DIGITAL WATERMARKING & CAPTION GENERATOR USING DEEP LEARNING

Dr. Archana Dehankar
Guide, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
archana.dehankar@pcenagpur.edu.in

[2]Tulsi Mundada
UG Student,Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
[2]tulsimundada@gmail.com

[3]Sakshi Kubde
UG Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
sakshikubde98@gmail.com

[4]Pranali Kadukar
UG Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
kadukarpranali@gmail.com

[5] Dnyananda Ittadwar
[5]UG Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
dnyanadaittadwar@gmail.com

[6] Khushi Sahu
[5]UG Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
khushisahu416@gmail.com

***Abstract* -Images represent the memories of vision that significantly affect the human encephalon, allowing us to remember specific details about a location, a particular person, or an object we instantly record. A detailed description of each image is required to get a clear idea of what the picture actually consists of because some of the images cannot be recognized. Deep learning and machine vision are used to comprehend the context of a picture and add the appropriate captions to it. It entails categorizing a photograph with English keywords using datasets made accessible during model training. Using the imagenet dataset, the CNN classifier Xception also is trained. Xception handles the retrieval of image features. These extracted characteristics will change the LSTM model to produce the caption for the image. When using machine learning-based methodologies, applications that automatically attempt to present captions or descriptions regarding pictorial and clip frames have a lot of potential. The caption of both images and videos is regarded as a clever issue in imaging science. General-purpose robot vision systems, automatically developing captions for pictures and videos for people with varying degrees of visual impairment, and many other application fields are among the application fields. Another feature of this project is to watermark the image digitally and provide the secured content to the user and the original image can be extracted whenever needed with a security key provided by the host or admin.**

***Keywords* - Image captioning, video captioning, Machine Learning, LSTM, neural network, image processing, Digital Watermarking, Encryption, Decryption.**

I. INTRODUCTION

Science and business have both benefited greatly from image analysis, and this trend will continue. It has been used in a variety of contexts, including scene interpretation and perception of sights, to mention a few. A large number of researchers depended on imaging methods that worked well on rigid components in controlled environments using specialized equipment before the introduction of deep learning. Convolutional neural networks powered by deep learning have recently had a beneficial and significant impact on picture captioning, enabling significantly more versatility. In this article, we discuss new developments in the field of deep learning-based image and video labeling. Convolutional neural networks powered by deep learning have recently had a significant and beneficial impact on the field of picture captioning, offering significantly more freedom. In relation to the topic of deep learning, we have chosen to highlight current developments in the area of image and video annotation in this proposed paper.

II. LITERATURE SURVEY

A. Verma, H. Saxena, et al [1], this research undertakes the task of caption development with an LSTM and RNN based establishes an approach that relies on the same to produce effective and relevant captions by correctly training the dataset. Our model was effectively trained using the Flicker8k dataset. The model's precision is evaluated using standard evaluation measures.

V. Agrawal, S. Dhekane, et al [2], using a variety of techniques, involving DL, CV, and NLP, among others, the task is to produce concise captions. The system that produces the captions in this research makes use of an encoder as well as a decoder, along with an attention technique. It uses an RNN called GRU to supply the proper caption after first extracting the characteristics of the photograph using a CNN with prior training called Inception V3. The proposed model generates captions using a well-positioned attention algorithm. The MS-COCO dataset is used to train the algorithm. The results show that the model is capable of producing text and fairly understanding images.

C. Amritkar and V. Jabade et al [3], the model of regenerating neurons is created. It is reliant on computer imagery and machine translation. This method results in organic phrases that ultimately describe the image. RNN and CNN are additional elements of this strategy. The RNN is used to create sentences, while the CNN algorithm is employed to extract characteristics from images. When provided an input image, the simulation has been trained to generate titles that almost verbatim describe the image. On different datasets, the model's accuracy as well as its fluency or comprehension of the language it picks up from its visual representations are evaluated.

E. Mulyanto et al [4], this study is crucial because there isn't an Indonesian corpus for picture captioning. This research will contrast the experimental results in the FEEH-ID dataset with datasets in English, Chinese, and Japanese using the CNN and LSTM models. The performance of the suggested model in the test set indicates promise with scores of 60.0, down for BLEU-1 and 28.9 for BLEU-3, which are higher than normal for Bleu assessment results in other language datasets. The algorithm used to combine CNN and LSTM.

