

# Somdyuti Paul

+91 7752894776 ✉ somdyuti@cai.iitrkgp.ac.in

## Current Position

---

**Indian Institute of Technology Kharagpur** **Kharagpur, WB, India**  
*Assistant Professor, Centre of Excellence in Artificial Intelligence* 2023 - Present

## Education

---

**The University of Texas at Austin** **Austin, TX, USA**  
*Doctor of Philosophy, Electrical and Computer Engineering* 2017 - 2022  
Advisor: Dr. Alan Bovik  
GPA: 3.8325/4.0

**Indian Institute of Technology Kanpur** **Kanpur, UP, India**  
*Master of Technology, Electrical Engineering* 2013 - 2015  
Advisor: Dr. Sumana Gupta  
GPA: 8.75/10.0

**West Bengal University of Technology** **Kolkata, WB, India**  
*Bachelor of Technology, Electronics and Communication Engineering* 2008 - 2012  
GPA: 9.12/10.0

## Professional Experience

---

**Amazon.com Services, Inc.** **Seattle, WA, USA**  
*Applied Scientist, Amazon Prime Video* 2022 - 2023

**Amazon.com Services, Inc.** **Seattle, WA, USA**  
*Applied Scientist Intern, Amazon Prime Video* Summer 2021

**Netflix, Inc.** **Los Gatos, CA, USA**  
*Research Scientist Intern, Video Algorithms Group* Summer 2019 and Summer 2018

**Tata Consultancy Services** **Noida, UP, India**  
*Systems Engineer, Multimedia, Graphics and Robotics Group* 2015-2017

## Research Experience

---

**University of Texas at Austin**

**Austin, TX, USA**

*Graduate Research Assistant*

*2017 - 2022*

- Developed a convolutional recurrent neural network based model to predict the convex hull representations of video shots, which was trained using a two-step transfer learning procedure. This reduces the pre-encoding time required to generate content-aware bitrate ladders for adaptive video streaming. Pre-encoding time was reduced significantly while the predicted convex hulls closely approximated the ground truth convex hulls.
- Developed a neural network based search free motion estimation model to generate block motion vector predictions. A hierarchical spatial transformer model was trained in a self supervised manner to give the motion vectors of varying block sizes in a single step. The approach outperformed conventional block matching based motion estimation algorithms in terms of speed, while maintaining comparable prediction accuracy and also improved the rate-distortion performance of the state of the art AV1 encoding when used to replace the search based motion estimation process in the SVT-AV1 encoder.
- Developed a block-based adaptive quantization method for the AV1 video codec to enhance the perceptual quality of the encoded videos. In order to exploit the property of contrast masking of the human visual system, a spatial patch based predictor of the contrast masking threshold was devised using steerable filtering. Subsequently, the video frames were segmented into different regions based on the predicted contrast masking thresholds. The value of the quantization parameter to be applied on each segment was chosen according to its predicted level of contrast masking to improve the perceptual quality and rate-distortion performance of the resulting encodes obtained using the SVT-AV1 encoder.
- Developed a deep learning based technique to speed up video encoding in the intra-mode for the VP9 encoder. A hierarchical fully convolutional neural network was employed to predict the intra-mode superblock partitions in a bottom-up fashion. Significant reduction in encoding time was achieved at the expense of a smaller decline in rate-distortion performance as compared to the fastest recommended speed setting of the reference VP9 encoder.
- Explored the applicability of a no-reference image quality metric called natural image quality evaluator (NIQE) to predict the quality of geophysical well-log images. By training the NIQE model on pristine well logs images, well logs with unacceptable image qualities for data retrieval were accurately classified from the ones having acceptable qualities.

**Amazon.com Services, Inc.**

**Seattle, WA, USA**

*Applied Scientist*

*2022-2023*

- Developed a visual attention based model to detect the presence of open mouths of characters in movie frames, which was used as a content curation tool for rejecting frames having such attributes to promote selection of aesthetically appealing frames for cover art generation.
- Developed a deep learning based frame selection model for movie frames based on the sharpness of salient regions. The model, trained on a synthetic blur dataset, learned to identify the presence of motion blurs and defocus blurs, as well as their relative location in the foreground and background of the frames. Frames having foreground motion blur were rejected while frames having sharp foreground and

defocus blurred background were favored for the purpose of cover-art generation.

**Amazon.com Services, Inc.**

**Seattle, WA, USA**

*Applied Scientist Intern*

*Summer 2021*

- Developed a recurrent neural network based framework for generating key action highlights from broadcast soccer videos. A bidirectional gated recurrent unit (GRU) based model was trained on a large database comprising hundreds of soccer games annotated with action timestamps. The trained model was able to detect 13 key soccer actions within 20 second video chunks to generate highlight clips, outperforming the precision-recall performance of concurrent action spotting models when applied to the task.

