Sangdon Park

sangdon.github.io
sangdon@postech.ac.kr
github.com/sangdon
Google Scholar

Research Interests

Artificial Intelligence (AI), Trustworthy AI, Uncertainty Quantification, and Computer Security — My research interest focuses on designing trustworthy AI systems by understanding from theory to implementation and by considering practical applications in computer security, computer vision, natural language processing, robotics, and cyber-physical systems.

Education

University of Pennsylvania — Philadelphia, USA
Ph.D. in Computer and Information Science — 2021

- Advisors: Insup Lee and Osbert Bastani
- Thesis: *Uncertainty Estimation Toward Safe AI*
- Committee: Kostas Daniilidis, Nikolai Matni, Edgar Dobriban, and Kilian Q. Weinberger

Seoul National University — Seoul, Korea
M.S. in Electrical and Computer Engineering — 2012

- Advisor: Kyoung Mu Lee
- Thesis: *Abnormal Object Detection by Transformed-Canonical Scene Generation*

Seoul National University — Seoul, Korea
B.S. in Computer Science and Engineering — 2010

- Thesis Advisor: Byoung-Tak Zhang
- Thesis: *Behavioral Intelligence for Crowd Avatar in 3D Virtual Worlds*

Employment

POSTECH — Pohang, Korea
Assistant Professor — Aug. 2023-Now

Georgia Institute of Technology — Atlanta, USA
Postdoctoral Researcher (Mentor: Taesoo Kim) — Sept. 2021-July 2023

Google Cloud AI — Sunnyvale, USA
Research Intern (Host: Kihyuk Sohn) — Summer 2020

Biointelligence Laboratory, Seoul National University — Seoul, Korea
Undergraduate Researcher — 2008-2010

Republic of Korea Army — Korea
Military Service — 2006-2008

Publications

[1] S. Li, **S. Park**, I. Lee, and O. Bastani, "TRAQ: Trustworthy Retrieval Augmented Question Answering via Conformal Prediction", in *Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**)*, 2024.

[2] H. Park, J. Hwang, S. Mun, **S. Park**, and J. Ok, "MedBN: Robust Test Time Adaptation against Malicious Test Samples", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**)*, 2024.

[3] W. Si, **S. Park**, I. Lee, E. Dobriban, and O. Bastani, "PAC prediction sets under label shift", in *The Twelfth International Conference on Learning Representations (**ICLR**)*, 2024.

[4] W. Si, S. Li, **S. Park**, I. Lee, and O. Bastani, "Angelic Patches for Improving Third-Party Object Detector Performance", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**)*, 2023.

[5] **S. Park**, O. Bastani, and T. Kim, "ACon^2: Adaptive Conformal Consensus for Provable Blockchain Oracles", in *Proceedings of the 32nd USENIX Security Symposium (**Security**)*, 2023.

[6] R. Kaur, K. Sridhar, **S. Park**, Y. Yang, S. Jha, A. Roy, O. Sokolsky, and I. Lee, "CODiT: Conformal out-of-distribution detection in time-series data for cyber-physical systems", in *Proceedings of the 14Th ACM/IEEE International Conference on Cyber-Physical Systems (**ICCPS**) (to appear)*, 2023.

[7] **S. Park**, X. Cheng, and T. Kim, "Unsafe's Betrayal: Abusing Unsafe Rust in Binary Reverse Engineering via Machine Learning", *arXiv preprint arXiv:2211.00111*, 2023.

[8] **S. Park**, E. Dobriban, I. Lee, and O. Bastani, "PAC Prediction Sets for Meta-Learning", in *Neural Information Processing Systems (**NeurIPS**)*, 2022.

[9] S. Li, **S. Park**, X. Ji, I. Lee, and O. Bastani, "Towards PAC multi-object detection and tracking", *arXiv preprint arXiv:2204.07482*, 2022.

[10] S. Jang, **S. Park**, I. Lee, and O. Bastani, "Sequential covariate shift detection using classifier two-sample tests", in *Proceedings of the 39th International Conference on Machine Learning (**ICML**)*, 2022.

[11] R. Kaur, S. Jha, A. Roy, **S. Park**, E. Dobriban, O. Sokolsky, and I. Lee, "iDECODe: In-distribution equivariance for conformal out-of-distribution detection", in *Association for the Advancement of Artificial Intelligence (**AAAI**)*, 2021.

[12] **S. Park**, S. Li, I. Lee, and O. Bastani, "PAC confidence predictions for deep neural network classifiers", in *International Conference on Learning Representations (**ICLR**)*, 2021.

[13] **S. Park**, O. Bastani, J. Weimer, and I. Lee, "Calibrated prediction with covariate shift via unsupervised domain adaptation", in *International Conference on Artificial Intelligence and Statistics (**AISTATS**)*, 2020.

[14] **S. Park**, O. Bastani, N. Matni, and I. Lee, "PAC confidence sets for deep neural networks via calibrated prediction", in *International Conference on Learning Representations (**ICLR**)*, 2020.