L. Abisha Anto Ignatius et al [5], the objects that have been recognised are given names using the semantic tags that are present in the image. The captions' ability to describe the objects more accurately is improved by including these factors contextual labels. The Sequence-to-Sequence language paradigm creates the captions one word at a time. The face identification algorithm uses the faces dataset, which includes the facial photos of 232 celebs, to find and identify celebrities' faces in pictures. The mentions of the people in the sentence were changed to their names to make personalized captions. Correlation and the Bilingual Evaluation Understudy measures were created to gauge the precision of the captions that were generated.

M. P. R, M. Anu et al [6], Using image descriptions is the best option for people whose work happen to be blind or who experience difficulty understanding visuals. If an individual's eyesight cannot be corrected, descriptive words can be generated as speech output when using a correlation-based picture caption generator. Today, the study of image processing will become more and more important, mostly for the sake of preserving lives.

S. Li and L. Huang et al [7], in the current encoding and decoding structure, the attention process is frequently used. The current image caption models, built on CNN and RNN, have issues like gradient explosion and are not very good at extracting important information from images. To solve these issues, the research suggests a procedure for creating context-based image captions. The procedure begins with labeling with SCST and LSTM, then progresses to feature extraction with SCST and scenario coding. The outcomes of the experiments show how successful the suggested approach is.

T-Y LIN et al [8], the information was thoroughly statistically analysed by the authors, who then contrasted it with PASCAL, SUNI, and ImageNet. Then, using a Deformable Parts Model, we show early functional testing for classifying identification and bounding box findings. The collection's images included 96 different object types that a 5-year-old could readily identify. Our dataset was developed using specialised software tools for subcategory recognition, case spotting, and instance segmentation, at a total of some million annotated occurrences in 3580 photographs.

A. Karpathy et al [9], the authors present an approach that can produce summaries of graphics and their areas in natural language. Our method uses records of images that have been explained in sentences to find the cross-modal connections between linguistic and visual data. Our alignment approach is based on CNN over image regions, concurrent ML algorithms throughout facial expressions and an organizational objective that coordinates the two categories using multimodal embedding.

P J TANG et al [10], the extra layers enhance and reserve the LSTM model. A weighted average method is used to combine the final forecast probability for each of the Soft maximum functions that are given the correct categorization layers during the test. Experimental results Flickr30K, MSCOCO, and collections show that our model is efficient and outperforms other comparable approaches on a number of assessment measures.

III. METHODOLOGY

In order to watermark a digital image, watermark information must first be embedded into a multimedia product. This information must then be

recovered from or recognized by the watermark in the information product. These methods guarantee image insertion, content validation, authentication, and tamper resistance.

The entire endeavor is broken up into two modules, the first of which is for creating a digital stamp and the second of which is for creating captions described below briefly. Here, the word "digital watermark" refers to the process of using digital tools to embed a watermark into an input image and hide it during display. The watermark can serve as a visual representation of the content's uniqueness.

Module 1 - Digital Watermarking:

The method of watermarking involves encrypting or digitally combining the content. An electronic watermark is a type of marker that is subtly incorporated into a signal that can tolerate noise, such as audio, video, or picture data. It is usually used to establish who owns the copyright to a particular signal.

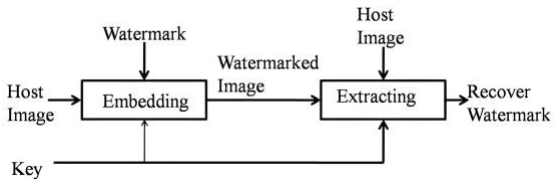

Figure 3.1 Digital Watermark Process Flow

According to the flow diagram provided, the picture is first received from the browser, sent for embedding, and then it appears as watermarked content from which the original image can be extracted using the security key. The ultimate image is shown once the aforementioned steps have been completed.

Module 2 - Caption Generator:

Compared to picture classification and object recognition, the job of automatically creating captions and detailing the image is noticeably more difficult. The narrative of a photo must include not only the items in the image but also the relationships between those items and the activities and characteristics that are depicted in the image [21].

