

# Saeed Khorram

E-mail: [khorrans@oregonstate.edu](mailto:khorrans@oregonstate.edu)

Website: <https://khorrans.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 Fall 2022)

*M.Sc in Computer Science* (June 2020)

Advised by Dr. Fuxin Li

## RESEARCH INTERESTS

Explainable AI, 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*

- **Counterfactual visual explanation:** Generating CF explanations from the latent space of generative models. (under-review CVPR 2022)
- **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. (ICMLW 2020 + ICML 2021)
- **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 (Dissertation + preprint 2020)
- **Integrated-Gradient optimized attribution (saliency) maps:** Saliency map generation methods that optimize masks using integrated gradient. (**I-GOS:** CVPRW 2019 + AAAI 2020. **iGOS++:** ACM-CHIL 2021 )
- **Deep feature embedding for automatic high-level concept extraction:** proposing a novel explanation module for extracting high-level concepts from the activation space of the deep networks (Journal of AI 2020)

## Automatic Medical Coding *Data scientist at Carrene, May - Dec 2017*

Research on automatic medical prescription analysis and coding.

## RECENT PUBLICATIONS

- 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’21)
- 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-perint).
- **Saeed Khorram**, Tyler Lawson, Li Fuxin. “IGOS++: Integrated Gradient Optimized Saliency by Bilateral Perturbations”. (ACM-CHIL’21)
- Zhongang Qi, **Saeed Khorram**, Li Fuxin. “Embedding Deep Networks into Visual Explanations”. (Journal of AI’20)
- Zhongang Qi, **Saeed Khorram**, Li Fuxin. “Visualizing Deep Networks by Optimizing with Integrated Gradients”. (CVPRW’19 - AAAI’20)

## DISSERTATION

- **Saeed Khorram**. “Toward Disentangling the Activations of the Deep Networks via Low-dimensional Embedding and Non-negative Factorization”

## CODING

{Python, Matlab, C}, {PyTorch, Keras, Tensorflow}, {JavaScript, HTML, CSS}, {Bash, Git}, {AWS, Kubernetes, Docker}