Image Captioning Using Deep Learning

Infosys Springboard Internship

Elevate medical diagnostics and streamline report generation through the power of deep learning-driven image captioning. This innovative project explores the development and deployment of advanced models to bring enhanced efficiency and precision to the healthcare industry.

Intern Name : Shreyas H S
Mentor Name : Sudheer Kumar Y

Project Overview and Objectives

1

Project Overview

Development and deployment of an image captioning system for radiology images using deep learning models.

2

Objective

Enhance medical diagnostics and report generation through automated image captioning.

Deep Learning Models Employed

Vision Encoder-Decoder Model

Integrates Google's ViT with GPT-2 for end-to-end image captioning.

BLIP Model

Conditional image captioning model developed by Salesforce.

CNN-RNN Model

Combines CNNs for feature extraction with RNNs for sequential caption generation.

Data Science Process

1

Data Acquisition

Collecting and organizing the ROCO dataset of radiology images and text captions.

2

Data Preprocessing

Cleaning, resizing, and normalizing images; tokenizing and cleaning text captions.

3

Model Training

Optimizing hyperparameters, leveraging transfer learning, and fine-tuning models.

Data Acquisition and Preprocessing

Dataset

ROCO (Radiology Objects in Context) dataset, containing radiology images and associated text captions.

Image Preprocessing

Loading, resizing, and normalizing images to fit model requirements.

Text Preprocessing

Tokenization, cleaning, and removal of special characters from the text captions.

Training Process and Optimization

1

Hyperparameter Tuning

Optimizing batch size, learning rate, and number of epochs for best performance.

2

Optimization Techniques

Employing Adam optimizer, early stopping, and model checkpointing to prevent overfitting.

3

Transfer Learning

Leveraging pre-trained models to accelerate training and improve accuracy.

Evaluation and Results

1

Evaluation Metrics

ROUGE and BLEU scores to assess the quality and coherence of generated captions.

2

Model Performance

Comparison of CNN-LSTM and VisionEncoderDecoderModel (BLIP) on validation and test sets.

3

Analysis

Transformer-based models demonstrate superior accuracy and caption coherence over CNN-based approaches.

Use Cases and Applications

- Diagnostic Assistance
- Medical Education
- Clinical Documentation
- Telemedicine
- Medical Research
- Quality Assurance
- Patient Communication
- Regulatory Compliance

THANK YOU

Thank you Infosys Springboard for the wonderful opportunity to show case our work!