影像分割 Image Segmentation — 實例分割 Instance Segmentation(1)

李謦伊 Follow
Jul 25 · 8 min read

在之前的文章中介紹過影像分割的相關應用及各種語義分割 (Semantic segmentation) 的演算法，本文將要來介紹實例分割 (Instance segmentation) 的代表演算法、paper、code。

目前的實例分割方法分為四種: 基於物件偵測 top-down 方法、基於語義分割bottom-up 方法、綜合 top-down 及 bottom-up、直接分割方法，而早期的實例分割方法偏向 top-down、bottom-up。

由語義分割的方法可看到 FCN 在各個演算法中都占了很重要的地位。然而為何 FCN 不能做到實例分割任務呢？因為卷積的平移不變性 (translation invariant)，導致每一個像素只能對應一種語義，因此 FCN 只能對於每個像素進行分類，並不能區分出獨立的物體。

平移不變性是指**即使目標物的位置發生變化也不會改變預測結果**，這個特性有利於 image classification 任務，就算目標物位置改變，其預測類別還是相同。詳細可參考知乎上的回答，Hengkai Guo 說明得非常仔細: 既然cnn對圖像具有平移不變性，那麼利用圖像平移（shift）進行數據增強來訓練cnn會有效果嗎？

基於語義分割**bottom-up** 方法

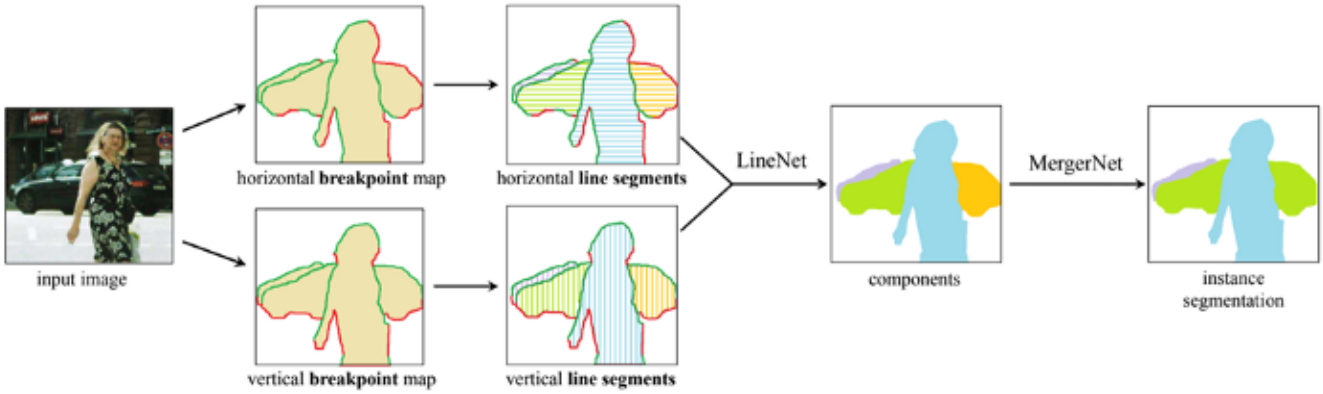
先來介紹基於語義分割 bottom-up 方法，該方法是先對影像進行 mask 預測，再將這些 mask 像素進行分組，產生各個實例。有以下缺點:

- 較依賴 mask 的預測質量，容易導致非最優的分割。
- 對於複雜場景的分割能力有限，因為 mask 是在低維特徵圖中提取的。
- 需要較複雜的後處理以產生各個實例。

由於這些缺點，基於 bottom-up 方法的演算法並不多，如 SGN、Semantic Instance Segmentation with a Discriminative Loss Function 等。

Sequential Grouping Networks for Instance Segmentation (SGN，ICCV 2017)

