Model Development Phase Template

Date	10 November 2024
Team ID	739939
Project Title	Image Caption Generator
Maximum Marks	5 Marks

Model Selection Report

For an image caption generator leveraging ResNet50, the model selection often involves using the pre-trained weights of ResNet50 (trained on a large image dataset like ImageNet) as the image encoder to extract visual features. These features are then fed into a decoder (often an RNN or Transformer-based network) that generates the caption.

Model Selection Report:

Model	Description
ResNet50	ResNet50 is a convolutional neural network architecture that is 50 layers deep and a popular choice for image recognition tasks. Its key innovation is the use of "residual connections" or "skip connections," which allow gradients to flow more easily through the network, mitigating the vanishing gradient problem and enabling the training of much deeper networks. These connections help the network learn identity mappings, making it easier to optimize and often leading to improved performance.