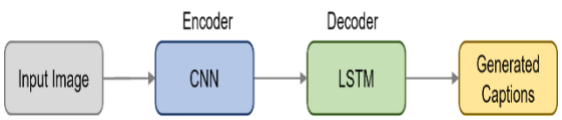

Figure 3.2 Caption Generation Process Flow

The foundation for common watermark techniques, which are broken down into procedures for incorporating and removing watermarks, is briefly revised in the first part of this study. The usual design criteria for judging the potency of watermarking systems are listed in the ensuing subsections. Because there are so many descriptions of similar apps, watermarking systems is now a very specialized area of study. Following that, a list of digital image watermarking techniques is provided according to their functional area.

The results of the research are then summarized in tabular form using the state-of-the-art methods described earlier. The foundation for common watermarking that have techniques, which are broken down into procedures for incorporating and removing watermarks, is briefly revised in the first part of this study. The usual design criteria for judging the potency of watermarking systems are listed in the ensuing subsections. Because there are so many descriptions of similar apps, watermarking systems is now a very specialised area of study. Following that, a list of digital image watermarking techniques is provided according to their functional area. The results of the research are then summarised in tabular form using the state-of-the-art methods described earlier.

The embedding and extraction steps of the digital image watermarking process are for an encrypted communications paradigm. The data embedding portion pre-processes the cover image before evaluating its entropy to determine what image's integrating ability data. The encoder then uses a secret key and an autofocus encoding method to embed a watermark image within the highly entropy-rich host image.The system then gathers data on a laser beam's phase and shape before creating the watermarked picture. Before beginning the watermark extraction procedure, the watermarked image must first undergo some pre-processing. The device then gathers data on the amplitudes and phases that shape the light source's beam patterns. The entropy of these radiation patterns is then calculated.

A large entropy value is selected during the watermark extraction to provide greater robustness and imperceptibility. The watermark picture is separated from the image with the watermark using the same key. The innovation demonstrates how straightforward, trustworthy, and undetectable it is to generate images that are input from the original image.

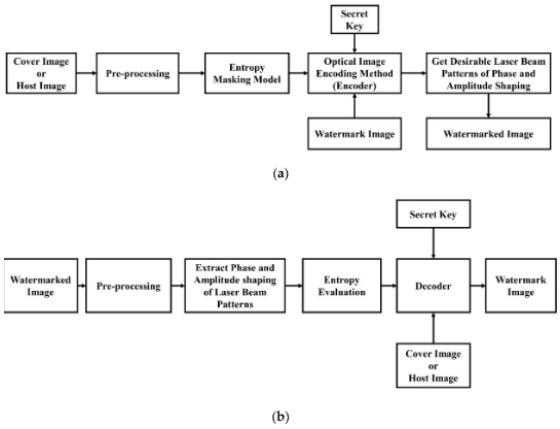

Fig. 3.3 (**a**) Watermark embedding, (**b**) watermark extraction.

As depicted in Fig. 3.3, the model is based on the Long short-term memory block, which is dependent on the LSTM without a peephole architecture. The LSTM's packet and gates are related in the ways listed below:

$$i_l = \sigma(W_{ix}x_l + W_{im}m_{l-1}) \quad (1)$$

$$f_l = \sigma(W_{fx}x_l + W_{fm}m_{l-1}) \quad (2)$$

$$o_l = \sigma(W_{ox}x_l + W_{om}m_{l-1}) \quad (3)$$

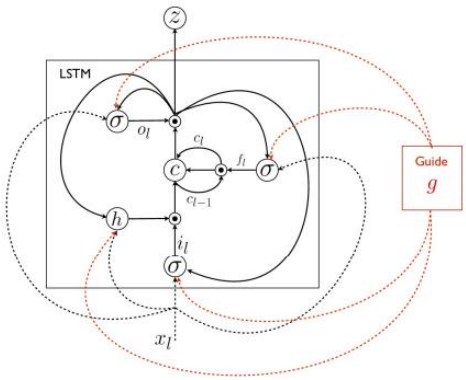

Fig. 3.4 LSTM Connection Diagram

Where the element-wise multiplication, hyperbolic tangent function, and sigmoid function are all noted. In the LSTM cell, the values il stand for the gate used for input, fl for the forgotten gate, ol for the gate that provides the output, cl for the current status of the memory cell in the unit, and ml to feed the hidden state, which also happens to be the result of the block processed in the LSTM.

