*Localization and Classification*

Nayem Alam   
ECSE415 – Intro to Computer Vision  
McGill UniversityMontreal, Canada  
260743549

Thomas Philippon  
ECSE415 – Intro to Computer Vision  
McGill UniversityMontreal, Canada  
260747645Tristan Bouchard   
ECSE415 – Intro to Computer Vision  
McGill UniversityMontreal, Canada  
260747124

Shawn Vosburg  
ECSE415 – Intro to Computer Vision  
McGill UniversityMontreal, Canada  
260743167Alex Masciotra   
ECSE415 – Intro to Computer Vision  
McGill UniversityMontreal, Canada  
260746829

*Abstract*— The final project allowed us to learn and implement machine-learning related algorithms to analyze a specific dataset. The two problems that were addressed are classification and localization. The classification task required is to extract features to classify the type of vehicle within an image. The localization task required us to pick objects of interest from a large image to classify them. The tools used in this project were a Jupyter Notebook with python3 programming language, and the MIO-TCD image dataset.

Keywords— Classifier, Localization, Machine-learning, SVM, Cross-validation

# Introduction

The purpose of this project was given to exercise the skills acquired from previous assignments to building a classification system using several algorithms, namely the SVM classifier & K-Nearest-Neighbor. The reason why these two methods were selected was because SVM was required by the project document & K-Nearest-Neighbor was implemented due to its simplicity. In order to initiate the project, students were paired in groups of 4-5, provided a source for acquiring a dataset [1] and implemented code within an open-source web application, Jupyter. Students were given 3 weeks to complete this project.

# Classification

## Part 2 – Data and Features

**\*\*\* From the teacher: Discuss the methods in detail; your goal is to convince the reader that your approach is performing the way you claim it does and that it will generalize to similar data.**

The dataset that was used is called the MIO-TCD-Classification set [1]. The dataset originally contained 519194 training images, that were a part of 11 categories. These categories are different types of vehicles and modes of transportation, such as: 10,346 samples of an articulated truck, 5120 samples of a single-unit truck, 50,906 samples of a pickup truck, 10,316 samples of a bus, 260,518 samples of a car, 1982 samples of a motorcycle, 9679 samples of a work van, 2284 samples of a bicycle, 1751 samples of non-motorized vehicles, 6292 samples of pedestrian and 160, 000 samples of backgrounds.

However, we decided to select 2200 samples of each of the categories and selected all the samples for every category that contained less than 2200, in total we have 23533. This choice was made because it’s large enough to acquire enough details but not large enough to prolong runtime. Moreover, all the images were resized to dimensions of 128x128 pixels for them to all be uniformized as it helped in optimizing the code.

The extracted features were the gradients of the images and these features were extracted using the Histogram of Oriented Gradient (HoG) feature descriptor to train our SVM and K-Nearest-Neighbor (KNN). This feature was chosen because it’s amongst one of the most popular object detectors [3], it provides a compressed and encoded version of our images while also maintaining the general shape of the object. The HoG method was to investigate the gradients in several different directions in computing a histogram of the resulting gradient change. In order to detect precise edges within images, the feature extraction hyperparameters that were used were a cell size of 8x8 pixels, block size with 2x2 cells, and 8 angular directions (every 45° along a unit circle). Since our images are 128x128 pixels, our features size is 16x16x8 which is 2048 dimensions. These parameters were used because they provide sharp HoG features while keeping the size of the features low. We selected 8 directions because it generalizes every direction appropriately without the gradients being repetitive.

There were different types of vehicles, meaning the training images were not consistent, such as the backgrounds weren’t consistent, and the pixel intensities weren’t consistent. However, the general shapes remained consistent, for example, a bicycle resembles a bicycle but not the shape of a car. Hence, the HoG descriptor suited the needs for this criterion.

## SVM Implementation

Explain how this classifier (Support Vector Machine, SVM) was implemented and why we made those choices?

We used the Support Vector Machine (SVM) classifier. In brief, this algorithm takes labelled training data and outputs an optimal hyperplane separating classes [2]. The SVM was implemented using the scikit-learn library (machine learning library). Three important parameters to the SVM are the kernel type, gamma, and penalty parameter C.

The kernel type that was chosen was the Radial Basis Function (RBF) type, because in order to measure the similarity between two sets of features. RBF basis its distance between two features exponentially, which allows for quick computation. Initially we considered another kernel type, linear but its computation time took about two times more than the RBF kernel type which entails this increases the cost of doing validation. Also, since backgrounds contain a lot of noise, our dataset is not distributed linearly.

