Title: Region Based Convolutional Neural Networks

URL: https://en.wikipedia.org/wiki/Region_Based_Convolutional_Neural_Networks

PageID: 63359350

Categories: Category:Deep learning, Category:Object recognition and categorization

---

Region-based Convolutional Neural Networks (R-CNN) are a family of machine learning models for computer vision , and specifically object detection and localization. [1] The original goal of R-CNN was to take an input image and produce a set of bounding boxes as output, where each bounding box contains an object and also the category (e.g. car or pedestrian) of the object. In general, R-CNN architectures perform selective search [2] over feature maps outputted by a CNN.

R-CNN has been extended to perform other computer vision tasks, such as: tracking objects from a drone-mounted camera, [3] locating text in an image, [4] and enabling object detection in Google Lens . [5]

Mask R-CNN is also one of seven tasks in the MLPerf Training Benchmark, which is a competition to speed up the training of neural networks. [6]

History

The following covers some of the versions of R-CNN that have been developed.

November 2013: R-CNN . [7]

April 2015: Fast R-CNN . [8]

June 2015: Faster R-CNN . [9]

March 2017: Mask R-CNN . [10]

December 2017: Cascade R-CNN is trained with increasing Intersection over Union (IoU, also known as the Jaccard index) thresholds, making each stage more selective against nearby false positives. [11]

June 2019: Mesh R-CNN adds the ability to generate a 3D mesh from a 2D image. [12]

Architecture

For review articles see. [1] [13]

Selective search

Given an image (or an image-like feature map), selective search (also called Hierarchical Grouping) first segments the image by the algorithm in (Felzenszwalb and Huttenlocher, 2004), [14] then performs the following: [2]

R-CNN

Given an input image, R-CNN begins by applying selective search to extract regions of interest (ROI), where each ROI is a rectangle that may represent the boundary of an object in image. Depending on the scenario, there may be as many as two thousand ROIs. After that, each ROI is fed through a neural network to produce output features. For each ROI's output features, an ensemble of support-vector machine classifiers is used to determine what type of object (if any) is contained within the ROI. [7]

Fast R-CNN

While the original R-CNN independently computed the neural network features on each of as many as two thousand regions of interest, Fast R-CNN runs the neural network once on the whole image. [8]

At the end of the network is a ROIPooling module, which slices out each ROI from the network's output tensor, reshapes it, and classifies it. As in the original R-CNN, the Fast R-CNN uses selective search to generate its region proposals.

Faster R-CNN

While Fast R-CNN used selective search to generate ROIs, Faster R-CNN integrates the ROI generation into the neural network itself. [9]

Mask R-CNN

While previous versions of R-CNN focused on object detections, Mask R-CNN adds instance segmentation. Mask R-CNN also replaced ROIPooling with a new method called ROIAlign, which can represent fractions of a pixel. [10]

References

Further reading

Parthasarathy, Dhruv (2017-04-27). "A Brief History of CNNs in Image Segmentation: From R-CNN to Mask R-CNN" . Medium . Retrieved 2024-09-11 .