Jiayi Chen

177 Yellowstone DR APT 308, Charlottesville, VA 22903, United States
jc4td@virginia.edu | (+1) 434-234-7342 | Google Scholar | Homepage

Education

University of Virginia — **United States**
Ph.D. in *Computer Science* (GPA 4.0/4.0) — 08/2018 – Present

Xi'an Jiaotong University — **China**
M.S. in *Control Science and Engineering* (Specialization: *Artificial Intelligence and Robotics*) — 09/2015 – 06/2018

Xi'an Jiaotong University — **China**
B.E. in *Automation Engineering* — 09/2011 – 06/2015
Special Class for the Gifted Young of China — 09/2009 – 06/2011

Technical Skills

Programming Languages: **Python** (8+ years), C++, SQL, Java, Matlab

Tools & Platforms: **PyTorch** (6+ years), **Tensorflow**, **Jax** (2+ years), Flax, Haiku, Gym, Robotic Operating System

Deep Neural Networks: **Large Language Models** (e.g. BERT/ViT/T5/CLIP/Llama/LayoutLM/GPTs), Graph Neural Networks, Deep Generative Models (e.g., VAE/GAN/Diffuser), Seq2Seq, HyperNetworks, etc.

Research Areas

Multimodal Machine Learning (Vision, Language, Audio, 3D Motion, Time Series), Multimodal **Large Language Models, Meta/Multitask** Learning, **Federated** Learning, **Graph** Mining, Model-efficient Transfer Learning (**Pruning/Distillation**)

Recent Publications (Since 2020)

Multimodal Generative AI / Multimodal LLMs:

[1] **Jiayi Chen**, Mia Shu, Shili Sheng, Aidong Zhang. "A Graphical Multimodal Codebook is All You Need: Toward Interpretable Large Language Models for Audio-visual Generation". *Preprint in Progress*.

[2] **Jiayi Chen**, Aidong Zhang. "On Disentanglement of Asymmetrical Knowledge Transfer for Modality-Task Agnostic Federated Learning". *Accepted* by **AAAI** 2024 (Oral). (main track)

[3] **Jiayi Chen**, Aidong Zhang. "On Hierarchical Disentanglement of Interactive Behaviors for Multimodal Spatiotemporal Data with Incompleteness". *Published* at **KDD** 2023 (research track).

[4] **Jiayi Chen**, Hanjun Dai, Bo Dai, Aidong Zhang, Wei Wei. "On Task-personalized Multimodal Few-shot Learning for Visually-rich Document Entity Retrieval". *Published* at Findings of **EMNLP** 2023. (long paper)

[5] Lijun Yu, Jin Miao, Xiaoyu Sun, **Jiayi Chen,** AG Hauptmann, Hanjun Dai, Wei Wei. "DocumentNet: Bridging the Data Gap in Document Pre-Training". *Published* at **EMNLP** 2023: Industry Track.

[6] Jiachen Yu, Li Jin, **Jiayi Chen,** Youzi Xiao, Zhiqiang Tian, and Xuguang Lan. "Deep Semantic Space guided Multi-scale Neural Style Transfer." *Published* at Multimedia Tools and Applications, 2022, 81(3): 3915-3938.

Multimodal Federated/Meta/Graph Learning:

[7] **Jiayi Chen**, Aidong Zhang. "Architecture-personalized Multimodal Collaborative Learning with Graph Generative HyperNetworks". *Under Review,* submitted December 2023.

[8] **Jiayi Chen**, Kishlay Jha, Aidong Zhang. "LifelongSkill: Toward Modality-varying Lifelong Learning with Diversity-aware Latent Knowledge Hypergraph". *Under Review,* submitted November 2023.

[9] **Jiayi Chen**, Aidong Zhang. "FedMSplit: Correlation-adaptive Federated Multitask Learning across Multimodal Split Networks". *Published* at **KDD** 2022. (research track)

[10] **Jiayi Chen**, Aidong Zhang. "Topological Transduction for Hybrid Few-shot Learning". *Published* at **WWW** 2022.

[11] **Jiayi Chen**, Aidong Zhang. "HetMAML: Task-heterogeneous Model-agnostic Meta-learning for Few-shot Learning across Modalities". *Published* at **CIKM** 2021. (full paper)

[12] **Jiayi Chen**, Aidong Zhang. "HGMF: Heterogeneous Graph-based Fusion for Multimodal Data with Incompleteness". *Published* at **KDD** 2020. (research track)

Work Experience

Google LLC **Mountain View, CA**
Software Engineer Intern (PhD) Summer 2023

- ***Mentors/Teams:*** Zhe Zhao (DeepMind), Shuo Yang (YouTube), Hussein Hazimeh (Google Research)
- ***Project***: "Topology-agnostic Knowledge Inheritance for Improved Knowledge Distillation." The task is to explore both *direct* and *indirect* knowledge from pre-trained Foundation Models for *recommendation* and *vision*-related downstream tasks. The objective is to harness parameter-Level knowledge of LLMs (e.g., pre-trained weights and their topological structure) to accelerate and improve knowledge distillation.
- ***Skills:*** Tensorflow, Jax, training using TPU, LLM for Recommendation, Distillation, Pruning
- ***Accomplishment:*** a research paper in progress

