Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

Construction of reliable image captioning system for web camera based traffic analysis on road transport application

R Dhaya
Professor, Department of Computer science and Engineering, King Khalid University, Kingdom of Saudi Arabia
E-mail: dhayavel2005@gmail.com

Abstract

The automated captioning of natural images with appropriate descriptions is an intriguing and complicated task in the field of image processing. On the other hand, Deep learning, which combines computer vision with natural language, has emerged in recent years. Image emphasization is a record file representation that allows a computer to understand the visual information of an image in one or more words. When it comes to connecting high-quality images, the expressive process not only requires the credentials of the primary item and scene but also the ability to analyse the status, physical characteristics, and connections. Many traditional algorithms substitute the image to the front image. The image characteristics are dynamic depending on the ambient condition of natural photographs. Image processing techniques fail to extract several characteristics from the specified image. Nonetheless, four properties from the images are accurately described by using our proposed technique. Based on the various filtering layers in the convolutional neural network (CNN), it is an advantage to extract different characteristics. The caption for the image is based on long short term memory (LSTM), which comes under recurrent neural network. In addition, the precise subtitling is compared to current conventional techniques of image processing and different deep learning models. The proposed method is performing well

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

in natural images and web camera based images for traffic analysis. Besides, the proposed algorithm leverages good accuracy and reliable image captioning.

Keywords: learning model, image processing

1. Introduction

In recent years, the study of computer vision in image processing has made significant advances, with the most prominent recent advancements being image classification and object recognition, respectively. An automated phrase or series of syllables is produced with response to image input [1, 2]. Image captioning is the term used to describe this process. The accuracy of produced image descriptions may have a significant effect on a variety of things, including the titles associated with news images, medical image description, text-based image retrieval, and the information that has been made available to visually impaired individuals, among other things. When it comes to image captioning, this kind of research has become the most important theoretical and practical implications [3]. Artificial intelligence has enhanced the complexity of image captioning, making this task not only more difficult but also more important than it was before. Figure 1 shows the sample natural images for captioning in many applications such as crowd identification and traffic analysis. Figure 1 shows some natural and web camera based images for captioning [4].

Images may be accompanied by a few lines of text, which is referred to as a cutline that explains or expands on the illustrated image. Captions are referred to as "heads" in the media, since they are the words that are read the most frequently in an article. Using picture descriptions has another advantage, in that they are often precise and concise, pointing readers to the most noticeable or important aspects displayed in the pictures [5]. We use implicit knowledge and text around the image to minimize the need for human involvement in the process. To put it another way, the system learns to generate captions from a dataset that has not been formally labelled or

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

categorized. Among the items in a dataset are news articles and the images that were linked to them [6]. A wide range of issues from several fields is covered, including technology, sports, education, politics, and a variety of other topics. Caption generation is often accomplished via the use of a two-stage framework, which consists of two stages: content selection and surface realization. Identification of the subject of the image and accompanying article is known as a content selection, while verbalizing the information is known as surface realization, when it comes to content selection and surface realization are the same thing [7, 8]. Figure 2 shows natural images contain road transport violence message content. The bus is not parking at station or stand. The next image contains that vehicle is moving during green signals in the traffic lights.

