

# YUE CAO

(+1)9797399821  $\diamond$  cyppsp@tamu.edu

Department of Electrical and Computer Engineering, TAMU  
College Station, Texas 77843-3128, USA

## EDUCATION

---

### Texas A&M University

*Sep. 2016 - Present*

Ph.D. in Electrical Engineering

(Current GPA: 4.0/4.0, Expected Graduation Date: Aug. 2021)

### University of Science and Technology of China

*Sep. 2012 - June 2016*

Bachelor of Science, Applied Physics

## PROFESSIONAL EXPERIENCE

---

### Trusted AI Group, IBM Research

*June 2020 - Aug. 2020*

Ph.D. Research Intern

## RESEARCH INTERESTS

---

- **Machine Learning:** Learning-based Optimization, Bayesian Optimization, Joint Embedding Learning
- **Computational Biology:** Protein Docking, Protein Design

## AWARDS AND HONORS

---

- Received the **NeurIPS** Travel Award. Oct. 2019
- Received the **Critical Assessment of Genome Interpretation (CAGI)** Fellowship. Nov. 2019
- Our team (Y. Cao and Y. Shen) ranked the **2nd** among **26** groups for difficult targets in the 3rd joint **CASP-CAPRI** (Critical Assessment of protein Structure Prediction and Critical Assessment of PRedicted Interactions), a community-wide experiment on comparative evaluation of protein structure prediction and protein docking methods. Apr. 2019
- Our team (Y. Cao and Y. Shen) ranked the **3rd/51** for difficult targets in the 7th **CAPRI** (Critical Assessment of PRedicted Interactions), 2017-2019
- Received the **First-class Award** for *Excellent Students* in University of Science and Technology of China. Sep. 2015
- Bronze Medal in the 4th **Asia-Pacific Informatics Olympiad**, May 2010
- First-class Award in the **National Olympiad in Physics** in China. Nov, 2011
- First-class Award in the **National Olympiad in Informatics** in China. Nov, 2011
- First-class Award in the **National Olympiad in Informatics** in China. Dec, 2010
- First-class Award in the **National Olympiad in Informatics** in China. Dec, 2009

## PUBLICATIONS

---

Preprints/Under Review

- **Y. Cao**, T. Chen, Z. Wang, and Y. Shen. *Bayesian Learning to Optimize: Quantifying the Optimizer Uncertainty*. Under review for **International Conference on Learning Representations**, 2021
- **Y. Cao**, P. Das, P. Chen, V. Chen., I. Melnyk, and Y. Shen. *Fold2Seq: A Joint Sequence(1D)-Fold(3D) Embedding-based Generative Model for Protein Design*. Under review for **International Conference on Learning Representations**, 2021
- **Y. Cao** and Y. Shen. *TALE: Transformer-based protein function Annotation with joint sequence-Label Embedding*. **bioRxiv** 315937, Under review for **Bioinformatics**, 2020
- R. Taftaf and other authors including **Y. Cao**. *ICAM1 initiates CTC cluster formation and lung metastasis of triple negative breast cancer*. Under review for **Nature Communications**, 2020

