PRCV Project 2: Content-based Image Retrieval

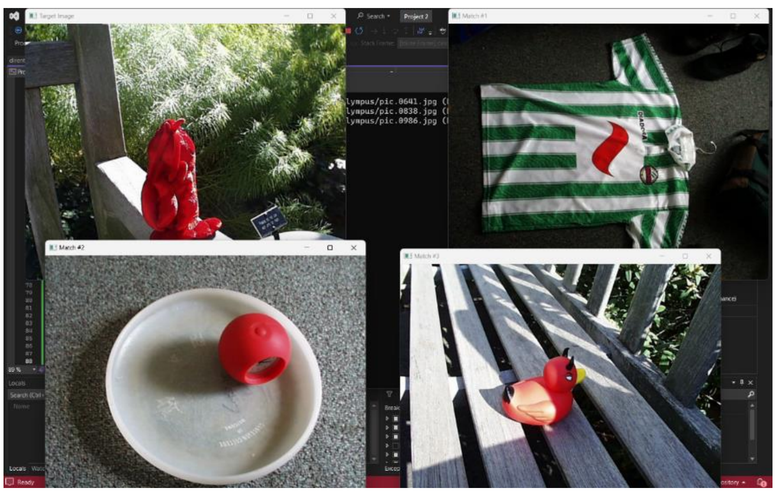
Basil Reji & Kevin Sani

Project Description

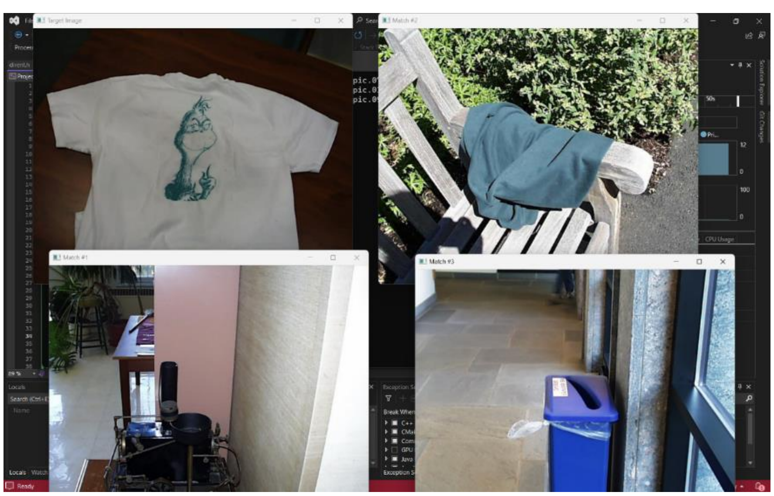
The main goal of the project is to put in place a content-based image retrieval (CBIR) system, in which visual content, as opposed to verbal metadata, is used to compare and match images. The system receives as inputs a target image, a database of images, a feature computation method, and a distance measure for comparison. Next, using the chosen distance metric, it computes features for the target image and the images in the database, compares them, and returns the top N matches. In addition to examining multiple distance measures to gauge image similarity, the project investigates a variety of feature extraction techniques, including histograms, texture analysis, and deep neural network embeddings. The objective of this procedure is to gather knowledge about various image comparison and retrieval algorithms, comprehend their advantages and disadvantages, and put into practice an effective and efficient CBIR system.

