Automatic Generation Image Captions based on Deep Learning and Neural Network

1rd Vamshi ponugoti
department of computer Science
University of central Missouri
Lees's Summit, Missouri
VXP28550@ucmo.edu

2rdGeeta Sree Doddapaneni
department of computer Science
University of central Missouri
Lees's Summit, Missouri
GXD49050@ucmo.edu

3rd Naveena Tangella
department of computer Science
University of central Missouri
Lees's Summit, Missouri
Nxt82010@ucmo.edu

4thThoyaja Kanala
department of computer Science
University of central Missouri
Lees's Summit, Missouri
txk77440@ucmo.edu

***Abstract*—We present a new method for automatic image captioning that utilizes a Recurrent Long Short-Term Memory (R-LSTM) model. Our approach takes an image as input and generates a descriptive caption using a sequence of words. The R-LSTM architecture is a neural network type that can model long-term dependencies in sequential data.**

We first discuss the process of image data pre-processing and feature extraction, where we use a Convolutional Neural Network (CNN) to extract the essential features of the image. We then describe the R-LSTM architecture, including the attention mechanisms that we use to incorporate contextual information, enhancing caption quality.

We evaluate the proposed model using a benchmark dataset, and compare its performance to other state-of-the-art approaches. Our results indicate that our R-LSTM model outperforms these other models in terms of caption quality, as measured by standard evaluation metrics.

We also highlight the challenges of building an effective automatic image captioning system, such as the difficulty of generating semantically meaningful captions and the requirement for large amounts of labeled data. Despite these challenges, we believe that our proposed approach represents a significant step forward in this field, and has the potential to significantly improve the accessibility and usability of visual content on the internet.

In conclusion, we propose an R-LSTM model as a promising approach for automatic image captioning. We hope that our research will inspire further studies in this area and that the continued development of accurate and effective automatic image captioning systems will have a significant impact on image understanding, content accessibility, and image search and retrieval.

***Index Terms*—LSTM,CNN,NLP, Dataset,Word Sense Disambiguation ,VGG.**

[1]

I. Introduction

There are a lot of sources for images, including television, the internet, news, and many others. While most of these images lack descriptions, people are nevertheless able to understand them on their own, whereas it is exceedingly difficult for machines to do so. For machines to understand, descriptions are necessary. The captioning of a natural scene is a well-known research subject that aids in providing a description of an image in the age of artificial intelligence [1]. This research is important for numerous reasons, including the fact that large organizations like Facebook and Google use it to find out where you are, what you're doing, and engage in many other similar activities.

Understanding of images requires the recognition of things, actions, and relationships. The ability of a machine to recognize some silent elements, such as the fact that people are waiting for a train even when the train is not on the platform, is difficult. The generated sentence must be valid in both syntax and semantics [2]. The features of a particular natural landscape can be obtained in order to have a thorough comprehension of it. The method utilized for this is extensively divided into two sets: (1) Deep Learning Based Methods (2) Conventional Machine Learning Based Methods. Massive labeled datasets, like ImageNet and deep learning, have the key benefit that deep convolutional neural networks (CNN) are frequently very useful. Computer vision has benefited greatly from image captioning, or the ability to recognize images, making computers useful for a range of tasks, such as early education, video tracking, sentiment analysis, and the rehabilitation of persons with obvious impairments. More study is being done on image captioning in the area of artificial intelligence. They must be able to be found, their connections must be seen, and the generator must be able to deduce the semantic information in natural language. In particular, the efforts to caption photographs make use of template-based techniques, which need describing a number of aspects, such as direct or indirect objects, as well as their connections and attributes.

Identify applicable funding agency here. If none, delete this.

[1]https://github.com/vxp28550/Final-project.git

The encoderdecoder pipeline, which consists of two easy phases, is the basic foundation for these techniques. First, CNN Image characteristics are used to deduce the image's encoding into rigidly timed embedding vectors. Second, a recurrent neural network decoder is typically employed to generate a linguistic description. Thanks in part to CNN's strengths in representation and RNN's in temporal modeling, neural network-based approaches that can infer new phrases are more widely used.

