Orion Tran

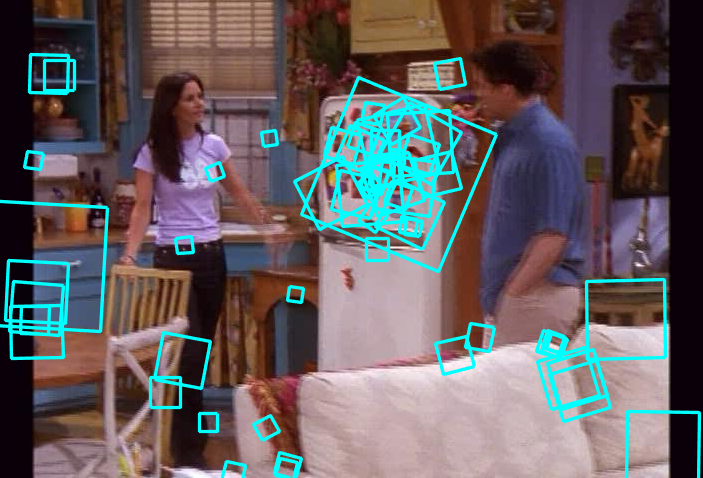
Toby Lee

1. The SIFT descriptor uses histograms to bin pixels within sub-patches according to their orientation. The bins that are recorded in a SIFT keypoint descriptor are part of a histogram that describe each sub-patch in the keypoint descriptor.
2. ReLU makes learning faster, simplifies backpropagation, and avoids saturation issues. From the convolution step, which provides activation maps, the non-linearity step takes those activation maps and non-linearizes them. ReLU, you can’t recover negative values and any positive values will be retained. Model complex non-linear functions, which is essential for deep learning neural networks. Thus without non-linear functions back-propagation will be a lot more difficult. Convolution filters are trained in a supervised manner through the back-propagation of classification errors. Without non-linear activation functions it will no longer be considered a deep network because, we are no longer able to learn, thus defeating the purpose of deep learning neural networks.

Part 2

1.

