Object Detection using HoG Features for Visual Situation Recognition

Efsun Sarioglu Kayi, Melanie Mitchell
Portland State University

August 31, 2017

Abstract

This project investigates a novel approach to building computer systems that can recognize visual situations. While much effort in computer vision has focused on identifying isolated objects in images, what people actually do is recognize coherent situations — collections of objects and their interrelations that, taken together, correspond to a known concept, such as "dog-walking", or "a fight breaking out", or "a blind person crossing the street". Situation recognition by humans may appear on the surface to be effortless, but it relies on a complex dynamic interplay among human abilities to perceive objects, systems of relationships among objects, and analogies with stored knowledge and memories. Enabling computers to flexibly recognize visual situations would create a flood of important applications in fields as diverse as autonomous vehicles, medical diagnosis, interpretation of scientific imagery, enhanced human-computer interaction, and personal information organization.

1 Introduction

Two previously studied approaches, brain-inspired neural networks for lower-level vision and cognitive-level models of concepts and analogy-making can be integrated for recognizing situations. System architecture is depicted in Figure 1. For Dog-Walking situation, there are three objects: Person, Dog and Leash. The relationships between them can be analyzed through Concept Network. A concept network for Dog-Walking situation is given in Figure 2. In this project, we focus on lower-level vision for efficient and effective object detection which would feed into cognitive-models for visual situation recognition.

2 Dataset

Dog-walking dataset consists of photographs, each of which is an instance of the visual concept (or situation) "Dog-Walking". The dog-walkers, dogs, and leashes labeled with ground truth bounding boxes. The number of files for each class is listed in Table 3.

2.1 Preprocessing

Label file of each image has the location of objects in that image. The negative crops are randomly selected from the remaining parts of the image with similar size of the positive crop. Cropped images are resized to a fixed width and height to make them comparable. For each class, 10%, 20%, 30%, 40%, 50% percentiles of width and height are used as resize options.

3 Experiments

Histograms of Oriented Gradients (HOG) [DT05] are edge orientation histograms computed on a dense grid of uniformly spaced cells. They are computed using MATLAB's extractHOGFeatures function on overlapping local contrast normalizations for improved performance. Figure 4 shows an image and its corresponding HOG features. Tables 5,6,7 show number of HOG features for different resize options for Dog, Leash and Person classes respectively. Tables 5, 6, and 7 lists resize options and the corresponding feature numbers for Dog, Leash and Person classes respectively.

