UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

Classical Method on New Datasets: Pedestrian Detection with HOG and CNN

Asher Mai, Matthew Tang, Sid Ahuja

Dec 6th, 2023

Table of Contents

Problem Statement

Detect whether a pedestrian is present in an image

1. Train a classifier on RGB images pixels.
2. Train a classifier with Histogram of Gradients (HoG) Feature Descriptors
3. Compare accuracy, precision, recall, and training time.

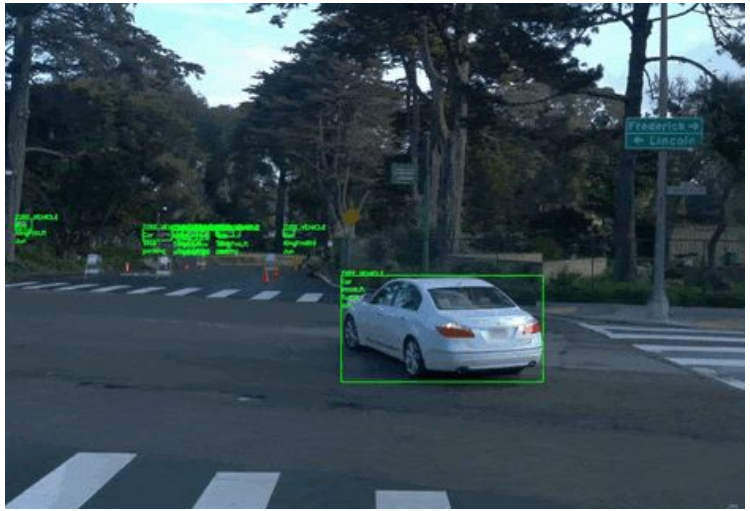
True (Pedestrian Present)

False (Pedestrian Absent)

Dataset

Waymo Road++ Dataset

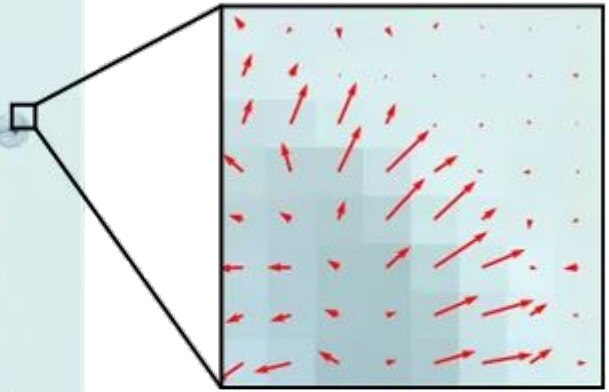
- Released at ICCV 2023, built on top of The Oxford RobotCar Dataset released in 2016.
- 198K frames from 1000 videos annotated
- Each video is 20 seconds long on an average
- Agent type labels, e.g. Pedestrian, Car, Cyclist, Large-Vehicle, Emergency-Vehicle etc.
- Due to hardware limitations, we used a subset of the dataset: 100 videos, which contain 18000 frames.



W. Maddern, G. Pascoe, C. Linegar and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset", The International Journal of Robotics Research (IJRR), 2016.
W. Maddern, G. Pascoe, M. Gadd, D. Barnes, B. Yeomans, and P. Newman, "Real-time Kinematic Ground Truth for the Oxford RobotCar Dataset", in arXiv preprint arXiv: 2002.10152, 2020.
https://sites.google.com/view/road-plus-plus/dataset

