**Lab3 Report**

1. **Project Review**

This project implements a CLIP-based multimodal image search engine that leverages the power of OpenAI’s CLIP model to allow users to retrieve images using either natural language or example image inputs. The interface is built with Gradio, providing an intuitive and interactive user experience.

Users can not only perform semantic search tasks in real-time but also preview the results directly in a gallery, manage a personal image collection through a Favorites system, and take further actions such as downloading results.

By integrating CLIP’s zero-shot vision-language capabilities, the system achieves semantic-level search without requiring class labels or prior annotations, enabling more flexible and natural interactions between users and image data.

1. **Project Functions**

|  |  |
| --- | --- |
| Function Category | Description |
| Search By Text | This feature allows users to input a natural language description. |
| Search By Image | Users can upload an image as a query to find the most visually and semantically similar results. |
| Result Overview | Once a search is performed, the matched results are displayed in a gallery format. |
| Add to Favorites | After selecting a preview image from the result gallery, users can click the "Add to Favorites" button to add it to Favorites. |
| Download Image | This feature allows users to download their favorite or selected search result images to their local device. |

**3.1 The Dataset Used**

The dataset adopted in this project is the GroceryStoreDataset, publicly available at <https://github.com/marcusklasson/GroceryStoreDataset>.

It contains a total of 5125 high-resolution natural images categorized into 81 different food-related classes. The dataset primarily consists of daily grocery items such as fruits, vegetables, beverages, packaged goods, and dairy products, offering a rich and diverse base for training and evaluating semantic image search performance.

Each class includes multiple sample images with varied lighting, backgrounds, and object orientations, which provides realistic variation for robust image embedding and matching.

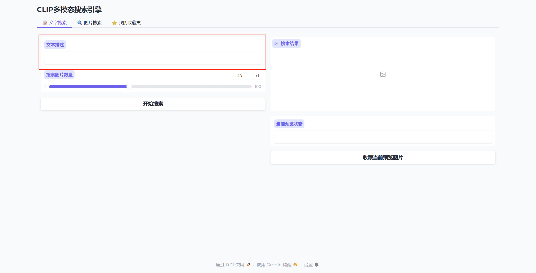
**3.2 How Does The Framework Reveal Five-Stage Search Framework?**

**3.2.1 Formulation**

The system offers two main input interfaces to initiate the search: a textbox for textual queries and an image uploader for visual queries. Users can freely input descriptive terms or drag-and-drop/upload an example image to serve as the query.

These inputs are converted into CLIP embeddings to perform the semantic search. In addition, once a query is submitted, the user is able to preview the input content (either the image or the text prompt) within the interface, ensuring full transparency and clarity before executing the search.

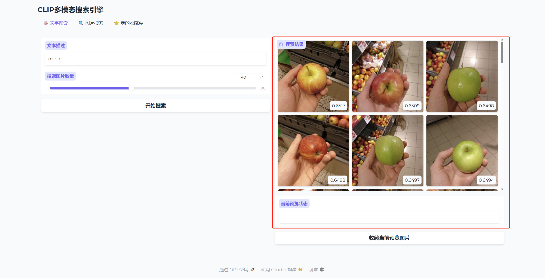
- Input box to upload a description text:



* Input box to upload an image:

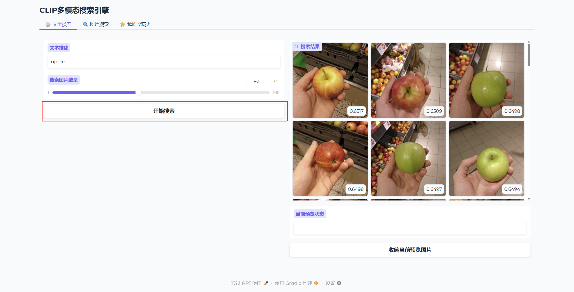


Users can preview the query in the searching window:



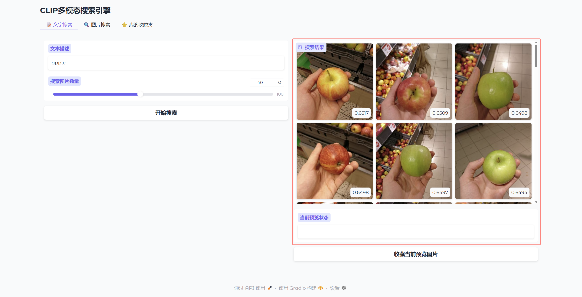
**3.2.2 Initiation Of Action**

After deciding the input method, users trigger the search action by clicking the appropriate search button. This invokes backend inference using the CLIP model and retrieves the most relevant results from the indexed dataset. The search operation is real-time and gives immediate feedback to the user, facilitating a fluid and responsive experience.

