Quest Journals
Journal of Software Engineering and Simulation
Volume 8 ~ Issue 1 (2022) pp: 06-12
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Nollywood Movie Sequences Summarization Using a Recurrent Neural Model

Ofut, Ogar Tumenayu[1]
Department of Computer Science
Cross River University of Technology Calabar, Nigeria.

Francis Sunday Bakpo
Department of Computer Science
University of Nigeria Nsukka, Nigeria

Debora U Ebem
Department of Computer Science
University of Nigeria Nsukka, Nigeria

Abstract
Summarization is the act of presenting the most significant information from an entire work like movie, audio, and text. Movie Sequences Summarization consist of meaningful representation of a movie which mostly enhance effective browsing of large movie collections as well as providing fast content accessing and indexing. In this paper, the propose system is targeted at summarizing Nollywood movie sequences by using a modify Recurrent Neural Network model. Long short-term memory (LSTM) and Bi-directional Long short-term memory (Bi-LSTM) were employed for the enhancement of recurrent neural networks model. The proposed model can condense huge movie and deliver an accurate but brief information about a movie in less time. A custom dataset for Nollywood movie was design following the full protocol of TVsum dataset.

Keywords: *Summarization; Movie; Dataset; Nollywood; Model; Long short-term memory; and Bi-directional Long short-term memory.*

***R**eceived 04 Jan, 2022; **R**evised 13 Jan, 2022; **A**ccepted 15 Jan, 2022 © The author(s) 2022.*
P**ublished with open access at **www.questjournals.org

I. Introduction

Automatic Summarization in general is currently being described as most challenging and interesting problems in the field of Natural Language Processing (NLP). The increasing volume of movie data makes analysis task difficult. Browsing tools are very demanding for the users to obtain a quick information about the movie content.

The application of movie sequences summarization model to Nollywood movie industry in place of movie previews to predict the genre of the movie is not just an interesting computer vision problem to solve, but also has extended benefits to stakeholders and decision makers in the industry such as: (i) automatically tagging the genre on content-hosting websites like YouTube and others; (ii) to reduce viewing time, (iii) when categorizing movies, summarization make selection process easier, (iv) automatic summarization improves the effectiveness of indexing, (v) automatic summarization algorithms are less biased than human summarizers, (vi) Personalized summaries are useful in question-answering systems as they provide personalized information [1].

Movie sequences summarization task aims at browsing a set of frames or segments from a visual sequence which contains most significant and informative movie scenes across the entire sequence. Not only is summarization useful for efficiently extracting the substance of data, it also serves many other applications such as video indexing [2], video retrieval [3], and anomaly detection [4].

The Nigerian movie industry called Nollywood is the second highest revenue earner in present day Nigeria. The twenty-first century Nigerian movie industry (Nollywood) produces approximately 2,000 movies a year, which arguably places it in third place on the global movie ranking. Thus, stakeholders in the movie industry are overwhelmed with an enormous and increasing amount of movie information, which often makes it very difficult to management, search, and retrieval of specific content. The arrangement of Nollywood movies according to their respective categories and subject matter is extremely important when searching for a specific movie. Currently, this categorization in Nollywood movie industry is dependent on movie preview or trailers.

Movie previews are intended to get the target audience excited about seeing a movie at the cinema. Movie trailers are growing more important in the marketing of films because in the past they were typically confined to theaters and screened during the previews for upcoming attractions [5]. Currently, the existing process of making movie preview in Nollywood goes through almost a manual process by the movie producers with rigorous steps. The existing process of movie preview for Nollywood movie industry is highly unreliable in application and hence a better way is required to create a summarized representation of the movie that is easily comprehensible in a short amount of time.

The abstractive summarization has great potential for generating a high-quality summary by using sequence-to-sequence (seq2seq) modeling [7]. Sequence-to-sequence modeling, a recurrent neural network model, takes a sequence of movie as input and generates a sequence of movie as output. The encoder–decoder architecture of the seq2seq model can be implemented by an RNN framework called long short-term memory (LSTM) [8]. It has been successfully applied to a variety of NLP -based problems like machine translation, headline generation, video summarization, and speech recognition.