The development of descriptive [3] image technology may one day enable blind people to "see" the outside world. It has recently acquired more attention and moved to the top among the list of the most important subjects in computer vision. Early methods for producing picture descriptions merged image data using statistical language models and static object class libraries. The authors offer a technique for automatically geotagging photos and use a dependency model to condense a number of web pages that mention image locations. L and co. I developed a network-scale n-gram technique that compiles potential terms and merges them to build sentences that explain images from scratch. I'd like to propose a language model built using the English Gigaword corpus. The hidden Markov model's parameters are then derived from these estimates. The picture description is produced using the sentence's most probable nouns, verbs, circumstances, and prepositions. The goal is to classify each possible region and pass it though a prepositional association function before using a conditionally random field's (CRF) forecast image tag to generate a natural language representation. A detector is used to identify objects in a picture. Recognition of objects is also carried out on images. created a 3D image analysis system to infer objects, features, and connection points in the image and transform them into a number of semantic trees.

The semantic trees of the trees were transformed back into images after learning the syntax to write written descriptions for these trees. The query expansion method proposed by Yagcioglu et al. involves retrieving images associated with from a sizable dataset [4] and using the distribution that has been stated in combination with the recovered images. This is one of several indirect methods for addressing picture description issues that have been proposed. The suggested descriptions are then rearranged by computing the cosine within the separated representation and the expanded query vector once the extended query has been produced using the expression. The input image's description is then taken from the closest description. Prior to the advent of the big data age and the popularity of the methods of deep learning, neural networks' efficacy and widespread use had advanced the area of photo description and opened up new opportunities.

II. Motivation

Generating a natural language description of an image is the goal of the popular research field of image captioning in computer vision and natural language processing. The purpose of image captioning is to give computers the ability to comprehend visual information and produce accurate, meaningful descriptions that resemble those of humans. The process of creating image captions requires a variety of skills, including object detection, scene comprehension, and language modeling. The system must first recognize the items, people, and other aspects in the image before using this knowledge to create a meaningful sentence that precisely depicts the scenario.

There are several possible uses for image captioning, including helping the blind, increasing search results, and improving user experience across a range of applications. For instance, image captioning can be used to create product descriptions on online storefronts, enabling consumers to look for certain objects based on their visual attributes. Additionally, picture captioning is an essential step in the creation of autonomous vehicles, robots, and other AI systems that must comprehend and communicate with their environment. Image captioning can aid in the development of more intelligent and responsive systems that are better able to comprehend and traverse challenging surroundings by giving robots the capacity to accurately describe the visual world.

Nevertheless, despite recent developments in image captioning research, a number of difficulties still need to be resolved, including how to deal with ambiguity in language and visual data, how to handle complex scenes with numerous objects and actions, and how to incorporate contextual information into the captioning procedure. Nevertheless, there are a lot of potential advantages to image captioning [7], and more study in this area will probably result in important developments in AI and other fields.

III. Main Contributions And Objectives

Describe user-visible aspects of the system that are not directly related with the functional behaviour of the system. Non-Functional requirements include quantitative constraints, such as response time (i.e. how fast the system reacts to user commands.) or accuracy (. e. how precise are the systems numerical answers.). The major non-functional Requirements of the system are as follows:

- **Usability**: The system is designed entirely with automated operations hence there is no user interference.

- **Reliability**: The system is more reliable because of the qualities that are inherited from the python platform. The code is built by using python which is more reliable.

- **Performance**: This system is developing in the high level languages and uses advanced front-end and back-end technologies .The response time on the client system is very less.

- **Supportability**: The system is designed is such a way that it is supportable on any platform and also it is supported on a wide range of hardware and software platform,which is having PVM, built in function.

IV. Proposed Framework

A. Proposed Framework

In this paper we have also suggested called Reference based long short term memory (R-LSTM) that main aim is by implementing reference information to give more descriptive caption for a query image. According to relation between image and words during the training phase, different weights are assigned. In addition to maximizing the agreement-score among the captions produced through the captioning methods and the reference data from the adjoining images of the intentional images that can limit the issue of not recognize correctly an image. The image In order to caption natural scenes, image captioning research Kiros et al. implemented this framework which is called encoder and decoder based to combined use image-text embedding model and multi-modal sentence generation models, for a query image, a description which is an output produced word by word similar to language translation. For encoding textual data a special RNN they used called Long Short Term Memory (LSTM) and for encoding visual data a deep Convolutional neural network used. Then, by using optimizing a pair-wise ranking loss, the visual content is encoded then is fed into an embedding space extended by Long Short Term Memory hidden states which encode textual data. In the embedding space, for decoding image features conditioned on feature vector of background word a structure content neural language is utilized that permit to generate sentence word after word. By the exact motivation from neural machine translation, Vinyals et al., first for encoding image as an encoder adopted deep CNN and next for decoding purpose as a decoder he used LSTM in RNN that will help in generating description from image features.

