ECCV 2018

Deep Feature Pyramid Reconfiguration for Object Detection

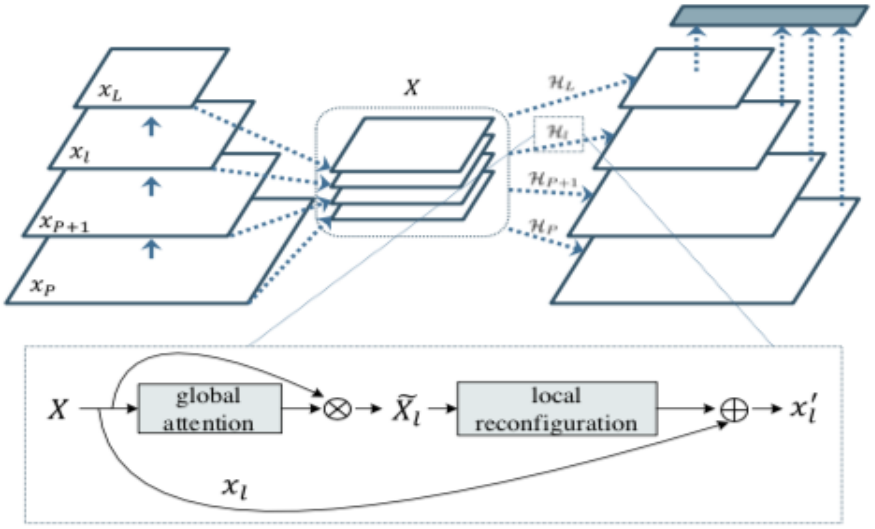
Tao Kong, Fuchun Sun, Wenbing Huang, and Huaping Liu

Overview

The concept of feature pyramid to learn multi-scale representation is generally used by the proposed detectors so far. The design of the feature pyramid is something that can be modified to push accuracy and make the model scale-invariant. Using the ConvNet feature hierarchy the author wants the network to learn information of interest for each pyramid level. The author formulates the feature transformation process as global attention and local reconfiguration problem as to deal with feature hierarchy from different scale

Concept

The experiment was conducted on PASCAL VOC 2007, PASCAL VOC 2012, and MS COCO datasets. ImageNet1k classification set was used for pretraining all network backbones and later fine-tuned on the detection dataset. VGG-16 and ResNet model pertain model was used.
Model is a combination of multiple feature maps, then generates features at a specific level, finally detect objects at multiple scales

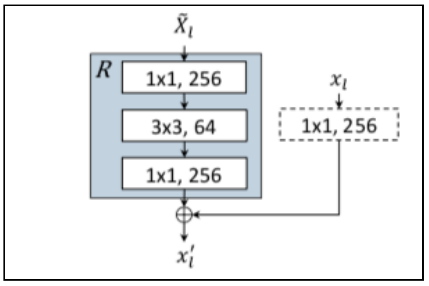


A building block illustrating global attention and local reconfiguration.

The author employs global attention to emphasize global information of the image followed by a local configuration to model local patch within the receptive field:

- **Global attention -** The global attention part is applied to extract more informative features and suppress less informative for a particular scale. As the basic model, Squeeze and Excitation block was used which consists of two steps, squeeze and excitation. The squeeze stage is signified as global pooling on each channel whereas the excitation stage consists of two fully connected layers followed by sigmoid activation taking the input from the squeeze stage.

- **Local reconfiguration -** The motivation of residual is to increase accuracy by increasing the network depth. The input of residual learning is taken from the above Global attention for feature hierarchy. It maps the feature hierarchy directly with the output feature map which is obtained by sliding the operation on input.



Strengths

The following are technical good points of concept used in the paper:

- Global - local configuration is helpful for extracting difficult features as it applies non-linear transformation on the input data

- Lightweight networks are used in doing global attention and local reconfiguration

- To make the method more efficient, the pyramidical processing for all scales are performed together rather than doing layer by layer.

- The advantages, modification, alternatives as well the scope of improvements are mentioned by the author while describing each step

- The details of the datasets used for training and testing are very clear from the paper. Also, several comparisons are made depending on accuracy as well as speed

- The author had made sure to list the advantage or reason to take any particular step making it easy for the reader to understand the intuition of the author.

Weakness

Using deep features drop the accuracy as well some hypothesis are made during the method which is not justified in the paper like semantic information is distributed among feature hierarchy and the residual learn block could select additional information by optimization. The paper is often smooth sail to read but according to me a separate figure explaining global attention would have added more comfortability for the reader to understand the method.

Scope of improvement

1 - As the author himself suggests that there is still room for improvement and potential of developing better pyramids to make the model scale invariant

2- Rather than using pretrained models, if the models are trained from scratch may give more relevant experimentation results

3 - This model could be experimented with the new object detection models.

REVIEWER

Rishab Khantwal

180100095