Figure 1. sample natural images for captioning in application

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

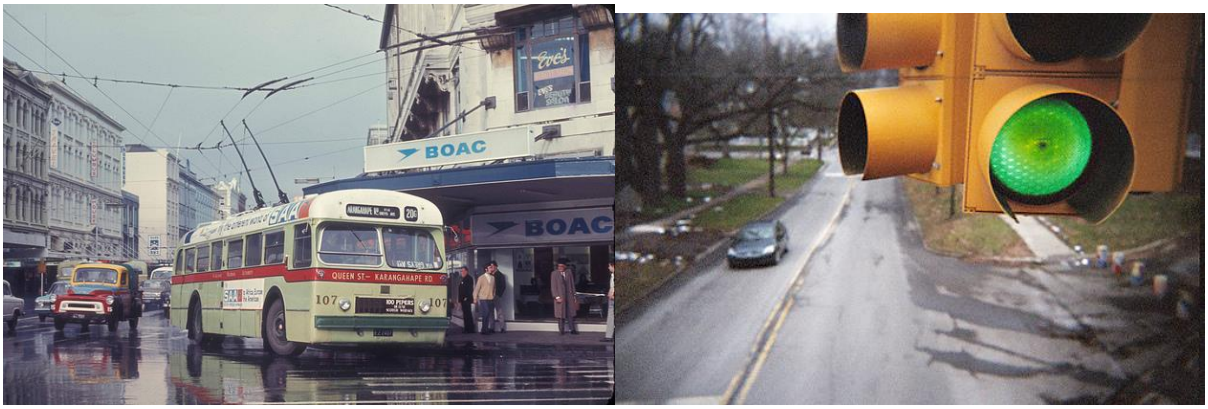

Figure 2. Violence issues on the road transport

The bulk of automated image captioning research is focused on three areas: template-based image captioning, retrieval-based image captioning, and novel image caption generation [9]. When creating image captions, a template-based system first detects the entities (i.e., objects, traits, or activities) present in a picture and then utilizes a present template to fill in the spaces between them. Search-based techniques first find images that are visually similar to one another, and then they choose a comparable image with a caption from among those images [10, 11]. They are capable of creating semantically accurate captions but are unable to generate image-specific as well as syntactically correct captions [12]. A alternative method would be to first select images containing relevant visual information, and then use a language model to produce picture captions from the visual data [13, 14]. When compared to the aforementioned techniques, the novel caption generating process may generate captions for an image that is more accurate (category 1 and 2). Several of the work pieces listed in this category were designed using machine learning and deep learning techniques, which are discussed in this article [15, 16]. There are many comparable frameworks in this field, one of which is the encoder-decoder framework for image captioning, which is an encoder-decoder framework for picture captioning [17].

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

2. Orgnization of the Research

The rest of this research paper is organized as follows: Section 3 discusses the preliminary work of automated picture captioning in depth. Section 4 outlines the planned effort for more accurate automated picture captioning. Section 5 illustrates some of the collected findings with captions and discusses them. Section 6 summarizes our study and discusses potential future tasks in the field of automated picture captioning.

3. Preliminaries

In a template-based approach to programming, predefined templates with slots that may be filled with various types of objects, properties and actions are utilized.

Farhadi et al opted to replace three template slot settings with a triplet of unique scene components in the Image Captioning assignment [18]. They extracted information from the items, characteristics, and connections among them as explained in Li et al. [19]. The technique that Kulkarni et al. used to infer objects, characteristics, and prepositions from a collection of phrases is a conditional random field (CRF). Template-based techniques may be used to generate grammatically accurate captions. Though predefined templates and caption lengths are included, they are not variables. To be visually appealing, the sentence pool or visual database of comparable pictures is searched for semantically relevant words, and those phrases are pasted into the image to make it more attractive [20].

When using the stacked-auxiliary-embedding technique to characterizing millions of poorly annotated pictures, Gong et al considered the application of the method in the context of their research paper [21]. The study conducted by Ordonez et al examined photographs on Flickr and discovered similar photographs; once this information was gathered, they returned with it, alongside descriptions, which was available for you to use in querying images and their associated

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

descriptions from among the hundreds of millions of photographs on Flickr [22]. Sun et al. state that according to semantic and visual similarity, keywords and imagery are clustered, and then caption similarity is used to identify target pictures within clusters of comparable photos [23].

The creation of unique captions employs machine learning and deep learning approaches. In practice, however, the process of acquiring visual information and utilizing language models to create picture captions is a very complex task.

