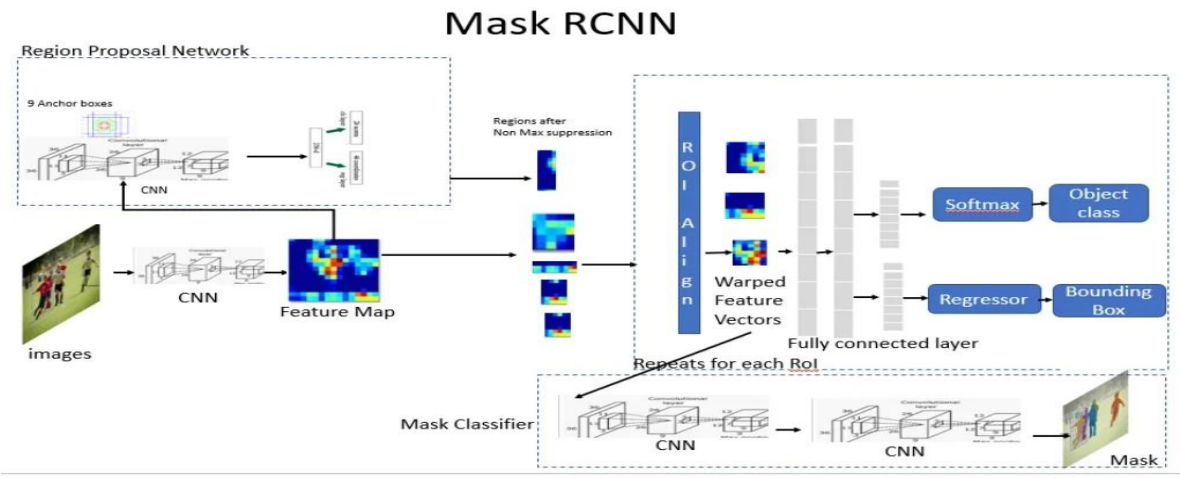
Part 3: Object Detection

1. Explain what is Mask R–CNN

A deep learning system called Mask R-CNN is applied to segmentation and object detection tasks. It is a development of the well-known Faster R-CNN method, which can identify items in a picture and provide the bounding boxes for those things. In addition to detecting the objects, Mask R-CNN gives each item a segmentation mask at the pixel level. It is able to pinpoint the precise pixels in a picture that correspond to each item.

There are two steps in the Mask R-CNN's fundamental design. The final object recognition and segmentation results are obtained in the second step by employing a fully convolutional network (FCN) to refine the collection of potential object proposals that were generated in the first stage using a Region Proposal Network (RPN).

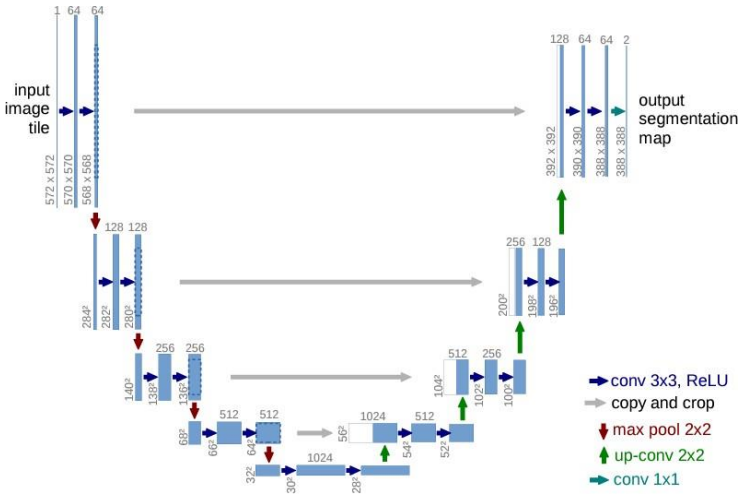


source: https://towardsdatascience.com/computer-vision-instance-segmentation-with-mask-r-cnn-7983502fcad1

The features from the backbone network (such as ResNet) are sent into the FCN branch of Mask R-CNN, which generates a feature map. Then, using this feature map, two outputs are produced: an object detection output (bounding box coordinates and class probabilities) and an instance segmentation output (a binary mask that shows which pixels are part of the object and which are not). Mask R-CNN employs a set of loss functions that promote precise object recognition and segmentation in order to train the model. For the object detection task, the loss function consists of a binary cross-entropy loss; for the segmentation task, it consists of a dice loss and a binary cross-entropy loss.All things considered, Mask R-CNN is an effective object recognition and segmentation technique has uses in domains including robotics, computer vision, and self-driving automobiles.

2. Explain U-Net

A popular convolutional neural network design for image segmentation applications, such as determining the borders of objects inside an image, is U-Net. The contracting path and the expanding path are the two primary components of the U-Net design. The max pooling layers that come after a sequence of convolutional layers make up the contracting route. These layers increase the amount of feature mappings while progressively lowering the input image's spatial resolution. The network may learn ever more intricate representations of the input picture through this approach.



Source :efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/1505.04597.pdf

The expanding route is made up of a sequence of convolutional layers that come after a series of upsampling layers. These layers progressively reduce the number of feature mappings while increasing the output's spatial resolution. The network can produce a segmentation mask with the same size as the original input picture thanks to this procedure. The skip connections between the contracting and expanding channels are a fundamental component of the U-Net design. By combining detailed geographical information from the expanding path with high-level semantic information from the contracting path, these connections enable the network to enhance the accuracy of the segmentation findings.

The output of the network is subjected to a pixel-wise softmax function at the conclusion to provide a probability map of the segmentation mask. After that, a binary mask of the segmented object may be created by thresholding the segmentation mask.

Numerous applications, including autonomous driving, satellite image processing, and medical picture segmentation, have made extensive use of the U-Net architecture. The U-Net design has the benefit of handling a wide range of picture sizes and shapes, which makes it appropriate for a variety of real-world situations. In addition, compared to other segmentation designs, it contains comparatively fewer parameters, which may facilitate training on smaller datasets. In conclusion, the U-Net architecture is a popular choice for image segmentation applications since it is a strong and adaptable convolutional neural network. It can enhance segmentation accuracy by combining specific spatial information with high-level semantic information through skip links between the contracting and expanding routes.

3. What is intersection over union (IOU)

A typical statistic for assessing the effectiveness of object identification and segmentation algorithms is Intersection over Union, or IoU. It calculates the overlap for a particular object between the ground truth bounding box or segmentation mask and the expected bounding box or mask.The area of union between the two regions is divided by the area of intersection between the anticipated and ground truth regions to determine the IoU:

IoU = Area of Intersection / Area of Union

The IoU equals 1.0 if there is a full overlap between the anticipated and ground truth zones. The IoU is equal to 0.0 if there is no overlap between the areas. IoU is frequently used to gauge how effectively an object identification or segmentation algorithm locates and categorizes things in an image. An algorithm's ability to recognize and find objects more precisely is indicated by a higher IoU, whereas a lower IoU suggests that the method is less accurate.