Google LLC **Mountain View, CA**
Software Engineer Intern (PhD) Summer 2022

- ***Mentors/Teams:*** Hanjun Dai (DeepMind), Bo Dai (DeepMind), Wei Wei (Cloud AI)
- ***Project***: "Efficient Few-shot Multimodal Information Extraction in Visually-rich Documents". Visually-rich documents consist of three modalities: *language*, *image*, and *layout structure* of contents. The task is to leverage a Multimodal Large Language Model and balance the general and task-specific knowledge to improve few-shot understanding of *new* document types and *user personal interested* concepts.
- ***Skills:*** Tensorflow, Jax, training using TPU, Few-shot learning, Personalization, Multimodal LLM
- ***Accomplishment:*** Successfully published **two** research papers at EMNLP 2023.

Selected Projects

*In addition to research projects, I've also worked on **NLP / CV / 3D Vision / Graphics / Robotics** related projects before PhD. Some of them are listed here.*

Toward Emotion Controllable Chatbot via Deep Generative Models **11/2018 – 03/2019**

- ***Skills***: Natural Language Generation, GAN, VAE, Reinforcement Learning
- Generate texts/language data whose semantics or attributes (e.g., emotional tones) can be controlled; **adversarially** trained the sequential data generator with Conditional GAN and VAE, while using **reinforcement learning (RL)** to promote the authenticity of the generated language.

Cognition Knowledge Guided 3D Scene Generation **03/2017 – 02/2018**

- ***Skills***: Computer Vision, Neural Style Transfer, 3D Non-photorealistic Rendering
- Studied human knowledge-guided **neural style transfer**, focusing on improving the illusion of space in generated images by simulating how artists harness skills to understand and reproduce a **3D scene** (e.g., geometric structures, lighting and shallow); also studied 3D non-photorealistic rendering based on the neural style transfer paradigm.
- Proposed an illumination-guided deep alignment method using CNN, Lighting Path Expression, and PatchMatch.
- Crafted a small-scale **2D-to-3D dataset**, including 3D objects rendered by multiple lighting conditions (using Maya); 2D images with lighting segments (using Photoshop and Matlab); and I draw artistic materials myself by CorelPainter.

Vision-guided Drone Autonomous Navigation & Search-and-rescue System **01/2016 – 12/2016**

- ***Skills***: Robotics, 3D Computer Vision, C++, Python, Robotic Operating System (ROS)
- Developed real-time **Object detection** and **3D Localization** modules, as a part of our Drone's autonomous navigation system. 3D poses of objects are computed from 2D images given the real-time camera postures using 3D Geometry.
- Developed an autonomous **Tracking-and-Landing** module for Drone's landing **safety** on a moving vehicle; developed a platform to manage multimodal sensors and commands.
- Won the **4th** place from 130+ in the **international** contest "2016 DJI Developer Challenge" held in New York, USA.

Efficient Human Action Recognition based on Video Compression Domain **12/2014 – 05/2015**

- ***Skills****:* OpenCV, FFMPEG, C++
- Learned a visual codebook of Motion Vectors (MV) from the raw compressed MPEG-4 video bitstreams.
- Leveraged MV codebook to fast detect Spatial-Temporal Interest Points, instead of learning from decoded image frames.

Selected Awards

- ACM KDD'23 Student Travel Award.
- **4th** place from World Top10 in the **International Contest** of "2016 DJI Drone Developer Challenge". (*Top 3%, Rank: 4/130*) Issued by UNDP (United Nations Development Programme), Ford Motor Company, DJI Technology Co., Ltd.
- **2nd** prize in 2016 China **National Contest** on "The 3rd Winner Cup on Smart-City Technology and Creative Design".
- **1st** prize in 2015 China **National Contest** on Traffic-Scene Image Understanding. (*Leadership*)
- Others: 2018 Outstanding Master's Thesis Award; 2016 Industrial Scholarship. (*Top 3%*)

Relevant Courses

Machine Learning and Computer Vision (A+), Natural Language Processing (A+), Reinforcement Learning (A+), Database Systems (A), Graph Mining (A+), Statistical Learning and Graphical Models (A), Computer Graphics (A+), Software Artifacts (A), Autonomous Mobile Robot (A+), Data Structures and Algorithms (A+), Computer Network (A), etc.

Additional

Peer Review Services: Reviewers / Area Chairs for NeurIPS, ICLR, AAAI, IJCAI, CIKM, WSDM, SDM, BIBM, etc.

Startup Experience: Joined a university **Startup** team (MachInsight) from 2015 to 2016, focusing on dealing with **(1) Computer Vision**-related projects for industry; **(2) Drone** (a type of Autonomous Aerial Robot) related software product driven by Computer Vision; and **(3) Augmented Reality** related product. E.g., at 2015 XJTU University Anniversary, we developed an augmented reality iOS app using Java, Swift, and Unity3D; in 2016, we developed a communicative module for Drone for connecting with Ground Vehicle to achieve enhanced visual tracking and safe landing.

Teaching Assistant: Program and Data Representation (Fall 2019, Spring 2020); Digital Image Processing (Fall 2015)