Name - Mayank Kumar Singh

Homepage - https://mayank-git-hub.github.io
GitHub - https://github.com/mayank-git-hub
Mobile - +91 84549 51729
Mail ID - mayank2220singh@gmail.com

One step at a time is the mantra with which I have been developing my Research profile. I have worked on Computer Vision, Speech, Singing, Music and Audio-Visual domains. Currently, I am working on Singing Voice Conversion and Automatic Speech Recognition technologies.

Patent

- **Text Publication to 3D Audio** - *SONY Global Corp., (Filing in Japan and US)*

Publications

- **Hierarchical Diffusion Models for Singing Voice Neural Vocoder** - *ICASSP 2023, SONY Global Corp.*
  - Proposed a mel to waveform audio vocoder based on prior grad which hierarchically works on low to high sample rate generation.
  - Achieved SOTA results on singing voice datasets NUS-48E and NHSS.
  - Samples are available at https://t-naoya.github.io/hdm/
- **Nonparallel Emotional Voice Conversion For Unseen Speaker-Emotion Pairs Using Dual Domain Adversarial Network & Virtual Domain Pairing** - *ICASSP 2023, SONY Global Corp.*
  - Proposed a voice conversion model which achieves both speaker identity as well as emotion conversion.
  - We beat the SOTA results in terms of MOS of quality and speaker identity conversion.
  - Samples are available at https://demosamplesites.github.io/EVCUP/
- **Cross-modal Face-and Voice-style Transfer** - *SONY Global Corp.*
  - Proposed a way to convert audio based on avatar profiles for metaverse applications.
  - Achieved SOTA results for avatar/face conditioned voice conversion and voice to face generation results.
  - Samples are available at https://t-naoya.github.io/xfavot/
- **Robust One-Shot Singing Voice Conversion** - *SONY Global Corp.*
  - Proposed a GAN based architecture for singing voice conversion and achieved SOTA results in terms of MOS quality and speaker identity conversion.
  - For practical applications, proposed a robustification method which allows the model to work even on artefact heavy source separated content.
  - Samples are available at https://github.com/t-naoya/rosvc
- **Source Mixing And Separation Robust Audio Steganography** - *ICASSP 2022, SONY Global Corp.*
  - Proposed a watermarking model which introduces imperceptible distortion and is robust to mixing and source separation algorithms.
  - Allows copyright claims of individual audio stems even if they are mixed.
- **Hierarchical Disentangled Representation Learning for Singing Voice Conversion** - *IJCNN 2021, SONY Global Corp.*
  - Proposed a multi-resolution hierarchical VQ-VAE architecture for SVC which improved the Quality MOS by 1.15 and Similarity MOS by 0.12
- **Improving Voice Separation by Incorporating End-to-End Speech Recognition** - *ICASSP 2020, SONY Global Corp. (Internship)*
  - Modified ConvTasNet to incorporate ASR features for enhanced **Audio Source Separation** on the AVSpeech dataset
  - Achieved State of the Art Results on it, beating Google and Oxford's implementation, increasing the SI-SNR by 3.7dB.
- **NENET: An Edge Learnable Network for Link Prediction in Scene Text** - *Guide Prof. Shubhasis Chaudhari, IIT Bombay (2020)*
  - Posted the paper "NENET: An Edge Learnable Network for Link Prediction in Scene Text" on arxiv.
  - Proposed a novel method of linking characters by creating a graph of characters and applying GNN.
  - Proposed a novel modification of GNN which outperforms other methods on link prediction task.
- **ISBI 2018: Diabetic Retinopathy, Segmentation of lesions**- *Guide Prof. Amit Sethi*

- Aim - Segmentation and classification of the lesions in patients of Diabetic Retinopathy
- Applied state-of-the-art algorithm **fusion-net** for segmentation and **Zoom-In Net** for classification.
- Competition Paper- https://mayank-git-hub.github.io/pdf/ISBI2018.pdf

Professional Experience (Japan)

- **SONY Research Japan, Research & Development Engineer** : *Jan 2021 - July 2022*
  - Developed a Singing Voice Conversion model based on StarGANv2-VC which is deployed on internal SONY CreativeAI tool.
  - Developed Audio Mixing and Sepration Robust Watermarking to facilitate audio stem copyright claims of SONY.
- **SONY Japan, Research Internship** : *May - July 2019*
  - Worked with Audio Technology Research Department in SONY Japan, Osaki to improve Deep Audio Visual Source Separation
  - Used **WaveNet** like architecture, Temporal Convolution for audio speech separation and used visual features for improving separation **SISNR** and surpassed current SOTA implementations.

