MUHAMMED HALBUTOGULLARI AND XIAOXIAO (ANNIE) QIN

PS3

PART 1

1)

The SIFT key point descriptor is created by first taking a 16x16 box around the keypoint, splitting it into 16 sub-blocks, and then creating an 8 bin histogram of the orientations of the gradients for each sub block. The result is the descriptor, a 128-bin histogram which is represented as a vector.

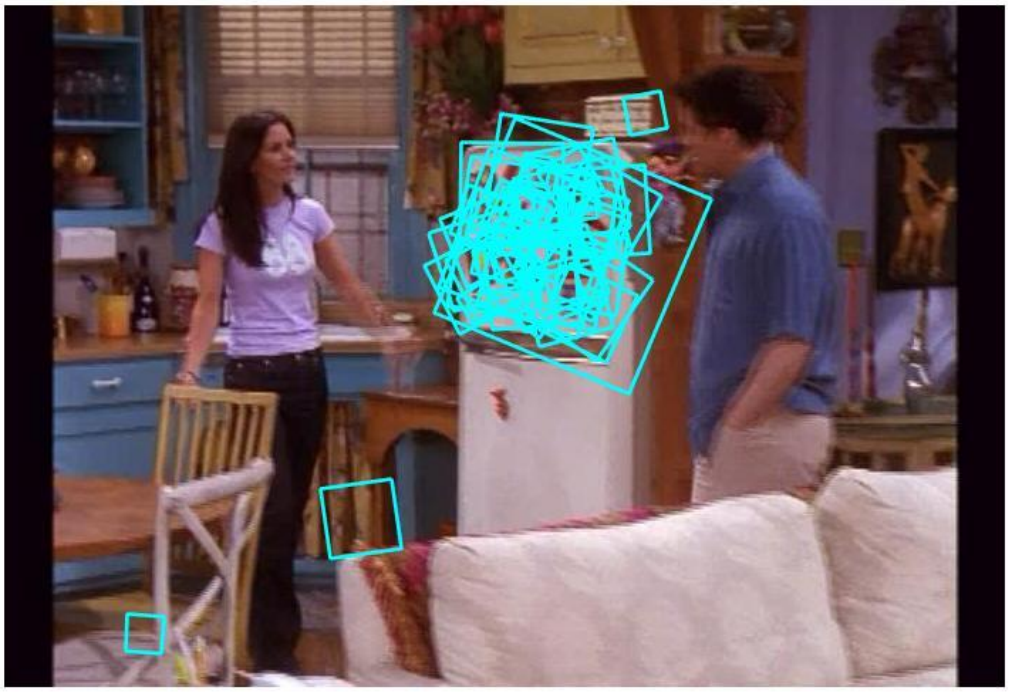
2) If the deep neural network did not have any non-linear activation functions, it would not be able to learn a complex non-linear function. It would only be able to learn a linear function at best. It would still be considered a deep network as long as it had a few layers, however it would be considered linear, and no longer non-linear.

PART 2

1) For this part, in we selected the region on the fridge door in the first image (like the example in the assignment), and this was the resulting matching patches in the second image. Overall our algorithm was exceptionally good (few incorrect matches) at identifying and thresholding the descriptors to concisely match the patches on the fridge from our selection to the patches representing the same fridge in the second shot.

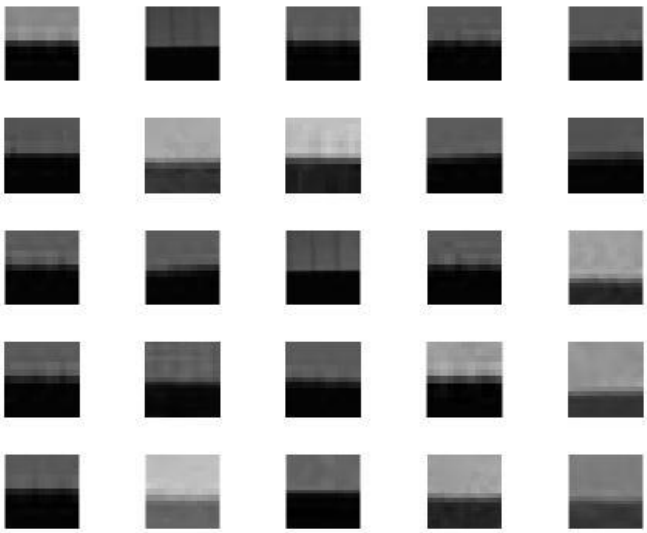
Selected Reigon of img1:

