Vedant **Joshi,**

Masters student

UC San Diego
linkedin.com/in/vedant-joshi-b822bb169
https://vedrocks15.github.io/
+1-(858)-305-4782 @ vejoshi@ucsd.edu

"**How to make machines understand ?**", is the question that drives my research every morning. Based on this question, my fundamental goal is to understand the inner workings of deep learning models & make a sincere contribution towards building dynamic as well as modular mechanisms that achieve human level generalisability by efficiently representing a multitude of input modalities into a single structured latent space & are contextually modifiable based on the situation at hand.

Research Interests

- Representational learning
- Self-supervised learning
- Building multi-modal latent spaces
- Open World Learning
- Reflectance Models
- Model quantization & pruning
- Model explainability & interpretability

Highlights

- Gold medalist for highest GPA in graduating class of 2017.
- Received Best Paper award at 8th ICSCC 2021 by IEEE.
- SIH-2020 Finalist under Airport Authority of India.
- Oracle Certified Java Associate Programmer
- Kaggle open source contributor for implementing latest research papers & frameworks.

Work Experience

Vision & Imaging Engineer-I

February 2023 - July 2023

Tonbo Imaging Pvt. Ltd.

- **IR object detection** : Achieved a **20%** improvement in detection performance on thermal images by re-engineering the base layers of **YOLOv5** which significantly improved the robustness of Tonbo's autonomous driving software stack for heads up displays.
  - Stabilized the training framework to support higher resolution feature maps in the **detection head** to improve the localization of tiny objects.
  - Developed wider **cross-stage partial** layers along with **selective kernel attention** modules to model dynamic receptive fields for improved multi-scale detection.
- **RGB to IR mapping** : Leveraged **Generative A.I.** based diffusion models to generate realistic thermal images from RGB videos to solve the data shortage problem for training perception models.
  - Implemented **subpixel-convolution** with **ICNR initialization** to reduce checkered artifacts & added **neural style transfer** loss to improve fine details in the generated thermal image domain.
- **Depth Estimation** : Reduced the error rate by **7%** in depth map generation from monocular videos using PoseNet & self-supervised learning for **Nvidia Xavier NX** devices to improve the spatial awareness of self-driving solutions.

Data Scientist - I

July 2022 - December 2022

Vedantu Innovations Pvt. Ltd.

- **Voice Nudges** : Developed a concatenative synthesis engine using **spectral smoothing methods** to stitch the name of a student into an audio nudge in the voice of the class teacher.
  - Defined appropriate **student behaviour metrics** to select an appropriate audio nudge covering scenarios from motivational to concerning in real time.
  - Experimental results indicated a **10% improvement** in student engagement & concentration after receiving personalised nudges during the live online class.
- **User Cohort Generation** : Improved the productivity of marketing teams by **25%** through generation of user behaviour clusters that captured intricate patterns in click-stream & user interaction data.
  - Implemented **TabularTransformers** along with the self supervised method **SCARF** to learn latent representations from unlabelled data.
  - Using **UMAP** & **K-Means**, identified key features in student behaviour clusters that served as actionable insights for the marketing teams to adjust their course selling approach.
- **Profanity Filtering** : Productionized a novel solution for syntactic matching of illict words in real time chats module that achieved a **10%** improvement in recall over regular expressions.
  - Implemented contrastive learning to learn noise invariant word embeddings & perform **zero shot learning** on LSTMs especially for detection of profane words in low resource languages.
  - Further added a modular cleaning pipeline that handled removal of domain specific tokens & close matches to improve the precision of our overall detection system by **4.5%**.

ML Researcher

July 2021 - June 2022

Vedantu Innovations Pvt. Ltd.

