Saeed Khorram

E-mail: khorrams@oregonstate.edu
Website: https://khorrams.github.io
LinkedIn: https://www.linkedin.com/in/saeed-khorram-7a0a8281/
Google Scholar: https://scholar.google.com/citations?hl=en&user=-zfeeKUAAAAJ/

Education

Oregon State University, Corvallis, Oregon, USA
Ph.D. in Computer Science (— Expected Winter 2023)
M.Sc in Computer Science (June 2020)
Advised by Dr. Fuxin Li

Research Interests

Explainable AI (XAI), Computer Vision, Feature Disentanglement, Generative Models, Unsupervised Learning, Reinforcement Learning.

Work And Research Experiences

2D/3D Computer Vision for Autonomous Vehicles
Machine Learning Research Engineer Intern at Scale AI, Jan - Apr 2021
Pre-labeling scenes based on Lidar and camera inputs of AVs.

eXplainable Artificial Intelligence (XAI)
Researcher at Oregon State University, Dec 2017 - Now

- **Learning from the long-tail data**: Using generative models to augment the tail classes.
- **Counterfactual visual explanation**: Generating CF explanations from the latent space of generative models.
- **Understanding recurrent policy networks by Finite-State Machines (FSM)**: Quantizing the representations of memory and observations of RNNs and analyzing them using FSM, known as Moore Machine.
- **Non-negative factorization for feature disentanglement:** Layer-wise feature disentanglement of deep networks by low-rank matrix factorization and a novel training scheme for deep networks using ADMM.
- **Integrated-Gradient optimized attribution (saliency) maps:** Saliency map generation methods that optimize perturbation masks using integrated gradient.
- **Deep feature embedding for automatic high-level concept extraction:** a novel explanation module for extracting concepts from the activation space of the deep networks

Recent Publications

- **Saeed Khorram**, Li Fuxin. "Cycle-Consistent Counterfactuals by Latent Transformations". (**CVPR 2022)**
- Li Fuxin, Zhongang Qi, **Saeed Khorram**, Vivswan Shitole, Prasad Tadepalli, Minsuk Kahng, Alan Fern. "From Heatmaps to Structured Explanations of Image Classifiers". (Applied AI Letters 2021)
- Mohamad H. Danesh, Anurag Koul, Alan Fern, **Saeed Khorram**. "Re-Understanding Finite-State Representations of Recurrent Policy Networks". (**ICML 2021**)
- **Saeed Khorram**, Xiao Fu, Mohamad H. Danesh, Zhongang Qi, Li Fuxin. "Stochastic Block ADMM for Training Deep Networks" (pre-print).
- **Saeed Khorram**, Tyler Lawson, Li Fuxin. "IGOS++: Integrated Gradient Optimized Saliency by Bilateral Perturbations". (ACM-CHIL 2021)
- Zhongang Qi, **Saeed Khorram**, Li Fuxin. "Embedding Deep Networks into Visual Explanations". (Journal of AI 2020)
- Zhongang Qi, **Saeed Khorram**, Li Fuxin. "Visualizing Deep Networks by Optimizing with Integrated Gradients". (**AAAI 2020**)
- **Saeed Khorram**. "Toward Disentangling the Activations of the Deep Networks via Low-dimensional Embedding and Non-negative Factorization" (M.Sc. Thesis)

Professional Services

- Reviewer for ICLR, ICML, NeurIPS, CVPR, ECCV, and AAAI.

Coding

- Python, PyTorch, Keras, Tensorflow, Matlab, C, JavaScript, HTML, CSS, Bash, Git, AWS, Kubernetes, Docker.