SOLO V1: Detecting, and Segmenting Objects

Tyler Edwards (SEAS '24), Nicholas Kuo (SEAS '24)

Mentor: Professor Jianbo Shi

University of Pennsylvania, Department of Computer and Information Science

Funded by Penn Undergraduate Research Mentorship program (PURM)

Background

SOLO (Segmenting Objects by Location) is an algorithm designed by a group at the University of Adelaide for object segmentation, where given an input image, each object within an image is highlighted and labelled as one of a few predefined categories. Object segmentation is important because it assigns meaning to the pixels in an image. Previous approaches to instance segmentation include a top-down approach using bounding boxes (Mask R-CNN) and bottom-up approach using pixel grouping with embedded vectors. Solo produces a grid of classifications, trying to predict whether an object falls into a specific subsection of the image. For each cell in the grid, a mask of the guessed object is proposed. Using non-maximal suppression the guesses are combined into a cohseive prediction.

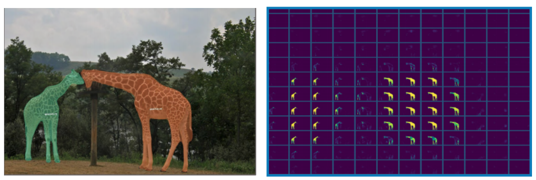

Figure 1: SOLO's Instance Segmentation
SOLO's instance segmentation of giraffes, along with its corresponding activation map of soft mask predictions

Objective

Given an input image of different shapes (squares, and circles) with randomized sizes, orientations, and positions, our goal was to program and train the SOLO algorithm to detect, classify, and segment each shape in the image. While SOLO is capable of identifying real-world objects given an input image, we focused the training of our model on simple shapes due to time and GPU constraints. The goal was for to build the model to accurately segment these images as a proof of concept.

References

Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, & Lei Li. (2020). SOLO: Segmenting Objects by Locations.

Tasks and Procedure

1) Create shapes dataset for data loading, using an open-sourced algorithm that generates random shapes in random locations, in addition to ground truth matrices for each shape's mask segmentation and instance category classification

2) Program mask and class loss functions as outlined by SOLO research paper

3) Design backbone and neck of SOLO model, importing pre-trained ResNet-18 backbone and building fully convolutional network (FCN)

4) Design classification and mask heads

5) Implement inference functions of non-maximum supression, confidence score assessment, and intersection over union

6) Design segmentation visualization using "seaborn" package to generate heatmap of SOLO model's predictions

7) Train SOLO algorithm on shapes dataset, adjusting hyperparmeters according to class and mask loss

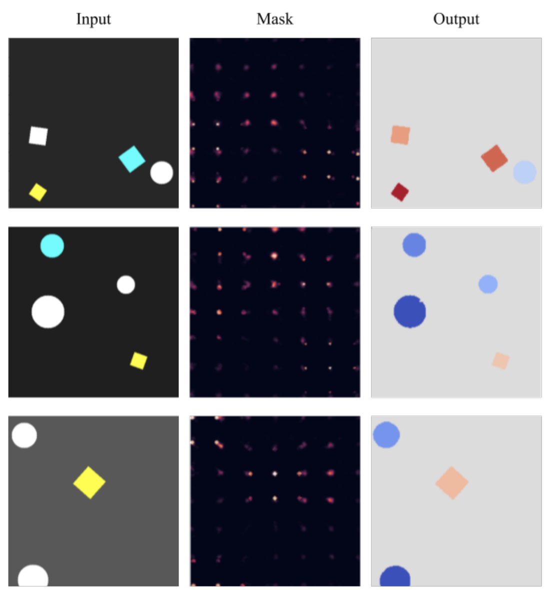

Figure 2: Example Predictions
Example input images (left). Soft mask predictions (middle). Final mask outputs (right). Blue areas are predicted circles, red are squares.

Results

Through the PURM process, we learned more about how to effectively conduct research with the help of a research mentor. Two to three times a week, Professor Shi would guide us through a topic within computer vision (Fig. 3) such as convolution, transformers, and feature pyramid networks. Through these lectures and assigned readings we built a knowledge base. From here, we decided to implement an image segmentation model (objects in an image get highlighted and classified) from scratch (Fig 2). Through this process, we learned how to synthesize multiple papers, use Google Colab as well as Pytorch, and work together on complex machine learning projects.

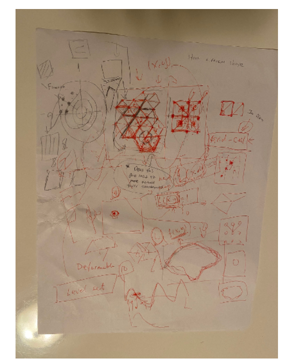

Figure 3: Professor Shi's Notes
Example notes made by Professor Shi during one of his lectures on hexagonal lattices and grid cells.

Conclusion

Our research this summer has given us the perfect springboard to continue work within computer vision. Tyler plans on working with Professor Shi on new projects discussed during PURM. The research this summer exposed him to many new ideas, and created many new questions related to graph theory, human memory, and generative adversarial networks. Nick plans to continue to learn about affordances, causality, and detection with tranformers, using this background in computer vision in his future software endeavors.