

LETTER OF MOTIVATION

Sarosij Bose

sarosijbose2000@gmail.com

Kolkata, India

December 13th, 2021

As Prof. Richard Szeliski ponders in his book: “.. However, despite all of these advances, the dream of having a computer interpret an image at the same level as a two-year old (for example, counting all of the animals in a picture) remains elusive. Why is vision so difficult?”. I was also always intrigued by this very fact: how does the human eye perceive real world 3D scenes and process it inside the brain? This curiosity naturally drove me into visual understanding and try to build systems that can meaningfully interpret real world video data. My primary research objective and interest lies in the area of Computer Vision and Machine Learning. I intend to further continue my research by pursuing a Ph.D. in the field of visual computing and computer vision.

I started my journey into video understanding by researching about Human Activity Recognition (HAR) at my home university. Part of what inspired me to take up HAR is that it is an integral part to achieving full video understanding and the existing drawbacks in this domain. The Two Stream architecture was a very popular backbone first proposed by Simoyan et al. in 2014. My task was to explore relatively efficient architectures suitable for edge deployment. Instead of keeping such an expensive temporal stream for optical flow processing, I proposed a simple late fusion between a spatiotemporal stream and another spatial stream (I3D and Xception). The findings from this work lead to a publication at the IEEE INDICON 2021 conference.

Meanwhile, I was very fortunate for the opportunity to get to work with Prof. Kunal Chaudhury at the Indian Institute of Science (IISc) also this summer. It enabled me to explore Convolutional Neural Networks (CNNs) and Fully Connected Networks (FCNs) at a much more fundamental level and understand the rigorous mathematical principles behind it (Lots of Matrix Theory, Linear Algebra and Semi definite programming). My task was to explore the robustness of these networks and find out the extent to which the upper bound can be made more tighter using Lipschitz regularization. This internship helped me appreciate neural networks at a much more deeper extent and work with various robustness schemes. I recently returned to IISc again this winter to work under Prof. Anirban Chakraborty, at the Visual Computing Lab where I am currently experimenting with some meta learning approaches to solve the Person Re-Identification problem (ReID).

Also, I and my friend undertook self driven research on our own over the summer. We were inspired by an interesting concept which we came across during our reading week known as “Unrolling” of CNNs. It explained how CNNs could be converted to FCNs without losing any spatial information. We then compared the performance of the “converted CNN” (which we name ResCNN) and a normal CNN and published our results at IEEE UPCON 2021.

At ETH Zurich, I believe the very experience of being surrounded by some of the best brains in the world in this trade and getting an international exposure will certainly mould me to gain a deeper insight in building AI models. Through this research internship, I want to go beyond building offline systems, and build generalizable and scalable real-time models such that they can be deployed in autonomous systems with ease. Particularly, Prof Siyu Tang and the VLG group’s work on human motion and their numerous applications in cloth modelling interests me and I feel it aligns with my background as well. I would be incredibly humbled to get an opportunity to intern at ETH Zurich and look forward to contribute to technologies that will lead the world to a more sustainable future.