

SWAPNIL MALLICK

Los Angeles, CA, USA

✉ smallick@usc.edu || 📞 +1 (213) 245-3649 || [🌐 LinkedIn](#) || [🌐 Homepage](#)

EDUCATION

University of Southern California, Los Angeles

Aug 2021 - Dec 2023

Master of Science, Computer Science

Overall GPA: 3.56/4

Maulana Abul Kalam Azad University Of Technology (WBUT), India

Jun 2016 - Jun 2020

Bachelor of Technology, Computer Science and Engineering

Overall GPA: 9.03/10

RESEARCH EXPERIENCE

University of North Carolina at Chapel Hill

Remote

UNC CV Group

March 2024 - Present

- Supervisor: Prof. Roni Sengupta ([Google Scholar](#))
- Currently collaborating on detecting student video lecture engagement and live deep fakes with passive monitor illumination
- Implemented a personalized ViT backbone architecture which shows a significant improvement in performance (74.95%) when compared with the previous architecture (62.05%) in dark lighting conditions
- Experimenting with diffusion models and SDS loss (per-image optimization) to improve the monitor image prediction quality
- Exploring the application of self and cross attention based architectures in detecting video engagement and live deep fakes with the ultimate goal of building a generalized model for the task

University of Southern California

Los Angeles, California

USC CPS-VIDA (Cyber-Physical Systems - Verification, Intelligence, Design, and Analysis)

March 2024 - Present

- Supervisor: Prof. Jyotirmoy Vinay Deshmukh ([Google Scholar](#))
- Currently working on a project on generating safety-critical scenarios for autonomous vehicle testing (**RV '24**)
- Explored Scenic to generate safety-critical scenarios
- Broadly exploring the application of generative models like, VQGAN, diffusion models for generating safety-critical scenarios for comprehensive self-driving car testing

University of Southern California

Los Angeles, California

USC CPS-VIDA (Cyber-Physical Systems - Verification, Intelligence, Design, and Analysis)

May 2023 - August 2023

- Supervisor: Prof. Jyotirmoy Vinay Deshmukh ([Google Scholar](#))
- Developed a safety monitoring framework for detecting the robustness of existing pedestrian detection models in low-light weather conditions using Timed Quality Temporal Logic (TQTL) (**RV '23**)
- Compared the robustness of vanilla YOLO and IA-YOLO in both foggy as well as night conditions
- Proposed TQTL as a quality metric that will help in debugging or improving the existing pedestrian detection models

Indian Statistical Institute

Kolkata, India

Computer Vision And Pattern Recognition (CVPR) Unit

July 2019 - May 2020

Certificate Link: <https://bit.ly/3S0wxYo>

- Supervisor: Prof. Ujjwal Bhattacharya ([Google Scholar](#))
- Designed three separate architectures for detecting handwritten Bengali numerals using Residual blocks, Inception blocks and LeNet-5 architecture
- Employed DCGAN to create additional samples to deal with the initial scarcity of training data
- Found that the simpler LeNet-5 architecture achieved a higher accuracy of $\sim 98.2\%$ on test data as compared to the architectures that used Inception block ($\sim 97.4\%$) and Residual block ($\sim 97.1\%$)

PUBLICATIONS

2. Vidisha Kudalkar, Navid Hashemi, Shilpa Mukhopadhyay, **Swapnil Mallick**, Christof Budnik, Parinitha Nagaraja, Jyotirmoy V. Deshmukh. Sampling-based and Gradient-based Efficient Scenario Generation. (**Accepted at the 24th International Conference on Runtime Verification, 2024**) ([Abstract](#)), ([Pre-print version](#))
1. **Swapnil Mallick***, Shuvam Ghosal*, Anand Balakrishnan, Jyotirmoy Deshmukh. [Safety monitoring for pedestrian detection in adverse conditions](#). 23rd International Conference on Runtime Verification, pp. 389-399. Springer Nature Switzerland, 2023. ([Paper](#))

