IMAGE CAPTION GENERATOR

**A Project Report**

## Submitted in partial fulfillment of the

Requirements for the award of the Degree of

**BACHELOR OF SCIENCE (INFORMATION TECHNOLOGY) By**

## Rhythm Shah

## UID - 20BIT053

**Under the esteemed guidance of**

**Mr. Wilson Rao and Ms. Bertilla Fernandes**



# DEPARTMENT OF INFORMATION TECHNOLOGY JAI HIND COLLEGE

***(Autonomous)* MUMBAI, 400020 MAHARASHTRA**

# 2022-23

**STATEMENT**

An image caption generator is a system that can automatically generate a natural language description of an image. It is a type of machine learning model that combines techniques from computer vision and natural language processing to generate captions for images.

The development of image caption generation models has been motivated by the growing need for image annotation in a variety of applications, such as image retrieval, visual search, and assistive technologies for the visually impaired. The goal of image caption generation is to generate captions that are both relevant to the image and grammatically correct, as well as semantically similar, contextually relevant, and informative to the human reader.

.

# PROBLEM DEFINITION

The problem of image caption generation can be defined as the task of automatically generating a natural language description of an image. It can be formalized as a mapping between an input image and an output text, where the output text is a sequence of words that describes the contents of the image.

There are a few different ways to approach this problem, but a common method is to use a deep learning-based model that combines techniques from computer vision and natural language processing. The model takes an image as input and generates a caption by:

1. Extracting visual features from the image using a convolutional neural network (CNN)
2. Feeding the visual features into a recurrent neural network (RNN) or a transformer-based model
3. Generating a sequence of words by sampling from the output of the RNN or transformer

The main challenge in this task is to generate captions that are both relevant to the image and grammatically correct. The model needs to be able to understand the contents of the image, as well as the relationships and attributes of the objects present in the image. It also needs to be able to generate natural language descriptions that are coherent and grammatically correct.

Additionally, the generated captions need to be semantically similar, contextually relevant and informative to the human reader, even in the presence of images with complex scenes, cluttered images, and images with low resolution or lighting. And, the model should be able to generalize well to handle unseen images and generate caption for that.

.

# OBJECTIVES

The main objective of an image caption generator is to generate a natural language description of an image. This can include identifying objects, people, scenes, and activities present in the image, as well as describing their relationships and attributes.

Some specific objectives of an image caption generator include:

1. Object Recognition: The model should be able to identify and label the objects present in the image.
2. Scene Understanding: The model should be able to understand the context and layout of the image, and describe the overall scene.
3. Language Generation: The model should be able to generate grammatically correct and coherent sentences that accurately describe the image.
4. Natural Language Understanding: The model should be able to understand the nuances of natural language, such as idiomatic expressions and figurative language.
5. Human-like Description: The model should be able to generate descriptions that are similar to those that a human would give, in terms of both content and style.
6. Multimodal Understanding : Model should be able to understand the multiple modalities present in the image and generate a caption accordingly.
7. Multilinguism : For certain applications, the model should be able to generate captions in multiple languages.
8. Robustness : The model should be able to generate captions even in the presence of complex scenes, cluttered images, and images with low resolution or lighting.
9. Domain-specific application: The model should be able to generate captions that are specific to a particular domain, such as wildlife, sports, or fashion.
10. Handling Novelty : Model should be able to generalize well to handle unseen images and generate caption for that.

# SCOPE

The scope of image caption generators is quite broad, as they have the potential to be used in a variety of applications. Some examples of the ways in which image caption generators can be used include:

1. Image Retrieval: Image caption generators can be used to generate captions for images, which can then be used to improve the accuracy and efficiency of image retrieval systems.
2. Visual Search: Image caption generators can be used to generate captions for images, which can then be used to improve the accuracy and effectiveness of visual search systems, allowing users to search for images based on their textual descriptions.
3. Assistive Technologies: Image caption generators can be used to generate captions for images, which can then be used to provide information to visually impaired individuals about the contents of an image through text-to-speech or braille displays.
4. Image Annotation: Image caption generators can be used to automatically generate captions for images, which can then be used to improve the efficiency and accuracy of image annotation systems.
5. Social Media: Image caption generators can be used to generate captions for images shared on social media, which can make it easier for users to understand the contents of an image.
6. Advertising: Image caption generators can be used to generate captions for images used in advertising, which can make the ads more attractive and informative.
7. Robotics: Image caption generators can be used to provide a robot with a natural language understanding of the environment, allowing the robot to navigate and interact with its surroundings.
8. Autonomous Cars : Image caption generators can be used to generate captions for images captured by cars, which can be used to improve the decision-making process of autonomous cars.
9. Surveillance : Image caption generators can be used to generate captions for images captured by surveillance cameras, which can be used to improve the surveillance systems.

Top of Form

Bottom of Form

# METHODOLOGY

This approach uses a deep learning-based model that combines techniques from computer vision and natural language processing. This approach typically consists of the following steps:

1. Collect a dataset of images and their corresponding captions. This dataset is used to train the model.
2. Pre-process the data by resizing images, splitting the data into training, validation and test sets, and tokenizing the captions.
3. Extract visual features from the images using a convolutional neural network (CNN). The CNN is usually pre-trained on a large dataset of images, such as ImageNet, and then fine-tuned on the image captioning dataset.
4. Feed the visual features into a recurrent neural network (RNN) or a transformer-based model. The RNN or transformer-based model is trained to generate a caption by sampling from the output of the CNN.
5. Train the entire model using backpropagation. The training process is typically done using a cross-entropy loss, which measures the difference between the generated caption and the ground-truth caption.
6. Fine-tune the model by adjusting the hyperparameters and architectures.
7. Generate captions for new images by forwarding the images through the CNN, and then feeding the visual features into the RNN or transformer-based model and generate a caption by sampling from the output.

# Software and Hardware Requirements

### Software Requirements:

Operating System: Windows

Technology: Python

IDE: Visual Studio Code

### Hardware Requirements:

Laptop/PC

RAM: 4GB and above

System type: 64-bit operating system, x64-based processor

**TESTING APPROACH**

Following are the testings approaches which are performed on the system:

1. **UNIT TESTING:** Unit Testing is a software testing method, in which each module in the system is tested individually, independent of other modules.
2. **SYSTEM TESTING:** System testing is a software testing in which the system as a whole, is tested completely together.
3. **INTEGRATION TESTING:** Integration testing is a software testing in which two or more modules are combined together and then tested.

# LIMITATIONS

### Following are some of the limitations of the system::

1. Lack of diversity
2. Limited understanding of context
3. Difficulty in handling visual complexities
4. Computationally expensive

# CONCLUSION

# BIBLIOGRAPHY