The paper by Vinyals et al. claims that LSTM is used as a decoder, while CNN is used as an encoder to create pictures. While the model has several positive features, such as its high degree of generalization, it also has a few disadvantages, including the fact that it tends to overfit rapidly, necessitating the usage of costly and cumbersome GoogLeNet with 22 hidden layers [24].

Karpathy et al studied the viability of computer software that uses natural language to provide a visual description. They utilize textual descriptions and photographic databases to seek out intermodal connections. The most first step in the writing process is aligning words with the visual regions of the picture. To form a whole sentence, sentence fragments are joined together using different modalities, which insert each piece. A model in a recurrent neural network was trained using training data generated from this description, which in turn was used to train a second model in a recurrent neural network [25].

They utilize a convolution neural network and an LSTM to describe the input image to extract maps. The fact that the picture classification is of poor quality is a failure of this study because the current benchmark of CNN models is the outdated and extremely costly Oxford VGGnet.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

Yu et al are among the few scientists who have studied categorization. They offer a two-sided cross-domain CTSVM method that contains user and item characteristics (CTSIF-SVMs) [26].

However, there is a significant drawback to the image subscription work, which motivated us for further essential research. The measurement used to test and loss for training is different. We utilize cross-entropy as a loss; however, measures are not distinct and cannot be used as a training loss directly. And log-like features may be seen to give every phrase the same weight, whereas, in fact, people regard different words with selective weights.

4. Methodologies

With developments in science and technology, as well as the necessity for human life development, robots have been utilized in an increasing number of fields. Self-driving robots can avoid obstacles, modify routes, and interact with humans automatically based on road conditions. Other tasks such as automatic parking may be done in addition to leverage a safe and efficient driving. In addition, the increase of safety events in the road transport through the traffic lights should be specific and controllable. The machine can tell humans what it observes and then processes according to the input of the machine. We need to rely on the automatic image description generation to perform the tasks [27].

4.1 Proposed CNN

With the assistance of the SVM method, we have built a deep CNN model for acquiring multi-image features for classification frameworks, which we have used in conjunction with the SVM technique. The captioning of the class label images is accomplished via binary cross entropy from the class label images. Initial training is done by using computer-processed images of traffic

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

lights around them in order to learn the multi-image characteristics with the use of different convolutional filters.

The sparse cross entropy is included in the label captioning in order to identify the multiclass pictures. The characteristics from the ImageNet dataset are being adjusted via the built architecture, which is working in conjunction with the SVM classifier, using pre-trained VGG-19 model parameters. Figure 3 shows the block diagram of proposed architecture with LSTM [28].

