

Arka Sadhu

Applied Research Scientist
Surreal Team
Meta

Sunnyvale, CA - USA

✉ ark.sadhu2904@gmail.com

📄 <https://theshadow29.github.io/>

[LinkedIn](#)

[GScholar](#)

[Github](#)

[1 Page CV](#)

Work Experience

06/03/2024- **Meta**, 1 Hacker Wy, Menlo Park, CA 94025.

Present Applied Research Scientist
Surreal Team

- Working in surreal team on Aria glasses to power Contextual AI agents.
- Contributed to multiple projects from Anonymization, LLM Agents, Retrieval Augmentation Generation (RAG), Long-Context Open-Loop Benchmarking, Streaming VLM design.
- Our project was shown to all top-executives in Meta including Mark Zuckerberg, Andrew Bosworth, Alexandr Wang, Chris Cox. My contribution was particularly towards the Contextual AI Agents.
- Proficient in Pytorch, distributed training, dataset and benchmark creation and evaluation.

Education

Aug'18-May'24 **University of Southern California.**

PhD Computer Science (Annenberg Fellow)

Advisor: Prof. Ram Nevatia

CGPA 4.0/4.0

July'14-May'18 **Indian Institute of Technology Bombay.**

Bachelor of Technology in Electrical Engineering (with Honors)

Minor in Computer Science

CGPA 9.44/10

Apr'12-Mar'14 **R. N. Podar High School.**

CBSE Higher Secondary

Internship Experience

May'22-Aug'22 **Meta AI, Menlo Park**, *Research Intern*,

Supervised by Animesh Sinha, Licheng Yu, Hugo Chen and Ning Zhang.

- Unaligned Video-Text Pre-training using Iterative Alignment

May'21-Aug'21 **Allen Institute for Artificial Intelligence (AI2), Seattle**, *Research Intern*,

Supervised by Tanmay Gupta, and Aniruddha Kembhavi.

- Unifying Video and Image Language Models

May'20-Aug'20 **Allen Institute for Artificial Intelligence (AI2), Seattle**, *Research Intern*,

Supervised by Mark Yatskar, Tanmay Gupta, and Aniruddha Kembhavi.

- Recognizing Situations in a Video

- May'18-Aug'18 **Wadhvani Institute for AI, Mumbai**, *Research Intern*,
Supervised by Prof. Subhasis Chaudhuri, Rahul Panicker.
 - Worked on automatic grading of ZN stained Tuberculosis Sputum Microscopic images and achieved less than 5% counting error.
- May'17-July'17 **University of Southern California, Los Angeles**, *Research Intern*,
Supervised by Prof. Ram Nevatia.
 - Developed clustering algorithms to detect spoofed images containing a scene similar to another image and achieved 85% accuracy on nimble dataset for baseline matching.
- May'16-July'16 **Aalto University, Finland**, *Research Intern*,
Supervised by Prof. Juho Kannala.
 - Developed robust loop closures for reconstruction of 3D indoor environments and analyzed contributions of various loss functions in the final generated point cloud.

Publications

- DiVE-k: Differential Visual Reasoning for Fine-grained Image Recognition
Raja Kumar, **Arka Sadhu**, Ram Nevatia
Under Submission ICLR26 [[Paper](#)]
- Detect Anything: An Open-Vocabulary Object Detection Model with Weak Supervision
Arka Sadhu, Ram Nevatia
Part of USC Thesis
- Leveraging Task-Specific Pre-Training to Reason across Images and Videos
Arka Sadhu, Ram Nevatia
WACV'24 [[Paper](#)]
- Unaligned Video-Text Pre-training using Iterative Alignment [[Paper](#)]
Arka Sadhu, Licheng Yu, Animesh Sinha, Yu Chen, Ram Nevatia, Ning Zhang
- Gradient-based Memory Editing for Task-Free Continual Learning [[Paper](#)] [[Code](#)]
Xisen Jin, **Arka Sadhu**, Junyi Du, Xiang Ren
Neurips 2021
arxiv:2006.15294 [cs.LG]
- Visual Semantic Role Labeling for Video Understanding [[Paper](#)] [[Code](#)]
Arka Sadhu, Tanmay Gupta, Mark Yatskar, Ram Nevatia, Aniruddha Kembhavi
CVPR 2021
arxiv:2104.00990 [cs.CV, cs.CL]
- Video Question Answering with Phrases via Semantic Roles [[Paper](#)][[Code](#)]
Arka Sadhu, Kan Chen, Ram Nevatia
NAACL 2021
arxiv:2104.03762 [cs.CV, cs.CL]
- Improving Object Detection and Attribute Recognition by Feature Entanglement Reduction [[Paper](#)]
Zhaoheng Zheng, **Arka Sadhu**, Ram Nevatia
ICIP 2021
- Utilizing Every Image Object for Semi-supervised Phrase Grounding [[Paper](#)]
Haidong Zhu, **Arka Sadhu**, Zhaoheng Zheng, Ram Nevatia
WACV 2021
arxiv:2011.02655 [cs.CV]
- Visually Grounded Continual Learning of Compositional Phrases [[Paper](#)][[Code](#)]
Xisen Jin, Junyi Du, **Arka Sadhu**, Ram Nevatia, Xiang Ren

