

# Chih-Yao Ma

PHD · GEORGIA TECH

☎ (+1) 404-907-6833 | ✉ [cyma@gatech.edu](mailto:cyma@gatech.edu) | 🏠 <https://chiyaoma.github.io/>  
| 💻 <https://github.com/chiyaoma> | 🔗 <https://www.linkedin.com/in/chi-h-yao-ma-9b5b3063/>

Machine Learning · Deep Learning · Computer Vision · Visual Understanding

## Education

### Georgia Tech

MS/PHD IN ELECTRICAL AND COMPUTER ENGINEERING

Atlanta, GA

Aug. 2014 - PRESENT

### NCTU (National Chiao Tung University)

BS/MS IN ELECTRICAL AND COMPUTER ENGINEERING

Taiwan

Sept. 2006 - June 2011

## Research & Project

### 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

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

- 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 2019 - Aug. 2019  | <b>Research Intern</b> , Facebook Research (with <a href="#">Marcus Rohrbach</a> , <a href="#">Yannis Kalantidis</a> , and <a href="#">Peter Vajda</a> ) | <i>Menlo Park, CA</i> |
| May 2018 - Aug. 2018  | <b>Research Intern</b> , Salesforce Research (with <a href="#">Caiming Xiong</a> and <a href="#">Richard Socher</a> )                                    | <i>Palo Alto, CA</i>  |
| May 2017 - Dec. 2017  | <b>Research Intern</b> , NEC Machine Learning Labs (with <a href="#">Asim Kadav</a> )  | <i>Princeton, NJ</i>  |
| Aug. 2014 - PRESENT   | <b>Graduate Research Assistant</b> , Georgia Tech  | <i>Atlanta, GA</i>    |
| Sept. 2012 - May 2014 | <b>Research Assistant</b> , CommLab, NCTU  | <i>Taiwan</i>         |

## Honor & Award

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

## Publication

- [Chih-Yao Ma](#), Zuxuan Wu, Ghassan AlRegib, Caiming Xiong, and Zolt Kira, “**The Regretful Agent: Heuristic-Aided Navigation through Progress Estimation**,” *Under review, 2019*. [\[PDF\]](#)
- Zuxuan Wu, Caiming Xiong, [Chih-Yao Ma](#), Richard Socher, and Larry Davis, “**AdaFrame: Adaptive Frame Selection for Fast Video Recognition**,” *Under review, 2019*. [\[arXiv\]](#)
- [Chih-Yao Ma](#), Jiasen Lu, Zuxuan Wu, Ghassan AlRegib, Zolt Kira, Richard Socher, and Caiming Xiong, “**Self-Monitoring Navigation Agent via Auxiliary Progress Estimation**,” *International Conference on Learning Representations (ICLR), 2019*. [\[OpenReview\]](#) [\[GitHub\]](#) [\[Project\]](#)
- [Chih-Yao Ma](#), Asim Kadav, Iain Melvin, Zolt 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, Zolt 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, Zolt 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*. [\[Patent\]](#)