fixed scale. Inspired by this idea, we propose the following constrained optimization problem for our NLPP,

$$\min_{\Theta} \frac{1}{2}\mathbf{tr}(\mathbf{HLH}^T) \quad \text{s.t.} \quad \mathbf{HDH}^T = \mathbf{I}_{p\times p}, \tag{12}$$

where $\mathbf{I}_{p\times p}$ is an identity matrix with size p. Note that the constrained problem (1) is not simply a manifold optimisation problem, but the constraint condition is specified through the so-called relaxed Stiefel manifold, see [Absil et al. 2008], which was implicitly introduced in the standard LPP formulation. It is a quite strong constraint which will result in the reduced uncorrelated dimensionality feature representation. Moreover, problem (12) is highly nonlinear over all the parameter Θ, and this will make it very difficult to solve the optimization model (12). To ease this issue, we relax this constraint by proposing the following optimization model for NLPP

$$\min_{\Theta} \frac{1}{2}\mathbf{tr}(\mathbf{HLH}^T) \quad \text{s.t.} \quad \text{diag}(\mathbf{HDH}^T) = \mathbf{I}_{p\times p}, \tag{13}$$

where $\text{diag}(\cdot)$ denotes the diagonal matrix from the diagonal elements of a matrix. This constraint domain is a manifold, i.e., the Oblique manifold [Absil et al. 2008]. In other words, the outputs on the last layer become the points on the Oblique manifold. As this manifold constraint is not directly applied on the optimization variables (i.e., weights and biases), (13) cannot be simply regarded as an optimization problem on the Oblique manifold. In the next subsection, we will optimize this new objective function.

3.2. Model Optimization

For our convenience, we introduce the following notation: Let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n] \in \mathbb{R}^{m\times n}$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n] \in \mathbb{R}^{m\times n}$, we define

$$\mathbf{U}\widetilde{\otimes}\mathbf{V} = [\mathbf{u}_1 \otimes \mathbf{v}_1, \mathbf{u}_2 \otimes \mathbf{v}_2, \cdots, \mathbf{u}_n \otimes \mathbf{v}_n]^T \in \mathbb{R}^{n\times m^2},$$

where $\mathbf{u}_i = [u_{i1}, u_{i2}, \cdots, u_{im}]^T$, $\mathbf{u}_i \otimes \mathbf{v}_i$ is the Kronecker product of vector $\mathbf{u}_i$ and $\mathbf{v}_i$ whose definition is

$$\mathbf{u}_i \otimes \mathbf{v}_i = [u_{i1}\cdot \mathbf{v}_i, u_{i2}\cdot \mathbf{v}_i, \cdots, u_{im}\cdot \mathbf{v}_i]^T \in \mathbb{R}^{m^2\times 1}$$

In order to solve model (13), we reformulate this constraint optimization as following unconstrained problem

$$\mathcal{J} = \frac{1}{2}\mathrm{tr}(\mathbf{HLH}^T) + \frac{\lambda}{2}\|\mathrm{diag}(\mathbf{HDH}^T) - \mathbf{I}_{p\times p}\|_F^2 \tag{14}$$

Note that (14) is not the Lagrange multiplier problem of (13), as a squared norm term $\|\mathrm{diag}(\mathbf{H}^T\mathbf{DH}) - \mathbf{I}_{p\times p}\|_F^2$ is added into the objective function.

Denote $\mathcal{J}_1 = \frac{1}{2}\mathrm{tr}(\mathbf{HLH}^T)$ and $\mathcal{J}_2 = \frac{\lambda}{2}\|\mathbf{Q}\|_F^2$, where $\mathbf{Q} = \mathrm{diag}(\mathbf{HDH}^T) - \mathbf{I}_{p\times p}$. Thus $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$. Clearly, we have the following differentials

$$d\mathcal{J}_1 = \frac{1}{2}(\mathbf{H}(\mathbf{L}+\mathbf{L}^T)) :^T d\mathbf{H} : \tag{15}$$

$$d\mathcal{J}_2 = \lambda\mathbf{Q} :^T d\mathbf{Q} : \text{ with } d\mathbf{Q} := \mathbf{I}_{p\times p}\widetilde{\otimes}\mathbf{I}_{p\times p} d\mathbf{Z} : \tag{16}$$

where $\mathbf{Z} = \mathbf{HDH}^T$, $\mathbf{I}_{p\times p}$ is the identity matrix and $\mathbf{H}:$ denotes the long column vector formed by concatenating the columns of $\mathbf{H}$. Further we have

$$\begin{aligned} d\mathbf{Z} : &= (\mathbf{I}_p \otimes \mathbf{HD})d\mathbf{H}^T : + (\mathbf{HD}^T \otimes \mathbf{I}_p)d\mathbf{H} : \\ &= (\mathbf{T}_{[p,p]}(\mathbf{HD}\otimes\mathbf{I}_p) + \mathbf{HD}^T \otimes \mathbf{I}_p)d\mathbf{H} : \\ &= (\mathbf{I}_{p\times p} + \mathbf{T}_{[p,p]})(\mathbf{HD}\otimes\mathbf{I}_p)d\mathbf{H} : . \end{aligned} \tag{17}$$

where $\mathbf{I}_p$ is a $p\times 1$ vector and the values of all elements are 1 , $\mathbf{T}_{[m,n]}$ is vectorized transpose transformation, which is defined as

$$\mathbf{X}^T := \mathbf{T}_{[m,n]}\mathbf{X} : \text{ for any matrix } \mathbf{X} \in \mathbb{R}^{m\times n}.$$

Combining with (16) and (17), we can obtain

$$d\mathcal{J}_2 = \lambda\mathbf{Q}^T(\mathbf{I}_{p\times p}\widetilde{\otimes}\mathbf{I}_{[p,p]})(\mathbf{I}_{p\times p} + \mathbf{T}_{[p,p]})(\mathbf{HD}\otimes\mathbf{I}_p) : d\mathbf{H} : .$$

Finally from

$$d\mathcal{J} = d\mathcal{J}_1 + d\mathcal{J}_2,$$

we can easily work out that

$$\begin{aligned} \frac{\partial\mathcal{J}}{\partial\mathbf{H}} =& \frac{1}{2}\mathbf{H}(\mathbf{L}+\mathbf{L}^T) \\ &+ \lambda\mathbf{Q}^T(\mathbf{I}_{p\times p}\widetilde{\otimes}\mathbf{I}_{p\times p})(\mathbf{I}_{p\times p} + \mathbf{T}_{[p,p]})(\mathbf{HD}\otimes\mathbf{I}_p). \end{aligned} \tag{18}$$

The derivative information $\frac{\partial\mathcal{J}}{\partial\mathbf{H}}$ calculated according to (18) can be backpropogated to the neural network.