Baseline Matching

Original Test Case

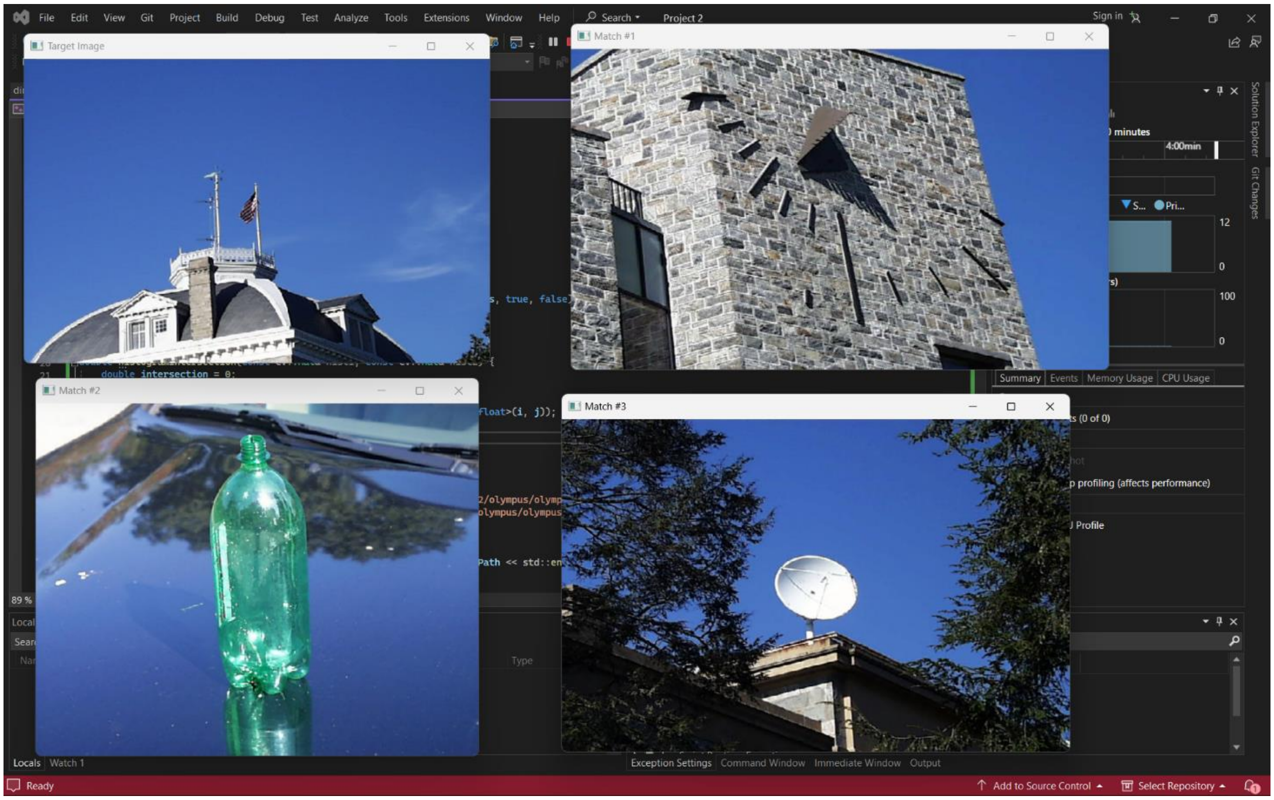


Second Test Case

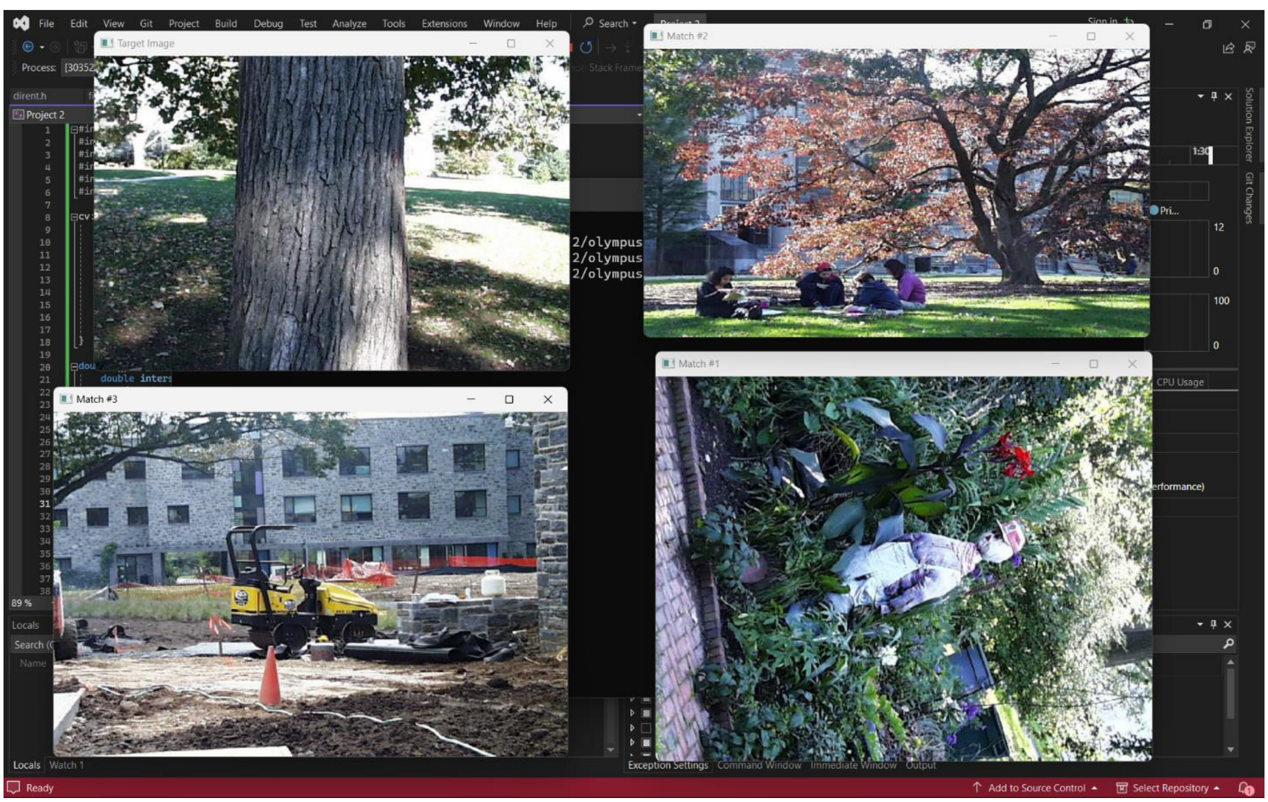


Histogram Matching

Original Test Case

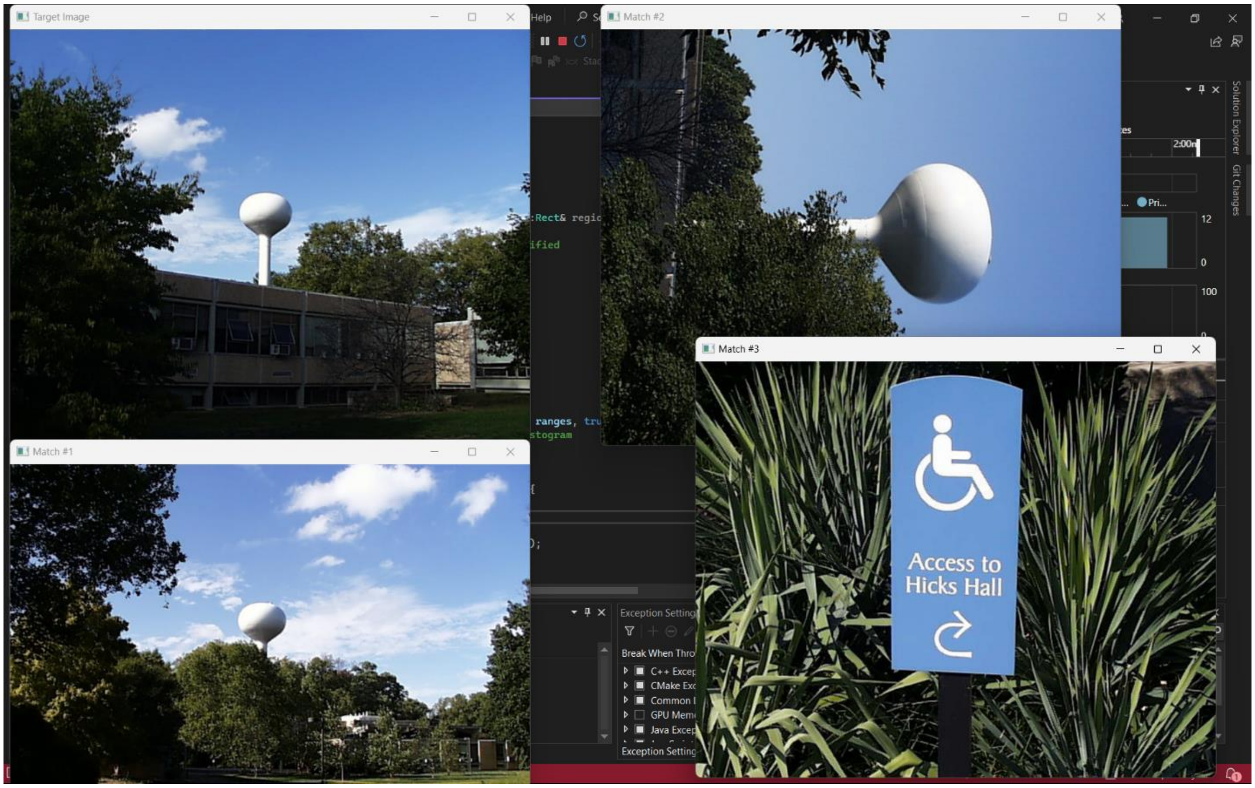


Second Test Case

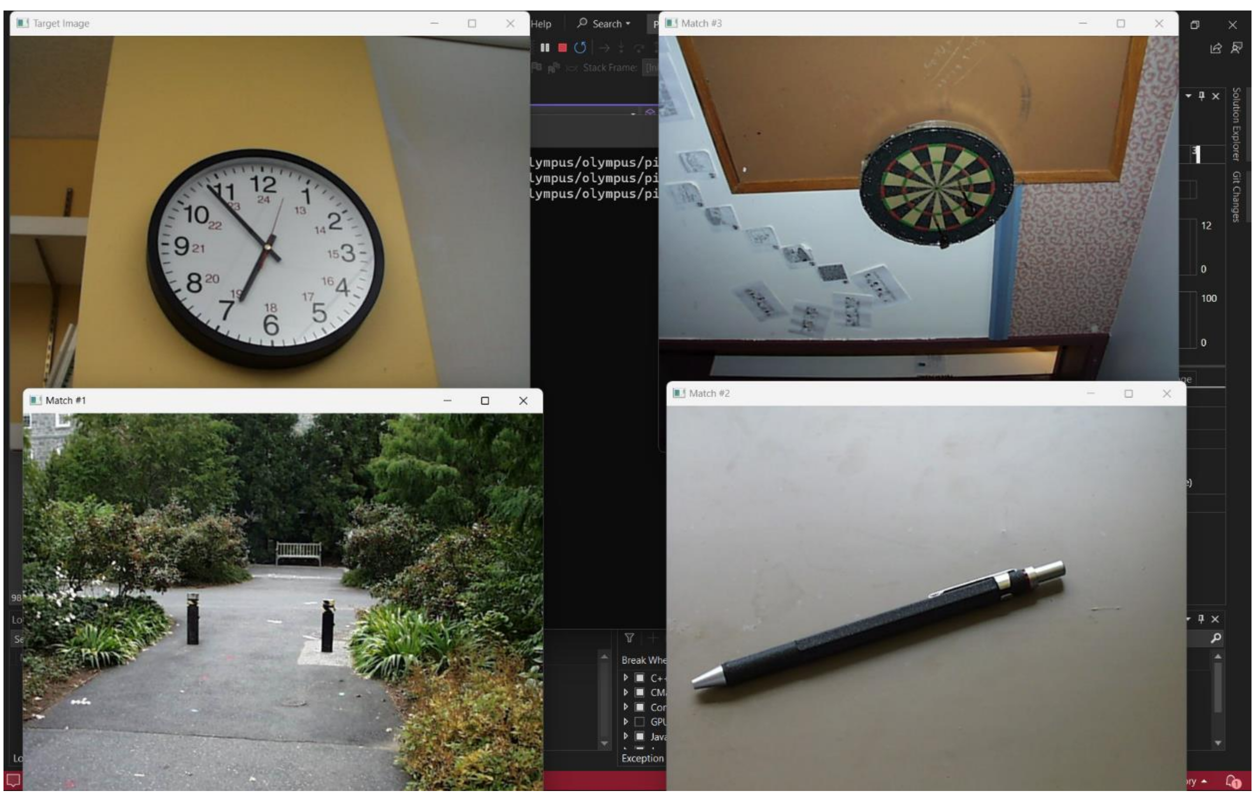


Multi-Histogram Matching

Original Test Case



Second Test Case



Matching Using Texture and Color

Image 535 matched using texture and color

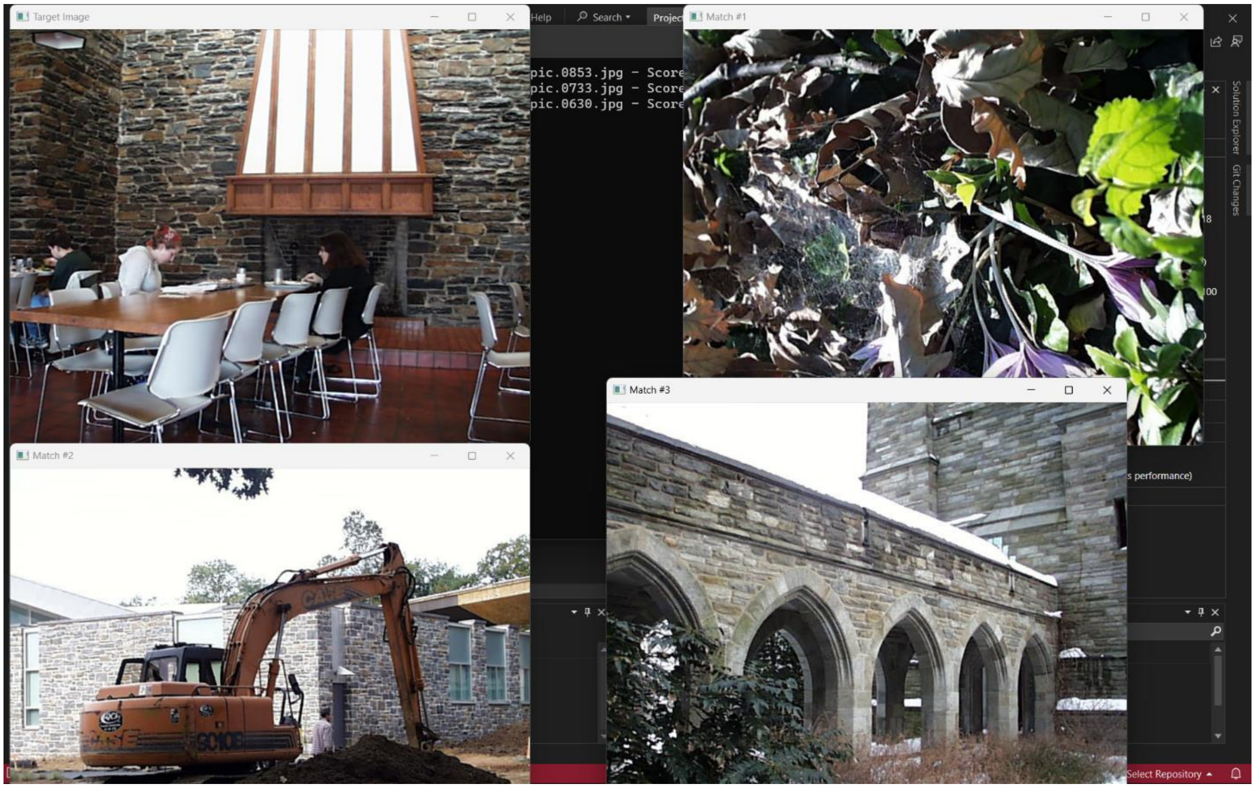