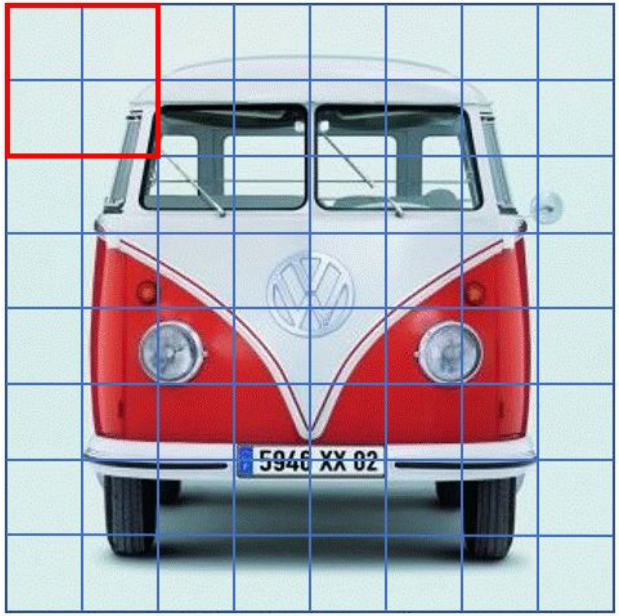
Brief Overview of HoG

1. Compute a gradient for each pixel with magnitude and direction



Nemutlu,Dahi. "HOG Feature Descriptor", Medium, Sep 9th 2022. https://medium.com/@dnemutlu/hog-feature-descriptor-263313c3b40d
N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

Brief Overview of HoG

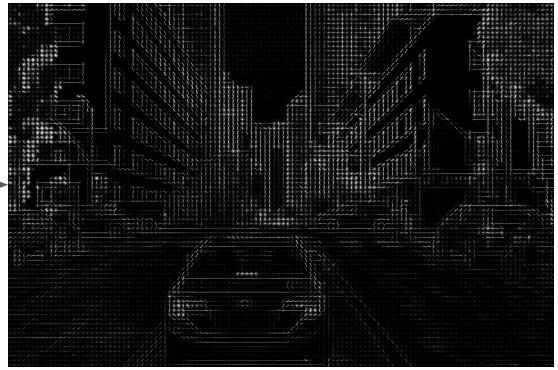


1. Compute a gradient for each pixel with magnitude and direction
2. Divide image into cells of 16x16 pixels.
3. Group the cells into 2x2 blocks.
4. For each cell, compute a Histogram of Gradients with 9 bins, each bin spam 20 degree. The sum of magnitudes in each bin will be the height.
5. Normalize cell histograms within each block.

The example on the left has 8x8 cells, which create 7x7 blocks.
The feature vector is of shape [7, 7, 2, 2, 9]

Nemutlu,Dahi. "HOG Feature Descriptor", Medium, Sep 9th 2022. https://medium.com/@dnemutlu/hog-feature-descriptor-263313c3b40d
N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

Methodology



HoG Visualization

RGB Image

ResNet18 → Has Pedestrian?

HoG Feature Descriptor
Shape: [23, 35, 2, 2, 9]

ResNet18 → Has Pedestrian?

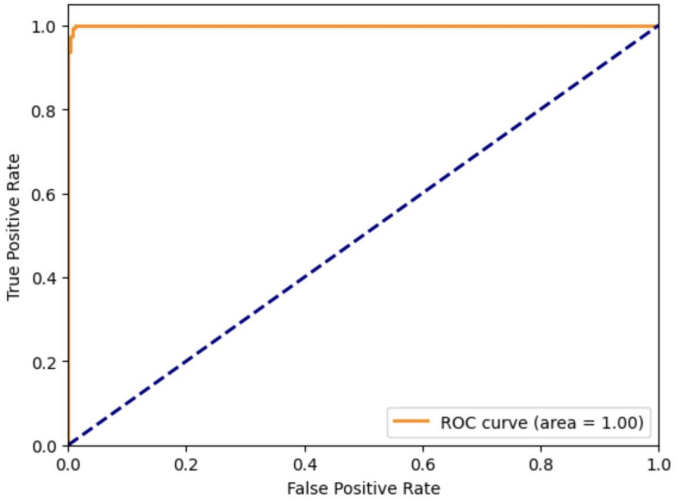
Fully-Connected Layer → Has Pedestrian?

