ISSN 2395-1621

# Tag Based Caption and Information Generation

[#1]Mr. Pratik Bhakkad, [#2]Ms. Priyanka Bhavsar, [#3]Ms. Kshitija Chinchkar,
[#4]Prof. M.A.R.Shabad

[1]pratikrbhakkad@gmail.com
[2]priyanka.bhavsar85@gmail.com
[3]kschinchkar@gmail.com
[4]muzaffar.shabad@gmail.com

[#1234]Department of Computer Engineering,

Savitribai Phule Pune University
Sinhgad Academy of Engineering,Pune,India.

**ABSTRACT**

**This paper is concerned with the task of automatically generating captions for images based on keywords, which is important for many image related applications. Examples include video and image retrieval as well as the development of tools that aid visually impaired individuals to access pictorial information. Our approach leverages the vast resource of pictures and information available on the web and the fact that many of them are captioned and collocated with thematically related documents. Our model learns to create captions from the keywords that are input by the user and based on this caption our system searches various websites using Google Analytics API for the related information. This information is analysed using sentiment analysis and a synopsis of relevant information is generated using K-mean clustering algorithm.**

ARTICLE INFO

**Article History**

Received: 30th November 2016

Received in revised form :

30th November 2016

Accepted: 2nd December 2016

**Published online :**

**3rd December 2016**

**Keywords: Caption generation, image annotation, summarization, Topic Models**

## I. INTRODUCTION

Image caption generation is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. Automatically describing the content of an image using properly formed English sentences is a very challenging task.

This task is even harder than the well-studied image classification or objects recognition tasks, for which people challenge and then achieve breakthrough in Large Scale Visual Recognition Challenge (ILSVRC) [1]. Indeed, to generate exact sentences, the model should not only detect the objects of interest contained in the image, but also analyze the relationship between these objects. In addition, the weakness of computing capacity restrict the success of complex models. Fortunately, thanks to the rapid development of computer vision and natural language processing technologies, with the application of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), recent works have made momentous progress, and present a unified method which dominates in image caption generation. [4]

The models based on deep convolutional networks and recurrent neural networks have dominated in recent image caption generation tasks. Performance and complexity are still eternal topic. Inspired by recent work, by combining the advantages of simple RNN and LSTM, we present a novel parallel-fusion RNN-LSTM architecture, which obtains better results than a dominated one and improves the efficiency as well. The proposed approach divides the hidden units of RNN into several same-size parts, and lets them work in parallel. Then, we merge their outputs with corresponding ratios to generate final results. Moreover, these units can be different types of RNNs, for instance, a simple RNN and a LSTM. By training normally using NeuralTalk1 platform on Flickr8k dataset, without additional training data, we get better results than that of dominated structure and particularly, the proposed model surpass Google NIC in image caption generation[2]

## II. LITERATURE SURVEY

### Existing System

The first approach to use neural networks for caption generation was Kiros et al., who proposed a multimodal Log-Bilinear model. But most of other recent works are different from it, which replace a feed-forward neural language model

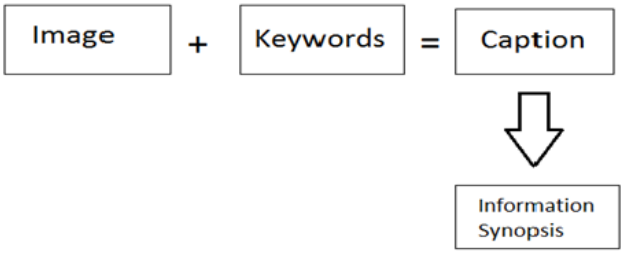
with a recurrent one. These works have some common key structures, and accordingly, we call those structures simply as dominated model or general model. In details, two of major parts, i.e., the CNN and RNN, play core roles in general model respectively.

Especially, for Flickr8k dataset, Mao et al. present a multimodal Recurrent Neural Network (m-RNN) model which contains a VGG-net CNN and a simple RNN. Karpathy&Li gain the similar BLEU scores as Mao on Flickr8k in this task. Besides, Vinyals et aluse LSTM instead of other RNNs in their model and unlike wisely show the image to RNNs at the beginning, leading to performance improvement finally. The temporary winner (Xu et alarchives the state-of-art performance. In this work, it presents an attention-based model and uses the CN-N feature extracted from fourth layers, which increases the computational cost. However, as the amount of training data increases, the model with less training time will be needed.[1]

## III. PROPOSED SYSTEM

The ultimate goal of our model is to improve performance in condition of promoting efficiency. Several recent works have shown that the dominated approach is universal and effective for image interpretation, which has powerful capacity in aligning visual and language data.

