

Model Optimization and Tuning Phase Template

Date	10 November 2024
Team ID	739889
Project Title	Image Caption Generator
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

Model optimization for an image caption generator aims to improve its performance (e.g., caption accuracy, fluency) and efficiency (e.g., training time, memory usage). This often involves techniques like tuning hyperparameters (learning rate, batch size), applying regularization (dropout, weight decay), and potentially using more efficient network architectures or optimization algorithms (e.g., AdamW).

Hyperparameter Tuning Documentation (8 Marks):

Model	Tuned Hyperparameters(Epochs)
ResNet50	In this image caption generator project, training involves iterating through the entire image-caption dataset multiple times, with each full pass being an epoch. The number of epochs is a hyperparameter that determines how long the model trains, aiming for optimal performance without overfitting or underfitting the data.

	<pre> Epoch 1/5 100/100 ————— 0s 49ms/step - accuracy: 1.9152e-04 - loss: 72144.0234 Accuracy for epoch 1: 0% 100/100 ————— 6s 49ms/step - accuracy: 1.9117e-04 - loss: 72658.7969 Epoch 2/5 100/100 ————— 0s 91ms/step - accuracy: 9.9295e-04 - loss: 315246.4688 Accuracy for epoch 2: 0% 100/100 ————— 9s 91ms/step - accuracy: 9.8621e-04 - loss: 315892.5938 Epoch 3/5 100/100 ————— 0s 57ms/step - accuracy: 9.6183e-06 - loss: 585496.3125 Accuracy for epoch 3: 0% 100/100 ————— 6s 57ms/step - accuracy: 1.1070e-05 - loss: 586189.5000 Epoch 4/5 100/100 ————— 0s 49ms/step - accuracy: 1.1384e-04 - loss: 864011.3125 Accuracy for epoch 4: 0% 100/100 ————— 5s 49ms/step - accuracy: 1.1426e-04 - loss: 864654.9375 Epoch 5/5 100/100 ————— 0s 74ms/step - accuracy: 4.8886e-05 - loss: 1113768.7500 Accuracy for epoch 5: 0% 100/100 ————— 7s 74ms/step - accuracy: 4.9949e-05 - loss: 1114264.5000 <keras.src.callbacks.history.History at 0x7ebda9b2b6a0> </pre>
--	---

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
ResNet50	<p>ResNet50 is a strong choice as the image encoder in this project due to its proven ability to extract robust and hierarchical visual features from images, thanks to its deep architecture and residual connections.</p> <p>Leveraging pre-trained ResNet50 weights allows the model to benefit from features learned on large image datasets, leading to better understanding of image content and ultimately, more accurate and relevant captions.</p>