

# Zihao Wang

zw1074@nyu.edu | (551)225-9955 | 204 10<sup>th</sup> St, Apt. 314, NJ 07302 | <https://zw1074.github.io>

## Educations

**New York University**, Courant Institute of Mathematics

*Master of Science in Data Science (3.80/4.0)*

**New York, NY**

*May 2017*

**Nanjing University**

*Bachelor of Science in Computational Mathematics (3.50/4.0)*

**Nanjing, China**

*Jun 2015*

## Related Courses

Machine Learning, Big Data, Deep Learning, Statistical Inference, NLP, Advanced Python, Fundamental Algorithm, Business Understanding for Data Science, Artificial Intelligence, Operating System, Optimization Theory, Linear Algebra, Probability Theory, Stochastic Processes.

## Skills

- Programming Skills: Python (Proficient), C/C++ (Proficient), SQL, Hadoop, AWS, Git, Theano, TensorFlow, Matlab
- Data Analytic Skills: Machine Learning, NLP, Deep Learning, Web Scrawling, Data Visualization
- Language: English, Chinese

## Professional Experience

**American International Group Inc.**

*Deep Learning Research Assistant, Intern*

**New York, NY**

*Aug 2016-Dec 2016*

- Contribute to the automatic car damage appraisal (ADA) project, especially for license plate detection on poor image dataset and heat map generation of damaged parts.
- Build an efficient license plate detector by using convolution neural network to generate saliency map and OpenCV to extract contour of license plate.
- Build an end-to-end Grad-CAM solution to generate heat map of damage part given a pre-trained model and image.
- Design a heat map generation API for efficiently using and easy updating.
- Implement these methods with Theano and Tensorflow and test them on both Linux and Windows system.
- The license plate detector achieves 15% top-5 error with 30ms/image speed.
- The map generation API has been merged into ADA project and proven to be more effective than previously used.

## Academic Projects

**Center for Data Science, NYU**

*Deep Learning: Efficient Auto-encoder for Physics Particle Collision Event*

**New York, NY**

*Sep 2016-Dec 2016*

- Build an auto-encoder for compressing 100GB CERN particle collision event.
- Evaluate model by calculating R2-score of reconstruction vector and MLP auto-encoder has 0.95 after adding threshold RELU.
- Apply well-trained auto-encoder to create an efficient anomaly detector.
- This work is supervised by Prof. Kyle Cranmer and presented in NIPS 2016 invited talk.

*NLP: Image Caption*

*Sep 2016-Dec 2016*

- Build an end-to-end attention-based image caption generator based on MS COCO dataset.
- Modify original model by calculating feature map attention instead of location attention.
- Gain less than 20 perplexity by using Google pre-trained GloVe word representation.

*Machine Learning: Duplication Detection*

*Feb 2016-May 2016*

- Build an end-to-end solution to detect duplication record in health care information system.
- Use 32-bit rolling hash to compute the representation value of string quickly.
- Create our own parallel filter algorithm to find possible duplication pairs which is much faster and more effective than local sensitive hashing (LSH) algorithm.
- Use T-SNE technique to visualize the feature vector and train a random forest model based on these vectors.

*Machine Learning: Yelp Restaurant Rating Prediction*

*Feb 2016-May 2016*

- Build a model to predict future rating level of restaurant based on business attributes, previous ratings and Yelp reviews.
- Divide the rating to 3 different levels based on the distribution to make a balance dataset.
- Use google pre-trained word2vec model to represent Yelp reviews as 300-dimension vector.
- Train a 0.86 micro-AUC score fine-tune logistic regression model by combining above feature.