

XIAOTIAN HAN

2808 CALDER AVE NE, APT 6146,
REDMOND, WA, 98052

919-638-8406
XIAOTIAN.SKY.HAN@GMAIL.COM

SUMMARY

Research Scientist with 6+ years professional experience in GenAI.

STRENGTHS AND SKILLS

- Tools: Pytorch, Azure (AzureML, COSMOS, Batch, etc), DeepSpeed, ONNX, ROS
- Languages: Python, C++, Javascript, Scala, Rust
- Expertise in problem-solving, model development and deployment.
- Passionate about new techniques, fast and continuous learner

PROFESSIONAL DEVELOPMENT

WORK EXPERIENCE

- Senior Research Scientist, ByteDance Inc, Bellevue, USA Aug. 2023 – Present
- Construct Multimodal LLM reasoning evaluation benchmark [InfiMM-Eval](#).
 - Built SOTA cross-attention style open source MLLM [InfiMM](#) series, from pretraining to instruction finetuning.
 - Filtering MATH/STEM/Code data from Common Crawl and parse into interleaved format for continual pretraining, then IFT with perceiver autoregressive architecture, achieve SOTA on MathVerse.

- Senior Applied Scientist, Microsoft Azure Cognitive Service, Redmond, USA Sept. 2022 – Aug. 2023
- Collaborated with Schwarz group to define auto-checkout problem, evaluation metric and deployment criterial, collected bootstrapping dataset, finetuned [Florence foundation model](#), deployed model, set up automatic feedback data collection and continuous model finetuning. Achieved 1st customer success with Florence foundation model.
 - Built [Florence foundation model customization service](#).

- Applied Scientist II, Microsoft Azure Cognitive Service, Redmond, USA May. 2020 – Sept. 2022
- Implemented common [Scene Graph Detection algorithms](#) for benchmarking, proposed a new architecture, and achieved SOTA on Visual Genome, Open Images and GQA VRD tasks. Extracted features for VinVL.
 - Built and maintained [Azure Kinect](#) Body Tracking model evaluation benchmark, compressed body tracking model and implemented cloud inference pipeline.
 - Deployed RGBD multi-camera system in lab environment, built RGBD2RGB automatic annotation system, arranged external actors for data collections. Host [multi-camera tracking workshop](#) in ICCV2021.
 - Led vendors collect and annotate large scale retail store products on shelf datasets as internal benchmarks for CPG vertical. Pretrained a dense object detector for universal product detection, achieved SOTA on SKU110k and Retail50k.
 - Built unified pipeline for synthetic data generation and model training for shelf product detection.

- Applied Scientist, Microsoft Bing Multimedia, Redmond, USA Jul. 2019 – May. 2020
- Trained and deployed Fashion and Home Furniture Object Detection Model for Bing Visual Search ([example](#)).
 - Implemented single object tracking prototype running on mobile devices in Bing app ([link](#)).
 - Optimized Bing image search recommendation to increase user engagement with sematic similar images. Pretrained image feature extraction model using contrastive loss with Bing index images.
 - Collaborated to develop MagGAN for changing fashion attributes. The model was used for Bing visual search image query reformulation. Defined metric and measurement set, set up UHRS with A/B testing for evaluation.

- Research Intern, Kwai, Seattle, USA Feb. 2019 – Jul. 2019
- Led the project of DouDiZhu Poker endgame. Implemented MiniMax Tree Search solver with caching to serve million users. Worked with production team to integrate feature into mobile game app ([link](#)). Built an online active learning system to help game designers generate poker endgames with real-time playing feedback.
 - Initiated DouDiZhu Reinforcement Learning research in Kwai. Implemented RL Poker simulator. Experimented with Counterfactual Regret Minimization, contributed to DouZero ([link](#)).

- Research Intern, Tencent, Shenzhen, China May. 2018 – Aug. 2018
- Implemented graph Effective Closeness algorithm in Angel-ML. ([link](#))
 - Implemented Markov Clustering Algorithm with Spark. Approximated 200M×200M matrix multiplication by Monte Carlo Random Walk with 40% speed up. Won Tencent Micro Innovation Award.
 - Implemented Multiplex Network Embedding on *Honor of Kings* social network for user friendship prediction.

PAPERS

- Han, Xiaotian, et al. " *InfiMM: Advancing Multimodal Understanding with an Open-Sourced Visual Language Model*. ([link](#))
- Han, Xiaotian, et al. " *Exploring the reasoning abilities of multimodal large language models (mllms): A comprehensive*

survey on emerging trends in multimodal reasoning. ([link](#))

- Han, Xiaotian, et al. "*InfiMM-Eval: Complex Open-Ended Reasoning Evaluation For Multi-Modal Large Language Models.*" ([link](#))
- Han, Xiaotian, et al. "*MMPTRACK: Large-scale densely annotated multi-camera multiple people tracking benchmark.*" Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023. ([link](#))
- Han, Xiaotian, et al. "*Image scene graph generation (sgg) benchmark.*" arXiv preprint arXiv:2107.12604 (2021). ([link](#))

EDUCATION

M.S. in Computer Engineer, Duke University

Aug. 2017 – Dec. 2018

B.S. in Computer Science, University of Science and Technology of China

Aug. 2013 – Jul. 2017

B.S. in Physics, University of Science and Technology of China

Aug. 2013 – Jul. 2017