BHAVIN JAWADE

3296 Main Street, Apt C1 · Buffalo, NY, 14214 · +1 716-495-9321 · bhavinjawade@gmail.com
LinkedIn - www.linkedin.com/in/bhavinjawade · **Website** - https://bhavinjawade.github.io ·
Github - github.com/bhavinjawade

EDUCATION

Universities	Degree	Program	Specialization	GPA	Year
University at Buffalo, SUNY	Doctor of Philosophy Advisor–Prof. Venu Govindaraju	Computer Science	Deep Learning	**3.8 / 4**	2nd year
University at Buffalo, SUNY	Master of Science Advisor–Prof. Sargur Srihari	Computer Science	Artificial Intelligence	**3.8 / 4**	May 2021
G.S. Institute of Tech & Sci	Bachelor of Engineering	Information Tech.		**3.8 / 4**	June 2019

WORK EXPERIENCE

Adobe Research | **Research Scientist Intern** | *San Jose, CA* May 2022 - Nov 2022

- Proposed a **Large-Scale Vision-Language contrastive pre-training** strategy for transformer-based encoders to learn embeddings that generalize over out-of-distribution task.
- Proposed a novel loss formulation that utilizes tags, metadata and clauses to create hierarchy of shared semantics that bring semantically similar vision and language embeddings closer to each.
- Built a distributed large-scale pretraining framework for vision-language training with Torch-distributed, web-datasets and gradient-checkpointing. Trained on 100 million image-text pairs on 160 A100 GPUs.
- Our method out-performed Open-AI's CLIP by 6.4% on zero-shot ImageNet retrieval when trained on 50 million image-text pairs, 9.2% when trained on 15 million subset and 8.4% when trained on 5 million sub-set.

Center for Unified Biometrics and Sensors (CUBS) | **Research Assistant** | *Buffalo, NY* Jun 2020 – Present

- NSF CITER Grant | Developed a multi-task learning based CNN network trained with Deep metric losses (Contrastive, AdaCos) along with minutiae loss to match contactless fingerprints captured using smartphone camera against legacy fingerprints.
  - Won **Russel Agrusa Research Innovation Award,** UB CSE, 2021 (https://cutt.ly/bCYPTd8)
  - Demonstrated the project **at Federal Identity Conference,** 2022
- NSF DIBBS Grant | Responsible for creating the full-stack machine learning framework called MLToolkit where the user can drag and drop the different elements to create and execute machine learning pipelines. (https://git.io/JtJ4d)
- Cross Modal Metric Learning for Text-to-Image Retrieval using Cross-Attention and Multi-Similarity Loss. Upcoming paper on using semantic proxies for generating contextual embedding hypersphere. Achieves results on par with SOTA on MSCOCO and Flickr30k datasets.

Persistent Systems | **Machine Learning Engineer** | *Pune, India* Jul 2019 - Dec 2019

- Built OCR ML application for banks to auto-analyze POS invoices saving $200,000 per annum & 100 hours/week.
- Developed an Invoice Management tool for a Supply Chain firm to automate the supplier-distributor financing process and decrease finance approval time by 43%.

IIT Madras, Department of Computer Science | **Research Intern** | *Remote* Feb 2019 - May 2019

- Advised by Dr. Rupesh Nasre, Professor IIT Madras.
- Designed an efficient algorithm to compute execution time for updating million node DAG in hierarchical task scheduling.
- Algorithm computes the result for a million node DAG & over ten thousand updates within 0.8 seconds with a near-linear O(n) computational complexity.

SELECTED RESEARCH PAPERS

NAPReg: Nouns as Proxies Regularization for Semantically Aware Cross-Modal Embeddings Aug 2022
IEEE/CVF Winter Conference on Applications of Computer Vision, 2023 ***(WACV 2023)***

Summary: We proposed NAPReg, a novel regularization formulation that projects high-level semantic entities i.e. Nouns into the embedding space as shared learnable proxies. We show that using such a formulation allows the attention mechanism to learn better word-region alignment while also utilizing region information from other samples to build a more generalized latent representation for semantic concepts. Experiments on MS-COCO, Flickr30k and Flickr8k demonstrate that our method achieves state-of-the-art results in cross-modal metric learning for text-image and image-text retrieval tasks.

Hear The Flow: Optical Flow-Based Self-Supervised Visual Sound Source Localization Aug 2022
IEEE/CVF Winter Conference on Applications of Computer Vision, 2023 ***(WACV 2023)***

Summary: In a video, often-times, the objects exhibiting movement are the ones generating the sound. In this work, we capture this characteristic by modeling the optical flow in a video as a prior to better aid in localizing the sound source. We further demonstrate that the addition of flow-based attention substantially im- proves visual sound source localization. We benchmark our method on standard sound source localization datasets and achieve state-of-the-art performance on the SoundNet Flickr and VGG Sound Source datasets.

RidgeBase: A Cross-Sensor Multi-Finger Contactless Fingerprint Dataset Jun 2022
IEEE/IAPR International Joint Conference on Biometrics, 2022 ***(IJCB 2022)*** (https://ijcb2022.org/#/accepted_papers)

Acquired first of its kind multi-finger, cross-sensor, multi-environment contactless fingerprint datasets with over 15 thousand contactless fingerprint images from 88 subsets. Proposed dataset is the largest publicly available smartphone based contactless fingerprint dataset. Benchmarking commercially available software such as Veri-finger and deep-learning based methods demonstrate that RidgeBase is a challenging dataset with more practical applicability.