The algorithm 1 summarizes the detailed procedure for optimizing our NLPP model.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate the performance of the proposed dimensionality reduction algorithms NLPP, we conduct several experiments for image recognition and clustering tasks on some public databases and simulation data. These experiments aim at demonstrating the performance of the proposed model in feature extraction ability by comparing with other related algorithms. The compared dimensionality reduction algorithms include the standard PCA, LPP, LLE, Laplacian Eigenmaps and Isomap. We regard PCA as our benchmark algorithm. All used codes of compared algorithms are from drtoolbox (Matlab Toolbox for Dimensionality Reduction), and are implemented on PC machine installed a 64_bit operating system with an Intel Core i3-2120 3.30GHz CPU machine with 8G RAM. Our NLPP algorithm is coded by in Python with Tensrorflow.

ALGORITHM 1: The optimization algorithm for NLPP model

Input: Given training sample set $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, the nearest neighbor parameters $K_{neighbor}$, the layer number M of neural network and the balancing penalty parameter λ, the learning rate α, the number of iterations K.

Output: The output of the K-th iteration $\mathbf{H} = \{\mathbf{h}_1, \cdots, \mathbf{h}_N\}$

```
Initialize {W^(m), b^(m)}_{m=1}^M and H^(0) = X, k = 0.
Constructing the Laplacian matrix L of data X by (4);
while k ⩽ K do
    for m = 1, 2, ··· , M do
        Computing H^(m) via H^(m) = g(W^(m−1) H^(m−1) + b^(m))
    end for
    Using (18) to calculate derivative ∂J/∂H
    for m = M − 1, M − 2, ··· , 1 do
        Calculating partial derivatives by using the standard BP algorithm ∂J/∂W^(m−1) and ∂J/∂b^(m−1).
        Updating W^(m) and b^(m) by gradient descent as (8) and (9).
    end for
    k = k + 1
end while
```

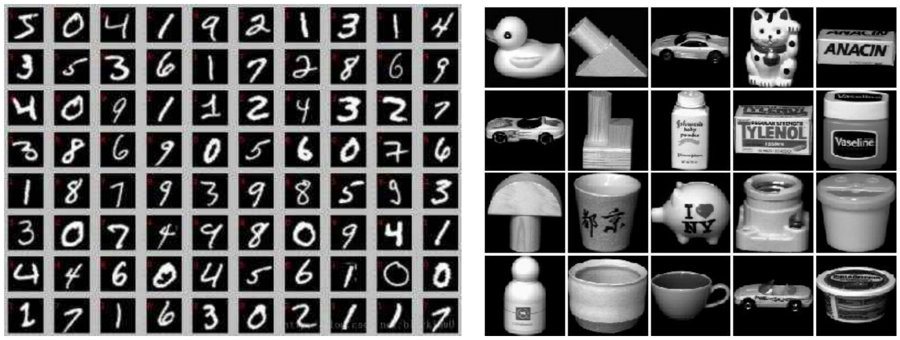
Fig. 2: Sample images of CMU-PIE database

4.1. Databases and Experimental Settings

The experiments of image recognition and clustering are conducted on five public databases: CMU Multi-PIE face image database[1], MNIST handwritten digital image database[2], Georgia Tech face image database[3], ORL face image database[4] and COIL20 object image database[5].

CMU-PIE database (Fig. 2) is composed of 68 individuals with 41,368 face images. The images are acquired with 13 poses, 43 illumination conditions and 4 expressions for each individual with size 64×64. For computational convenience, we choose the ninth pose (Pose09) for the experiment and all images are further resized to 32×32 and convert the RGB into gray image. Pose09 has 1632 images for 68 people, with 24 images each person. In image recognition, eighteen images of each person are used for training, and the remaining for test.

[1] https://www.flintbox.com/public/project/4742/
[2] http://yann.lecun.com/exdb/mnist/
[3] http://www.anefian.com/research/face_reco.htm
[4] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[5] http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php



(a) (b)

Fig. 3: Sample images of (a) MNIST database and (b) COIL20 database

(a) (b)

Fig. 4: Sample images of (a) Georgia Tech face database and (b) ORL face database

MNIST database (Fig. 3(a)) consists of 70000 handwritten digits images, which has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. All the examples in the database have been centered and size-normalized into a fixed-size of 28×28. All examples in the training set and test set are used for image recognition tasks.

COIL20 database (Fig. 3(b)) is an object image set with 20 categories, and each category contains 144 images (in which 72 images is with background information and another 72 images is without background). We choose the object images in which the background has been discarded (and the images consist of the smallest square that contains the object), then resize each image from 128×128 to 32×32 and use 60 of them for training, the left 12 for test in the image recognition task.

Georgia Tech database (Fig. 4(a)) is a face image database which consists of 50 subjects with 15 natural color face images for each subject. All face images in the database are with cluttered background taken at resolution 640×480 pixels in two or three sessions between 06/01/99 and 11/15/99 at the Center for Signal and Image Processing at Georgia Institute of Technology. The images show frontal and tilted faces with different facial expressions, lighting conditions and pose. Each image is manually labeled to determine the position of the face in the image. In the experiment, all the images are resized to 32×32 and converted into gray scale. For each subject, we choose 12 images for training and the left 3 images for test.