In this research work, we intend to focus on how to modify Recurrent Neural Network (RNN) with the ability to understand how the task of movie sequences summarization are done. Recurrent neural networks represent an important class of machine learning techniques that are specialized for processing sequential data [6].

The expected output of the summary is usually made up of a set of video clips which is an abstract of the original video with some editing process. The main reason for movie sequence summarization is to speed up browsing of a large collection of movie data, and achieve efficient access and representation of the video content. By watching the summary, users/stakeholder can make quick decisions on the usefulness of the video.

II. Related work

Video or movie summarization task has been studied for about two decades. During these years, many approaches have been developed by exploring cues ranging from low-level visual inconsistency, attention [9] to high-level semantic change of concepts [10] and entities in videos [11]. However, most of these studies focus on unsupervised leaning technique. Recently, the research focus has been extending to supervised learning approaches [12] and [13], which aims at explicitly learning the summarizing capability from the human labels. Supervised approaches have always produced better performance than unsupervised ones.

In the past few years, the Recurrent Neural Network (RNN) – a type of neural network that can perform calculations on sequential data (e.g. sequences of words) – has become the standard approach for many Natural Language Processing tasks. In particular, the sequence-to-sequence model with attention [14].

One of the researches done by [15] pioneered the application of LSTM for supervised video summarization to model the variable-range temporal dependency among video frames to derive both representative and compact video summaries. With the propose method the authors was able enhance the capacity of the LSTM with the determinantal point process which is a probabilistic model for diverse subset selection. Zhang et al. [15] develop a bidirectional LSTM to predict the probability of each shot to be selected. Furthermore, a hierarchical architecture of LSTMs is constructed to deal with the long temporal dependencies among video frames [16]. They were limited in capturing the video structure information, because the shot was obtained by fixed length segmentation.

In another work [17] the author constructed a deep rank model relying on two CNNs .ie., AlexNet [18] and C3D [19] stitched with two Multi-Layer Perceptron (MLP) behind their final pooling layers. With frames or subshots as input, the deep rank model produce a ranking score. Normally, a higher score indicates higher probability of that frame or subshot to be chosen as summary. LSTM is employed in [20] to model the video sequence and rank the video subshots, which has achieved the state-of-the-art results in video summarization. However, to address the weakness in long temporal dependency, the input sequence to the LSTM is created by the mean uniform sampling of frame features, which will advocate for inevitable information loss. Actually, the proposed approach in this research is essentially developed to solve this problem.

The essence for applying recurrent neural networks to enhance movie sequences summarization is due to several factors. Firstly, it is a theoretical possibility to approximate arbitrary functions by using recurrent neural networks [21]. Secondly, it is the simplicity with which the learning process scales. Inverse learning algorithm error propagation makes it easy to read the gradient of the loss function by model parameters and do it in parallel. Thirdly, the growth of computing power scalability of learning processes would not cost anything if it were not for the ability to scale all the computations.

In this research, we will analyze in details the current methods of enhancing recurrent neural networks (RNN) to be apply in selecting a few frames that are deemed useful for a summary based on the information that they are conveying from the original movie, and also propose a number of new algorithms that solve this problem more efficiently.

III. The propose Method

Our propose method is made up two distinct layers RNN model, the first layer is an LSTM that tailored toward shots segmentation, the second layer Bi-LSTM to exploit the inter-subshot temporal dependency and ascertain whether a particular subshot to be included in the summary or not.

The proposed method is an unsupervised movie summarization, where a smaller number of frames are selected from the original movie on the basis of threshold value. The method is demonstrated in the following stages: The first stage is input the movie file. The second stage is splitting the input movie file into frames (f1, f2,…fn). The third stage is movie processing that entails key frames extraction on the basis of threshold values. In achieving the third stage we employ a machine learning approach where the machine is train with some distinct features that identified an interesting frame to be selected to the summary. The classification of the trained features is then fed into our modified recurrent neural networks (RNN) to apply in selecting a few frames that are deemed useful for the summary based on the information that they are conveying from the original movie. The final stage is the output of the new summarized movie. The diagram below (Figure 3.1) shows the overall framework of the proposed method.

