# Arushi Rai

 $\Box +1$  (650) 279 2949

arushirai1999@gmail.com 🖸 arushirai1 🖬 arushi-r-568887b3

## Education \_\_\_\_

#### University of Pittsburgh

Doctorate in Computer Science

Relevant Coursework: Vision-Language Models for Computer Vision (Advanced Topics in AI), Advanced Topics in Natural Language Processing (Large Language Models), Introduction to NLP, Machine Learning, Artificial Intelligence

#### **Illinois Institute of Technology**

Bachelor of Science in Computer Science

Honors: Magna Cum Laude Relevant Coursework: Philosophy and Psychology of Language, Geospatial Vision and Visualization, Linear Algebra, Data Mining, Probability and Statistics, Regression, Online Social Networks Analysis

# Publications

[1] A. Rai and A. Kovashka, "VEIL: Vetting Extracted Image Labels from In-the-Wild Captions for Weakly-Supervised Object Detection." To appear at The 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL), March 2024.[pdf]

[2] A. Rai and A. Kovashka, "Improving Language-Supervised Object Detection With Linguistic Structure Analysis," presented at Workshop on Open-Domain Reasoning Under Multi-Modal Settings (O-DRUM), held in conjunction with the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2023. [pdf]

### Experience \_

#### **Graduate Research Assistant**

University of Pittsburgh

• My research is focused on using language as supervision for vision tasks such as object detection, visual reasoning, and action quality assessment. Some challenges include handling noisy supervision from naturally occurring captions and vision-language models. See projects below for more details and publications above.

Machine	Learning	Engineer	Intern	(Computer	Vision)
Gather.ai					

- Improved case counting pipeline by 13% on difficult, irregular cases by targeting intermediate stack segmentation and mean-shift clustering
- Implemented **instance segmentation approaches** like Panoptic DeepLab and increased robustness to viewpoint shifts through perspective augmentations and fusing depth cues with RGB images through attention-based multi-modal fusion techniques.

#### Computer Vision Research Intern (REU) May 2020 - August 2020 University of Central Florida — Center for Research in Computer Vision Link to Report

• Extended self-supervised learning frameworks such as DeepCluster and XDC in Pytorch to handle multimodal data from HowTo100M like video clips and narrations to achieve 5% improvement on long-complex activity recognition.

# Highlighted Projects

Weakly Supervised Object Detection with Noisy Supervision — 2022-2023

- Effectively used abundant noisy caption sources for weakly supervised object detection using torch, BERT, and Wetectron to achieve a 9-point Mean Average Precision (mAP) improvement on PASCAL VOC-07 for weakly supervised object detection.
- Increased precision of noisy labels (reducing false positives) for WSOD by developing a visual presence model conditioned on caption context that bootstraps YOLO and VinVL-C4, to increase label precision by 44.9%.
- Mitigated imbalanced label distribution to facilitate model convergence through weighted frequency sampling and improved performance by 3 mAP on PASCAL VOC-07 on top of gains due to better label quality.

Can Vision-Language Models understand Memes? -2021

• Designed experiments and applied hypothesis testing to measure effect of pragmatic and discourse attributes on pretrained vision-language models like CLIP and LXMERT by computing vision-language alignment.

Search Feature for Socratic Diary - 2021

• Reduced performance cost from polynomial round robin comparison to linear comparison by generating custom (SimHash) indexes in Firestore to scale question search.

July 2017 - Dec 2020 GPA: 3.76

August 2021 - Current

August 2021 - Current

May 2022 - August 2022

GPA: 3.75