Table I: Recognition accuracy (%) on CMU-PIE face database

Reduced Dimension		NLPP	LPP	PCA	LLE	LE	Isomap	AE
10	1-NN	87.75	13.25	**95.25**	86.25	33.5	91.25	79.25
	SVM	68.38	11.27	59.07	45.34	44.61	69.85	**94.75**
40	1-NN	**100**	28	97	99.75	65.5	97.5	92.75
	SVM	**99.02**	24.75	94.36	97.31	86.03	92.64	98.75
70	1-NN	**100**	42.75	97.25	99.75	80.5	98.25	92.5
	SVM	**100**	34.06	98.28	98.52	89.95	95.34	99.25
100	1-NN	**100**	52.5	97	99.75	87	98.25	96
	SVM	**100**	50	99.26	99.02	86.52	96.81	99.25

ORL database (Fig. 4(b)) contains 40 distinct individual and each individual consists of 10 images. For some individuals, the images were taken at different times, illumination changes, facial expressions variation (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position. Each image is with a resolution of 92×112, and 256 gray levels of per pixel. We normalize all the images into a size 32×32, and 8 of each subject are for training and the left 2 images for test.

In our experiments, we use the sigmoid activation function of the neural network which is defined as follows:

$$g(z) = \mathrm{sigmoid}(z) = \frac{1}{1+e^{-z}}.$$

In our NLPP, we train a three layer fully connected network network with a learning rate of $\alpha = 3 \times 10^{-5}$ and the weight coefficient regulariser $\lambda = 1$. The original feature vector of an image is the concatenation of intensity of all pixels.

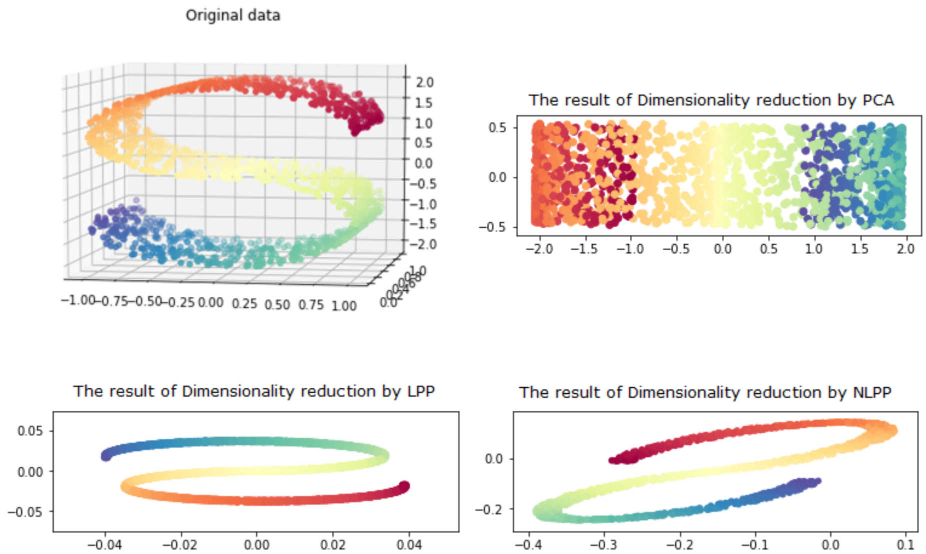
4.2. Experiments Based on Simulation Data

To prove that our proposed algorithm can preserve the local structure in the low-dimension space, we conduct the experiment on the simulated S-curve and Swiss roll datasets. The simulated data are generated by `sklearn` toolbox in Python. Each of the two datasets contain 1500 sample points and the noise parameter is set to 0.5. The results in Fig. 5(a) and Fig. 5(b) show that LPP can preserve the local structure of the simulated data while the benchmark PCA cannot. Comparing with the results of LPP and PCA, it can be easily seen that dimensionality reduction results of NLPP have the same local structure as LPP. This demonstrates NLPP's ability of preserving local structures.

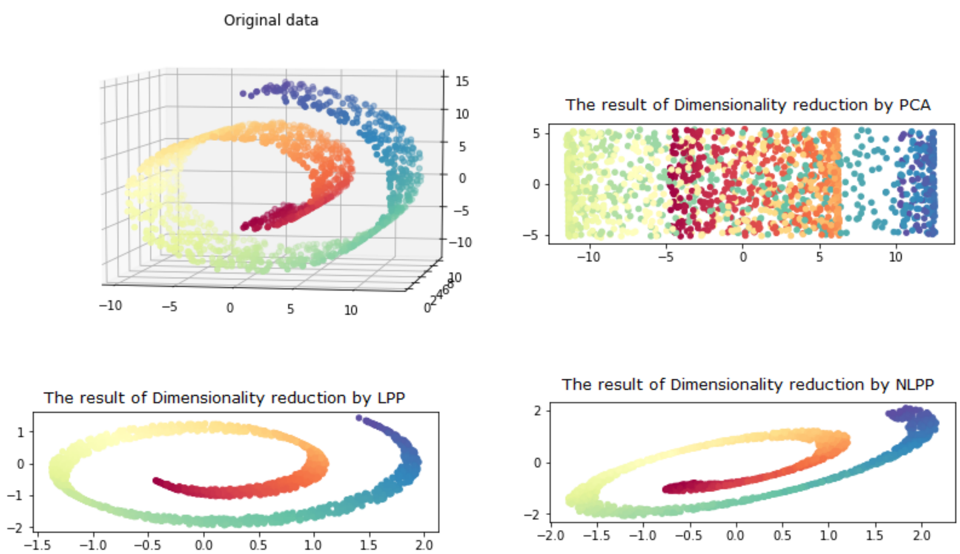
Fig. 6 is the result of blobs experiment, aiming to test the discriminating power of NLPP. It shows three clusters of isotropic Gaussian data blobs. We reduce it to 1-dim with the algorithm of PCA, LPP and our proposed NLPP. Obviously, the experiment results show that both PCA and LPP cannot separate the yellow and purple blobs, while NLPP can distinguish them clearly. The experiments demonstrate that NLPP has the better discriminating power than PCA and LPP.

4.3. Image Recognition and Cluster

To evaluate the feature representation ability of NLPP dimensionality reduction method, we implement classification and clustering experiments by our proposed method on practical datasets.



