KEHAN QI

Mail: kehan.qi@stonybrook.edu

EDUCATION

Stony Brook University PhD Student
Biomedical Informatics *Aug 25 2025 - present*

- **Research Team**: Led by Professor Chao Chen. [homepage]
- **MR Image Reconstruction**: Propose a topology-based MR image reconstruction method and validate on open datasets.
- **OCTA Image Translation**: Propose a deep learning method for translating OCT volume to OCTA.

University of Chinese Academy of Sciences Graduate Study
Master of Engineering in Computer Technology *Sept 01 2018 - June 26 2021*

- **Research Team**: Led by Professor Shanshan Wang [homepage]
- **Brain Stroke Segmentation in MR Images**: Employ neural networks to segment brain stroke in MR images. Produced papers: 1 MICCAI as 1st author, 1 MICCAI as 3rd author, 1 IEEE Access as 3rd author.
- **MR Image Reconstruction and Segmentation**: Utilize a two-module neural network and re-weighted loss to segment and reconstruct MR images simultaneously. Produced papers: a pre-print paper as 1st author.
- **Reconstructed MR Image Quality Assessment**: Employ a neural network to assess MR image quality. Produced papers and patents: a pre-print paper as 1st author, a US patent as 3rd author.

Zhejiang University Undergraduate Student
Bachelor of Engineering in Measurement Control Technology and Instruments *Sept 01 2013 - July 15 2017*

WORK EXPERIENCE

Stori Full-time Employee, Hangzhou, China
Data Engineer *Apr 20 2023 - July 31 2024*

- **Low-latency ML Inference System for Risk Control**: Designed and deployed a real-time ML inference system for transaction-level risk control. Used AWS DMS + Flink + Kinesis + Lambda + SageMaker for cross-account model invocation with 1-second average latency. Optimized inference pipeline for throughput and latency.
- **Real-time Data Infrastructure**: Built real-time data pipelines supporting ML model invocation, data monitoring, and downstream query API integration using Flink, Kinesis, Lambda, DynamoDB, and Elasticsearch. Served as backend for real-time financial indicators and risk flag triggering.
- **Team Leadership and Standards**: Established internal coding and deployment standards, CI/CD pipeline, and AWS CDK infrastructure templates. Mentored two junior engineers and led weekly sprint planning and code reviews.

Amazon Full-time Employee, Beijing, China
Software Development Engineer *Aug 02 2021 - Feb 10 2023*

- **Applied ML System Engineering**: Designed and implemented an automated pipeline for weekly product classification updates using ML models deployed on AWS SageMaker. Integrated Lambda, SNS, S3, and DynamoDB to support scalable, production-level ML inference and ingestion with tens of millions of products.
- **Impact Analysis via Distributed Processing**: Built large-scale PySpark pipelines to evaluate financial impacts of updated classification models. Analyzed billions of records to compute fee deltas pre- and post-deployment across multiple dimensions (product, seller, category). Applied Spark job optimization (e.g., executor tuning, broadcast disabling, RDD reuse) to reduce runtime to within 20 minutes.
- **Future Fee Prediction System**: Developed inference-based fee projection system utilizing classification results. Performed batch processing on 1.5B+ records with AWS Glue and Redshift, and optimized TPS throttling to support SageMaker-based fee computation. Enabled daily updates within a 24-hour SLA.

Tencent Intern, Shenzhen, China
Research Intern *June 18 2020 - Sept 04 2020*

- **Main Responsibility**: Develop new methods for medical image processing.
- **Project registrated medical image quality analysis**: a) Detect landmarks from registrated CT images. b) Train a neural network to predict registrated image quality score, with lanmarks and registrated image as input. c) A Chinese patent produced.

SELECTED PROJECTS

LLM-based Automatic Review System (Stony Brook University) Ongoing
Developer, PhD Research Project *May 2025 – Present*

- **Objective**: Design and build a full-stack pipeline for automatic scientific manuscript review using LLM.
- **Pipeline**: Includes model selection, fine-tuning (LoRA, SFT), inference optimization, latency monitoring, SageMaker deployment, auto-retraining, and A/B testing.
- **Result**: To support paper review action and performance evaluation in realistic environment with custom UI.

Papers and Publications

- Lanting Yang, **Kehan Qi**, Peipei Zhang, Jiaxuan Cheng, Hera Soha, Yun Jun, Haochen Ci, Xianliang Zheng, Bo Wang, Yue Mei, Shihao Chen*, and Junjie Wang*. "Diagnosis of Forme Fruste Keratoconus Using Corvis ST Sequences with Digital Image Correlation and Machine Learning." Bioengineering 11.5 (2024): 429.
- Shanshan Wang, Hairong Zheng, **Kehan Qi**, Chuyu Rong, and Xin Liu. "Image data quality evaluation method and apparatus, terminal device, and readable storage medium." U.S. Patent Application No. 18/546,425.
- Dong Wei,**Kehan Qi**, Yuexiang Li, Jiawei Chen, Kai Ma, and Yefeng Zheng. "Image registration quality evaluation model training method, device and computer equipment", Chinese patent, Application No. CN202011308476.3. 2022.
- **Kehan Qi**, Haoran Li, Chuyu Rong, Yu Gong, Cheng Li, Hairong Zheng, and Shanshan Wang*. "Blind Image Quality Assessment for MRI with A Deep Three-dimensional content-adaptive Hyper-Network". arXiv preprint arXiv:2107.06888 (2021).
- **Kehan Qi**, Yu Gong, Xinfeng Liu, Xin Liu, Hairong Zheng, and Shanshan Wang*. "Multi-task MR Imaging with Iterative Teacher Forcing and Re-weighted Deep Learning". arXiv preprint arXiv:2011.13614 (2020).
- **Kehan Qi**, Hao Yang, Cheng Li, Zaiyi Liu, Meiyun Wang, Qiegen Liu, and Shanshan Wang*. "X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-range Dependencies". Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22. Springer International Publishing, 2019.
- Hao Yang, Weijian Huang, **Kehan Qi**, Cheng Li, Xinfeng Liu, Meiyun Wang, Hairong Zheng, and Shanshan Wang* "CLCI-Net: Cross-Level Fusion and Context Inference Networks for Lesion Segmentation of Chronic Stroke". Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22. Springer International Publishing, 2019.
- Xin Liu, Hao Yang, **Kehan Qi**, Pei Dong, Qiegen Liu, Xin Liu, Rongpin Wang*, and Shanshan Wang*. "MSDF-Net: Multi-scale deep fusion network for stroke lesion segmentation". IEEE Access 7 (2019): 178486-178495.

Skills

- **Data Processing Techniques**: Spark, Flink, Hive, MySQL, No-SQL
- **Amazon Web Service (AWS) Skills**: Glue, Elastic Map Reduce (EMR), Lambda Function, Message Queueing Service (SQS), Managed Service for Apache Flink, Application Programming Interface (API) Gateway, Virtual Private Cloud (VPC), Database Migration Service (DMS), Simple Storage Service (S3), SageMaker
- **Deep Learning Techniques**: SFT, LoRA, RLHF, RAG