Anurag **Bagchi**

(+412) 287 3504 | anuragba@andrew.cmu.edu | miccooper9 | miccooper9 | Website

Experience

Carnegie Mellon University — *Pittsburgh*

Research Assistant under Dr. Martial Hebert, RI-SCS — *Aug 2023 - Present*

- Leveraging **Text-to-Video** & **Text-to-Image Diffusion** models for **Video Understanding** & **Robotics** with **Toyota Research Institute**
- Authored **ReferEverything** (cool results **here**!), improving **Zero-shot Generalization** in videos from **objects** to ambiguous, dynamic **non-object concepts** (like waves in the ocean) by **32%**.

Bytedance AI Lab — *Sinapore*

Computer Vision Engineer, TikTok Brand Safety — *March 2023 - Aug 2023*

- I worked in **Prof. Song Bai's** team, where I have **launched 11 Multi-Modal** models using **Vision**, **Audio** and **text** for various challenging policy detection tasks and **automated** the iteration process for **21 ASR models**.
- My work directly **enhanced TikTok's ability to detect policy violations** on its platform at scale.

TikTok R&D Singapore — *Sinapore*

Machine Learning Engineer, Video & Push Recommendation — *March 2021 - Feb 2023*

- Created the entire ML pipeline to introduce and leverage **user-feedback** from notifications as a **training signal** for TikTok's Recommendation systems. **If you have used TikTok and received a Video notification with a feedback option, that was me**.
- Created new recommendation models and improved existing ones to significantly increase, **DAU (+0.07%)**, **user staytime(+3%)**, **click-through-rate (+15%)** and reduce **system latency(-3%)** and **memory-usage(-2.5%)**.
- End-to-End ML at scale, from problem formulation & feature engineering to model design , A/B testing and deployment.

IIIT Hyderabad — *Hyderabad, India*

Research Assistant under Dr. Ravi Kiran and Dr. Makarand Tapaswi, CVIT — *Oct 2020 - Feb 2021*

- Proposed and implemented the **first ever Audio-Visual** framework for **Temporal Action Localisation (TAL)** in Videos achieving new SoTA on **ActivityNet-v1.3** and **Thumos14** datasets at the time of publication.
- Extensively investigated **Graph Networks** and **Vision-language** models for Action understanding

Samsung Research — *Bangalore, India*

Intern , Imaging R&D Team — *May 2019 - July 2019*

- Received a **Full time offer** for modifying different functions in the mobile camera service suite using **ToF depth data**.
- I was also awarded the **Samsung Advanced** certificate for solving a 3 hour long open-ended coding challenge on the first try.

Indian Statistical Institute Kolkata — *Kolkata, India*

Research Intern under Dr. Swagatam Das, ECSU — *2018 - 2020*

- Researched, developed and investigated **Differential Evolution** based methods for **Adversarial attacks** on SoTA image classifiers.
- Implemented and experimented with **Learning based Clustering** algorithms

Artificial Intelligence Lab, Jadavpur University — *Kolkata, India*

Research Assistant under Dr. Amit Konar — *2018 - 2020*

- Designed and implemented a Vision actuated system for detecting and grasping objects with brain commands.
- This work was later published in the Journal of Biomedical Signal processing and Control
- Published two other papers on the application of Convolutional Neural Networks in Brain computer Interfacing.

Education

Carnegie Mellon University — *Pittsburgh, USA*

Master of Science in Robotics, School of Computer Science — *2023 - Present*

- **Advisor : Dr. Martial Hebert, 100% scholarship**

Jadavpur University — *Kolkata, India*

Bachelor Of Electronics and Telecommunication Engineering — *2016 - 2020*

- Academia **CGPA : 9.35/10** | Standardised Test Scores : **GRE (331/340)** **TOEFL iBT (113/120)**

Skills & Interests

Programming Python, C, C++, MATLAB

Deep Learning Keras, PyTorch, TensorFlow, OpenCV, High-Performance clusters

Interests Computer Vision, Multimodal Learning, Reinforcement Learning, Deep Learning, Machine learning, A.I.

