A Deep Learning for the Generation ofTextual Story Corresponding to a Sequence of Images

B Venkat Raman, Nagaratna P Hegde, NenavathVenkateshNaik, Allu Siva Kishore Reddy

Abstract ***Generating a short story for a sequence of images is much more interesting than generating a single line textual description for an image. Story generation involves relating the meaning of the previous image and the current image and continuing this through out the sequence of images. This can be helpful for better understanding of the situation. In this paper we present our idea of generating story using a CNN model which is pre trained on MSCOCO dataset that can detect objects and concepts of language modelling and NLP text pre-processing techniques . We used a custom stories dataset in which we manually labelled every sentence in every story. Number of sentences in the generated story is equal to the number of images. The results are quite accurate in many cases for a small custom stories dataset and the performance is expected to increase with a bigger dataset.***

KeyWords : Image Classification, Convolution Neural Network, Long Short Term Memory, Language Modelling, Text Pre-Processing.

I. INTRODUCTION

Short story generation for sequence of images is the problem in a image caption generator[1,2] because a simple caption generator cannot take sequence of images as input i.e generation of caption for single image only. But our neural story generation is capable of generating a story by understanding the content of each image and by relating those sequence image contents. In these few years many approaches came to generate the captions[1,2] and presented a successful and impressive outcomes but they did not shown an impressive outcome in story generation because most of these techniques are based on RNN(recurrent neural networks) [3].

So we used an advanced network of RNN [3] i.eLSTM(Long Short Term Memory) [4] which is capable of having long term dependency between the words of any sentences. We used CNN(Convolutional Neural Network) [5] to classify the objects which are present in the image. So in our project we give sequence of images as input to the

CNN [5] and it will process those images and classifies them and we assume two labels for each one-hot vector combination of objects in the image and those labels [2] goes through the LSTM network and generate a textual description as story which is our output.

CNN: Person and Horse
LSTM: a person riding a horse
Story: a person riding a horse and happily sitting on it may be he is going to his home

II. LITERATURE REVIEW

Here we used some already existed concepts for the purpose of Story Generation which are mentioned below.

2.1 Convolutional Neural Network

Rapid advances in computer vision are enabling brand new applications which are producing best results in our daily life by the help Convolutional neural network [5] which is involved in some applications like Image Classification [6],Object Detection, Neural Style Transfer, Face Recognition [7] etc. had used the concept of CNN.

In our Paper, we have used Convolutional Neural Network for the purpose of classifying the objects and tells what type of objects are present in a particular image. In recent years many approaches had came for image classification, all those approaches does the same thing i.e., classifying the object by using CNN.

For objects classification we used yolov3 [8] which is having an incremental developmet in classifying the objects [6] and it will give much accurate results when we compared to it's previous version i.e yolo,yolov2 and the dataset we used is MSCOCO dataset [9] which is a trained on 80 objects.

Revised Version Manuscript Received on 10 September, 2019.

B Venkat Raman, Research Scholar at Osmania University, Hyderabad and Assistant Professor, CSE, RGUKT, Basar, Telangana, India.
(Email: venkat521@yahoo.co.in)
Nagaratna P Hegde, Professor at Vasavi College Of Engineering, Hyderabad, Telangana, India.
(Email: nagaratnaph@gmail.com)
NenavathVenkateshNaik, UG Scholar,Department Of Computer Science &Engineering,RGUKTBasar, Telangana, India.
(Email: venky.b131286@gmail.com)
Allu Siva Kishore Reddy, UG Scholar,Department Of Computer Science &Engineering,RGUKTBasar, Telangana, India.
(Email: kishorereddy8466@gmail.com)

2.2 Long Short Term Memory

Long Short Term Memory is usually called as "LSTM" [4] is special kind and advanced network of RNN.Recurrent Neural Network(RNN) [3] do not support the long-term dependencies that means RNN doesn't have capabilities of remembering text or words for a long time.But our 'Story Generation' needs a powerful architecture like LSTM which supports long-term dependencies unlike RNN.Vanishing Gradient and Exploding Gradients limitations also making RNN as a weak architecture so we are moving with LSTM which can overcome the problems of RNN.