Image 535 using Histogram Matching

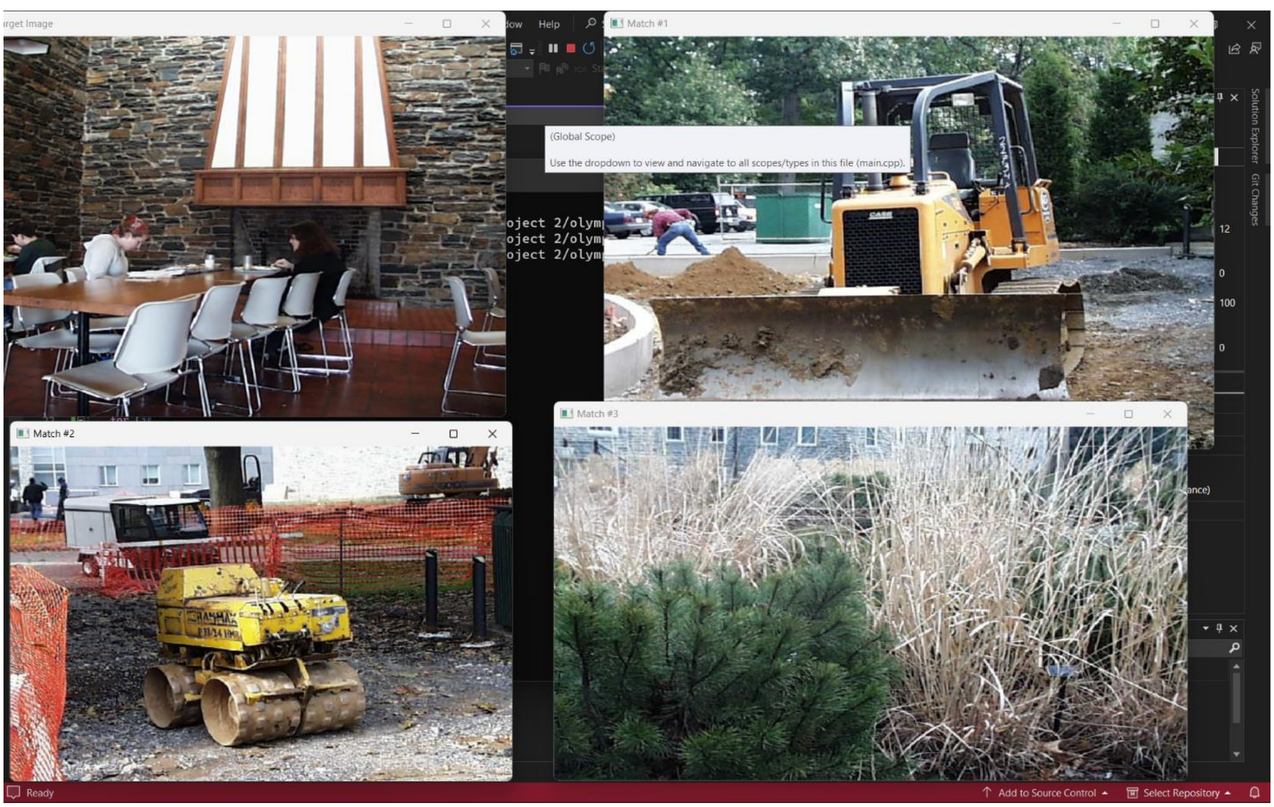
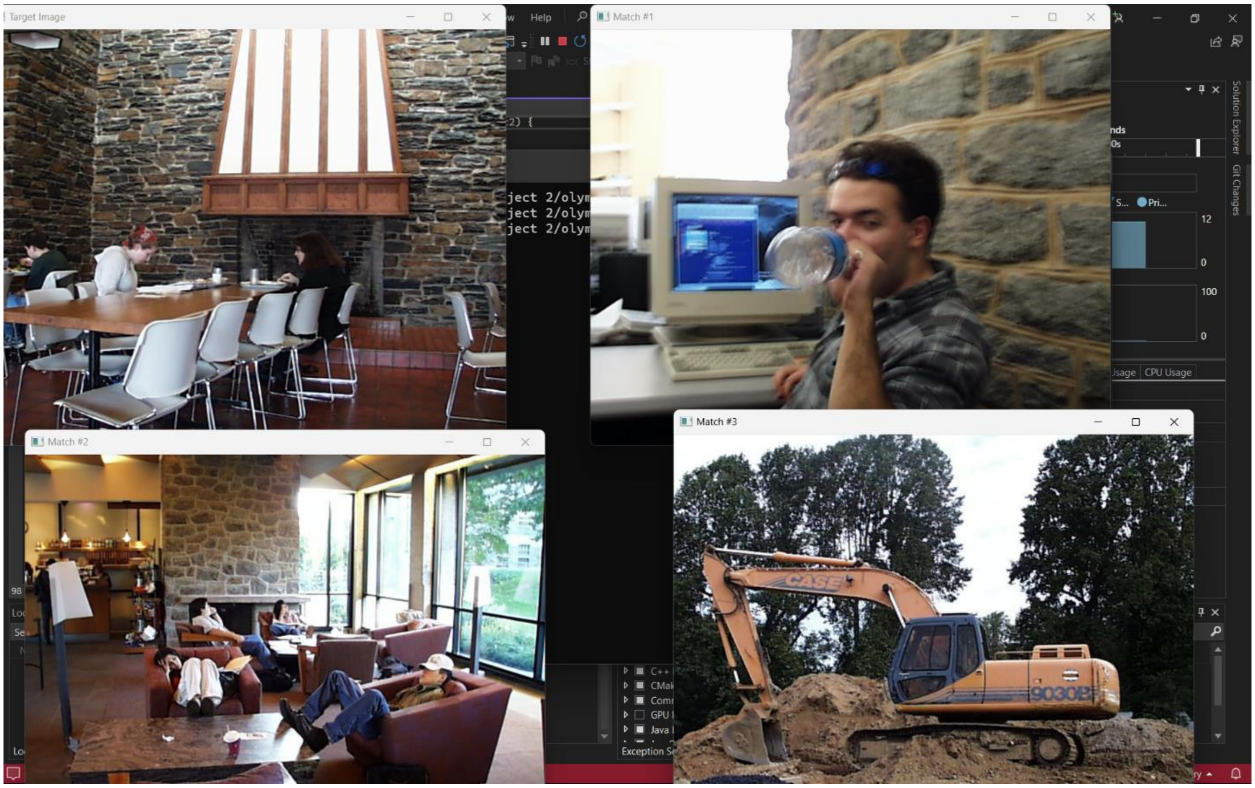


Image 535 using Multi Histogram Matching



Deep Network Embeddings

Image 893 using Deep Network Embeddings

As you can see here object detection is conducted here the algorithm matched images containing fire hydrants making it easier to match and identify them. Different color fire hydrants is also getting matched here.

