IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMAGE CAPTION GENERATOR

Using Convolutional Neural Networks And Long Short Term Memory

[1]Mr. Jadhav Chaitanya Chandrakant, [2]Mr. Pandey Shiva Jitendra, [3]Mr. Khade Nitin Narayan, [4]Ms. Harshada Sonkamble

[1]Student, [2]Student, [3]Student, [4]Assistant Professor

[1]Computer Engineering

[1]Vishwatmak Om Gurudev College Of Engineering, Mohili - Aghai, India

Abstract: When you see an image, your brain can easily tell what the image is about, but can a computer tell what the image is representing? Computer vision researchers worked on this a lot and they considered it impossible until now! With the advancement in Deep learning techniques and availability of huge datasets and computer power, we can build models that can generate captions for an image. This is what we have implemented in this Python based project where we have used the deep learning techniques of CNN (Convolutional Neural Networks) and LSTM (Long short term memory) which is a type of RNN (Recurrent Neural Network) together so that using computer vision computer can recognize the context of an image and display it in natural language like english.

Index Terms **- Computer vision, Deep learning, Convolutional Neural Networks, Long short term memory, Recurrent Neural Network.**

I. INTRODUCTION

Image caption generator is a task that involves computer vision and natural language processing concepts to recognize the context of an image and describe them in a natural language like English. In this Python based project, we will have implemented the caption generator using CNN (Convolutional Neural Networks) and LSTM (Long short term memory). The image features will be extracted from Xception which is a CNN model trained on the imagenet dataset and then we feed the features into the LSTM model which will be responsible for generating the image captions.

Convolutional neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easy. It can handle the images that have been translated, rotated, scaled and changes in perspective.

LSTM stands for Long short term memory, they are a type of RNN (recurrent neural network) which is well suited for sequence prediction problems. Based on the previous text, we can predict what the next word will be. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

II. LITERATURE SURVEY

1. SkinVision: SkinVision is a mobile application available for Android and IOS which lets you confirm whether a skin condition can be skin cancer or not.
2. Google Photos: Google Photos is an android application which makes use of an image caption generator to classify photos into Mountains, Sea, etc.
3. Picasa: Picasa is a cross platform image organizer which organizes images and identifies you and your friends in a group picture.
4. Tesla/Google Self Drive Cars: All Self driving cars are using image caption generators so that computers can drive cars safely and efficiently.
5. Adobe Photoshop: It is an image editing application which uses image captioning for providing editing recommendations.
6. Facebook: Facebook is an online social media and social networking service which automatically generates captions of images you have uploaded.
7. ShutterStock: ShutterStock is a stock photography application where you can sell images and it automatically generates tags for images using an image caption generator.

IV. PROPOSED SYSTEM

This project is proposed to be used in robotics, using image caption generator robotics can determine the surrounding conditions and show intelligence like humans, using image caption generator robots can make better decisions by understanding the environmental conditions.

This project is also proposed to be used in self driving cars. by installing the image caption generator made by us in the car accidents can be eliminated. Using an image caption generator computer will detect the approaching object and automatically apply brakes to stop the car. Self driving cars can also detect the traffic signal using an image caption generator, when the traffic signal is red, the computer will detect it and stop the car and follow the traffic laws.

To eliminate car accidents this project is very useful and plays a very vital role, image caption generators can also be installed in regular cars to provide automatic braking feature. Sometimes during night drive the illumination generated by approaching vehicle headlight directly incident on the eyes of the driver and drivers vision is lost for sometime which can result in accident. This can be eliminated by installing an image caption generator in the car and using an led array in car headlights. With the help of an image caption generator computer will be aware of the approaching objects and will not spread light in regions occupied by an approaching object.

V. PROJECT ARCHITECTURE

Convolutional Neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix and CNN is very useful in working with images. CNN is basically used for image classifications and identifying if an image is a bird, a plane or Superman, etc. It scans images from left to right and top to bottom to pull out important features from the image and combines the feature to classify images. It can handle the images that have been translated, rotated, scaled and changes in perspective.

LSTM stands for Long short term memory, they are a type of RNN (recurrent neural network) which is well suited for sequence prediction problems. Based on the previous text, we can predict what the next word will be. It has proven itself effective from the traditional RNN by overcoming the limitations of RNN which had short term memory. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

So, to make the image caption generator model, we have merged these architectures. It is also called a CNN-RNN model.