ACADEMIC PROJECTS

Analysing Twitter Data on Rhino Conservation using Natural Language Processing

Project Report: <https://bit.ly/4eJ5mcX>

November 2023

- This study, conducted as group project for CSCI 461 at USC, was designed to demonstrate the utility of social media data in monitoring public opinion regarding illegal rhino poaching and trade
- Various toolkits, including Pattern, TextBlob, VADER, and Twitter-roBERTa-base, were utilized to gauge the sentiment expressed in the tweets; used SHAP values to interpret the behavior of the models/toolkits
- In parallel, topic modeling techniques such as LDA and BTM were applied to gain deeper insights into the prevalent discussion topics and identify the user categories participating in these conversations
- The exploratory data analysis conducted on rhino-related tweets revealed that the volume of such tweets exhibited a comparative increase from 2014 to 2019, reaching a peak in 2018
- For assessing the sentiment of tweets in conservation-related discussions, the results obtained did not provide a clear indication of the effectiveness of relying on pretrained NLP tools
- The topic modeling analysis revealed a dual nature of discussions on rhinos, encompassing both general themes such as 'Poaching' and 'Horn/Ivory Trade', as well as event-specific topics like the 'Last White Male Northern Rhino Extinction' and 'Horn Stealing/Veterinarian'

CHIP: Contrastive Hierarchical Image Pretraining

Project Report: <https://bit.ly/4en84EA>

May 2023

- This study, conducted in group of four students as the course project of CSCI 566 at USC, proposed a few-shot three-level hierarchical contrastive loss-based classification model that can classify an object of any unseen animal class into a relatively general category in a hierarchically based classification
- The proposed model was trained on a subset of the ImageNet (ILSVRC-12) dataset containing only the animal classes (366 classes in total) with 1300 images in each class
- The trained model was evaluated on a dataset created by us that contained 20 unseen animal classes not present in the ImageNet dataset
- The trained model, evaluated on 20 unseen animal classes, could classify 14 of them to their correct parent and super parent class

Adaptive Cruise Control System of Self-driving Car

Project Code Repository: <https://bit.ly/48UcXTI>

December 2022

- Designed an adaptive cruise control system of a self-driving car using YOLOv5 model for object detection and Monodepth2 for distance estimation
- The Monodepth2 model was used to output a disparity map of the bounding box region of the detected lead car from which the value at the centroid position of the bounding box was taken to get the estimated distance
- This was then fed to the PID controller of the ego car, which has to maintain a safe distance to the lead car
- The system was tested in CARLA simulator Town 6 and the ego car was found to always maintain a safe distance to the lead car

Stock Price Prediction of a Company using LSTM

Project Code Repository: <https://bit.ly/3MOREpN>

June 2022

- Leveraged an LSTM architecture to predict the stock price of a company
- Measured the performance of the model using RSME score
- Received highest RMSE score of 7.25 on unknown test data after running five times

A Multi-modal Approach to Speech Emotion Recognition

Project Report: <https://bit.ly/4gKNMqD> | Project Code Repository: <https://bit.ly/3BuQu3u>

May 2022

- This study, conducted as the course project of CSCI 544 at USC, uses audio and text as modalities from the IEMOCAP dataset for human emotion recognition
- A VGG frame architecture was trained from scratch to detect emotions from audio and a pretrained BERT was finetuned to predict the emotions from text transcriptions
- The text-only and audio-only architectures were combined to design a multimodal architecture for emotion recognition
- The multimodal architecture achieved a considerably higher accuracy of 76.9% as compared to the models that used either text (72.4%) or audio (47.8%) as the modality

Cipher Text Prediction using BiLSTM

Project Code Repository: <https://bit.ly/3QiJ5Jw>

March 2022

- Designed a binary classifier that can classify whether a piece of text is encoded or not
- Converted text into word embeddings and employed BiLSTM for classification task
- The classifier achieved $\sim 87.5\%$ accuracy on training data and $\sim 85\%$ accuracy on test data

COURSEWORK

- | | | |
|-----------------------|------------------------------------------------|-------------------------------------|
| • Machine Learning | • Deep Learning | • Autonomous Cyber-Physical Systems |
| • Applied NLP | • Information Retrieval and Web Search Engines | • AI for Sustainable Development |
| • Database Systems | • Foundations of AI | • Algorithms |
| • Operations Research | • Computer Networks | • Automata Theory |
| • Operating Systems | • Information and Coding Theory | • Numerical Methods |

