

Sai Kumar Dwivedi January 11, 1993

Contact: saidwivedi@gmail.com • +49-152-54696873

Webpage: saidwivedi.in • G-Scholar: [Profile](#) • Github: github.com/saidwivedi • LinkedIn: [saidwivedi](#)

Education

International Max Planck Research School for Intelligent Systems PhD in Computer Science Supervisors: Dr. Michael J. Black, Dr. Dimitris Tzionas Collaborator: Dr. Cordelia Schmid Research Topic: Perceiving humans interacting with objects from images or videos.	2021 - Present
National Institute of Technology, Rourkela Bachelors & Masters in Computer Science and Engineering	2011 - 2016

Experience

Mercedes-Benz R&D India <i>Computer Vision Researcher</i> 2D Human Pose Estimation for Smart Car Interiors Developed a novel real-time headpose estimation algorithm from a single monocular IR camera which was later deployed as Side Mirror Selection in 2021 Mercedes EQS and Rear Sunblind Control in 2021 Mercedes S-Class.	SEP 2017 - FEB 2020
Intel Corporation <i>Machine Learning Engineer</i> Deep Learning Algorithms for Edge Devices Developed Binary and Ternary CNNs similar to XNOR-Net, addressing performance issues in deeper networks by implementing innovative quantization techniques, optimizing weight representation for efficiency, particularly beneficial for edge device applications.	JAN 2016 - AUG 2017

Selected Publications

- **TokenHMR: Advancing Human Mesh Recovery with a Tokenized Pose Representation**
[S.K. Dwivedi*](#), [Y. Sun*](#), [P. Patel](#), [Y. Feng](#), [M.J. Black](#) (* equal contribution)
In Proceedings of Computer Vision and Pattern Recognition (CVPR 2024)
TokenHMR addresses the paradox of declining 3D accuracy of HPS methods with increasing 2D precision by introducing a Threshold-Adaptive Loss Scaling (TALS) loss and reformulating the problem as token prediction.
- **ChatPose: Chatting about 3D Human Pose**
[Y. Feng](#), [J. Lin](#), [S.K. Dwivedi](#), [Y. Sun](#), [P. Patel](#), [M.J. Black](#)
In Proceedings of Computer Vision and Pattern Recognition (CVPR 2024)
ChatPose integrates large language models to comprehend and reason about 3D human poses from images or textual descriptions, leveraging world knowledge and body language.
- **POCO: 3D Pose and Shape Estimation using Confidence**
[S.K. Dwivedi](#), [C. Schmid](#), [H. Yi](#), [M.J. Black](#), [D. Tzionas](#)
In Proceedings of International Conference on 3D Vision (3DV 2024), Oral
POCO is a novel framework that can be applied to common human pose and shape regressors, extending them to estimate the method's confidence in the result without any downside.
- **Detecting Human-Object Contact in Images**
[Y. Chen](#), [S.K. Dwivedi](#), [M.J. Black](#), [D. Tzionas](#)
In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR 2023)
HOT tackles the lack of a reliable approach for detecting human-object 2D contact in images by introducing a dataset of 2D contacts and developing a contact detector guided by part-attention.
- **Learning to Regress Bodies from Images using Differentiable Semantic Rendering**
[S.K. Dwivedi](#), [N. Athanasiou](#), [M. Kocabas](#), [M.J. Black](#)
In Proceedings of International Conference on Computer Vision (ICCV 2021)
DSR introduces a novel Differentiable Semantic Rendering loss that utilizes semantic clothing information to improve 3D human body estimation and thus surpassing prior state-of-the-art methods.

Research Internship

- Max Planck Institute for Intelligent Systems** | *Research Assistant* JUL 2020 - AUG 2021
3D Human Pose and Shape Estimation
Explored human pose and shape estimation from images using weak supervisory signals like clothing segmentation and formulated novel semantic rendering loss.
- Tallinn University of Technology** | *Research Exchange Student* SEP 2014 - JUN 2015
Digital System Design for AI
Developed machine learning algorithms for edge devices. One such notable work was designing a neural network based face-recognition system on FPGA.
- Indian Institute of Technology, Gandhinagar** | *Research Intern* APR 2014 - JUN 2014
Virtual Reality Modules for Autism
Designed virtual reality modules for physiology-based affect sensitive social communication system for children with an autism spectrum disorder. The system adapted according to the response and recommended corrective modules for improved learning.
-

Other Publications

- **ProtoGAN: Towards Few Shot Learning for Action Recognition**
S.K. Dwivedi, V. Gupta, R. Mitra, S. Ahmed, A. Jain
In Proceedings of International Conference on Computer Vision Workshops (ICCVw 2019)
ProtoGAN framework addresses the challenge of few-shot learning or FSL for action recognition and generalized FSL by synthesizing additional examples for novel categories using class prototype vectors.
 - **Out-of-Distribution Detection for Generalized Zero-Shot Action Recognition**
D. Mandal, S. Narayan, S.K. Dwivedi, V. Gupta, S. Ahmed, F. Khan, L. Shao
In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR 2019)
Our novel framework incorporates an out-of-distribution detector for generalized zero-shot action recognition to distinguish between seen and unseen action categories, achieving significant improvements over existing methods.
 - **Progression Modelling for Online and Early Gesture Detection**
V. Gupta, S.K. Dwivedi, S. Ahmed, A. Jain
In Proceedings of the International Conference on 3D Vision (3DV 2019), Oral
Our simple yet effective multi-task learning framework addresses the issue of online and early gesture detection by modelling the gesture progression along with frame level recognition.
-

Awards and Honours

- IMPRS-IS Scholar - 2022:** Max Planck Institute for Intelligent Systems, Germany
RA Fellowship - 2021: Max Planck Institute for Intelligent Systems, Germany
Outstanding Performer Award - 2018: Mercedes-Benz R&D, India
Employee Recognition - 2017: Intel Corporation, India
Student Fellowship - 2015: Erasmus Mundus European Union Program, Estonia
Student Fellowship - 2014: IIT Gandhinagar, India
-

Academic Service

- CVPR Reviewer:** 2022, 2023
ECCV/ICCV Reviewer: 2022, 2023, 2024
3DV Reviewer: 2021, 2022, 2023
-

Patents

- Intelligent Airbag Deployment using Head Pose Estimation, *Filled*
Progression Modelling for Online and Early Gesture Detection, *Filled*
Hand-pose estimation from IR camera inside car, *Filled*
-

Skills

- PyTorch • Torch
 - Python • Lua • C++ • C
-