The gamma parameter was chosen

## Part 2.1 - Classifier Evaluation (Cross-validation)

In order to evaluate our classifiers, we used cross-validation (describe the cross-validation method **-10pts** and how we performed cross-validation) …

To evaluate the performance, the following metrics were obtained **(15pts)**:

* **(5pts)** Average classification accuracy across validations, including standard deviation:
* **(5pts)** Average precision and recall across validations: Are these values consistent with accuracy? These values are consistent with accuracy because … **AND** are they more representative of the dataset? Moreover, they are more representative of the dataset because … **AND** In what situations would you expect precision and recall to be a better reflection of model performance than accuracy? Finally, in order for precision and recall to be a better reflection of model performance than accuracy, it would mean that …
* **(5pts)** A confusion matrix on a validation set can be seen in Fig. 1 (plot matrix as an image, make a confusion matrix – maybe explain it a little bit 1-2 sentences?). **AND** are any of the classes difficult for your classifier?



1. *Confusion Matrix of Validation Set.*

The confusion matrix demonstrates the accuracy and recall of the SVM validation on the training data. The y-axis represents the actual label of the training image, and the x-axis represents the label that we predict using our SVM. The values along the diagonal are probabilities, describing the chance for the SVM to correctly classify the image label.

Observing the results in the confusion matrix, we can see that the SVM has difficulties identifying non-motorized vehicles, as it has the lowest recall value of correct identification, at 0.511.

**Include well-documented code (10pts)**. Moreover, to understand our classification approach practically, we included a well-documented code along with this report.

# Localization

## Part 3 – Localizer Implementation (Engineering Decisions)

In this part, we generated bounding boxes for the previous objects of interest. Describe the contents of the dataset (number of samples and bounding box size for each label, contents – **5pts**).

What localizer did we use (describe the localization method – **10pts**), why and how? **AND** describe the method from the input images to the set of output bounding boxes.

## Part 3.1 - Localizer Evaluation

We evaluated our localizer by computing the DICE coefficient for the predicted vs. true bounding boxes and when he had multiple boxes in one image, we matched the boxes that would maximize the mean DICE. **(5pts)** The distribution of DICE coefficients over our validation sets can be reported as follows:

* **(10pts)** Report the distribution of DICE coefficients over our validation sets.
* **(10pts)** In order to evaluate our classifier, we used the localization predicted by our localizer. The following metrics were obtained when comparing with our localizer and classifier vs. the classification data and classifier:
* The accuracy of our localization and classifier was: …, whereas the accuracy of the classification data and classifier was: …
* The prediction of our localization and classifier was: …, whereas the prediction of the classification data and classifier was: …
* The recall of our localization and classifier was: …, whereas the recall of the classification data and classifier was: …

By analyzing our results, we can see that there is (or not) a difference between the accuracy, prediction and recall because … (why or why not?). Should the 'background' label of the classifier be included when evaluating the performance of the localizer, and why/why not?

We also used cross-validation (describe your cross-validation approach - **5pts**).

**Include well-documented code (5pts)**. Finally, to understand our localization approach practically, we included a well-documented code along with this report.

# Deep Learning (Bonus) – Part 4

* Schematic of architecture **(1pt)**
* Description of training **(2pts)**
* Evaluation of performance (as described in the relevant tasks’ section) **(1pt)**
* Description of validation **(3pts)**
* Comparison with the methods from Sections 2 and 3 **(1pt)**
* Code with a description of the environment **(2pts)**

# Conclusion

The following project allowed us to dive into machine learning by understanding how to train a program using a classification and localization algorithm. The initial part of this experiment was to train a support vector machine classifier (SVM) in order to classify given images to 11 categories. What did we find in classification, anything significant? Finally, we implemented a localization method, using \_\_\_ localizer. We were able to classify the images and localize the objects using bounding boxes. Our code is also included with the report for reference.

References

1. MIO-TCD. (2018). Retrieved from http://podoce.dinf.usherbrooke.ca/challenge/dataset/
2. Patel, S. (2018). Chapter 2 : SVM (Support Vector Machine) — Theory – Machine Learning 101 – Medium. Retrieved from https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72
3. Dalal, N. and Triggs, B., “Histograms of Oriented Gradients for Human Detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, San Diego, CA, USA.