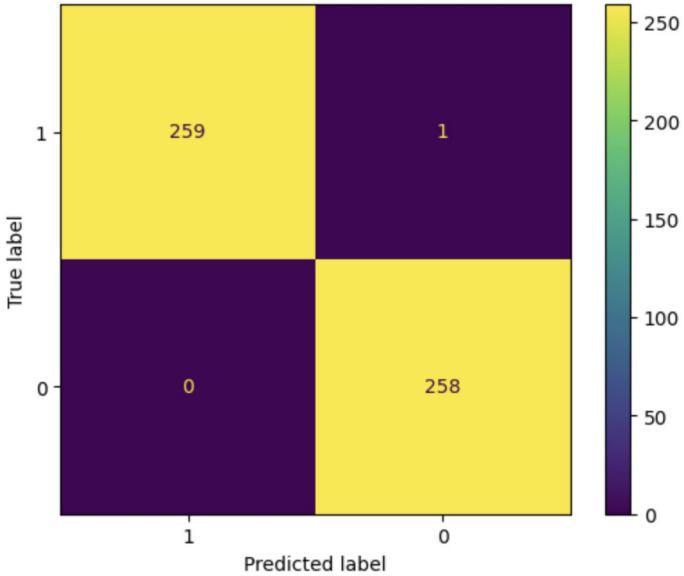
1-layer CNN → Has Pedestrian?