[15] **S. Park**, R. Ivanov, J. Weimer, and I. Lee, "From verification to learning for defense against adversarial examples in neural networks", *Korea Cyber-security Competition*, 2018.

[16] **S. Park**, J. Weimer, and I. Lee, "Resilient linear classification: An approach to deal with attacks on training data", in *International Conference on Cyber-Physical Systems (**ICCPS**)*, 2017.

[17] J. Oh, T. M. Howard, M. R. Walter, D. Barber, M. Zhu, **S. Park**, A. Suppe, L. Navarro-Serment, F. Duvallet, A. Boularias, *et al.*, "Integrated intelligence for human-robot teams", in *International Symposium on Experimental Robotics (**ISER**)*, 2016.

[18] **S. Park**, W. Kim, and K. M. Lee, "Abnormal object detection by canonical scene-based contextual model", in *European Conference on Computer Vision (**ECCV**)*, 2012.

Scholarships and Awards

- NeurIPS23 Outstanding Reviewer Award 2023
- ICCPS23 Best Paper Award finalist 2023
- NeurIPS21 Outstanding Reviewer Award (top 8% of reviewers) 2021
- Korea cyber-security competition best paper award ($4,500) 2018
- PhD fellowship at University of Pennsylvania 2013-2021
- Distinguished MS Dissertation Award at Seoul National University 2012
- Academic Performance Scholarship 2009
- National Science and Engineering Undergraduate Scholarship 2003–2008

Service

- **Area Chair**
  NeurIPS24
- **Reviewer**
  NeurIPS21-23, ICML21-23, ICLR22-24, Journal of the Royal Statistical Society: Series B
- **External Reviewer**
  S&P21, S&P22, Security22, Security23, Security24, NDSS24

Projects (selected)

- **D3: Debloating, Dialecting and Diversification for Attack Resilient Software with Real-time Constraints** for Technology Innovation Institute, Abu Dhabi, UAE Sept. 2021 - Aug. 2023
  We exploit sound measurement from a drone or a drone swarm to detect adversarial attacks.
- **Assured Autonomy** for DARPA Jun. 2018 - Aug. 2021
  My task is to provide a probabilistic guarantee on the correctness of uncertainty such that it can be deployed in the real world.
- **Security and Privacy-Aware Cyber-Physical Systems** for NSF-Intel Mar. 2016 - Aug. 2019
  My aim was to devise a robust learning algorithm for classification by maximizing example-margins of neural networks such that the learned neural net is robust to adversarial examples.
- **High-Assurance Cyber Military Systems (HACMS)** for DARPA Sep. 2015 - Feb. 2016
  My goal was to automatically tune PID attitude controllers by identifying dynamics of an unmanned aerial vehicle (UAV) such that PID controllers with initially zero PID parameters can control the attitude of the UAV. Check out!
- **Robotics Collaborative Technology Alliance** for U.S. Army Research Lab Sep. 2013 - Dec. 2014
  My goal was to detect a facade and doors for proposing the direction-to-observe of an unmanned ground vehicle.

Teaching

- **Instructor** at POSTECH Spring 2024
  Discrete Mathematics (CSED261)
- **Instructor** at POSTECH Fall 2023
  Trustworthy ML (AIGS703L / CSED703L)
- **Teaching Assistant** at University of Pennsylvania Spring 2015
  Machine Perception (CIS580)
- **Teaching Assistant** at University of Pennsylvania Fall 2014
  Computer Vision and Computational Photography (CIS581)

- **Teaching Assistant** at Seoul National University — Fall 2010
  Linear Algebra for Electrical Systems
- **Instructor** at Seoul National University — Feb. 2008
  1st Free Computer Education for Gwanak-gu Community Youth

Skills

- **Libraries:** `PyTorch`, `Matplotlib`, `TensorFlow`, `foundry`, `Web3.py`, `ROS`
- **Programming language:** `Python`, LaTeX, `MATLAB`, `C/C++`, `Solidity`, `Rust`
- **Communication language:** Korean, English

Talks

- **From Verification to Learning for Defense against Adversarial Examples in Neural Networks**
  KAIST CS — Aug. 2018
  Hanyang University — Aug. 2018
  KIISC — Aug. 2018
- **Uncertainty Quantification via PAC Prediction Sets**
  DGIST — Dec. 2021
- **PAC Prediction Sets for AI Safety**
  ICML Workshop DFUQ 2022 Invited Talk — Jul. 2022
- **Uncertainty Learning for Trustworthy and Secure AI**
  POSTECH AI/CSE — Mar. 2023
  KAIST EE — April 2023
  SNU CSE — April 2023
  Korea University CSE — July 2023
  SNU IPAI — July 2023
  SNU Frontier Summer School — Aug. 2023
  UNIST IB — Oct. 2023
  KAIST — Jan. 2024
  CAU AI — May 2024
  Korea GSS — June 2024
- **Conformal Prediction for Trustworthy AI**
  Korean AI Association Winter School — Feb. 2024
- **Trustworthy AI: A Compositional Perspective**
  POSTECH AI/CSE — April 2024