The parameter values of the series at time step l is represented by the variable. The term "model parameters" refers to the quantity gives the loss function, where St is the generated sentence at time t. This loss is uniformly minimized for all LSTM and embedding of words settings.

IV. RESULT

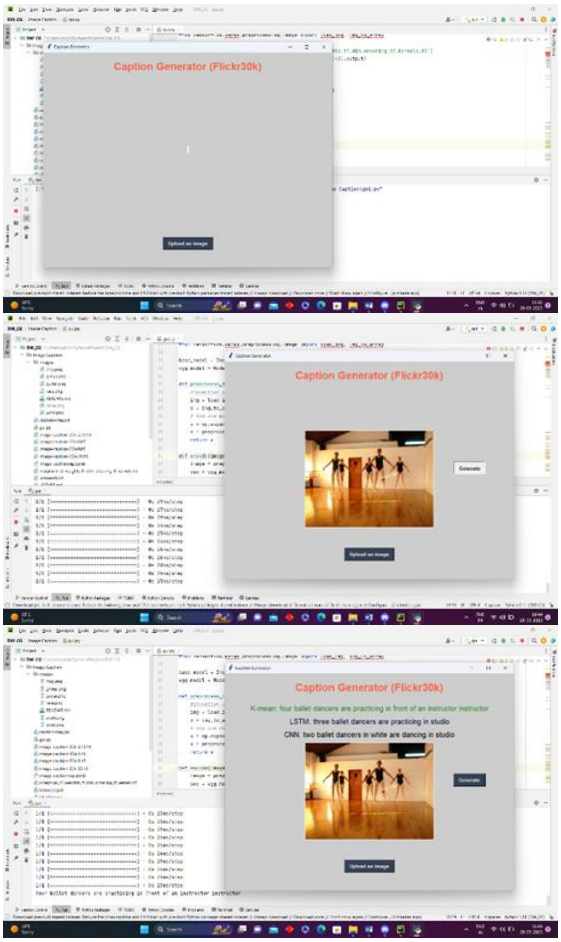

Fig. 4.1 Module 1 Result (Generating Captions for input images

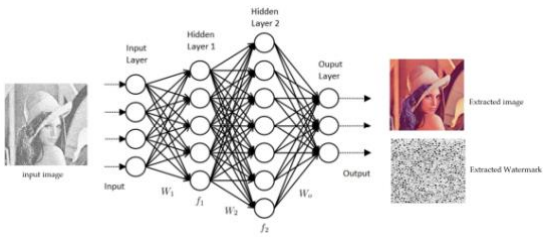

Fig. 4.2 Module 2 Result (Digital Watermarking)

V. CONCLUSION

We have overcome previous restrictions in the field of image captioning by developing a model based on LSTM-based CNN capable of screening and extracting information from any given image and translating it to an interrelationship-line phrase focused in the natural language of English. We are pleased to have avoided overfitting the data, even though it is recognized that this can be difficult to do. The majority of the focus was on the algorithmic foundation of different attention techniques. By

doing so, we can say that we succeeded in developing a model that is a greatly improved version of every other picture caption generator previously made available. And also with the fresh concept of digital watermarking technique to keep the user data safe and secure.

VI. FUTURE SCOPE

As depicted in the image captioning result of our project, we are able to implant a camera in the shoe's front face in future to capture real-time environment video and obtain an audio conversion of that generated caption a means to connect it wirelessly to the blind person's Bluetooth in-ear. This can help the blind person to stop or change their path or seek help of others to get out of the situation easily. The only difference now that this Arduino equipment is being used is that the annotations will be generated in a dynamic environment and made to be played on the blind person's Bluetooth device so that he can go out with more caution. This will undoubtedly reduce accidents and mishaps specifically involving blind people.

IV. REFERENCES

1. A. Verma, H. Saxena, M. Jaiswal and P. Tanwar, "Intelligence Embedded Image Caption Generator using LSTM based RNN Model," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 963-967, doi: 10.1109/ICCES51350.2021.9489253.

2. V. Agrawal, S. Dhekane, N. Tuniya and V. Vyas, "Image Caption Generator Using Attention Mechanism," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9579967.

