Research Article

CONTENT BASED IMAGE RETRIEVAL (CBIR) FOR STORING OF PRODUCTS

[1]Dr. Kannan. V,

[2] Dr. Mrs. Sangeetha. P,

[1]Professor, Department of Management Studies, Kumaraguru College of Technology, Coimbatore

[2]Associate Professor, D J Academy for Managerial Excellence, Coimbatore

Article History: Received:11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: With rise in complexity of materials and lack of necessary skills among the mass population, it is necessary to find a system that can help the employees to identify the materials without the help of additional labour, training and manuals. Hence, a system has been created to identify the material and give similar materials when the model is provided with just an image of the material in question. This being an unlinked database helps in maintaining privacy of the company. Convolutional neural networks are used for object detection and then captioning is done by Recurrent neural network which also compares the generated caption with the provided database and return the best match. LSTM and NLP aid the process of caption generation and search. The dataset used is MSCOCO 2014. The evaluation metrics is BLEU which returned a score of 70.3%. The whole idea is easy to combine with concepts like KANBAN and facilitate the layout design of the company to a great flexibility. By reducing the time spent on training and sorting the efficiency of the firm is increased. The system created for identifying the material can be interlinked with centralised inventory management system to help track the material in the production process. The layout of the store was not optimised which resulted in delay in the production process. This can be rectified using a software like ARENA to design the store layout. With a system in place for material deduction the steps followed in stores can be reduced.

Keywords: NLP, MSCOCO 2014, KANBAN, BLEU.

I. Introduction

India is the leading provider of Data solutions to the world with the country having a huge labor force and skill force. The country has a growth rate of 34% in this sector. The country is producing numerous start-ups and has a vast array of companies which are taking in huge number of employee's day to day. These companies have a huge product mix which is hard to keep track. The company employee must know what a product is, what are it is similar products etc. Human learning and recognition are not as efficient to adapt the rapid changes. Here, a simple reverse search engine which incorporates a caption generator, helps in identifying the product in question with an image of it and gives a similar image to find where the product is and to store it in a separate data base. To train and teach about the company products is a tedious process. Hence, a system has been proposed with low cost and highly customizable to needs of the company and maintain its privacy. With growing population, the need for the product gets increased which in turn results in an increase in production. The production increase directly translates the need for storing more materials for production. The material can be stored and retrieved efficiently with image capturing technique. This also helps in removing unused products constantly from stores.

Objectives of the study

- To create a model that can identify and caption a product from an image
- To create an end-end image search system that has good accuracy and closeness to natural language understanding.
- To train the neural network to be adaptive in nature to reduce the work of RNN by using LSTM.
- To reduce the picking time of material in the stores to start the production process
- To change the store layout based on output from the image capturing based on ABC and FSN analysis.

II. Review of Literature

Andrej Karpathy and Li Fei Fei (2014) in their study Deep visual semantic alignments for generating image descriptions presented a model that generates natural language annotations of images and their boundary boxes of objects. This approach learns about inter-modal co- respondence between description and image data. This is a combination of CNN for image and bi-directional RNN over sentences. Datasets used are Flickr8k, Flickr30k, MSCOCO.

Google team (2015) - the whole model was evaluated using BLEU on datasets like PASCAL, FLICKR30k, SBU, COCO. The model used LSTM and transfer learning. The increase in score are as follows PASCAL (29-59), FLICKR 30k (56-66), SBU (19-28). Human score on PASCAL is 60.

Guoyong Duan, Jing yang, Yulong Yan (2016) in their paper CBIR research emphasized the importance of CBIR and advantages with the key technologies used in it. They also introduced a new technology that combines colour, texture, shape for an image retrieval. This also focuses on feature extraction.

Jerzy Buliński, Czesław Waszkiewicz, Piotr Barczewski (2013) - Based on the product list and information connected with the sale level, the goods were grouped into categories to achieve better financial results.

III. Research Methodology

Tools Used

The tools used for the study are TensorFlow, Keras, Anaconda, Python 3.0, ABC and FSN analysis , KANBAN

i) Tensor Flow

Tensor Flow is an open-source package that is used in python for machine learning and deep learning concepts. It is designed by google brain team and is used for variety of tasks.

ii) Keras

Keras is a deep learning library written in python that is run on top of TensorFlow. It enables fast experimentation with deep neural networks.

iii) Anaconda

Anaconda is an open source IDE for Python and R. It simplifies environment creation and package installations.

iv) Python 3.0

Python 3.0 is the latest release series of python that is a high-level general-purpose language.

v) NLTK

The Natural Language Toolkit, or more commonly NLTK is a text classified that uses feature extraction and converts them to vectors.

