Information Retrieval Assignment 2

Dhyan Vimalkumar Patel
Roll No. 2021041

March 18, 2024

Introduction

This project aims to develop a Multimodal Retrieval System that leverages the capabilities of Convolutional Neural Networks (CNNs) and text processing techniques to analyze and retrieve relevant images and text reviews. Focused on processing and comparing multimodal data (images and text) for similarity, the system performs a series of steps including image preprocessing, feature extraction, text preprocessing, and TF-IDF score calculation. The endeavor explores the integration of visual and linguistic content to enhance retrieval accuracy and relevance, providing insights into the effectiveness of multimodal data analysis.

1 Image Feature Extraction

Libraries Used

The preprocessing stage utilized one primary library:

- **Keras' InceptionV3 and Model**: A pre-trained InceptionV3 model is initialized globally to extract features from images. This model, known for its depth and accuracy in image recognition tasks, is loaded without its top classification layer, making it adaptable for feature extraction.

- **PIL for Image Processing**: The Python Imaging Library (PIL), particularly its Image and ImageEnhance modules, is used for opening, converting, resizing, and enhancing the contrast and brightness of images.

- **NumPy and Requests for Data Handling:**: NumPy is utilized for its efficient array manipulation capabilities, transforming image data into a format suitable for the model. The requests library fetches images from specified URLs, while BytesIO handles byte streams from fetched images.

Functionality

Preprocessing and feature extraction procedure:

- The `preprocess and extract features` function embodies the preprocessing workflow, including image resizing to 224x224 pixels (a requirement for InceptionV3), contrast and brightness enhancement for image quality improvement, and conversion of images into arrays for model processing.

- The InceptionV3 model then predicts on the preprocessed image data, extracting deep features indicative of the image's content. These features are reshaped into a 1D array to standardize output, facilitating easier comparison and retrieval tasks.

2 Text Feature Extraction

Text Preprocessing and TF-IDF Computation:

- **nltk for Natural Language Processing**: The Natural Language Toolkit (NLTK) provides tools for text preprocessing, including tokenization, lemmatization, and stop word removal, which are essential for cleaning and standardizing text data.

- **TF-IDF Score Calculation**: This process begins with term frequency (TF) computation, assessing the importance of words within documents. Inverse document frequency (IDF) calculation follows, measuring how much information the word provides across the entire document corpus. Combining TF and IDF into TF-IDF scores quantifies the relevance of words within each document, aiding in the identification of significant textual features.

3 Image and Text Retrieval

Cosine Similarity for Feature Comparison:

- **Image Feature Similarity**: Cosine similarity measures the cosine of the angle between two vectors, in this case, the feature vectors extracted from images. This metric effectively quantifies the similarity between images, with a higher cosine value indicating greater similarity.
- **Text Feature Similarity**: Similarly, cosine similarity is applied to the TF-IDF vectors representing text data, facilitating the retrieval of text documents that are most relevant to a given query based on their content.

Retrieval Mechanisms:

- **Finding Similar Images**: The `find most similar` function iterates through the dataset of pre-extracted image features, comparing each with the features of the input image to identify the most similar images.
- **Finding Similar Reviews**: The `find most similar reviews` function performs a parallel process for text, comparing the TF-IDF vector of the input text against those of the dataset to find the reviews most similar in content.

4 Results and Analysis

Execution and Output Display:

- Upon receiving user input, the system preprocesses the input image and text to compute their features and TF-IDF scores, respectively.

- It then employs the similarity measures to retrieve and rank similar images and reviews from the dataset, showcasing the top matches and their similarity scores.
- The output includes a composite similarity score, averaging the image and text similarity scores to provide a holistic measure of relevance.

Saving and Serializing Data:

- **Efficient Data Handling**: By serializing the preprocessed and computed data into pickle files, the system ensures quick loading and processing of this data for retrieval tasks, minimizing computation time during user interactions.