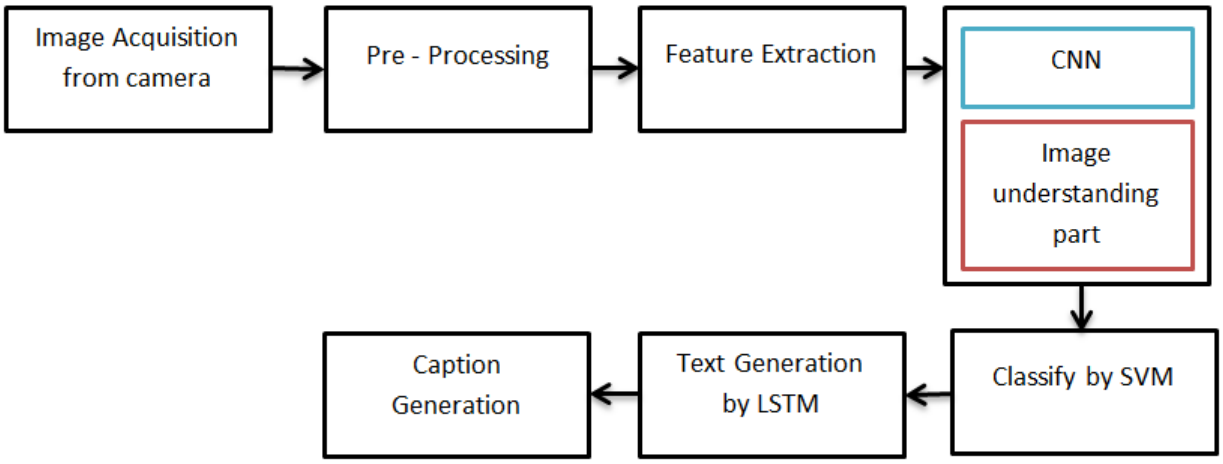

Figure 3. Block diagram of proposed architecture

4.2 Pre-processing

Among the components of the proposed algorithm framework are data preprocessing, which is a critical step in cleaning up the data for input pictures in order to extract the different characteristics from the images for captioning purposes. The preprocessing portion may be used to modify the properties and features of the picture data that has been entered. The essential issue is that captions are correctly predicted throughout training time, and that the target variable anticipated in the output model is accurately predicted during training time. As part of the dataset's

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

array design, we have already included certain encode words to aid with rapid prediction. The output captioning information is tallied word by word, sentence by sentence. To support our suggested algorithms, which are titled Picture to word and word to image, we have built our own word Python dictionaries. It is saying that a separate piece of work is offered in every aspect of the pictures as a number index, which will be used to give captioning for the corresponding input image [29].

4.3 Training

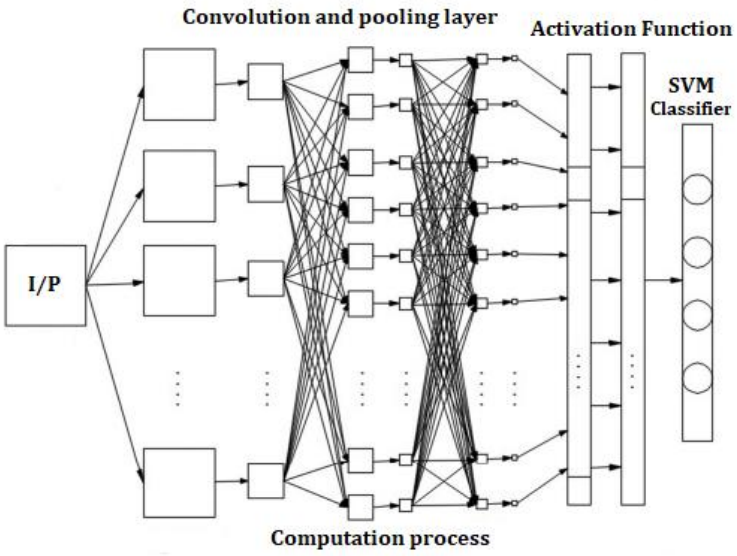

Figure 4. Proposed architecture

This research work has trained a neural network using CNN along with a variety of filters for different image characteristics present in the input photos that were provided. To facilitate training and testing, the input datasets are divided into two parts: training and testing. The provided input images are trained by using a computer process to learn the image characteristics and the learned image features are then utilized to create multi label images from the given input images.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

By using intelligent extraction, the image characteristics are extracted clearly and compared to one another for the purpose of identifying the picture being analyzed. Figure 4 shows our proposed framework architecture for classification of the given input image. The prompt library is included in the picture captioning to ensure proper captioning. As a result, the planned work is divided into two phases, which are as follows:

1. Image classification via the use of a classifier
2. Captioning images via data processing

4.4 The process of creating image captions

Since the advent of deep learning and machine learning, the image captioning method has become more sophisticated, requiring a greater number of particular characteristics from the input pictures. The text creation part is based on LSTM model [28, 29]. In practice, the written description of a group of phrases will be given for input pictures in order to provide a name to the group of sentences in English, as described above. The description of a picture in this suggested framework is currently limited to the English language alone at the time of writing. When an input picture has certain characteristics that are marked, the textual description of English for that image is carried out automatically and in an efficient manner.