**Tata Consultancy Services Research**

**Noida, UP, India**

*Systems Engineer*

*2015 - 2017*

- Developed a deterministic policy gradient based actor-critic deep reinforcement learning framework for path planning of a robotic manipulator to reach a target object in a simulated 3D environment. A convolutional neural network (CNN) trained on a stereo image pair was used to predict the optimal actions in the form of efforts applied to the robots joints to reach the target in a continuous action space. The trained agent developed an implicit perception of depth of the target object, thereby succeeding in effectively guiding the manipulator to reach the target irrespective of the target position, or the initial robot pose. Unlike existing approaches, no additional cues such as state of the robot in the form of joint angles, or the 3D coordinates of the target were required.

**Indian Institute of Technology Kanpur**

**Kanpur, UP, India**

*Graduate Student*

*2014 - 2015*

- Developed an approach to colorize grayscale videos as a volume, relying on the motion and texture cues obtained from three dimensional steerable pyramid decomposition of a video shot. Scribble colors were bidirectionally propagated from the keyframes across spatiotemporal patches over the entire volume, according to the priorities dictated by the dominant orientation response of the steerable pyramid at each pixel. The colorization results were robust to occlusion, and the method significantly improved accuracy as well as computational cost of video colorization by dispensing with the computation and usage of motion vectors to transfer colors between adjacent frames.

## Publications

---

### Journals - published:

- **Somdyuti Paul**, Andrey Norikin and Alan C. Bovik, "Self-Supervised Learning of Perceptually Optimized Block Motion Estimates for Video Compression," *IEEE Transactions on Image Processing* (Early Access), December 2022.
- **Somdyuti Paul**, Andrey Norikin and Alan C. Bovik, "On Visual Masking Estimation for Adaptive Quantization Using Steerable Filters," *Signal Processing: Image Communication*, vol. 96, pp. 116290, August 2021.
- Dae Y. Lee, **Somdyuti Paul**, Christos G. Bampis, Hyunsuk Ko, Jongho Kim, Se Y. Jeong, Blake Homan and Alan C. Bovik, "A Subjective and Objective Study of

- Space-Time Subsampled Video Quality," *IEEE Transactions on Image Processing*, vol. 31, pp. 934-948, December 2021.
- **Somdyuti Paul**, Andrey Norkin and Alan C. Bovik, "Speeding up VP9 Intra Encoder with Hierarchical Deep Learning Based Partition Prediction," *IEEE Transactions on Image Processing*, vol. 29, pp. 8134-8148, July 2020.
- **Somdyuti Paul** and Alan C. Bovik, "Image Statistic Models Characterize Well Log Image Quality," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1130-1134, February 2019.
- **Somdyuti Paul**, Saumik Bhattacharya and Sumana Gupta, "Spatiotemporal Colorization of Video Using 3D Steerable Pyramids," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 8, pp. 1605-1619, August 2017.
- **Somdyuti Paul**, Saumik Bhattacharya and Sumana Gupta, "Selection of Keyframes for Video Colorization using Steerable Filtering," *Sadhana*, vol 42, no. 10, pp. 1685-1692, September 2017.

**Journals - in review:**

- **Somdyuti Paul**, Andrey Norkin and Alan C. Bovik, "Efficient Per-Shot Convex Hull Prediction By Recurrent Learning," arXiv preprint arXiv:2206.04877 (submitted in *IEEE Transactions on Image Processing*).

**Conferences:**

- **Somdyuti Paul** and Lovekesh Vig, "Deterministic Policy Gradient Based Robotic Path Planning with Continuous Action Spaces," *IEEE International Conference on Computer Vision Workshops*, October 2017.

## Teaching Experience

---

<b>Indian Institute of Technology Kharagpur</b>	<b>Kharagpur, WB, India</b>
Course instructor for Deep Learning Foundations and Applications	Spring 2024
Course instructor for Machine Learning Foundations and Applications	Spring 2024

## Academic Honors and Achievements

---

George J. Heuer, Jr. Ph.D. Endowed Graduate Fellowship from Cockrell School of Engineering, University of Texas at Austin (2017-2018).

## Technical Skills

---

**Programming languages:** Python, C, MATLAB

**Software packages and libraries:** Keras, Tensorflow, PyTorch, Scikit-learn, libvpx, SVT-AV1, FFmpeg, OpenCV, ROS, Gazebo

## Professional Activities

---

Reviewer (IEEE Transactions in Image Processing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Journal on Selected Areas in Communication, Signal Processing: Image Communication)