(a)



(b)

Fig. 5: (a) Dimensionality reduction results on S curve and (b) Dimensionality reduction results on swiss roll

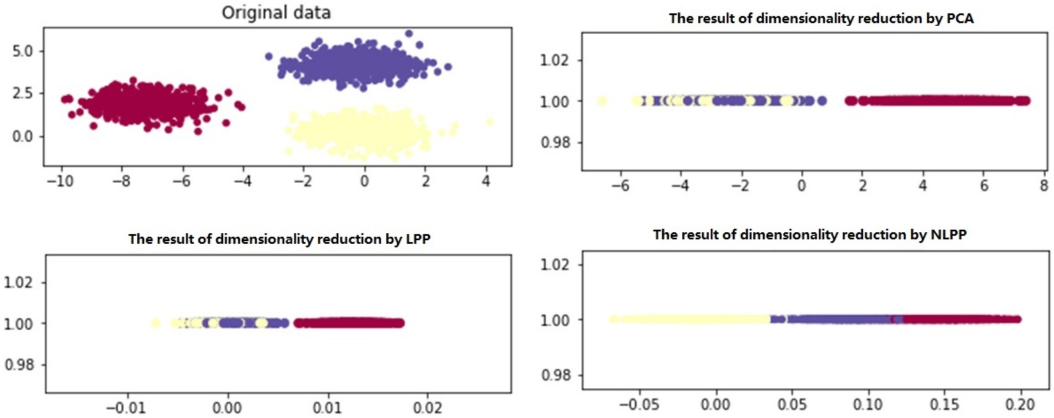


Fig. 6: Results of Gaussian blobs simulation experiment

Table II: Recognition accuracy (%) on Georgia Tech face database

Reduced Dimension		NLPP	LPP	PCA	LLE	LE	Isomap	AE
10	1-NN	**98**	83.33	97.33	88.4	89.12	91.2	92.6
	SVM	86.67	78.33	80	92	88	81.33	**96.67**
40	1-NN	**98.67**	89.33	97.33	92.92	90.8	92	93.2
	SVM	**98.67**	83.67	80	94.67	97.33	92.67	98.67
70	1-NN	**98.67**	89.33	97.33	89.58	90.72	93.6	93.2
	SVM	**98.67**	86.33	94	94.67	98	94.67	98.67
100	1-NN	**98.67**	96	98	88.5	89.67	92.6	93
	SVM	**98.67**	92.33	96	98	97.33	97.33	98.67

Table III: Recognition accuracy (%) on ORL face database

Reduced Dimension		NLPP	LPP	PCA	LLE	LE	Isomap	AE
10	1-NN	**90.83**	80.83	90	88.33	76.67	85	90
	SVM	71.67	68.33	86.67	89.17	89.17	**90**	87.24
40	1-NN	**94.17**	83.33	90.8	93.3	92.5	90	93.3
	SVM	**93.33**	87.5	90.67	93.33	93.33	92.5	90.3
70	1-NN	**93.3**	85	90	92.5	88.33	88.33	93.3
	SVM	**95.83**	89.33	93.33	95	91.67	93.33	95
100	1-NN	**95**	85.83	90.8	92.5	80	86.67	92.5
	SVM	**96.67**	91.67	94.17	96.67	90	93.33	96.67

In image recognition experiment, we divide the data set $\mathbf{X}$ into training and testing sets $\mathbf{X} = \{\mathbf{X}_{train}, \mathbf{X}_{test}\}$. The data set $\mathbf{X}_{train}$ is used to train our neural network and simultaneously obtain the low-dimensionality representation $\mathbf{H}_{train}$. Then, we get $\mathbf{H}_{test}$ through the trained neural network. Finally, $\mathbf{H}_{train}$ and $\mathbf{H}_{test}$ are sent to the nearest neighbor classifier (NNC) [Cover and Hart 1967] and Support Vector Machine(SVM) [Cortes and Vapnik 1995] to acquire the final classification accuracy.

Table IV: Recognition accuracy (%) on COIL20 database

Reduced Dimension		NLPP	LPP	PCA	LLE	LE	Isomap	AE
10	1-NN	69.92	37.92	**95.83**	90.83	77.92	94.17	94.92
	SVM	59.17	33.33	77.08	88.33	88.33	74.58	**93.51**
40	1-NN	**95.83**	72.5	90.42	92.92	93.75	95	94.08
	SVM	**95.42**	70.83	89.17	89.17	93.06	91.25	94.58
70	1-NN	**94.59**	70.41	86.25	89.58	88.33	94.17	94.08
	SVM	**96.67**	70.83	89.17	82.08	96.25	89.17	95.21
100	1-NN	**95.83**	68.75	78.75	82.5	70	92.08	93.67
	SVM	**94.17**	68.33	92.92	81.25	91.25	86.25	92.92

Table V: Recognition accuracy (%) on MNIST database

Reduced Dimension		NLPP	LPP	PCA	LLE	LE	Isomap	AE
10	1-NN	**95.35**	37.6	90.8	88.4	89.12	91.2	92.6
	SVM	54.8	30.4	76	74.6	**88**	83.8	81.4
40	1-NN	**95.49**	46.4	94.92	92.92	90.8	92	93.2
	SVM	**89.43**	33.2	87.5	88.6	88.8	85.6	87.2
70	1-NN	**94.92**	49.2	94.22	89.58	90.72	93.6	93.2
	SVM	**89.53**	36.7	89	89.4	89.8	88.2	88.8
100	1-NN	**95.49**	48.4	94.57	88.5	89.67	92.6	93
	SVM	**92.13**	45.4	89.8	91.6	90.2	89.4	89

Table VI: Clustering accuracy and NMI (%) on COIL20 database (given 20 classes)

Reduced Dimension	NLPP		LPP		PCA	
	ACC.	NMI	ACC.	NMI	ACC.	NMI
10	58.34	62.25	**69.55**	**82.53**	69.41	79.39
20	63.58	68.45	**81.57**	**90.78**	63.63	74.37
30	62.64	71.65	**81.51**	**89.09**	63.67	74.98
40	70.8	76.87	**72.63**	**87.56**	61.58	70.48
50	**69.86**	76.54	69.59	**84.62**	55.41	69.54
60	**68.37**	**77.16**	58.78	76.30	53.34	69.02
70	**69.04**	**76.05**	48.26	73.10	50.19	67.07
80	**68.91**	**76.33**	40.63	67.29	49.69	66.37
90	**69.61**	**78.19**	29.34	60.14	50.23	66.37
100	**69.44**	**77.31**	28.78	56.27	46.67	65.27