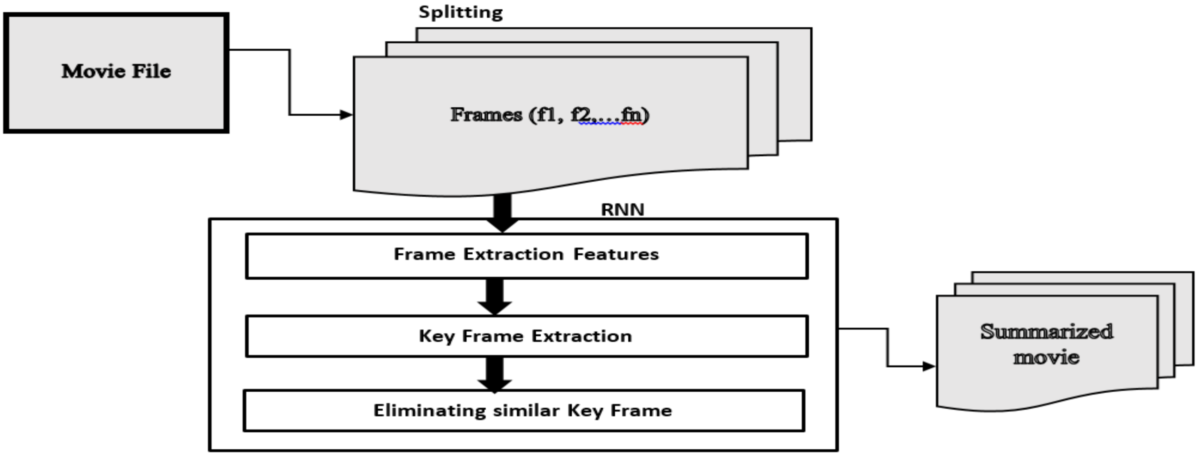


Figure 1: framework of the proposed method

The typical RNN should work efficiently in sequence modelling. However, it is really difficult to train due to gradient vanishing problem. LSTM is designed to address this issue, which is the most popular option of standard RNN. Specifically, LSTM is an extension from standard RNN with an extra memory cell, which is applied to selectively memorize the previous inputs. This variant of LSTM constitutes the first layer of our model.

The second layer of RNN model is the Bi-directional LSTM (Bi-LSTM), which is composed of a forward and a backward LSTM. Bi-LSTM is then employed to exploit the inter-subshot temporal dependency and ascertain whether a particular subshot is valuable to be a key subshot or not to be included in the summary. The bi-directional LSTM is made up of a forward LSTM and a backward LSTM. The main advantage between them is that the backward LSTM operates reversely. The two layers exploit the intra-subshot and inter-subshot temporal dependency, respectively, and the output of the second layer is utilized to predict the confidence of each subshot to be selected into the summary. The motivation for formulating the proposed framework is to improve its potential to handle long-range temporal dependency of the movies in other to address the peculiar needs for automatic Nollywood movie sequences summarization.

IV. Dataset

To the best of our knowledge there is no publicly available dataset suitable for the purpose of automatic summarization of Nollywood movie sequences. We therefore collected a new data sample from YouTube, Netflix and IROKOTV in line with a benchmark dataset Title-based Video Summarization (TVSum) to create a custom dataset for Nollywood movies summarization. TVSum is a benchmark dataset that is employ for movie sequences validation summarization techniques [23]. Our Nollywood custom dataset is therefore design following the full format of TVsum dataset and it's imbedded into the original TVsum dataset to be use for the training of our model.

