IMAGE CAPTION AND SPEECH GENERATION USING LSTM AND GTTS API

S.Pooja Sai Sree - 17wh1a0549, D.Manisha - 17wh1a0524, S.Susmitha - 17wh1a0529, Guide - Mr. K.Bhargav Ram

Department of Computer Science and Engineering

BVRIT HYDERABAD College of Engineering for Women

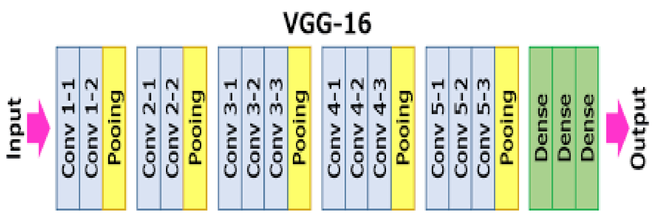
ABSTRACT

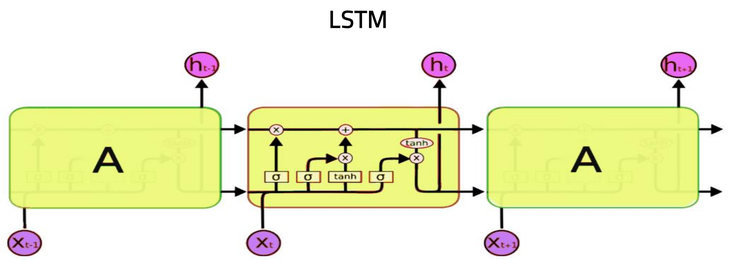
The main aim of the project is to develop Image caption and speech generator, which generates captions or descriptions for an image according to the content observed and converts description into speech.

INTRODUCTION

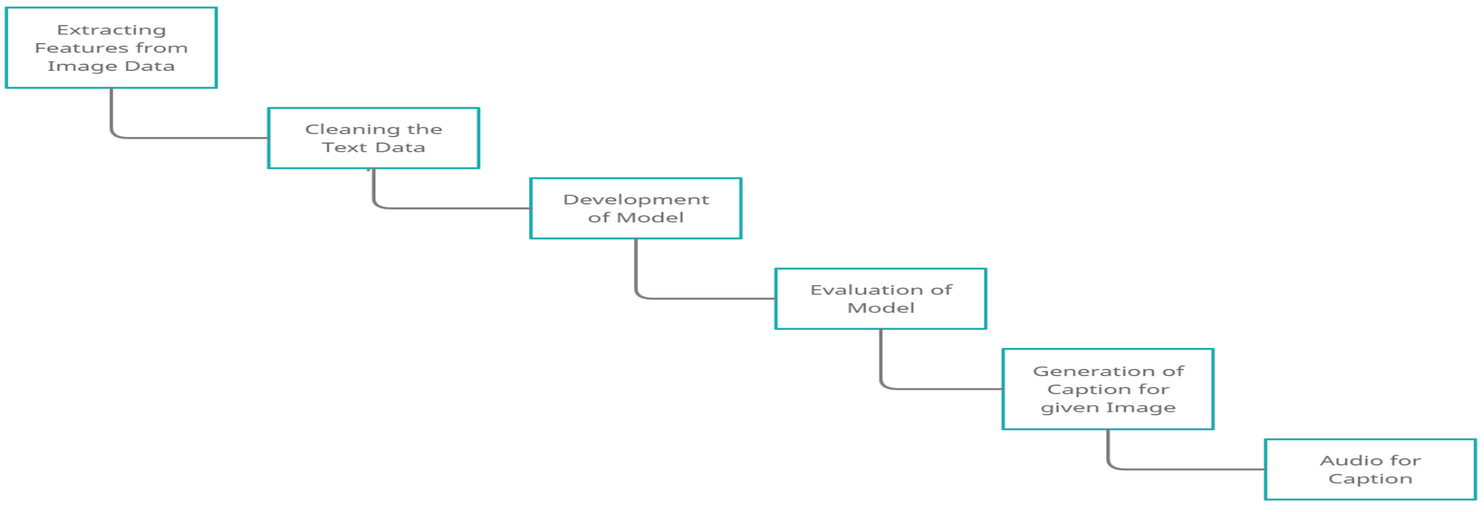
Caption generation is one of the challenges in rapid developments of machine learning. It includes generating descriptions from the content observed in the image. The traditional machine learning algorithms were not successful in doing this task. But the deep learning models showed greater impact compared to machine learning models. This is due to their property of capturing the connection present on the relevant image and their ability to generalize is much better than traditional methods. The main objective is to create image caption and speech generator which has its applications in assistance for visually impaired, media and publishing houses and social media posts.

METHODOLOGY





ARCHITECTURE



EVALUATION OF MODEL

```
reference_translation, candidate_translation = list(), list()
for img_id, desc_list in test_descriptions.items():
  y = description_generator(model, tokenizer, test_features[img_id], max_length)
  references = [d.split() for d in desc_list]
  reference_translation.append(references)
  candidate_translation.append(y.split())
print('BLEU-Score: %f' % (corpus_bleu(reference_translation, candidate_translation, weights=(1.0, 0, 0, 0))*100))
```

```
BLEU-Score: 52.759237
```

RESULTS

The quality of predicted descriptions by the image caption generator is measured using BLEU(Bilingual Evaluation Understudy) evaluation metric.
BLEU evaluates the closeness of Machine translations to Human Reference translations.

Achieved BLUE score : 52.7

CONCLUSION

A deep learning approach for the captioning of images and the GTTS API for the conversion of generated description into speech is implemented. The sequential API of Keras was used with Tensorflow as a backend to implement the deep learning architecture to achieve an effective BLEU score of 0.52 for model. The Bilingual Evaluation Understudy Score, or BLEU for short, is a metric for evaluating a generated sentence to a reference sentence.

REFERENCES

[1] B.Krishnakumar , K.Kousalya , S.Gokul , R.Karthikeyan and D.Kaviyarasu, “IMAGE CAPTION GENERATOR USING DEEP LEARNING”, International Journal of Advanced Science and Technology, Vol. 29, No. 3s, (2020), pp. 975-980.

[2] Marc Tanti, Albert Gatt, Kenneth P. Camilleri, “What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption Generator?”, arXiv:1708.02043v2 [cs.CL], 25 Aug 2017.

[3] Srikanth Tammina, “Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images”, International Journal of Scientific and Research Publications, Volume 9, Issue 10, October 2019