Image 893 using Baseline Matching

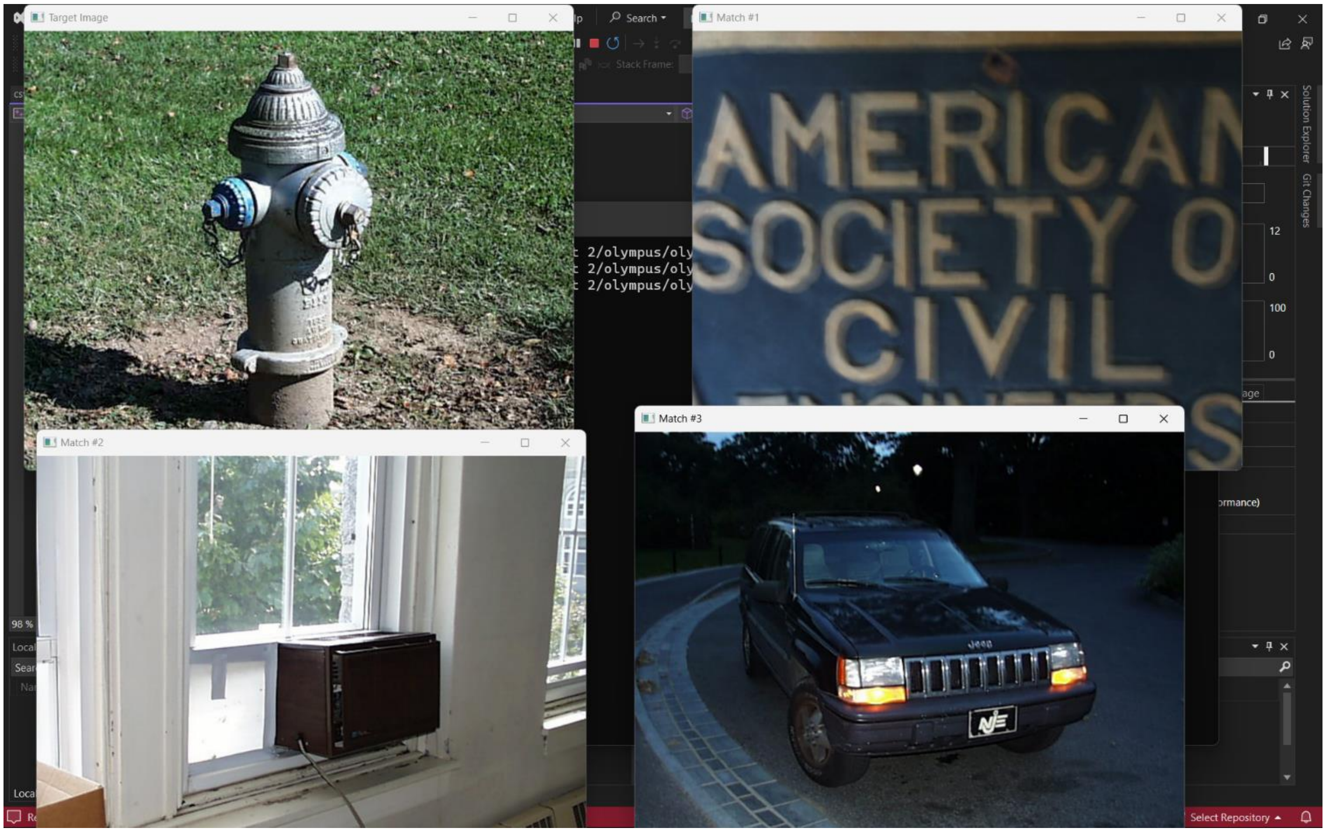


Image 893 using Histogram Matching

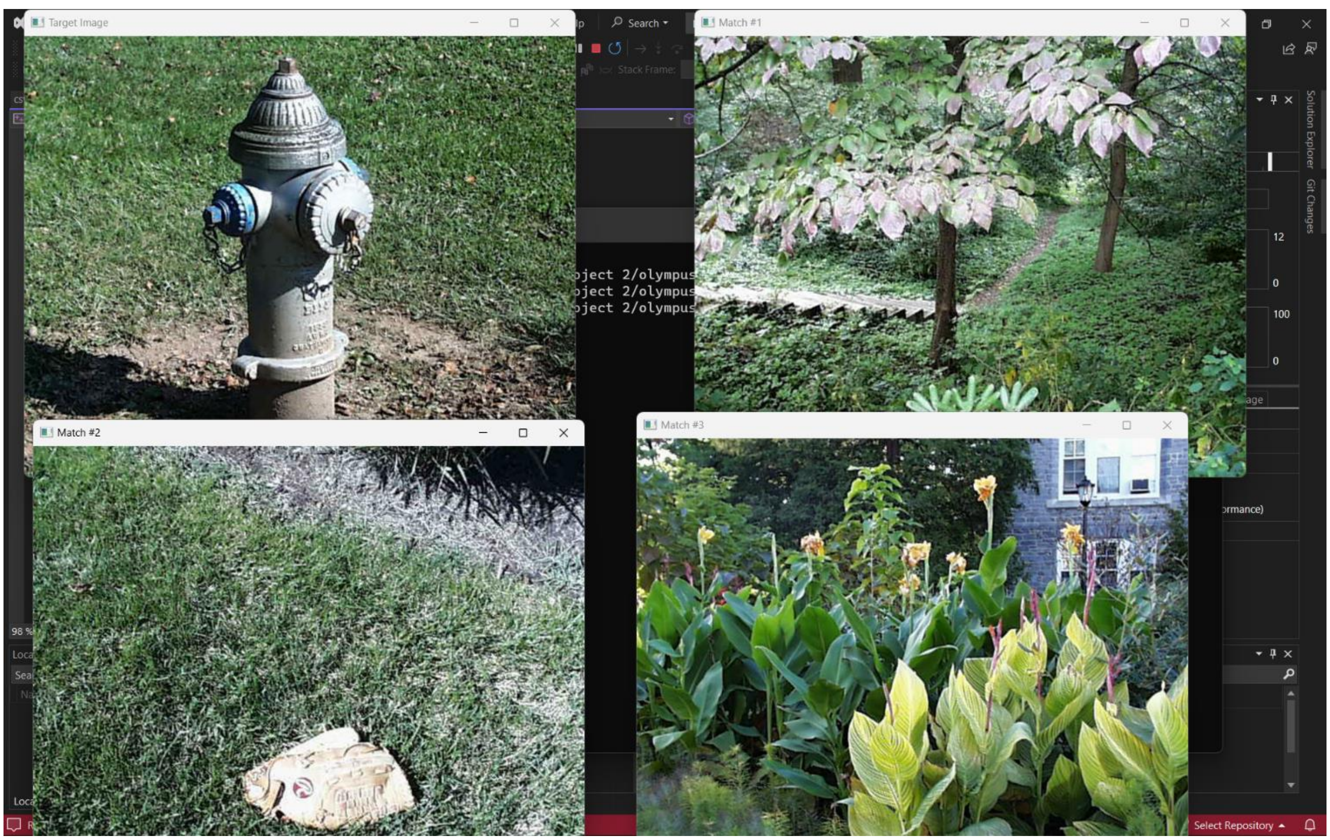


Image 893 using Multi Histogram Matching

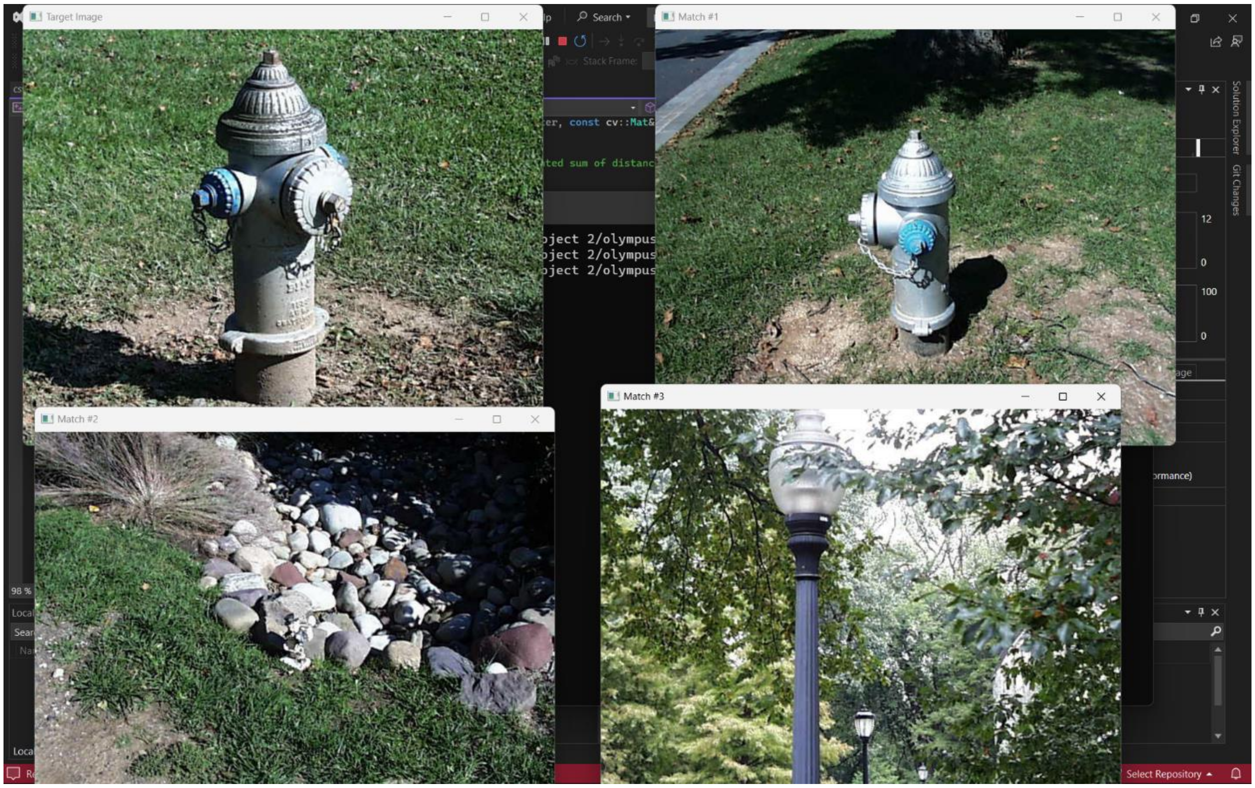


Image 164 using Deep Network Embeddings

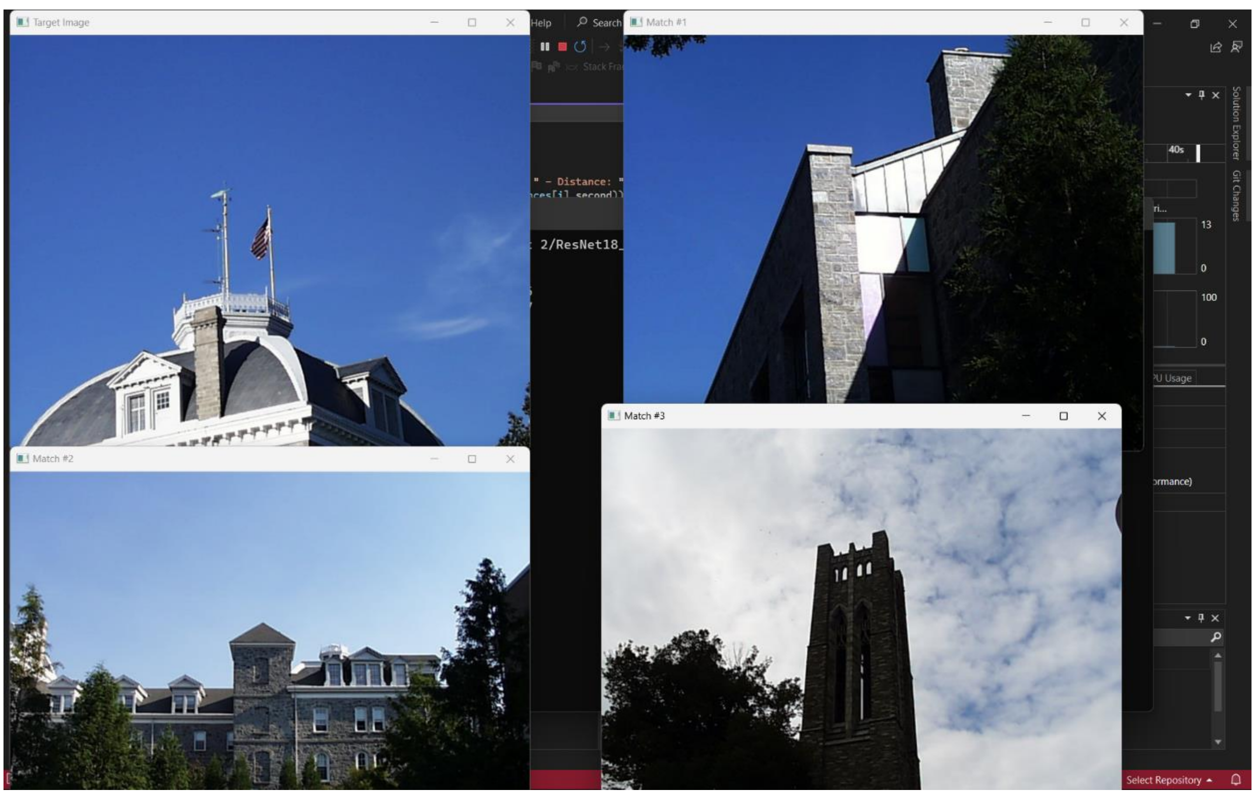


Image 164 using Baseline Matching

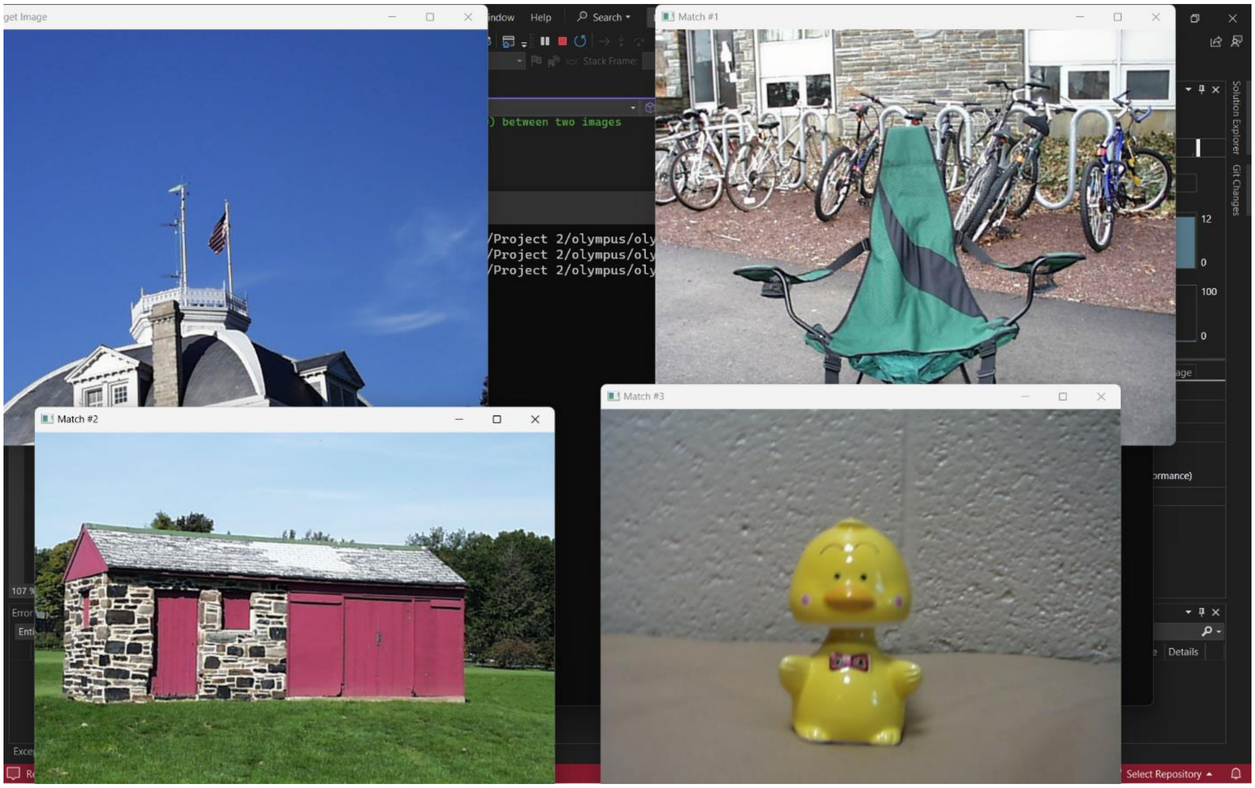


Image 164 using Histogram Matching

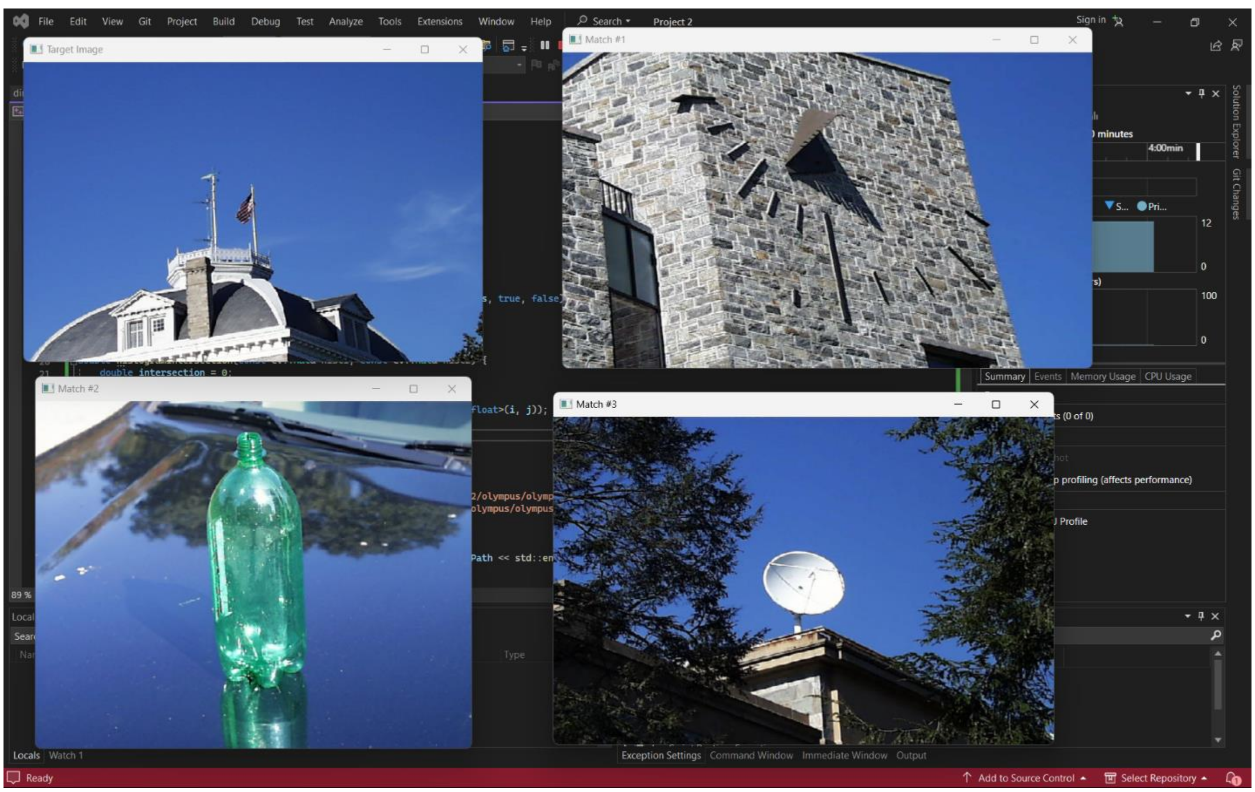
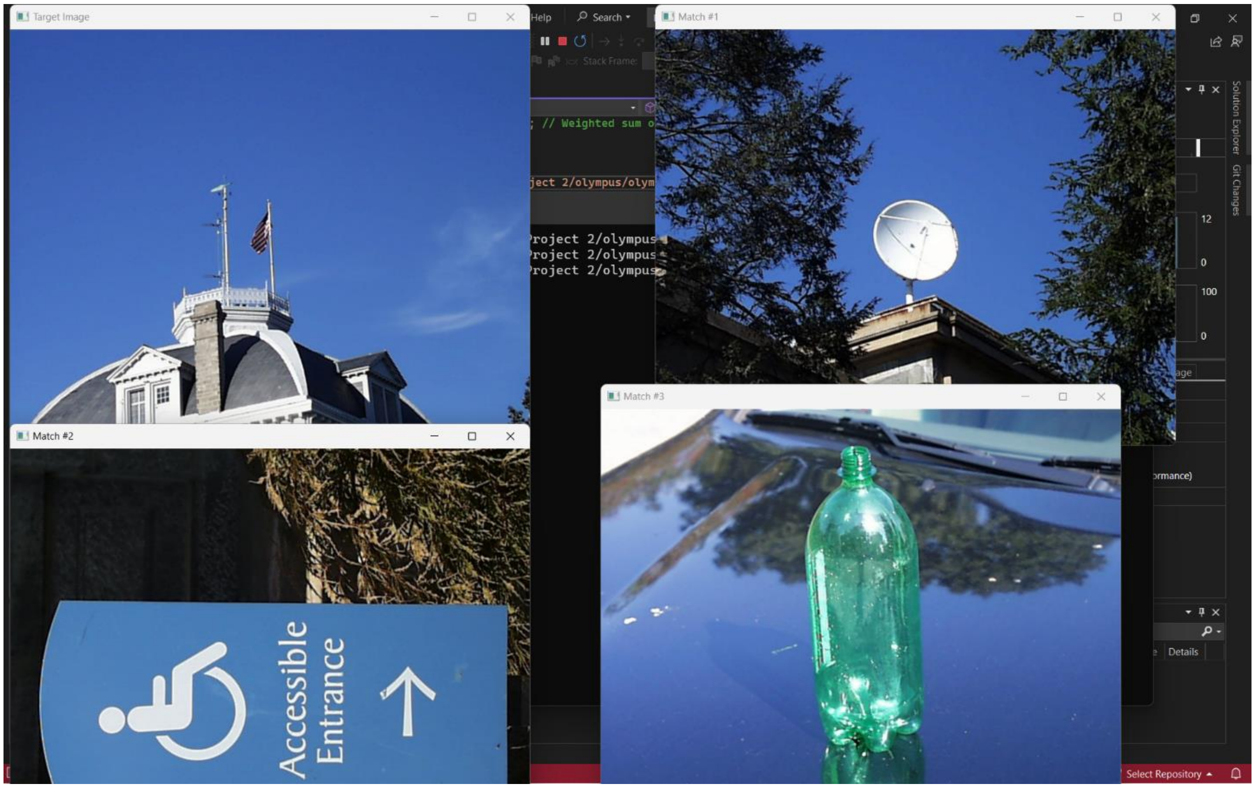