- **Cluster Cleaner** : Improved the quality of matches returned by elastic search engine by **40%**, through creation of a text cleaning pipeline that used **n-gram SimHashing** & **Levenstein** distances to remove noisy results that were added due to misreads by MathPixOCR.
- **SSL Search Engine** : Architected a real time graph based search engine by learning a compressed latent space of doubt images through **self supervised learning** along with domain specific augmentations that lead to a **87%** reduction in time towards solving a doubt posted on the platform.
  - The **augmentations** modelled noise functions generated from the real world camera & this helped the model capture appropriate feature invariances to return high quality matches.
  - Implemented SOTA frameworks such as **BYOL & SimCLR** with an innovative cross temperature scaling mechanism for InfoNCE loss to mine hard negative samples during pretraining.
- **Doubts Clustering Module** : Reduced the load on academic experts by repurposing the search engine to an unsupervised clustering module using **UMAP & HDBSCAN** which lead to a reduction in redundant images from **9.2 lakh** instances to **2.5 lakh** clusters in the doubts image repository.

Deep Learning Intern

September 2020 - April 2021

Vedantu Innovations Pvt. Ltd.

- **Image Denoisers** : Experimented with image de-noising/skewing models using **UNET segmentation** & **VAEs** to create binarised images that reduced the character error rate by **5%** during text extraction by Tesseract OCR. The cleaned binary images also reduced the memory costs on AWS instances by **32%**.
- **Image 2 Latex Markup** : Ported legacy Tensorflow code to 2.0 along with **attention implementation** from scratch for the paper Image to Markup Generation. The generated latex markup from images helped us create focused latex only search spaces for each subject & improved the search recall by **6%**.
- **Subjects Classifier** : Enhanced the F1-score on bi-directional embedding vectors from fine tuned **BERT** by **14%** for the task of subject classification on low resource classes by text normalisation & **LDA topic modelling**.

Research Intern

September 2020 - April 2021

TCS Rapid Labs

- **Vision based lip reading** : Generated a **25%** character error rate on the task of single word lip reading from videos by re-modelling the **LipNet** model from word to character level along with an efficient **CTC loss** implementation that helped patients suffering from hearing deformities.
  - Developed functions to achieve speaking speed normalisation by frame count streching & pre-processed each video frame to lip region only via the Dlib package.
  - Proposed FYEO, an **attention** based LipNet model, which reduced the character error rate by **2.5%** & improved model transparency through heat maps that showcased predicted character & time frame alignment.

Education

Masters in Computer Science

September 2023 to June 2025

UC San Diego

- GPA : $4.0/4.0$
- Relevant Courses : Computer Vision, Probabilistic Reasoning, Recommender Systems

Bachelor of Technology Honours Computer Science

August 2017 to April 2021

IIIT Kottayam

- Gold medal for receiving the highest grade in the batch of 2017. Acquired CGPA : $9.82/10.0$
- 3 out of 8 semesters with perfect grade of 10.0

Skills

Programming languages	Python, Java, C, C++
Deep Learning Frameworks	TensorFlow, Keras, PyTorch, NLTK, Scikit-learn, Nvidia-TensorRT, JAX, MATLAB, ONNX
Development Tools	Visual Studio Code, Google Colab, Docker, Git, Pandas, Numpy
Operating Systems	Windows, Linux, Mac OS X
Embedded Boards	RaspberryPi, Arduino (Nano, UNO, Mega), Nvidia Jetson Nano, NodeMCU
Domain Knowledge	Computer Vision, NLP, multi-modal learning, Micro Ariel Vehicles, IoT

Academic Projets

Open World Object Detection — Prof. Zhuowen Tu

Machine Learning, Perception, and Cognition Lab — October 2023 - Current

Improving the generalization ability of **one stage object detectors** from closed to open vocabulary through novel distillation strategies that leverage the knowledge of large scale pre-trained **vision language models**.

- Adapting the detection head of **YOLOv7** from **softmax** based classification to **CLIP** style semantic embedding generation to support novel instance classification during inference.
- Reducing the domain gap between the image level objective of **vision language models** & the instance level objective of detection models to generate an improved supervised signal during training.

Python | TensorFlow | Distributed Training | OpenCV | PyTorch | Tensorboard

Glaucoma Progression Detection — Dr. Mark Christopher

Shiley Eye Institute UCSD — November 2023 - Current

Developing multi-modal image classifiers from OCT & Fundus images to detect explainable features for progression of glaucomma in patients with high occular hypertension.