Better Similarity Score and Reasoning:

- To determine which retrieval technique—image or text—yields a better similarity score, several factors must be considered, such as the nature of the dataset, the consistency of the data modalities, and the quality of feature extraction and text processing methods.
- **Image Retrieval**: The effectiveness of image retrieval largely hinges on the diversity and distinctiveness of visual features across the dataset. InceptionV3, being a deep learning model pre-trained on ImageNet, is adept at extracting a wide array of features from images, making it a powerful tool for distinguishing between visually diverse images. Image retrieval is likely to perform well if the dataset contains images with clear, distinguishable features.
- **Text Retrieval**: Text Retrieval shines with rich, varied textual content, utilizing pre-processing and TF-IDF vectorization to distinguish documents based on nuanced differences in their topics and language.

  Note

  The output prioritises the most similar picture first, regardless of whether the image is the input itself, and the same is true for the review text. Consideration was given to a composite of these factors when ranking them.The cosine similarity score is given just for the retrieved image, not the average.

Reasoning for Better Performance:

- If the dataset includes images with high visual similarity but textual content that varies significantly, text retrieval might yield more distinctive and relevant results. This is because the nuanced differences in text can provide a better basis for differentiation than visual features alone.

- Conversely, if the dataset consists of visually distinct images but the text is relatively uniform or less descriptive, image retrieval could perform better, leveraging the strong feature extraction capabilities of InceptionV3.

Challenges:

- **Quality of Data**: The performance of both retrieval systems can be significantly affected by the quality of the input data. Noisy, incomplete, or irrelevant data can lead to poor feature extraction and text processing results.

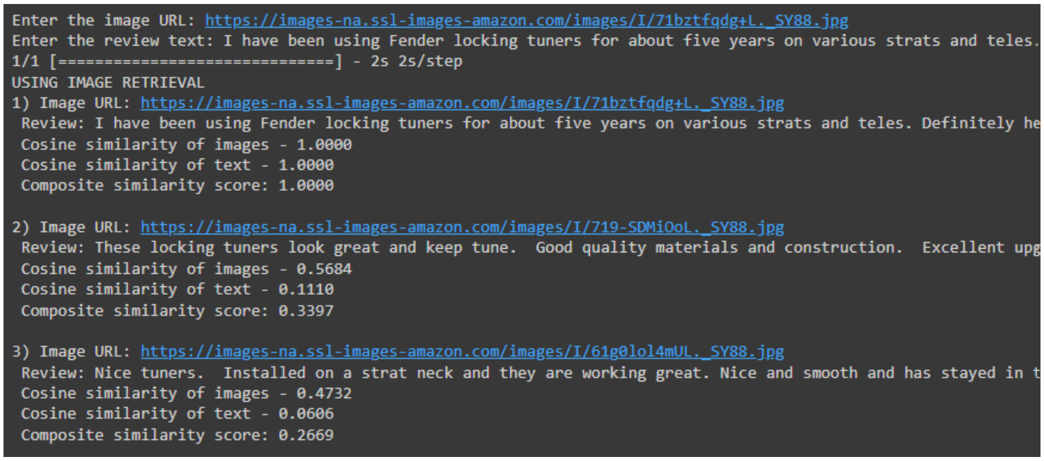
- **Feature Extraction Limitations**: While InceptionV3 is powerful, it may not capture all relevant features for specific domains or specialized image types, potentially leading to suboptimal image retrieval results.

- **Text Variability**: Natural language is highly variable and context-dependent. The effectiveness of text preprocessing and TF-IDF vectorization can be impacted by synonyms, polysemy, and the specificity of the domain language.

Potential Improvements:

- Enhancements can be made through advanced preprocessing techniques, model fine-tuning on domain-specific datasets, adoption of more sophisticated text vectorization methods (like word embeddings or transformer-based models), and developing hybrid retrieval systems that combine both image and text features for a more nuanced understanding and retrieval performance.