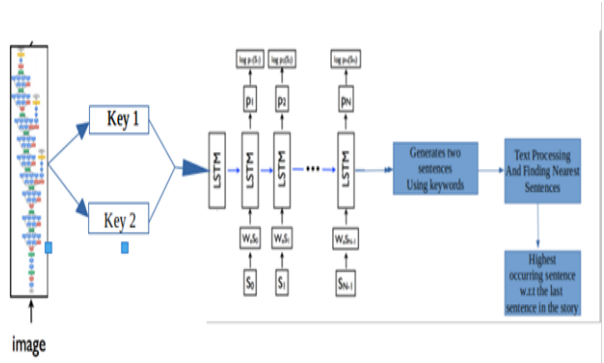
An LSTM [4] just like RNN [3] which is also used for sequential data and it will be used to predict [10] the next word of sentence when we pass one word as input which is a previous word of the predicted word.LSTM also used in Language Models[10,11] also for the purpose of text generation.LSTM and RNN are used for characters level predictions also.

We are using LSTM in a Language model [10,11] in our project in order to predict [10] the next word in the sequence of a word.

III. PROPOSED METHODOLOGY

Here we came with our own proposed methodology which is helping in generating Stories for sequence of images.

3.1 Flow of processes step-by-step



First we will give an image as input from a web cam and that image will be passed to CNN [5] for image classification [6].for each image we take two keywords based on objects combinations.[Image classes are Person,Dog,Bird and Cat].These keywords represents multiple objects present in the input image.For Example if cat and dog present in a given input image then we will assign two keywords like "cat and dog" , dog and cat".

Those two keyword from CNN will be given to LSTM [4] and we generate two sentences [10,11] through LSTM where as the LSTM is trained on some sentences or on some text.i.e "sentence1" is for "keyword1" and "sentence2" is for "keyword2" respectively.

Here we are generating two sentences because if we generate only a single sentence then we will don't have a chance to compare that single sentence with any other sentence and we need to append that sentence to our generating story even if that sentence is suitable or not suitable. then that generated story may be not related to our stream of images.

So, instead of generating a single sentence, we are generating two or more sentences so that we can compare those sentences with previously fixed sentence in a story and we will get a related sentence so that generated story will be have meaningful description about the images.

LSTM [4] can't generate the exact sentence we want because it is trained on so many stories. So it just predicts the next word and forms as a sentence and that sentence may not be related to our images descriptions.

In order to generate the nearest sentence of LSTM generated sentence we are taking help of some NLP text pre-processing techniques [12] those are mentioned below.

3.2 Removing StopWords

A stop word [12] is a commonly used word (such as "the", "a", "an", "in") that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.

We don't want these words to take valuable processing time. For this we can remove them easily, by grouping those words as a list and those can be consider to be stop words [12]. Natural Language Toolkit(NLTK) in python has some list of stopwords stored in 16 different languages. You can find them in the nltk_data directory. ./nltk_data/ corpora/stopwords file.

```
importnltk
fromnltk.corpus import stopwords
set(stopwords.words('english'))
```

which displays the stop words in englishlanguage.we can even add extra stop words of our choice to the list by modifying that list or .txt. file in the stopwords directory.

Example

Berfore removing stop words

1. Rama is the king of ayodhya
2. Rama's wife is seetha

After removing stop words

1. Rama king ayodhya
2. Rama wife seetha

stop words : is,the,of, 's, is

3.3 Lemmatization

Lemmatization is nothing but it is the process of grouping the different inflected kinds of a word.so the meaning of those different inflected words can be analysed to a single meaning.Lemmatization is similar to stemming but is group the words which can bring context to that wordrocks : rock , corpora : corpus , better : good

3.4 Unique Words

After the process of Removing StopWords and lemmatization the remaining words are considered to be Unique Words

Example

1. Rama is the king of ayodhya
2. Rama's wife is seetha

unique word : Rama,king,ayodhya,wife,seetha

3.5 Text to Vector

After performing the above process we convert all the unique words to a vector and save this in a .csv file.

3.6 After text processing

Text pre-processing [12] will compare two LSTM generated sentences with other sentences [which we used to write the stories dataset] and find the nearest sentence for those two.

For the first image in the sequence , among the two generated and matched sentences we select and fix the highest occurring starting sentence of the stories. For the remaining images in the sequence , we try to find the highest occurring sentence in stories among the two matched with respect to the previous fixed sentence of the story and append that to the output story. In this way the story will be generated.

3.7 Language Modelling

With the latest techniques and developments of Natural Language Processing(NLP) so many problems are easy to solve and becoming successful in executing it. Language modelling [10,11] is a method which uses deep learning models for text generation.

with the help of Language Modelling[10,11] Text Generation is a quit easy task which can be used to explain any situations in a written format as a sentence or a paragraph which is a type of language model.LanguageModelling is the main problem of so many tasks of NLP which is used in speech-to-text,summarization of text, and convertionalsystems.that means a language model will learns occurrence of likelihood words by taking the previous words as reference and generate the next words .it can be used useful at single character level,n-gram level,sentence or paragraph level.

