

Information Retrieval Assignment 2

Dhyan Vimalkumar Patel

Roll No. 2021041

March 3, 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' ResNet50 and Model:** A pre-trained ResNet50 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 pre-processing workflow, including image resizing to 224x224 pixels (a requirement for ResNet50), contrast and brightness enhancement for image quality improvement, and conversion of images into arrays for model processing.
- The ResNet50 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.

4 Image and Text Retrieval

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.

5 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.

6 Sample Output

```
Enter the image URL: https://images-na.ssl-images-amazon.com/images/I/71bztfdg+L.\_SY88.jpg
Enter the review text: Loving these vintage
WARNING:tensorflow:5 out of the last 1992 calls to <function Model.make_predict_function.<locals>.predict_1
1/1 [=====] - 3s 3s/step
USING IMAGE RETRIEVAL
1) Image URL: https://images-na.ssl-images-amazon.com/images/I/71bztfdg+L.\_SY88.jpg
Review: using fender locking tuner five year various strats teles definitely help tuning stability way fast
Cosine similarity of images - 1.000
Cosine similarity of text - 0.0000

2) Image URL: https://images-na.ssl-images-amazon.com/images/I/719-SDMi0oL.\_SY88.jpg
Review: locking tuner look great keep tune good quality material construction excellent upgrade guitar drill
Cosine similarity of images - 0.434
Cosine similarity of text - 0.0000

3) Image URL: https://images-na.ssl-images-amazon.com/images/I/71KBeByN-9L.\_SY88.jpg
Review: hold tune ok string clamp loosen tuning update theyve better time go
Cosine similarity of images - 0.261
Cosine similarity of text - 0.0000

Composite similarity scores of images: 0.565
Composite similarity scores of text: 0.565
Final composite similarity score: 0.565
```

Figure 1: Output - Image Retrieval

```
USING TEXT RETRIEVAL
1) Image URL: https://images-na.ssl-images-amazon.com/images/I/81q5+IxFVUL.\_SY88.jpg
Review: loving vintage spring vintage strat good tension great stability floating bridge want spring way gc
Cosine similarity of text - 0.487
Cosine similarity of images - 0.178

2) Image URL: https://images-na.ssl-images-amazon.com/images/I/71YEX7X28kL.\_SY88.jpg
Review: say im loving
Cosine similarity of text - 0.408
Cosine similarity of images - 0.112

3) Image URL: https://images-na.ssl-images-amazon.com/images/I/81eTaHUrNeL.\_SY88.jpg
Review: great little compact mixer super versatile loving basement studio
Cosine similarity of text - 0.236
Cosine similarity of images - 0.139

Composite similarity scores of images: 0.377
Composite similarity scores of text: 0.377
Final composite similarity score: 0.377
```

Figure 2: Output - Text Retrieval

7 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.