# Yilin Lyu

2850 Broadway 4N, NEW YORK, NY, 10025

(+1) 929-386-8160 | yl3832@columbia.edu | github.com/yl3832 | linkedin.com/in/yilin-lyu/

## Education

**Columbia University** — *New York, NY*

M.A. in Statistics | GPA: 3.9/4.0 | Data Science Track | TA of Graduate level Deep Learning Course — *Sep. 2017 - Dec. 2018*

- Courses: Neural Networks and Deep Learning Research, Data Science For Industry, Statistical Machine Learning, Reinforcement Learning
- TA of ECBM E4040 Neural Networks and Deep Learning: design and teach Machine Learning basics and environment setup recitations

**Central University of Finance and Economics (CUFE)** — *Beijing*

B.S in Statistics | Major GPA: 3.67/4.00 — *Sep. 2013 - Jun. 2017*

- Courses: Regression Analysis, Multivariate Statistics, Bayesian Statistics, Statistical Computing, Data Management, Data Mining
- Research: Research Assistant and Survey Inquirer at CUFE Economics Lab | Summer Research Program at University of Michigan (ICPSR)
- Overseas: Exchange Student in Management Science at City University of Hong Kong | GPA 3.50/4.00

## Professional Experience

**Wayfair** — *Boston, MA*

Data Scientist | Pricing Optimization - Competitive Pricing — *May. 2019 - Present*

- Data scientist at the Algo group, I collaborated with a team of fellow data scientists and engineers to develop an algorithmic pricing engine that optimize prices on 70% of Wayfair revenue ($4.7B)

**Memorial Sloan Kettering Cancer Center (MSKCC)** — *Montvale, NJ*

Medical Image Processing Research Intern | Department of Medical Physics — *Jun. 2018 - Oct. 2018*

- Focused on a medical image (CT,PET and MRI) denoise and deblur project under the supervision of Yulin Song, PhD
- Re-implemented and modified DeblurGAN (2017) model architecture in Tensorflow with a creative new initialization phase for algorithm convergence; combined brand new ideas from the current academic literature such as Noise2Noise (Nvidia 2018) and CartoonGAN (CVPR 2018) to the base model and achieved state-of-art results
- Developed a new training phase and trained from scratch on a paired medical image dataset (14,329 pairs totally) created by ourselves using an open source dataset (TCIA) and images generated by our well-designed motion blur simulator
- International conference papers submitted | Current acceptance: CISP-BMEI 2018 | Conference presentation: Beijing, Oct.2018

**Microsoft Research Asia (MSRA)** — *Beijing*

Research Intern | Web Search and Data Management Group — *May. 2016 - Sept. 2016*

- Focused on the project of Digital ME, a digital AI helper dealing with repetitious work for diversified industries including business, healthcare organizations and education institutions, etc
- Integrated abundant unstructured text data from industries in R and Python and performed simplified datasets; enriched text information by ProBase and strongly facilitated conceptualization process which was used by other projects of short text understanding at Microsoft
- Assisted implementation of Deep Learning models and Machine Learning approaches,trained, tested and compared performance; best model with over 95% accuracy was used by the AI
- Selected as the recipient of Star of Tomorrow Award (Ranked Top5% among all interns) from Microsoft

## Skills

**Proficient:** Python, TensorFlow, PyTorch, Keras, R, LaTeX, SQL, SPSS **Basic:** Matlab, SAS, Tableau, Stata, Bloomberg

## Research and Project Experience

**Professor Zoran Kostic Deep Learning Lab : Voice Therapy Project** — *Columbia University*

Research Assistant — *May. 2018 - May.2019*

- Worked with and under the supervision of Prof.Zoran Kostic in the EE department; focused on a project to build automatic speech pathology detecting system from scratch based on Machine Learning and Deep Learning; conducted interdisciplinary cooperation with Columbia Medical Center and Columbia Nursing School
- Collected and preprocessed 10,000+ segments of pathological voice recordings and tested accuracy using machine learning algorithm (SVM Classifier, Bagging and Boosting) for initial results
- Designed and implemented multiple CNN (Convolutional Neural Network) models in the framework of Tensorflow; trained and tuned models to achieve a current 91% accuracy on four clinical stages

**Professor Zoran Kostic Deep Learning Lab : Smart City - Traffic Project** — *Columbia University*

Research Assistant — *Sept. 2018 - Dec.2018*

- Joined a smart city project on vehicles and pedestrians detection and trajectory prediction at complex busy intersections at NYC
- Gathered, annotated and processed massive video frames shot by multiple cameras at different intersections; applied Mask R-CNN (ICCV 2017) for object detection and combined it with pixel level Optical Flow (OpenCV) for tracking prediction

**Image Blur-Removal with Generative Adversarial Networks** — *Columbia University*

ECBM E6040 Course Project — *Apr. 2018 - May. 2018*

- Re-implemented the paper Deblur GAN: Blind Motion Deblurring Using Conditional Adversarial Networks in TensorFlow
- Proposed and implemented a novel training phase; achieved similar results as the method in original paper but in much shorter training time; results were greatly used as the foundations of following research

## Others

**Interests:** Cooking, Basketball (School Captain), Traditional Chinese Painting (National Silver Award)