5 Sample Output

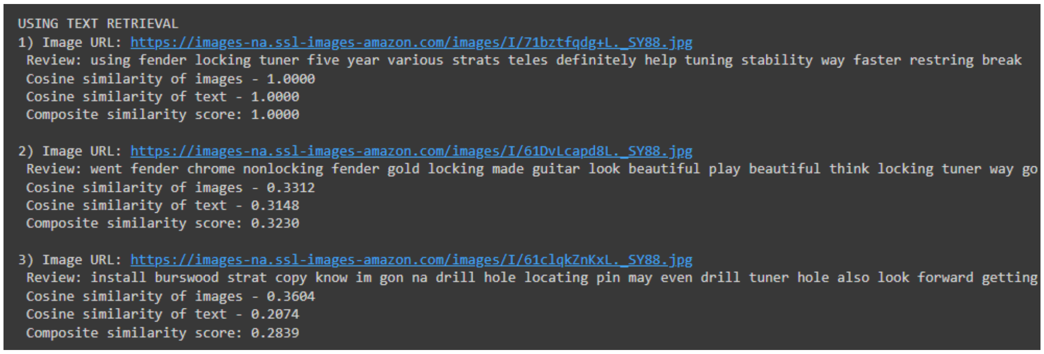


```
Enter the image URL: https://images-na.ssl-images-amazon.com/images/I/71bztfqdg+L._SY88.jpg
Enter the review text: I have been using Fender locking tuners for about five years on various strats and teles.
1/1 [==============================] - 2s 2s/step
USING IMAGE RETRIEVAL
1) Image URL: https://images-na.ssl-images-amazon.com/images/I/71bztfqdg+L._SY88.jpg
 Review: I have been using Fender locking tuners for about five years on various strats and teles. Definitely he
 Cosine similarity of images - 1.0000
 Cosine similarity of text - 1.0000
 Composite similarity score: 1.0000

2) Image URL: https://images-na.ssl-images-amazon.com/images/I/719-SDMiOoL._SY88.jpg
 Review: These locking tuners look great and keep tune.  Good quality materials and construction.  Excellent upg
 Cosine similarity of images - 0.5684
 Cosine similarity of text - 0.1110
 Composite similarity score: 0.3397

3) Image URL: https://images-na.ssl-images-amazon.com/images/I/61g0lol4mUL._SY88.jpg
 Review: Nice tuners.  Installed on a strat neck and they are working great. Nice and smooth and has stayed in t
 Cosine similarity of images - 0.4732
 Cosine similarity of text - 0.0606
 Composite similarity score: 0.2669
```

Figure 1: Output - Image Retrieval



```
USING TEXT RETRIEVAL
1) Image URL: https://images-na.ssl-images-amazon.com/images/I/71bztfqdg+L._SY88.jpg
 Review: using fender locking tuner five year various strats teles definitely help tuning stability way faster restring break
 Cosine similarity of images - 1.0000
 Cosine similarity of text - 1.0000
 Composite similarity score: 1.0000

2) Image URL: https://images-na.ssl-images-amazon.com/images/I/61DvLcapd8L._SY88.jpg
 Review: went fender chrome nonlocking fender gold locking made guitar look beautiful play beautiful think locking tuner way go
 Cosine similarity of images - 0.3312
 Cosine similarity of text - 0.3148
 Composite similarity score: 0.3230

3) Image URL: https://images-na.ssl-images-amazon.com/images/I/61clqkZnKxL._SY88.jpg
 Review: install burswood strat copy know im gon na drill hole locating pin may even drill tuner hole also look forward getting
 Cosine similarity of images - 0.3604
 Cosine similarity of text - 0.2074
 Composite similarity score: 0.2839
```

Figure 2: Output - Text Retrieval

6 Conclusion

This detailed report highlights the system's robust approach to handling and analyzing multimodal data through advanced preprocessing, feature extraction, and retrieval techniques. By leveraging deep learning models for image analysis and sophisticated text processing methods for natural language data, the system effectively bridges the gap between visual and textual content, offering a comprehensive solution for multimodal retrieval tasks.