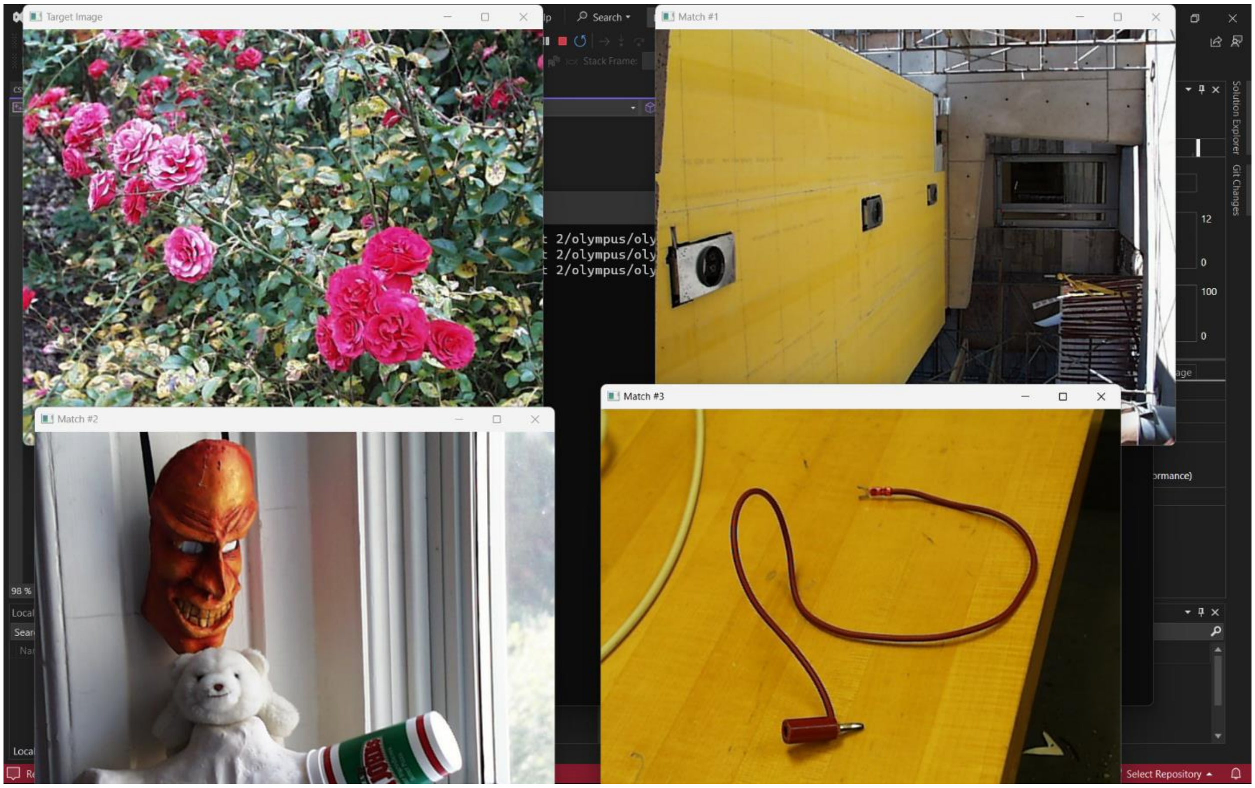


Image 164 using Multi-Histogram Matching



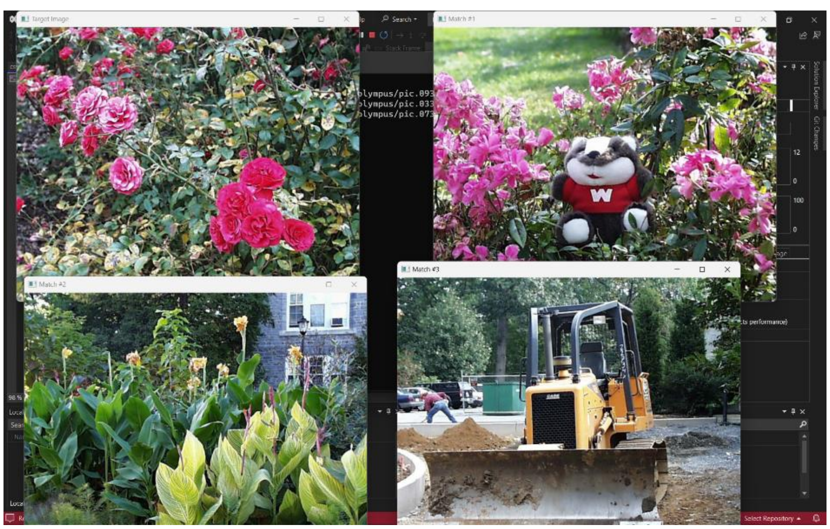
Deep Network Embeddings and Classic Features

Image 1072 using Baseline Matching



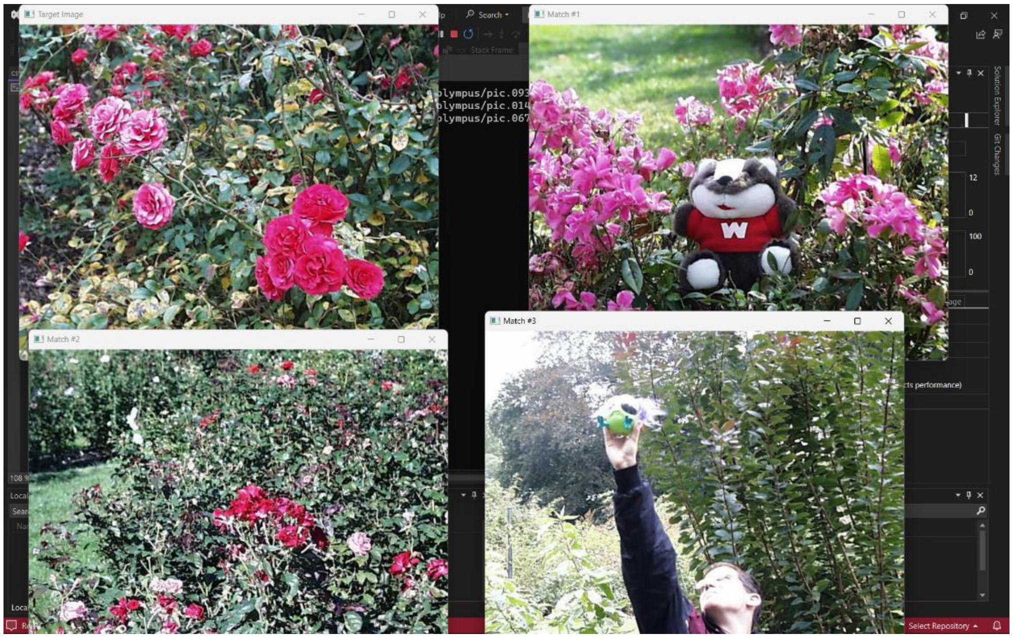
The baseline approach, here using simple color and texture features, has returned images that match some aspects of the target image, such as the presence of red objects and scenes. However, they are not related to the target image of roses.