Attribute De-biased Vision Transformer (AD-ViT) for Long-Term Person Re-identification Sept 2022
IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2022 ***(AVSS 2022)***

Summary: We propose an Attribute De-biased Vision Transformer (AD-ViT) to provide direct supervision to learn identity-specific features. Specifically, we produce attribute labels for person instances and utilize them to guide our model to focus on identity features through gradient reversal. Our experiments on LTCC and NKUP datasets shows that the proposed work consistently outperforms the state-of-the-art methods.

Multi Loss Fusion for Matching Smartphone Captured Contactless Finger Images Sept 2021
IEEE International Workshop on Information Forensics and Security, 2021 ***(WIFS 2021)*** | https://cutt.ly/RUCxbLj

Summary: Cross Sensor Contactless fingerprint matching using weak supervision based dual encoder CNN with AdaCos Angular Margin Loss. Proposed method achieved 3.1% improvement on 2 benchmarks datasets with respect to state of the art methods and commercial off the shelf software.

Low computation in-device geofencing using hierarchy-based searching for offline usage Nov 2018
IEEE International Conference on Inventive Computation Technologies, 18' ***(ICICT 18')*** | https://cutt.ly/fjWMULt

Summary: Proposed a fast, efficient, low computation geo-fencing algorithm that can run on the edge device without depending on internet connectivity. The algorithm was 20% faster and 33% more accurate than some commercial geofencing APIs with a validation margin of less than 3 meters.

ACHIEVEMENTS AND AWARDS

- Winner - **Adobe Code Jam 2022** | *(Coding Competition)* | (https://cutt.ly/SCYSoul)
- Won **Russell Agrusa Computer Science Innovation Award 2021** | (*Project Competition)* (https://cutt.ly/bCYPTd8)
- Won Blackstone launchpad **Best Idea Award 2021** | *(Idea Competition)* (https://cutt.ly/ZCYDV1u)
- Second Position - **Maple Ridge Hackathon** (Govt. of British Columbia) | *(Hackathon)* (https://cutt.ly/OCYP4bz)
- Represented India at **Microsoft Asia Summit at Taiwan 2017**
- **Project Award** for *elogbook* by Hon' Minister of State (Education Minister, Govt of M.P., India)
- Grand Finalist **Smart India Hackathon**, 2019 | *(Hackathon)*

- Winner of **WittyHacks Hackathon 2018** | *(Hackathon)*
- Represented College at **ACM ICPC Regionals 2017**, IIITM | *(Coding Competition)*

SELECTED PROJECTS

Semantic Embedding Hypersphere for Text-to-Image Retrieval (Computer Vision | Deep Learning) Nov 2021

- Proposed a novel approach to embed semantic proxies in the loss's feature space to learn discriminative cross-modal embeddings. Approach extracts vision and textual features from Vision transformer and BERT optimizes a deep metric learning loss with novel semantic proxy supervision**.**
- Utilized bottom-up Attention network to extract visual representations and used BERT & CLIP for text embeddings.
- Attained 4.2% improvement in R@1 and 1.8% improvement in R@10 on MSCOCO and Flickr30k datasets.
- **Tech:** Pytorch, Python, Sklearn, NLTK, Deep Learning, LSTM, Deep Metric Learning

Attention based Neural Image Captioning (Computer Vision | Deep Learning) Oct 2020

- Implemented Show Attend and Tell - Neural Image Captioning model with adaptive attention.
- Improved it by implementing Adaptive Attention Mechanism. Used ResNet 101, DenseNet 201 and VGG 16 CNN.
- Attained 0.39 BLEU-4 Score with 88% accuracy on captioning and 94.5 Top-5 accuracy on classification.
- **Tech:** Pytorch, Python, Sklearn, NLTK, Deep Learning, LSTM, BLEU-4.

Now You See Me - The Blind Project | Project video http://bit.do/nysmvideo Jan 2019

- Built an app to provide semantic visual analysis of the environment in voice feed to a visually impaired person.
- Developed a multi-threaded architecture with computer vision capabilities built over TensorFlow, SSD algorithm, OpenCV.
- **Project won Blackstone Launchpad's Ideas Competition.**
- Tested with 60 visually impaired students at National Institute for blind.
- **Tech:** Android, Tensorflow Lite, Cognitive Services, Celery, Firebase, Maps API, Auth0.

Lip Motion Biometrics (Computer Vision | Deep Learning).

- Modeled lip motion as a spatial-temporal learning problem to investigate if movement of lips can help identify people.
- Used a spatial transformer network along with a ResNet101 encoder and a LSTM and transformer-based decoders.
- The method was evaluated on VGG's LRW and VoxCeleb dataset.
- **Tech:** Pytorch, Python, Sklearn, NLTK, Deep Learning, LSTM, BLEU-4, Machine Learning, Neural Networks, Azure.

Reinforcement Learning - Actor Critic Methods Jan 2020

- Trained a CNN-based Deep Q Network, Dueling Network, and Policy gradient algorithms - Advantage Actor-Critic to play Atari Games at a human level performance. Improved DQN reward by using dueling from 9200 to 30250 on roadrunner.
- Multiagent Reinforcement Learning algorithm to solve a Ship Docker Problem, A2C convergence in just 6 hours of training.
- **Tech:** Pytorch, Tensorflow, OpenAI Gym, AWS.

POSITIONS OF RESPONSIBILITY

- Serving as President of Computer Science Graduate Student Association (CSE-GSA (UB)) | 2021-22
- Lead organizer of HackIndore 2018, the largest hackathon of Central India.
- Co-lead Facebook Developer Circle Indore
- City Head - Microsoft Student Partner
- Founder and Head for SGSITS' first techno-learning club (#include)
- Head, Design - Entrepreneurship Cell SGSITS
- Vice-Captain - Aeromodelling club SGSITS