So we used the concept of LanguageModelling[10,11] for the purpose of text generation i.e sentence generation in our project.

For this language modelling we selected advanced neural network of Recurrent Neural Networks(RNN) [3] i.e LSTM [4] which is having capabilities of long term dependency of the words in a sentence.

3.7.1 How we are using Language Modelling in Our Project ???

When we give an image as input to the CNN it will generate keywords i.e class names or other names which are manually set by us based on the combination objects in an image.

On other side we will train our languages model on some text, paragraphs, sentences or some stories.

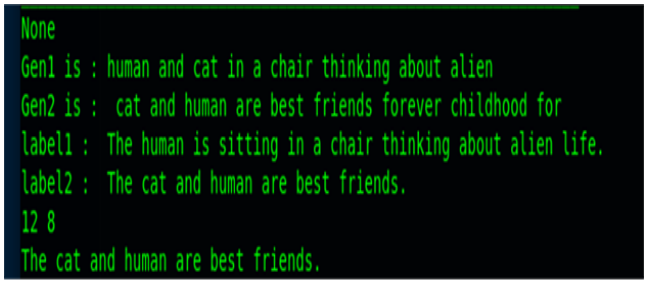
So the keyword will be passed to the Language modellingi.e LSTM [4] by taking that keyword as input it will generate next highest probability [10] word as output which is sequence to the received keyword. In this way the process goes on and forms a sentences with those predicted words by our language modelling.

IV. OUTPUT RESULTS

The tested results for a small stories dataset may not be accurate but we believe that for a big stories dataset the output story will be more accurate.

4.1 Input Image-1 :

First sentence of the output story



4.2 Input Image-2 :

Second sentence of the output story

V. CONCLUSION

Our thought on "Story Generation for Image Sequence" will be useful for the blind people in some situations like if any blind people want to go somewhere then they just captures their surroundings so that our story generator just take those images and understand the content of the images and tells them the situations of their surroundings as a story.

Short story for an image sequence is much more understandable and relatedness among the image content increases. Captioning of an image does not depend on the

previous image or the next image whereas in story generation current output sentence depends on the previous output sentence thus maintaining the relation among the images resulting in much more meaningful story.this can be used for computer story generation and narration in the coming future.

REFERENCES

1. OriolVinyals, AlexanderToshev,SamyBengio, DumitruErhan.: Show and Tell: A Neural Image Caption Generator. https://arxiv.org/abs/1411.4555.(2014)
2. Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, Wei Xu.: CNN-RNN: A UnifiedFramework for Multi-label Image Classification. https://arxiv.org/abs/1604.04573.(2016)
3. Zachary C. Lipton, John Berkowitz.: A Critical Review of Recurrent Neural Networksfor Sequence Learning. https://arxiv.org/abs/1506.00019.(2015)
4. Klaus Greff, Rupesh K, Srivastava, Jan Koutn´ık, Bas R. Steunebrink, J¨urgenSchmidhuber.: LSTM: A Search Space Odyssey.https://arxiv.org/abs/1503.04069.(2017)
5. Keiron O'Shea, Ryan Nash.: An Introduction to Convolutional NeuralNetworks. https://arxiv.org/abs/1511.08458.(2015)
6. M Manojkrishna, M Neelima, M Harshali, M VenuGopalaRao.: Image classification using Deep learning.(2018).
7. AshwinBhandare, Maithili Bhide, PranavGokhale, RohanChandavarkar.: Applications of Convolutional Neural Networks.(2016)
8. Joseph Redmon, Ali Farhadi.: YOLOv3: An Incremental Improvement.https://arxiv.org/abs/1804.02767.(2018)
9. Tsung-Yi Lin, Michael Maire, Serge Belongie, LubomirBourdev, Ross Girshick, James Hays, PietroPerona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollar.: Microsoft COCO: Common Objects in Context. https://arxiv.org/abs/1405.0312.(2015)
10. YoshuaBengio, RéjeanDucharme, Pascal Vincent, Christian Jauvin.: A Neural Probabilistic Language Model.(2003)
11. Thomas Cherian, AkshayBadola,VineetPadmanabhan.: Multi-cell LSTM Based Neural Language Model.https://arxiv.org/abs/1811.06477.(2018)
12. Jose Camacho-Collados, Mohammad TaherPilehvar.: On the Role of Text Preprocessing in Neural Network Architectures:An Evaluation Study on Text Categorization and Sentiment Analysis. https://arxiv.org/abs/1707.01780.(2018).