5. Results & Discussion

We employed the image model and were able to construct extremely equivalent subtitles when compared to human-generated subtitles. When employing the proposed CNN and pre-trained VGG net models, all possible items in the image have an equal probability of being recognized and correctly identified [29]. Figure 5 demonstrates how our proposed system effectively captures

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

and classifies vehicle material that violates traffic rules. The picture is receiving high-quality captioning that is very correct and trustworthy in additional traffic road photos.

Figure 5. Sample results obtained for traffic pictures by proposed work

Here, CNN has been utilized in conjunction with an SVM classifier to optimize the learning process by progressively reducing the amount of learning that converges faster as the learning process proceeds further. To accomplish our objectives, we coupled an SVM classifier to classify for improved optimization with regulatory methods like as L_2 and discontinuation. When it comes to filters, several of them compare the technique of including convolution layers with various parts. Due to the unchanging information in the pictures, the other algorithms are very successful when it comes to natural image description [30]. However, our suggested method works well with web camera-based traffic analysis pictures. Additionally, the picture captioning is extremely helpful and provides a more comprehensive explanation than other current algorithms. Figure 6 illustrates some of the testing pictures in our dataset, denoted by the numbers image 1, image 2, and image 3.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

Figure 6. Test images (1, 2 &3)

A questionnaire includes 20% of images for testing and 80% of generated descriptions for training the dataset for new models, with the remaining images utilised for research. We will conduct further testing for image captioning using a benchmark dataset that receives less attention. Table 1 summarises the obtained results and includes an accuracy calculation.

Table 1. Obtained image captioning

Methods		Text Generation			Caption	Overall Accuracy
		Feature – 1	Feature - 2	Feature - 3		
Image1	RNN	Person	Jump with basket ball	shooting	The person shoots the basketball	85%
	LSTM	Sportsman	Basket ball	playing	Sportsman is playing basketball	88%
	Proposed CNN	Person alone	Basketball	Playing in ground	The person alone is playing basketball in the ground	93%
Image2	RNN	Child guitar	Small	playing	Small child is playing guitar	80%
	LSTM	Child guitar	Sitting	playing	Child is playing guitar at sitting	78%
	Proposed CNN	Boy holding guitar	Seashore	Playing	The boy is holding and playing a guitar at seashore	95%
Image3	RNN	Traffic red	Car	Man standing	The car, man standing at near traffic red	75%
	LSTM	On the road	Car, traffic red	Man standing	Car and man standing on the road nearly traffic red	74%
	Proposed CNN	Traffic red	Single car	Moving	Single car is moving in traffic red	91%

Observers are asked to evaluate whether or not the description generated automatically properly describes the image's content. When the proposed framework is used, correct results are achieved, showing that 91% of subtitles are created properly with moderate care for the model,

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

while 95% are generated properly with soft care for the network-trained model. Figure 7 demonstrates the overall accuracy of an image captioning.

