(12) PATENT APPLICATION PUBLICATION
(19) INDIA
(22) Date of filing of Application :21/10/2025
(21) Application No.202541101451 A
(43) Publication Date : 28/11/2025

(54) Title of the invention : IMAGE CAPTIONING USING DEEP LEARNING

(51) International classification	:G06N0003045000, G06N0003080000, G06N0003044000, G06F0040300000, G06F0040560000
(31) Priority Document No	:NA
(32) Priority Date	:NA
(33) Name of priority country	:NA
(86) International Application No Filing Date	: :01/01/1900
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**
1)R.V.R. & J.C. COLLEGE OF ENGINEERING
Address of Applicant :R.V.R. & J.C. COLLEGE OF ENGINEERING, CHOWDAVARAM – 522 019. Chowdavaram Andhra Pradesh India
(72)**Name of Inventor :**
1)Mr. R. VEERAMOHANA RAO
2)D. N. V. L. HARSHITHA
3)G. ROHITH
4)G. NANDINI BAI

(57) Abstract :
ABSTRACT [0014] Image captioning is the task of generating textual descriptions from visual content using a combination of computer vision and natural language processing. This project implements a deep learning-based image captioning system that employs the VGG16 convolutional neural network (CNN) for feature extraction and a recurrent neural network (RNN) with LSTM units for sentence generation. The Flickr8k dataset is used to train and evaluate the model. Each image is processed through the VGG16 model to extract high-level features, which are then input to the LSTM-based decoder to generate relevant captions. Captions are pre-processed using tokenization and padding techniques. The model is evaluated using BLEU scores to compare generated captions with human-annotated references. The findings have applications in the areas of content indexing, assistive technology for the visually impaired, and enhancing user interfaces on image-centric systems.
[0015] This approach demonstrates the effectiveness of combining pre-trained visual encoders with sequential language models. The results are promising and indicate potential for real-world applications such as image indexing, accessibility tools for the visually impaired, and content-based image retrieval. Overall, the system achieves meaningful alignment between visual input and natural language output. It effectively bridges the gap between image understanding and language generation. This work serves as a foundation for more advanced multimodal AI systems. A Streamlit web app is created that generates image captions using two models: a custom VGG16+LSTM and a pretrained BLIP transformer. It supports image upload/URL input, feature extraction, and real-time caption comparison with a user-friendly interface.

No. of Pages : 9 No. of Claims : 6