CONCEPTS

The following are the concepts used in this study:

- Deep learning
- Neural networks
- Image captioning
- CNN
- RNN
- LSTM
- Transfer learning

Deep Learning

Deep learning is a subset of machine learning that involves artificial intelligence and uses multiple layers to extract features from input. The word deep learning was introduced by Rina Dechter during the year 1986. The base concept of multilayer perceptrons was a paper by Alexey ivakhnenko. This can be used for many purposes from image data extraction to text data extraction. The word deep in deep learning refers to no of layers. This uses neural networks to accomplish its tasks. The learning part which is the most important part of the network can be supervised i.e., a human input is required all the time unsupervised where no human interference is required for completion. The learning can also be semi-supervised where partial human help is needed. The deep learning suffered from many numbers of errors and these problems were sorted out once the concept of back propagation was introduced by Yann LeCun et al in 1989.

NEURAL NETWORKS

These are computing systems that are modeled on basis of a human brain. The advent of neural network started as early as 1950s when mathematicians proposed the idea of replicating a human neuron. Neural networks are essentially a set of neurons like structures working on a task which makes them efficient and since they are on par with human nerves their accuracy is increasing day by day.

IMAGE CAPTIONING

Image captioning is a process by which images are processed and captioned accordingly to what is in the image. The caption must be capable of providing the objects in the image, relationship between them, action done by them and most importantly describe them as simple as it can be with accuracy and removing un-necessary information. This is done by optimizing the results. An image can have number of captions but choosing the best one that describes them is the task. This is done by NLP. A caption should encompass all the objects and their association with each other and the action they present in the image.

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

CONVOLUTION NEURAL NETWORKS

A type of neural network that involves using concept of convolutions. This is best suited for image processing. The term convolution is a mathematical operation done on two functions to produce a third function.

RECURRENT NEURAL NETWORKS

A type of artificial neural network used in natural language processing and designed to recognize a data's sequential characteristics and use patterns to predict next likely scenario. Unlike feedforward neural networks RNN can use their internal memory to process inputs. They can use the input again and again in comparison to solve complex problems inputs. They use output of the previous time step (feedback loops) to process a sequence of data that inform the final output, which can also be a sequence of data. These feedback loops allow information to persist; the effect is often described as memory.

LONG-SHORT TERM MEMORY

These are an extension of RNN and are used to extend their memory and filter out unwanted inputs and provide only necessary inputs to RNN. This is like a gated cell. This has three gates a forget gate, input gate, and remember gate

TRANSFER LEARNING

Transfer learning is a technique by which a model trained for one purpose is re-used for another purpose of same or different nature. This is achieved by removing the final dense layers to retain the trained phase of the model. This is a popular technique in the field of deep learning as it saves training cost, skips training phase, retains trained knowledge and enables the model to be used with less computation costs.

ABC analysis

A - **Items:** These are 5 – 10% of the items with a 70 – 75 % value.

B - **Items:** These are 10 – 15% of the items with a 10 – 15 % value.

C - **Items:** These are 80 -90% of the items with 5 – 10 % value.

FSN analysis

F-S-N Analysis is based on the annual usage of items and are classified as F (fast-moving), S (slow-moving) and N (non-moving).

Kanban

Kanban is a lean method to manage and improve work across human systems. This approach aims to manage work by balancing demands with available capacity, and by improving the handling of system-level bottlenecks by using image capturing technology.

IV. ANALYSIS

SOURCE OF DATA

The Training of the model is using secondary data sources. Here we used MSCOCO (Microsoft Common Object in Context) 2014 dataset.

DATASET DESCRIPTION

Microsoft Common Objects in Context (MS-COCO) is a dataset created by a team of Microsoft employees in 2014. The data set features 80 different class of objects and over l,60,000 images and over 2.5 million label instances.

METHODOLOGY

The proposed methodology has two phases

PHASE 1: Image captioning phase to generate text captions for images.

PHASE 2: Text similarity based ranking phase on image captions to identify similar images.

