Yao Qin

Email: yaoqindlut@gmail.com

Education

University of California, San Diego *2015.09 - 2020.01*
Doctor of Philosophy, Department of Computer Science and Engineering
Advisor: Prof. Garrison Cottrell

University of California, San Diego *2015.09 - 2017.12*
Master of Science, Department of Computer Science and Engineering
Advisor: Prof. Garrison Cottrell

Dalian University of Technology *2011.09 - 2015.06*
Bachelor of Science, Department of Electrical Engineering
Advisor: Prof. Huchuan Lu

Research Experience

Research Scientist, Google Brain, New York, USA *2020.01 - present*

Research Assistant, UC San Diego, USA *2015.09 - 2020.01*
Advised by Prof. Garrison Cottrell

Research Intern, Google Brain, Toronto, Canada *2019.04 - 2019.10*
Advised by Geoffrey Hinton, Colin Raffel and Nicholas Frosst

Research Intern, Google Brain, California, USA *2018.10 - 2019.01*
Advised by Ian Goodfellow, Colin Raffel and Nicholas Carlini

Research Intern, Google Brain, California, USA *2018.07 - 2019.10*
Advised by Suharsh Sivakumar and Raghu Krishnamoorthi

Research Intern, Microsoft Research, Cambridge, UK *2017.06 - 2017.09*
Advised by Antonio Criminisi and Aditya Nori

Research Intern, NEC Lab, New Jewsey, USA *2016.06 - 2016.09*
Advised by Haifeng Chen and Dongjin Song

Research Assistant, Dalian University of Technology, China *2014.03 - 2015.06*
Advised by Prof. Huchuan Lu

Publications (**Note**: * below denotes equal contribution) Google Scholar

Preprints

3. A. Balashankar, X. Wang, **Y. Qin**, N. Thain, B. Packer, E. Chi, A. Beutel. Improving Robustness through Pairwise Generative Counterfactual Data Augmentation. *Under Review* (**EMNLP**), 2022.
2. **Y. Qin**, N. Frosst, C. Raffel, G. Cottrell and G. Hinton. Deflecting Adversarial Attacks. *Preprints*, 2019.
1. Ian Goodfellow, **Yao Qin**, David Berthelot. Evaluation Methodology for Attacks Against Confidence Thresholding Models. *Preprints*, 2018.

Conferences & Journals

14. J. Zhao, X. Wang, **Y. Qin**, J. Chen, K. Chang. Investigating Ensemble Methods for Model Robustness Improvement of Text Classifiers. *Findings of Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2022.

13. **Y. Qin**, C. Zhang, T. Chen, B. Lakshminarayanan, A. Beutel, X. Wang. Understanding and Improving Robustness of Vision Transformers through Patch-based Negative Augmentation. *Advances in Neural Information Processing Systems* (**NeurIPS**), 2022.

12. J. Gu, V. Tresp, **Y. Qin**. Are Vision Transformers Robust to Patch-wise Perturbations? *European Conference on Computer Vision* (**ECCV**), 2022.

11. **Y. Qin**, X. Wang, A. Beutel, E. Chi. Improving Uncertainty Estimates through the Relationship with Adversarial Robustness. *Advances in Neural Information Processing Systems* (**NeurIPS**), 2021.

10. **Y. Qin**, X. Wang, B. Lakshminarayanan, E. Chi, A. Beutel. What are Effective Labels for Augmented Data? Improving Robustness with AutoLabel. *ICML Workshop on Uncertainty and Robustness in Deep Learning* (**ICML-UDL**), 2021.

9. T. Wang, X. Wang, **Y. Qin**, B. Packer, K. Li, J. Chen, A. Beutel, E. Chi. CAT-Gen: Improving Robustness in NLP Models via Controlled Adversarial Text Generation. *Conference on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020.

8. **Y. Qin***, N. Frosst*, S. Sabour, C. Raffel, G. Cottrell and G. Hinton. Detecting and Diagnosing Adversarial Examples with Class-Conditional Capsule Reconstructions. *International Conference on Learning Representations* (**ICLR**), 2020.

7. **Y. Qin**, N. Carlini, I. Goodfellow, G. Cottrell and C. Raffel. Imperceptible, Robust and Targeted Adversarial Example for Automatic Speech Recognition. *International Conference on Machine Learning* (**ICML**), 2019.

6. **Y. Qin**. Imperceptible Adversarial Example for Automatic Speech Recognition. *ACL Student Research Workshop* (**ACL-SRW**), 2019.

