In order to run the SVM algorithm, we need to extract features from the data set. Here, we explore some of the classical features used in image recognition. 
- Color Histogram: represents the distribution of colors in an image.
- Histogram of Oriented Gradients (HOG): the technique counts occurrences of gradient orientation in localized portions of an image, which captures shape information. (Figure 1)
- Local Binary Pattern (LBP): LBP gives a histogram of the frequency of each higher illuminance for each patch (Figure 1). It emphasizes texture information within each patch.

We mainly use SVM and CNN models in our implementation. Different features used in image processing are used and compared to the CNN-SVM model.

### Feature extraction

**SVM results on HOG and LBP**

![Figure 2: HOG feature](image2.png)

![Figure 3: LBP feature](image3.png)

![Figure 4: Confusion Matrix for HOG (left) and LBP (right)](image4.png)

**CNN-SVM**

Given the small dataset, we choose to fine-tune the CNN model of BVLC Reference CaffeNet. The architecture of CNN model is shown below.

<table>
<thead>
<tr>
<th>Layers</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>fc7 relu7 drop7</td>
<td>InnerProduct ReLU Dropout</td>
</tr>
<tr>
<td>fc6 relu6 drop6</td>
<td>InnerProduct ReLU Dropout</td>
</tr>
<tr>
<td>conv5 relu5 pool5</td>
<td>Convolution ReLU Pooling</td>
</tr>
<tr>
<td>conv4 relu4</td>
<td>Convolution ReLU</td>
</tr>
<tr>
<td>conv3 relu3</td>
<td>Convolution ReLU</td>
</tr>
<tr>
<td>conv2 relu2 pool2 norm2</td>
<td>Convolution ReLU Pooling LRN</td>
</tr>
<tr>
<td>conv1 relu1 pool1 norm1</td>
<td>Convolution ReLU Pooling LRN</td>
</tr>
</tbody>
</table>

Table 1: Architecture of CNN model

We fixed the first fifth layers and fine tune on the last two layers. We choose the model after 3000 iterations. The learning accuracy curve on Validate Set is shown below.

![Figure 5: Accuracy Curve](image5.png)

To combine CNN and SVM, we extract the sixth layer from the model as the input features in SVM. The result (confusion matrix on the test set) is shown below.

![Figure 6: Confusion Matrix using CNN features: fc6 no fine-tuning (top left) fc7 no fine-tuning top right) fc6 fine-tuning (down left) fc7 fine-tuning (down right)](image6.png)

### Conclusion

Comparing the confusion matrices generated using different features, we observe that:
- HOG feature can not distinguish between different food with similar shapes, such as chicken and bread, while these categories can be well separated base on their different textures.
- Compared to the standard image classification features, the CNN features result in a net improvement in prediction. Although features like HOG and LBP provide local information about the image, they assume fixed patch size thus might not capture all the information available; CNN deals with this problem by providing a more flexible framework.

### References


### Acknowledgements

We’d like to thank Intel Labs Pittsburgh for the publication of the food image data set. We also want to thank all CS229 staff for their support during the quarter.