Revolutionizing Image Analysis with Automated Captioning

Combines the power of computer vision and natural language processing

PRESENTED BY:
AVINAV

Agenda

01 Define Project goals

02 Collect And Prepare Data

03 Choosing Deep Learning Framework

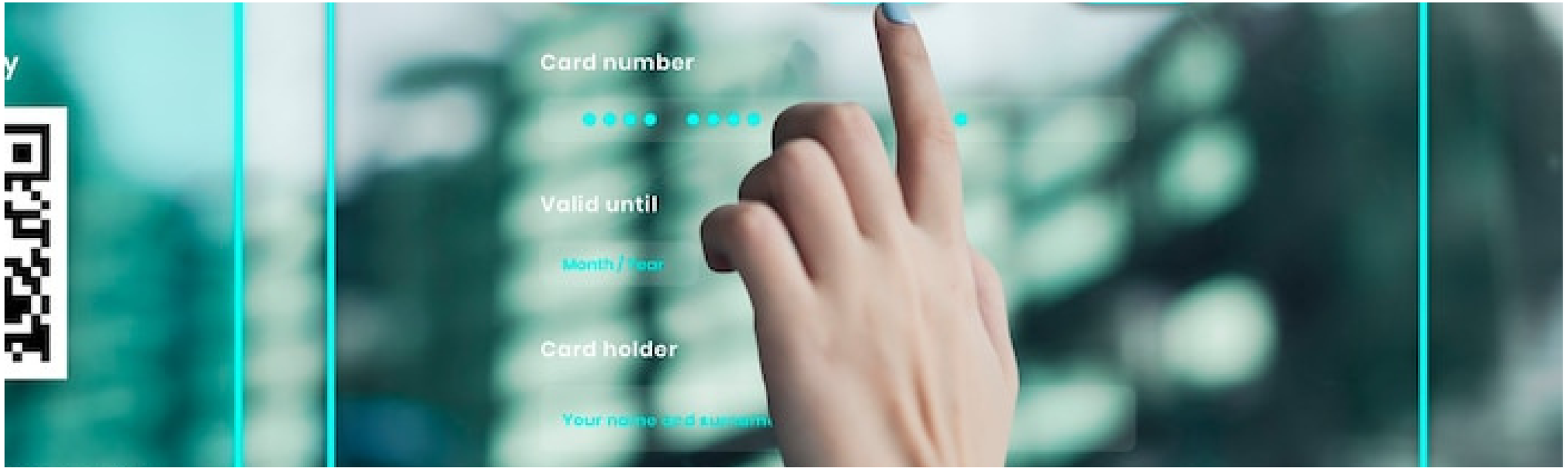
04 Implementing LSTM and Other Models

05 Train,Test and Validate Images

06 Generate Captions

ABOUT THE PROJECT

Automated Captioning is revolutionizing image analysis. With the help of image caption generator technology, images can now be automatically annotated with descriptive captions, transforming them into data that can be analyzed and searched. This presentation will explore the technology behind automated captioning and how it is transforming the field of image analysis.



AUTOMATED CAPTIONING

Automated Captioning technology has several benefits, including improving efficiency by automating the process of image annotation, improving accuracy by reducing human error, and enabling new insights by transforming images into data that can be analyzed and searched.

Automated Captioning technology is being used in a variety of applications, from automated image tagging in e-commerce to medical image analysis in healthcare. It is also being used to improve image search capabilities on social media platforms and search engines.

SCOPE

The ultimate purpose of Automated caption generator is to make users experience better by generating automated captions.It is used in various areas like:

- Image Indexing
- Social Media
- Website Ranking
- Visually Impaired Person
- Object Identification
- SEO Purpose

BACK TO AGENDA

TECHNOLOGIES USED

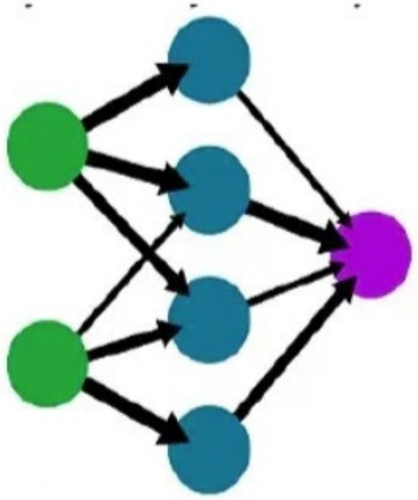
- Python
- Deep Learning
- Jupyter Notebooks
- Keras Library
- Numpy
- Natural Language Processing
- Tensorflow
- Pillow
- tqdm

AIM & OBJECTIVE

- To extract features from images using Convolutional Neural Network techniques.
- To train the model on dataset so that it can detect multiple assets in an image like objects,things.
- To train model for generating captions for images at character ,word or sentence level.
- To convert the generated textual captions into an audio format.

EXAM

1 .A Boy is Playing Cricket.
2. A Boy Holding the Cricket Bat.



?

CNN and RNN

Encoder

The Convolutional Neural Network(CNN) can be thought of as an encoder. The input image is given to CNN to extract the features. The last hidden state of the CNN is connected to the Decoder.