EMNLP 2020

arxiv:2005.00785 [cs.CL]

- Video Object Grounding using Semantic Roles in Language Description [[Paper](#)] [[Code](#)]

Arka Sadhu, Kan Chen, Ram Nevatia

CVPR 2020

arXiv:2003.10606 [cs.CV]

- Zero-Shot Grounding of Objects from Natural Language Queries [[Paper](#)] [[Code](#)]

Arka Sadhu, Kan Chen, Ram Nevatia

ICCV 2019 (Oral)

arXiv:1908.07129 [cs.CV]

Invited Talks

Jun'21 **CVPR Activity Recognition Workshop**, *Summer21*.

“Introducing VidSitu Challenge at CVPR'21 [[Slides](#)]

Jan'20 **USC-WiSE Seminar Series**, *Spring20*.

“The Need for Language in Vision” [[Slides](#)]

June'19 **Workshop on Media Forensics at CVPR'19**, *Summer'19*.

“Recap of Synthetic Realities - Audio Visual Fakes Workshop held at ICML'19” [[Slides](#)]

Service

Oct'25 **CVPR'26**, *Reviewer*.

Oct'25 **ICRA'26**, *Reviewer*.

Oct'25 **HackwithBay – Financial Agents Hackathon**, *Finalist Judge*.

Sep'25 **ICLR'26**, *Reviewer*.

Aug'25 **AAAI'26**, *Reviewer*.

Jun'25 **Neurips'25 Datasets and Benchmark**, *Reviewer*.

May'25 **Neurips'25**, *Reviewer*.

May'25 **ARR'25 May**, *Reviewer*.

May'25 **BMVC'25**, *Reviewer*.

Apr'25 **ICCV'25**, *Reviewer*.

Feb'25 **ARR'25 Feb**, *Reviewer*.

Feb'25 **COLM'25**, *Emergency Reviewer*.

Feb'25 **ICML'25**, *Emergency Reviewer*.

Jan'25 **CVPR'25**, *Emergency Reviewer*.

Dec'24 **ARR'24 Dec**, *Reviewer*.

Oct'24 **ARR'24 Oct**, *Reviewer*.

Aug'24 **ICLR'24**, *Reviewer*.

Jul'24 **EMNLP'24**, *Reviewer*.

Jul'24 **Neurips'24**, *Reviewer*.

Jun'24 **ARR'24 June**, *Reviewer*.

Jun'24 **ALVR'24 Workshop**, *Reviewer*.