As the proposed NLPP is an nonlinear dimensionality reduction method based on LPP, we compare the performance of NLPP with LPP, and the standard nonlinear dimensionality reduction method such as LLE, Laplacian Eigenmaps and Isomap. We also compare NLPP with PCA which is a typical linear dimensionality reduction algorithm. Meantime, the NLPP is a kind of dimensionality reduction based on deep learning algorithm, so we compare it with the Auto-Encoder (AE). The image recognition experiment is based on COIL20 object image database, the CMU Multi-PIE face image database, the MNIST handwritten digit image databases, Georgia Tech face image database and the ORL face image database. In the experiments, the training and test samples are chosen as the way described above, and the Nearest Neighbor classifier (1-NN) and Support Vector Machine(SVM) are adopted for classification.

Table VII: Clustering accuracy and NMI(%) on MNIST database (given 10 classes)

Reduced Dimension	NLPP		LPP		PCA	
	ACC.	NMI	ACC.	NMI	ACC.	NMI
10	39.55	25.24	23.58	9.58	**57.1**	**50.71**
20	42.81	31.98	22.7	9.52	**55.75**	**48.82**
30	40.98	33.24	22.95	10.17	**46.57**	**41.23**
40	**46.37**	**39.05**	24.4	10.11	44.56	38.79
50	**46.74**	**38.43**	23.95	10.44	41.49	38.13
60	**49.95**	**42.49**	24.68	11.16	37.25	35.88
70	**53.13**	**44.26**	22.62	9.41	35.46	32.65
80	**51.45**	**45.57**	22.7	9.62	31.05	31.71
90	**55.28**	**48.67**	23.5	9.87	30.65	29.90
100	**57.51**	**49.13**	24.25	10.48	28.9	29.19

Table VIII: Clustering accuracy and NMI (%) on CMU-PIE database (given 68 classes)

Reduced Dimension	NLPP		LPP		PCA	
	ACC.	NMI	ACC.	NMI	ACC.	NMI
10	**22.21**	**50.76**	17.34	46.78	18.12	49.81
20	**22.58**	**55.32**	20.5	50.12	20.59	50.28
30	**26.7**	**57.22**	22.04	53.12	23.65	53.04
40	**27.94**	**58.42**	22.17	53.45	25.86	54.09
50	**30.31**	**58.67**	23.67	55.52	26.83	55.59
60	**30.13**	**60.01**	25.03	56.48	26.5	57.05
70	**30.13**	**59.80**	25.89	57.45	26.83	57.74
80	**30.34**	**61.67**	27.67	59.22	27.43	60.58
90	**31.69**	**60.46**	29.63	60.36	27.48	60.01
100	**31.83**	**61.29**	29.79	60.40	27.65	61.23

Table I reports the experimental results of seven feature extraction algorithms, NLPP, PCA, LPP, LLE, Laplacian Eigenmaps, Isomap and Auto Encoder on the CMU-PIE database with the reduced dimension ranging from 10 to 100. It demonstrates that the recognition accuracy of our proposed NLPP has a slight advantage against the other algorithms. We have highlighted the best results for different algorithms in all dimensions. Further, we set contrast experiment on Georgia Tech database and the ORL face database. In Georgia Tech database, the data have cluttered background, different facial expressions and pose. In the ORL face database, the data have more changes on facial expressions and pose. Table II and Table III show the recognition accuracy on Georgia Tech database and the ORL face database respectively. It can be seen obviously that our NLPP have good performance on the two database and is robust as the dimension becomes higher. These results can illustrate that our NLPP has better ability for features extraction. Especially, our NLPP has better performance than LPP in almost all cases. This demonstrate that the nonlinear LPP based on deep neural network can extract better representation feature for classification.

In order to verify the robustness of our algorithm for rotating objects, we set the experiment on COIL-20 database which consists of 20 objects shot from different angles. Table IV shows the results of recognition. Our NLPP reaches the best performance as the dimension rises and then keeps robust when the results of other compared algorithms get worse. We also set an experiment on MNIST database and Table V shows the experimental results, which demonstrate that NLPP performs better than other compared algorithms.

The clustering experiments are conducted on COIL20 database, MNIST database and CMU-PIE database. The compared methods include LPP, for the purpose showing the benefit of using deep learning in LPP, and PCA as the benchmark algorithm for dimensionality dimension. The features extracted by the three dimensionality reduction algorithms are sent to Normalized cuts (N-cut) [Shi and Malik 2000] for clustering, and the cluster numbers are set as the subset numbers of each database respectively.

Table VI shows the clustering accuracy and NMI [Cover and Thomas 1991] of different algorithms on COIL20 database. We can find from this table that LPP and PCA have some advantages in the case of lower dimension but worse in the case of high dimension. On MNIST database, Table VII reports that the results of NLPP is better than those of LPP, this demonstrates that the our NLPP has better feature extraction ability than LPP, but PCA still has slight advantages in the cases of low dimensions. Table VIII shows that, on CMU-PIE database, NLPP performs better than LPP and PCA in every dimension. These results demonstrate that as the dimensions rises, the features extracted by our NLPP can maintain a relatively good clustering performance.

In general, the proposed NLPP is comparable with the state of the art algorithms. All these experiments show that the new NLPP is more robust in cases of a higher reduced dimensions, so our conclusion is, for a safety guard, a slight larger size on the output layer may guarantee a better feature extraction capacity of the neural networks at least for classification and clustering tasks.

5. CONCLUSIONS

In this paper, we proposed a nonlinear LPP model (NLPP). This model extended the unsupervised LPP to handle nonlinear data by incorporating a deep neural network mapping to replace the linear mapping. To avoid trivial solution and encourage the deep network to produce structurally meaningful features, a new manifold constraint regular term is introduced. The optimization algorithm of NLPP model is derived.