- CNN is used for extracting features from the image. We will use the pre-trained model Xception.
- LSTM will use the information from CNN to help generate a description of the images.

VI. ADVANTAGES

1. Assistance for visually impared
    a. The advent of machine learning solutions like image captioning is a boon for visually impaired people who are unable to comprehend visuals.
    b. With AI-powered image caption generators, image descriptions can be read out to the visually impaired, enabling them to get a better sense of their surroundings.
2. Recommendations in editing
    a. The image captioning model automates and accelerates the closed captioning process for digital content production, editing, delivery, and archival.
    b. Well-trained models replace manual efforts for generating quality captions for images as well as videos.
3. Media and Publishing Houses
    a. The media and public relations industry circulate tens of thousands of visual data across borders in the form of newsletters, emails, etc.
    b. The image captioning model accelerates subtitle creation and enables executives to focus on more important tasks.
4. Self driving cars
    a. By using the captions generated by image caption generator self driving cars become aware of the surroundings and make decisions to control the car.
5. Reduce vehicle accidents
    a. By installing an image caption generator in the vehicles, vehicles can stop by applying the automatic brake when an object in the surrounding is detected.

VII. FUTURE SCOPE

1. Sheer complexity
    a. Everything has the potential to be scaled up in terms of power and complexity. With technological advancements, we can make CPUs and GPUs cheaper and/or faster, enabling the production of bigger, more efficient algorithms. We can also design neural nets capable of processing more data, or processing data faster, so it may learn to recognize patterns with just 1,000 examples, instead of 10,000. Unfortunately, there may be an upper limit to how advanced we can get in these areas—but we haven't reached that limit yet, so we'll likely strive for it in the near future.
2. New applications
    a. Rather than advancing vertically, in terms of faster processing power and more sheer complexity, neural nets could (and likely will) also expand horizontally, being applied to more diverse applications. Hundreds of industries could feasibly use neural nets to operate more efficiently, target new audiences, develop new products, or improve consumer safety—yet it's criminally underutilized. Wider acceptance, wider availability, and more creativity from engineers and marketers have the potential to apply neural nets to more applications.

3. Integration
    a. The weaknesses of neural nets could easily be compensated if we could integrate them with a complementary technology, like symbolic functions. The Hard part would be finding a way to have these systems work together to produce a common result—and engineers are already working on it.

VIII. CONCLUSION

Image caption generator is a task that involves computer vision and natural language processing concepts to recognize the context of an image and describe them in a natural language like English. So, to make an image caption generator model, we have merged these architectures. It is also called a CNN-RNN model.

- CNN is used for extracting features from the image. We will use the pre-trained model Xception.
- LSTM will use the information from CNN to help generate a description of the image.

Convolutional Neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix and CNN is very useful in working with images.CNN is basically used for image classifications and identifying if an image is a bird, a plane or Superman, etc.It scans images from left to right and top to bottom to pull out important features from the image and combines the feature to classify images. It can handle the images that have been translated, rotated, scaled and changes in perspective.

LSTM stands for Long short term memory, they are a type of RNN (recurrent neural network) which is well suited for sequence prediction problems. Based on the previous text, we can predict what the next word will be. It has proven itself effective from the traditional RNN by overcoming the limitations of RNN which had short term memory. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

IX. ACKNOWLEDGEMENT

The success and final outcome of the project requires a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the working of our project. All that we have done is only due to such supervision and assistance and we would not forget to thank them.We respect and thank Dr. Rajesh Jaware Sir for providing us an opportunity to do the project and giving us all support and guidance.We owe a deep gratitude to our project guide Prof. Harshada Sonkamble madam , who took keen interest in our project work and guided us all along by providing all the necessary information for developing a good system. We would not forget to remember our Head of Department Prof. Anup Maurya Sir for their encouragement and more over for their timely support and guidance.We are thankful to and fortunate enough to get constant encouragement, support and guidance from all teaching staffs of our department which helped us in our project work.

X. REFERENCES

1. https://docs.python.org/3.8/
2. https://www.jetbrains.com/pycharm/documentation/
3. https://www.tensorflow.org/api_docs/python/tf/all_symbols
4. https://keras.io/guides/
5. https://numpy.org/doc/
6. https://pillow.readthedocs.io/en/stable/
7. https://tqdm.github.io/
8. https://jupyterlab.readthedocs.io/en/stable/