

Chih-Yao Ma

PHD · GEORGIA TECH

☎ (+1) 404-907-6833 | ✉ cyma@gatech.edu | 🏠 <https://chiyaoma.github.io/>
| 🐙 GitHub | 📄 Google Scholar | 🔗 LinkedIn

Machine Learning · Deep Learning · Computer Vision · Visual Understanding

Education

Georgia Tech

PHD IN ELECTRICAL AND COMPUTER ENGINEERING

Atlanta, GA

Aug. 2014 - Dec. 2019 (expected)

National Chiao Tung University (NCTU)

BS/MS IN ELECTRICAL AND COMPUTER ENGINEERING

Taiwan

Sept. 2006 - June 2011

Research & Project

Grounded Visual Captioning

PHD

Georgia Tech & Facebook Research

Feb. 2019 - May 2019

- Proposed a novel training regimen to enforce visual captioning models to be visually grounded using PyTorch.
- Improved grounding accuracy 40% and 20% respectively on the Flickr30k dataset without using ground-truth annotations.

The Regretful Navigation Agent

PHD

Georgia Tech

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

PHD

UMD & Georgia Tech

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

RESEARCH INTERN

Salesforce Research

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

RESEARCH INTERN

NEC Labs

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

RESEARCH INTERN

NEC Labs

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

PHD

Georgia Tech

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

PHD

Georgia Tech

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

- 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 2019 - Aug. 2019	Facebook Research , Research Intern, with Marcus Rohrbach (FAIR), Yannis Kalantidis (AML), and Peter Vajda (Mobile Vision)	Menlo Park, CA
May 2018 - Aug. 2018	Salesforce Research , Research Intern, with Caiming Xiong and Richard Socher	Palo Alto, CA
May 2017 - Dec. 2017	NEC Machine Learning Labs , Research Intern, with Asim Kadav , Iain Melvin , and Hans Peter Graf	Princeton, NJ
Aug. 2014 - PRESENT	Georgia Tech , Ph.D., with Ghassan AlRegib and Zsolt Kira	Atlanta, GA
Sept. 2012 - May 2014	CommLab , Research Assistant, NCTU, with Hsueh-Ming Hang	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, Yannis Kalantidis, Ghassan AlRegib, Peter Vajda, Marcus Rohrbach, and Zsolt Kira, “**Learning to Generate Grounded Image Captions without Localization Supervision**,” *Technical Report*, 2019. [\[arXiv\]](#) [\[GitHub \(coming soon\)\]](#) [\[Project\]](#)
- Chih-Yao Ma, Zuxuan Wu, Ghassan AlRegib, Caiming Xiong, and Zsolt Kira, “**The Regretful Agent: Heuristic-Aided Navigation through Progress Estimation**,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019 (**Oral**). [\[arXiv\]](#) [\[GitHub\]](#) [\[Project\]](#) [\[ML@GT\]](#)
- Zuxuan Wu, Caiming Xiong, Chih-Yao Ma, Richard Socher, and Larry Davis, “**AdaFrame: Adaptive Frame Selection for Fast Video Recognition**,” *IEEE Conference on Computer Vision and Pattern Recognition (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**,” *International Conference on Learning Representations (ICLR)*, 2019 (**Top 7% of reviews**). [\[arXiv\]](#) [\[OpenReview\]](#) [\[GitHub\]](#) [\[Project\]](#) [\[Poster\]](#) [\[ML@GT\]](#)
- 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. [\[Paper\]](#)

Patent

US

- Asim Kadav, Chih-Yao Ma, Iain Melvin, and Hans Peter Graf, “**Spatio-temporal interaction network for learning object interactions**,” *Publication No.: US20190019037A1*, *Publication Date: Jan. 17, 2019*. [\[Patent\]](#)
- 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*. [\[Patent\]](#)