## Published

- **Y. Cao** and Y. Shen. *Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking*. **Journal of Chemical Theory and Computation**, 16 (8), 5334-5347, 2020 (IF=5.31)
- **Y. Cao**, T. Chen, Z. Wang, and Y. Shen. *Learning to Optimize in Swarms*. **Advances in Neural Information Processing Systems** 32, 15018-15028, 2019 (acceptance rate: 21.6%)
- M. Karimi\*, S. Zhu\*, **Y. Cao\*** and Y. Shen. *De Novo Protein Design for Novel Folds Using Guided Conditional Wasserstein Generative Adversarial Networks (gcWGAN)*. **Journal of Chemical Information and Modeling** (<https://doi.org/10.1021/acs.jcim.0c00593>) (\* Equal Contributions) (IF=4.55)
- **Y. Cao**, Y. Sun, M. Karimi, H. Chen, O. Moronfoye, and Y. Shen. *Predicting Pathogenicity of Missense Variants with Weakly Supervised Regression*. **Human Mutation**, 40(9), 1579-1592, 2019 (IF=5.36)
- **Y. Cao** and Y. Shen. *Energy-based Graph Convolutional Networks for Scoring Protein Docking Models*. **Proteins**. <https://doi.org/10.1002/prot.25888>; 2020
- M. Kawaguchi, N. Dashzeveg, **Y. Cao**, Y. Jia, X. Liu, Y. Shen, and H. Liu. *Extracellular Domains i-ii of CD44 Mediate Its Trans-homophilic Dimerization and Cluster Aggregation*. **Journal of Biological Chemistry**, 295 (9), 2640-2649, 2020
- M. S Cline, G. Babbi, S. Bonache, **Y. Cao**, et al. *Assessment of Blind Predictions of the Clinical Significance of BRCA1 and BRCA2 Variants*. **Human Mutation**, 40, 1546-1556, 2019.
- M. Lensink and other authors including **Y. Cao**. *Blind prediction of Homo- and Hetero- Protein Complexes: The CASP13-CAPRI Experiment*. **Proteins: Structure, Function, and Bioinformatics**, doi:10.1002/prot.25838, 2019
- A. Voskanian and other authors including **Y. Cao**. *Assessing the Performance of in-silico Methods for Predicting the Pathogenicity of Variants in the Gene CHEK2 among Hispanic Females with Breast Cancer*. **Human Mutation**, 40, 1612-1622, 2019
- X. Liu and other authors including **Y. Cao**. *Homophilic CD44 Interactions Mediate Tumor Cell Aggregation and Polyclonal Metastasis in Patient-derived Breast Cancer Models*. **Cancer Discovery**, 9(1): 96113, 2019

## INVITED PRESENTATIONS

- **Y. Cao**, T. Chen, Z. Wang, Y. Shen. *Learning to Optimize in Swarms*. (Poster) **Advances in Neural Information Processing Systems (NeurIPs)**, Dec. 2019, Vancouver, Canada

- **Y. Cao**, Y. Sun, M. Karimi, H. Chen, O. Moronfoye, and Y. Shen. *Predicting Pathogenicity of Missense Variants with Weakly Supervised Regression*. **Critical Assessment of Genome Interpretation (CAGI) Workshop**, Dec. 2019, San Francisco, USA
- **Y. Cao** and Y. Shen. *Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking* (Presented by Yang Shen). **Intelligent Systems for Molecular Biology (ISMB)**, July 2019, Basel, Switzerland
- M. Karimi\*, S. Zhu\*, **Y. Cao\*** and Y. Shen. *De Novo Protein Design of Novel Folds using Guided Conditional Generative Adversarial Networks (gcWGAN)* (Poster). **Intelligent Systems for Molecular Biology (ISMB)**, July 2019, Basel, Switzerland
- **Y. Cao** and Y. Shen. *Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking* (Presented by Yang Shen). **7th CAPRI Evaluation Meeting**, April 2019, Hinxton, UK
- **Y. Cao** and Y. Shen. *Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking* (Poster). **Modeling of Protein Interaction (MPI)**, November 2018, Lawrence, KS, USA
- **Y. Cao** and Y. Shen. *Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking* (Poster). **Bioinformatics and Cancer Symposium**, Sep. 2018, College Station, TX, USA

## TECHNICAL SKILLS

---

<b>Programming Languages:</b>	Python, C++, Bash Scripts
<b>Deep Learning Frameworks:</b>	Pytorch, Tensorflow
<b>Operating Systems:</b>	Linux, Mac OS, Windows
<b>Other Computer Skills:</b>	Git, PyMOL, Latex, CHARMM

## SELECTED COURSES

---

- **Machine Learning:** Machine Learning with Networks, Statistical Machine Learning, Probabilistic Graphical Modeling, Reinforcement Learning
- **Optimization:** Linear Programming
- **Statistics & Maths:** Information Theory, Statistical Inference, Graph Theory
- **Algorithms:** Analysis of Algorithm, Data Structures

## SOCIAL ACTIVITIES

---

- Distribute free furniture at Grace Bible Church, College Station *June 2018*
- Participate in the e-sports college competition: Cstarleague, as the leader of team TWELFTH MAN DOTA and representing TAMU *Sep. 2019*
- Lead the team of Department of Physics to win the silver medal of the soccer competition in USTC *Mar. 2015*
- Get sponsorship of Lejia Market for the dance party of Department of Physics, as the member of the financial office of Student Union at USTC *Oct. 2014*
- Send flyers to promote legal knowledge at Jinhu community, Wuhu, Anhui, China *Oct. 2013*