Image 1072 using Histogram Matching



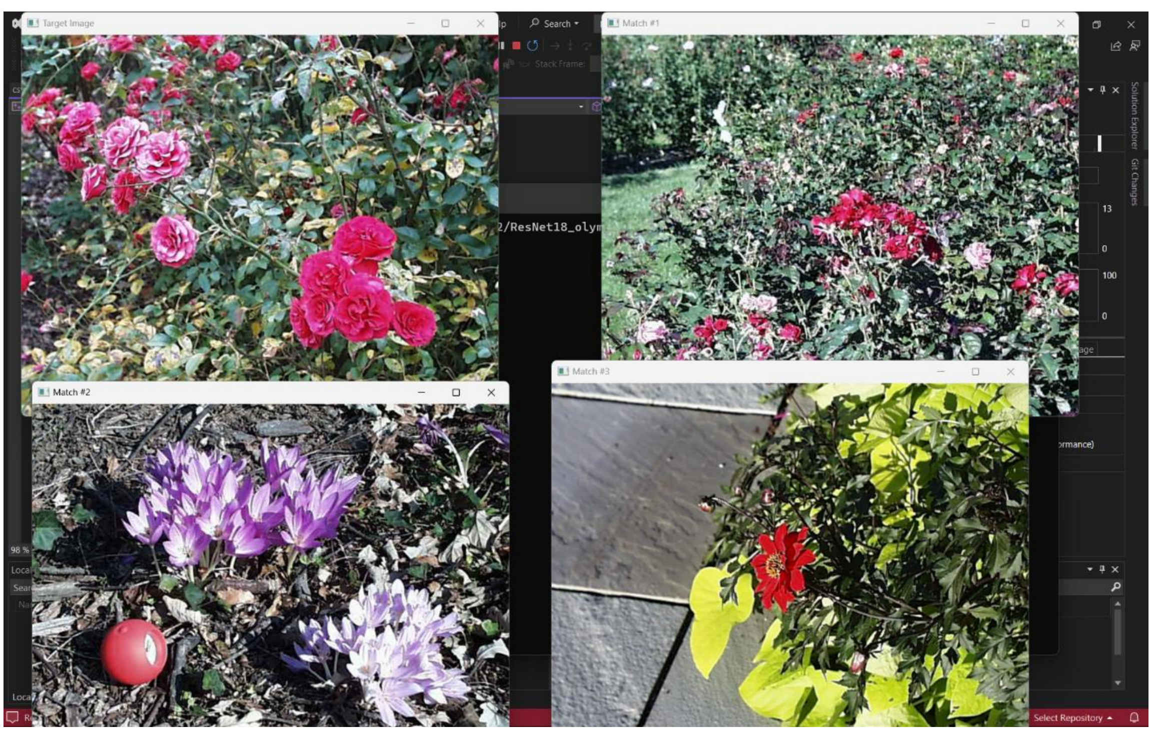
Histogram matching seems to focus more on the color distribution. The images retrieved share a similar color palette with the target image, one common pattern is there are green color and grasses common in these images and all of them are outdoor scenes. But they still miss the context of the scene.

Image 1072 using Multi-Histogram Matching



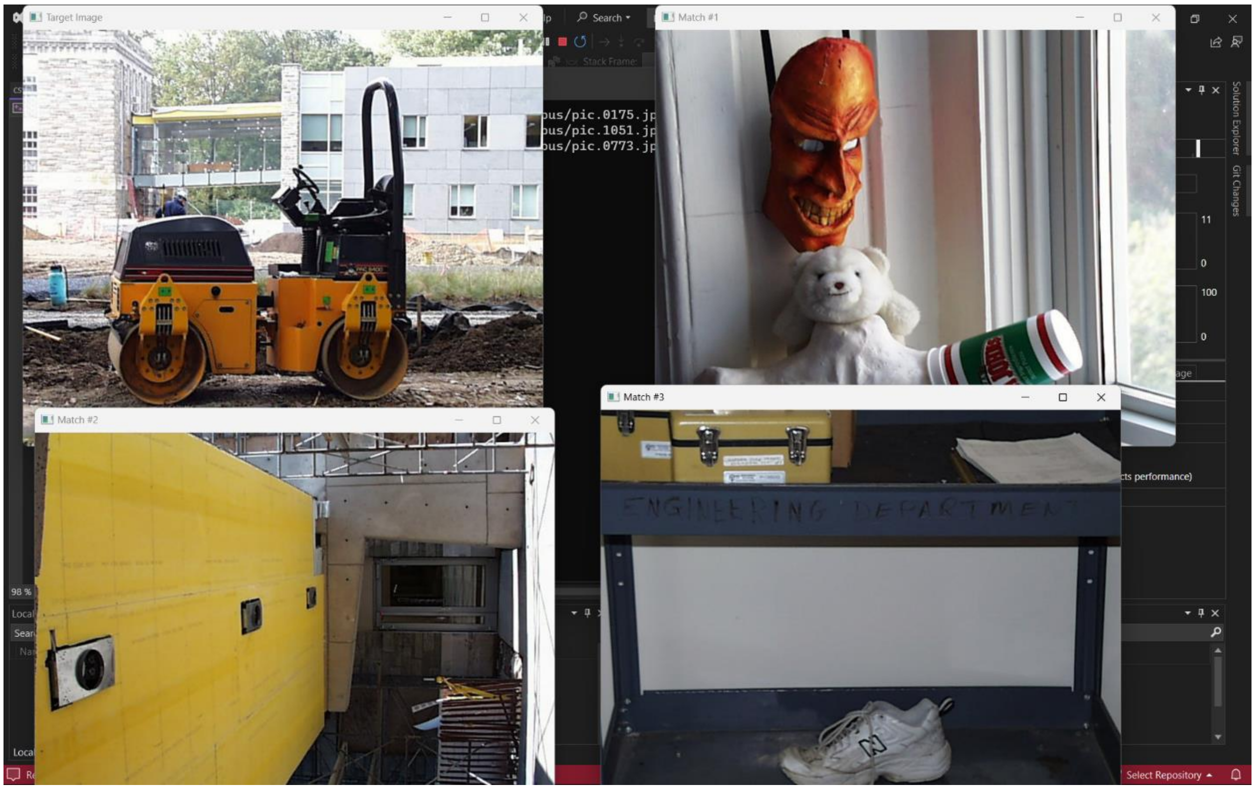
This approach considers multiple aspects of the image, different color spaces and additional texture features. It is more accurate than normal histogram matching. The target image is of roses but the multi-histogram approach here is prioritizing the greenery, possibly due to its larger area in the target image, but is not sensitive enough to the red roses, which are the focus. These matches suggest that the algorithm has picked up on the texture or another feature that's present in both the target and the matched image, but it does not have the color accuracy for the red roses.

Image 1072 using Deep Network Embeddings



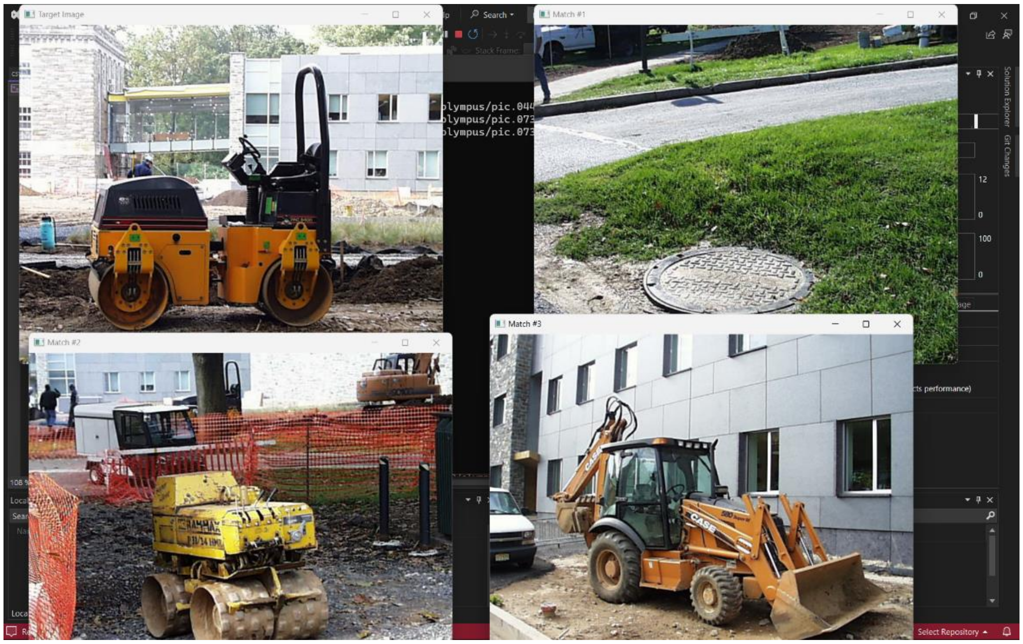
DNN Embeddings captures the scene with red flowers and green foliage, which closely mirrors the red roses against green leaves in the target image. This suggests that the DNN has successfully learned to identify the specific features of roses. Also shows the flowerbed, with different types of flowers. The match still retains the botanical theme, which is a relevant characteristic of the target image. The third match also displays a single red flower among green leaves, which, while not as close a match as the first two images, still resonates with the target due to the color and the single flower focus.

<u>Image 734 using Baseline Matching</u>



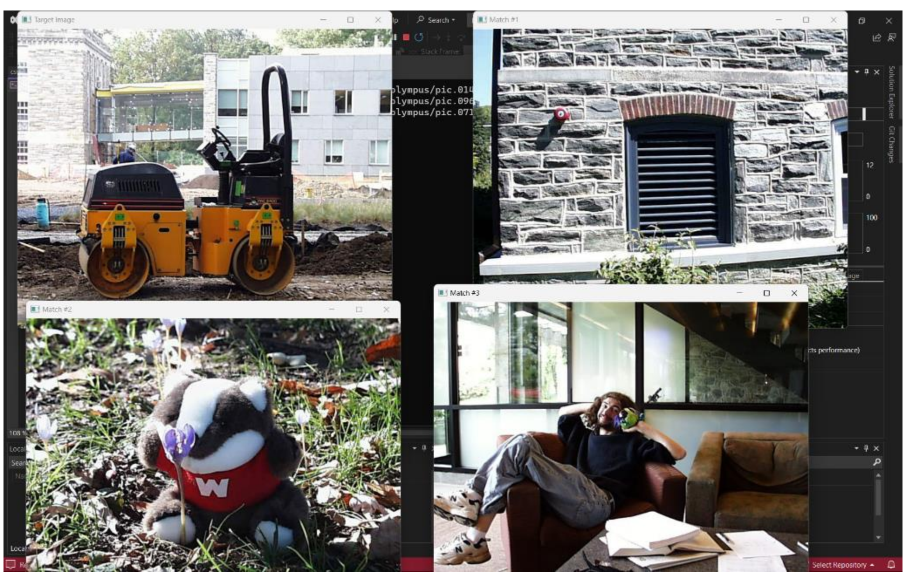
The first match shows an object with a significant amount of yellow, similar to the construction vehicle in the target image. However, the context and overall scene are entirely different, indicating that the baseline matching is focusing on the predominant color. The yellow background in the second match is likely what the algorithm picked up on. However, the subject matter is unrelated to construction or vehicles. The third match displays an indoor scene with various objects, none of which resemble the construction machinery in the target image.

