Karan Sikka

Senior AI Scientist, SRI International

Email: karan.sikka1@gmail.com **Web:** `http://ksikka.com`

Google Scholar: Link

Professional Experience

- Senior Scientist
  - Center for Vision Technologies
  - SRI International, Princeton, USA
  - Feb'22-Present
- Scientist
  - Center for Vision Technologies
  - SRI International, Princeton, USA
  - Sept'16-Feb'22
- Student Associate
  - Vision and Learning Group
  - SRI International, Princeton, USA
  - Jan'16-Mar'16
- Student Intern
  - Augmented Reality Team
  - Qualcomm R&D, San Diego, USA
  - Jun'11-Sept'11
- Summer Intern
  - Medical Informatics Lab, RWTH Aachen, Germany
  - May'09-July'09

Education

- *PhD, ECE*, 2016
  University of California, San Diego
  - *Advisor: Dr. Marian S. Bartlett*
  - *Thesis:* Latent Dynamic Space-Time Volumes for Predicting Human Facial Behavior in Videos
- *MS, ECE*, 2012
  University of California, San Diego
- *B.Tech., ECE*, 2010
  Indian Institute of Technology Guwahati

Publications

- Sahu, P., Sikka, K., and Divakaran, A. (2024). Pelican: Correcting Hallucination in Vision-LLMs via Claim Decomposition and Program of Thought Verification (available soon).
- Chen, Y., Sikka, K., Cogswell, M., Ji, H., and Divakaran, A. (2024). Measuring and improving chain-of-thought reasoning in vision-language models (NAACL).
- Rajvanshi, A., Sikka, K., Lin, X., Lee, B., Chiu, H. P., and Velasquez, A. (2024). Saynav: Grounding large language models for dynamic planning to navigation in new environments (ICAPS).

- Chen, Y., Sikka, K., Cogswell, M., Ji, H., and Divakaran, A. (2024). Dress: Instructing large vision-language models to align and interact with humans via natural language feedback (CVPR).

- Sur, I., Sikka, K., Walmer, M., Koneripalli, K., Roy, A., Lin, Xiao, Divakaran, A. and Jha, S. (2023). TIJO: Trigger Inversion with Joint Optimization for Defending Multimodal Backdoored Models (ICCV)

- Jin, Y., Lee, Y. C., Sharma, K., Ye, M., Sikka, K., Divakaran, A., and Kumar, S. (2023). Predicting information pathways across online communities (KDD).

- Gwilliam, M., Cogswell, M., Ye, M., Sikka, K., Shrivastava, A. and Divakaran, A., (2023). A Video is Worth 10,000 Words: Training and Benchmarking with Diverse Captions for Better Long Video Retrieval (ArXiv).

- Ye, M., Sikka, K., Atwell, K., Hassan, S., Divakaran, A. and Alikhani, M. (2023). Multilingual Content Moderation: A Case Study on Reddit (EACL).

- Walmer, M., Sikka, K., Sur, I., Shrivastava, A., and Jha, S. (2022). Dual-Key Multimodal Backdoors for Visual Question Answering (CVPR).

- Sahu, P., Sikka, K., and Divakaran, A. (2022). Challenges in Procedural Multimodal Machine Comprehension: A Novel Way To Benchmark (WACV).

- Sahu, P., Sikka, K., and Divakaran, A. (2021). Towards solving multimodal comprehension (ArXiv).

- Kiourti, P., Li, W., Roy, A., Sikka, K., and Jha, S. (2021). MISA: Online Defense of Trojaned Models using Misattributions (ACSAC).

- Cadigan, J., Sikka, K., Ye, M., and Graciarena, M. (2021). Resilient Data Augmentation Approaches to Multimodal Verification in the News Domain (ICCV Workshops).

- Sikka, K., Sur, I., Jha, S., Roy, A., Divakaran, A. (2020). Detecting Trojaned DNNs Using Counterfactual Attributions (ArXiv).

- Sikka, K., Huang, J., Silberfarb, A., Nayak, P., Rohrer, L., Sahu, P., Byrnes, J., Divakaran, A., Rohwer, R. (2020). Zero-Shot Learning with Knowledge Enhanced Visual Semantic Embeddings (ArXiv).

- Mithun, N., Sikka, K., Chiu, H., Samarasekera, S., Kumar, R. (2020). RGB2LIDAR: Towards Solving Large-Scale Cross-Modal Visual Localizations (ACMM). (**Best Paper Candidate**)

- Sikka, K., Silberfarb, A., Byrnes, J., Sur, I., Chow, E., Divakaran, A., Rohwer, R. (2020). Deep Adaptive Semantic Logic (DASL): Compiling Declarative Knowledge into Deep Neural Networks (ArXiv).