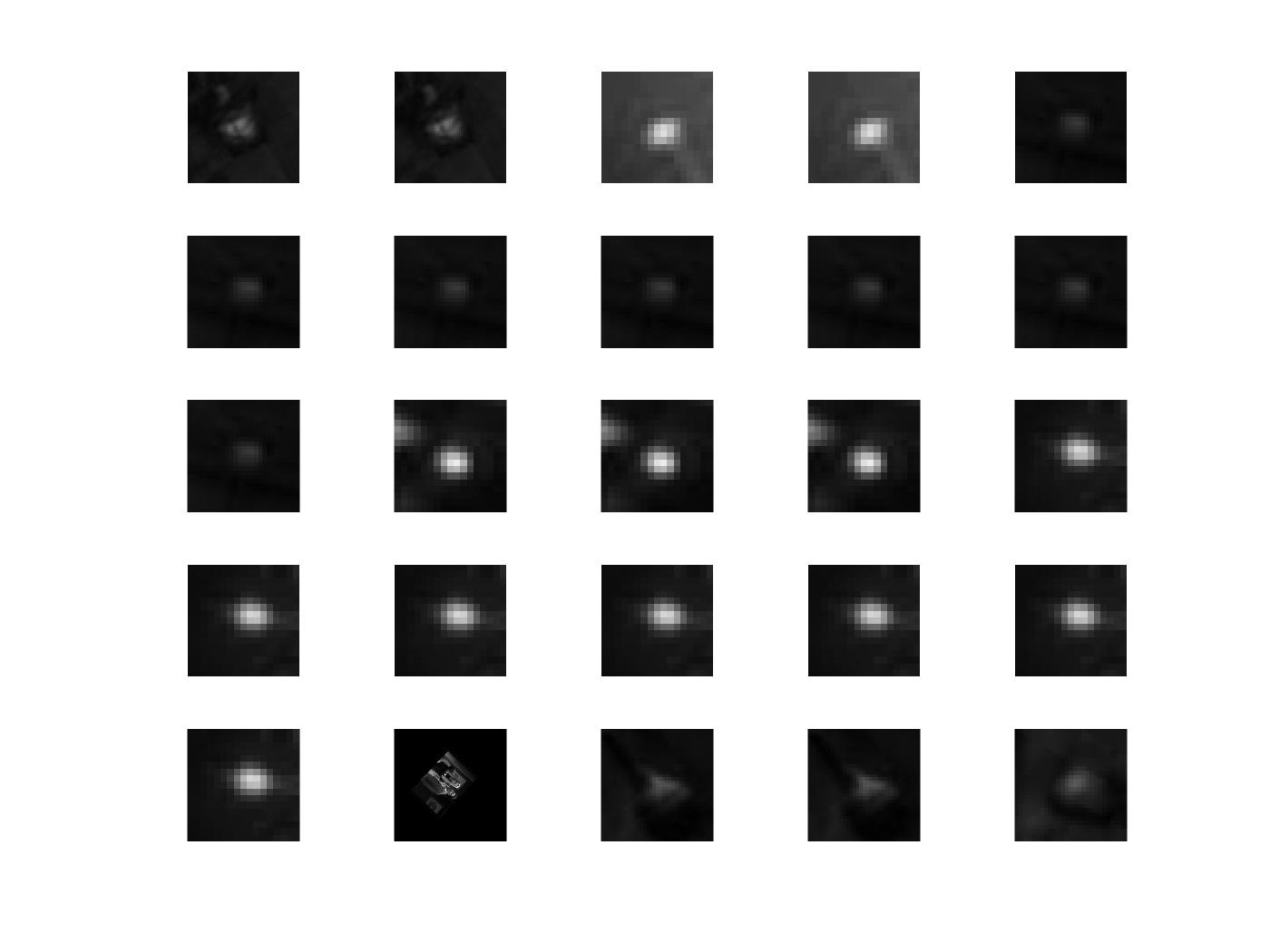




2.

Word 1

My script found two words whose mean centers were as far apart as possible using Euclidean distance. What we got was two words where one is a relatively dark patch with a lightish center and another with slightly lighter color and a horizontal edge. Between the dark and light colors of the two words and the contents of the middle, we can see the two words can be considered quite distinct



Word 2

A close up of a gate

Description automatically generated

3.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

The first and last queries were able to find five images that to a human look very similar. In the case of the first query, the primary differences between the photos is the body position of the turkey headed person, so there were few features that were wildly different between the resulting images. For the third query, the man’s facial expression was the primary difference and there were few enough features there that the normalized scalar product calculation was able to find very similar matchs. The second query, we can see that there are only four really similar images, and in them, the background features seem to be what drove the matches. In the final image, the features of the white robed woman were what drove the matches. It’s possible that this match was formed since the

4.

|  |  |
| --- | --- |
| A picture containing photo  Description automatically generated |  |
| A screen shot of a person  Description automatically generated |  |
| A close up of a person  Description automatically generated |  |
| A screen shot of a person  Description automatically generated |  |
| A screen shot of a person  Description automatically generated |  |

As we can see in the first two examples, choosing a credit line, all our resulting images have a credit line as well. In the first region query, all five of the results have the same credit line in varying backgroun since the features used were likely revolving around the words and not the background. In the second query region, we get three images with the same credit line and two with different credit lines. This is likely in part due to the fact that there are not five images with the same credit line in the frame set and also the features of the background of the query region are affecting the normalized scalar product. In the third set, we see essentially near perfect results where cabinet in the back is found in five different images all with a different foregound (different positions of the man) including one where the region of interest is obscured. Finally, the fourth image was run using a region of just the woman’s face there wer can see the results are not great. This is likely because there are too few features on such a small region that can consistently be found in multiple images. The image did get one correct image first, but the following four are not in correct. It is possible te woman’s complexion might have played a role in affecting the normalized scalar product. We can see in the addition fifth query region using the same image, if we use the whole body where more features containing a pattern are used, the results are much better where only the last image isn’t identical, but still contains the woman and the pattern

5.

First Image

A person looking at the camera

Description automatically generated



A screenshot of a cell phone

Description automatically generated

Second Image

A person posing for a picture

Description automatically generated



A screenshot of a cell phone

Description automatically generated

Given the first query image, the bag of words method was unable to come up with any images that would be considered similar to the first, but looking at the resulting images from the deep learning data, we can see that many of the images have the same subject, the women. Additionally, deep learning images also mimic the darker lighting present in the original query image contrary to many of the bright images resulting from the bag of words images. Similar results are present for the images of the second query image. The subject is maintained in eight of the ten deep learning images while none of the bag of words images depict the same subject. Bag of words results were unable to maintain either the subject or the background and in most cases, both whereas the deep learning images were able to produce results that were very similar to the input image.