TITLE	AUTHOR	APPROACH	DESCRIPTION	PROS	CONS
DEEP LEARNING APPROACHES ON IMAGE CAPTIONING: A REVIEW (BASE PAPER)	Taraneh Ghandi, Hamidreza Pourreza, Hamidreza Mahyar	They included various methods such as attention-based methods, deep learning approaches, and the utilization of Transformers, scene graphs, and vision language pre-training methods in image captioning. These methods aim to improve the quality of image captions by emphasizing relevant parts of the input image, capturing complex relationships between objects, and learning from joint visual and textual information.For better performance in generating the captions.	The paper provides a structured review of deep learning methods in image captioning, addressing challenges in the field, ranking method performance, and suggesting future research directions.	The pros of using deep learning approaches in image captioning include capturing complex relationships between objects, understanding visual content, generating natural language descriptions, and achieving significant improvements in overall performance.	Need fixing the problem where the words describing an image don't match it exactly. Making sure the data we use to teach computers about images isn't unfairly representing certain types of pictures. finding ways to teach computers to understand both images and words together so they can describe pictures accurately, and developing better tools to check how well computers are doing at describing images.
Pattern Recognition Letters	Songtao Ding, Shiru Qu, Yuling Xi, Arun Kumar	They tried to implement a novel image captioning model based on high-level image features, a bottom-up attention model, mainstream image caption generation algorithm combining CNN image recognition and NLP structured model, a generation model based on a deep recurrent architecture, and an attention-based model using two attention mechanisms.	The paper presents a novel image captioning model based on high-level image features and a bottom-up attention mechanism driven by low-level and advanced image features, aiming to improve image caption generation by mimicking human visual attention mechanisms.	The pros of the methodologies in the paper include addressing challenges in caption generation, effectively combining low-level and high-level features for attention detection, imitating the human visual attention system, and achieving good performance on benchmark datasets.	Limited diversity of descriptions due to hard coding and pre-built sentence templates Difficulty in accurately describing specific behaviors in images with complex backgrounds Challenge in precise object detection with relatively small size
Generating Image Captions based on Deep Learning and Natural language Processing	Smriti Sehgal, Jyoti Sharma, Natasha Chaudhary	Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Natural Language Processing (NLP), VGG16 model	The paper discusses the development and implementation of an Image Caption Generation model using deep learning algorithms like CNN and RNN, alongside NLP, with a focus on aiding visually impaired individuals and those with short-sightedness. The model is implemented in Python,	The pros of the Image Caption Generator model include automation of caption generation, aiding visually impaired individuals, making web pages dynamic, and organizing files efficiently.	Feature maps capturing positions very precisely, leading to different feature maps with small variations in the input image

Image Caption Generation using Convolutional Neural Network and Long Short Term Memory	Afeefa Nazneen N Z and Dr. Shreedhara K S	Deep learning techniques, vast datasets, and computer power are combined to implement the proposed project.	Combining Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) techniques, developing a model using Deep Learning (DL) techniques, vast datasets, and computer power, and ensuring that the generated captions for photographs will only contain words from the proposed model's vocabulary.	The proposed model's data-dependent nature ensures that the generated captions only contain words from the model's vocabulary, enhancing the accuracy and relevance of the generated captions.	The amount of training data required for CNN and LSTM models can be substantial, which may pose challenges in terms of data collection and processing.
Every Picture Tells a Story: Generating Sentences from Images	Ali Farhadi, Mohsen Hejrati , Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, David Forsyth	The paper demonstrates the ability of automatic methods to generate short descriptive sentences from images	Introduces a dataset, a novel representation intermediate between images and sentences, and a discriminative approach for sentence annotation. The model is learned to maximize scores along the path identified by a triplet, using the stochastic subgradient descent method for optimization.	Discusses the representation of sentences and the computation of similarity between sentences and triplets, providing a comprehensive approach to generating descriptive sentences from images	Lack of a dataset with images and corresponding sentences for evaluation Difficulty in quantitative evaluation of generated sentences
Image Caption Generating Deep Learning Model	Aishwarya Maroju, Sneha Sri Doma, Lahari Chandarlapati	The ResNet model is used to extract image features, which are then passed to the LSTM networks for caption generation	The paper utilizes a Convolutional Neural Network (ResNet) as an encoder to access the image features and a Recurrent Neural Network (Long Short Term Memory) as a decoder to generate captions for the images	The proposed model aims to address the vanishing gradient problem commonly encountered in traditional CNN-RNN models	The model may require a substantial amount of training data and iterations to effectively learn the intricate relationships.