**CS512 Computer Vision**



**PROJECT PROPOSAL**



**Sukanta Sharma - A20472623**

**Vidya Sudharshana -**

1. Main Paper

This project is based on paper titled “Seeing without Looking: Contextual Rescoring of Object Detections for AP Maximization” by Lourenc¸o V. Pato, Renato Negrinho and Pedro M. Q. Aguiar, available on IEEE Xplore.

1. Problem Statement

Most of the object detectors lack on context: class predictions are independent of detections. Here with reference of the said paper we will try to reimplement and incorporate context in object detection by post-processing the output of an arbitrary detector to rescore the confidences of its detection.

The normal convolution method to detect object usually process the whole image to find suitable object proposal and these are classified independently. On the other side, humans used to have a broad range of contextual cues to recognize objects, by occurrence statistics and relative object locations and sizes.



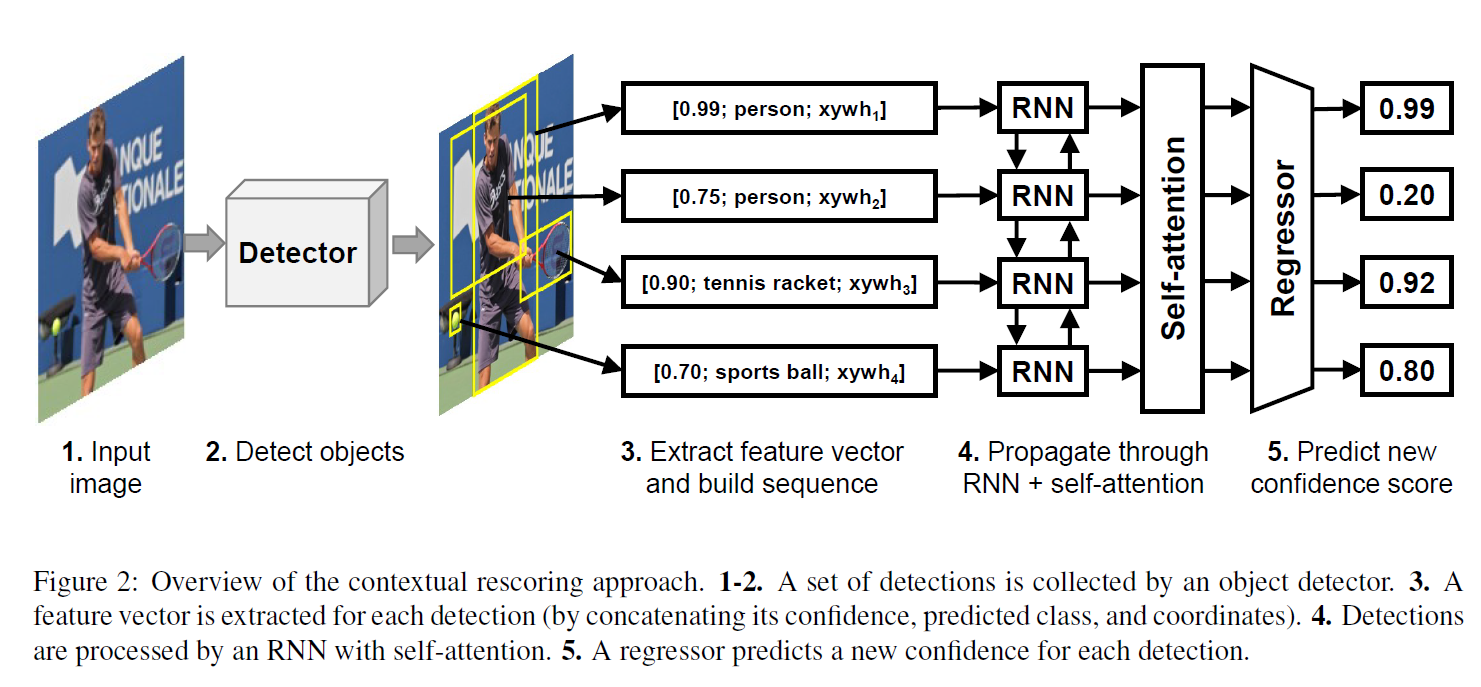
With some error analysis, it can be observed that this current object detector create some error in the suitable outcomes but that can be mitigated by the use of context. There are five steps involved in this contextual rescoring approach: 1. Input Image 2. Detect Objects 3. Extract Feature Vector and Build Sequence 4. Propagate through RNN + self-attention 5. Predict New Confidence Score.

Figure 1: Detection confidences before (left) and after (right) contextual rescoring. High-confidence detections inform the topic of the image. False positives have their confidences reduced (only suitcase and the umbrella are in the ground truth). The line thickness of a bounding box is proportional to its confidence.

The work division for each member is going to be mentioned on the following section. However, the complete detailed division will be given in the final report.

1. Approach

The approach for this project will be similar to the milestones from the paper “Seeing without Looking: Contextual Rescoring of Object Detections for AP Maximization” by Lourenc¸o V. Pato, Renato Negrinho and Pedro M. Q. Aguiar, available on IEEE Xplore as it is shown in the following figure:



* 1. Rescoring Target

AP is computed for each class separately at various IoU thresholds (0.5, 0.55, …, 0.95). Increasing IoU thresholds reward better localization by requiring a detection to be closer to a ground truth to be considered true positive. AP is computed by determining true and false positive by matching each detection with a ground truth. Post this we will perform the following to solve the problem - Greedy maximization of AP, Matching detections with ground truths, Optimal confidence values, Target AP. This part will be performed by Sukanta Sharma.

* 1. Model Architecture

The set of detections is mapped to a sequence of features x ∈ RL×N that is fed to our model that computes the rescored confidences ˆy ∈ RL. Each rescored confidence in ˆyi is generated by conditioning on x (i.e., the whole set of detections). This will be followed by performing Feature Extraction, Recurrent Neural Networks, Self-attention, Regressor, Loss Function. This part will be performed by Vidya Sudharshana.

1. Data

To test all the previously mentioned steps and to generate a classification model, some of the following datasets are going to be used. Existing detectors on MS COCO [22] to generate detections for train2017 (118k images) for training, val2017 (5k images) for model selection, and test-dev2017 (20k images) for evaluation. As baseline detectors, MMDetection’s implementations of Cascade R-CNN and Faster R-CNN with ResNet-101 and ResNet-50 [16] backbones are used.

1. References

Papers of reference:

* Seeing without Looking: Contextual Rescoring of Object Detections for AP Maximization” by Lourenc¸o V. Pato, Renato Negrinho and Pedro M. Q. Aguiar, available on IEEE Xplore.

Other useful links:

<https://docs.opencv.org/master/>

<https://docs.python.org/3/>

https://github.com/LourencoVazPato/seeing-without-looking/