

# Detector Explanation Toolkit - Human-Centric Evaluation

## Task Description

Deepan Chakravarthi Padmanabhan

January 23, 2022

Firstly, thank you for your time. I assure you that I will use the answers solely for research purposes without disclosing any user identity. The evaluation includes two tasks. Task I: Questions 1-7. Task II: Questions 8-10.

### 1 Task I: Which Robot's explanation is better?

- An artificial intelligence (AI) agent performing the task of localizing and classifying all the objects in an image is called an object detector.
- The output from an object detector to detect a single object includes the **bounding box** representing the maximum rectangular area occupied by the object and the **class name** representing the category of the object inside the bounding box. The output is called detection.
- $(x\_left\_top, y\_left\_top)$  and  $(x\_right\_bottom, y\_right\_bottom)$  are the two coordinate points to represent a bounding box. The class name of the object is represented as a text label near the bounding box as shown in Figure 1.

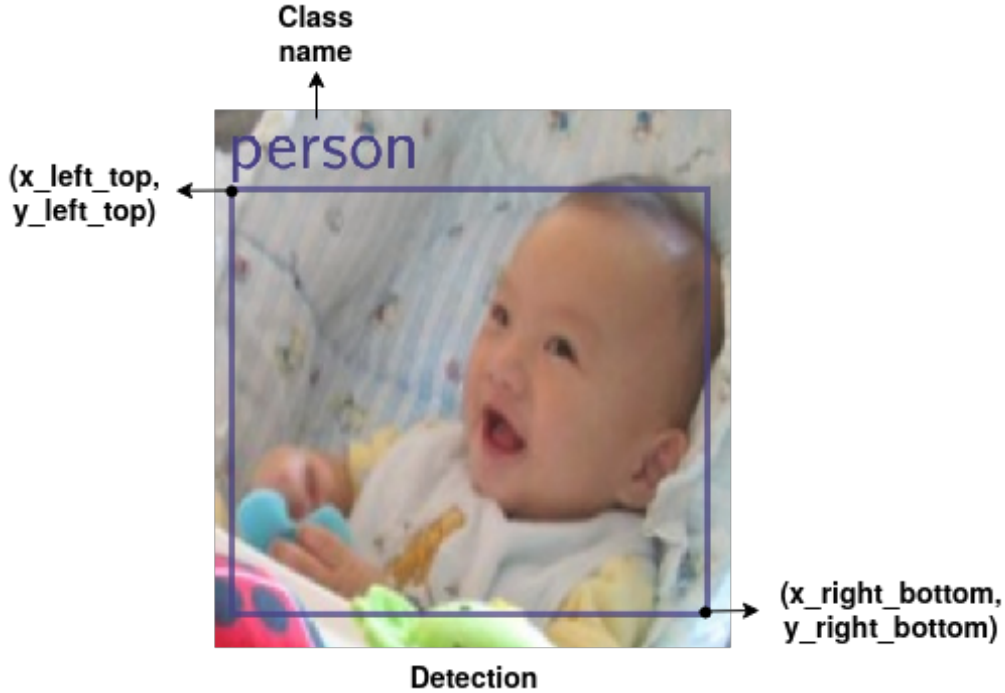


Figure 1: An illustration of a detection output from a detector.

- Therefore, each detection is made of two decisions (predictions), namely, bounding box coordinates decision and classification decision.
- In this study, the reason for a particular decision, say bounding box coordinates or class prediction, in a single detection, is shown.
- This reason behind the decision-making process is given by the explanation. In this task, the explanation is generated by two different robots, **Robot A** and **Robot B**. The explanation images are provided for classification and bounding box decisions separately.

- The explanation for a particular decision is provided by highlighting the pixels important for the decision-making process. The color bar provided in Figure 2 on the right of the explanation image indicates the pixel importance value.

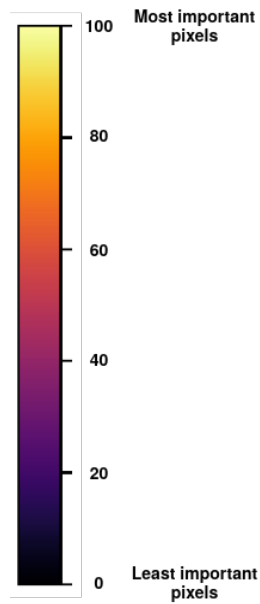


Figure 2: A heatmap representing the importance of pixels for a particular decision.