Matching patches in img2:

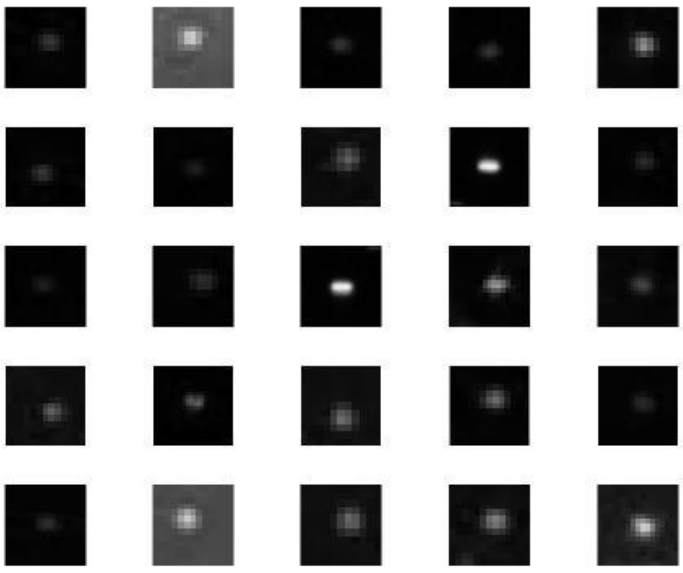


2) We chose the second and sixth most common words in our training set, and found each of their 25 most common descriptors:

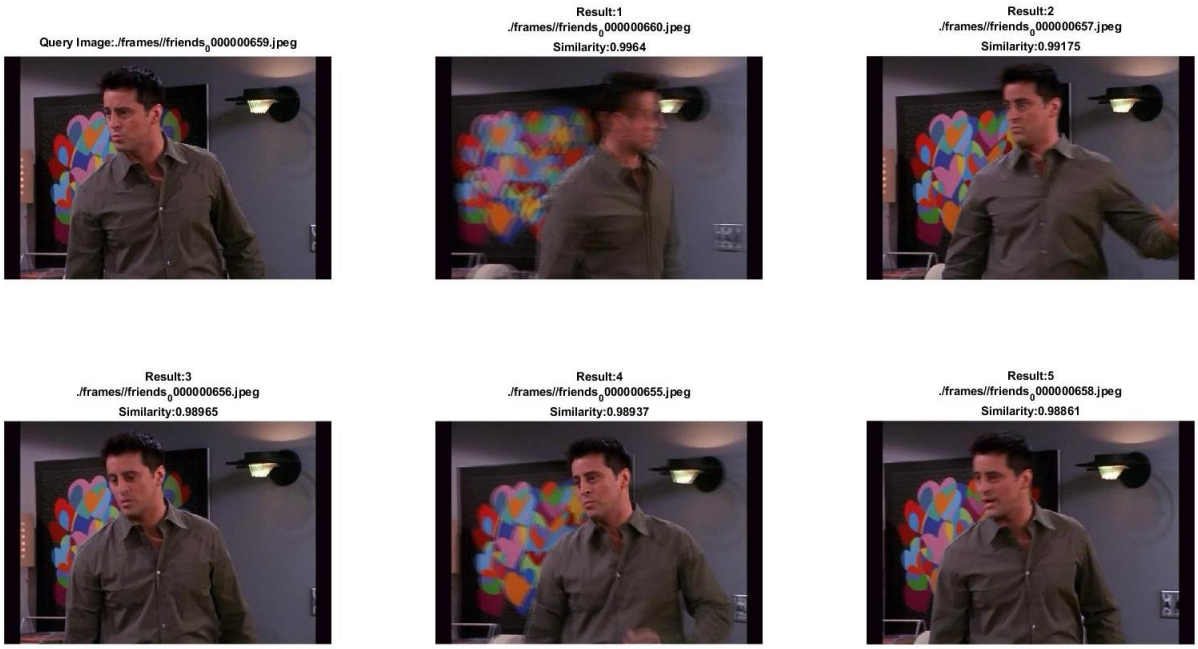
The second most common word was a horizontal edge of sorts. Most of the examples are probably the edges of tables and other objects (especially those with a black background):



The sixth most common word was a small white circle. It is hard to tell exactly what these patches are but they are likely things like shiny spots, light bulbs and other bright spots among darker backgrounds.