The experiments on simulation data demonstrated that the proposed NLPP model not only preserve local structure of original data, but also have discriminating power in low-dimension space obviously. The experiments of image recognition and clustering on several public databases aim to test the ability of feature extraction and representation. The experimental results of image recognition and clustering tasks have verified the effectiveness and superiority of the proposed method in terms of linear LPP and some dimensionality reduction methods.

Acknowledgments

This research was supported by National Natural Science Foundation of China under Grant 61772048 and 61672071, in part by the Beijing Natural Science Foundation under Grant 4172003, 4162010 and 4184082. The work of Junbin Gao is supported by the University of Sydney Business School ARC Bridging Fund (2017).

REFERENCES

P. A Absil, R Mahony, and R Sepulchre. 2008. *Optimization algorithms on matrix manifolds*. Princeton University Press.

P. L. Bartlett. 1998. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory* 44(2) (1998), 525–536.

D. Cai, X.F. He, J.W. Han, and H.J. Zhang. 2006. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing* 15, 11 (2006), 3608–3614.

S.Y. Chang, W. Han, J.L. Tang, G. J. Qi, C. C. Aggarwal, and T. S. Huang. 2015. Heterogeneous Network Embedding via Deep Architectures. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), 119–128.

P. Comon. 1994. Independent component analysis, a new concept? *Signal Processing* 36, 3 (1994), 287–314.

C. Cortes and V. Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

T. M. Cover and P. E. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13(1) (1967), 21–27.

T. M. Cover and J. A. Thomas. 1991. *Elements of information theory*. John Wiley & Sons.

X.F. He, D. Cai, S.C. Yan, and H.J. Zhang. 2005a. Neighborhood preserving embedding. *IEEE International Conference on Computer Vision* 2 (2005), 1208–1213.

X.F. He, D. Cai, S.C. Yan, and H. J. Zhang. 2005b. Neighborhood Preserving Embedding. *IEEE International Conference on Computer Vision* 2, 23 (2005), 1208–1213.

H. Hoffmann. 2007. Kernel PCA for Novelty Detection. *Pattern Recognition* 40(3) (2007), 863–874.

Thomas Hofmann, Bernhard Schlkopf, and Alexander J. Smola. 2008. Kernel Methods in Machine Learning. *Annals of Statistics* 36, 3 (2008), 1171–1220.

J. L. Hu, J. W. Lu, and Y. P. Tan. 2014. Discriminative deep metric learning for face verification in the wild. *IEEE International Conference on Computer Vision and Pattern Recognition* (2014), 1875–1882.

J Huang, PC Yuen, WS Chen, and JH Lai. 2004. Kernel subspace LDA with optimized kernel parameters on face recognition. *IEEE International Conference on Automatic Face and Gesture Recognition* (2004), 327–332.

I.T. Jolliffe. 1986. Principal component analysis and factor analysis. In *Principal component analysis*. Springer, 115–128.

A. Krizhevsky, L. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems* (2012), 1097–1105.

J.A. Lee and M. Verleysen. 2007. *Nonlinear dimensionality reduction*. Springer Science & Business Media.

J. Li, H.F. Liu, H.D. Zhao, and Y. Fu. 2017. Projective Low-rank Subspace Clustering via Learning Deep Encoder. *International Joint Conference on Artificial Intelligence* (2017), 2145–2151.

M. Luo, F. Nie, X. Chang, Y. Yang, A. Hauptmann, and Q. Zheng. 2016. Avoiding Optimal Mean Robust PCA/2DPCA with Non-Greedy ℓ_1-Norm Maximization. *International Joint Conference on Artificial Intelligence* (2016), 1802–1808.

Y. J. Ma, C. Shang, F. Yang, and D. X. Huang. 2014. Latent subspace clustering based on deep neural networks. *International Symposium on Advanced Control of Industrial Processes* (2014), 502–507.

P. Magdalinos, C. Doulkeridis, and M. Vazirgiannis. 2011. Enhancing clustering quality through landmark-based dimensionality reduction. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5, 2 (2011), 1–44.

B. Mikhail and N. Partha. 2009. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Processing Systems* 14(6) (2009), 585–591.

R. Ohbuchi, K. Osada, T. Furuya, and T. Banno. 2008. Salient local visual features for shape-based 3D model retrieval. *IEEE International Conference on Shape Modeling and Applications* (2008), 93–102.

X. Peng, J.S. Feng, J.W. Lu, W.Y. Yau, and Z. Yi. 2017. Cascade Subspace Clustering. *The Association for the Advance of Artificial Intelligence* (2017), 2478–2484.

X. Peng, S.J. Xiao, J.S. Feng, W.Y. Yau, and Z. Yi. 2016. Deep Subspace Clustering with Sparsity Prior. *International Joint Conference on Artificial Intelligence* (2016), 1925–1931.

S. Roweis and L. Saul. 2000. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290, 5500 (2000), 2323–2326.

J. Shi and J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 8 (2000), 888–905.

N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

L. Sun, B. Ceran, and J. Ye. 2010. A scalable two-stage approach for a class of dimensionality reduction techniques. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2010), 313–322.

J. Tenenbaum, V. Silva, and J. Langford. 2000. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290, 5500 (2000), 2319–2323.

F. Tian, Q. Cui B. Gao, E. H. Chen, and T. Y.Liu. 2014. Learning deep representations for graph clustering. *The Association for the Advance of Artificial Intelligence* (2014), 1293–1299.

M. W. Trosset and C. E. Priebe. 2008. The out-of-sample problem for classical multidimensional scaling. *Computational Statistics and Data Analysis* 52, 10 (2008), 4635–4642.

L. Van der Maaten, E. Postma, and J. Van den Herik. 2009. Dimensionality reduction: a comparative. *Journal of Machine Learning Research* 10 (2009), 66–71.

L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus. 2013. Regularization of neural networks using dropconnect. *International Conference on Machine Learning* (2013), 1058–1066.

D.X. Wang, P. Cui, and W.W.Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1225–1234.