<u>Image 734 using Histogram Matching</u>



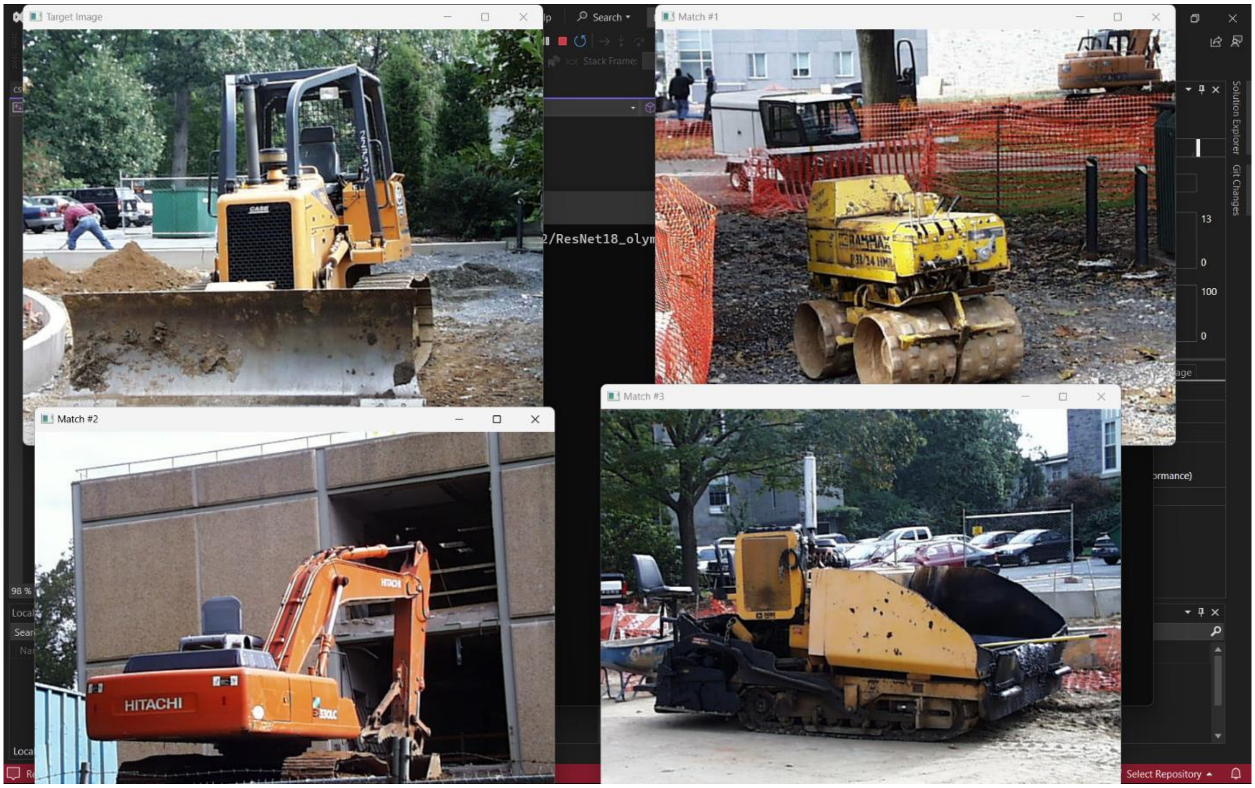
The matches are of outdoor scenes seems to have a similar distribution of colors to the target image, particularly the grays, which is be why it was selected. Features construction equipment, which is contextually relevant to the target image. The color tones of the machinery and the outdoor setting are likely what made the matches. The final match shows a construction vehicle, which is similar in context and function to the target image's subject. The colors, particularly the orange of the vehicle, and the outdoor environment contributed to this match.

<u>Image 734 using Multi-Histogram Matching</u>



Multi-Histogram: In this case multi histogram seems to be highly inefficient. The first matching image shows a stone building with a security camera, which is contextually different from the construction scene of the target. However, the presence of similar grayscale values may have contributed to the match. The second match depicts a stuffed animal in an outdoor setting. The outdoor context is a match, but the content is quite different. The third match is also completely different from the target image.

Image 734 using DNN Matching



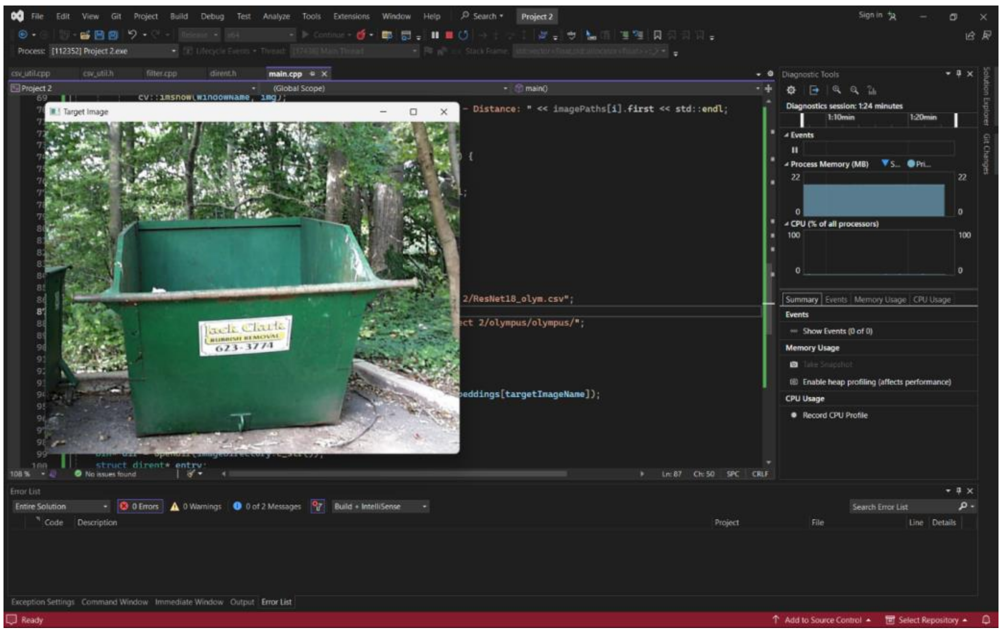
From the above result we can intuitively say DNN Matching functions better than CBIR. DNNs, particularly those pre-trained on extensive datasets like ImageNet, which contains over a million labeled images across a thousand categories, have the advantage of recognizing a vast array of features that span from basic textures and patterns to complex objects and scenes. This enables them to extract high-level semantic information and not just rely on direct pixel comparisons or color distributions. For instance, in the case of the Olympus dataset, if the images within it were part of the training data for the ResNet model, the DNN would have an inherent advantage due to its training on a diverse set of images, which likely includes a variety of environments and objects. This training allows the DNN to be more robust in uncontrolled environments where the background, lighting, and positioning of subjects vary greatly, often rendering classical methods less reliable.

DNN focuses more on object detection while CBIR is more on colors and textures, we can see a massive difference here itself. The above two cases of images show us that the results obtained from DNN matching provided a more matching output to the target image.

Custom Matching

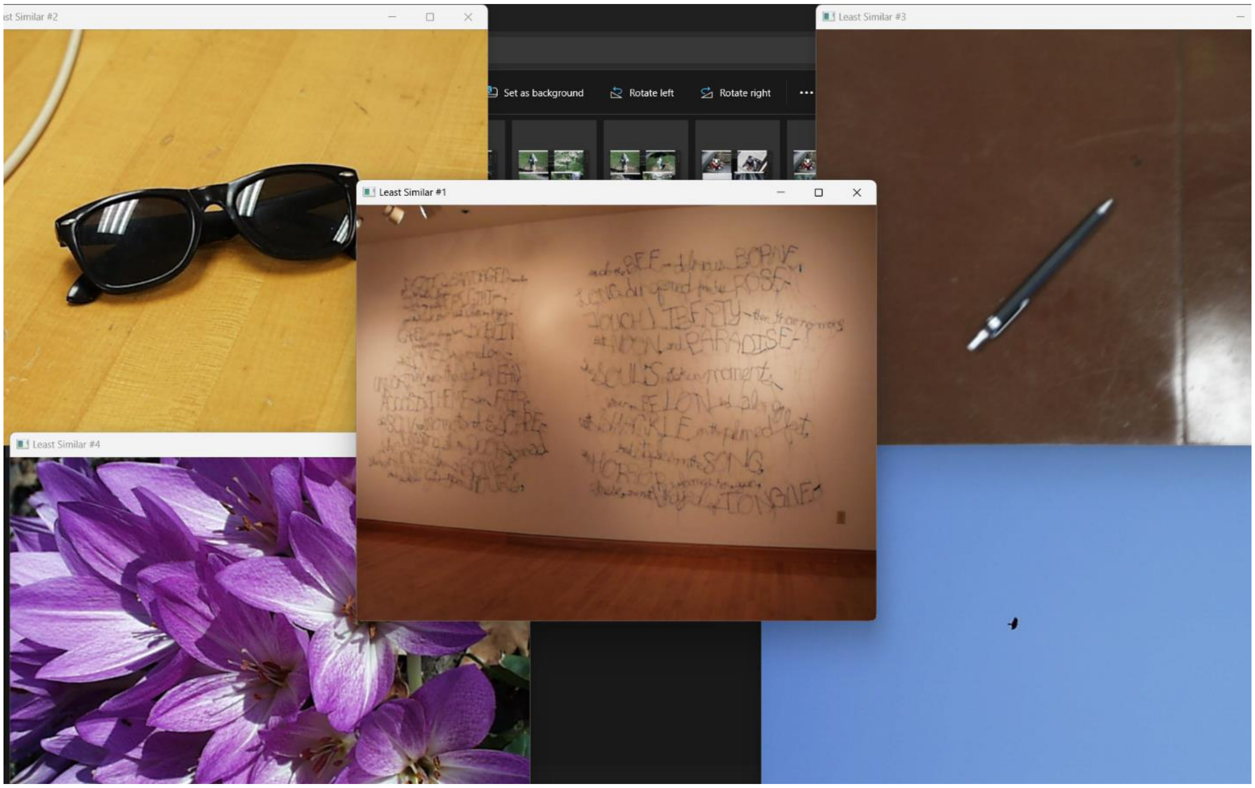
Performs a custom image matching by combining deep neural network (DNN) embeddings and color histogram comparisons. This function first retrieves DNN embeddings for the target image and compares them with embeddings of images in a dataset to calculate cosine distances. Additionally, it computes color histograms for both the target and dataset images to measure similarity based on color distribution. The final similarity score for each image is a combination (average) of the DNN embedding distance and the color histogram intersection. The function sorts all images based on this combined score and returns the top N matches along with a selection of top N least similar matches, effectively utilizing both texture (via DNN embeddings) and color information for content-based image retrieval.

