Milestone 2

Varun Chitturi , Brandon Kleinman , Aditya Sirohi , Runyu Tian

Evaluation measures:

Watermark detection measures:

- **Precision**: Indicates the proportion of true positive predictions among all positive predictions.

  Precision = (TP)/(TP+FP)

- **Recall (Sensitivity):** Reflects the proportion of actual positives correctly identified

  Recall = (TP) / (TP+FN)

- **F1 Score:** The harmonic mean of Precision and Recall.

  F1 = 2 * (Precision * Recall) / (Precision + Recall)

Captioning measures:

- **BLEU**: Assesses the overlap of n-grams between the predicted and reference captions.
- **METEOR:** Considers synonymy and stemming, providing a more nuanced evaluation compared to BLEU.
- **ROUGE:** Focuses on the overlap of word sequences, particularly useful for longer text outputs.

Watermark Dataset preparation:

`CocoCaptionMixedWMDataset` contains a mixture of watermarked images (created by embedding signatures using the `StegaStampEncoder`) paired with their corresponding signatures, and original images with captions.

Simple Baseline:

We use the nearest neighbor based approach in our simple baseline. Initially, we calculate the Euclidean distance between each validation image and all images in the training dataset using pixel values across all channels. After identifying the most similar training image, we assign the watermark detection label from that training image and assign the caption based on the corresponding training image's caption.

Strong Baseline:

For each image, the CNN-based StegaStampDecoder endeavors to extract the embedded signature. This is accomplished by processing the image through the decoder and applying a predefined bit threshold to determine the presence of a watermark. If a watermark is detected, the decoded signature is utilized as the generated caption. Conversely, if no watermark is identified, captions are generated using a pre-trained ViT-GPT2 language model based on the unwatermarked images.

Performance:

	Precision	Recall	F1	Meteor	BLEU	ROUGE1	ROUGE2	ROUGEL
Simple	**0.549**	**0.656**	**0.598**	**0.247**	**0.01**	**0.452**	**0.01**	**0.4**
Strong	**0.999**	**0.954**	**0.976**	**0.347**	**0.262**	**0.526**	**0.164**	**0.505**