3) The first image's matches were good, as the top 5 frames were within the same scene as the query. The matching likely worked well due to the vibrant colors in the background. It was interesting to see the second result match with our query, because it was quite blurry [we wouldn't have expected such a high similarity score from a blurry image (although it is clearly from the same scene)]. This result could have been better if the blurry image was given less of a similarity score than the other matches which are much more "similar" to the query.

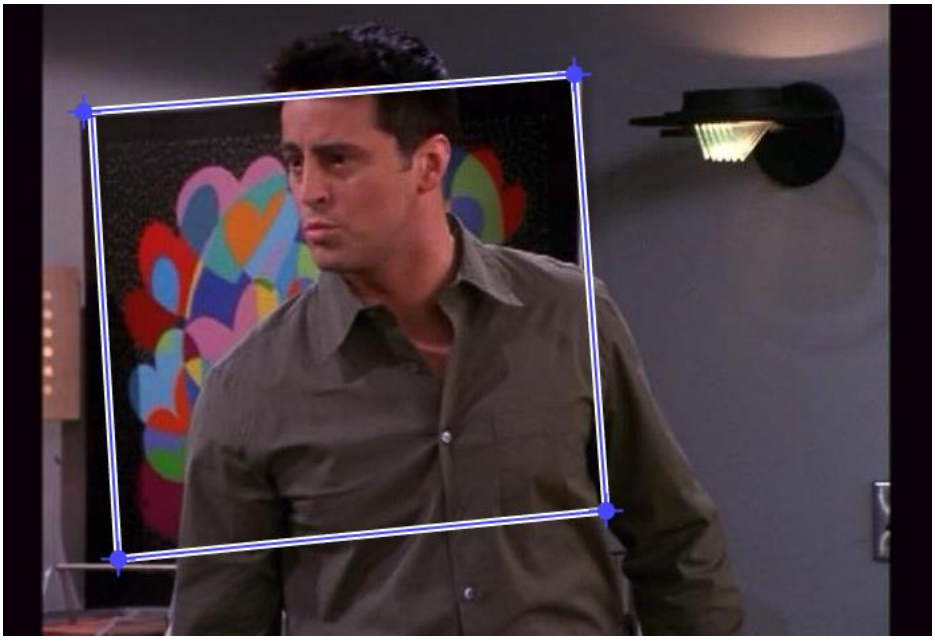


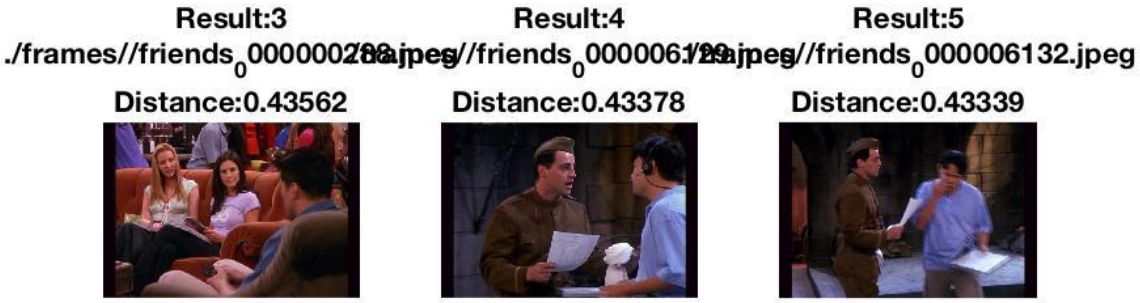
Our second search was also successful. The top 5 results returned are all exactly from the same scene as the query image. The success is likely due to the very uniquely colored shirt vs the plain background in this scene. This is an ideal result for our image searching.

The third query results were the worst of the three as it contained a single mismatched frame. The algorithm worked pretty well for the first few (returns frames from the same scene), and only the final image it matched was a mismatch. It is understandable why this mismatch may have occurred: the women's clothing and the general scene colors of the query frame very closely resemble the patterns and colors of the buildings in the fifth matching frame.