Q. Wang and Q. Gao. 2016. Robust 2DPCA and Its Application. *IEEE International Conference on Computer Vision and Pattern Recognition Workshops* (2016), 1152–1158.

Z. W. Wang, D. Liu, J. C. Yang, W. Han, and T. Huang. 2015. Deep networks for image super-resolution with sparse prior. *IEEE International Conference on Computer Vision* (2015), 370–378.

L. Xie, D. Tao, and H. Wei. 2016. Multi-View Exclusive Unsupervised Dimension Reduction for Video-Based Facial Expression Recognition. *International Joint Conference on Artificial Intelligence* (2016), 2217–2223.

MINGYAN YANG
北京市朝阳区平乐园
100号北京工业大学
机电学院
证件号:
110108199702046023
Beijing, China,
Beijing 100124
中国
yangmingyan9724@126.com
(86) 1067396333
13693052017
NEEA ID: **3205129**
ETS ID: 10109520
余额: RMB¥ 42
退回这部分钱

查看成绩

Registration Number: 0000000031356644

Test	Test Date	Reading	Listening	Speaking	Writing	Total
TOEFL iBT	Sun Oct 29 08:50:54 EDT 2017	25	26	22	27	100

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.

ETS GRE

TEST TAKER SCORE REPORT

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

MINGYAN YANG

Address: BEIJING UNIVERSITY OF TECHNOLOGY, NO.100 PINGLEYUAN, CHAOYANG DISTRICT, Beijing, 100124 China

Email: yangmingyan9724@126.com
Phone: 86-13693052017
Date of Birth: February 4, 1997
Social Security Number (Last Four Digits):
Gender: Female
Intended Graduate Major: Computer Science (0402)

Most Recent Test Date: May 13, 2018

Registration Number: 3333132
Print Date: July 26, 2018

Your Scores for the General Test Taken on May 13, 2018

Verbal Reasoning	Quantitative Reasoning	Analytical Writing
Your Scaled Score: 156	Your Scaled Score: 170	Your Score: 4.5
130 – 170	130 – 170	0 – 6
73rd Percentile	96th Percentile	82nd Percentile

Your Test Score History

General Test Scores

	Verbal Reasoning		Quantitative Reasoning		Analytical Writing	
Test Date	Scaled Score	Percentile	Scaled Score	Percentile	Score	Percentile
May 13, 2018	156	73	170	96	4.5	82
April 14, 2018	143	20	168	94	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

Designated Score Recipient(s)

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

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name: **Mingyan Yang**

Program: **Computer Science: MS**

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

Recommender Name: **Jie Lin**

Organization Name: **Xi an Jiaotong University**

Title: **Associate Professor**

E-mail Address: **jielin@mail.xjtu.edu.cn**

Telephone Number: **8613519183355**

Relationship to Applicant: **course instructor**

Certification (Date): **10-23-2018**

*"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."

西安交通大學 电信学院
SCHOOL OF ELECTRONIC AND INFORMATION ENGINEERING
XI'AN JIAOTONG UNIVERSITY
XI'AN 710049, P.R. CHINA

Recommendation Letter for Mingyan Yang

Date: September 7, 2018

To whom it may concern,

I am an Associate Professor of Computer Science at the Xian Jiaotong University. I am writing this letter to recommend Mingyan Yang strongly for admission to the graduate program at your University. I believe Mingyan's learning capability, research skills, along with her personality, would make her an ideal candidate for this program.

As the teacher of the course, named computer graphics, I have known Mingyan from beginning of her junior year, when she joined my class. During the course, I also supervised her to carry out the experiment on implementing many different drawing algorithms. Mingyan performed excellently in my class. Although this course was considered to be difficult for most students, Mingyan distinguished herself by excellent academic standing and learning skills, which won her the opportunity to gain a relatively high score of 86/100 in the final exam. With this experience in mind, I believe that Mingyan would be able to quickly adapt to the studies in your university, just like what she did at Xi'an Jiaotong University and my class.

Not only did she perform well in class, she also accomplished assignments I left them after class very well. Her well-finished assignments always went beyond the required criteria, writing neatly and conceiving of original and unique viewpoints. Based on assignments and experiments of my course, Mingyan demonstrated ability in self-learning and great cooperation with other students. Mingyan was easily able to apply what she learned to solve the problem that she faces. Her prominent characters such as strong self-learning capacity, persistent spirit of exploring knowledge and great team spirit fully revealed her ability in conducting studies and researches.

Additionally, Mingyan is also a good planner and well-organized individual. When an experiment is assigned for her, she never required me to urge her to do the experiment work, but always completed the task in advance. I believe the good time management skills and strong ability to work efficiently will enable her to continuously perform well in future graduate study and research.

Also, Mingyan reaches out to me very frequently. Through these conversations, I can see that she cherishes great dream and has always been craving for state-of-the-art knowledge in the field of computer science. In addition, her English ability and programming ability should be more than sufficient to support her following study and research.

With the acquaintance with Mingyan Yang, I believe she is a promising student and an ideal candidate for this program. Therefore, I am here wholeheartedly recommending her in this letter. If you need further information concerning this letter or have other questions, please feel free to contact me at jielin@mail.xjtu.edu.cn.

Your favorable consideration will be highly appreciated.

Sincerely yours.
Dr. Lin Jie Jie Lin

Associate Professor of School of Electronic and Information Engineering, Xi'an Jiaotong University
jielin@mail.xjtu.edu.cn / 86-13519183355
No.28, Xianning West Road, Xi'an, Shaanxi, 710049, P.R. China

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name: **Mingyan Yang**

Program: **Computer Science: MS**

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

Recommender Name: **Yuanqi Su**

Organization Name: **Xi an Jiaotong University**

Title: **Lecturer**

E-mail Address: **yuanqisu@mail.xjtu.edu.cn**

Telephone Number: **13700230551**

Relationship to Applicant: **course instructor and research supervisor**

Certification (Date): **10-24-2018**

*"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."

