Caption Generator For Images Powered By Watson Visual Recognition

**1. Introduction**

**1.1 Overview**

Image Captioning refers to the process of generating textual description from an image – based on the objects and actions in the image. The ability to recognize image features and generate accurate, syntactically reasonable text descriptions is important for many tasks in computer vision.

**1.2 Purpose**

Image Captioning can be used in the following scenarios.

Self-driving cars — Automatic driving is one of the biggest challenges and if we can properly caption the scene around the car, it can give a boost to the self-driving system.  
Aid to the blind — We can create a product for the blind which will guide them travelling on the roads without the support of anyone else. We can do this by first converting the scene into text and then the text to voice. Both are now famous applications of Deep Learning.  
CCTV cameras are everywhere today, but along with viewing the world, if we can also generate relevant captions, then we can raise alarms as soon as there is some malicious activity going on somewhere. This could probably help reduce some crime and/or accidents.

**2. Proposed Solution**

This Caption Generation for Images project aims at building an application which takes input as image analyses it and generate the captions in the form of speech. To achieve this, we will be using IBM Services like node-red service to build a web UI where user uploads a picture. This picture is analyzed by visual recognition service and the analyzed description is then converted in to text to speech service using text to speech service.

**3. Theoretical Analysis**

**3.1 Block Diagram**

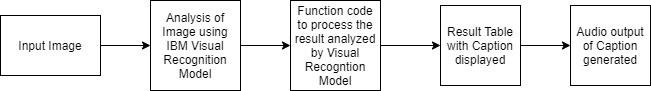


Fig 1: Block diagram for Caption Generation for Images

As shown in Fig. 1, image is passed as input to the visual recognition model to analyze the image.Then the analysis done by IBM tool Visual Recognition model is given as input to the function node which will process the result to be displayed.Then the result table is displayed which has the classification details of the image along with the caption.

**3.2 Hardware/Software Designing**

IBM cloud

IBM Visual Recognition Model

Node red for UI

**4. Experimental Investigations**

* Using UI page created using IBM Node red tool, browse button is given so that the user selects the image to be captioned.
* Then the image is passed as input to the Visual recognition model for the classification.
* The classified data of image is displayed in result table.
* The audio is played for the caption generated for the image.

**5. Flowchart**

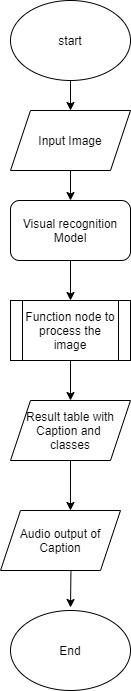
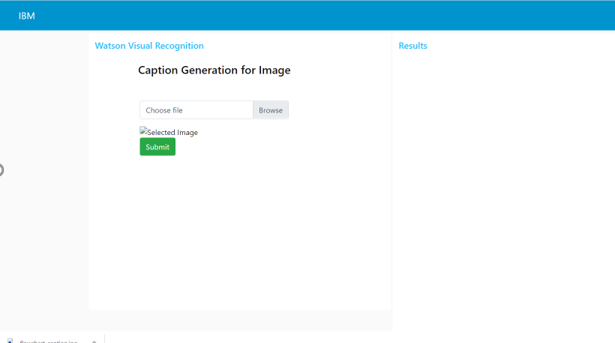


Fig 2: Caption Generation flow chart

**6. Results**

The Fig 3 shows the screenshots of the Covid-19 Essentials Chatbot.



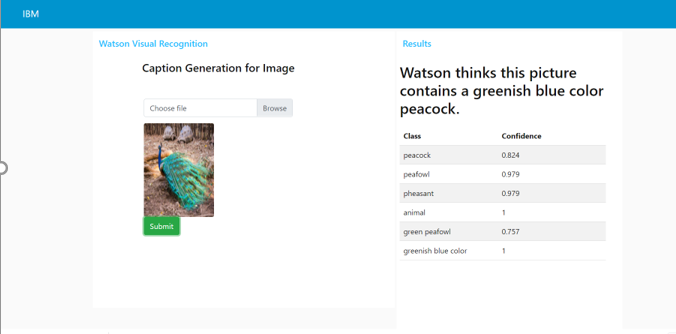


Fig 3: Result Screenshots of Image Captioning

**7. Advantages and Disadvantages**

**7.1 Advantages**

Image Captioning can be used

1. In editing applications

2. Can be used in virtual assistants

3. Can be used by impaired persons

4. Can be used for social media

**7.2** **Disadvantages**

1. It may be difficult to classify the images accurately for the sensitive changes in the image which would not fool a human observer.

2. Datasets are not trained to the Visual recognition model, so might be difficult to caption complicated images having less clarity.

**8. Applications**

1. Use of Image Captioning in Self-driving cars — Automatic driving is one of the biggest challenges and if we can properly caption the scene around the car, it can give a boost to the self-driving system.  
2. Use of Image Captioning as Aid to the blind — We can create a product for the blind which will guide them travelling on the roads without the support of anyone else. We can do this by first converting the scene into text and then the text to voice. Both are now famous applications of Deep Learning.

3. CCTV cameras are everywhere today, but along with viewing the world, if we can also generate relevant captions, then we can raise alarms as soon as there is some malicious activity going on somewhere. This could probably help reduce some crime and/or accidents.

4. Twitter Image Analysis

5. Used in Natural Processing applications

**9. Conclusion**

The Caption Generation for Image is a basic application to generate the caption to the image given as input. The image is classified using Visual Recognition model and the result table with various classes and their confidence is displayed along with the suitable caption for the image.

**10. Future Scope**

In this project, caption generation happens for a given single image at a time. Hence, this can be enhanced to generate the captions for more number of images at the same time.

And Visual Recognition Model can be trained using data set of various images so that Image captioning can be done accurately.