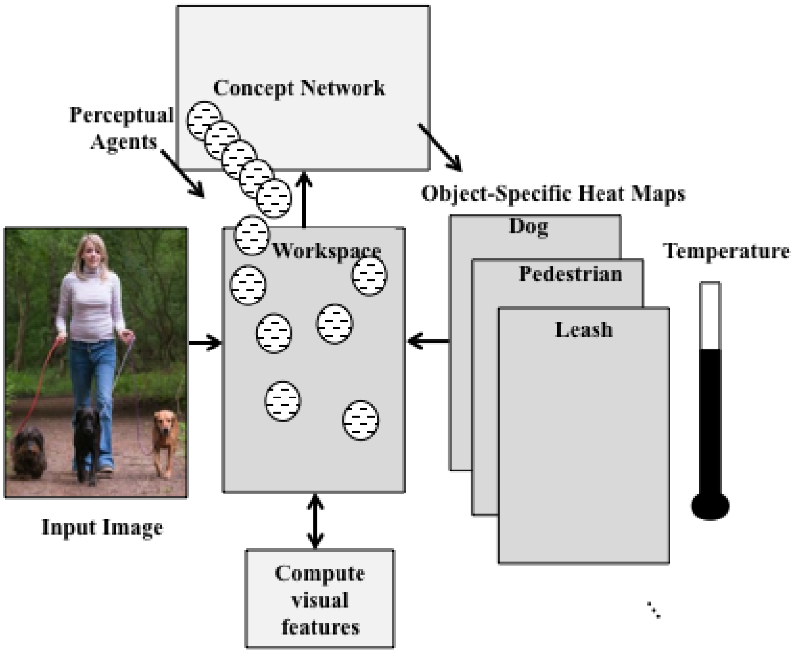

Figure 1: Situate Architecture

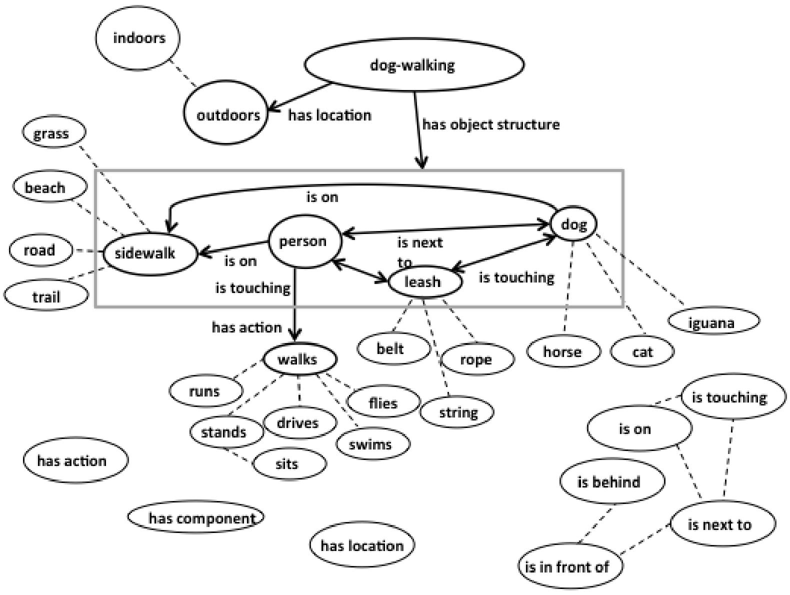

Figure 2: Concept Network

	Dog	Leash	Person
Positive	584	584	558
Negative	584	584	583
Total	1,168	1,168	1,141

Figure 3: Dog-Walking dataset

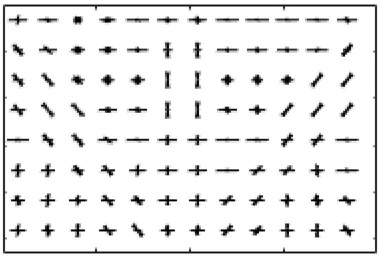

Figure 4: HOG Features

Resize	Width	Height	Features
10%	72	82	2,592
20%	96	106	4,752
30%	123	127	7,056
40%	143	155	10,368
50%	221	221	24,336

Figure 5: HOG features for Dog class

Resize	Width	Height	Features
10%	46	58	864
20%	65	85	2,268
30%	89	112	4,680
40%	112	132	7020
50%	209	193	20,700

Figure 6: HOG features for Leash class

Resize	Width	Height	Features
10%	102	145	6,732
20%	137	221	14,976
30%	173	272	23,760
40%	211	355	38,700
50%	331	519	90,720

Figure 7: HOG features for Person class

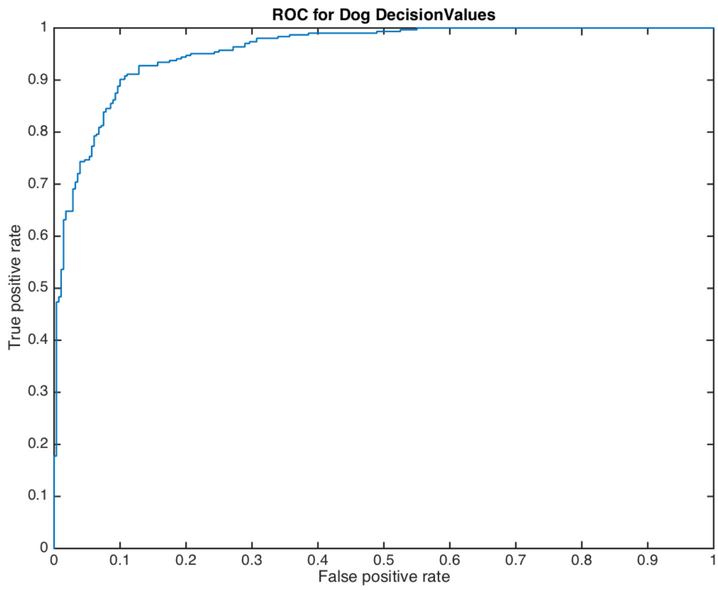

Figure 8: ROC plot for Dog class

Library for Support Vector Machines (LIBSVM) MATLAB extension is used [CL11]. Using the images resized to their average width and height, SVM is trained using 50% of the data and tested on the remaining 50%. The ROC graphs using decision values are displayed in Figures 8, 9 and 10 corresponding to Dog, Leash and Person classes. Corresponding AUC values are listed in Table 11.

