UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Seminar

Thursday May 25, 2023 at 11am in Reynolds 1101

Jeetendra (Viraj) Guntamukkala

Line Labeling of Polyhedral Scenes: Comparing Performance and Properties of Different Neural Architectures

Advisor: Dr. Neil Bruce
Advisory: Dr. Minglun Gong

Abstract:

Many advancements have been made to train machines to recognize and interact with objects through understanding their geometry and the configuration of lines that make up their structure. A classic problem in computer vision addresses this in considering the configuration of edges among three separate categories, convex, concave and outer or occluding edges. Knowing the properties of each edge yields the capability to recognize and interact with objects. It is unclear to what extent different classes of neural networks are capable of relating geometry across an object to produce valid labeling of lines.

To investigate this problem, we introduce a dataset generator that automatically, without any manual input, generates random two-dimensional polyhedral scene images, along with ground truth labels for the object's edges. Through this generated dataset, we are able to perform a comprehensive benchmark of the performance of various state-of-the-art Convolutional Neural Networks (CNN), Recurrent Neural Networks, and Vision Transformers performing semantic segmentation to classify edge category.

The results show that frameworks with decoders that gradually recover spatial information exhibit superior segmentation performance over frameworks with lightweight decoders using bi-linear interpolation. In particular, the U-Net and Swin-UNet architectures perform the best among the evaluated models. Moreover, results show that encoding using transformers slightly outperforms encoding via CNNs but the performance is much closer than one might anticipate from performance on other problems in computer vision.

We also notice that recurrent algorithms can provide equal or better performance than some CNN and Transformer based algorithms, with significantly fewer parameters. This presumably arises from the possibility to perform constraint propagation between edges in an iterative fashion. As a whole, this work can help bring about advancements in network design for challenging vision problems and also demonstrates some canonical properties important to model performance for challenging vision problems relating to the 3D structure of objects.