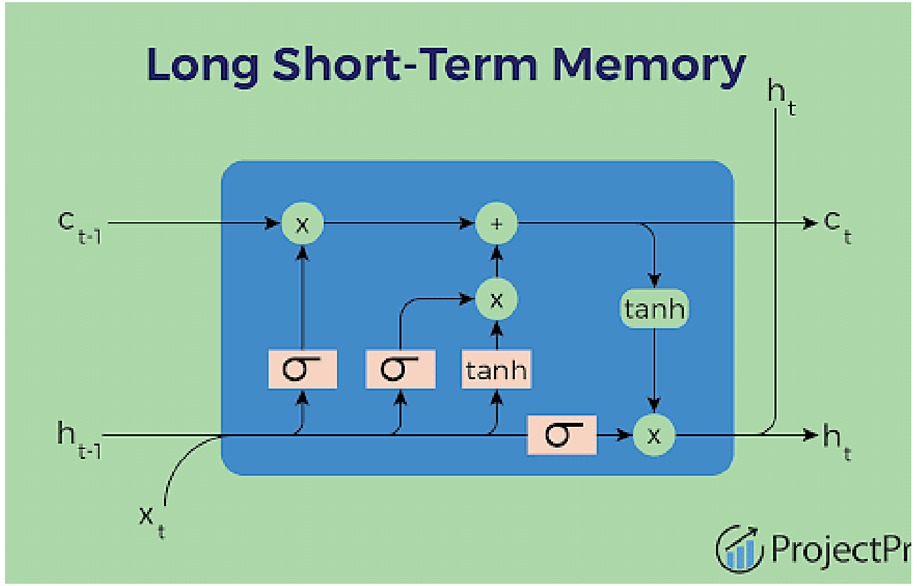
Decoder

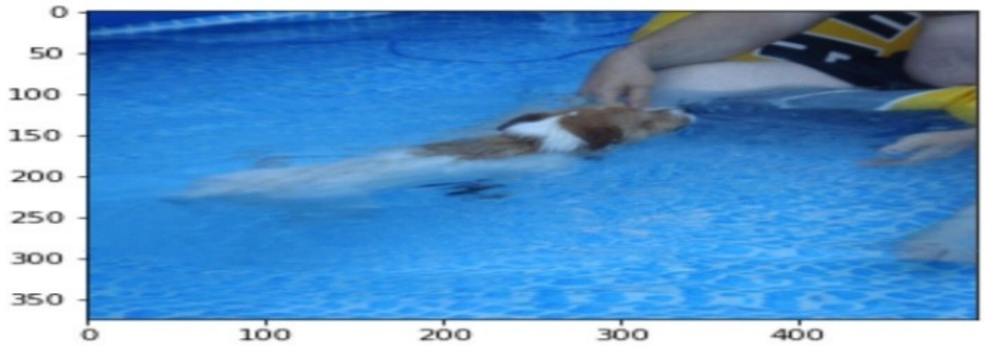
The Decoder is a Recurrent Neural Network(RNN) which does language modelling up to the word level. The first time step receives the encoded output from the encoder and also the <START> vector.

Flickr 8K. A collection of 8 thousand described images taken from flickr.com

LONG SHORT TERM MEMORY

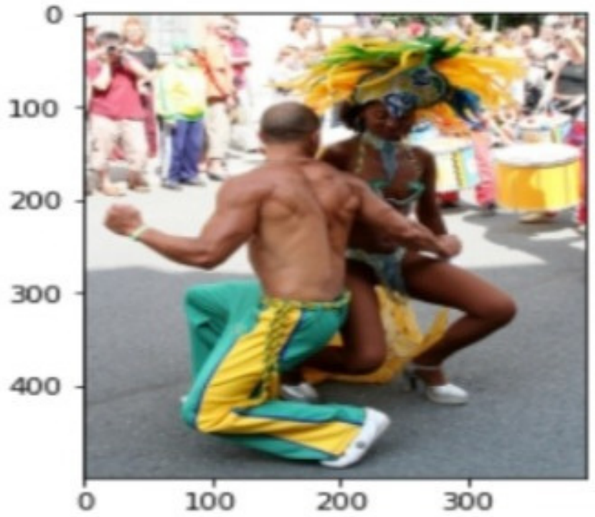
LSTM stands for long short-term memory networks, used in the field of Deep Learning. It is a variety of recurrent neural networks (RNNs) that are capable of learning long-term dependencies, especially in sequence prediction problems. LSTM has feedback connections, i.e., it is capable of processing the entire sequence of data, apart from single data points such as images. This finds application in speech recognition, machine translation, etc. LSTM is a special kind of RNN, which shows outstanding performance on a large variety of problems.





```
Referance Captions:
a brown and white dog swim towards some in a pool
A dog in a swim pool swim toward sombody we cannot see .
A dog swim in a pool near a person .
Small dog be paddle through the water in a pool .
A small brown and white dog be in a pool .
Predicted Caption:
A boy be jump into a pool .
bleu score:  0.32347562464306545
```

SCREENSHOTS



```
Referance Captions:
A man and a woman in festive costume dance .
A man and a woman with feather on her head dance .
A man and a woman wear decorative costume and dance in a crow
one performer wear a feathered headdress dance with another p
Two person be dance with drum on the right and a crowd behind
Predicted Caption:
A man in a red shirt be sit on a street .
bleu score:  0.6076795808137692
```

```
Referance Captions:
A couple of person sit outdoors at a table with an umbrella and talk .
Three person be sit at an outside picnic bench with an umbrella .
Three person sit at an outdoor cafe .
Three person sit at an outdoor table in front of a building paint like the Union Jack .
Three person sit at a picnic table outside of a building paint like a union jack .
Predicted Caption:
A man in a white shirt be jump in a park .
```

CONCLUSION

Automated Captioning technology is revolutionizing image analysis, enabling new insights and improving efficiency and accuracy. As this technology continues to develop, it will undoubtedly have a significant impact on a wide range of fields, from e-commerce to healthcare.

REFERENCES

- https://machinelearningmastery.com/develop-a-deep-lEarning-caption-generation-model-in-python/
- https://www.youtube.com/watch?v=yk6XDFm3J2c
- https://wepik.com
- https://www.tutorialspoint.com/google_colab/index.html


TH
AN
K
YO