B. Advantages of proposed system

Image captioning is a boon for visually impaired people who are unable to comprehend visuals. With AI-powered image caption generator, image descriptions can be read out to visually impaired, enabling them to get a better sense of their surroundings.

C. Life Cycle Model

In our project we used waterfall model because it consists of 5 steps:

- **Requirements**: In this phase, design, function, purpose are understood, and the requirements are recorded.
- **Design**: The requirement specifications from first phase are studied in this phase and system design is prepared. System Design or Software architecture helps in specifying hardware, technology and system requirements.
- **Execution**: With inputs from system design, the software is split into small units. This is the coding phase where the requirements are converted into units which are integrated to create the completed product.

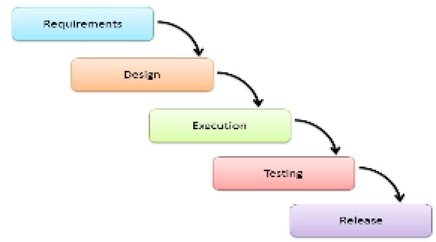

Fig. 1. Life cycle method

- **Testing**: This product goes through various stages of testing to make sure there are no errors and all the requirements are complete. Testing is done so that the client does not face any problem during the installation of the software.
- **Release**: Once testing is done, and the product is found to be complete with no errors, it is deployed in the customer environment or released into the map.

D. DESIGN

1) Architecture: The final model for generating captions for images is a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which takes two inputs - images and corresponding captions. For the RNN layers, one word is inputted into each layer, and the model learns to predict the next word, optimizing itself through the caption data. The image features are obtained from a pre-trained VGG16 model and saved in a file for correlation with the captions. The outputs from both the image features and LSTM layers are then combined and fed into a decoder model to generate the final captions, with the last layer having a size equal to the length of the vocabulary. The categorical cross-entropy method is used for this model to predict the probability of each word, and the Adam optimizer is used to update the weights of the network during optimization.

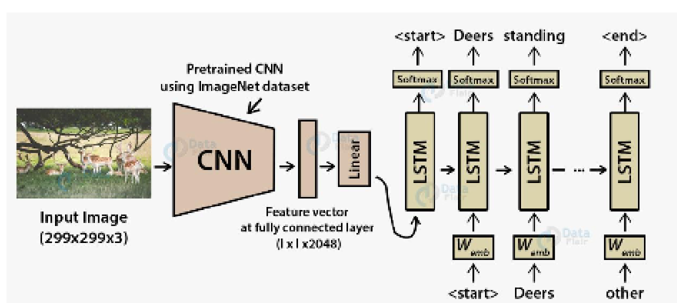

Fig. 2. Model-Image Caption Generator

E. Dataflow Diagram

Data flow diagrams are graphical representations of how data moves through a business information system. They illustrate the processes involved in transferring data from inputs to storage and generating reports. These diagrams can be categorized into two types: logical and physical. The logical data flow diagram shows how data flows through a system to achieve specific business functions. On the other hand, the physical data flow diagram describes the practical implementation of the logical data flow

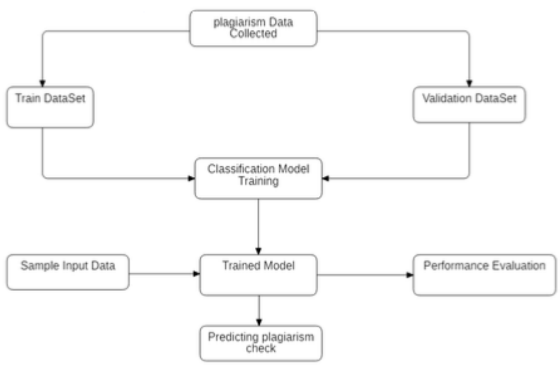

Fig. 3. Data Flow

Design engineering deals with the Unified Modelling Language (UML) which is a standard language for writing software blueprints. The UML is a language for

- Visualizing
- Specifying
- Constructing
- Documenting the artifacts of a software intensive system.

The UML is a language which provides vocabulary and the rules for combining words in that vocabulary for the purpose of communication. A modelling language is a language whose vocabulary and the rules focus on the conceptual and physical representation of a system. Modelling yields an understanding of a system.

V. IMPLEMENTATION

A. Working of Project

1) Dataset: Initially, our model was trained on the Flickr30k dataset, which consisted of 31783 images, each with five captions. However, we encountered difficulties with the model's generalization due to the small number of training samples and the repetition of the "A man..." template in each caption. To address this, we shifted to the larger MSCOCO (2014) training dataset [9], which included 82780 images, each with five ground truth captions. For offline evaluation, we utilized the Karpathy split3, which, although not a standardized split, has been frequently used by researchers to report their results. This split consisted of 5000 images.