3.1 SVM Feature Selection

Once a classifier is built using SVM, its features can be selected based on their weights w (Recursive Feature Elimination (RFE) [GWBV02]. Since number of features is high, instead of removing features one by one in a recursive fashion, its top features were selected based on w^2. A new classifier was trained using the selected features and ROC and AUC results were computed for the test dataset. Using 50% training and 50% testing data and resized images of size 10%. Corresponding AUC values are displayed in Figures 12, 13 and 14 and listed in Table 15. ROC plots with feature selections are shown in Figures 16, 17, and 18.

3.2 Effect of HOG Cell Size

HOG features are computed on given cell sizes in pixels. The default size is 8×8. We also explored 2×2 and 4×4 cell sizes to capture small-scale details. In addition, we evaluated combination of them as below:

- 2×2 + 4×4
- 2×2 + 8×8

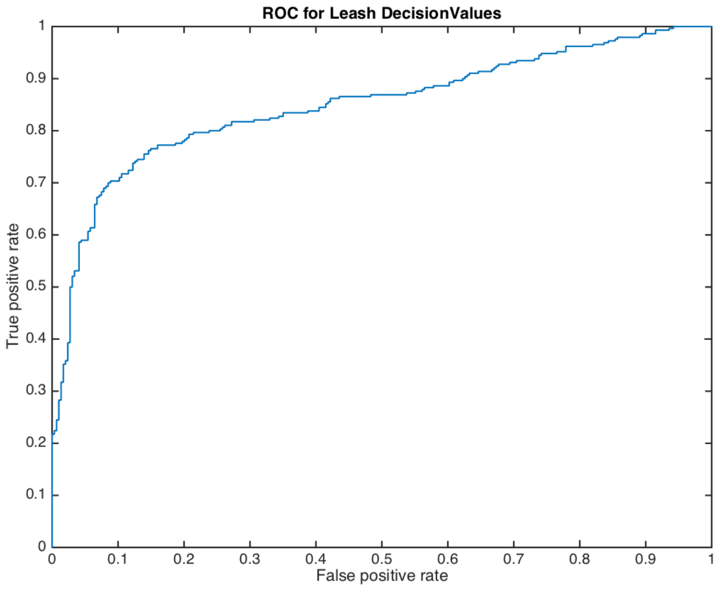

Figure 9: ROC plot for Leash class

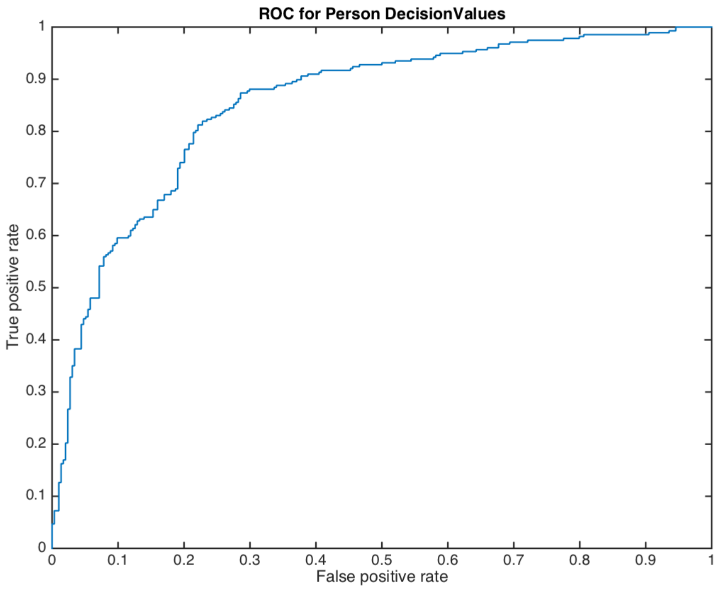

Figure 10: ROC plot for Person class

Class	10p	20p	30p	40p	50p
Dog	96.76%	96.99%	95.79%	96.56%	95.72%
Leash	81.90%	78.33%	81.73%	81.54%	84.89%
Person	84.08%	85.09%	84.92%	84.19%	85.28%