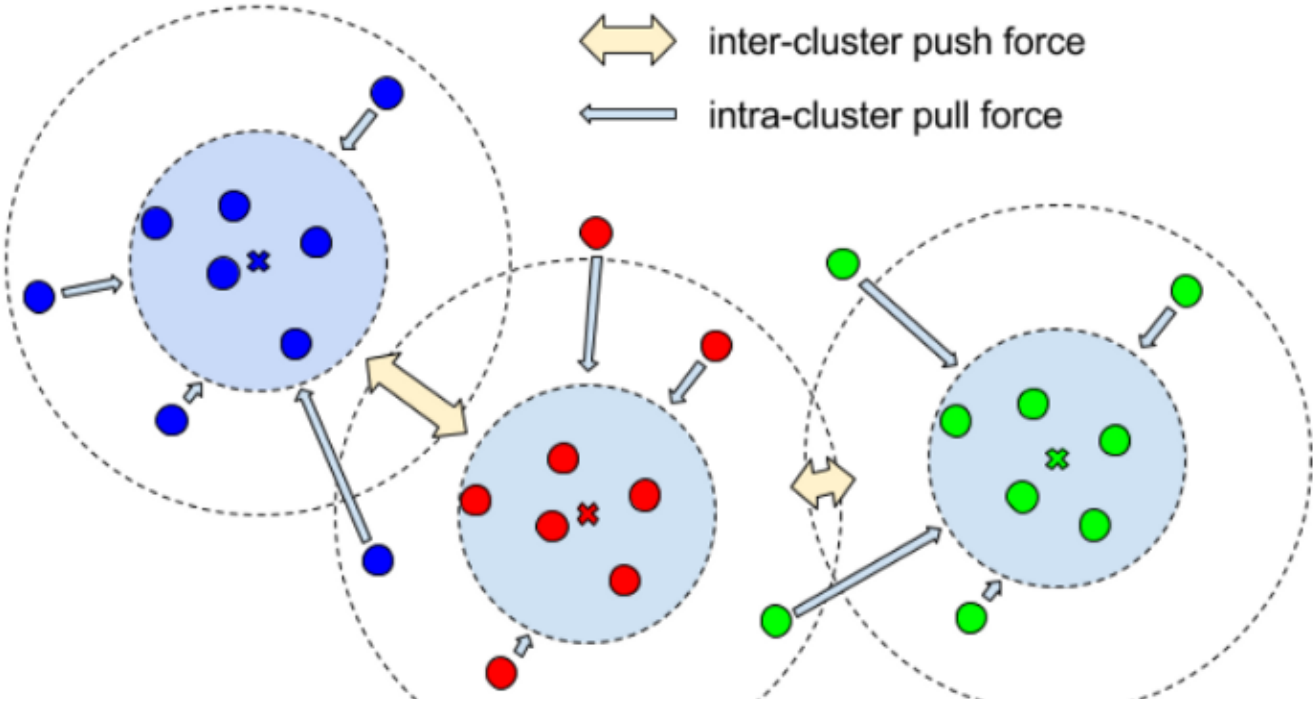
SGN 採用三個網路來進行預測：第一個網路沿著像素預測水平和垂直的物體斷點來產生分割線，第二個網路將這些水平和垂直的分割線進行組合，第三個網路則是將這些組合的分割線聚合成一個實例。



Semantic Instance Segmentation with a Discriminative Loss Function (CVPR2017)

Github: https://github.com/Wizaron/instance-segmentation-pytorch

透過 Discriminative Loss (variance、distance、regularization loss) 訓練模型，使得同一個實例的像素更加靠近、不同的實例像素盡可能地遠離，最後再使用 Mean-shift 進行聚類。



基於物件偵測 **top-down** 方法

top-down 方法就是先得到物件檢測框，再對框內的像素進行 mask 預測。代表演算法有 DeepMask、SharpMask、InstanceFCN、FCIS、Mask R-CNN、Mask Scoring R-CNN、YOLACT