- Kruk, J., Lubin, J., Sikka, K., Lin, X., Jurafsky, D., Divakaran, A. (2019). Integrating Text and Image: Determining Multimodal Document Intent in Instagram Posts. Conference on Empirical Methods in Natural Language Processing (EMNLP).

- Ray, A., Sikka, K., Divakaran, A., Lee, S., Burachas, G. (2019). Sunny and Dark Outside?! Improving Consistency in VQA through Entailed Question Generation. Conference on Empirical Methods in Natural Language Processing (EMNLP).

- Datta, S., Sikka, K., Roy, A., Ahuja, K., Parikh, D., Divakaran, A. (2019). Align2Ground: Weakly Supervised Phrase Grounding Guided by Image-Caption Alignment. International Conference on Computer Vision (ICCV).

- Kaur, P., Sikka, K., Wang, W., Belongie, S., Divakaran, A. (2019). FoodX-251: A Dataset for Fine-grained Food Classification. Computer Vision and Pattern Recognition Workshop (CVPRW).

- Seymour, Z., Sikka, K., Chiu, H., Samarasekera, S., Kumar, T. (2019). Semantically-Aware Attentive Neural Embeddings for Image-based Visual Localization. British Machine Vision Conference (BMVC).

- Sikka, K., Bramer L., Divakaran, A. (2019). Deep Unified Multimodal Embeddings for Understanding both Content and Users in Social Media Networks(ArXiv).

- Bansal, A., Sikka, K., Sharma, G., Chellapa, R., Divakaran, A. (2018). Zero-Shot Object Detection (ECCV).

- Ahuja, K., Sikka, K., Roy, A., Divakaran, A. (2018). Understanding Visual Ads by Aligning Symbols and Objects using Co-Attention (CVPRW).

- Kaur, P., Sikka, K., Divakaran, A. (2017). Combining Weakly and Webly Supervised Learning for Classifying Food Images (ArXiv).

- Kar, A., Rai, N., Sikka, K., Sharma, G. (2017). AdaScan: Adaptive Scan Pooling in Deep Convolutional Neural Networks for Human Action Recognition in Videos (CVPR).

- Sikka, K., Sharma, G. (2017). Discriminatively Trained Latent Ordinal Model for Video Classification (IEEE PAMI).

- Malmir, M., Sikka, K., Forster, D., Fasel, I., Movellan, J., Cottrell, W, Garrison. (2017). Deep Active Object Recognition by Joint Label and Action Prediction (CVIU).

- Sikka, K., Sharma, G., Bartlett, M. (2016). LOMo: Latent Ordinal Model for Facial Analysis in Videos (CVPR). (**Spotlight Presentation**)

- Malmir, M., Sikka, K., Forster, D., Movellan, J., Cottrell, W, Garrison. (2015). Deep Q-learning for Active Recognition of GERMS: Baseline Performance on a Standardized Dataset for Active Learning (BMVC). (**Acceptance Rate: 33%**)

- Sikka, K., Giri, R., Bartlett, M. (2015). Joint Clustering and Classification for Multiple Instance Learning (BMVC). (**Acceptance Rate: 33%**)

- Sikka, K., Dhall, A., Bartlett, M. (2015). Exemplar Hidden Markov Models for Classification of Facial Expressions in Videos (CVPRW).

- Sikka, K., Ahmed, A., Diaz, D., Goodwin, M., Craig, K., Bartlett, M., Huang, J. (2015). Automated Assessment of Children's Post-Operative Pain Using Computer Vision (Pediatrics). (**Impact Factor: 5.3**)

- Dhall, A., Joshi, J., Sikka, K., Goecke, K. and Sebe, N. (2015). The More the Merrier: Analysing the Affect of a Group of People In Images (IEEE FG).

- Sikka, K. (2014). Facial Expression Analysis for Estimating Pain in Clinical Settings (ICMI, Doctoral Consortium).

- Dhall, A., Goecke, R., Joshi, J., Sikka, K. and Gedeon, T. (2014). Emotion recognition in the wild challenge 2014: Baseline, data and protocol (ICMI).