SKILLS

- **Programming:** Python, C, C++
- **ML Frameworks:** Keras, TensorFlow, PyTorch, scikit-learn
- **Data Visualization:** Matplotlib, seaborn
- **Database:** MySQL, PostgreSQL, MongoDB
- **Software Tools:** Anaconda, Jupyter Notebook, L^AT_EX, CARLA Simulator, KNIME, WEKA, RapidMiner, Tableau, Gurobi Optimization

CERTIFICATIONS

- Deep Learning Specialization (DeepLearning.AI) ([Link](#))
 - ☐ Neural Networks and Deep Learning ([Link](#))
 - ☐ Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization ([Link](#))
 - ☐ Structuring Machine Learning Projects ([Link](#))
 - ☐ Convolutional Neural Networks ([Link](#))
 - ☐ Sequence Models ([Link](#))
- Natural Language Processing Specialization (DeepLearning.AI) ([Link](#))
 - ☐ Natural Language Processing with Classification and Vector Spaces ([Link](#))
 - ☐ Natural Language Processing with Probabilistic Models ([Link](#))
 - ☐ Natural Language Processing with Sequence Models ([Link](#))
 - ☐ Natural Language Processing with Attention Models ([Link](#))
- MATLAB (Vanderbilt University) ([Link](#))
- Bayesian Statistics (UC Santa Cruz) ([Link](#))
- Introduction to Big Data (UC San Diego) ([Link](#))
- Big Data Integration and Processing (UC San Diego) ([Link](#))
- Big Data Modeling and Management Systems (UC San Diego) ([Link](#))
- Machine Learning with Big Data (UC San Diego) ([Link](#))
- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning (DeepLearning.AI) ([Link](#))
- Convolutional Neural Networks in TensorFlow (DeepLearning.AI) ([Link](#))

BLOG POSTS

- A Survey On Data Driven Approaches To Scenario Generation For Autonomous Vehicle Testing
Link: https://swapnilmallick.github.io/blogs/scenario_generation.html
- A Guide To Generative Adversarial Network
Link: <https://swapnilmallick.github.io/blogs/gan.html>

The fate of our dream of an autonomous future largely depends on the safety and reliability of the autonomous systems we develop. Despite great advances in research and development of robotics and autonomous systems, there have been instances where these systems fail in real-world settings. Since most of them are safety-critical and mission-critical, building robust and reliable autonomous systems is of utmost importance to gain public trust before they are deployed in real-world settings. During my doctoral studies, I am looking for an opportunity to continue researching in the broad domain of autonomous driving, particularly scenario generation for safe and secure self-driving cars. I believe that exploring the idea of generating safety-critical scenarios using generative models, such as latent diffusion models, and examining the effectiveness of evolutionary algorithms like quality-diversity optimization algorithms in searching the latent space of these generative models to produce realistic and diverse scenarios for self-driving car testing could be an exciting area of research. Apart from that, creating a framework for open vocabulary video panoptic segmentation using text-2-video diffusion models is also another area that I am interested in exploring, which could significantly contribute to the perception module of the autonomy stack. Having said this, I am very flexible and also open to working on any other related problems that would allow me to use the skills I have learned so far and augment them further.

I have been fortunate enough to gain some valuable research experience across diverse domains of Computer Vision and Formal Methods with applications in Character Recognition, Safe and Secure Autonomous Systems, and Video Engagement and Deep fake Detection.

Character Recognition. During my first research project, I worked towards detecting handwritten Bengali numerals under the guidance of **Prof. Ujjwal Bhattacharya** at the CVPR Unit of the **Indian Statistical Institute**. Not having had any background in the field of deep learning and computer vision, I taught myself the basic concepts and, in turn, started reading papers related to document analysis and various neural network architectures. I designed three separate architectures - one mimicking the LeNet-5 network and the other two using Residual blocks and Inception blocks. Due to limited training data, synthetic data was generated using DCGAN. The findings from the experiments with the different architectures indicated that the simplest LeNet-5 architecture attained the highest accuracy of approximately 98.2% with the architectures using Residual and Inception blocks achieving lower accuracies of 97.1% and 97.4% respectively. In spite of being rudimentary, I became familiar with the research process through this project. I came to appreciate the importance of communicating my findings, and refined my ability to run experiments.