Figure 11: AUC values

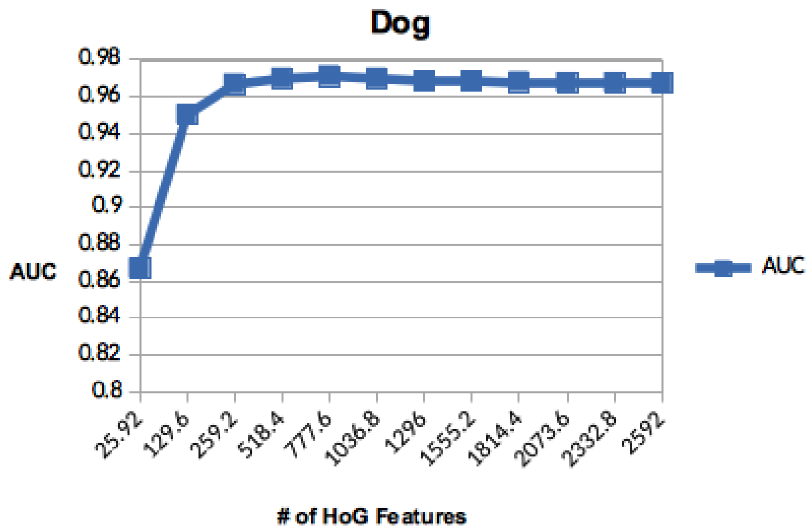

Figure 12: AUC values for Dog class with different number of features

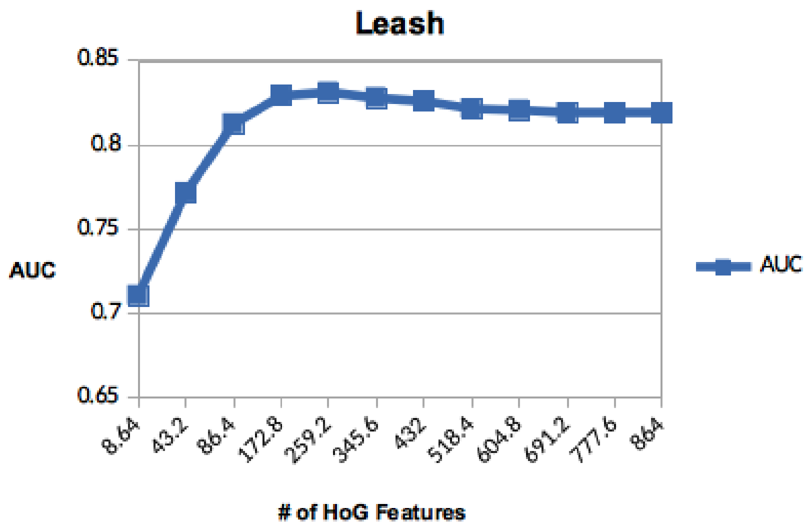

Figure 13: AUC values for Leash class with different number of features

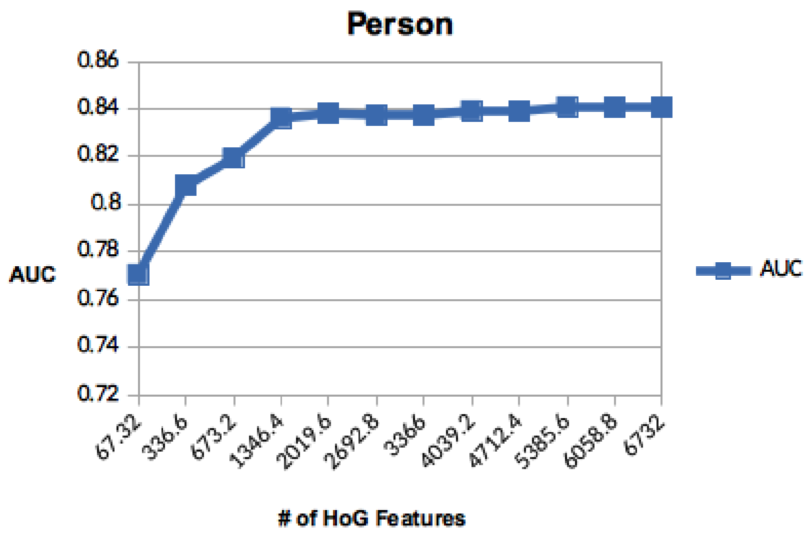

Figure 14: AUC values for Person class with different number of features

AUC	Top Selected Features (%)											Total # of features
	1	5	10	20	30	40	50	60	70	80	90	100
Dog	0.87	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	2,592
Leash	0.71	0.77	0.81	0.83	0.83	0.83	0.83	0.82	0.82	0.82	0.82	0.82	864
Person	0.77	0.81	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	6,732

