

# U-CAM: Visual Explanation using Uncertainty based Class Activation Maps (Supplementary Material)

Badri N. Patro   Mayank Lunayach   Shivansh Patel   Vinay P. Namboodiri  
 Indian Institute of Technology, Kanpur  
 { badri,mayank1,shivp,vinaypn }@iitk.ac.in

## 1. Introduction

This is the supplementary material for the paper ‘U-CAM: Visual Explanation using Uncertainty based Class Activation Maps’.

## 2. Questions for Figure 5(c, d)

Table 1 and 2 have the questions corresponding to ID’s present in the Figure 5 of the main paper.

Question	ID
What does the person in this picture have on his face?	1
How many baby elephants are there?	2
What is in the bowl?	3
Is the television on or off?	4
What color is the walk light?	5
Which way is its head turned?	6
How many people are riding on each bike?	7
What animal is in this picture?	8
What color is the road?	9
What color is the boy’s hair?	10

Table 1. Reference for the Figure 5(c) of the main paper.

## 3. Evaluation Methods

**Accuracy:** VQA dataset has 3 type of answers: *yes/no*, *number* and *other*. The evaluation is carried out using two test splits, i.e test-dev and test-standard. The question in corresponding test split are of two types: Open-Ended and Multiple-choice. Our model generates a single word answer on the open ended task. For each question there are 10 candidate answer provided with their respective confidence level. This answer can then be evaluated using accuracy metric defined as follows:

## Algorithm 1 Rank Correlation Procedure

- 1: **procedure** : (Initialization)
- 2:  $P_H$ : Probability distribution of Human Attention Map
- 3:  $P_M$ : Probability distribution of our model
- 4: **Rank**:
- 5:   Compute Rank of  $P_H$  :  $R_H$
- 6:   Compute Rank of  $P_M$  :  $R_M$
- 7: **Rank Difference** :
- 8:   Compute difference in rank between  $R_H$  &  $R_M$  :  $Rank_{Diff}$
- 9:   Compute square of rank difference  $Rank_{Diff} : S_{Rank-Diff}$
- 10: **Rank Correlation**:
- 11:   Compute Dimension of  $P_M$  :  $N$
- 12:   Compute Rank Correlation using :

$$R_{Cor} = 1 - \frac{6 * S_{Rank-Diff}}{N^3 - N}$$

- 13: **end procedure**

Question	ID
Is this wheat bread?	1
Is the cat looking at the camera?	2
Is this chair broken?	3
Are these animals monitored?	4
Does the cat recognize someone?	5
Is the figurine life size?	6
Is the smaller dog on a leash?	7
Is this in the mountains?	8
Is the woman sitting on the bench?	9
Is the church empty?	10

Table 2. Reference for the Figure 5(d) in the main paper.

$$Acc = \frac{1}{N} \sum_{i=1}^N \min\left(\frac{\sum_{t \in T^i} \mathbb{I}[a_i = t]}{3}, 1\right) \quad (1)$$

Where  $a_i$  the predicted answer and  $t$  is the annotated answer in the target answer set  $T^i$  of the  $i^{th}$  example and  $\mathbb{I}[\cdot]$  is the indicator function. The predicted answer  $a_i$  is correct if at least 3 annotators agree on the predicted answer. If the predicted answer is not correct then the accuracy score depends on the number of annotator that agree on the answer. Before checking accuracy, we need to convert the predicted answer to lowercase, number to digits and punctuation & article to be removed. **Rank Correlation:** The detailed algorithm is given in Algorithm 1.