

EDUCATION

International Institute of Information Technology, Hyderabad (IIIT-H)

- Master of Science in Computer Science Engineering

(August '21 - May '23)

Institute of Technology Nirma University, Ahmedabad

- Bachelor of Technology, Instrumentation and Control Engineering

(July '14 - June '18)

PUBLICATIONS

• Grounded Video Situation RecognitionZeeshan Khan, CV Jawahar, Makarand Tapaswi*Neural Information Processing Systems (NeurIPS 2022)* [[Project Page](#)]**• More Parameters? No Thanks!**Zeeshan Khan, Sukruth Kartheek, Vinay Namboodiri, CV Jawahar*Association for Computational Linguistics (ACL 2021)(findings)* : [[Paper](#)][[Code](#)]**• DeepHS-HDRVideo: Deep High Speed High Dynamic Range Video Reconstruction**Zeeshan Khan, Mukul Khanna, Parth Shettiar, Shanmuganathan Raman.*International Conference on Pattern Recognition (ICPR 2022)* [[Oral presentation](#)][[Paper](#)] [video](#)**• Appearance Consistent Human Pose Transfer Transfer via Dynamic Feature selection**Ashish Tiwari, Zeeshan Khan, Shanmuganathan Raman.*International Conference on Pattern Recognition (ICPR 2022)* [[Paper](#)]**• Exploring Pair-wise NMT for Indian Languages.**Sukruth Kartheek, Sreedhar Rajpurohit, Sai Himal Allu, Aman Singhal, Zeeshan Khan, Vinay Namboodiri, CV Jawahar*International Conference on Natural Language Processing, (ICON 2020)* : [[Paper](#)]**• FHDR: HDR Image Reconstruction from a Single LDR Image using Feedback Network**Zeeshan Khan, Mukul Khanna, Shanmuganathan Raman.*IEEE Global Conference in Signal and Information Processing, 2019* : [[Oral Presentation](#)] [[Paper](#)] [[Code](#)]

RESEARCH EXPERIENCE

CVIT, IIIT Hyderabad

Mentors: Prof. C.V. Jawahar, Prof. Makarand Tapaswi, Prof Vinay Namboodiri

Research Fellow

(August '20 - Present)

• Grounded Video Situation Recognition (GVSR) ~ NeurIPS 2022

- Proposed GVSR- A holistic video understanding framework, that combines semantic role captioning with spatio-temporal grounding, for understanding relational reasoning among multiple entities and events across time in short movie clips.
- Designed a 3-stage Transformer model, which hierarchically models all the objects, actions, and semantic roles across an entire video to answer structured questions like (who, did what, to whom, with that, where)
- Improved the spatio-temporal representations and enabled grounding without box-supervision, by encoding object-action representations together and exploiting Transformers attention. Improved fine-grained semantic role captioning performance by an absolute gain of 22% over prior SOTA (VidSitu CVPR 2021).

• Multilingual Neural Machine Translation(MNMT): ~ ACL 2021

- Studied the long standing problem of overcoming negative interference from long tailed languages in MNMT, and proposed to learn language specific and language invariant submodules to overcome it. Achieved State-of-the-art performance on 8+ languages on the multilingual TED talks dataset.
- Proposed a continual learning training scheme via iterative pruning and retraining of Transformer, to find language specific parameters from the existing multilingual parameters, without using any new language specific adaptive layers.

• Exploring Pair-wise NMT for Indian Languages : ~ ICON 2020

- Developed State-of-the-art NMTs for 11 low resource Indian Languages. Achieved rank #1 on the workshop on Asian Translation (WAT) leaderboard for multiple language pairs.
- Proposed a back-translation filtering mechanism, to filter the noisy back-translated monolingual corpora and retain only the high quality training data to finetune an MNMT to a single language pair.

