Soumya Shamarao Jahagirdar

 in linkedin.com/in/soumya-jahagirdar   soumyasj.github.io   github.com/soumyasj
 +91-9591163792   @ soumya.jahagirdar@research.iiit.ac.in

RESEARCH INTERESTS

Computer Vision, Machine Learning, Deep Learning, Video Understanding, Multi-modal learning : Vision + Language, Visual Question Answering, Scene-text Understanding

EDUCATION

2021-Present	**International Institute of Information Technology (IIIT), Hyderabad** M.S. by Research, Computer Science and Engineering, CGPA : 8.67/10 Advisors : Prof. C V Jawahar and Prof. Dimosthenis Karatzas
2017-2021	**KLE Technological University, Hubballi** BE, Computer Science and Engineering, CGPA : 9.01/10

PUBLICATIONS

> <u>Soumya Jahagirdar</u>, Minesh Mathew, Dimosthenis Karatzas, C.V. Jawahar, **Understanding Video Scenes through Text : Insights from Text-based Video Question Answering**, ICCV-W 2023 (Spotlight).

> <u>Soumya Jahagirdar</u>, Minesh Mathew, Dimosthenis Karatzas, C.V. Jawahar, **Watching the News : Towards VideoQA Models that can Read**, WACV 2023.

> <u>Soumya Jahagirdar</u> , Shankar Gangisetty, Anand Mishra, **Look, Read and Ask : Learning to Ask Questions by Reading Text in Images**, ICDAR 2021 (Oral).

> Shamanthak Hegde, <u>Soumya Jahagirdar</u>, Shankar Gangisetty, **Making the V in Text-VQA Matter**, CVPR-W 2023.

> Charani Alampalle, Shamanthak Hegde, <u>Soumya Jahagirdar</u>, Shankar Gangisetty, **Weakly Supervised Visual Question Answer Generation**, CVPR-W 2023.

> Girish Hedge, Tushar Pharale, <u>Soumya Jahagirdar</u>, Vaishakh Nargund, Ramesh Tabib, Uma Mudenagudi, Basavaraja Vandrotti, Ankit Dhiman, **DeepDNet : Deep Dense Network for Depth Completion Task**, CVPR-W 2021.

PATENT

> Uma Mudenagudi, Girish Hedge, Ramesh Tabib, <u>Soumya Jahagirdar</u>, Tushar Pharale, Basavaraja Vandrotti, Ankit Dhiman, Vaishakh Nargund **Method and Device of Depth Densification using RGB Image and Sparse Depth**, Patent Number **WO2022103171A1**.

EXPERIENCE

September 2022 November 2022	**Research Intern, COMPUTER VISION CENTER, CVC, UAB, Spain** > Completed an internship at CVC under the guidance of Prof. Dimosthenis Karatzas and Prof. Ernest Valveny. Worked on Multi-modal learning, specifically video question answering on the videos containing text.
February 2020 July 2020	**Research Assistant, THINK & INK EDUCATION AND RESEARCH FOUNDATION, BANGALORE, India** > Worked as Research Assistant under the guidance of Prof. B. A. Patil, Founder and CEO of Think & Ink Education and Research Foundation, on magnitude spectrum and descriptive statistics-based soil classification. Also developed GUI for the same application for the farmers to use.
June 2019 July 2019	**Summer Intern, INDIAN INSTITUTE OF TECHNOLOGY DELHI, India** > Worked with Dr. Prem Kumar Kalra and his Ph.D. students on the project "Activity recognition for Suturing by Neuro Surgeon" which was a collaborative project with AIIMS, Delhi.

</> TECHNICAL SKILLS AND COURSES

Technical Skills

> Deep Learning : Pytorch, Tensorflow, NLTK, MMF, Transformers, OpenCV, Numpy, scikit-learn, Flask
> Programming/Scripting : Python, C, C++, Java, R, Bash, HTML, CSS

Courses

> Computer Vision, Introduction to NLP, Digital Image Processing, Statistical Methods in AI, Information Retrieval and Extraction

🖥 PROJECTS

VIDEO QUESTION ANSWERING ON VIDEOS CONTAINING TEXT   DR. C V JAWAHAR, DR. DIMOSTHENIS KARATZAS
2021 - PRESENT

- Presently working on multi-modal learning : video question answering on the videos containing text.
- Proposed a new task of obtaining answers by reading text in the news videos.
- Proposed a new dataset : NewsVideoQA, which contains question-answer pairs from news videos across different topics based on the textual content.
- Benchmarked and repurposed multiple baselines for this task.
- Achieved comparable results to multi-modal methods by repurposing a Text-only model : BERT-QA on two datasets – NewsVideoQA and M4-ViteVQA.

`Pytorch` `OpenCV` `MMF` `Lightning` `Huggingface` `SimpleTransformers`

TEXT-BASED VIDEO RETRIEVAL WITH MULTILINGUAL CAPABILITIES   DR. C V JAWAHAR
JANUARY 2023 - JULY 2023

- Set up a server and curated a large-scale dataset of videos from domains such as educational and news videos.
- Built a simple yet effective TF-IDF-based search engine using OCR tokens present in the videos. This also incorporates queries in Multiple Indian languages.
- The dataset collection also contains videos in Indian languages such as Telugu, Kannada, Malayalam, Hindi, etc.

`OpenCV` `scikit-learn`

TEXT BASED VISUAL QUESTION GENERATION   DR. SHANKAR GANGISETTY, DR. ANAND MISHRA
2020 - 2021

- Proposed a novel task of generating questions based on text present in the images.
- Proposed OCR-consistent visual question generation model that Looks into the visual content Reads the scene text, and Asks a relevant and meaningful natural language question (OLRA).
- Introduced OCR-aware loss function to preserve the OCR consistency while generating questions.

`Pytorch` `OpenCV` `NLTK`

DEPTH DENSIFICATION USING RGB AND SPARSE SENSING   DR. UMA MUDENGUDI, SRIB
2020 - 2021

- Proposed a novel DRS (Dense-Residual-Skip) Autoencoder for the task of depth densification from RGB image and sparse depth information.
- Introduced new Gradient-Aware-Mean-Square-Error (GAMSE) loss function for better depth predictions at edges.
- Developed ARCore android application to obtain sparse depth data from mobile phones.

`Pytorch` `OpenCV` `ARCore` `SDK`

👥 ACHIEVEMENTS/CO-CIRRICULAR ACTIVITIES

> Moderator and conducted tutorials in Summer School of AI, CVIT, IIIT-H (2022-2023).
> Volunteer of ICFHR-2022 Conference.
> Finalist of PRISMGlanz, Samsung Research India Bangalore, India (2020).
> Brand Ambassador of Smart Knower (2020).
> Secured 3rd position in BUILDATHON, product building event at KLE Technological University (2019).
> Organized inter-college coding event : TECHNOPHILIA (2019).
> Department (CSE) Co-curricular committee member at KLE Technological University (2018).