2) Syntax Analysis: Syntax identification involves checking whether a language follows its grammatical rules. Parsing, stemming, and lemmatization are some commonly used techniques in this process.

3) Semantic Analysis: In NLP, algorithms are utilized to comprehend the intended meaning. Techniques such as Word Sense Disambiguation, Named Entity Recognition, and Natural Language Generation (NLG) are employed for this purpose.

B. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) are commonly used for visual recognition tasks, with multiple convolutional layers and fully connected layers. CNNs take advantage of the 2D structure of images through local connections, tied weights, and pooling techniques, producing translation invariant features. The benefits of using CNNs include ease of training and fewer parameters compared to other networks with equivalent hidden states. The VGG network, a deep CNN for large scale image recognition, is used in this work with 16 or 19 layers, achieving similar classification error rates for the validation and test sets. The extracted image features are used in the caption generation process. Long Short-Term Memory (LSTM) [10] is a type of recurrent neural network used for modeling the transient dynamics in a sequence. Traditional RNNs can have difficulty learning long-term dynamics due to vanishing or exploding gradients, but LSTM addresses this issue by using a memory cell that stores information for an extended period. Gates control the update timing of the cell state, and the number of connections between the memory cell and gates represent variables.

C. Model deployment

The deployment stage of a model involves putting it into production use, which is usually delegated to a data engineer or database administrator after the data scientist has chosen a reliable model and specified its performance requirements. The deployment workflow may vary depending on the business infrastructure and the problem being solved. The implementation, testing, and maintenance of infrastructure components for proper data collection, storage, and accessibility fall under the responsibility of a data engineer. This includes translating the final model from high-level programming languages to low-level languages that better integrate with the production environment.

To measure the performance of the model, A/B testing is usually conducted by the data engineer. This testing can show how customers engage with a model used for personalized recommendations and how it correlates with a business goal. On the other hand, when dealing with smaller amounts of data, a database administrator is responsible for putting the model into production.

The deployment process also depends on whether the data science team performed the earlier stages manually using inhouse IT infrastructure or automatically with machine learning as a service (MLaaS) products. MLaaS is a cloud platform that provides tools for data preprocessing, model training, testing, and deployment, as well as forecasting. Popular MLaaS include Google Cloud AI, Amazon Machine Learning, and Azure Machine Learning by Microsoft, and they differ in the number of provided ML-related tasks, which depends on the level of automation.

Deployment on MLaaS platforms is automated and results can be bridged with internal or other cloud corporate infrastructures through REST APIs. The way analytical results are received, either in real-time or at set intervals, should also be taken into consideration.

VI. GENERATION OF SENTENCE WITH LSTM

The approach of generating sentences in a neural network is inspired by the encoder-decoder principle used in machine translation models. This involves using an encoder to map variable sequences of words in natural language to distributed vectors, which are then used by a decoder to generate a new sequence of words in the target language. During training, the goal is to optimize the translation process to achieve a high probability of generating a natural sentence in the source language. When generating captions for images, the objective is to maximize the length and quality of the caption produced for a given image.

A. Strategy: matching the problem with the solution

During the initial phase of a machine learning project, the focus is primarily on establishing strategic objectives, which involve identifying a problem, defining the scope of work, and planning the development process. A business analyst is responsible for assessing the viability of a software solution and establishing the necessary requirements, while a solution architect oversees the development process to ensure that these requirements are incorporated into the solution. Essentially, the solution architect's role is to ensure that the requirements set by the business analyst serve as the foundation for the new solution.

B. Dataset preparation and preprocessing

Data serves as the fundamental building block for any machine learning project. The second phase of implementing such a project is intricate and comprises various steps, such as data collection, selection, preprocessing, and transformation. Each of these steps involves a series of procedures to be followed.

C. Data Collection

The role of a data analyst is crucial in the implementation of a machine learning project as they are responsible for identifying relevant sources of data, collecting comprehensive data, interpreting it, and analyzing the results using statistical techniques. The type of data required varies depending on the specific prediction task at hand. The amount of data needed for a machine learning problem cannot be precisely determined since each problem is unique. The selection of attributes used in building a predictive model depends on their predictive value. It is generally advisable to collect as much data as possible during the data collection phase, as the more data available, the better the chances of accurate model training. Additionally, publicly available datasets can be used to complement internal data sources, and there are several platforms, such as Kaggle, Github, and AWS, that provide free datasets for analysis.