Computer Vision and Graphics Lab, IIT Gandhinagar

Mentor: Prof. Shanmuganathan Raman

Research Assistant

(February '19 - June '20)

• DeepHS-HDRVideo Deep High Speed High Dynamic Range Video Reconstruction ~ **ICPR 2022**

- Proposed to generate high FPS HDR video, from a sequence of low FPS alternating exposure LDR frames for the first time.
- Used pre-trained video frame interpolation models to generate multiple high and low exposure frames in the LDR domain, and merged them at each timestep to generate a high speed HDR video at arbitrarily high frame rates.

• Appearance Consistent Human Pose Transfer via Dynamic Feature Selection ~ **ICPR 2022**

- Proposed a Novel 3-stream GAN consisting of image, pose, and appearance pathways. That progressively transforms a human image from a source to a target pose. Used deformable convolutions, non-local attention, and adaptive instance normalisation for appearance transfer. Achieved State-of-the-art results on Fashion dataset.

• FHDR: HDR Image Reconstruction from A Single Exposure LDR Image using Feedback Network ~ **GlobalSIP 2019**

- Proposed a novel Feedback CNN, for HDR image generation from a single exposure LDR image. Achieved State-of-the-art performance, with significant improvement over prior methods.
- Designed a novel Dense Feedback Block using hidden states of RNN, to transfer high-level information to low-level features. LDR to HDR representations are learned in multiple iterations via a feedback loop.

Raxter @ <https://raxter.io>, Gandhinagar

Research Assistant

(Jan '18 - Jan '19)

Mentor: Prof. Sourish Dasgupta ~ CEO at Raxter (Asst. Professor @ DAIICT)

• RAxBot: Reinforcement Learning(Q-learning) based Scholarly Article Recommendation Engine: [[Pre-print](#)]

- Designed and implemented RAxBot: The core Personalised Scholarly Article Recommendation Engine behind **raxter.io**, providing scalable content delivery and adaptive recommendations.
- Proposed a novel Q-learning based framework that exploits the user selection pattern history and takes action to predict the user behaviour. Based on the prediction a query is generated and sent to the Elasticsearch server for content retrieval.

• Scholarly Article Recommendation Engine using a Query Augmentation Framework:

- Implemented a personalised recommendation engine using content based filtering and a query augmentation technique.
- Designed a complex 3-level Elasticsearch query. i) Topic-level using LDA topic models ii) Word Embedding level using Word2vec model, and iii) Noun Phrase level. Performed extensive experimentation to determine the query weights.

CONSULTANCY PROJECTS

Azure Knowledge Corporation, Ahmedabad

Technical Consultant

(Jul '20 - Aug '20)

Project Coordinator : Abhilash Mankad ~ COO at Azure

• Automatic Advertisement and Signboard Detection in Mobility

- Designed and developed a deep learning based tool to automate the filtering of advertisements and signboards on streets in the wild. Trained a YOLO-V3 object detection algorithm for the given task.
- Further developed a re-trainable tool using transfer learning, which allows it to adapt to new datasets and classes.

Honeywell International India Pvt. Ltd., Bangalore

Research Consultant, with IITGN

(Oct '19 - Mar '20)

Project Coordinator : Manjuprakash Rama Rao ~ Director, Architecture and Innovation

• Synthetic Data Generation for using Human Pose Transfer [[Architecture-diagram](#)] [[Code](#)]

- Proposed a 2-stream GAN framework for Human Pose Transfer, involving foreground and background image generation. FG path deals with pose transfer and BG path inpaints the background for generating consistent target FG and BG images.
- Extended the Progressive Attention Transfer Network(PATN)(CVPR-2019) by proposing FG and BG losses for self-supervised generation of target FG and BG masks, allowing to separately model the FG and the BG paths.

