A Case for Neuro-Symbolic AI : 2 Instances of Image Generation (Vision) & Learning to Build Complex Structures (Robotics)

Dr. Parag Singla*

Department of Computer Science and Engineering, Indian Instiute of Technology, Hauz Khas, New Delhi

Email: parags@cse.iitd.ac.in

The last decade has seen phenomenal growth in application of neural models to a variety of problems, including those in Computer Vision, NLP and Robotics & Sciences, among other domains. One of the recent research directions in the deep learning world has been around the problem of incorporating symbolic & object-centric representation in neural networks, to exploit the underlying problem structure, and enable them to be more accurate by exploiting domain knowledge and attain effective generalisation while also being more interpretable. In this talk, we will present two different problem instances from 2 different application domains, viz., (a) image generation (b) learning to construct grounded spatial concepts. We will also discuss corresponding solutions that exploit the underlying problem structure for building effective neuro-symbolic architectures to solve these problems. First, we deal with the problem of text-conditioned image generation. Our approach reposes the problem of image generation, as one of generation followed by (multi-step) editing, where the editing specifically focuses on objects & their properties which do not align with original instruction, identified via an MLLM. In the second instance, we look at the problem of learning complex structures from data (e.g., a staircase is a sequence of towers of increasing height), and explore how these can be learned by a robotic agent from human demonstrations, in the form of executable (python) programs. Our approach combines MCTS for guiding program search, with LLMs to aid generalisation of learned programs. For each of these problem instances, we will have a discussion on potential application of the ideas presented to translational science.