Recommendation Letter for Mingyan Yang

Yuanqi Su Yuanqi Su.
Lecturer of Department of Computer Science and Technology, Xi'an Jiaotong University
yuanqisu@mail.xjtu.edu.cn / 86-13700230551
No.28, Xianning West Road, Xi'an, Shaanxi, 710049, P.R.China

Dear Sir or Madam,

This letter serves for Miss Mingyan Yang's Admission into her intended master program in the United States. Being her class instructor and research project advisor since this April, I have frequently contacted with Mingyan and obtained in-depth understanding about this promising lady. Therefore, I provide this recommendation letter to support your further evaluation on her.

Since April 2018, I taught Mingyan the *Computer Vision & Pattern Recognition* for one semester. During our regular communication, I noticed she was highly motivated to the area of computer vision. Never missing any single class, Mingyan held the positive learning attitude towards every task I assigned. Usually, after each class, I would ask my students to do one programming exercise and another one in written calculation, both of which were challenging. Though there were only three days from these exercises, plus so much other homework to do, Mingyan always turned in her work on time and with high quality, demonstrating she spent extra time after class to deal with every problem. What's more, due to her great interest in computer vision, she emailed me several times to present her willingness to join my research team for self-development. The one-semester contact was not that long, however, I believed I saw Mingyan's passion and perseverance on being a researcher.

Impressed by what she achieved on my course, I was glad to ask Mingyan to join my project of *Generation of SMPL 3D Human Body Model Attitude and Recognition in Real Scene.* Not letting me down again, Mingyan clearly knew how to be a qualified researcher. Each week, she reported her progress, asked questions and discussed with me about next-stage task. Even during the summer vacation, she still emailed me once a week to provide her own thoughts towards current situations. What's more, facing with difficulties, she never gave up. Instead, spending lots of time in the library, she searched lots of academic references to seek for answers. From the very beginning until the end, Mingyan, with her behaviors and attitudes, constantly reminded me, she was the brave and determined person who could defeat problems and reach the final destination.

In our regular discussions, Mingyan clarified to me several times that computer vision was her dream research area for that it was more and more widely used in our life. Moreover, as a movie lover, Mingyan was excited that computer vision was closely related to movie effects. It was so fantastic that she discovered the combination between hobbies and her majors, which made me more certain that she could achieve more.

Based on what I have observed, I am certainly confident Mingyan is fully prepared for the graduate study. Her exceptional study skill, academic enthusiasm and determination to overcome any difficulty could definitely help her go further and steadily.

Northwestern | THE GRADUATE SCHOOL

Recommendation Form

The Graduate School Northwestern University Evanston, IL 60208-1113

Applicant Name: **Mingyan Yang**

Program: **Computer Science: MS**

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

Recommender Name: **Junbin Gao**

Organization Name: **The University of Sydney**

Title: **Professor**

E-mail Address: **junbin.gao@sydney.edu.au**

Telephone Number: **61286274856**

Relationship to Applicant: **research advisor**

Certification (Date): **10-25-2018**

*"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."

Professor Junbin Gao

Discipline of Business Analytics
Business School

The University of Sydney
NSW 2006, Australia

Tel: +61 2 8627 4856
Email: junbin.gao@sydney.edu.au
http://www.sydney.edu.au

13 September 2018

Dear Admission Officer,

Recommendation Letter for Ms Mingyan Yang

I am glad to recommend to you Ms Mingyan YANG, who is a student at Xi'an Jiaotong University. As Professor of Big Data Analytics at the University of Sydney Business School, I was invited by Beijing University of Technology as the High-level Overseas Talent, working periodically in Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology and supervising students in carrying out research projects. I got to know Mingyan in January 2018 when she started her research project *Locality Preserving Projection via Deep Neural Network* in our lab and became more familiar with her as she spent her summer vacation continuing conducting and completing this project under my supervision. Through close contact with her, I am highly impressed by her great passion and vast potential in research and strongly believe that she will be a valuable member of your esteemed program.

Mingyan is an inquisitive researcher with strong curiosity and desire of exploring new knowledge and she has demonstrated strong research ability and vast research potential in the research. She was good at using resources in the Internet to search and find the required information and materials. When faced notions and models that she hasn't known before or she couldn't fully understand, she would find related materials quickly through the Internet. In a number of occasions, I observed that she exchanged ideas and discussed them with the graduate students in our group. She was modest to learn and never felt ashamed to ask questions. If she couldn't figure them out by herself, she never put it aside but consulted me or other graduate students till she could fully comprehend them and interpret them clearly.

She is also a quick learner, with aspirations to master the new knowledge and technology. I noted she self-taught herself python, a programming language popular in AI fields. As a beginner, although she faces lots of problems when programming, she never gives it up until finding a way through and fixing bugs in programs. As a former IT professor I have good judgement over person's programming skill. From my observation I find her completed programs are in high standard with great readability and neatness. Her enquiring mind as well as outstanding self-learning ability is an accurate reflection of her great potential to succeed in the graduate program she is applying for. I am sure that she has already possessed the ability of becoming a qualified researcher.

Mingyan has good command of English and she excels at English literature review. To better conduct the research, she seized every opportunity to learn the model included in the research project through searching and reading related English papers and literatures and contributed a lot to the literature review part of our project. Although she didn't have any experiences in writing English papers before, she learned how to write an English academic paper and greatly improved her English writing ability through this project. I believe Mingyan's excellent English skills will give her a great advantage in quickly adapting the new language environment when pursuing graduate study in the US.

Furthermore, owing to her strong communication and cooperation abilities, she got along very well with other graduate students in our lab. On the professional side, she always asked them questions about difficult theoretical derivation, had heated discussions with them and exchanged their ideas about the research. On the life side, they often gathered together to talk about their plan upon graduation.

Based on my appreciation of such a promising young lady who seeks for her goal bravely, I strongly support her decision of higher academic pursuit without reservation. And I sincerely hope my recommendation will receive your favourable consideration. Feel free to contact me if I am of any further help.

Truly yours,

JBgao

Digitally signed by Junbin Gao
DN: cn=Junbin Gao, o=The University of Sydney, ou, email=junbin.gao@sydney.edu.au, c=AU
Date: 2018.09.14 08:07:36 +10'00'

Junbin Gao
Professor of Big Data Analytics