Jun'24 **WACV'25**, *Reviewer*.
May'24 **BMVC'24**, *Reviewer*.
Mar'24 **ECCV'24**, *Reviewer*.
Dec'23 **ACL'24**, *Reviewer*.
Nov'23 **CVPR'24**, *Reviewer*.
July'23 **WACV'24**, *Reviewer*.
June'23 **EMNLP'23**, *Reviewer*.
May'23 **Neurips'23**, *Reviewer*.
May'23 **BMVC'23**, *Reviewer*.
Mar'23 **ICCV'23**, *Reviewer*.
Jan'23 **TPAMI**, *Reviewer*.
Jan'23 **ICML'23**, *Reviewer*.
Jan'23 **ACL'23**, *Reviewer*.
Jan'23 **AURO**, *Reviewer*.
Nov'22 **CVPR'23**, *Reviewer*.
Aug'22 **BMVC'22**, *Reviewer*.
Jun'22 **WACV'23**, *Reviewer*.
Jun'22 **EMNLP'22**, *Reviewer*.
May'22 **Neurips'22**, *Reviewer*.
Mar'22 **ECCV'22**, *Reviewer*.
Nov'21-Nov'22 **ACL Rolling Review**, *Reviewer*.
Nov'21 **ICLR'22**, *Emergency Reviewer*.
Nov'21 **CVPR'22**, *Reviewer*.
Oct'21 **IJCV**, *Reviewer*.
Sept'21 **AAAI'22**, *Reviewer*.
June'21 **CVPR'21 ActivityNet Workshop VidSitu Presentation**, [[Website](#)].
June'21 **NAACL'21 ALVR Workshop Program Committee**, [[Website](#)].
June'21 **BMVC'21**, *Reviewer*.
May'21 **EMNLP'21**, *Reviewer*.
April'21 **IROS'21**, *Reviewer*.
March'21 **ICCV'21**, *Reviewer*.
Feb'21 **ACL'21**, *Reviewer*.
Dec'20 **CVPR'21**, *Reviewer*
Awarded Outstanding Reviewer, [[Link](#)].
Oct'20 **AAAI'21**, *Reviewer*.
May'20 **BMVC'20**, *Reviewer*
Awarded Outstanding Reviewer, [[Link](#)].
March'20 **ICPR'20**, *Reviewer*.
August'19 **WACV'20**, *Reviewer*.

June'19 **ICML'19 Workshop organizer**,
Synthetic Realities - Audio Visual Fakes, [[Website](#)].

Teaching Experience

Spring'21 **CSCI 570: Analysis of Algorithms**,
University of Southern California,
Teaching assistant with Prof. Victor Adamchik.

Fall'19 **CSCI 677: Advanced Computer Vision**,
University of Southern California,
Teaching assistant with Prof. Ram Nevatia.

Academic and Technical Achievements

- Awarded Outstanding Reviewer at CVPR'21.
- Awarded Outstanding Reviewer at BMVC'20.
- Awarded Annenberg Graduate Fellowship at University of Southern California, 2018.
- Placed 3rd at the IFood Challenge which was a part of the FGVC5 (Fifth Workshop on Fine Grained Visual Categorization) as a part of CVPR 2018.
- Awarded Viterbi-India Scholarship to pursue research at Viterbi School of Engineering (Summer'17)
- Awarded Aalto Research Assistant fellowship to pursue research at Aalto University (Summer'16)
- Received the Best Idea Award for our project on Virtual Reality at the Tech and RnD Expo 2015.
- Secured All India Rank 134 in JEE Advance 2014 among 1.5 lakh selected candidates.

Technical Skills

- **Programming Languages:** C++, Python, Javascript, Matlab, Bash, Embedded-C, Assembly
- **Deep Learning Frameworks:** Pytorch.
- **Computer Vision Frameworks:** OpenCV, Scikit-Image, Scikit-Learn, Pillow
- **Web Dev:** JQuery, D3JS, Bootstrap, Flask

Notable Open Source Repositories

Feb'20-Present **Research Advice List**, [[Repository](#)].

- A curated list of presentations and write-ups for research and writing.

Sep'18-Present **Awesome-Grounding**, [[Repository](#)].

- A curated list of grounding papers.

Dec'18-Jan'19 **Visual Commonsense Pytorch**, [[Code](#)].

- Pytorch implementation of BERT model used for Visual Commonsense dataset.

Notable Projects

Feb'19-Apr'19 **VQA using scene-graphs**, *Course Project* [[Report](#)] [[Code](#)].

- Investigated using scene-graphs for visual-question answering. Combined the scene-graph generation process along with existing VQA methods for interpretable reasoning.