Academic Service

Reviewer **NeurIPS 2024, ICLR 2025**, **TPAMI**, Elsevier (IF : **8.139**), IET Computer Vision (IF : **1.95**) *Aug 2021 - Present*

Publications

ReferEverything: Towards Segmenting Everything We Can Speak of in Videos

ICLR 2025

Anurag Bagchi, Zhipeng Bao, Yu-Xiong Wang, Pavel Tokmakov, Martial Hebert *Under Submission*

- Text-to-Video Diffusion model, Refer Segmentation, Vido-Language grounding
  [**View paper**][**Project Page**]

Video Diffusion Models Learn the Structure of the Dynamic World

Under Submission

Zhipeng Bao, **Anurag Bagchi**, Yu-Xiong Wang, Pavel Tokmakov, Martial Hebert *CVPR 2025*

- Text-to-Video Diffusion model, Video understanding
  [**View paper**]

Hear Me Out: Fusional Approaches for Audio Augmented Temporal Action Localization

VISIGRAPP 2022 Oral

Anurag Bagchi, Jazib Mahmood, Dolton Fernandes, Ravi Kiran Sarvadevabhatla

- Video Understanding, Temporal Action Localisation, Graph Neural Networks, Audio-Visual fusion, SoTA on **ActivityNet- v1.3** and **Thumos14**
  [**View paper**][**View code**]

Autonomous grasping of 3-D objects by a vision-actuated robot arm using Brain–Computer Interface

Biomedical Signal Processing and Control

Arnab Rakshit; Shraman Pramanick *; **Anurag Bagchi***; Saugat Bhattacharyya *IF : 5.076*

- RGB-D Grasp prediction, Object Detection, Inverse Kinematics, BCI, EEG classification [**View paper**]

Projects

Vision engine for object-attribute detection

Object detection - Image Classifiaction

Independent *June 2020*

- Developed a felxible framework for a two stage vision engine that performs coarse-grained object detection from RGB images and classifies the detected objects based on fine-grained attributes. Faster-RCNN is used for the initial detection of objects followed by parallel ResNet-50 based image classifiers that each detect the state of a different object attribute in a highly modular fashion.

Deep Power K-means for high dimensional clustering

Autoencoders - Deep Clustering - Representation Learning

Indian Statistical Institute, ECSU *Aug 2020 – Dec 2020*

- Developed a deep clustering framework based on Power K-means that jointly optimises the power means objective together with the auto-encoder loss from Deep K-means while simultaneously learning low dimensional cluster representations in each iteration. The differentiable surrogate in deep K-means is effectively replaced with the kolmogorov mean, while the annealing step is modified to achieve a series of smoother alternatives to the K-means objective.

Novel sparse whitebox adversarial attack on image classifier networks using differential evolution.

Deep Differential Evolution - Adversarial Attacks - Resnet - VGG

Indian Statistical Institute, ECSU *Jan 2020 – Aug 2020*

- Implemented and investigated a tunable Differential Evolution based scheme for whitebox adversarial attacks on multiple state-of the-art classifiers like Resnet 50, 101 and VGG-16.. The parameters for evolution are learned using a gradient descent scheme to reach the ideal balance between exploitation and exploration and increase the fooling rate.

Achievements

2023	**Full(100%) Scholarship for MS-Research**, Carnegie Mellon University	*Pittsburgh, USA*
2020	**University Rank 1 in Final semester (grade : 10/10)**, Jadavpur University	*Kolkata, India*
2019	**Samsung Advanced Programming Certificate**, Samsung Research	*Bangalore, India*
2016	**Top 0.1%(99.9 percentile)**, State Engineering Entrance Exam	*West Bengal*
2016	**Top 0.4%(99.6 percentile)**, National Engineering Entrance Exam	*India*

Relevant Coursework

Undergraduate Linear Algebra, Pattern Analysis and Machine Intelligence,Data Structures and Algorithms, Digital Image Processing, Probability and Statistics, Operating Systems, Digital Signal Processing

Graduate 3D Vision, Robot Learning, Advanced Computer Vision, Math for Robotics, Mechanics of Manipulation