4.1 Nollywood Movies Custom Data pre-processing

A preliminary processing of data in order to prepare it for the primary processing or for further analysis is carried out at this stage by utilizes existing tools and methodologies employed by TVsum. Firstly, a full Nollywood movie file is automatically divided into a set of shots and scenes, where high-level visual and sound features are annotated manually for each shot. The high-level visual features especially the semantic concepts chosen by considering their frequency of occurrences within the domain of movie and their ease of detection using a machine learning classifier. Visual features such as semantic event/actions are annotated manually.

Speech transcripts of each shot are selected using Fast Forward Moving Picture Expert Group (FFMPEG). Movie shorts contain spoken content as audio data along with their starting and ending timings in the corresponding movie sequence. Since a movie is segmented into a set of shots based on the visual properties of the movie, audio sound in movie shorts have the same starting and ending timings with the shorts. As such, speech transcripts of each shot are synchronized with the video frames.

4.1.1 Steps for creating a Nollywood custom dataset

i. Collect data: Nollywood movie files of deferent genres were collected from YouTube and IROKO TV.

ii. prepare data: Take sample shots from each video files collected

iii. Annotate or label data: The movie data was annotated in deferent ways of annotating video data based on interest: highlighting objects of interest in video frame and labelling them.

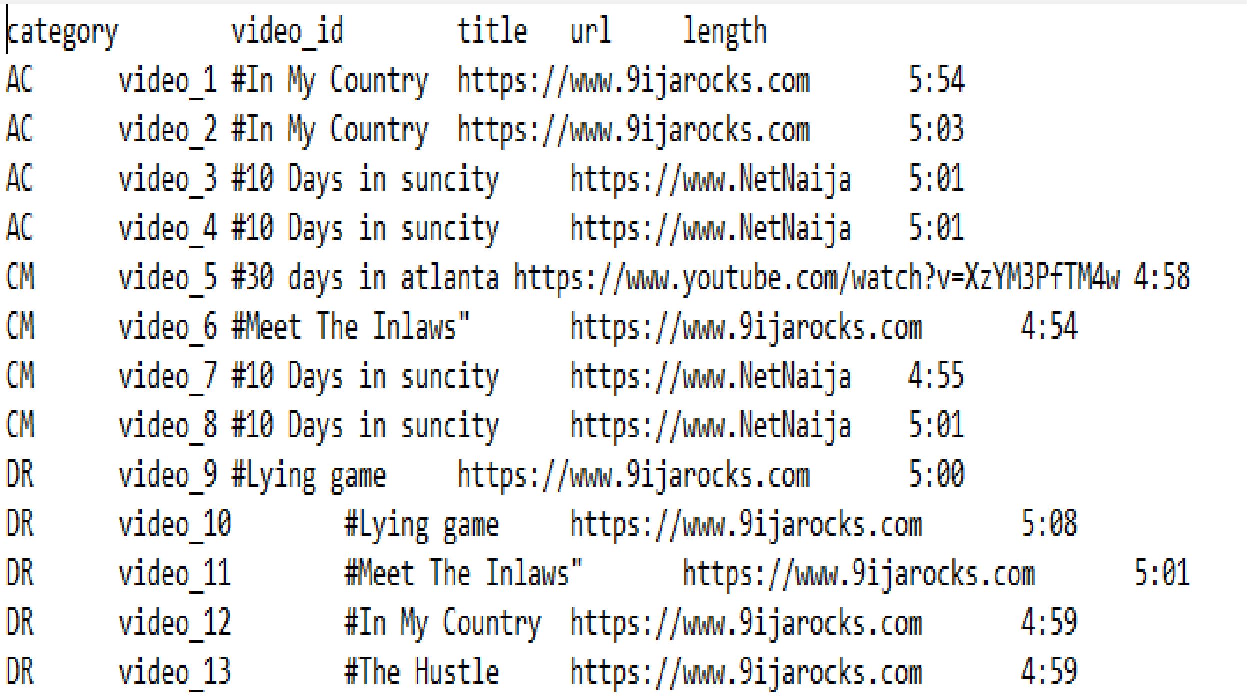
iv. export the annotated data into a file format (eg. JSON, TSV, CSV or txt etc)

v. create an ML model.

vi. used the prepared dataset to train the model.