- Developed a novel small scale dataset using patient's historical features & lab readings to generate an aligned OCT-Fundus mapping along with a **signal overlap metric** to help select most informative instances during training.
- Implemented self-supervised pre-training approaches to leverage unlabelled fundus images to find optimal set of weights for **vision transformers** & improve the final downstream F1 score for the task of Glaucoma detection.

Python | TensorFlow | Distributed Training | HuggingFace | PyTorch | Tensorboard

Coco Layers : Edge Based Inference

Orzico Funded Project — December 2019 - April 2021

Developed an edge based real time object detection system which served as a pluggable module for any robotic implementation to pluck coconuts directly from trees.

- Curated a novel, small scale, manually annoted coconut images dataset using drones at multiple locations in Kerala. The dataset successfully captured a spectrum of **natural lighting conditions, occlusions & coconuts of various shapes & sizes** that occur in the wild.
- Completed a comparative study on mixed precision & quantization aware training for **SSD MoileNetV2, YOLOv3 & tensorRT optimised tiny YOLOv4** for real time inference on RaspberryPi 3B+ & Nvidia Jetson Nano.
- Developed a model deployment pipeline for quantized models with efficient frame buffer handling mechanisms & thread level multiprocessing to achieve a **frame rate of 22 FPS** in real scenarios.

Python | TensorFlow Lite | Nvidia TensorRT | OpenCV | Linux

Lemon Picking Drone

Project Report — IIIT-Kottayam — August 2019 - December 2019

Created an **autonomous drone with a 3D-printed peripheral arm** that assisted in picking lemons suspended in mid-air along with a live video feed streaming service.

- Developed a quad-copter from scratch & an inter-process communication system between pixHawk flight controller system & Raspberry Pi to complete **GPS guided autonomous flights**.
- Implemented a light weight **color segmentation** script to perform lemon tracking by edge devices. A gstreaming pipeline was written to broadcast the live detection to a user interface.
- The 3D arm control was handled by an embedded C subroutine that controlled the **micro controller Arduino UNO** along with a combination of stepper motors.

Python | Embedded C | PixHawk | MissionPlanner | Parallel Computing

Publications

1. **Vedant Sandeep Joshi**, Sivanagaraja Tatinati, Yubo Wang, *Looking For A Match : Self-supervised Clustering For Automatic Doubt Matching In e-learning Platforms* preprint : arXiv:2208.09600, https://doi.org/10.48550/arXiv.2208.09600

2. **Vedant Sandeep Joshi**, Sivanagaraja Tatinati, Yubo Wang, *YZR-net : Self-supervised Hidden representations Invariant to Transformations for profanity detection* preprint : arXiv:2211.15532, https://doi.org/10.48550/arXiv.2211.15532

3. **Vedant Sandeep Joshi**, Ebin Deni Raj, *FYEO : A Character Level Model For Lip Reading*, IEEE/8th International conference on Smart Computing and Communications, Accepted Sep 6 2021, https://doi.org/10.1109/ICSCC51209.2021.9528104

4. **Vedant Sandeep Joshi**, Ebin Deni Raj, Jeena Thomas, *Quantized Coconut Detection Models with Edge Devices*, Journal of Interconnection Networks [Scopus Indexed Journal], Accepted November 2 2021, https://doi.org/10.1142/S0219265921440102

References

Dr. Sivanagaraja Tatinati
Manager - Data Science, Vedantu Innovations Pvt. Ltd
@ tatinati.sivanagaraja@vedantu.com
On Request

Dr. Ebin Deni Raj
Associate Dean (Academic Affairs), IIIT Kottayam
@ ebindeniraj@iiitkottayam.ac.in
+91 (0) 482-2202154

Dr. J. V. Bibal Benifa
Assistant Professor, IIIT Kottayam
@ benifa@iiitkottayam.ac.in
+91 (0) 482-2202163