Introduction
The objective of this assignment is to create a multimodal retrieval system - image and text based on the data provided.

Data Preprocessing
The dataset comprises image IDs and corresponding image links. To make the dataset more usable, we converted it into a more structured format, facilitating efficient processing. The dataset structure was crucial for subsequent image feature extraction.

1 Image Feature Extraction
3.1 Basic Image Pre-processing
To prepare the images for feature extraction, we employed basic image pre-processing techniques. This involved contrast adjustment, resizing, geometric orientation alterations, random flips, brightness adjustments, and exposure modifications. These operations collectively enhanced the robustness of the subsequent feature extraction process.

3.2 CNN Architecture Selection
The choice of a pre-trained CNN architecture is crucial for effective feature extraction. We considered ResNet, VGG16, Inception-v3, and MobileNet, all pre-trained on the ImageNet dataset. After careful consideration, we opted for ResNet due to its proven performance in image recognition tasks.

3.3 Feature Extraction
Utilizing the ResNet architecture, we extracted relevant features from the images in the training set. The pre-trained weights of ResNet were leveraged to capture hierarchical features, ensuring the representation of distinctive patterns in the images. Fine-tuning was unnecessary as the pre-trained ResNet already demonstrated strong generalization capabilities.

3.4 Feature Normalization
Normalization of the extracted features is a critical step to ensure consistent and comparable representations. The feature vectors obtained from ResNet were normalized to bring them within a standard range, promoting stability and convergence during subsequent retrieval tasks.

2 Text Pre-processing

2.1 Textual data often requires pre-processing to ensure that it is in a suitable form for feature extraction. The following techniques were implemented on the given text reviews:

Lower-Casing
All text reviews were converted to lowercase to ensure uniformity and prevent case sensitivity issues during subsequent processing steps.

Tokenization

Tokenization involved breaking down the text into individual words or tokens. This step is essential for building meaningful representations of the text.

Removing Punctuations
Punctuation removal was performed to eliminate unnecessary characters that do not contribute significantly to the semantic meaning of the text.

Stop Word Removal
Stop words, commonly used but generally uninformative words, were removed from the text to focus on the more meaningful content.

Stemming and Lemmatization
Stemming and lemmatization were applied to reduce words to their base or root form, aiming to further simplify and standardize the vocabulary.

2.2TF-IDF Calculation
Term Frequency (TF)
Term frequency measures the frequency of a term within a document. It is calculated as the ratio of the number of occurrences of a term to the total number of terms in a document.

Inverse Document Frequency (IDF)
Inverse document frequency measures the importance of a term across the entire dataset. It is calculated as the logarithm of the total number of documents divided by the number of documents containing the term.

TF-IDF Scores
The TF-IDF scores for the textual reviews were calculated using the product of term frequency and inverse document frequency. These scores provide a numerical representation of the importance of terms in individual reviews relative to the entire dataset.

To facilitate the development of separate retrieval systems for images and text, the extracted features were organized and stored in separate dictionaries.

The utilization of cosine similarity for both image and text-based retrieval systems provides a robust and efficient method for measuring the similarity between queries and stored features. The top 3 relevant results, based on cosine similarity scores, offer users concise and accurate retrieval outcomes.

Challenges Faced
1) Some links were not valid
2) For an image id there were multiple links, so I was confused how to take cosine similarity of image links as there were multiple links.