Fei Xia

Room 260, Packard Building, 350 Serra Mall, Stanford, CA 94305-9515 ⋄ feixia@stanford.edu

EDUCATION

Stanford University, Stanford, CA, USA *2016.9 - Present*
PhD Student, Department of Electrical Engineering,
Current Coursework: Machine Learning, Deep Learning in Genomics and Biomedicine

Tsinghua University, Beijing, China *2012.8 - 2016.7*
Bachelor of Engineering(*Anticipated*) Department of Automation
Cumulative GPA: **94.1/100**, Class Rank: 1/144 in the Department of Automation

Stanford University, Stanford, CA, USA *2015.7 - 2015.9*
Undergraduate Visiting Research Assistant in the Department of Electrical Engineering
The Chinese Undergraduate Visiting Research (UGVR) Program, with only 18 students selected from China

Georgia Institute of Technology, Atlanta, GA, USA *2014.8 - 2014.12*
Exchange Student in the School of Electrical and Computer Engineering
GPA: **4.0/4.0**

Courses related to my research interests:
Tsinghua University: Data Structure and Algorithms: (98), Computer Language and Programming: (92), C++ Program Design and Training: (98), Interdisciplinary Research Training (in Bioinformatics): (92). Calculus A1: (95), Calculus A2: (99), Linear Algebra 1: (93), Linear Algebra 2: (95), Probability and Statistics: (100), Operation Research: (93);
Georgia Institute of Technology: Stochastic Processes: (4.0/4.0), Signals and System Analysis: (4.0/4.0), Digital Signal Processing: (4.0/4.0), Computer Vision: (4.0/4.0).

RESEARCH EXPERIENCES

Stanford University, Stanford, CA, USA *2015.7 - Present*
Information Systems Laboratory, Department of Electrical Engineering
Research Assistant, Advisor: **Professor David Tse**

Project 1: *De novo* **DNA Sequence Assembly from Barcoded Reads**
- Established the information-theoretic bounds for a third generation sequencing technology, 10X. Discovered that closely spaced interleaved repeats are the main bottleneck for this read model.
- Designed algorithms to take advantage of barcoded linked reads in order to generate better assembly than what is currently available.
- Experimented on Human Chromosome 21, and boosted N50 of state-of-the-art assembler by 30%.

Ongoing project: A *de novo* **Sparse String Graph Assembler for PacBio Reads**
- Generated **finished** assembly at accuracy 99.9% for *E.Coli* based on sparse string graph methods, with details in publication [1][2].
- Extended NSG(Not-So-Greedy) algorithm to a regime when triple repeats are all-bridged and interleaved repeats are bridged, i.e. information-theoretic bound for perfect assembly.

Megvii Inc., Beijing, China *2016.3-2016.7*
DTR(Detection, Tracking, Re-identification) Group
Research Intern, Mentor: **Chi Zhang, Chief Scientist**

Project 1: Pedestrain Parsing Models
- Built an deep convolutional neural network model based on Holistically-Nested Edge Detection model and adapted it for pedestrian parsing.

Project 2: Pixel Level Domain Transfer for Pedestrian Re-identification

- Built a generative adversial network that transfer from pedestrian domain to upper-cloth domain, and used that model for pedestrian re-identification.
- Both models were incoporated into company's API for downstream applications.

Georgia Insitute of Technology, Atlanta, GA, USA *2014.8 - 2014.12*
Sun Lab, School of Computational Science & Engineering, College of Computing
Research Assistant, Advisor: **Professor Jimeng Sun**

Project 1: Epilepsy Seizure Prediction Based on EEG Data
- Built an analytic model for epilepsy seizure prediction based on EEG data. Developed methods for feature extraction from EEG data and dealing with imbalanced datasets.
- Participated in the Kaggle Competition, achieved AUC of 0.7298, and ranked in the top 8% (out of 504 teams)

Project 2: Cost Estimation for Cloud-Based Analytic Machine Learning Pipeline
- Proposed models and conducted experiments in order to do estimation for running time, cost and progress of cloud-based analytical pipeline.

Tsinghua University, Beijing, China *2014.2 - 2015.2*
Knowledge Engineering Group, Department of Computer Science and Technology
Research Assistant, Advisor: **Professor Jie Tang**

Project: Continuous Time Information Network Mining for Diffusion Cascades
- Designed models that consider indirect influence and structural influence for continuous-time information diffusion in networks
- Proposed gradient descent methods for learning models.
- Experimented on Sina Weibo dataset and increased AUC by 0.3 compared to baseline algorithms.

Tsinghua University, Beijing, China *2013.1 - 2014.7*
Tinker@Home Group, Texas Instruments-Tsinghua Future Robots Club
Team Leader

- Worked on developing a robust face detection and recognition system for robots.
- Developed an ROS package for face detection, alignment, archive building, and recognition.
- Participated in RoboCup 2014 in João Pessoa, Brazil, won 10$^{\text{th}}$ place in @Home League.

PUBLICATIONS AND MANUSCRIPTS

[1] G. Kamath*, I. Shomorony*, **F. Xia***, T. Courtade, D. Tse. " HINGE: Long-Read Assembly Achieves Optimal Repeat Resolution." *bioRxiv, 2016.* (*equal contributions)

[2] I. Shomorony, G. Kamath, **F. Xia**, T. Courtade and D. Tse, "Partial DNA Assembly: A Rate-Distortion Perspective." *ISIT 2016.*

[3] **F. Xia**, *et al.* "Human-aware mobile robot exploration and motion planner." *Proceeding of IEEE SouthEastCon 2015.*

AWARDS

2016 Stanford Graduate Fellowship, Stanford University
2015 Chang Jiong Scholarship (Highest honor in Dept. of Automation, 1/560)
2014 Fang Chongzhi Scholarship (Highest honor in Dept. of Automation, 1/560)
2014 China Scholarship Council Excellent Undergraduate Fellowship
2014 Tsinghua Sparks Program Fellowship (Highest Academic Honor, 50/3000)
2014 Bronze Prize, International Genetically Engineered Machine Competition
2013 National Southwest Associated University Scholarship (1/560)
2012 Tsinghua University Outstanding Freshman Scholarship
2011 Gold Medal in 25$^{\text{th}}$ National Chemistry Olympics Contest (Ranking 8/92000)

TECHNICAL STRENGTHS

Programming Languages	Proficient with C/C++, Python, MATLAB, Java
Research Skills	Familiar with state-of-the-art machine learning, statistics, neuroscience signal processing and NGS data analytics.
Additional Skills	ROS(Robot Operating System), vim, git, cmake, gcc, LaTeX, bash, MPI, OpenMP

LANGUAGE SKILLS

English Excellent listening, speaking, reading and writing abilities

- TOEFL iBT 109/120 (Reading 30, Listening 29, Speaking 24, Writing 26)
- GRE Verbal 155/170, Quantitative 170/170, Analytical Writing 4.0/6.0