Mask R-CNN is a deep learning algorithm for object detection and instance segmentation, which was introduced by Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick in 2017. Mask R-CNN framework is an extension of the Faster R-CNN model by adding a mask branch which allows the model to predict pixel level masks for object instances in an image.

Mask R-CNN framework is a two stage model. The first stage is called Region Proposal Network (RPN), it proposes regions of an image that are likely to contain objects. The second stage is Region based CNN (RCNN), it refines and localizes these proposals. The key innovation that make Mask R-CNN successful is the addition of a third branch of the Faster R-CNN architecture, which predicst a binary mask for each object proposal. This branch enables Mask R-CNN to perform instance segmentation by generation pixel-level masks for each object in the image. Mask R-CNN is trained end-to-end with a multi-task loss function that includes a classification loss, a bounding box regression loss, and a mask prediction loss. Mask R-CNN is trained in a two-stage process. In the first stage, the RPN is trained to generate high-quality region proposals. In the second stage, the RCNN and mask branch are trained using the region proposals generated by the RPN. The RCNN is trained to classify the proposals into object categories and to regress the bounding box coordinates, while the mask branch is trained to predict a binary mask for each proposal. Mask R-CNN has been applied to a wide range of computer vision tasks, including object detection, instance segmentation, and human pose estimation. It has achieved state-of-the-art performance on benchmarks such as Pascal VOC or COCO, which is a large-scale object detection, segmentation, and captioning dataset. Its ability to perform both tasks simultaneously makes it useful in a variety of applications such as autonomous vehicles, robotics and surveillance.

In conclusion, Mask R-CNN is a powerful deep learning algorithm for object detection and segmentation that builds on the Faster R-CNN architecture by adding a third branch. It has achieved state-of-the-art performance on several benchmark datasets demonstrates its effectiveness in real world scenarios. As deep learning research continues to progress it is likely that we will see further improvements in the accuracy and efficiency and explore new applications for Mask R-CNN and related techniques.

References:

K.He, G. Gkioxari, P. Dollar and R.Girshick, (2017) “Mask R-CNN”, Proceedings of the IEEE International conference on computer vision.

Girshick, R. (2015). “Fast R-CNN”. ICCV.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems.

Tutku Tashkan