We propose a parallel-fusion RNN-LSTM architecture that contains two major structures without additional parts compared to the general model. The part of image representation is based on CNN while the part of caption generation is based on RNN structures. We apply them to extract image features and align visual and language data respectively.[4]




## IV. SYSTEM ARCHITECTURE

The user will input the image and the tags related to the image. Based on that processing will be done automatically and caption and information synopsis will be generated.

For processing the algorithms used are:

1. Google analytics API
2. Text Annotation algorithm
3. K-means clustering algorithm
4. Sentiment Analysis algorithm

Google analytics API is used to fetch the data from various websites related to the tags input by the user. Using different queries this algorithm works. This algorithm is used only for searching purpose.

K-means clustering algorithm is used to organize the data into different clusters based on similar attributes. The algorithm segregates the huge data into small clusters which can be efficiently analysed.

Sentiment analysis is used to understand the data. It ranks the sentences with positive and negative points based on the tone of the data. It also helps to generate the synopsis from the bulky data.

The system will generate captions based on the given image; many of the search engines deployed on the web retrieve images without analyzing their content, simply by matching user queries against textual information. We propose framework consisting of content selection. Content selection makes use of dictionaries that specify a mapping between words and image regions or features uses human written templates or grammars for producing textual output. There are two stages in the proposed work they are Initial stage in which the images are obtained and the feature are extracted from the images and they are compared with the other images in the database. The second stage is the annotation of the tag to the images through which the image are obtained and the tag provided along with the similar images are returned to the query image. The feature extraction and the comparison of the visual feature are performed and the images are obtained according to the visual structures. Many of the images are provided with the same features and they are tagged. If the visual features of the images are equal then the tag of that image are retrieved and the result image are displayed as image with the tag. The ultimate goal of our model is to improve performance in condition of promoting efficiency. Several recent works have shown that the dominated approach is universal and effective for image interpretation.[3]

## V. ADVANTAGES

- Real time System
- Automated
- Efficient and more accurate
- Simple design

## VI. APPLICATIONS

- Generating captions and content to the preferred images.
- To provide affordable and reliable articles.
- The captions are used to enhance the information of the image.
- By this method we can also retrieve the details of the barcode by giving the barcode image

## VII. FUTURE SCOPE

The query image given is first annotated and then prioritized and then the image is matched with the article and finally generate caption for the given image. The future work can be done by using the bar code .Using barcode the image can be searched according to the code which is used to know the name of the image and the type of the image.

## VIII. CONCLUSION

We have presented extractive and abstractive caption generation models. A key aspect of our approach is to allow both the visual and textual modalities to influence the generation task. This is achieved through an image annotation model that characterizes pictures in terms of description keywords that are subsequently used to guide the caption generation process. Simply extracting a sentence from the document often yields an inferior caption. Our experiments also show that a probabilistic abstractive model defined over phrases yields promising results. It generates captions that are more grammatical than a closely related word-based system and manages to capture the gist of the image (and document) as well as the captions written by journalists. This paper contains the system of annotating images and generating captions to the preferred images. The motivation of creating this system is to provide affordable and reliable articles. The captions are used to enhance the information of the image.

Sentiment analysis is used to understand the data. It ranks the sentences with positive and negative points based on the tone of the data. It also helps to generate the synopsis from the bulky data.

## IX. ACKNOWLEDGEMENT

It is an incidence of great pleasure in submitting this project report. Making this project reality takes many dedicated people and it is great pleasure to acknowledge the contribution of entire computer department. We take this opportunity to express profound gratitude and ineptness for the personal involvement and constructive criticism provided beyond technical guidance during project to our Guide Prof. M.A.R. Shabad and project coordinator Prof. Kiran Avhad. of Computer department. We shall ever be grateful to them for encouragement and suggestions given by them from time to time. We should like to thank H.O.D Prof. B.B.Gite of Computer Department for providing the necessary facilities during the period of working of this project. We should like to thank Principal Prof. K.P. Patil for providing the necessary facilities during the period of working of this project.

## REFERENCES

[1]Unhua Mao, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang and Alan Yuille, "Deep captioning with multimodal recurrent neural networks (m-rnn)", ICLR, 2015.

[2]Yansong Feng, Member, IEEE, and Mirella Lapata, Member, IEEE, “Automatic Caption Generation for News Images”, VOL. 35, NO. 4, APRIL 2015

[3]Andrej Karpathy and Li Fei-Fei, "Deep visual-semantic alignments for generating image descriptions", pp. 3128-3137.

[4] Y. Gu, X. Qian, Q. Li. “Image Annotation by Latent Community Detection and Multikernel Learning”. IEEE Transactions on Image Processing 24(11): 3450-3463 (2015)