Target Image

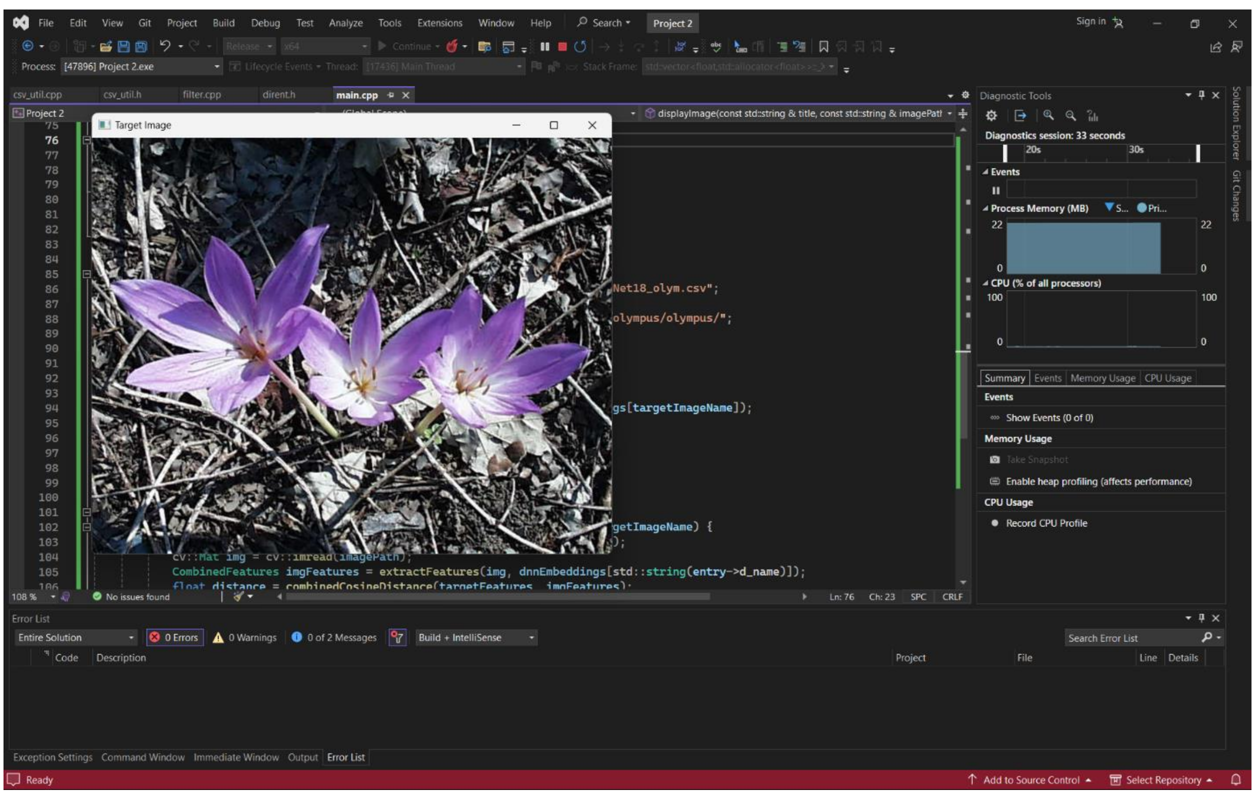


Top 5 Similar Images

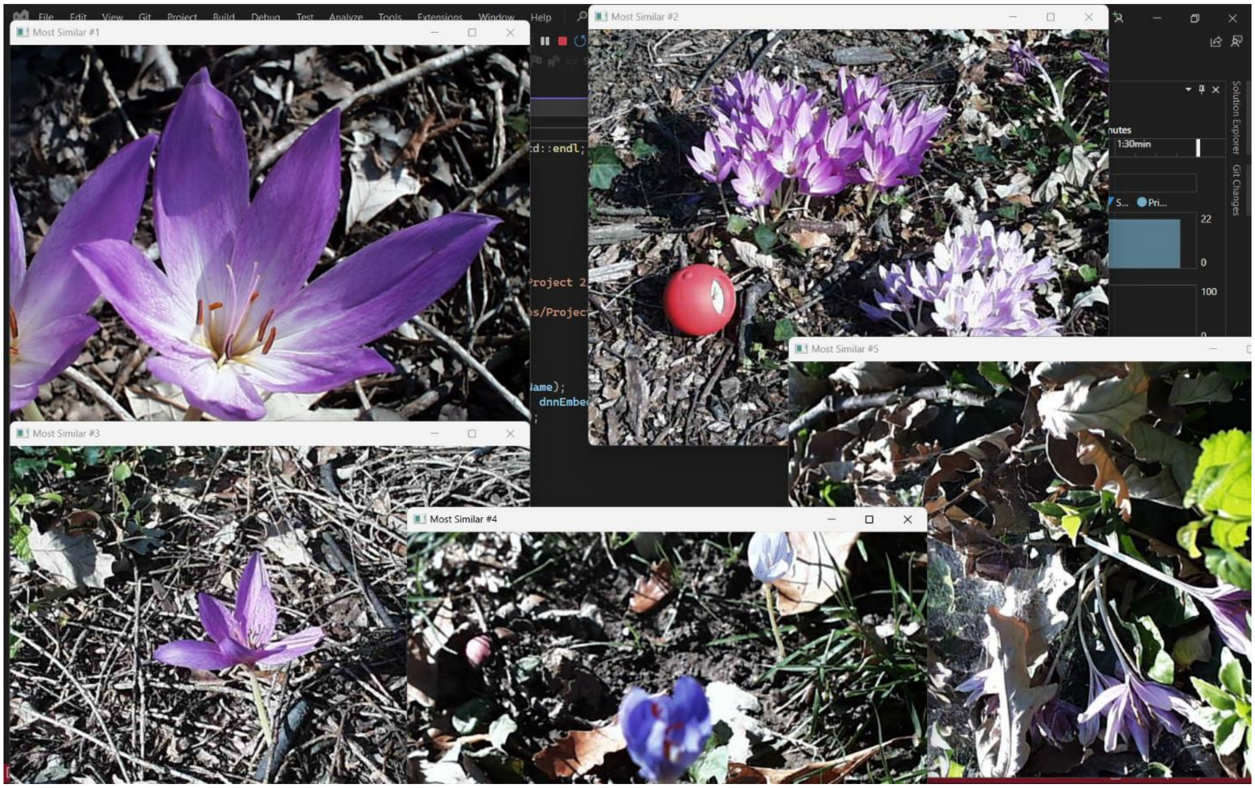
Top 5 Least Similar Images



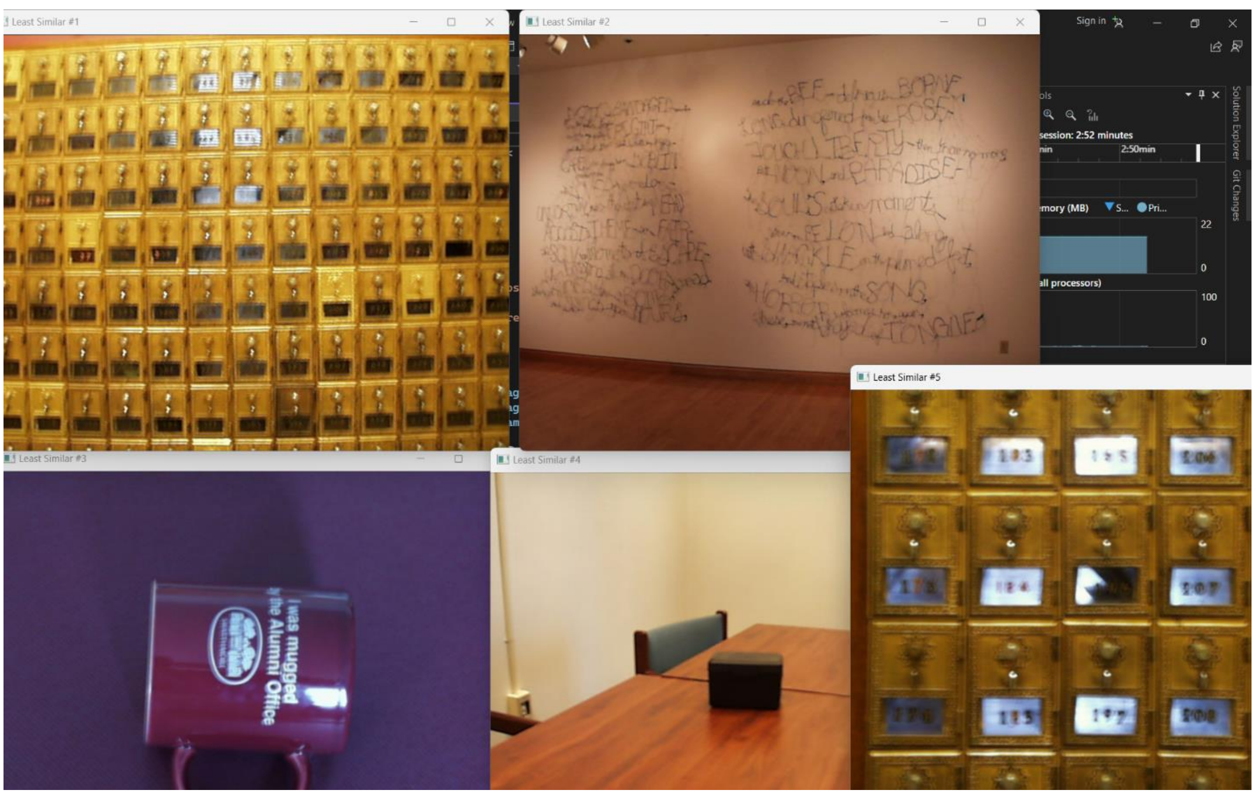
2nd Target Image



Top 5 Similar Images



Top 5 Least Similar Images



What We Learned

This project gave me the chance to get my hands dirty in the field of image analysis and processing, with a special emphasis on content-based picture retrieval. I developed a deeper grasp of how photos may be compared and measured based on their visual content by putting different feature extraction techniques and distance measures into practice. Through trying out several approaches, I was able to understand the subtleties and compromises associated with picture retrieval systems, like the harmony between accuracy and computational complexity. Additionally, I became aware of the ability of pre-trained models to capture complex visual representations while dealing with deep neural network embeddings. All things considered, this project improved my knowledge of programming, image processing methods, and the practical use of machine learning. It emphasized the continuous developments in computer vision and emphasized the significance of testing and assessment in the design of efficient image retrieval systems.

Acknowledgement

1. OpenCV Documentation : https://docs.opencv.org/4.x/
2. BoostMyTools Youtube Channel: https://www.youtube.com/@BoostMyTool
3. Computer Vision: Algorithms and Applications, 2nd Edition by Rick Szeliski