

Tong XIAO

Research Scientist at Meta

I am passionate about developing **multi-modality foundation models** and applying them to help extend **human capabilities**, build **autonomous machines**, and ultimately, engineer **humanoids with general intelligence**. My expertise and unique experience encompass:

- Hands-on experiences of cutting-edge foundation models – Llama 3 Vision, diffusion models, alignment, RLHF.
- End-to-end product delivery utilizing computer vision (CV) and deep learning (DL) technologies.
- A decade-long commitment to large-scale CV and DL research and applications.
- Proficiency in the engineering aspects of DL – distributed training, scalable data I/O, efficiency optimization, etc.

Education

08/2013 - 09/2017	The Chinese University of Hong Kong Ph. D. in Electronic Engineering, advised by Prof. Xiaogang Wang
08/2009 - 08/2013	Tsinghua University, Beijing B. Eng. in Computer Science and Technology

Experience

10/2017 - Present	Research Scientist @ Meta Menlo Park / Burlingame, CA
	- Llama 3 . Co-led Llama 3 Vision post-training: SFT, DPO, Reward Modeling, and Rejection Sampling. Solved challenging issues: DPO loss goes to 0, how to keep the capabilities of the text-only model, deterministic data loading, etc.
	- Gen AI . Led the research-to-production of EmuEdit (https://emu-edit.metademolab.com/), a text-driven freeform image editing diffusion model.
	- VR Face Tracking . Co-led a v-team of 15 people. Shipped face tracking to Meta Quest Pro. Focused on system design, data collection & annotation, distributed ML model training, and evaluation. Touched on hardware co-design and ML model deployment.
08/2013 - 09/2017	Research Assistant @ MMLAB The Chinese University of Hong Kong
	- ILSVRC2016 champions . 1st place in ImageNet detection, video object tracking, and Places2 scene parsing competitions. I pre-trained the foundation CNN.
	- Online Instance Matching Loss . https://arxiv.org/abs/1604.01850 , one of the earliest papers introducing contrastive loss to visual representation learning. Affected later self-supervised learning works such as Memory Bank and MoCo.
05/2014 - 08/2014	Intern @ Institute of Deep Learning (IDL), Baidu Beijing
	Published a CVPR15 paper on training a CNN for image classification using noisy web data.
07/2012 - 03/2013	Intern @ Face++, Megvii Beijing
	Built the face detection and tracking systems for <i>Crows Coming</i> (https://youtu.be/Xlq47P97fpI), the first motion-based game on iPhone.

Professional Activities

Journal Reviews

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI): 2021, 2019
International Journal of Computer Vision (IJCV): 2018, 2017
IEEE Transactions on Cybernetics: 2018
IEEE Transactions on Multimedia (TMM): 2017
IEEE Transactions on Neural Networks and Learning Systems (TNNLS): 2017
IEEE Transactions on Image Processing (TIP): 2017
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT): 2017, 2016
IET Computer Vision: 2018
Pattern Recognition: 2017
Multimedia Tools and Applications (MTAP): 2017
Computer Vision and Image Understanding (CVIU): 2016

Conference Reviews

SIGGRAPH: 2023
NeurIPS: 2024, 2022
ICLR: 2024, 2022
CVPR: 2024, 2022, 2021, 2020, 2019 (**Outstanding Reviewer Award**), 2018, 2017
ICCV: 2021, 2019, 2017
ECCV: 2020, 2018, 2016 (**Outstanding Reviewer Award**)
VCIP: 2017

Awards

2015 - 2016	Champions of Multiple Tracks	ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
	Results (CUImage and CUVideo entries): https://goo.gl/NzFr1V and https://goo.gl/pz5syg	
	News (in Chinese): https://goo.gl/6nPvmm and https://goo.gl/wJqHaX	
2010 - 2012	Undergraduate Scholarship	Tsinghua University
	Awarded to top 10% students each year under overall evaluation	
05/2010	Champion	14th Tsinghua University Artificial Intelligence Contest
	Develop AI programs for a computer game and compete with other players	
	Website: https://ai.net9.org News (in Chinese): http://goo.gl/AJGBW2	
12/2008	First Prize, 3rd Place	National Olympiad in Informatics (Province)
	Competitors design and implement algorithms to solve problems	

Skills

Programming Languages

Proficient in C/C++, Python. Experienced with Java, JavaScript, Matlab.