Results

Raw Image Resnet

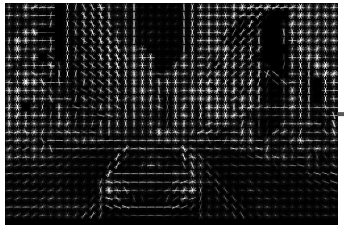
ResNet18



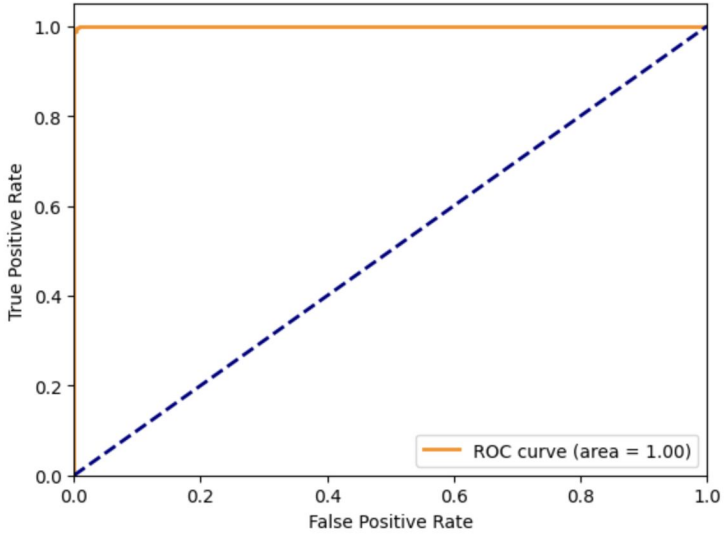


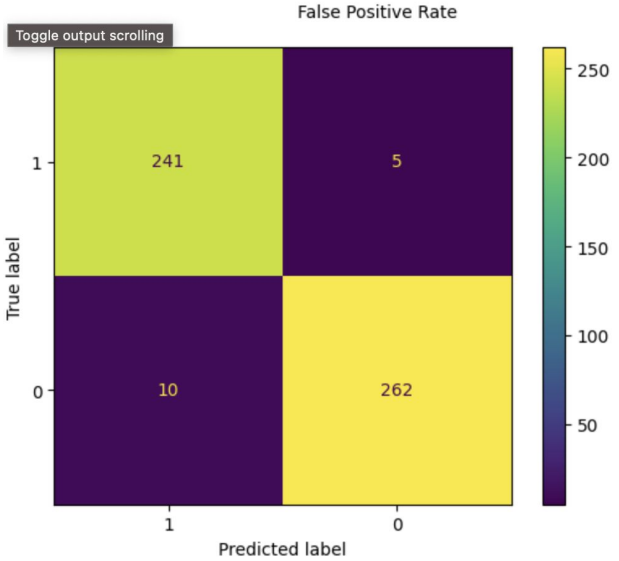
Results

HOG Resnet



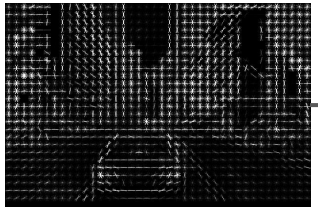
ResNet18



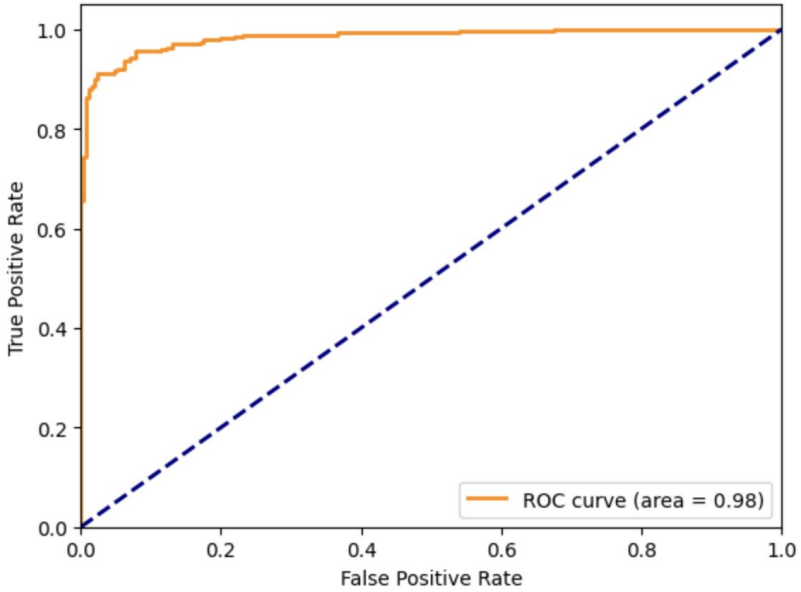


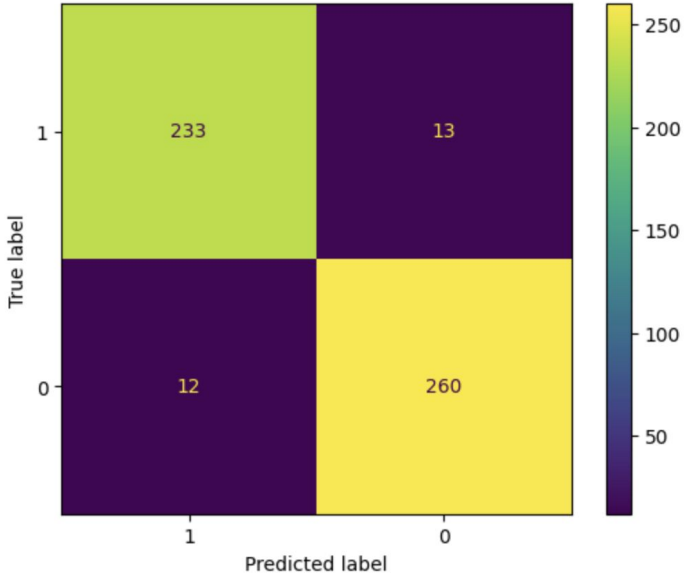
Results

HOG FC-layer Baseline



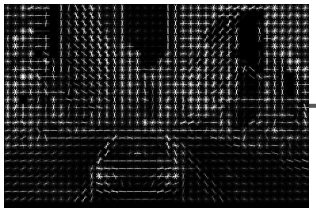
Fully-Connected Layer



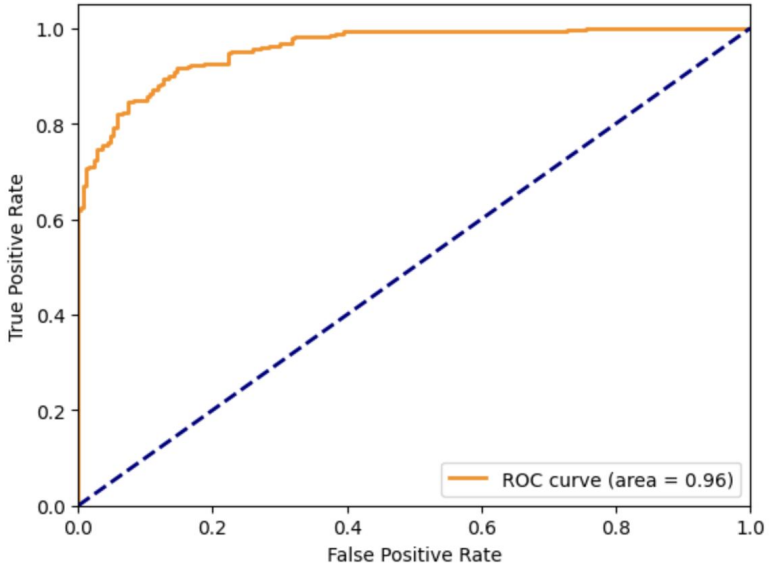


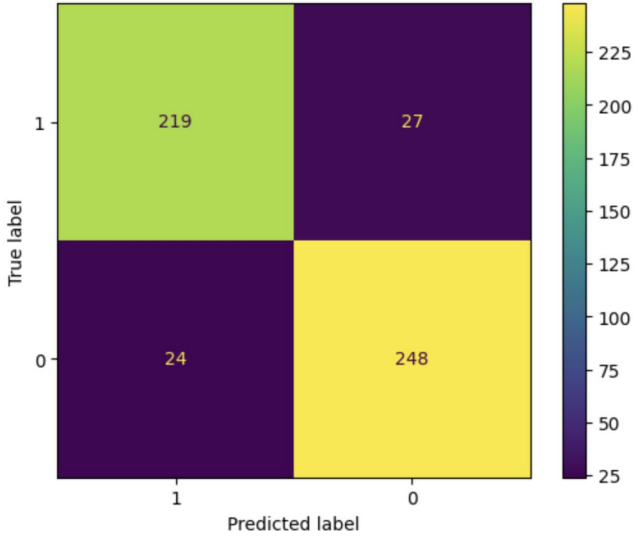
Results

HOG 1-layer Conv



1-layer CNN



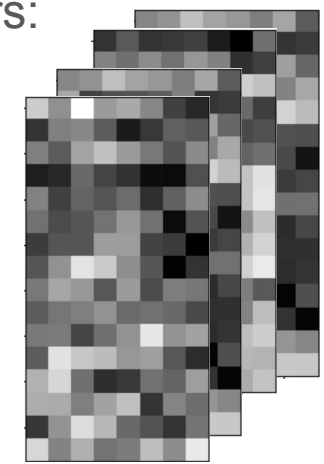


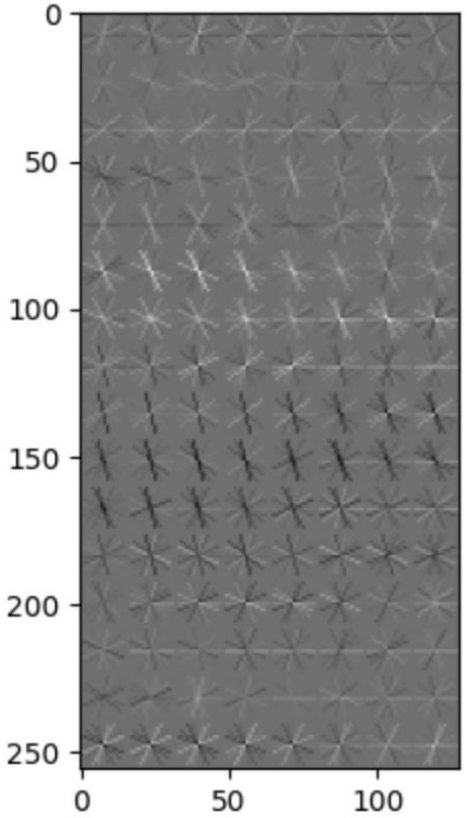
Results

HOG 1-layer Conv

Visualization of the pedestrian filter

9 Filters:





Model Comparison

	Training time (s) *	Precision	Recall	Accuracy (%)
RGB ResNet18	475.88	**1.0**	**0.9924**	**99.6**
HoG Resnet18	3.80	0.9846	0.9771	98.1
HoG Fully-Connected	**2.48**	0.9510	0.9472	93.4
HoG 1-layer Conv	3.42	0.9012	0.8902	90.15

* Training time for 1 epoch on a GPU

Conclusion

Key Takeaways

- We find that extracting HoG features on the dataset and training on those features can have large magnitudes of **training time reduction**.
- This method has **comparable accuracies** to training on raw RGB images.
- Extracting the HOG features only takes **113ms (8.83 fps)** on a CPU, which can be sped up even further with a GPU. This enables applications of this pedestrian detection method in **real time** systems such as autonomous vehicles.