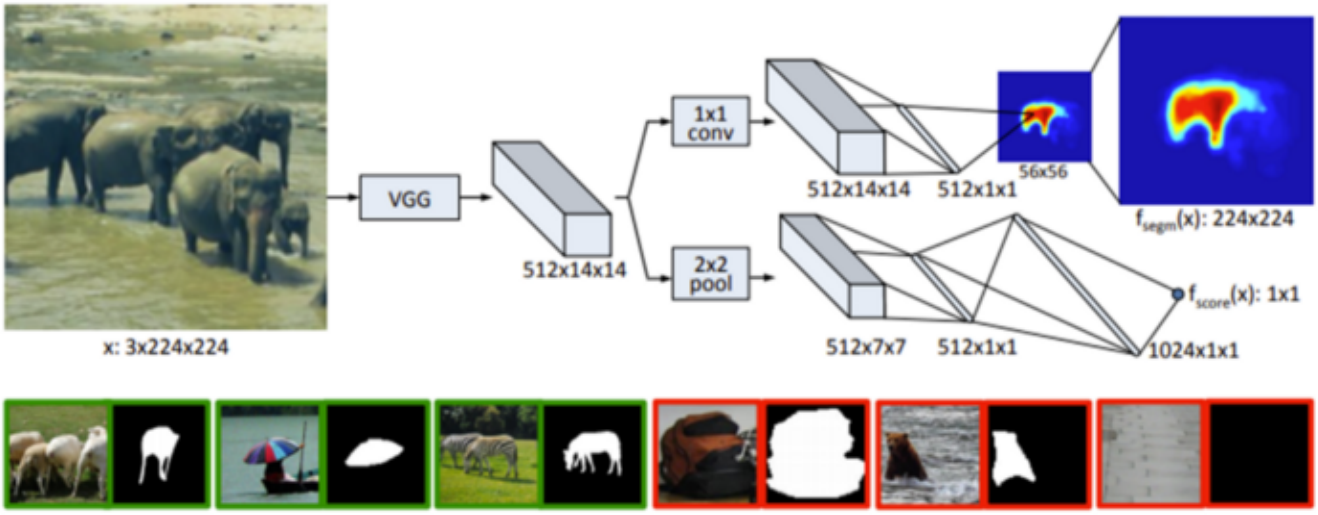
DeepMask (2015)、SharpMask (2016)

Github: https://github.com/facebookresearch/deepmask

DeepMask 是實例分割的始祖，其網路架構為使用 VGG-A 作為 backbone，去除全連接層及最後一個 max-pooling 層，再分為 Mask、Score 分支同時訓練。