PHASE 1

Image captioning is done with the help of neural networks. Neural networks used here are convolution neural network (CNN) and long-short term memory (LSTM) to generate captions. Input image is fed into a pre-trained CNN model which converts image objects into feature vectors using NLTK which is then fed to LSTM which removes the un-necessary captions and data and feeds the necessary data to RNN. The above process is simple as the output of CNN models which is a filtered feature vector of the input image is fed into RNN.

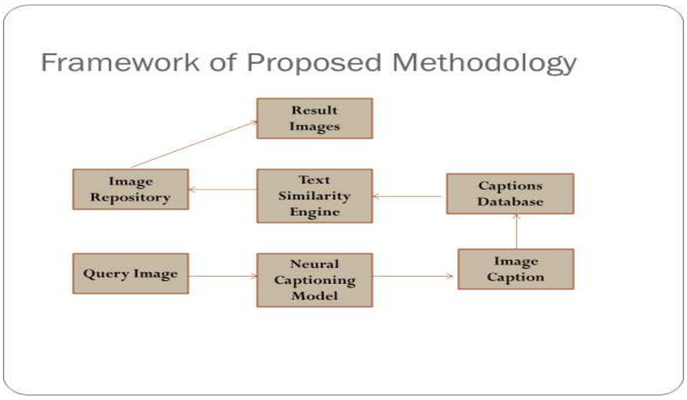
PHASE 2

This phase is divided into sub phases

- Part of speech (POS) tagging of generated captions (the image caption is encoded into a vectorised form using word to vector process).
- Running similarity query on the POS tagged and indexed database of images.
- Query will match noun and weigh them to verbs and scores them. Then Retrieve the top 5 images that fit the similarity query and display the results.

FRAMEWORK

The framework processes the input images using neural captioning model and generates the image captions. The image ID is stored with the corresponding caption with part of speech tag in a database. When a user submits a query image, image caption is generated for this image using captioning model. A database query is run with the caption of the query image to find the most relevant or similar description and the result is displayed in sorted order of relevance of the description.



V. RESULTS

Findings

- It is found out that improper material classification technique is followed in most of the industries.
- Improper layout exists for storing the materials without proper input from technology.

Suggestions

- ABC and FSN analysis can be coupled with KANBAN to keep the material which are used in day-to-day production only inside the stores.
- Proper layout for stores can be developed by image capturing along with ARENA software**.**

LIMITATIONS

The model is limited by many factors both internal and external.

INTERNAL FACTORS

The internal factors that limit the model are as follows:

- RNN is prone to overfitting. The model is time consuming than human.

EXTERNAL FACTORS

- Limited by computational power
- Model is limited to training dataset classes
- Culture aspects and norms are not covered in dataset.
- The test data needs to be within the class of the train data.
- The accuracy of the model depends completely on the computational capacity and training data size.
- The memory used by the model is higher than another model.

VI. REFERENCES

1. C. Rashtchian, J. Hockenmaier, and D. A. Forsyth. Every picture tells a story: Generating sentences from images. In ECCV (4), 2010.

2. I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In NIPS, 2014.

3. J. Mao, W. Xu, Y. Yang, J. Wang, and A. L. Yuille. Deep captioning with multimodal recurrent neural networks. In ICLR, 2015.

4. Keiron O'Shea1and, Ryan Nash ``An Introduction to Convolutional Neural Networks", Aberystwyth University Ceredigion, UK.

5. Mussarat Yasmin, Muhammad Sharif, Sajjad Mohsin "Use of Low-Level Features for Content Based Image Retrieval: Survey." Research Journal of Recent Sciences Vol. 2(11), 65-75, November 2013.

6. R. Kiros, R. Salakhutdinov, and R. S. Zemel. Multimodal neural language models. In ICML, 2014.

7. X. Chen and C. L. Zitnick. Mind's eye: a recurrent visual representation for image caption generation. In CVPR, 2015.

8. Jerzy Buliński, Czesław Waszkiewicz, Piotr Barczewski "Utilization of ABC/XYZ analysis in stock planning in the enterprise" Annals of Warsaw University of Life Sciences – SGGW Agriculture N0 61 (Agricultural and Forest Engineering) 2013 : 89 – 96

9. Felix T.S. Chan, H.K. Chan "Improving the productivity of order picking of a manual-pick and multi-level rack distribution warehouse through the implementation of class-based storage" Expert Systems with Applications 38 (2011) 2686-27