**Design Document**

**Nov. 26, 2018, Francis:**

Limited the number of images used for classification to 2000 since the smallest category for testing contained 1751 images (non-motorized\_vehicle). This results in much faster computing (up to 75% faster) and produces a better confusion matrix. This is due to certain categories accounting for a much more important proportion of images compared to some other categories.

Once the features array is divided into k=10 subsets, for one subset at the time, the classifiers are built using the other 9 subsets, the predictions are computed, and the evaluations are done. The accuracy, precision, and recall for every category are calculated using the formulas given. Then, the average accuracy of the subset is computed. Finally, the average accuracy for all the subsets is computed and the confusion matrix is obtained on the last subset. Since only one confusion matrix is needed, it doesn’t matter which subset it is obtained from. Updates are being displayed as the computations are done to allow the user to keep track of the progress.

Implemented option to load or build / save classifiers. This allows to skip the building of classifiers when it is not necessary, and since that’s the part which takes the most time, it can help a lot.