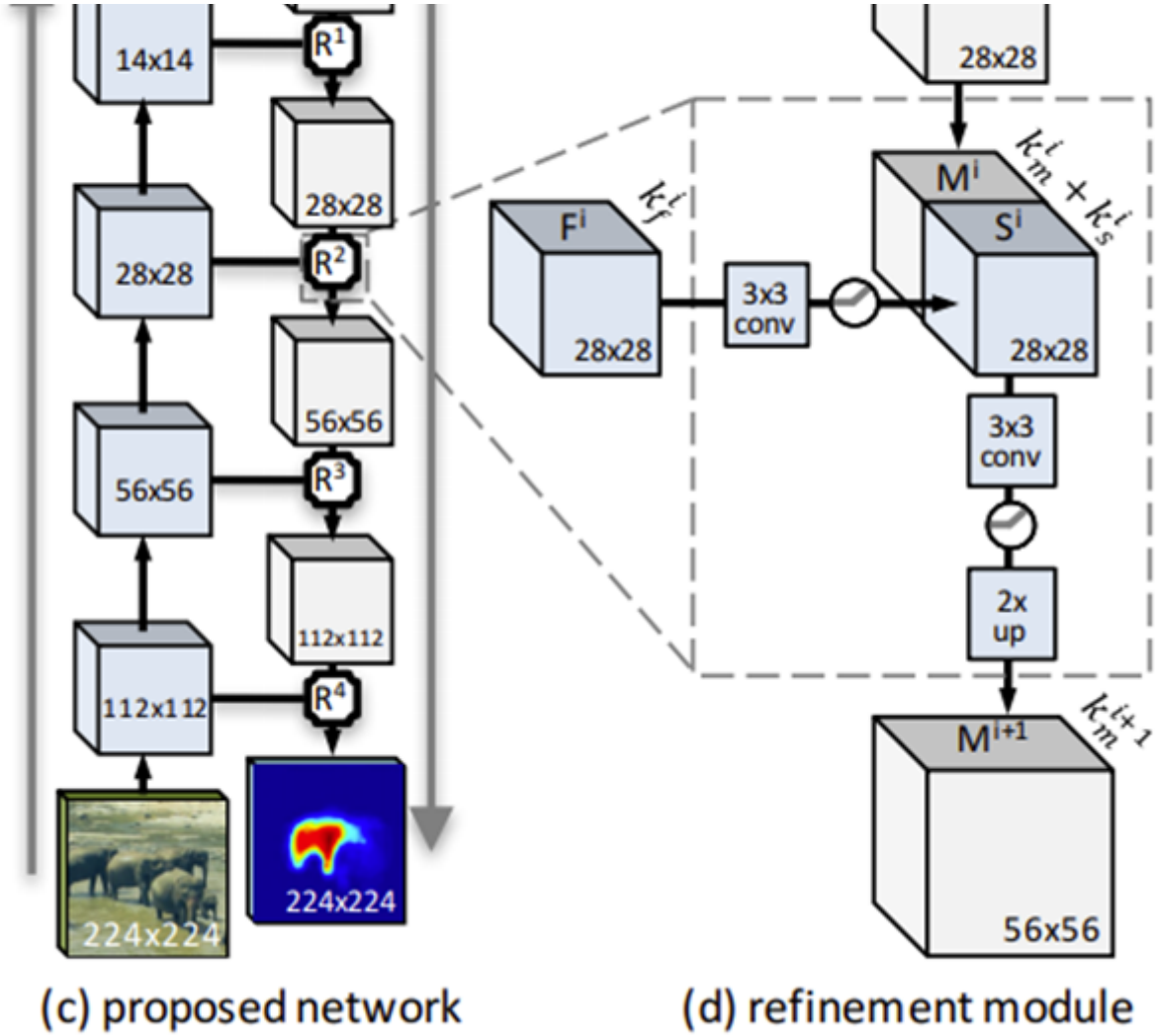
Mask 用來實現分割任務，輸入 Image patch 預測該 patch 的每個像素是否屬於目標物中心。Score 則是對目標物的預測分數，必須滿足兩個限制:

1. 目標物位於 Image patch 中心點附近
2. 目標物有一定的範圍包含在 Image patch



DeepMask

SharpMask 以 DeepMask 為基礎做改進，加入了 refinement 來解決輸出 mask 較粗糙的問題。其思路為將 high-level 的物體訊息融合 low-level 的特徵能夠得到較精細的 mask。

