4. GRoIE Architecture

The FPN is an architecture commonly used to extract features from different image resolutions. Starting from a region produced by the RPN, for each scale, a fixed-size RoI is pooled from the region.

The resulting “n” feature maps are, first, separately pre-processed with the objective to apply a preliminary elaboration to the pooled regions and gives the network an additional degree of freedom which is specific for each image scale.

Then, merged into a single feature map.

Finally, post-processing is applied to extract global information, jointly considering all the scales.

This architecture grants an equal contribution of each scale and benefits from the information embodied in all FPN layers by overcoming the limitations inherent in the arbitrary choice of a single FPN layer.

It is worth noting that this procedure is valid for both object detection and instance segmentation.

5. Results

For this experiment, we considered the networks that best represent the two-stage networks [3, 4, 5, 6] and we have thus replaced only the standard RoI extraction modules with GRoIE in its most performing configuration: sum as aggregation function, 5x5 convolution for pre-processing and attention module for post-processing.

It is rather evident that the introduction of GRoIE as RoI extraction layer strongly contributes to an improvement in precision in all the tested architectures.

In these graphs, it can be seen that in later epochs the positive effect of GRoIE increases, suggesting that it can arguably be even higher with more training epochs.

Summary

A novel RoI extraction layer called GRoIE is proposed, with the aim of a more generic, configurable and interchangeable framework for RoI extraction in two-stage architectures for instance segmentation. GRoIE is introduced to the major state-of-the-art architectures to demonstrate its superior performance with respect to traditional RoI extraction layers.

While preliminary, the results reported in this paper are quite promising and seem to indicate the potentiality of GRoIE as novel extraction layer. As a consequence, our future works will concentrate on exploiting the modularity of GRoIE to further enhance the quality of the output features to improve the overall accuracy of different computer vision applications.