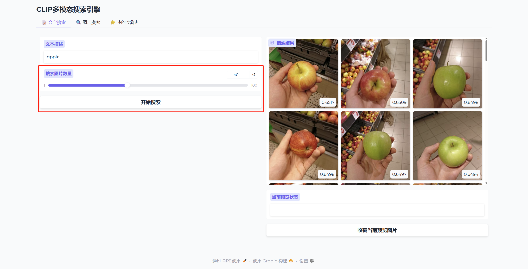


**3.2.3 Review of results**

Upon completion of a search, the interface dynamically presents the matched images in a scrollable, selectable gallery. This layout allows users to visually examine the result set and understand how closely each image matches the original query. Each result is accompanied by similarity scores or labels, and selecting an image updates the preview section, where the current selection is displayed in detail.



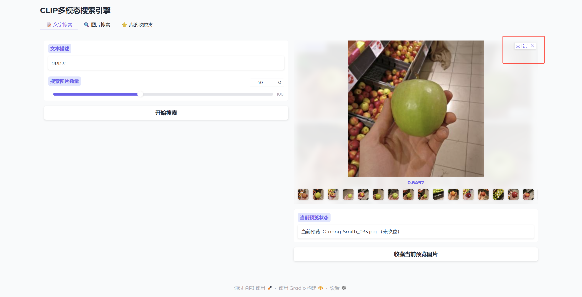
**3.2.4 Refinement**

The framework supports iterative refinement through adjustable parameters. A slider is provided to allow users to set the number of top-k results returned in each search. If the initial result set is not satisfactory, the user can tweak the slider to increase or decrease the scope of the results and rerun the search. This encourages exploratory search behavior and offers flexibility to adapt to different needs.

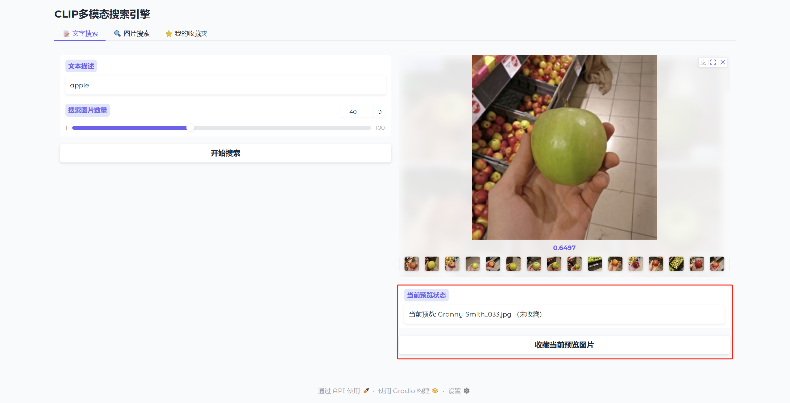
**3.2.5 Use**

Users can take some actions to the images.

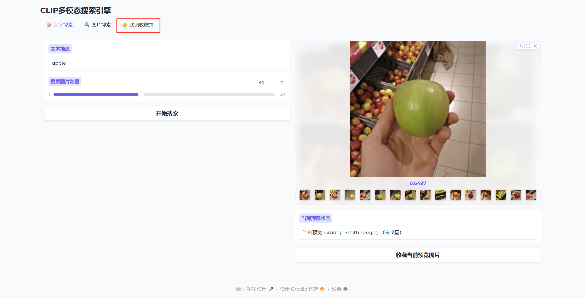
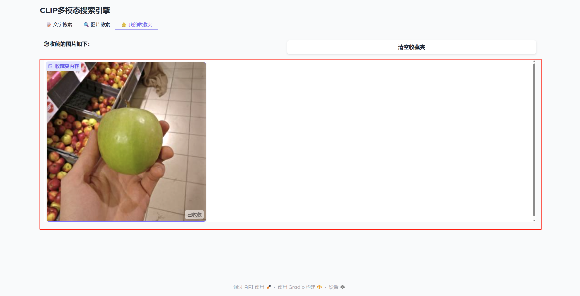
After the user searches for results, he can click on the image in the preview box to select it, and can choose to maximize the image or download it to a local computer.



After the user searches for the results, he can click to select the image in the preview box, and then click the "Add current preview image to favorites" button to add the current preview image to the favorites. At the same time, whether the current image has been added to the favorites will be displayed in the status bar below the preview box.



Users can click "My Favorites" at the top of the interface to go to the Favorites interface and view the pictures currently in their collection.

After entering the favorites interface, users can click the "Clear Favorites" button to clear the files saved in the current favorites.