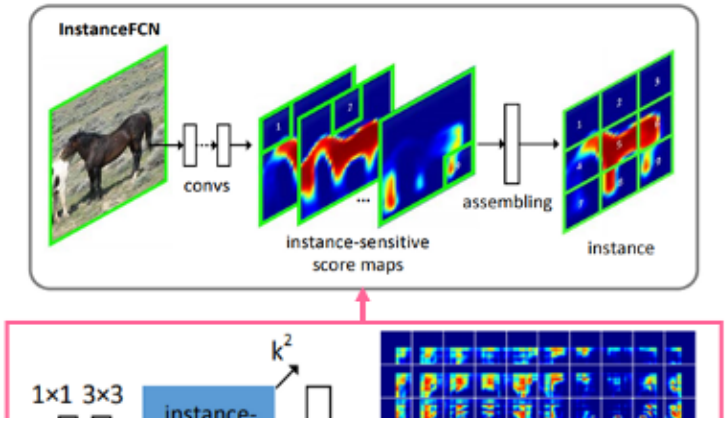


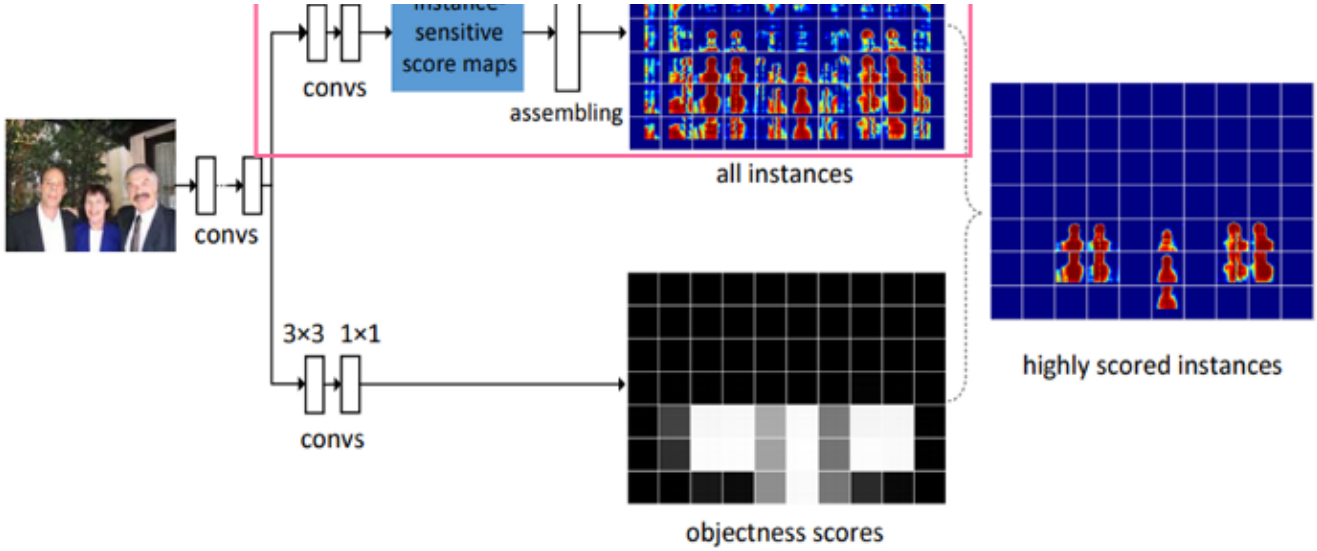
SharpMask

InstanceFCN (ECCV 2016)

InstanceFCN 提出了 Instance-sensitive score map 用來預測每個像素的值屬於哪個類別的相對位置，藉由預測相對位置解決目標物重合的問題，以區分出獨立的物體。接著將 score map 輸入至 Instance assembling module 得到所有 instance 結果。

網路架構如下圖，使用 VGG 作為 backbone 提取特徵後，分成兩個分支: 一個用於 Instance-sensitive score map、另一個用於輸出 objectness score，目的是要分類以像素為中心的滑動視窗是否為實例。

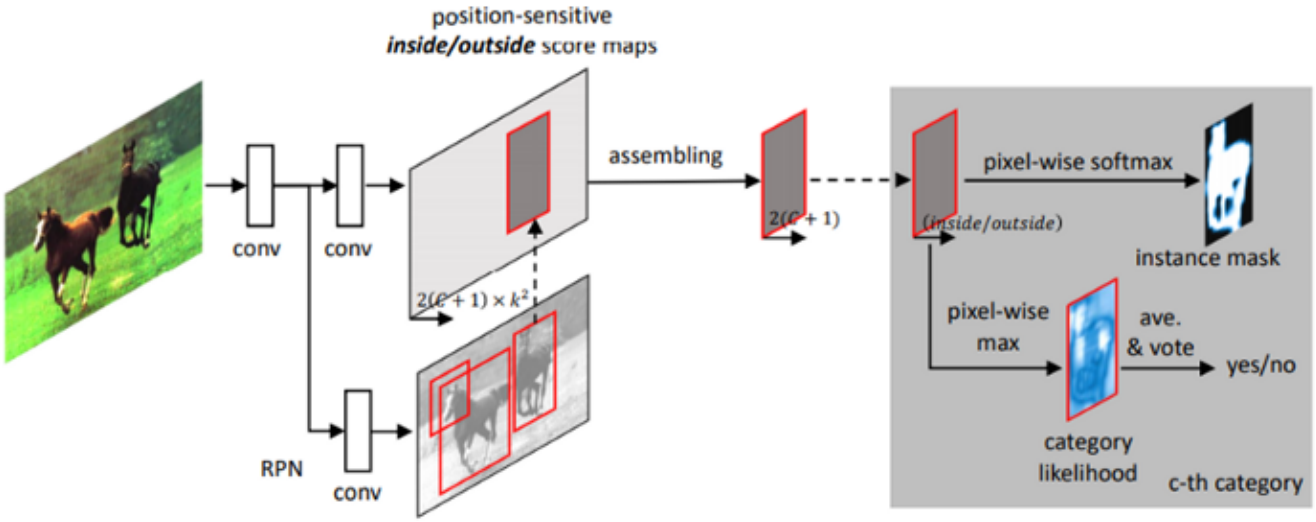




FCIS (2016)

