

# Siqi Wang

Boston | [siqiwang@bu.edu](mailto:siqiwang@bu.edu) | <https://cs-people.bu.edu/siqiwang/> | <https://github.com/SunnySiqi>

## EDUCATION

- Boston University** **Boston, MA**  
Ph.D. in Computer Science | Overall GPA: 3.95/4.0 Aug. 2020 - Present
- Brown University** **Providence, RI**  
M.S. in Computer Science | Overall GPA: A Aug. 2018 - May 2020  
*Coursework Highlights: Introduction to Computer Graphics, Probability for Computing and Data Analysis, Data Science*
- Beihang University (BUAA)** **Beijing, China**  
B.E. in Computer Science | Overall GPA: 3.7/4.0 Sep. 2014 - Jul. 2018  
*Coursework Highlights: Linear Algebra, Data Structure and Algorithms, Operating System, Machine Learning*
- Uppsala University** **Uppsala, Sweden**  
National Excellent Undergraduate Exchange Project Aug. 2016 - Jan. 2017  
*Coursework Highlights: Database Management, Software Engineering, Artificial Intelligence, Image Processing*

## TECHNICAL SKILLS

**Language:** Python, Bash, C/C++, Java, R, SQL, MATLAB

**Framework/ Tools:** PyTorch, TensorFlow, Scikit-learn, Lightning, Git

## RESEARCH EXPERIENCE

- Boston University AI Research Group (AIR) - Image and Video Computing Lab (IVC)** **Boston, MA**  
Research Assistant Dec. 2023 - Present
- **Learning Robustness to In-Domain Noise and Out-of-Domain Generalization** Dec. 2023 - Present  
Proposed a new task Robust2All to address simultaneous learning with in-domain noise and out-of-domain generalization. Created a comprehensive benchmark utilizing three real-world datasets and one synthesized noisy dataset. Evaluated a range of older and state-of-the-art (SOTA) methods from **Learning with Noisy Labels (LNL)** and **Domain Generalization (DG)**, as well as their combination, reveals unexpected outcomes. Identified challenges from unbalanced noise sources and domain-specific sensitivities, highlighting the limitations of traditional LNL sample selection strategies.
  - **Learning with Noisy Labels (LNL)** Aug. 2022 - Present
    - (i) Proposed a novel task known as Learning with Noisy Labels and noise source distribution Knowledge (**LNL+K**), assuming a certain understanding of potential sources of label noise. Explored several baseline LNL+K methods integrating noise source knowledge into state-of-the-art LNL methods across diverse datasets and noise types, yielding performance enhancements of **5-15%** compared to the unadapted methods.
    - (ii) Introduced an integrated framework named COMBO, comprising three key components: **noise modeling**, source knowledge identification, and **enhanced noise detection**. Methods within the COMBO framework exhibit improvements of up to **10%** in top-1 classification accuracy in synthesized noise datasets and **3-5%** in real-world noisy datasets compared to state-of-the-art techniques.
  - **Cell Representation Learning** Oct. 2021 - May 2022  
Worked on cell morphology classification using Painted Cell Datasets. Applied a **Mixture-of-Experts** approach in embedding learning, training a range of experts specialized in **addressing technical variations** arising from the microscopy image collection process. Utilized Treatment **Exemplars** to represent the dataset's distribution in training each minibatch, optimizing GPU usage. Achieved a remarkable performance enhancement of

5.5-11% in cell treatment classification compared to state-of-the-art methods.

- **Fashion Outfit Feature Compatibility Learning**

Jan. 2021 - Aug. 2022

Explored stylistic compatibility prediction for fashion items using the **Polyvore Dataset**, employing **contrastive learning**. Introduced **EVO (Embedding Variance Operator)** with **transformers** to measure compatibility gains from adding a new object to an existing outfit (a group of objects). Achieved a top performance of **63.1%** on the FITB (Fill-In-The-Blank) task, surpassing the state-of-the-art benchmarks.

## **Brown University Visual Computing Lab**

**Providence, RI**

- **Style-Aware 3D Interior Scene Completion**

Jan. 2019 - Jul. 2021

Proposed a **self-supervised** model that learns to recommend style compatible objects for given scenes. Designed a **conditional CNN** where the generation of object embeddings are conditioned by scenes and placed in a scene-specific embedding space. Designed and implemented an interactive tool to collect user evaluation.

