

**AMRUTVAHINI COLLEGE OF ENGINEERING,
SANGAMNER**

DEPARTMENT OF COMPUTER ENGINEERING

2023-2024

**Project Synopsis
on**

**“Image Captioning system using deep neural network based
on encoder-decoder framework”**



**BE Computer Engineering
BY**

Group Id-B04

Mr. Mayur Gadakh (4136)

Mr. Gaurav Chaudhari (4117)

Ms. Akanksha Gaikwad (4141)

Ms. Shivanjali Dhage (4126)

Prof. R. S. Gaikwad

Project Guide

Dept. of Computer Engineering

Dr. D. R. Patil/ Dr. R. G. Tambe

Project Coordinator

Dept. of Computer Engineering

Prof. R. L. Paikrao

H.O.D

Dept. of Computer Engineering

- **Title:** Image Captioning system using deep neural network based on encoder-decoder framework.

- **Domain and Sub-domain:** Deep Learning, Computer Vision.

- **Objectives:**

1. To study the deep learning techniques like CNN and RNN.
2. To develop a deep learning-based image caption generator that can accurately describe the contents of an image in natural language.
3. To create a user-friendly interface for interacting with the image captioning system, making it accessible and usable.

- **Abstract:**

This project focuses on the development of an advanced deep neural network-based framework for image captioning, incorporating a "Convolutional Neural Network (CNN)" as the encoder and a "Recurrent Neural Network (RNN)" as the decoder. The proposed model excels in capturing intricate visual details and understanding the contextual relationships among objects within images. In this neural network-based system, CNN is employed as the image feature extractor, effectively preserving spatial information and recognizing objects. And, the RNN decoder is used to predict words and generate coherent and contextually relevant sentences based on the extracted image features. The project aims to improve the caption quality and the handling of complex visual contexts.

- **Keywords:**

Deep Learning, Convolutional Neural Network (CNN), Image Captioning, Recurrent Neural Networks (RNN).

- **Problem Definition:**

The current image captioning systems often struggle to provide detailed and contextually relevant descriptions for the images. To address this limitation, there is a need to develop an enhanced captioning system using deep learning techniques, which can generate more accurate and informative captions, enhancing the value and accessibility of an image content.

- **List of Modules:**

1. Encoder (CNN)
2. Attention Mechanism
3. Decoder (RNN)
4. Post-processing

- **Current Market Survey:**

The market for deep neural network-based encoder-decoder frameworks in image captioning saw rising demand for automated image description solutions across industries. Commercial APIs facilitated easy integration, while industry-strength frameworks like Python-Keras and TensorFlow were popular for custom model development. Transformer-based pre-trained models gained traction. Scalability and domain-specific customization remained vital, with a diverse player base, including tech giants and startups, driving innovation. Staying updated with the latest reports, industry news, and research developments in the rapidly evolving AI and computer vision landscape was essential for market insights.

- **Scope of the Project:**

Automatic image captioning using deep neural network encoder-decoder frameworks has extensive potential. It can enhance accessibility, content indexing, e-commerce, healthcare, education, and more by generating descriptive image captions. This technology streamlines processes, improves user experiences, and finds applications across diverse domains.

- **Literature Survey:**

1. M. M. Rahman, A. Uzzaman and S. I. Sami, “Implementing Deep Neural Network Based Encoder-Decoder Framework for Image Captioning,” 2021 *IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)*, Dhaka, Bangladesh, 2021, pp. 26-31.
2. Kavitha, P. V., and V. Karpagam, “Deep Learning Techniques for Automatic Image Captioning.” *Disruptive Technologies for Big Data and Cloud Applications: Proceedings of ICBDDC 2021* (2022): pp 167-175.

3. M. S. Alam, V. Narula, R. Haldia and G. Nikam Ganpatrao, “An Empirical Study of Image Captioning using Deep Learning,” *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2021, pp. 1039-1044.
4. Hrga, Ingrid, and M. Ivašić-Kos. “An overview of image caption generation methods.” In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 995-1000. IEEE, 2019.
5. Ningthoujam, Chitrapriya, and Tejbanta S. Chingtham, “Comprehensive Comparative Study on Several Image Captioning Techniques Based on Deep Learning Algorithm.” *In Contemporary Issues in Communication, Cloud and Big Data Analytics: Proceedings of CCB 2020*, pp. 229-240. Springer Singapore, 2022.

- **Software and Hardware Requirement of the Project:**

Software:

1. Deep learning frameworks
2. Python programming environment
3. Image processing libraries

Hardware:

1. High-performance GPU (Graphics Processing Unit)
2. Computer system with sufficient memory and processing power

- **Contribution to Society:**

The image captioning systems have a potential to provide more accurate and efficient way of generating captions for images. This could have a wide range of applications, such as assisting visually impaired individuals in understanding the content of images, improving image search engines, and enhancing the accessibility of social media platforms.

- **Probable Date of Project Completion:** Feb. 2024

- **Outcome of the Project:**

1. Implementation of CNN and RNN for image captioning systems.
2. Learning deep neural-network based encoder-decoder framework.