Safe and Secure Autonomous Systems. Robust perception algorithms are a vital component for safe and reliable operation of autonomous systems such as self-driving vehicles. While taking the Autonomous Cyber-Physical Systems course at USC, I came across **Prof. Jyotirmoy Deshmukh's** work on proposing Timed Quality Temporal Logic (TQTL) as a formal language to express spatio-temporal properties of a perception algorithm processing a video and use it as an alternative metric for testing the robustness of existing perception algorithms in the absence of ground truth labels. While their framing of the problem was insightful, I realized that evaluations had not been carried out in diverse and challenging weather conditions where there is a higher chance of the perception systems failing. Hence, I reached out to the authors of the paper and led a collaborative effort to discuss the idea and have a better understanding of the intricacies of their work. Under the guidance of Prof. Deshmukh and his doctoral students, I, along with a fellow graduate student, was able to propose TQTL as an able quality metric for differential testing of perception systems in adverse weather conditions as well. This work resulted in a publication that got accepted at the 23rd International Conference on Runtime Verification. This project gave me hands-on experience of building a collaboration to improve upon previous work. Starting with an intuitive idea, I was able to discuss, iterate, and improve on it until we reached a valid solution.

After finishing my master's degree at USC, I am currently continuing my research at USC under the guidance of Professor Deshmukh, focusing on scenario generation for comprehensive autonomous vehicle testing. Initially, I explored Scenic - a probabilistic language for safety-critical scenario

generation. After that, I am broadly studying the application of generative models like, diffusion models for generating safety-critical scenarios for self-driving car testing. It is allowing me to conduct research independently on data-driven approaches for safety-critical scenario generation.

Video Engagement and Deep Fake Detection. A significant issue with increased virtual communication is the struggle of instructors to understand student attentiveness in online lectures. Being an educator herself, my mother shared with me that students can engage in other activities like playing games, watching other videos etc. while turning on the webcam and looking straight at the monitor. Another major concern with widespread virtual communication is the rise of live deep fake avatars. However, both deep fake virtual avatars and students disengaging from class content often share one underlying physical phenomenon, i.e. the lighting on their faces does not match the lighting emitted from the monitor. Luckily I got the chance to work on this problem after I approached **Prof. Roni Sengupta** at **UNC-Chapel Hill**. During my collaboration with Prof. Sengupta, I improved the performance of the existing personalized ViT backbone architecture for a single lighting condition from roughly 62% to around 74.95% by making architectural changes. Apart from that, I have been exploring the possibility of using self and cross-attention-based architectures for personalized detection and experimenting with diffusion models and SDS loss (per-image optimization) to improve the monitor image prediction quality. The final objective of our research is to build a generalizable video engagement and live deep fake detection model that works in diverse lighting conditions. This project is helping me augment my knowledge of Computer Vision, improve my skills of communicating and presenting my findings, and how to approach a unique and unsolved problem. I am also learning how to approach a broad goal by identifying small, concrete sub-problems.

While at USC, I credited diverse courses, including **Autonomous Cyber-Physical Systems, Machine Learning, Deep Learning, Natural Language Processing, Information Retrieval, and AI for Sustainable Development**. I have also worked on a diverse range of projects as a graduate student at USC. I have participated in a project that attempts to demonstrate the utility of social media data in monitoring public opinion regarding illegal rhino poaching and trade. Apart from that, I have also engaged in projects related to multimodal approaches for human speech emotion recognition, designing an adaptive cruise control system for self-driving cars, and few-shot image classification to name a few.

If given the opportunity to pursue doctoral studies at Columbia University, I would be honored to conduct research under the supervision of **Prof. Baishakhi Ray, Prof. Yunzhu Li, Prof. Hod Lipson**, and **Prof. Aleksander Holynski**. I feel that my research interest matches with that of **Prof. Baishakhi Ray** given her previous contribution to developing a diffusion based framework, called CTG, for realistic and controllable traffic simulation using Signal Temporal Logic. I would also be very happy to be guided by **Prof. Yunzhu Li** whose research interests lies at the intersection of computer vision, robotics and machine learning and **Prof. Hod Lipson** who has made innumerable contributions to building evolutionary robots. Apart from that, I would like to be advised by **Prof. Aleksander Holynski**, whose research focus is in computer vision and deep generative models like diffusion models, once he joins the university.