3. C. Amritkar and V. Jabade, "Image Caption Generation Using Deep Learning Technique," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697360.

4. E. Mulyanto, E. I. Setiawan, E. M. Yuniarno and M. H. Purnomo, "Automatic Indonesian Image Caption Generation using CNN-LSTM Model and FEEH-ID Dataset," 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Tianjin, China, 2019, pp. 1-5, doi: 10.1109/CIVEMSA45640.2019.9071632.

5. L. Abisha Anto Ignatious., S. Jeevitha., M. Madhurambigai. and M. Hemalatha., "A Semantic Driven CNN – LSTM Architecture for Personalised Image Caption Generation," 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 2019, pp. 356-362, doi: 10.1109/ICoAC48765.2019.246867.

6. M. P. R, M. Anu and D. S, "Building A Voice Based Image Caption Generator with Deep Learning," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 943-948, doi: 10.1109/ICICCS51141.2021.9432091.

7. S. Li and L. Huang, "Context-based Image Caption using Deep Learning," 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2021, pp. 820-823, doi: 10.1109/ICSP51882.2021.9408871.

8. T-y lin, m maire, s belongie et al., "Microsoft COCO:Common Objects in Context", Proceedings of the 2014 Euro-pean Conference on Computer Vision, pp. 740-755, 2014.

9. A Karpathy and F-F. Li, "Deep visual-semantic alignments for gen-erating image descriptions", Proceedings of the 2015 Interna-tional Conference on Computer Vision and Pattern Recognition, pp. 3128-3137, 2015. [Google Scholar]

10. P J Tang, H L Wang and K S Xu, "Multi-objective layer-wise optimization and multi-level probability fusion for image description generation using LSTM", Acta Automatica Sinica, vol. 44, no. 7, pp. 1237-1249, 2018. [Google Scholar]

11. Suma, V. "A Novel Information retrieval system for distributed cloud using Hybrid Deep Fuzzy Hashing Algorithm." JITDW 2, no. 03 (2020): 151-160.

12. Manoharan, Samuel. "Supervised Learning for Microclimatic parameter Estimation in a Greenhouse environment for productive Agronomics." Journal of Artificial Intelligence 2, no. 03 (2020): 170-176.

13. Tanwar Poonam & Rai Priyanka," A proposed system for opinion mining using machine learning, NLP and classifiers", IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 9, No. 4, December 2020, pp.

726~733 ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i4.pp726- 733.

14. H. R. Tavakoli, R. Shetty, B. Ali, and J. Laaksonen, “Paying attention to descriptions generated by image captioning models,” in Proceedings of the IEEE Conference on International Conference on Computer Vision, pp. 2506– 2515, Venice, Italy, October 2017.

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

16. A. Shankar and A. Kannammal, "A Hybrid Of Watermark Scheme With Encryption To Improve Security Of Medical Images," *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, 2021, pp. 226-233, doi: 10.1109/ICICV50876.2021.9388616

17. R. A. Hussain, M. E. Abdulmunem and A. M. J. Abdul-Hossen, "Propose Image Encryption Watermarking Algorithm Based on Frequency and Geometric Transform," *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, Baghdad, Iraq, 2019, pp. 143-147, doi: 10.1109/SCCS.2019.8852591.

18. Z. Yang, G. Liqun and R. Ping, "A second-generation wavelet transform digital watermarking encryption algorithm," *2008 Chinese Control and Decision Conference*, Yantai, Shandong, 2008, pp. 207-210, doi: 10.1109/CCDC.2008.4597300.

19. M. T. Mathew and P. R. Geetharanjin, "A Reversible Watermarking and Encryption for ensuring the Authenticity, Integrity and Protection of Medical Images," *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2020, pp. 340-345, doi: 10.1109/ICCCA49541.2020.9250722.

20. K. -x. Yin, L. -j. Zhang, Z. -s. Wang, X. -l. Fu and L. Chang, "Digital Watermarking Algorithm Based on Chaotic Encryption and HVS," *2009 International Conference on Multimedia Information Networking and Security*, Wuhan, China, 2009, pp. 134-137, doi: 10.1109/MINES.2009.202.

21. C. Amritkar and V. Jabade, "Image Caption Generation Using Deep Learning Technique," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697360.