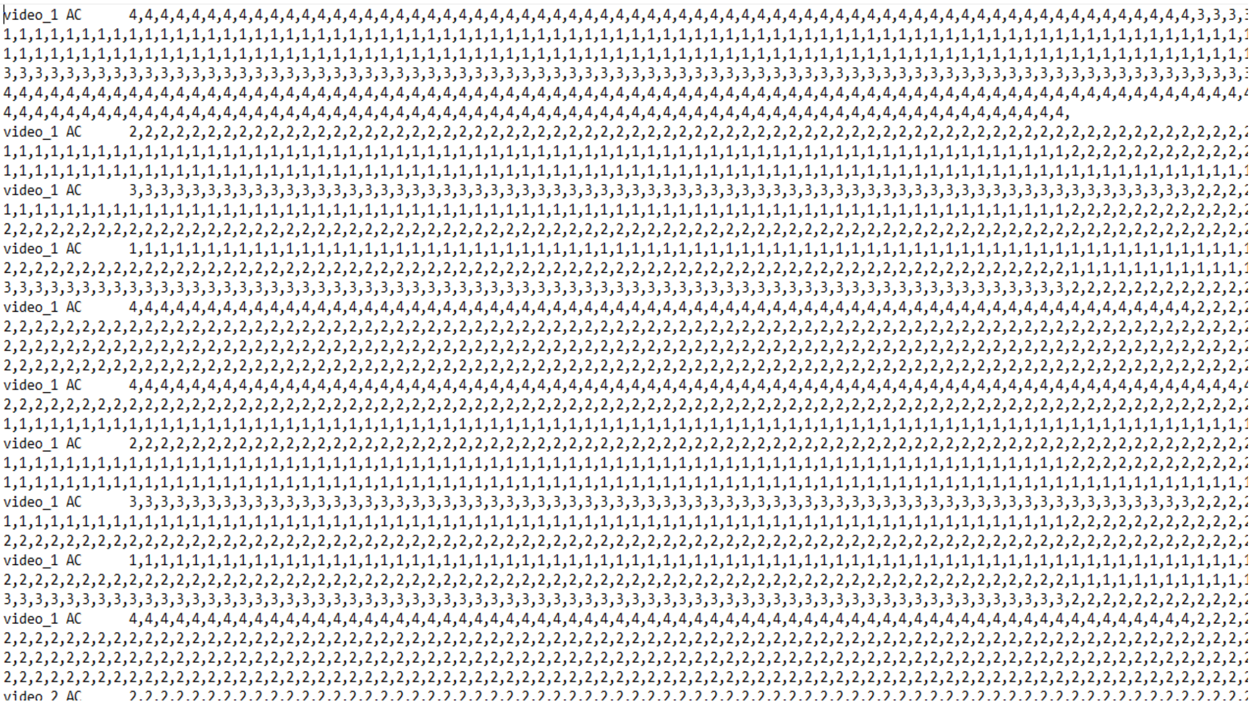
vii. validate and test model.

Figure 2. Sample data for Nollywood custom dataset



```
category        video_id        title   url     length
AC      video_1 #In My Country  https://www.9ijarocks.com       5:54
AC      video_2 #In My Country  https://www.9ijarocks.com       5:03
AC      video_3 #10 Days in suncity     https://www.NetNaija    5:01
AC      video_4 #10 Days in suncity     https://www.NetNaija    5:01
CM      video_5 #30 days in atlanta https://www.youtube.com/watch?v=XzYM3PfTM4w 4:58
CM      video_6 #Meet The Inlaws"       https://www.9ijarocks.com       4:54
CM      video_7 #10 Days in suncity     https://www.NetNaija    4:55
CM      video_8 #10 Days in suncity     https://www.NetNaija    5:01
DR      video_9 #Lying game     https://www.9ijarocks.com       5:00
DR      video_10        #Lying game     https://www.9ijarocks.com       5:08
DR      video_11        #Meet The Inlaws"       https://www.9ijarocks.com       5:01
DR      video_12        #In My Country  https://www.9ijarocks.com       4:59
DR      video_13        #The Hustle     https://www.9ijarocks.com       4:59
```

