# Mask R-CNN and Its Applications


## By Ryan Murphy

### CS420: Artificial Intelligence


## Introduction

The goal of this research is to explain the technology behind the machine learning model Mask R-CNN. Its technological background, improvements, and current implementations will also be explored

**What is Mask R-CNN? [3]**

Mask R-CNN (Mask Region-Based Convolutional Neural Net) is a neural net model used to perform the image processing task of instance segmentation on images.

Figure 1. Source: [3]

- Image Processing: Performing tasks on a digital image to achieve a function
- Instance segmentation consists of two main tasks when given an image:
    - Correctly identify all objects in an image
    - Precisely segmenting each instance of the object in an image
- Each pixel has to be classified in the image as either:
    - Part of an object
    - Background
- Supervised machine learning model
- Training set must be provided to the neural net for it to recognize objects
    - Training set consists of an image, a mask, and an object name

**What are the benefits of using Mask R-CNN over other Instance Segmentation Models?**

- Pixel accurate instance segmentation due to mask generation
- Extension of Faster R-CNN:
    - Well documented
    - Real-Time Classification
- Easy training process, transfer learning is also possible

**What are the basics of How Mask R-CNN works? [1]**

- Mask R-CNN is an extension off of Faster R-CNN
- Faster R-CNN only places bounding boxes around objects in the image
    - Two Outputs: a class label and bounding box
- Mask R-CNN adds a third output in the form of a mask branch

## Technological Background

**What Technology is Mask R-CNN built upon? [1][3]**

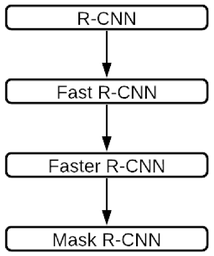


Figure 2.

Mask R-CNN is based off of Faster R-CNN, which utilizes pieces from the older versions of the R-CNN technology.

**R-CNN**: original model developed

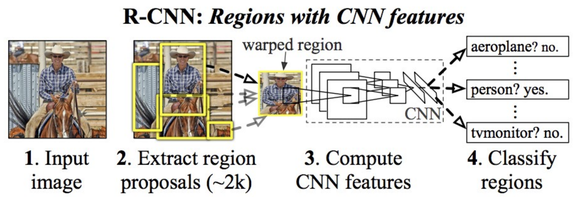


Figure 3. Source: Towards Data Science

- ~2000 random bounding boxes are generated
- Each region is then passed into a CNN
- CNN classifies any objects in the ROI
- Very slow due to ROI processing time

**Fast R-CNN [4]**: Changed the way that ROIs are generated

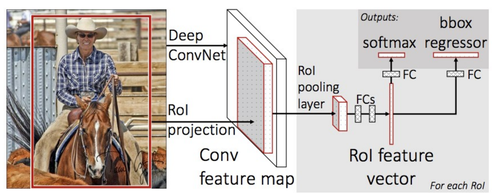


Figure 4. Source: [4]

- Entire image is fed into the CNN
- Generates a convolutional feature map
- The ROIs are then chosed based off of this feature map (Selective Search)
- Much faster than R-CNN
- Region Proposals are still the bottleneck in this model

**Faster R-CNN [6]**: Selective Search is replaced by a Region Proposal Network (RPN)

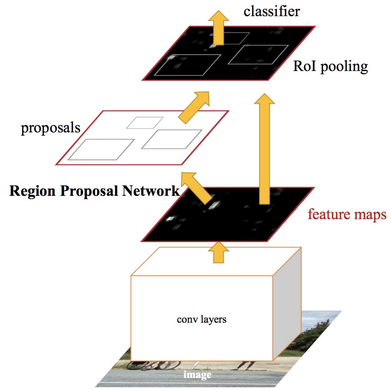


Figure 5. Source: Towards Data Science

- Model that Mask R-CNN is based on
- The whole image is fed into a CNN
- An RPN then predicts the region proposals

## What Does Mask R-CNN Change? [1][3]

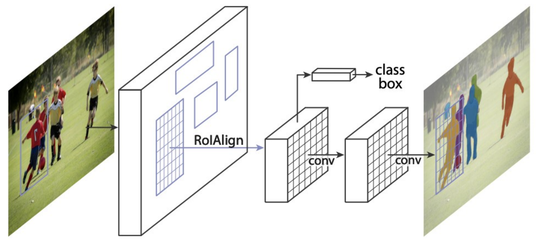


Figure 6. Source: [3]

- Mask R-CNN generates ROIs similar to Faster R-CNN.
- Object class and bounding box are also generated.
- Next the bounding box is refined using ROIAlign
- ROIAlign works by calculating the best bounding box based off of a series of projected bounding boxes
- Finally, a binary mask is generated using the refined bounding box.

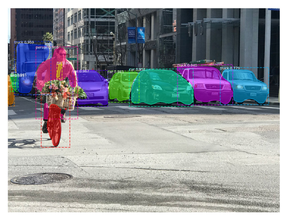

Figure 7. Source: [3]

## Current Applications of Mask R-CNN

**Fruit Detection for Strawberry Harvesting Robots [7]:**

- Used to detect strawberries for harvesting robots

Figure 8. Source: [7]

- Utilized over classical image processing identification due to variability:
    - Stems or leaves obstructing view of strawberry
    - Varying light intensity
    - Reflections and glare
    - Color variations (Ripe and Unripe)
- Mask R-CNN was able to accurately detect the different categories of strawberry as well as their locations in a given image within 1.2mm

**Automatic Knee Meniscus Tear Detection [2]:**

- Mask R-CNN as well as other instance segmentation methods were used in order to classify MR images of the knee
- Within these images, the tear presence, location, and orientation needed to be identified
- Mask R-CNN was capable of detecting the location and orientation of the tears

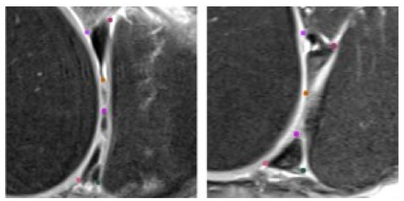

Figure 9. Source: [2]

## Conclusion

- The current state of Mask R-CNN makes it useful for any instance segmentation task
- Its speed and accuracy outmatch many of the other instance segmentation models available
- Its current uses will only continue to expand as more industry tasks become automated, necessitating effective machine vision

## Bibliography:


[1] Bharti P., Pramanik A., 2019, Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey: Advances in Intelligent Systems and Computing, ASIC vol. 999.

[2] Couteaux V., Si-Mohamed S., Nempont O., et al., 2018. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diagnostic and Interventional Imaging 100(4): 235-242.

[3] He K., Gkioxari G., Dollar P., Girshick R., 2017. Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV): 2961-2969.

[4] Girshick R., 2015. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV): 1440-1448. [5] Bharti P. and Pramanik A., Deep Learning Techniques-R-CNN to Mask R-CNN: A Survey by Bharti P. and Pramanik A.

[5] Khan M., Akram T., Zhang Y., and Sharif M., 2021. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. Pattern Recognition Letters 143: 58-66

[6] Ren S., He K., Girshick R., and Sun J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. NIPS.

[7] Yu Y., Zhang K., Yang L., and Zhang D., 2019. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Computers and Electronics in Agriculture 163. This paper has a connection to the Mask R-CNN paper by He K., et al.