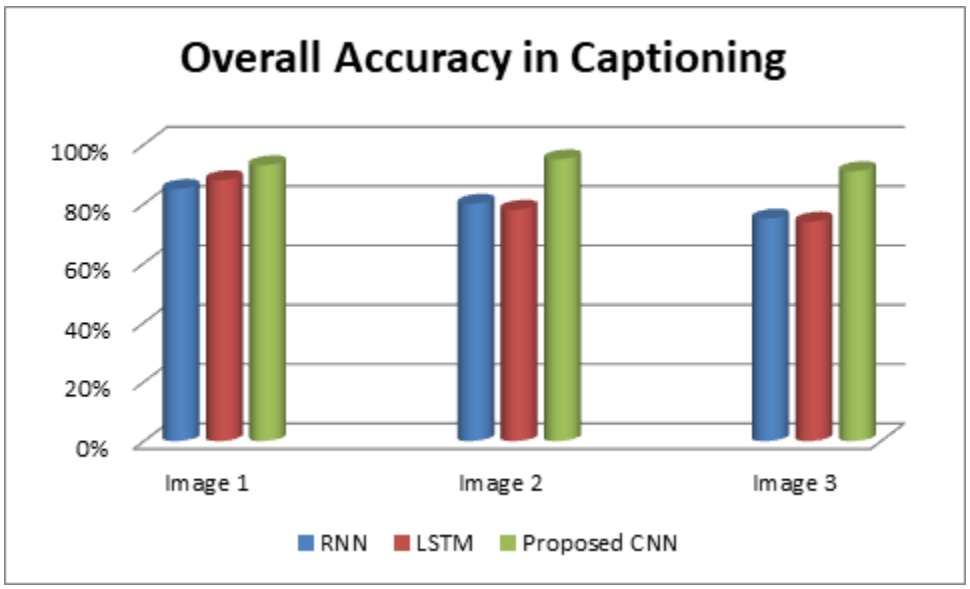

Figure 7. Overall Accuracy for Image Captioning

6. Conclusion

Thus, the suggested CNN demonstrates accurate captioning in a variety of traffic light-based pictures captured through web camera. It outperforms other algorithms in terms of accuracy, when captioning the images taken by using a web camera. Additionally, the suggested system produces more accurate and relevant captions for static natural images than previous algorithms. It established and quantified the accuracy by using the benchmark dataset. The obtained findings show that, although the overall loss reduces rapidly at the beginning of training, the rate of loss changes slightly afterwards. Additionally, the image annotation results and subject model are included into the language model along with a suitable title for the news images. Furthermore, the researchers want to enhance their performance by adding word vectors into a broader range of data

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

sources, such as news articles and other online sources. However, other configurations may be trained to improve the model's capacity to caption images.

References

[1] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." International Conference on Neural Information Processing Systems Curran Associates Inc. 1097-1105. (2012)

[2] Sungheetha, Akey, and Rajesh Sharma. "A Comparative Machine Learning Study on IT Sector Edge Nearer to Working From Home (WFH) Contract Category for Improving Productivity." Journal of Artificial Intelligence 2, no. 04 (2020): 217-225.

[3] Salunke, Vipul, and Suja Sreejith Panicker. "Image sentiment analysis using deep learning." In Inventive Communication and Computational Technologies, pp. 143-153. Springer, Singapore, 2021.

[4] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A Systematic Approach." Journal of Soft Computing Paradigm (JSCP) 2, no. 04 (2020): 226-235.

[5] Girshick, Ross, et al. "Region-based Convolutional Networks for Accurate Object Detection and Segmentation." IEEE Transactions on Pattern Analysis & Machine Intelligence 38.1:142-158. (2015)

[6] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation–A Systematic Approach." Journal of Artificial Intelligence 3, no. 01 (2021): 1-16.

[7] Mistry, Mihir, Ameya Apte, Varad Ghodake, and S. B. Mane. "Machine Learning Based User Interface Generation." In International Conference on Intelligent Computing, Information and Control Systems, pp. 453-460. Springer, Cham, 2019.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

[8] Adam, Edriss Eisa Babikir, and A. Sathesh. "Construction of Accurate Crack Identification on Concrete Structure using Hybrid Deep Learning Approach." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 85-99.

[9] Devlin, Jacob, et al. "Language Models for Image Captioning: The Quirks and What Works." Computer Science (2015)

[10] Kottilingam, Dr. "A DYNAMIC ROUTING MODEL FOR HYBRID ELECTRIC VEHICLES." Journal of Electrical Engineering and Automation 1, no. 1: 50-57.

[11] Lekshmy, V. Geetha, R. Athira Krishnan, and S. Aparnna. "Role of Non-textual Contents and Citations in Plagiarism Detection." In Proceedings of International Conference on Intelligent Computing, Information and Control Systems, pp. 601-611. Springer, Singapore, 2021.