Figure 4.3: Nollywood custom dataset description



```
video_1 AC      4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,3,3,3,3
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,
video_1 AC      2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
video_1 AC      3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
video_1 AC      1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1
3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,2,2,2,2,2,2,2,2,2,2,2,2
video_1 AC      4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
video_1 AC      4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
video_1 AC      2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
video_1 AC      3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,2,2,2,2
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
video_1 AC      1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1
3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,2,2,2,2,2,2,2,2,2,2,2,2
video_1 AC      4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
video_2 AC      2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2
```

Figure 4.4: Nollywood custom dataset annotation

We selected 10 categories from the TRECVid Multimedia Event Detection (MED) task [24] and a collection of 13 Nollywood movies (5 Drama, 4 Actions, and 4 Comedy category) from YouTube, Netflix and IROKO TV using the category name as a search query term. From the search results, we chose videos using the following criteria: (i) under the Creative Commons license; (ii) duration is 2 to 10 minutes; (iii) contains more than a single shot; (iv) its title is descriptive of the visual topic in the video. We collected videos representing various genres, including drama, comedy, action. This composition gave us a custom dataset that was impendent into the pre-train TVSum dataset following the same file format, frame size and others in order to suit our purpose.

V. Limitation

The major limitation of this research work would border on time, finances and availability of previously research work aim at developing the Nollywood movie industry. The memory requirement on specific devices for training and evaluation our model will also post some limitation. The propose solution to this problem is an online Graphics Processing Units (GPUs) Google Colab that was adopted to significantly accelerate the training process of our machine learning model.

The framework is not intended for special-purpose video summarization such as security/surveillance due to the difference in focus of such applications video summarization in the surveillance field focuses more on the detection of certain suspicious events than on user satisfaction.

VI. Conclusion

In this work we proposed a novel recurrent neural network model for the summarization of Nollywood movie sequences. Our framework adopted the combination of both LSTM and Bi-direction LSTM as an effective solution to the classification problem for a large number movie frames. Our LSTM layer was created to establish a sliding operation that enables short LSTM to process long videos. This act is expected to avoids long temporal dependency exploitation among thousands of frames, which can moderate the problem of vanishing gradient. While the bidirectional LSTM jointly captures the forward and backward information in frame sequence, which can detect the shot boundary effectively.

In order to achieve accurate and precise Nollywood movie summaries we created and added a custom dataset to TVsum dataset. This became necessary due to the fact that existing benchmark datasets do not contain Nollywood movie data frames that could be used for our model training in order yield good prediction for Nollywood movie summarization.

The full implementation of our framework shall produce an automatic Nollywood movie sequences summarization model that would be capable of generating a short and fluent summary of a longer movies, which shall mine appropriate information from the input movie to utilize the relevant information faster.

References

[1]. Juan-Manuel Torres-Moreno. (2014). *Automatic Text Summarization*. London: ISTE ; Hoboken, NJ : Wiley, 2014.

[2]. Richang Hong, Lei Li, Junjie Cai, Dapeng Tao, Meng Wang, and Qi Tian. Coherent semantic-visual indexing for largescale image retrieval in the cloud. IEEE Transactions on Image Processing, 2017.

[3]. Erkun Yang, Cheng Deng, Wei Liu, Xianglong Liu, Dacheng Tao, and Xinbo Gao. Pairwise relationship guided deep hashing for cross-modal retrieval. In AAAI, pages 1618–1625, 2017.