Columbia University has a strong and diverse faculty body and close collaborations across research groups. Hence, I believe Columbia will provide me with an excellent environment and important collaborations for conducting doctoral studies. I feel that my previous research experience and alignment of research interests with faculty makes me a good fit for Columbia's graduate program.

“The race is not always to the swift, but to those who keep on running.”

The story of the slow, gritty and humble *tortoise* winning the race against a fast and complacent hare attempts to impart the same message. In the hustle and bustle of our daily lives, many of such principles - extremely simple yet profound - tend to wane with the breeze of life. However, this particular message made an indelible impression on the mind of a baby *tortoise* living close to the shores of the Bay of Bengal in India. He was awestruck to see the determination of the *old tortoise*, whom he considered as a father figure, to set an example to the rest of the world. In elementary school, an early encounter with a language called Logo changed the world for this baby *tortoise*. To see another turtle making geometric shapes on the screen made this young *tortoise* excited to venture into the unknown world of computers.

Over the course of time, this *tortoise* grew up to attend college in order to undergo formal training in the field of computer science and engineering. He took small steps at a time to learn the nitty gritty of the building blocks of a computer. Then, one day in the year 2018, the news of the trio of Yoshua Bengio, Geoffrey Hinton, and Yann LeCun winning the coveted Turing Award opened a new door in front of him. The effectiveness of deep neural networks as a key component of computing left him fascinated. He soon enrolled in an introductory course in artificial intelligence in his college to test his interest in the field. In the process, he taught himself Prolog and started writing simple rule-based programs. In no time he got immersed in the fast flowing stream of artificial intelligence and started researching the topic of character recognition. He started enjoying swimming in these waters and made up his mind to swim further.

Suddenly, the dark cloud of a global pandemic brought a storm and raised the water level. The *tortoise* saw everyone struggling to keep swimming with many succumbing to the high tides. With the storm abating, he finally managed to cross the ocean and reached the shores of Southern California after some delays and hindrances. In a land of valorant Trojans, the *tortoise* initially struggled miserably to find his footing. Everytime he tried swimming in the shores, he got washed away by the majestic surfing waves. Dejected and defeated, the *tortoise* felt as if he had fallen into a deep trench, as if he had lost the race. But then he received a valuable perspective from the *old tortoise*. The *old tortoise* said that when he ran the race he never viewed it as a competition with the hare. He treated it as a journey - he was just running his own race. He was just aiming to improve himself, to be better than what he was before and in turn testing his determination to finish a task which seemed impossible to the rest of the world. This brought a radical change in the way the *tortoise* perceived things. He stopped seeing failure as the direct opposite of success but as a precursor to success. He continued his diligent attempt to improve himself by trying not to repeat the same mistakes committed before. In this process, he came to this realization that the Trojans also had the same motto of succeeding but their way of portraying it was different. Slowly with continued efforts, the *tortoise* started to swim with the surfing waves along the shores of Southern California. In the meantime, he also found some gentle and considerate academic mentors who patiently guided and helped him in taking further steps in the world of research.

The *tortoise* dreams to serve the world of academia, remaining occupied in research as well as being the guiding light for his succeeding generations by sharing the little experience he would have gained till then. Besides being a research advisor, he wishes to be a good mentor for his students helping them navigate swiftly through the problems they would face in their daily lives because he believes that the productivity of students is proportional to their wellbeing. He feels pursuing doctoral studies will help him in multiple ways to attain his goals. First of all, he feels that he would gain first hand experience of teaching through teaching assistantship opportunities. He sees these opportunities as a way to refine his soft skills and at the same time test his time management and organizational skills. Apart from that, conducting research during doctoral studies will help him gain in-depth knowledge in his domain of interest which, in turn, will help him conduct research independently. To fulfill his dream, he wishes to continue his academic pursuit at Columbia University where students are equally passionate, resilient and focused towards achieving their goals.