Deep Learning Related

PyTorch, Caffe, CUDA, MPI, BLAS, distributed training, scalable data I/O

Open-ReID

<https://github.com/Cysu/open-reid>

An open source PyTorch library for person re-identification created by me

Publications

The llama 3 herd of models

Llama team, **Core Contributor**

arXiv preprint arXiv:2407.21783, 2024

Camera Reprojection for Faces

James Allan Booth, Elif Albuz, Peihong Guo, **Tong Xiao**

US Patent 11562535, 2023

System and Method for Rendering Three Dimensional Face Model based on Audio Stream and Image Data

Tong Xiao, Sidi Fu, Mengqian Liu, Peihong Guo, Shu Liang, Evgeny Zatepyakin

US Patent 11113859, 2021

System and Method for Applying an Expression to an Avatar

Elif Albuz, Melinda Ozel, **Tong Xiao**, Sidi Fu

US Patent 10970907, 2021

Geometric Correspondence Fields: Learned Differentiable Rendering for 3D Pose Refinement in the Wild

Alexander Grabner, Yaming Wang, Peizhao Zhang, Peihong Guo, **Tong Xiao**, Peter Vajda, Peter M Roth, Vincent Lepetit

European Conference on Computer Vision (ECCV), 2020

Person Re-identification with Deep Kronecker-Product Matching and Group-shuffling Random Walk

Yantao Shen, **Tong Xiao**, Shuai Yi, Dapeng Chen, Xiaogang Wang, Hongsheng Li

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019

Order-Aware Generative Modeling Using the 3D-Craft Dataset

Zhuoyuan Chen, Demi Guo, **Tong Xiao**, Saining Xie, Xinlei Chen, Haonan Yu, Jonathan Gray, Kavya Srinet, Haoqi Fan, Jerry Ma

IEEE International Conference on Computer Vision (ICCV), 2019

End-to-End Deep Kronecker-Product Matching for Person Re-Identification

Yantao Shen, **Tong Xiao**, Hongsheng Li, Shuai Yi, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Deep Group-Shuffling Random Walk for Person Re-Identification

Yantao Shen, Hongsheng Li, **Tong Xiao**, Shuai Yi, Dapeng Chen, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Video Person Re-Identification With Competitive Snippet-Similarity Aggregation and Co-Attentive Snippet Embedding

Dapeng Chen, Hongsheng Li, **Tong Xiao**, Shuai Yi, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Joint Detection and Identification Feature Learning for Person Search

Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **Spotlight**, 2017

Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification

Tong Xiao, Hongsheng Li, Wanli Ouyang, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

Identity-Aware Textual-Visual Matching with Latent Co-attention

Shuang Li, **Tong Xiao**, Hongsheng Li, Wei Yang, Xiaogang Wang

IEEE International Conference on Computer Vision (ICCV), 2017

Learning Deep Neural Networks for Vehicle Re-ID with Visual-spatio-temporal Path Proposals

Yantao Shen, **Tong Xiao**, Hongsheng Li, Shuai Yi, Xiaogang Wang

IEEE International Conference on Computer Vision (ICCV), 2017

Learning from Massive Noisy Labeled Data for Image Classification

Tong Xiao, Tian Xia, Yi Yang, Chang Huang, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

Person Search with Natural Language Description

Shuang Li, **Tong Xiao**, Hongsheng Li, Bolei Zhou, Dayu Yue, Xiaogang Wang

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

Object Detection in Videos with Tubelet Proposal Networks

Kai Kang, Hongsheng Li, **Tong Xiao**, Wanli Ouyang, Junjie Yan, Xihui Liu, Xiaogang Wang
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

Crafting GBD-Net for Object Detection

X Zeng, W Ouyang, J Yan, H Li, **T Xiao**, K Wang, Y Liu, Y Zhou, B Yang, Z Wang, H Zhou, X Wang
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017

T-CNN: Tubelets with convolutional neural networks for object detection from videos

K Kang, H Li, J Yan, X Zeng, B Yang, **T Xiao**, C Zhang, Z Wang, R Wang, X Wang, W Ouyang
arXiv preprint arXiv:1604.02532, 2016

Convolutional Neural Networks with Low-rank Regularization

Cheng Tai, **Tong Xiao**, Yi Zhang, Xiaogang Wang, Weinan E
International Conference on Learning Representations (ICLR), 2016

DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification

Wei Li, Rui Zhao, **Tong Xiao**, Xiaogang Wang
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014