- Aug'18-Nov'18 **Subreddit Classification**, *Course Project* [[Report](#)] [[Code](#)],
Advisor: Prof. Jonathan May and Prof. Nanyun Peng.
- Used PRAW to scrape titles from different subreddits and created a large dataset with 400k samples and the task is to classify them on the basis of titles. Provided baselines based on TF-IDF and [ULMFiT](#). Published as a [Kaggle Dataset](#)
- Mar'18-May'18 **Inference Networks for Structured Prediction Energy Networks**,
Course Project [[Report](#)] [[Code](#)],
Advisor: Prof. Sunita Sarawagi.
- A tensorflow implementation to perform multi-label classification experiments from the paper Learning Approximate Inference Networks for Structured Prediction. Performed further analysis on the knowledge base FIGMENT dataset. Further extended the paper to include WGAN like training for more stable training.
- Aug'17-Nov'17 **Voice Conversion using GANs**, *Course Project* [[Report](#)] [[Code](#)],
Advisor: Prof. Preethi Jyothi.
- Participated in the Voice Conversion Challenge 2018 in both parallel and non-parallel corpus track. Showed superior performance of voice synthesis using GANs and sentence embeddings of transcriptions and achieved mean opinion score of 3.8 which was a significant improvement over a simple Variational AutoEncoder.
- Aug'17-Nov'17 **Learning to Run:NIPS Challenge 2017**, *Course Project* [[Report](#)] [[Code](#)],
Advisor: Prof. Shivaram Kalyanakrishnan.
- Explored standard algorithms from Deep Deterministic Policy Gradient to Proximal Policy Optimizations on the open simulator challenge to provide correct activation functions for the muscles of the skeleton allowing it to run. Our best model achieved score of 32 points. Also proposed a solution using reward shaping and made ablation studies on the effect of parameter noise and layer normalization at train time.
- Aug'17-Nov'17 **Neural Networks with Memory**, *Course Project* [[Code](#)],
Advisor: Prof. Ganesh Ramakrishnan.
- Designed an OpenAI gym maze environment where the agent needs to collect a key and then open a door (a toy version of Montezuma's Revenge) with varying maze sizes. Experimentally verified that only deep networks with explicit memory (Memory Q-Network and Recurrent Memory Q-Network) learn the optimal policy for a 7x7 grid.
- Mar'17-Apr'17 **Image Noise Modeling via Skellam Distribution**,
Course Project [[Report](#)] [[Code](#)],
Advisor: Prof. Ajit Rajwade.
- Modeled the noise characteristics of a DSLR-camera via difference based imaging of a colored chart. Used the model to perform background subtraction and edge detection and verified its usability in the wild.
- Oct'16-Nov'16 **Document Scanner via Image Stitching**, *Course Project* [[Report](#)] [[Code](#)],
Advisor: Prof. Ajit Rajwade.
- Stitched images of overlapping parts of a large document to produce a high resolution picture. Implemented Homography Transformation, Multi-View Blending and Bundle Adjustment from scratch in python.
- May'15-Jul'15 **Virtual Reality Controller and Headset**, *Technical Project* [[Code](#)] [[Video](#)].
- Designed a low cost (~ \$20) headwear and controller for virtual reality applications. Demoed it on the popular game of Counter-Strike. This won the Best Idea Project 2016 in the Technical and Research Exposition 2015, IIT Bombay.

Relevant Courses

Computer Vision
Optimization

Natural Language Processing
Information Extraction

Reinforcement Learning
Advanced ML

Extra-curricular Activities

- Apr'16-Apr'17 **Web Manager**, *Student Technical Activities Body*, IIT Bombay.
 - Developed and Managed the Website of STAB IIT Bombay with other web applications.
 - Helped in development of OKSP (Online Knowledge Sharing Platform).
- June'15-Apr'16 **Convener Robotics Club**, *Student Technical Activities Body*, IIT Bombay.
Conducted sessions on robotics and talks on Augmented and Virtual Reality
- Aug'15-Apr'16 **Course in Japanese Language**, *IIT Bombay*.
Learnt how to read and write Katakana, Hiragana and basic Kanji and helped in organization of Japanese events.
- May'15-Jun'15 **Android App Workshop**, *IIT Bombay*.
Attended 6 lectures on developing an android app from scratch.
- May'15-Jun'15 **Entrepreneurship Bootcamp**, *IIT Bombay*.
Attended 6 lectures from distinguished entrepreneurs on how to build a successful startup.
- June'14-Apr'15 **Volunteer**, *National Service Scheme*, IIT Bombay.
Taught under privileged children in the Phulenagar Slum (3 hours per week).