Github: https://github.com/msracver/FCIS

FCIS 是第一個 end-to-end 實例分割模型，基於 Faster RCNN 改進，去除了 ROI-Pooling、全連接層，且在影像分割及分類的任務之間共享參數。此外，參考 InstanceFCN 提出了 inside/outside score maps 達到同時進行分割與分類的預測。在 COCO 2016 實例分割比賽獲得了冠軍。

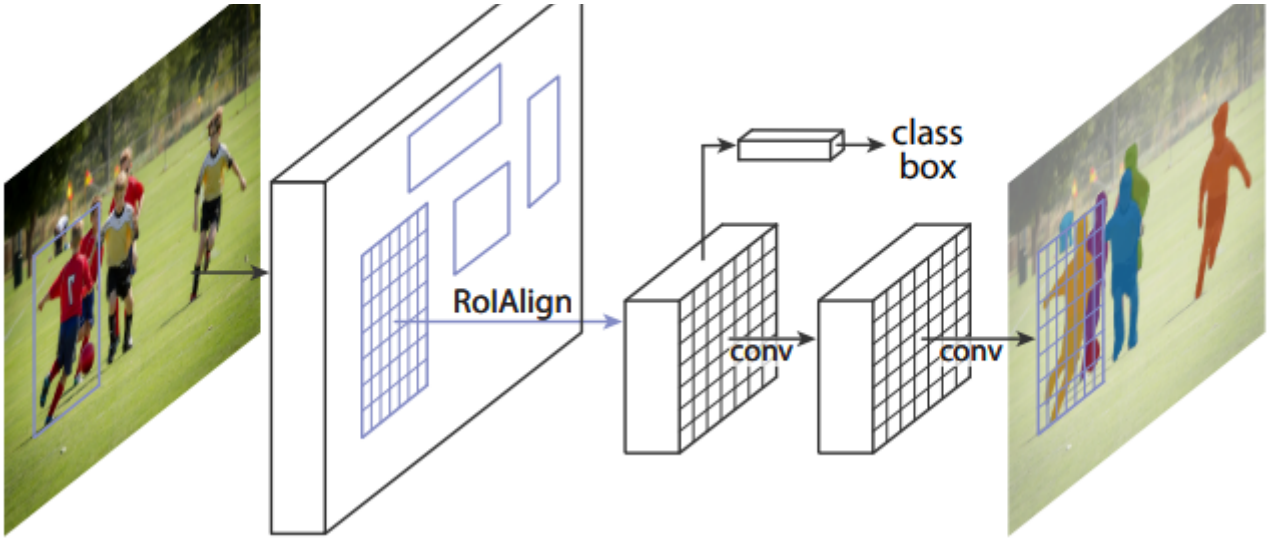


Mask R-CNN (ICCV 2017)、Mask Scoring R-CNN (CVPR 2019)

