

Project Proposal for Visual Question Answering System

Title: Enhancing Accessibility Through Visual Question Answering

Authors: Rajib Lochan Das (50546929)

Introduction

This project aims to develop a Visual Question Answering (VQA) system that helps visually impaired users interact with their environment by answering questions posed in natural language about their surroundings. By leveraging advancements in computer vision and natural language processing, this system will interpret images and provide descriptive answers, thus enhancing accessibility and interaction for users with visual impairments.

Problem Statement

The primary challenge is to create a VQA system capable of accurately interpreting visual data and generating coherent responses to user-posed questions. The system will utilize the VQA v2.0 dataset, which includes images paired with question-answer sets. We aim to measure the system's accuracy, precision, and recall, employing standard metrics used in machine learning for image and language processing tasks.

Technical Approach

The project will use a combination of Convolutional Neural Networks (CNNs) for image recognition and Long Short-Term Memory (LSTM) networks for processing the questions. An integration layer will fuse the outputs of both networks to generate answers. The model's performance will be compared against baseline models in the field to evaluate improvements.

Milestones

1. Literature Review and Feasibility Study (7 days): Research existing VQA models and select appropriate technologies.
2. Data Acquisition and Preprocessing (3 days): Download and preprocess the VQA v2.0 dataset.
3. Model Development (2 days): Develop the CNN and LSTM models and integrate them.
4. Initial Testing and Refinement (7 days): Test the model on a validation set and refine based on performance metrics.
5. Final Evaluation and Documentation (5 days): Perform extensive testing, document the project, and prepare the final report.

Societal Benefits

This VQA system will significantly aid visually impaired individuals, allowing them a greater understanding of their visual surroundings, thus promoting inclusivity and accessibility.

Ethical Considerations

The project will adhere to ethical AI guidelines, ensuring privacy in data handling, fairness in model outcomes, and transparency in algorithms.

Evaluation of Results

The system will be evaluated based on its accuracy, precision, and recall, comparing its performance against existing benchmarks. Qualitative evaluations will also be included, with user feedback sessions to determine the system's practical efficacy.