Figure 15: AUC values with feature selection

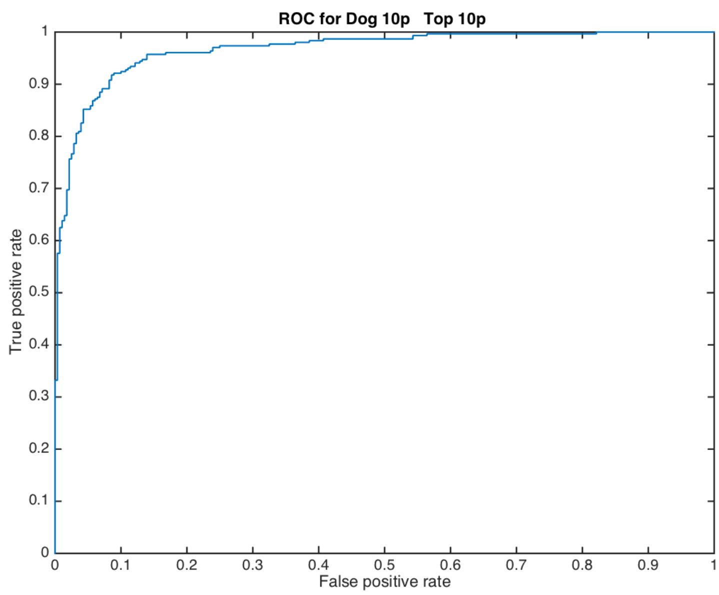

Figure 16: ROC graph for Dog class using top 20% features

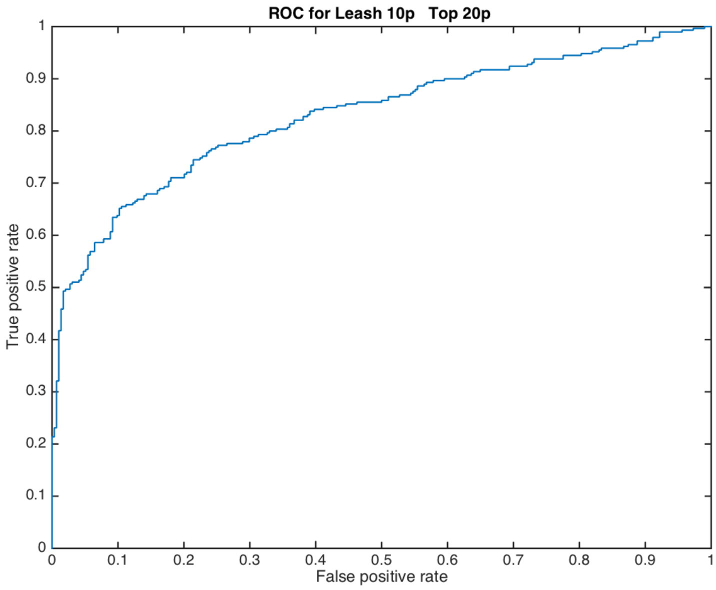

Figure 17: ROC graph for Leash class using top 20% features

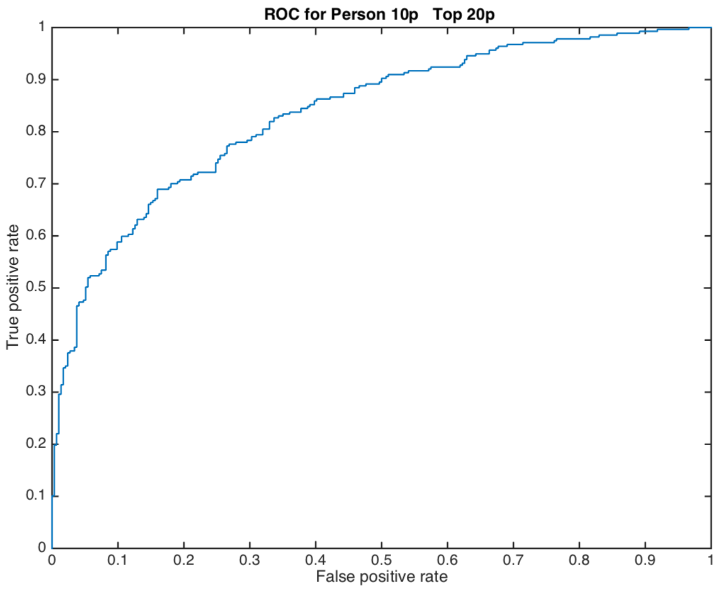

Figure 18: ROC graph for Person class using top 20% features

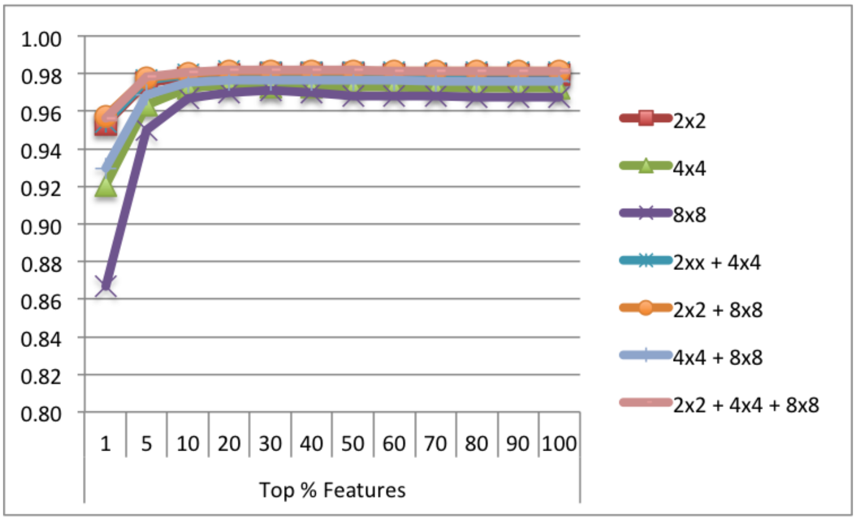

Figure 19: AUC values using top HOG features with different cell sizes for Dog Class

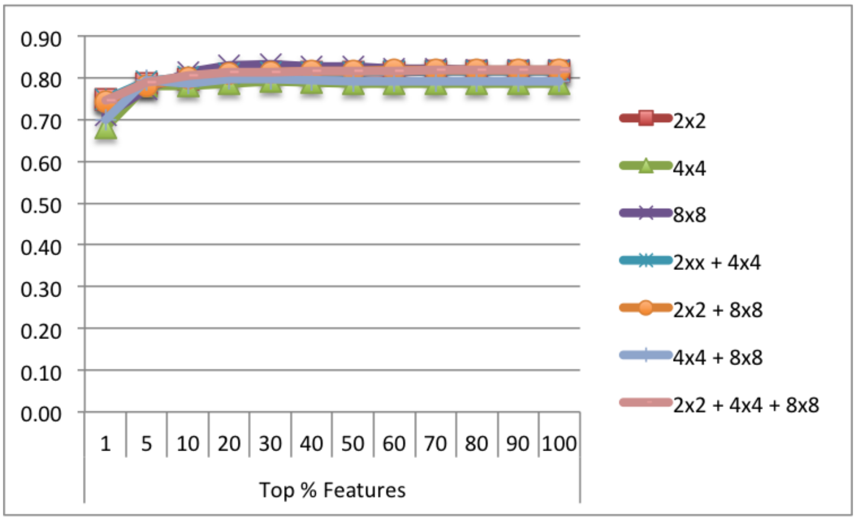

Figure 20: AUC values using top HOG features with different cell sizes for Leash Class

- $4{\times}4 + 8{\times}8$
- $2{\times}2 + 4{\times}4 + 8{\times}8$

Smaller cell sizes also correspond to bigger feature sizes. Accordingly, we used the same SVM feature selection mechanism explained in Section 3.1. Using 50% training and 50% testing data with images resized to 10% of their class sizes, SVM was trained with HOG features with different cell sizes. Figures 19, 20, and 21 show the AUC values for Dog, Leash and Person classes for different cell sizes and top selected features. Tables 22, 23, and 24 lists the AUC values for Dog, Leash and Person classes using different cell sizes and top selected features.