Github: https://github.com/matterport/Mask_RCNN

Github: https://github.com/zjhuang22/maskscoring_rcnn

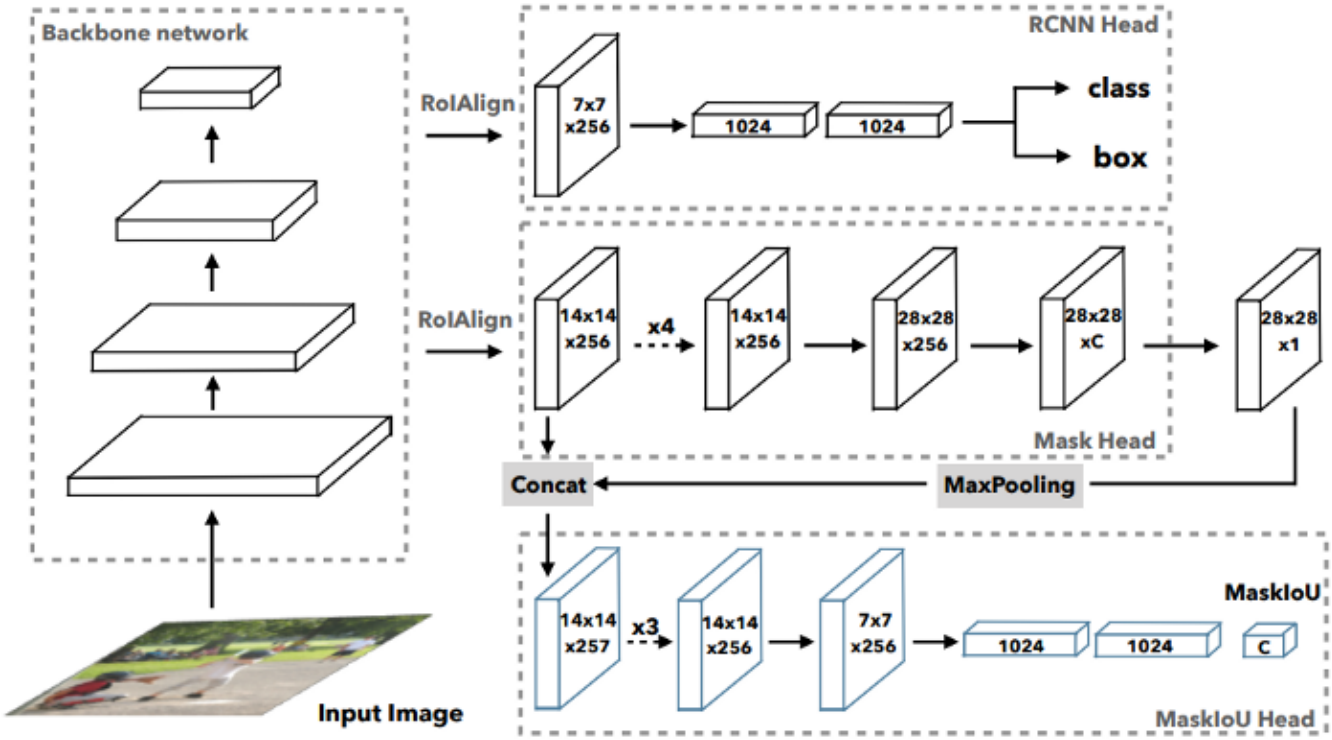
Mask R-CNN 是一個 two-stage 方法，以 Faster R-CNN 為基礎提出了 RoIAlign 代替 ROI-Pooling，並增加了一個全卷積來預測 mask，也就是在每個 ROI 上使用 FCN。



Mask R-CNN

但 Mask R-CNN 存在一個問題，對於 mask 質量分數也採用 class 的 confidence，但事實上這樣的打分方式並不合理，confidence 越高也可能存在 mask 的質量很差的情況。

為了改善這個問題而提出了 Mask Scoring R-CNN — 基於 Mask R-CNN 加入了 MaskIoU 分支，用於預測輸入 mask 與 ground truth mask 的 IOU，作為對 mask 質量的打分。MaskIoU 分支的輸入由 RoIAlign 得到的 feature map 及 mask 分支的輸出進行 concat 所組成。



Mask Scoring R-CNN

YOLACT、YOLACT++、YolactEdge

YOLACT (You Only Look At CoefficienTs) 是首個能夠實現 real-time 的實例檢測模型，詳細可參考: YOLACT (You Only Look At CoefficienTs) 系列介紹

由於內容篇幅過多，剩下的方法會在之後的文章介紹~~

相關文章

影像分割 Image Segmentation — 語義分割 Semantic Segmentation(1)

影像分割 Image Segmentation — 語義分割 Semantic Segmentation(2)

Deep Learning Artificial Intelligence Segmentation Instance Segmentation