Professional Experience (India)

- **SONY Research India, Research & Development Engineer** : *Oct - Dec 2020 & Aug 2022 - Ongoing*
  - Worked in collaboration with Audio Technology Research Department in SONY Japan, Osaki to develop Singing Voice Conversion Models.
- **Co-Founded Autonise AI** : *Sep 2018 - Dec 2021*
  - Founded a team of 8 with the vision to act as Technical Consultant in the field of Machine Learning.
  - Targeted the domains - **Text Detection and Recognition**, **Quant Algorithms**, **Facial Segmentation**
  - Changed the startup direction in May 2020 to target the Ed-Tech domain for college students.
  - Managed a team of 11 to create and teach 2 month Machine Learning and Web Development Interactive projects
  - Achieved a target audience of size 300.
- **HDFC Life, Research Internship** : *May - July 2018*
  - Automated customer interaction by automating questions asked using **Reinforcement Learning**.
  - **Feature Engineering** and **Clustered Customer data** for extracting useful statistics and analysis of the algorithm

Projects

Research Projects

- **Segmentation of Medical Image** - *Guide Prof. Amit Sethi*
  - Applied **NN**, **SSNMF**, **NMF**, **SVM** algorithms to do pixel-level segmentation on Hyperspectral Images
  - Implemented the initial steps for detecting cancer by segmenting epithelium, stromal and goblet cells.
- **Whole Slide Image Stitching using DC motor video** - *Guide Prof. Amit Sethi*

Other Projects

- **Kaggle Competition: *iMaterialist Challenge (Furniture) at FGVC5***
  - An orthodox classification competition with **highly skewed class size** and high intra class and low inter class variation.
  - Trained **ResNet-152**, **NASNet** model using extensive class specific data augmentation.
  - Got a **rank of 30** under the team name 'Artificial incoherence'
- **Text Detection and Recognition on Documents**
  - Implemented CRAFT for Text Detection on https://github.com/autonise/CRAFT-Remade which has 44 forks and 162 stars on GitHub
  - Implemented Pixel-Link for Text Detection on https://github.com/mayank-git-hub/Text-Recognition
  - Achieved an F1-score of 74% which is **6% more than Google's on our custom dataset** consisting of passports, aadhar cards, driving license cards and other docs which we annotated using our annotation tool built using javascript.
- **Web Development**
  - Designed and deployed a learning platform http://autonise.com (Deprecated)
  - The stack used is Ionic-Angular for UI, Django for backend and MongoDB for database.
  - Incorporated RazorPay payment gateway with instant UPI based referral bonus system.

Scholastic Achievements

- Secured All India Rank of **108 in JEE Advance 2016 in General Category** among 198,228 candidates
- Secured All India Rank of **1484 in JEE Mains 2016 in General Category** among 1,207,058 candidates
- Recipient of prestigious **KVPY fellowship with All India Rank of 363 (/60,000)**

Education

Institution	Specialisation	Year	GPA/Percentage
Indian Institute of Technology, Bombay	Electrical Engineering, B.Tech	2020	8.84

Technical skills

- **Programming Languages:**
  Proficient in: C, C++, Python, JAVA, Javascript, Angular
  Specific libraries for Machine Learning - Tensorflow, Pytorch
  Also basic ability with: MATLAB, Shell Script, Arduino, NgSpice, VHDL, AutoCad, Solidworks.
  Server Side - Django, NGINX, Flask, AWS, Kotlin, Django-Channels, MongoDB
  Client Side - Ionic, Android-Studio(JAVA & Kotlin), HTML, JS, D3JS, Three JS, ES6, React, Angular