4) Region_query

result 1:





This is our result 1, obliviously, it is much less accurate than the full frame query. We are only take the region of it to compute our similarity based bag of words. However, based on the region, the most similar one captured his face, and others all contains some kind of black and red color that matches the drawing behind the person's head. In addition, the color of the shirt also captured as we can see from the result 4 and 5.

result 2:

Result:1
./frames//friends_0000001112.jpeg
Distance:0.61046

Result:2
./frames//friends_0000000113.jpeg
Distance:0.56492

Result:3
./frames//friends_0000001112.jpeg
Distance:0.55829

Result:4
./frames//friends_0000001110.jpeg
Distance:0.54742

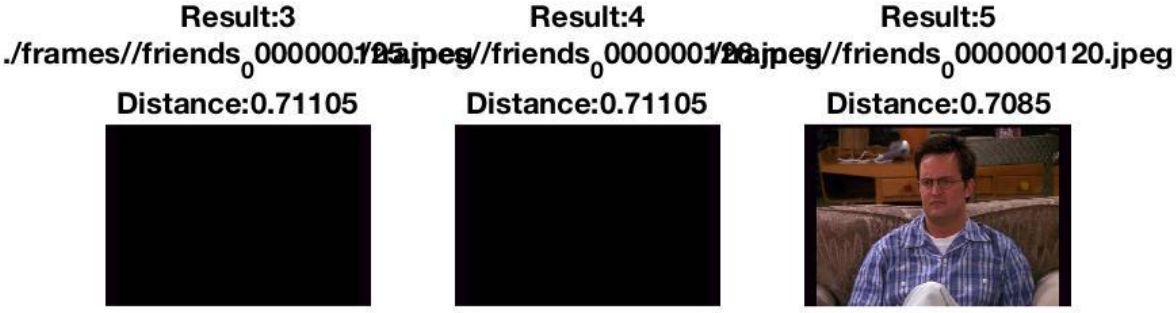
Result:5
./frames//friends_0000000105.jpeg
Distance:0.52678

The result 2 is more accurate. We marked the windows as our region to compute the similarity. The result was a success as every similar frames of ours captured the same scene. In addition, the curtain and person's shirt also captured.

result 3:

Result 3 is also success. We marked the blue shirt as our region query. Every similar frame captured the blue shirt in exactly the same scene albeit different angles. In addition, by comparing the distance/similarity of this result to the results in the previous results, it is obvious that this one is much larger, as this gives the result of 0.7, and others are showing only 0.5/0.4 (similarity score) results.

Result 4:



These results are our least consistent and least successful one. However, given the region we are comparing, is captured the person's face, and as well as the black colors in the background / the women's hair, it is understandable is why it is giving the black images here. In addition, the pattern on the wall behind the person's head is similar to the pattern on the sofa behind the person's back in the result, which means it does captured some details, but because of the small region it is given, it cant do so well, as evidenced in the results.

5)

Our bag of words(above) VS. deepFC7(below) for query image 394

From this result we can clearly see the result that using AlexNet is much better result. The result with our bag of words, we can tell that it defintely has some part similarity. For instance, there is a purple wall in the original picture, and in the similar frame, they all contains some kind of purple wall. Our result is based on the frame's sift value, while the AlexNet was pre-trained on the 1000-class ImageNet classification task, and we are using it to extract the layer-7 activation features for each image. Thus, for our result that using sift to calculated the histogram counts to figure our the similarity, it may only contains partially similar image. While AlexNet is giving us a much better result, what with more parts that are similar to each other, since it is a pre-trained ImageNet.

Query Image 4503:

Our bag of words(above) vs. deepFC7(below)

For this second result, not surprisingly, the AlexNet is also giving a similar result. Our result contains some similarity, for instance, all of them have a red square wall in them. However, our result couldn't consistent with the person's face, and every object in the image. It is comparing the similarity based on the bag of words result. Thus, it may only contains piece by piece similarity. However, the AlexNet pre-trained model could recognize and judge different object so accurately in order to judge the similarity. In addition, from our previous practice of using sift to find similarity, we found that for our implementation, the blurry picture are hard to judge. But with the AlexNet, it doesn't have this problem.