4 Conclusion

We have shown effective and efficient object detection using HOG features and SVM for Dog-Walking situation. Using SVM feature selection techniques, top 10 to 20% of top features performed as well as using the entire dataset. However, using different cell sizes for HOG did not make much difference in classification performance.

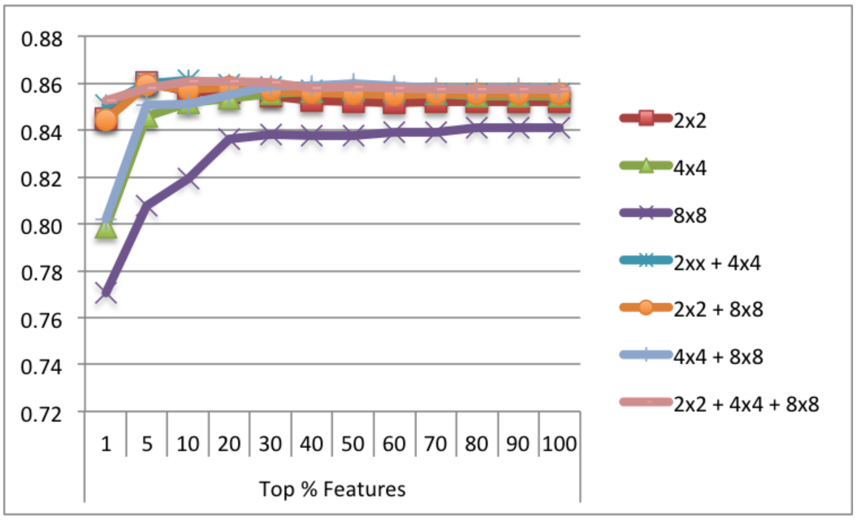

Figure 21: AUC values using top HOG features with different cell sizes for Person Class

Dog	Top % Features											
AUC	1	5	10	20	30	40	50	60	70	80	90	100
2x2	0.95	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
4x4	0.92	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
8x8	0.87	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
2xx + 4x4	0.95	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
2x2 + 8x8	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
4x4 + 8x8	0.93	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
2x2 + 4x4 + 8x8	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Figure 22: AUC values for Dog class for different cell sizes and top selected features

Leash	Top % Features											
AUC	1	5	10	20	30	40	50	60	70	80	90	100
2x2	0.75	0.79	0.80	0.81	0.81	0.81	0.81	0.81	0.82	0.82	0.82	0.82
4x4	0.68	0.78	0.78	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
8x8	0.71	0.77	0.81	0.83	0.83	0.83	0.83	0.82	0.82	0.82	0.82	0.82
2xx + 4x4	0.75	0.79	0.80	0.81	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.82
2x2 + 8x8	0.74	0.78	0.80	0.81	0.81	0.82	0.82	0.82	0.82	0.82	0.82	0.82
4x4 + 8x8	0.70	0.79	0.79	0.80	0.80	0.80	0.79	0.79	0.79	0.79	0.79	0.79
2x2 + 4x4 + 8x8	0.75	0.79	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.82	0.82	0.82

Figure 23: AUC values for Leash class for different cell sizes and top selected features

	Top % Features											
AUC	1	5	10	20	30	40	50	60	70	80	90	100
2x2	0.84	0.86	0.86	0.86	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
4x4	0.80	0.85	0.85	0.85	0.86	0.86	0.86	0.86	0.86	0.85	0.85	0.85
8x8	0.77	0.81	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
2xx + 4x4	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
2x2 + 8x8	0.84	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
4x4 + 8x8	0.80	0.85	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
2x2 + 4x4 + 8x8	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86

Figure 24: AUC values for Person class for different cell sizes and top selected features

References

[CL11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 886–893, Washington, DC, USA, 2005. IEEE Computer Society.

[GWBV02] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1-3):389–422, March 2002.