International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Image And Video Caption Generation using Machine Learning

Prof. Kamble. D. R[1], Anpat Vijay Vishnu[2] , Chavan Bhagyashri Nanaso [3] , Chavan Rushikesh Nitin[4] , Kshirsagar Nagesh Bharat[5]

1,2,3,4 ,5 SB Patil college of Engineering, Indapur ,Computer Engineering

---

Abstract - Image Grounded web straggler is the way toward looking through data by exercising affiliated images. The tremendous means of images are accessible on the web in that a large number of the images are contain as with named and without named caption. Our task is to induce an automatic caption for the images grounded on the image content. To produce an image caption, originally, the content of the image should be completely understood; and also, the semantic information contained in the image should be described using a expression or statement that conforms to certain grammatical rules. Therefore, it requires ways from both computer vision and natural language processing to connect the two different media forms together, which is largely gruelling. The paper targets producing mechanized eulogies by learning the contents of the image. At present images are clarified with mortal supplication, and it turns out to be nearly unbelievable task for tremendous databases. The picture information base is given as donation to a deep neural network Convolutional Neural Network encoder for creating caption which extricates the highlights and craft out of our image and intermittent Neural Network decoder is employed to interpret the.

***Key Words*:** Deep Learning, part of speech, image captioning, multi-task learning

1. INTRODUCTION

To grease inquiries in areas similar across-modal reclamation and the backing of visually disabled people image captioning which aims to link image with language has come a hot exploration content. An image captioning model needs to not only fete the salient objects, their attributes, and object connections in an image, but also organize these types of information into a syntactically and semantically correct judgment. With the advances of Neural Machine restatement, recent captioning models generally borrow the encoder decoder frame to "restate" an image into a judgment, and promising results have been achieved. In recent times, experimenters have made significant advances in some areas of computer vision understanding, similar as image bracket, point bracket, object discovery and recognition, scene recognition, action recognition, etc. still, having a computer automatically induce natural language descriptions for an image remains a delicate and grueling task. This task connects the two relatively different media forms, taking that computers not only have a correct and comprehensive understanding of the visual content in the image, but also use mortal language to combine and organize the semantics of the image. The subtasks of image captioning, i.e., relating semantic rudiments similar as visual objects, object attributes, scenes, are innately grueling, and organizing words and expressions to express this linked information adds indeed more difficulty to the entire task.

2. PROPOSED SYSTEM

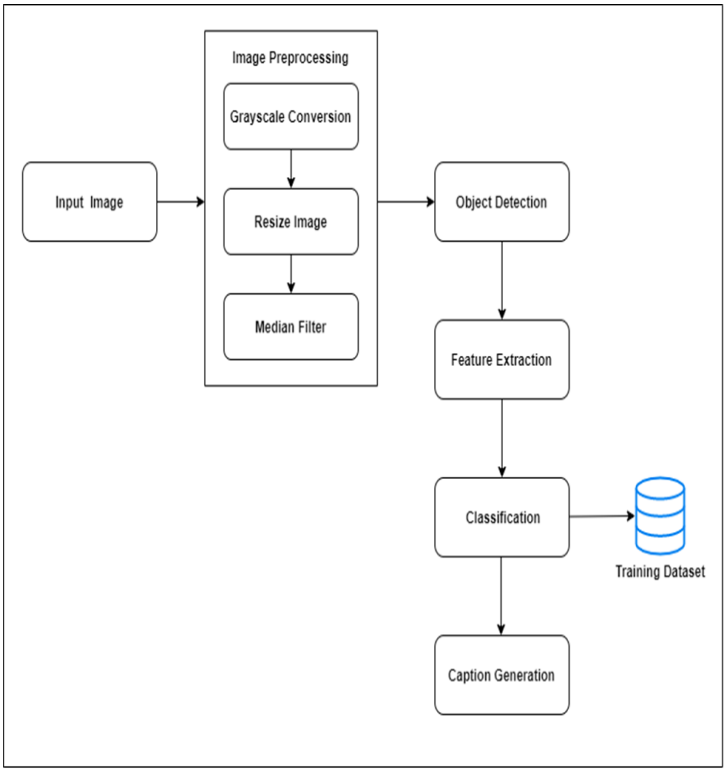
2.1 PROBLEM STATEMENT:

To build and implement Image and Video Captioning Using Machine Learning

2.2 EXISTING SYSTEM

- Existing Methods tend to yield overly general captions and consist of some of the most frequent words/phrases, resulting in inaccurate and indistinguishable descriptions.
- The existing approaches generally employ a data mining-based encoder-decoder architecture and resort to a reinforcement learning method to optimize this task.