- In task 1, the author requests you to rate Robot A's explanation by comparing it against Robot B's explanation in terms of understandability and meaningfulness of the explanation.
- A few classification decision explanations, Figure 3 and Figure 4, are provided below:

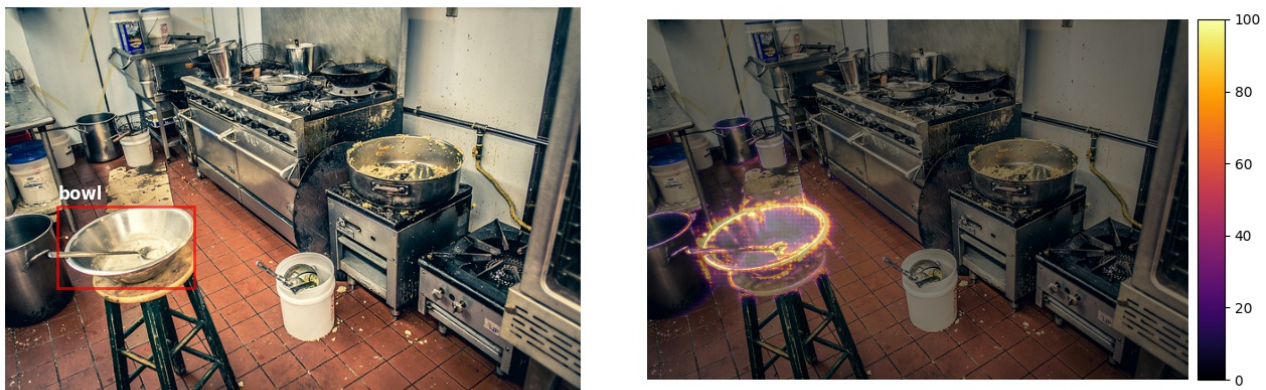


Figure 3: A bowl detection made by the detector (shown in the left). An explanation for the bowl classification decision (right). The most of the important pixels highlight the object detected. The pixel importance values of objects other than the detected object are very less and negligible.

- A few bounding box coordinate explanations, Figure 5 and Figure 6, are provided below:

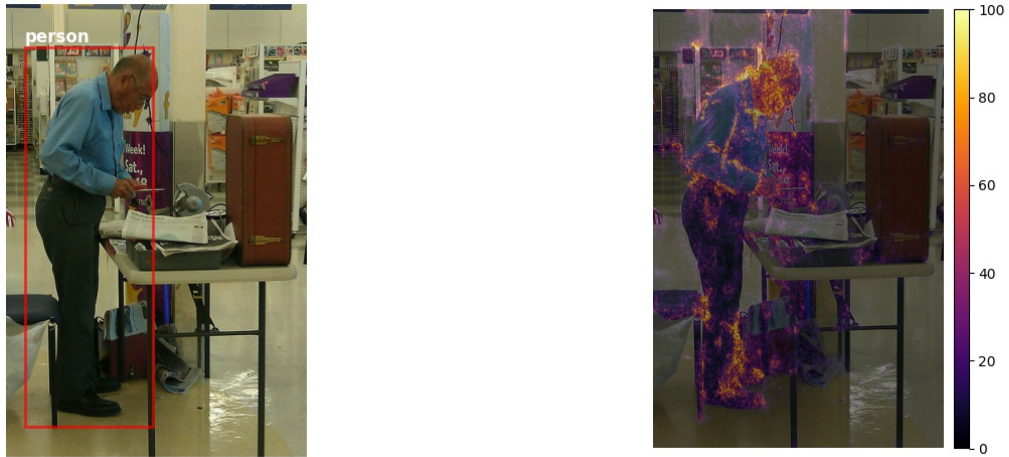


Figure 4: A person detection made by the detector (left). An explanation for the person classification decision (right). The most of the important pixels highlight the object detected. However, the explanations highlight pixels other than the detected object and is highly noisy.



Figure 5: A person detection made by the detector (left). An explanation for the person y\_left\_top coordinate prediction (right). The most of the important pixels highlight the object detected. In addition, the explanation is coherent with the bounding box coordinate as the explanation highlights region near the y\_left\_top.



Figure 6: A person detection made by the detector (left). An explanation for the person x\_right\_bottom coordinate prediction (right). The important pixels highlight the object detected. However, the explanation highlights numerous pixels outside the the detected object and is highly noisy.

## 2 Task II: Which method is better to summarize all detections and corresponding explanations?

- Each image shown in this task includes all the detection made by the detector. Similar to the previous task, each detection is represented as shown in Figure 1.
- In addition to all detections, each image illustrates the explanation for a particular decision, say bounding box coordinate or classification result, for all objects detected by the detector.
- In order to map the detection and the respective explanation, the same colors are used.
- The explanations are represented using 4 different methods. However, visually, across the 4 methods, the important pixels responsible for a particular decision are highlighted using either dots, ellipses, or irregular polygon.
- For ellipses and irregular polygon, the pixels inside the ellipse and irregular polygon are the important pixels responsible for the decision-making process.
- One of the options is “None of these”. This option can be selected when the detection and corresponding explanation of multiple objects illustrated in all 4 images are confusing and illegible to coherently understand the detection and the corresponding explanation.