INTERNSHIPS

Google Summer of Code with Sugarlabs (GSoC'17)

Intern

(May '17- July '17)

Mentor: Walter Bender ~ Founder Sugarlabs, ex-Executive Director of The MIT Media Lab

• Say No To GTK2: Graphical User Interface Refactoring : [[Blog with Code](#)]

- Completed the Google Summer of Code program '17 and was rewarded a sum of **2400\$**.
- Refactored the GUI and migrated 9 major applications of Sugarlabs from GTK2 to GTK3 toolkit in Python, enabling further growth and development of the applications. Also, Ported from GST 0.10 to GST 1.0.

Relevant Coursework with Projects at IIIT

- **Computer Vision**

- Multiview Geometry (determining homography in different settings)
- Camera parameter matrix estimation using DLT, RANSAC, and Zhang's method (**Implemented in assignment**)
- Lucas-Kanade Optical Flow with multi-scale coarse to fine refinement (**Implemented in assignment**)
- Graph-cut for segmentation (**Implemented in assignment**)
- Viola Jones for face detection, Eigenfaces, Person Detection
- SIFT, SURF, Harris Corner detectors.
- Object Detection using CNN (Faster RCNN)
- Text to Image generation using GAN (**Implemented in project**)

- **Digital Image Processing**

- 1st order 2nd order Edge Detection, Bit quantization (**Implemented in assignment**)
- Smoothening, Sharpening, Cartoonishing, Histogram Equalization and Matching (**Implemented in assignment**)
- Low Pass Filtering, Bilateral Filters, High Boost Filtering (**Implemented in assignment**)
- Fourier Transform, Inverse Fourier Transform, Fast Fourier Transform (**Implemented in assignment**)
- Morphological Operations and connected components for segmentation (**Implemented in assignment**)
- Hough Transform for line and circle detection (**Implemented in assignment**)

- **Introduction to Natural Language Processing**

- Tokenization, Stemming, Lemmatisation, pos-tagging
- Simple N-gram statistical Language models (**Implemented in assignment**)
- Kneser Ney, Witten bell smoothing in n-gram statistical Language modelling (**Implemented in assignment**)
- Word2Vec, Bag of words, skip gram Language modelling (**Implemented in assignment**)
- RNN, LSTM (**Implemented in assignment**)
- Machine Translation using sequence to sequence modelling (**Implemented in assignment**)

- **Advanced Natural Language Processing**

- Subword tokenization using byte pair encoding.
- Attention Mechanism and Transformer
- Sequence to Sequence models using Attention (RNN, LSTM) (**Implemented in assignment**)
- ELMO (**Implemented in assignment**)
- BERT, BART, ROBERTA
- Summarization, Machine Translation using sequence to sequence model attention
- Pointer Generator Network (**Implemented in assignment**)
- Transfer and Align: A 2-step approach for finetuning BERT for Answer Selection Task (**Implemented in Project**)

- **Topics in Applied Optimization**

- Basics of multivariate Calculus (**written assignment**)
- Convex Optimization
- Gradient Descent (**Implemented in assignment**)
- Lagrangian Dual, Conjugate, Duality.
- Solving Optimization Problems in the dual formation (**written assignment**)
- Gradient plots of mathematical functions (**Implemented in assignment**)
- RMSPROP, ADAGRAD, ADAM
- SVM with the perspective of optimization (**Implemented in assignment**)
- Recommendation system using SVD (**Implemented in assignment**)
- GAN's Min Max Optimization

- **Introduction to Cognitive Science**

- Perception and Cognition
- Brain, Neurons, and Sensory System
- Vision, Top-down Bottom-up attention in human visual system
- Multimodal Interaction in brain
- Memory creation and retrieval
- Learning- behaviourist theory vs cognitivist theory
- Towards Artificial General Intelligence using Inductive Biases from Humans (Term paper- Best Paper in a class of 100+ students)

Relevant Course Taken Online

- **Stanford CS-229n Machine Learning**
- **Stanford CS-231n Deep Learning for Computer Vision**
- **Stanford CS-224n Natural Language Processing with Deep Learning**
- **Linear ALgebra MIT (Gilbert Strang)**
- **Data Structures and Algorithms (Coursera by Robert Sedgewick)**