D. Data visualization

Presenting a vast amount of information in graphical form makes it more comprehensible and easier to analyze. Therefore, it is essential for a data analyst to possess skills in creating various visual aids such as slides, diagrams, charts, and templates. For instance, a sales-by-year chart can be an effective way to represent sales data visually.

E. Labeling

Supervised machine learning involves training a predictive model on historical data that has predefined target answers. To accomplish this, an algorithm must be provided with information about the target attributes or answers to look for in a dataset, which is referred to as labeling. Labeling datasets can be a time-consuming and challenging task, particularly when thousands of records need to be labeled for the machine learning model to function correctly. For example, if the machine learning algorithm needs to classify different types of bicycles in images, the dataset needs to be clearly defined and labeled accordingly. One way to streamline the labeling process is to engage domain experts to assist in labeling the data.

F. Transfer learning

Transfer learning is an alternative approach to labeling large datasets, where previously labeled training data can be repurposed. This technique involves leveraging knowledge and insights gained from solving similar machine learning problems by other data science [11] teams. A data scientist must identify which elements of the existing training dataset can be used for a new modeling task. Transfer learning is commonly used to train neural networks, which are models utilized for image or speech recognition, image segmentation, and human motion modeling.

G. Data selection

Once all the necessary information has been collected, a data analyst selects a subset of the data that is relevant to the defined problem. For example, if a business is interested in its customers' geographical location, it is unnecessary to include their personal information like cell phone or bank card numbers in the dataset. However, purchase history would be a critical attribute to consider when building a predictive model. This subset of data includes the attributes that must be considered when developing a predictive model. In smaller data science teams, a data scientist may be responsible for data collection, selection, preprocessing, and transformation, as well as model building and evaluation.

H. Data preprocessing

Preprocessing is a crucial step in preparing data for machine learning. The goal is to transform raw data into a structured and clean format that can be used effectively in a machine learning model. This enables data scientists to obtain more accurate results from their models. The process involves various techniques, such as data formatting, cleaning, and sampling.

I. Data formatting

When data is collected from multiple sources and individuals, data formatting becomes increasingly important. Data scientists must begin by standardizing the format of the records. They need to ensure that variables representing each attribute are consistently recorded. This consistency also applies to attributes represented by numeric ranges. By maintaining data consistency, data scientists can ensure the accuracy and reliability of their machine learning models.

J. Data cleaning

In this stage, a data scientist performs several procedures to improve the quality of data by removing irrelevant information and inconsistencies. They can use imputation techniques to fill in missing data and detect outliers - data points that significantly differ from the rest of the distribution. If an outlier indicates incorrect data, the data scientist either removes or corrects them. Additionally, they delete incomplete and irrelevant data objects.

K. Data anonymization

In certain cases, data scientists are required to remove or obscure attributes that contain confidential information, especially when dealing with sensitive data from industries such as healthcare or banking.

L. Data sampling

When dealing with large datasets, data analysis requires more time and computational resources. To address this, a data scientist may use data sampling to choose a smaller, yet representative subset of the data to build and run models. This approach allows for faster and more efficient analysis while still producing accurate results

M. Data transformation

The last step of preprocessing involves transforming and combining data to prepare it for machine learning [12] or data mining. Feature engineering, as it is also known, can involve scaling or normalizing the data, as well as decomposing or aggregating attributes. This process ensures that the data is in a suitable format for analysis and modeling.

N. Scaling

Numeric attributes of data may vary across different ranges, such as millimeters, meters, and kilometers. Scaling refers to the process of converting these attributes to a common scale, typically between 0 and 1 or 1 and 10, to ensure consistency in their magnitude.

O. Decomposition

When dealing with complex concepts represented by features, it may be challenging to discover patterns in data. In such cases, a technique called decomposition can be used. In decomposition, higher-level features are transformed into lower-level ones, and new features based on existing ones are created. This technique is commonly used in time series analysis. For instance, to forecast the demand for air conditioners per month, a market researcher may decompose the data that represents quarterly demand.

P. Aggregation

Aggregation is a technique that combines multiple features into a single feature that represents them all. For example, age information about customers can be aggregated into age categories like 16-20, 21-30, 31-40, and so on for demographic segmentation [13]. This helps in reducing the dataset size without losing valuable information. Preparing and preprocessing data is a gradual and time-consuming process that involves selecting appropriate techniques and iterating as necessary based on the business problem and the quality and quantity of data available.