## **INDUSTRY EXPERIENCE**

### **Etsy**

**Brooklyn, NY**

PhD Research Intern of International Search Team

May. 2024 - Aug. 2024

- **Multi-task Multi-distribution Search Embedding Learning**

Propose the SEQ+MD framework, which integrates sequential learning for **multi-task learning (MTL)** and feature-generated region-mask for **multi-distribution input**. This approach leverages the sequential order within tasks and accounts for regional heterogeneity, enhancing performance on multi-source data. Evaluations on in-house data showed a 2% increase on the high-value engagement including **add-to-cart and purchase** compared to state-of-the-art baseline models. Additionally, the multi-regional learning module is "plug-and-play" and can be easily adapted to enhance other MTL applications. This work has been submitted to a conference and is currently under review.

### **Wayfair**

**Boston, MA**

Data Scientist Intern of 3D Innovation Team, Vision and Speech Platform

May. 2019 - Aug. 2019

- **Style-Aware Scene Object Recommendation**

Initiated and played a leading role in the scene completion project. Analyzed 3D scene data in **3ds Max**. Combined scene data with other product information with **SQL**. Applied **triplet loss** for object embedding optimization and made visualizations with **D3**.

## **PUBLICATIONS**

- **Siqi Wang**, Audrey Zhijiao Chen, Austin Clapp, Sheng-Min Shih, Xiaoting Zhao, "SEQ+MD: Learning Multi-Task as a SEquence with Multi-Distribution Data", arXiv, 2024.
- **Siqi Wang**, Bryan A. Plummer, "LNL+K: Learning with Noisy Labels and Noise Source Distribution Knowledge", ECCV, 2024.
- **Siqi Wang**, Chau Pham, Bryan A. Plummer, "A Unified Framework for Connecting Noise Modeling to Boost Noise Detection", arXiv, 2023.
- Zitong Chen, Chau Pham, **Siqi Wang**, Michael Doron, Nikita Moshkov, Juan C. Caicedo, Bryan A. Plummer, "CHAMMI: A benchmark for channel-adaptive models in microscopy imaging", NeurIPS, 2023
- **Siqi Wang**, Manyuan Lu, Nikita Moshkov, Juan C. Caicedo, Bryan A. Plummer, "Anchoring to Exemplars for Training Mixture-of-Expert Cell Embeddings", arXiv, 2021.
- **Siqi Wang**, Tongyu Yue, Bo Lang, "An Efficient Adaptive Algorithm for Removal of Impulse Noises", CISP-BMEI, 2017.
- Yang Liu, Lei Huang, **Siqi Wang**, Xianglong Liu, Bo Lang, "Efficient Segmentation for Region- based Image Retrieval Using Edge Integrated Minimum Spanning Tree", IEEE-ICPR, 2016.

## TEACHING EXPERIENCE

### Boston University

- Artificial Intelligence, grader (CS640 Fall2022 Instructor: Iddo Drori)
- Image and Video Computing, grader (CS585 Spring2022 Instructor: Bryan Plummer)
- Machine Learning, Teaching Assistant(CS542 Fall2021 Instructors: Kate Saenko, Bryan Plummer)
- Deep Learning, Grader (CS523 Summer 2021 Instructor: Sarah Adel Bargal)
- Introduction to Algorithms, Teaching Assistant (CS330 Spring 2021 Instructors: Dora Erdos, Steve Homer)
- Introduction to Computer Science , Teaching Assistant

### Brown University

- Machine Learning , Teaching Assistant (CS1420 Spring 2019 Instructor: Stephen Bach)

## HONORS & AWARDS

Doreen Green UTAship for Women in CS	Feb. 2019
Scholarship of Academic Excellence, BUAA	2015, 2016, 2017, 2018
Second Prize in “Feng Ru Cup” Competition of Academic and Technological Works, BUAA	May 2017

## LEADERSHIP

President of International Department of Student Union in CS Department	Jul. 2015 - Jun. 2016
President of Volunteer Services in CS Department	Sep. 2014 - Sep. 2015