EE 58J | Data Mining for Visual Media | Spring 2020
Assignment II

Product Image Recognition Challenge 2020
Part II

Here is the second part of the Product Image Recognition Challenge 2020 organized in the context of the EE 58J course. The challenge will continue during the whole semester and consists of experimenting with some standard visual descriptors and statistical classifiers for the task of automatically recognizing the label of a product package from its image. Typical examples of such images are shown below.

In this assignment, you will use only the **Confectionery** Subset of the Vispera-SKU101-2019 dataset (©Vispera Information Technologies), which contains 101 different product SKUs (stock keeping unit as called in retail business, these stand for object classes in visual recognition terminology). The number of classes and the associated number of image instance are tabulated below. In the dataset folder structure, each class instances are placed under their corresponding folder. The dataset is not separated into training and test sets. In your subsequent experiments in this assignment, use 80% of the whole dataset as training instances (roughly 80 instances/class) and the remaining 20% as test instances (roughly 20 instances/class). Note that you should randomly sample each class to split the set into training and test subsets.

Product Category	**Nr. Classes**	**Nr. Instances**
Soft drinks-I	20	2030
Household	20	2081
Confectionery	**20**	**2065**
Ice cream	21	2140
Soft drinks-II	20	2045

Tools. You will use the RFS filter bank for feature extraction (m-file to generate the filter bank can be downloaded from here). The RFS filter bank consists of 2 anisotropic filters (an edge and a bar filter, at 6 orientations and 3 scales), and 2 rotationally symmetric ones (a Gaussian and a Laplacian of Gaussian) as shown in the next page (38 filters in total).

Preprocessing. Prior to the experiments, don't forget to size-normalize images to 128-by-128 pixels.

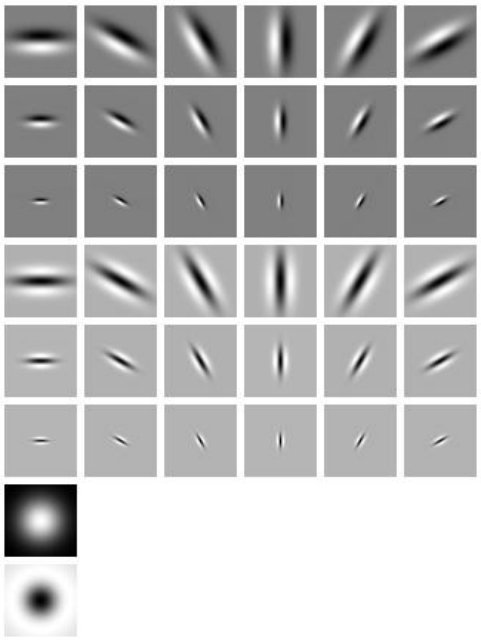
Feature Extraction. Step-by-step instructions for feature extraction are as follows:

(1) Filter (convolve) each of the three RGB channels with each of the 38 kernels in the RFS filter bank. You will obtain 38 filters x 3 channels = 114 feature maps (each of size 128-by-128).
(2) Apply the Rectified Linear Unit non-linearity ReLU(x) = max(0, x) to each pixel x in each feature map.
(3) Spatially pool each distinct non-overlapping 8x8 windows using the local max operator in each window (that is, just keep the maximum value within the window) in a given feature map. This will reduce the size of the feature map to (128/8)x(128/8)=16x16=256 features. Apply this to all feature maps.

At the end of these steps, you will obtain a feature set of dimension 256x38x3=29184 for each image.

Classifier. Train a Minimum Distance Classifier and a Naive Bayes Classifier on top of the original 29184 features of the previous step. Experiment with different density estimation methods (simple parametric estimation with a chosen model, histogram or kernel density estimation) for the Naïve Bayes Classifier.

RFS Filterbank



Reporting. You will report the performance in each case in terms of classification accuracy measured on the testing set. You may also experiment with different spatial pooling size options (instead of 8x8) in feature extraction. Classification accuracy is simply defined as the ratio of correctly classified testing instances to all available testing instances. You are required to provide neatly readable performance data tables letting the user understand the effect of changing a parameter. Don't try to dump all data into a single table unless you

found a good way to present it. You may use multiple tables as necessary. Think of nice visualizations, which can help the reader interpret the results.

At the optimal performance point, you're also asked to report the confusion matrix depicting the confused classes (search your resources for a definition of the confusion matrix). You may want to provide a picture-like heat map as the confusion matrix due to the high number of classes.