Q. Dataset splitting

When using a dataset for machine learning, it is recommended to divide it into three separate subsets: training set, test set, and validation set.

R. Training set

A model is trained and optimized by a data scientist using a training set to learn its necessary parameters from the data.

S. Test set

To evaluate the effectiveness of a trained model and its ability to generalize, a data scientist uses a test set, which contains data not previously used for model training. Overfitting occurs when a model is too complex and has memorized the training data instead of learning from it. To avoid this, it is essential to use separate datasets for training and testing.

1) Validation set: The validation set plays a crucial role in adjusting a model's hyperparameters, which are the high-level [14] structural settings that cannot be directly learned from data. These parameters can determine a model's complexity and its ability to identify patterns in data. Typically, the dataset is divided into a training set (80%) and a test set (20%). The training set is further divided, with 20% used to form a validation set. However, some experts recommend a 66% training and 33% testing split. The size of each subset is dependent on the overall size of the dataset.

T. Dataset-splitting

A data scientist can improve a model's potential performance by using a larger amount of training data. Likewise, the use of more testing data for model evaluation can enhance its generalization capability and overall performance.

U. Modeling

At this stage, a data scientist trains multiple models to determine which one yields the most precise predictions.

1) Model training: Once data preprocessing is complete and a dataset has been partitioned into three subsets, a data scientist can begin training various models to determine the one that yields the most precise predictions. Model training involves feeding an algorithm with training data, which the algorithm processes to produce a model capable of finding a target value in new data. This target value is the output of a predictive analysis. The two primary types of model training are supervised and unsupervised learning, which are selected depending on whether the task is to forecast specific attributes or group data objects based on similarities.

V. Supervised learning

Supervised learning is a technique used to process labeled data or data with target attributes. Prior to training [15], these attributes are already mapped in historical data. By utilizing supervised learning, a data scientist can address classification and regression problems.

VII. Results

Fig. 4. Output: A group of giraffes standing next to each other

Fig. 5. Output: A bird sitting on a branch of tree

Fig. 6. Output: A large white bird flying in the sky

Fig. 7. Output: A chair next to a wooden table

References

[1] Graves, A.; Liwicki, M.; Fernández, S.; Bertolami, R.; Bunke, H.; Schmidhuber, J. (May 2009). "A Novel Connectionist System for Unconstrained Handwriting Recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence.

[2] Li, Xiangang; Wu, Xihong (2014-10-15). "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition

[3] Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le, Quoc V.; Norouzi, Mohammad; Macherey, Wolfgang; Krikun, Maxim; Cao, Yuan; Gao, Qin (2016-09-26). "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

[4] Mayer, H.; Gomez, F.; Wierstra, D.; Nagy, I.; Knoll, A.; Schmidhuber, J. (October 2006). A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 543–548.

[5] Rodriguez, Jesus (July 2, 2018). ”The Science Behind OpenAI Five that just Produced One of the Greatest Breakthrough in the History of AI”. Towards Data Science. Archived from the original on 2019-12-26. Retrieved 2019-01-15.

[6] Li, Xiangang; Wu, Xihong (2014-10-15). ”Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition

[7] K. Cho, B. Van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empir-ical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014.

[8] Rashtchian C, Young P, Hodosh M, Hockenmaier J. (2010) Collecting image annotations using Amazon’s Mechanical Turk. In: Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon’s Mechanical Turk. Association for Computational Linguistics.

[9] D. Lin, An information-theoretic definition of similarity, in: Proceedings of the Fifteenth International Conference on Machine Learning

[10] Dewa Made Sri Arsa, I.P.A. Bayupati and Kadek Sastrawan ,” Detection of fake news using deep learning CNN–RNN based methods”

[11] Anjali Samad,Bhagyanidhi and Vaibhav Gautam, ”An Approach for Rainfall Prediction Using LSTM Neural Network.

[12] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate (2014), arXiv preprint arXiv: 1409.0473.

[13] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neu-ral networks, in: Advances in neural information processing systems, 2014, pp. 3104–3112 [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Van- houcke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015

[14] H. Liu and P. Singh. ConceptNet - A practical common-sense reasoning toolkit. BT technology journal, 22(4):211–226, 2004.

[15] Young P, Lai A, Hodosh M, Hockenmaier J (2014) From image descriptions to visual denotations: new similarity metrics for semantic inference over event descriptions. Trans Assoc Comput Linguist 2:67–78