2.3 SYSTEM ARCHITECTURE



2.4 ALGORITHMS

1. Input Image:

Here we will upload the Input Image.

2. Image Pre-processing:

In this step we will applying the image pre-processing methods like gray scale conversion, image noise removal.

3. Image Feature Extraction:

In this step we will applying the image object and edge detection methods to extract the image features from image.

4. Image Classification:

In this step we will applying the image classification methods

5. Result:

In this step will show the final generate caption result

2.5 MATEMATICAL MODULE

Relevant mathematics associated with the Project:

Let S be the Whole system S= I, P, O
I-input
P-procedure
O-output
Input (Video)
I=Input as Image
Where ,
Dataset Image
Procedure (P),
P=I,
Using I System perform operations and calculate the prediction
Output(O)-
O = Caption of given Image or Video.

2.6 SYSTEM REQUIRMENTS

2.6.1 SOFTWARE REQUIRMENTS

- Operating System - Windows
- Front End - HTML, Bootstarp,CSS
- Language - Python.
- IDE - Annaconda,Pycharm or jupyter

2.6.2 HARDAWRE REQUIRMENTS

- Processor - Intel i3/i5/i7
- Speed - 2.80 GHz
- RAM - 8 GB
- Hard Disk - 40 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse

2.7 ADVANTAGES

1. Time Saving.
2. Increased Efficiency and Capability.
3. Automated Identification.
4. Improve user experience.

2.8 APPLICATIONS

1. Education System.
2. Natural Language Processing.
3. Automobile.
4. Image indexing

3. CONCLUSIONS

In this paper, we propose a new deep neural network(NDNN) model to ameliorate the image captioning styles. The NDNN explores the relationship in the visual attention and learns the attention transmission medium through a acclimatized LSTM model, where the matrix- form memory cell stores and propagates visual attention, and the affair gate is reconstructed to sludge the attention values. Combined with the language model, both of the generated words and the visual attention areas gain memory in the space. We bed the NDNN model in three classical attention- grounded image captioning fabrics, and acceptable experimental results on the MS COCO and Flicker dataset demonstrate the superiority of the proposed NDNN.

REFERENCES

1. J. Lu, C. Xiong, D. Parikh, and R. Socher, "Knowing when to look: Adaptive attention via a visual sentinel for image captioning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3242–3250.

2. P. Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6077–6086.

3. L. Chen et al., "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5659–5667.

4. T. Yao, Y. Pan, Y. Li, Z. Qiu, and T. Mei, "Boosting image captioning with attributes," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 4904–4912.

5. X. Yang, K. Tang, H. Zhang, and J. Cai, "Auto-encoding scene graphs for image captioning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 10685–10694.

6. M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Paying more attention to saliency: Image captioning with saliency and context attention," *ACM Transactions on Multimedia Computing, Communications, andApplications*, vol. 14, no. 2, p. 48, 2018.

7. M. Yang, W. Zhao, W. Xu, Y. Feng, Z. Zhao, X. Chen, and K. Lei, "Multitask learning for cross-domain image captioning," *IEEE Transactionson Multimedia*, vol. 21, no. 4, pp. 1047–1061, 2018.

8. X. Xiao, L. Wang, K. Ding, S. Xiang, and C. Pan, "Deep hierarchical encoder-decoder network for image captioning," *IEEE Transactions onMultimedia*, 2019.

9. J. H. Tan, C. S. Chan, and J. H. Chuah, "Comic: Towards a compact image captioning model with attention," *IEEE Transactions on Multimedia*, 2019.

10. X. Li and S. Jiang, "Know more say less: Image captioning based on scene graphs," *IEEE Transactions on Multimedia*, 2019.

11. Z. Zhang, Q. Wu, Y. Wang, and F. Chen, "High-quality image captioning with fine-grained and semantic-guided visual attention," *IEEETransactions on Multimedia*, vol. 21, no. 7, pp. 1681–1693, 2018.

12. M. Tanti, A. Gatt, and K. P. Camilleri, "Where to put the image in an image caption generator," Natural Language Engineering, vol. 24, no. 3, pp. 467–489, 2018