- Sikka, K., Dhall, A. and Bartlett, M. (2014). Weakly Supervised Pain Localization and Classification with Multiple Segment Learning. The Best of Face and Gesture 2013 (Image and Vision Computing). (**Impact Factor: 1.6**)

- Dhall, A., Sikka, K., Littlewort, G., Goecke, R. and Bartlett, M. (2014). A Discriminative Parts Based Model Approach for Fiducial Points Free and Shape Constrained Head Pose Normalisation In The Wild (WACV). (**Acceptance Rate: 40%**)

- Sikka, K., Dykstra, K., Sathyanarayana, S., Littlewort, G. and Bartlett, M. (2013). Multiple Kernel Learning for Emotion Recognition in the Wild. ICMI'13. (**Best Paper Award**)

- Sikka, K., Dhall, A., and Bartlett, M. (2013). Weakly Supervised Pain Localization using Multiple Instance Learning. IEEE FG'13. (**Best Student Paper Honorable Mention Award**)('Acceptance Rate (Oral): 12%)

- Sikka, K., Wu, T., Susskind, J., and Bartlett, M. (2012). Exploring Bag of Words Architectures in the Facial Expression Domain (ECCV Workshops). ('Oral Presentation, Acceptance Rate: 33%)

- Singh, P. K., Sinha, N., Sikka, K., and Mishra, A. K. (2011). Texture information-based hybrid methodology for the segmentation of SAR images. International Journal of Remote Sensing (Taylor and Francis).

- Sikka, K., and Deserno, T. M. (2010). Comparison of algorithms for ultrasound image segmentation without ground truth (SPIE Medical Imaging).
- Sikka, K., Sinha, N., Singh, P. K., and Mishra, A. K. (2009). A fully automated algorithm under modified FCM framework for improved brain MR image segmentation (Magnetic resonance imaging).

Honors

- **Best Student Paper Honorable Mention Award** at IEEE International Conference on Automatic Face and Gesture Recognition 2013 (IEEE FG).
- **Best Paper Award** at EmotiW'13 Challenge, International conference on Multimodal Interaction 2013 (ICMI).
- Runner's up at EmotiW'13 Challenge, International conference on Multimodal Interaction 2013 (ICMI).
- Awarded a travel grant at International Conference on Multimodal Interaction 2014 (ICMI) (Doctoral Consortium).
- Jacobs Scholarship UCSD- three year fellowship- highest form of recognition for any PhD candidate in ECE Dept.
- Awarded travel grant at IEEE Automatic Face and Gesture Recognition, 2013 (IEEE FG) (Doctoral Consortium).
- Awarded an SPIE Contingency Student Travel Grant for SPIE Medical Conference, 2010.
- DAAD- German Academic Exchange Service, Undergraduate Internships, 2009.

Workshop/Challenge Organizer

- Organizer, iFood 2018 Challenge, FGVC Workshop, CVPR 2018, Salt Lake City, USA
- Organizer, Emotion Recognition In The Wild Challenge and Workshop (EmotiW 2014), ICMI 2014

Conference Reviewer/Program Committee

- ICML, Neurips, ICLR, AAAI, ECCV, CVPR, ICCV, WACV, ACCV, ACII, ICMI
- ACMMM (2019, 2020 Area Chair)

Journals Reviewer

- Transactions on Pattern Analysis and Machine Intelligence
- Transactions on Image Processing
- Transactions on Neural Networks
- Transaction on Affective Computing
- Transaction on Multimedia
- Computer Vision and Image Understanding
- International Journal on Computer Vision
- Pattern Recognition Letters
- Image and Vision Computing

Technical Talks

- Invited speaker at ICGVIP conference, 2020
- Invited speaker at Madima workshop, IJCAI 2018
- Computer Vision and Pattern Recognition Conference (CVPR), Salt Lake City, USA, 2018

- Arizona State University, Phoenix, USA, 2018
- Computer Vision and Pattern Recognition Conference (CVPR), Las Vegas, USA, 2016
- Indian Institute of Technology Kanpur (IIT Kanpur), India, 2016.
- Max Plank Institute of Informatics, Saarbrueken, Germany, 2016
- Stanford Research Institute (SRI) International, Princeton, USA, 2016
- Computer Vision and Pattern Recognition Conference (CVPR), Boston, USA, 2015
- Imperial College London (ICL), London, UK, 2015
- International Conference on Multimodal Interaction (ICMI), Istanbul, Turkey, 2014
- Automatic Face and Gesture Recognition (IEEE FG), Shanghai, China, 2013
- European Conference on Computer Vision (ECCV), Florence, Italy, 2012
- University of California San Diego, San Diego, USA, 2012

Funding Agencies

List of agencies that have funded my work at SRI International with details of my role

- DARPA
  - PI, Civil Sanctuary link
  - Co-PI, Modeling Influence Pathways link
  - Technical contributor, Knowledge Management at Scale and Speed link
  - Technical contributor, Semantic Forensics link
  - Technical Lead, SocSim (SBIR) link
- AFRL
  - Technical Lead, AFRL Mesa
- IARPA
  - Technical Lead, Trojans in Artifical Intelligence, link

June 20, 2024