[4]. Yachuang Feng, Yuan Yuan, and Xiaoqiang Lu. Learning deep event models for crowd anomaly detection. Neurocomputing, 219:548–556, 2017

[5]. Stephen, G.(Jan 13, 2012). *The Art of First Impressions: How to Cut a Movie Trailer. film maker magazine*. http://filmmakermagazine.com

[6]. Koutras, P., Zlatintsi, A. Elias I, Athanasios K., Petros M., and Alexandros P. (2015). *Predicting audio-visual salient events based on visual, audio and text modalities for movie summarization*. In Proc. ICIP. IEEE, 4361–4365.

[7]. Nallapati, R.; Zhou, B.; Santos, C.; Gulcehre, C.; Xiang, B.(2016) "Abstractive text summarization using sequence-to-sequence RNNs and beyond. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning; Association for Computational Linguistics: Berlin, Germany, 2016; pp. 280–290.

[8]. Zhang, Y., Liu, Q., and Song, L.(2018). *Sentence-State LSTM for Text Representation.* In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics Association for Computational Linguistics: Melbourne, Australia, 2018; Volume 1, pp. 317–327.

[9]. Chong Wah Ngo, Yufei Ma, and Hongjiang Zhang, (2005) "*Video summarization and scene detection by graph modeling,*" IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 2, pp. 296–305.

[10]. Xun Xu, Timothy M. Hospedales, and Shaogang Gong (2017). Discovery of shared semantic spaces for multiscene video query and summarization, IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 6, pp. 1353–1367.

[11]. Adway Mitra, Soma Biswas, and Chiranjib Bhattacharyya (2017). *Bayesian modeling of temporal coherence in videos for entity discovery and summarization*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 3, pp. 430–443.

[12]. Boqing Gong, Weilun Chao, Kristen Grauman, and Fei Sha, (2014). *Diverse sequential subset selection for supervised video summarization*, in Advances Neural Inf. Process. Syst., 2014, pp. 2069–2077.

[13]. Gygli, M.; Grabner, H.; Riemenschneider, H.; and Van Gool, L. 2014. Creating summaries fromuser videos. In ECCV, 505–520. Springer.

[14]. Abigail See, Peter J. Liu, and Christopher D. Manning (2017). *Get to the point: Summarization with pointer-generator networks,* In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1073–1083,.

[15]. Ke Zhang, W.-L. Chao, F. Sha, K. Grauman (2016). *Video summarization with long short-term memory*, in: European conference on computer vision, Springer, 2016, pp. 766–782.

[16]. Zhao B., X. Li, and X. Lu.(2017) Hierarchical recurrent neural network for video summarization. In Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017, pages 863–871,.

[17]. Ting Yao, Tao Mei, and Yong Rui. (2016) Highlight Detection with Pairwise Deep Ranking for First-Person Video Summarization. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, 982–990.

[18]. Alex Krizhevsky, Ilya Sutskever, and Georey E. Hinton. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States. 1106–1114.

[19]. Du Tran, Lubomir D. Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. (2014) C3D: Generic Features for Video Analysis. 2014CoRR abs/1412.0767.

[20]. Huan Yang, Baoyuan Wang, Stephen Lin, David P. Wipf, Minyi Guo, and Baining Guo. (2015) Unsupervised Extraction of Video Highlights via Robust Recurrent Auto-Encoders. In 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015. 4633–4641.

[21]. Koutras, P., Zlatintsi, A. Elias I, Athanasios K., Petros M., and Alexandros P. (2015). *Predicting audio-visual salient events based on visual, audio and text modalities for movie summarization*. In Proc. ICIP. IEEE, 4361–4365.

[22]. Zhao, Bin & Liu, Wei & Lu, Xiaoqiang. (2018). HSA-RNN: Hierarchical Structure-Adaptive RNN for Video Summarization. 7405-7414. 10.1109/CVPR.2018.00773.

[23]. Yale Song, Jordi Vallmitjana, Amanda Stent, and Alejandro Jaimes (2015). *TVSum: summarizing web videos using titles*, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 5179–5187, 2015

[24]. Smeaton, A., P. Over, and W. Kraaij (2006). Evaluation campaigns and TRECVid. In MIR, 2006.