IMAGE CAPTION GENERATION

using Deep Learning

DESCRIPTION

The aim of this project is to empower a bipedal humanoid robot with the ability to perceive its surroundings visually and describe them in natural language, bridging the gap between vision and communication. By integrating deep learning, computer vision and natural language processing (NLP) techniques, the humanoid is capable of generating real-time, contextually accurate captions for the scenes it observes, significantly enhancing its situational awareness and human-robot interaction capabilities.

Humanoid Robot

SOLUTIONS provided

Automated Visual Understanding

The humanoid can automatically perceive and understand its environment through images captured by its onboard camera, without any human intervention.

Real-Time Image Captioning

By using a pre-trained VGG16 model and an LSTM-based language model, the robot can generate real-time descriptive captions about what it sees.

Human-Robot Interaction Enhancement

By describing its environment in human language, the humanoid becomes much easier and more natural for people to interact with, especially for users who are not technically trained.

Real-Time Processing on Embedded Hardware

The solution is optimized to work on low-power, real-time embedded systems (like Raspberry Pi), making it cost-effective and deployable in mobile, lightweight robots.

ROADMAP

01 DATASET COLLECTION AND PREPROCESSING

Collect an image-caption dataset (Flickr8k). Clean and tokenize the captions .Resize and normalize images to fit the input requirements of the VGG16 model.

02 FEATURE EXTRACTION USING PRE-TRAINED VGG16

Pass images through a pre-trained VGG-16 model, followed by extracting and saving high-level feature vectors.

03 TRAINING THE CAPTION GENERATOR MODEL

Design a model that combines image features and text sequences.
Use an Embedding Layer for captions, followed by an LSTM network to learn how to generate text based on image features.

04 REAL-TIME CAPTION GENERATION

Predict a caption word-by-word using the trained LSTM model.

05 OPTIMIZATION AND DEPLOYMENT ON THE HUMANOID

Optimize the model for fast, real-time processing on embedded hardware (e.g., Raspberry Pi)

PROBLEMS SOLVED

LACK OF VISUAL UNDERSTANDING IN HUMANOID ROBOTS

Most humanoid robots lack the ability to interpret and describe what they see in natural language.

LIMITED HUMAN-ROBOT INTERACTION

Communication between humans and robots is often limited to commands or simple responses.

INACCESSIBILITY FOR VISUALLY IMPAIRED USERS

Visually impaired individuals cannot benefit from robot assistance if the robot cannot describe its surroundings.

CONTRIBUTIONS

ABHIK RAJGARIA

- Hardware
- Hardware-Software Integration
- Coding

SMRIDDHI PARASHAR

- Software coding
- Research

Humanoid Robot

THANK YOU

ABHIK RAJGARIA
E22CSEU1357
SMRIDDHI PARASHAR
E22CSEU1524