5. **Y. Qin**, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi and A. Nori. Autofocus Layer for Semantic Segmentation. *International Conference on Medical Image Computing & Computer Assisted Intervention* (**MICCAI**), 2018. (**Oral presentation**, 4% acceptance rate)

4. **Y. Qin***, M. Feng*, H. Lu and G. Cottrell. Hierarchical Cellular Automata for Visual Saliency. *International Journal of Computer Vision* (**IJCV**), 2017

3. **Y. Qin**, D. Song, H. Chen, W. Cheng, G. Jiang and G. Cottrell. A Dual- Stage Attention-Based Recurrent Neural Network for Time Series Prediction. *International Joint Conference on Artificial Intelligence* (**IJCAI**), 2017

2. Q. Pan, **Y. Qin**, Y. Xu, M. Tong and M. He. Opinion Evolution in Open Community. *International Journal of Modern Physics C, 1750003*, 2016.

1. **Y. Qin**, H. Lu, Y. Xu and H. Wang. Saliency Detection via Cellular Automata. In *Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2015

Patents

1. **Y. Qin**, X. Wang, B. Lakshminarayanan, E. Chi, A. Beutel. What are Effective Labels for Augmented data? Improving Robustness with AutoLabel.

2. D. Song, H. Chen, G. Jiang, **Y. Qin**. Dual Stage Attention based Recurrent Neural Network for Time Series Prediction.

Teaching & Mentoring

Teaching Assistant

1. CSE253: Neural Networks for Pattern Recognition (Winter 2019), UC San Diego
2. CSE190: Neural Networks and Deep Learning (Fall 2017), UC San Diego

Student Mentorship

- Zhouxing Shi (PhD at UCLA)
- Jieyu Zhao (PhD at UCLA → Incoming Assistant Prof. at USC)
- Ananth Balashankar (PhD at NYU → Research Scientist at Google)
- Jindong Gu (PhD at University of Munich → Postdoc at University of Oxford)
- Tianlu Wang (PhD at UVA → Research Scientist at FAIR)

Selected Awards

Rising Star in EECS	*MIT, 2021*
UCSD GSA Travel Grant	*UC San Diego, 2019*
MICCAI Travel Award	*MICCAI, 2018*
NIPS Women in Machine Learning Travel Award	*NIPS WiML, 2017, 2016*
Departmental Fellowship	*UC San Diego, 2015*
Outstanding Undergraduate Student Award	*Liaoning Province, China, 2015*
HIWIN Elite Scholarship (*top 15 students university-wide*)	*China, 2014*
Honorable Mention of Mathematical Contest in Modeling	*International, 2013*
National Scholarship	*China, 2013, 2012*

Selected Invited Talks

Leading a Breakout Session: Robustness of Machine Learning	@ *WiML Un-Workshop at ICML 2022*
Improving Calibration through the Relationship with Adversarial Robustness	@ *ITA Workshop, 2022*
Understanding and Improving Robustness of Machine Learning Models	@ *UCSB/CMU/USC/MPI, 2022*
What are Effective Labels for Augmented Data? Improving Robustness with AutoLabel	@ *UCSD, 2020*
Detecting, Diagnoising, Deflecting and Designing Adversarial Attacks	@ *Google/FAIR/Amazon/Apple, 2019*
Imperceptible, Robust and Targeted Adversarial Example for ASR	@ *Salesforce, 2019*

Professional Services

Fellowship & Proposal Reviewer

∗ (Reviewer) Google PhD Fellowship in North America and Europe *2021-2022*

∗ (Reviewer) Google Award for Inclusion Research Program (Faculty proposal) *2021*

Journal Reviewer

∗ (Reviewer) IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

∗ (Reviewer) Transaction of the International Society for Music Information (TISMIR))

Conference Reviewer/Area Chair

∗ (Reviewer) International Conference on Learning Representations (ICLR) *2018-2021*

∗ (Reviewer) Advances in Neural Information Processing Systems (NeurIPS) *2020-2021*

∗ (Program Committee) AAAI Conference on Artificial Intelligence (AAAI) *2018-2022*

∗ (Reviewer) Conference on Computer Vision and Pattern Recognition (CVPR) *2020-2022*

∗ (Reviewer) Internatial Conference on Computer Vision (ICCV) *2021*

∗ (Area Chair) Workshop for Women in Machine Learning (WiML) *2019-2021*