Chih-Yao Ma

PhD · Georgia Tech

(+1) 404-907-6833 | cyma@gatech.edu | https://chihyaoma.github.io/

| https://github.com/chihyaoma | https://www.linkedin.com/in/chih-yao-ma-9b5b3063/

Machine Learning · Deep Learning · Computer Vision · Visual Understanding

Education

Georgia Tech — *Atlanta, GA*

MS/PhD in Electrical and Computer Engineering — *Aug. 2014 - PRESENT*

NCTU (National Chiao Tung University) — *Taiwan*

BS/MS in Electrical and Computer Engineering — *Sept. 2006 - June 2011*

Research & Project

The Regretful Navigation Agent

Georgia Tech

PhD — *Sept. 2018 - Nov. 2018*

- Equipped a navigation agent with *Regret Module* to decide when to *rollback* or *forward* using PyTorch.
- Proposed a *Progress Marker* allows the agent to access the progress estimate on each navigable direction.
- Set a new state-of-the-art performance on the Vision-and-Language Navigation task (5% SR ↑ and 8% SPL ↑).

Adaptive Frame Selection for fast Video Understanding

Georgia Tech

PhD — *Sept. 2018 - Nov. 2018*

- Developed a data-efficient agent for adaptively selecting video frames for video classification PyTorch.
- Matches the performance of using all video frames with only ~8 frames per video on FCVID and ActivityNet.

Self-Monitoring Visual-Textual Co-Grounded Navigation Agent

Salesforce Research

Research Intern — *May 2018 - Sept. 2018*

- Introduced a self-monitoring agent consists of a visual-textual co-grounding module and progress monitor using PyTorch.
- Set a new state-of-the-art performance on the Vision-and-Language Navigation task (8% absolute success rate ↑).

Grounded Objects and Interactions for Video Captioning

NEC Labs

Research Intern — *Sept. 2017 - Dec. 2017*

- Dynamically and progressively discover higher-order object interactions as the basis for video captioning using PyTorch.
- Achieved state-of-the-art performance on large-scale video captioning dataset: ActivityNet Captions.

Higher-Order Object Interactions for Video Understanding

NEC Labs

Research Intern — *May 2017 - Sept. 2017*

- Proposed generic recurrent higher-order object interactions module for video understanding problems with PyTorch and MXNet.
- Achieved state-of-the-art performance on large-scale action recognition dataset: Kinetics.

Long-term Video Classification in YouTube-8M

Georgia Tech

PhD — *Jan. 2017 to May 2017*

- Implemented and adapted various RNNs and MANNs (LayerNorm, RHN, Hierarchical RNN, NAS, and DNC) in TensorFlow.
- Benchmarked accuracy and speed in modeling long-term video content.

Activity Recognition with RNN and Temporal-ConvNet

Georgia Tech

PhD — *May. 2016 to Mar. 2017*

- Proposed two networks to integrate spatiotemporal information: temporal segment RNN and Inception-style Temporal-ConvNet.
- Achieved state-of-the-art performance on UCF101 and HMDB51 using Torch.

Partially Occluded Object Tracking with RGB-D Cameras *Georgia Tech*

PhD *Nov. 2014 to Dec. 2016*

- Cooperated with Walmart and SoftWear in developing an over-head vision system for closed loop control in sewing industry.
- Developed a color histogram and frequency domain based approach to track multiple partially occluded objects using Kinect depth sensor network.

Learning-based Saliency Model with Depth Information *NCTU*

Research Assistant *Dec. 2012 to Aug. 2013*

- Established and released an eye-tracking database for 3D images.
- Utilized high, mid, low level and depth features to predict how human beings look at the contents of different images.
- Proposed an SVM based saliency model for 3D content which outperformed the state-of-the-art approaches on different datasets.

Work Experience

May 2018 - Aug. 2018	**Research Intern**, Salesforce Research (with Caiming Xiong and Richard Socher)	*Palo Alto, CA*
May 2017 - Dec. 2017	**Research Intern**, NEC Machine Learning Labs (with Asim Kadav)	*Princeton, NJ*
Aug. 2014 - PRESENT	**Graduate Research Assistant**, Georgia Tech	*Atlanta, GA*
Sept. 2012 - May 2014	**Research Assistant**, CommLab, NCTU	*Taiwan*

Honor & Award

2015	**High-Tech Talent Scholarship**, granted for 126,000 USD, Ministry of Science and Technology	*Taiwan*
2011	**Academic Achievement Award**, Rank #2, Institute of Electro-Optical Engineering, NCTU	*Taiwan*

Publication

- Chih-Yao Ma, Zuxuan Wu, Ghassan AlRegib, Caiming Xiong, and Zsolt Kira, "**The Regretful Agent: Heuristic-Aided Navigation through Progress Estimation**," *Under review for CVPR 2019.* [PDF]
- Zuxuan Wu, Caiming Xiong, Chih-Yao Ma, Richard Socher, and Larry Davis, "**AdaFrame: Adaptive Frame Selection for Fast Video Recognition**," *Under review for CVPR 2019.* [arXiv]
- Chih-Yao Ma, Jiasen Lu, Zuxuan Wu, Ghassan AlRegib, Zsolt Kira, Richard Socher, and Caiming Xiong, "**Self-Monitoring Navigation Agent via Auxiliary Progress Estimation**," *Under review for ICLR 2019.* [OpenReview]
- Chih-Yao Ma, Asim Kadav, Iain Melvin, Zsolt Kira, Ghassan AlRegib, and Hans Peter Graf, "**Attend and Interact: Higher-Order Object Interactions for Video Understanding**," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.* [arXiv] [Project] [Blog] [Poster]
- Chih-Yao Ma, Asim Kadav, Iain Melvin, Zsolt Kira, Ghassan AlRegib, and Hans Peter Graf, "**Grounded Objects and Interactions for Video Captioning**," *ViGIL Workshop in Neural Information Processing Systems (NeurIPS), 2017.* [arXiv]
- Chih-Yao Ma, Min-Hung Chen, Zsolt Kira, and Ghassan AlRegib, "**TS-LSTM and Temporal-Inception: Exploiting Spatiotemporal Dynamics for Activity Recognition**," *Signal Processing: Image Communication, 2017.* [arXiv] [GitHub] [Project]
- Chih-Yao Ma and Hsueh-Ming Hang, "**Learning-based Saliency Model with Depth Information**," *Journal of Vision 2015, 15(6):19.* [Paper]
- Chih-Yao Ma, Yu-Cheng Chang, and Yi-Pai Huang, "**Multi-Zone Digital Crosstalk Reduction by Image Processing in 3D display**," *Journal of Display Technology (JDT), Vol. 10, No. 6, pp. 488-493, June 2014.* [Paper]

Patent

US

- Chih-Yao Ma and Ghassan AlRegib, "**Automated and Robust Fabrics Tracking with Buckling Detection and Occlusion-prone Capabilities**," *Georgia Tech Research Corporation. No.: GTRC 7340, 2016.*
- Chih-Yao Ma, Yu-Cheng Chang, and Yi-Pai Huang, "**3D Display Panel and Pixel Brightness Control Method Thereof**," *Publication No.: US20120320097, Publication Date: Dec.20, 2012.* [Patent]

Skill

Deep Learning Frameworks PyTorch, TensorFlow, Torch, MXNet, Caffe

Programming Languages Python, Lua, C/C++, Matlab