[12] Adam, Edriss Eisa Babikir. "Evaluation of Fingerprint Liveness Detection by Machine Learning Approach-A Systematic View." Journal of ISMAC 3, no. 01 (2021): 16-30.

[13] Fang, H., et al. "From captions to visual concepts and back." Computer Vision and Pattern Recognition IEEE, 1473-1482. (2015)

[14] Kumar, Pranay, and S. Revathy. "An Automated Invoice Handling Method Using OCR." In Data Intelligence and Cognitive Informatics, pp. 243-254. Springer, Singapore, 2021.

[15] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 1-9.

[16] Cho, Kyunghyun, et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." Computer Science (2014)

[17] Bhile, Amey Arvind, and Varsha Hole. "Real-Time Environment Description Application for Visually Challenged People." In International Conference on Computer Networks and Inventive Communication Technologies, pp. 326-332. Springer, Cham, 2019.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

[18] Farhadi, M. Hejrati, M. A. Sadeghi et al., "Every picture tells a story: generating sentences from images," in Computer Vision – ECCV 2010, K. Daniilidis, P. Maragos, and N. Paragios, Eds., pp. 15–29, Springer, 2010.

[19] S.M. Li, G. Kulkarni, T. L. Berg, A. C. Berg, and Y. J. Choi, "Composing simple image descriptions using web-scale n-grams," in Proceedings of the Fifteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, pp. 220–228, Portland, Oregon, USA, 2011.

[20] G. Kulkarni, V. Premraj, S. Dhar et al., "Baby talk: understanding and generating image descriptions," in CVPR means IEEE Conference on Computer Vision and Pattern Recognition, pp. 2891–2903, 2011.

[21] Y. Gong, L. Wang, M. Hodosh, J. Hockenmaier, and S. Lazebnik, "Improving image sentence embeddings using large weakly annotated photo collections," in European Conference on Computer Vision, pp. 529–545, Springer, 2014.

[22] V. Ordonez, G. Kulkarni, and T. L. Berg, "Im2Text: Describing images using 1 million captioned photographs," Advances in Neural Information Processing Systems, pp. 1143–1151, 2011.

[23] C. Sun, C. Gan, and R. Nevatia, "Automatic concept discovery from parallel text and visual corpora," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2596–2604, Santiago, Chile, 2015.

[24] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: a neural image caption generator," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156–3164, Boston, MA, USA, 2015.

[25] Karpathy and L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, Stanford University, 2017.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021

Journal of Trends in Computer Science and Smart technology (TCSST) (2021)
Vol.03/ No. 02
Pages: 114-130
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2021.2.004

[26] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, “SVMs Classification based two-side cross domain Collaborative Filtering by inferring intrinsic user and item features,” Knowledge-Based Systems, vol. 141, pp. 80–91, 2018.

[27] Smys, S., and Wang Haoxiang. "Naïve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." Journal of ISMAC 3, no. 01 (2021): 40-49.

[28] Hochreiter, Sepp, and J. Schmidhuber. "Long Short-TermMemory."Neural Computation 9.8: 1735-1780. (1997)

[29] Tripathi, Milan. "Analysis of Convolutional Neural Network based Image Classification Techniques." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 100-117.

[30] Manoharan, J. Samuel. "A Novel User Layer Cloud Security Model based on Chaotic Arnold Transformation using Fingerprint Biometric Traits." Journal of Innovative Image Processing (JIIP) 3, no. 01 (2021): 36-51.

Author's Biography

R Dhaya is currently a Professor, in the Department of Computer science and Engineering at King Khalid University, in the Kingdom of Saudi Arabia. His major area of research includes Image and Video Processing Algorithms, Computer Vision, Motion Analysis, Stereo Vision, Object Recognition, computer graphics, photo interpretation, image retrieval, Embedded Image Processing and Real-time image and video processing applications.

ISSN: 2582-4104
Submitted: 18.05.2021
Revised: 21.06.2021
Accepted: 10.07.2021
Published: 17.07.2021